Radio Resource Management Techniques for Multi-tier Cellular Wireless Networks

by

Amr Adel Nasr Abdelnasser

A Thesis submitted to The Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

December 2015

Abstract

There is a prolific increase in the penetration of user devices such as smartphones and tablets. In addition, user expectations for higher Quality of Service (QoS), enhanced data rates and lower latencies are relentless. In this context, network densification through the dense deployment of small cell networks, underlaying the currently existing macrocell networks, is the most appealing approach to handle the aforementioned requirements. Small cell networks are capable of reusing the spectrum locally and providing most of the capacity while macrocell networks provide a blanket coverage for mobile user equipment (UEs). However, such setup imposes a lot of issues, among which, co-tier and cross-tier interference are the most challenging.

To handle co-tier interference, I have proposed a semi-distributed (hierarchical) interference management scheme based on joint clustering and resource allocation (RA) for small cells. I have formulated the problem as a Mixed Integer Non-Linear Program (MINLP), whose solution was obtained by dividing the problem into two subproblems, where the related tasks were shared between the Femto Gateway (FGW) and small cells. As for cross-tier interference, I have formulated RA problems for both the macrocell and small cells as optimization problems. In particular, I have introduced the idea of "Tier-Awareness" and studied the impact of the different RA policies in the macrocell tier on the small cells performance. I have shown that the RA policy in one tier should be carefully selected. In addition, I have formulated the RA problem for small cells as an optimization problem with an objective function that accounts for both RA and admission control (AC). Finally, I have studied cloud radio access network (C-RAN) of small cells which has been considered as a typical

realization of a mobile network which is capable of supporting soft and green technologies in Fifth Generation (5G) networks, as well as a platform for the practical implementation of network multiple-input multiple-output (MIMO) and coordinated multi-point (CoMP) transmission concepts.

Acknowledgements

In the name of "ALLAH" the Most Gracious, the Most Merciful. To him we belong and all the praise and thanks are due. After thanking Almighty "ALLAH" for his blessing and guidance to complete this work, I would like to express my sincere gratitude to my advisor Prof. Ekram Hossain for his continuous support to me during my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

I would like also to thank my thesis committee: Prof. Jun Cai, Prof. Rasit Eskicioglu, and Prof. Xianbin Wang for their insightful comments and encouragement.

I would like to express my gratitude to Prof. Attahiru Alfa for his interesting discussions and sincere advice. I would like also to thank the Faculty of Graduate studies for supporting me with the University of Manitoba Graduate Fellowship (UMGF) and a number of other scholarships. I would like to thank the staff at the Department of Electrical and Computer Engineering. Special thanks to Amy Dario for her kind help. I would like to thank my colleagues and fellow researchers for the stimulating discussions and for all the fun we have had during my PhD study.

On the personal side, special thanks and gratitude to my beloved parents for their continuous support and encouragement throughout my career. I just can not thank them enough for every single moment and effort they have invested on my path to earn this degree. I would like also to express my appreciation and gratefulness to my beloved wife for her tireless support, understanding, patience, and countless sacrifices since I started this program. Last but not least, I would like also to express my gratitude to my little kids for all the joy and innocence that they have brought to

my life. I would not deny the great influence those little angels have had on my enthusiasm and motivation to attain such an honor that would make them proud once they are old enough to read and comprehend these lines.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5G	Fifth generation
AC	Admission control
AWGN	Additive white Gaussian noise
BBU	Base band unit
BS	Base station
CAPEX	Capital expenditure
CH	Cluster head
C-RAN	Cloud radio access network
CSG	Closed subscriber group
CSI	Channel state information
CoMP	Coordinated multi-point transmission
D2D	Device to device
eNB	Evolved node B
FAP	Femto access point
FGW	Femto gateway
FUE	Femto user equipment
HeNB GW	Home eNB gateway
HetNets	Heterogeneous networks
ILP	Integer linear program

IQ	In-phase and quadrature
KKT	KarushKuhnTucker
MBS	Macro base station
MCC	Mobile cloud computing
MIMO	Multiple-input multiple-output
MINLP	Mixed integer nonlinear program
MUE	Macro user equipment
OFDMA	Orthogonal frequency-division multiple access
OPEX	Operating expenditure
QoS	Quality-of-service
RA	Resource allocation
RAN	Radio access network
RF	Radio frequency
RRH	Remote radio head
SCA	Successive convex approximation
SDP	Semi-definite program
SINR	Signal to interference plus noise ratio
SOCP	Second order cone program
SUE	Small cell user equipment
UE	User equipment

List of Symbols

$\mathcal{A}$	The set of available sub-channels for allocation
B	Index of the macrocell
$\mathbb{C}^{r\times t}$	The set of all $r \times t$ matrices with complex entries
$\mathcal{C}$	The set of clusters of femtocells
c_l	The l^{th} set of femtocells
$C_{s,max}$	Maximum fronthaul capacity available at a cell s
$d_{i,j}$	Distance between cell i and UE j (km)
D	Large positive constant
$\mathbf{D}$	Eigen values matrix
$\boldsymbol{d}$	Sub-gradient
d^n	Element of the sub-gradient $\boldsymbol{d}$
$\mathcal{E}$	The set of edges between each two pair of vertices
$\mathcal{F}$	Set of all SUEs
F	Number of SUEs
f	Index of an SUE
$\mathcal{F}_s$	Set of all SUEs served by small cell s
$g^n_{i,j}$	Channel gain of the link between UE j served by BS i on sub-channel n
g	Lagrange dual function

g_s	Lagrange dual function for small cell s
$g^n_{k_i,j}$	The channel gain between user k_i and base station j on sub-channel n
$\mathbf{h}^n_{i,j}$	The complex channel gain between small cell i and UE j on sub-channel n
$H^n_{i,j}$	The fast fading coefficient between cell i and UE j on sub-channel n
$\mathbf{h}^n_{\mathcal{S}_f,i}$	The channel vector between a UE i and the set of RRHs $\mathcal{S}_f$, serving SUE f, on sub-channel n
I^n_m	Maximum tolerable interference level on sub-channel n allocated to MUE m
I_{max}	Upper limit on the maximum tolerable interference level I^n_m
I_{th}	Equal value for the maximum tolerable interference level I^n_m , $\forall m \in \mathcal{M}, n \in \mathcal{N}$
$I_{th,L}$ and $I_{th,H}$	Lower and upper limits on I_{th} in the bisection method
$I_{th,M}$	Mean of $I_{th,L}$ and $I_{th,H}$
k_s	UE belonging to a small cell s
k_B	A UE belonging to macrocell B
$\mathcal{L}$	Lagrangian function
$l_{i,j}$	The path-loss between cell i and UE j
$\mathcal{M}$	Set of all MUEs
M	Number of MUEs
m	Index of an MUE
$\mathcal{M}_{cand}$	The set of candidate MUEs for sub-channel allocation
$\mathcal{N}$	Set of all available sub-channels
N	Number of sub-channels
n	Index of a sub-channel
$N_{a,c}$	Number of allocated sub-channels
$\mathcal{N}_m$	Set of all sub-channels allocated to MUE m

$\mathcal{N}_j$	The set of sub-channels allocated to UE j
N_o	Noise power
$P_{i,j}^n$	Power allocated to the link between UE j served by BS i on sub-channel n
$\tilde{P}_{i,j}^n$	Actual power allocated to the link between UE j served by BS i on sub-channel n
$P_{B,max}$	Total macrocell power
$P_{s,max}$	Total small cell power
$P_{k_i,i}^n$	The power assigned to the link between user k_i and base station i on sub-channel n
q_i	The maximum number of allocated sub-channels for a UE i
$\mathbf{Q}$	Eigen vectors matrix
$\mathbf{r}_i$	Unit norm random vectors
R_B	Coverage radius of macrocell B
R_m	Data rate requirement of MUE m
R_f	Data rate requirement of SUE f
$R_{ach,m}$	The achieved data rate for MUE m
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{N\times N}$	Set of real $N \times N$ matrices
$\mathbb{R}_+^{N\times N}$	Set of non-negative real $N \times N$ matrices
$\mathbb{R}^N$	An N-dimensional real-valued column vector
$\mathbb{R}_+^N$	An N-dimensional real-valued non-negative column vector
$\mathcal{S}$	Set of small cells/femtocells
S	Number of small cells/femtocells
s	Index of a small cell/femtocell

$\mathbb{S}_{+}^{n\times n}$	Set of non-negative symmetric real matrices
u_j^n	The transmitted information symbol with unit variance for UE j on sub-channel n
$\mathcal{V}$	The set of vertices
$w_{i,j}^{+}$	The degree to which two vertices i and j are similar
$w_{i,j}^{-}$	The degree to which two vertices i and j are different
$\mathbf{W}^{+}$	Matrix of the similarity weights
$\mathbf{w}^{-}$	Vector of the difference weights
$\mathbf{w}_{i,j}^n$	The precoding vector at small cell i corresponding to the signal transmitted to SUE j on sub-channel n
$\mathbf{w}_{\mathcal{S}_f,f}^n$	The precoding vector at the set of RRHs $\mathcal{S}_f$ corresponding to the signal transmitted to SUE f on sub-channel n
w_L^{-} and w_H^{-}	Lower and upper bound values on the range of weights
$x_{i,j}$	A binary decision variable to indicate if vertices i and j are in the same cluster
$y_{s,f}$	Admission control variable for SUE f served by small cell s
y_f	Admission control variable for SUE f
Z_{OPT}	Optimal objective function value
Z_{RR}	Objective function value after randomization
z	Iteration number
$[x]^{+}$	$\max(x, 0)$
$\lfloor x \rfloor$	The nearest integer less than or equal to x
$\lvert\mathbf{Y}\rvert$	The cardinality of the set $\mathbf{Y}$
$[.]^T$	The transpose operation
$[.]_{i,j}$	The element in row i and column j of a matrix

$\mathbf{x}^H$	Hermitian transpose
$\|\mathbf{x}\|$	Euclidean norm
$\text{diag}(\mathbf{x})$	the diagonal matrix constructed from the elements of the vector $\mathbf{x}$
$\mathbf{1}_{r\times t}$	the matrix of all ones whose dimensions are $r \times t$
$\mathbf{0}_{r\times t}$	the matrix of all zeros whose dimensions are $r \times t$
$\|\theta\|$	The absolute value of the complex scalar θ
α	Lagrange multiplier corresponding to the data rate constraint
α^n	Scale factor for small cells transmission powers on sub-channel n
$\gamma^n_{B,m}$	Received SINR of an MUE m served by macrocell B on sub-channel n
$\gamma^n_{s,f}$	Received SINR of an SUE f served by small cell s on sub-channel n
$\gamma^n_{k_s,s}$	The unit power SINR of an FUE k_s served by femtocell s on sub-channel n
$\Gamma^n_{i,j}$	Sub-channel allocation indicator for sub-channel n allocated to UE j served by BS i
Γ^n_j	The sub-channel allocation indicator for sub-channel n and UE j
$\Gamma^n_{k_s,s}$	Sub-channel allocation indicator for sub-channel n allocated to the link between user k_s and femtocell s
Δf	Sub-channel bandwidth
δ	Termination tolerance in bisection method
$\boldsymbol{\eta}$	Lagrange multiplier associated with the cross-tier interference
η^n	Element of the Lagrange multiplier $\boldsymbol{\eta}$
ϵ	Weighting factor
λ	Lagrange multiplier corresponding to the total power constraint
μ^n_f	Auxiliary SINR variable for SUE f on sub-channel n
$\Psi_{st}(x)$	Unit step function
ω^n_j	The additive white Gaussian noise (AWGN) in sub-channel

	n at UE j with zero mean and variance N_o
Ω	Number of antennas at a cell
υ	Maximum cluster size
$\vartheta_{i,j}^{n}$	The received signal for a UE j served by cell i on sub-channel n
$\xi_{k_B}^{n}$	Interference constraint for MUE k_B served by macrocell B on sub-channel n
$\xi_{k_j}^{n}$	Interference constraint for FUE k_j served by femtocells j on sub-channel n
ζ_{f}^{n}	Auxiliary variable

Publications

- Journal Publications:

1. **A. Abdelnasser** and E. Hossain, "Resource allocation for an OFDMA Cloud-RAN of small cells underlaying a macrocell," *IEEE Transactions on Mobile Computing*, to appear.
2. **A. Abdelnasser**, E. Hossain, and D. I. Kim, "Tier-aware resource allocation in OFDMA macrocell-small cell networks," *IEEE Transactions on Communications* , vol. 63, no. 3, pp. 695-710, Mar. 2015.
3. **A. Abdelnasser**, E. Hossain, and D. I. Kim, "Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1628-1641, Mar. 2014.
4. E. Hossain, M. Rasti, H. Tabassum, and **A. Abdelnasser**, "Evolution toward 5G multi-tier cellular wireless networks: an interference management perspective," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 118-127, 2014.

- Conference Publications:

1. **A. Abdelnasser** and E. Hossain, "On resource allocation for downlink power minimization in OFDMA small cells in a cloud-RAN," to be presented in *IEEE Global Communications Conference (Globecom'15)*, San Diego, CA, USA, 6-10 Dec., 2015.
2. **A. Abdelnasser** and E. Hossain, "Two-tier OFDMA cellular cloud-RAN: Joint resource allocation and admission control," to be presented in *IEEE Global Communications Conference (Globecom'15)*, San Diego, CA, USA, 6-10 Dec., 2015.
3. **A. Abdelnasser** and E. Hossain, "Joint resource allocation and admission control in OFDMA-based multi-tier cellular networks," in Proc. of *IEEE Global Communications Conference (Globecom'14)*, Austin, TX, USA, pp. 4689-4694, 8-12 Dec., 2014.
4. **A. Abdelnasser** and E. Hossain, "Joint subchannel and power allocation in two-tier OFDMA HetNets with clustered femtocells," in Proc. of *IEEE International Conference on Communications (ICC'13)*, Budapest, Hungary, pp. 6002-6007, 9-13 June 2013.
5. **A. Abdelnasser** and E. Hossain, "Subchannel and power allocation schemes for clustered femtocells in two-tier OFDMA HetNets," in Proc. of *IEEE International Conference on Communications Workshops (ICC'13)*, Budapest, Hungary, pp. 1129-1133, 9-13 June 2013.

Chapter 1

Introduction

Mobile wireless communication has experienced exponential growth over the past decade, pushed by the abundance and the proliferation of smart devices. There is a huge consensus in the wireless industry that this trend is anticipated to continue. Moreover, the wireless industry is preparing itself for a 1000-fold increase in traffic demand, by the year 2020 [1]. Hence, efficient measures are required to meet this remarkable growth in mobile traffic demand. Conventional homogeneous wireless networks are not expected to be able to handle such requirements. In this context, the concept of heterogeneous wireless networks (HetNets) with base station (BS) densification has been introduced. A HetNet uses a mix of macrocells and high density of deployed low power small cells in order to bring the network closer to end-users. This way, radio link quality can be enhanced owing to the reduced distance between the transmitter and the receiver, and the larger number of cells allows for more efficient spectrum reuse and, therefore, larger data rates. HetNets with multi-tier deployments pose critical challenges to the resource allocation (RA) process.

The target of this thesis is to investigate issues related to RA in multi-tier HetNets. More specifically, clustering and coordination/cooperation between small cells are

studied as well as the impact of different RA decisions of one tier on the other tier performance.

In the following sections I provide a background on HetNets with small cell deployments. I discuss the characteristics of small cells and the challenges in their deployment. Then, I outline the RA process in HetNets with small cell deployments. Finally, I summarize the main contributions of this thesis.

1.1 Multi-tier HetNets

1.1.1 HetNet Architecture

A HetNet is a network consisting of a variety of BSs with, possibly, different wireless access technologies, each of them having different characteristics in terms of transmit power, coverage radius, and backhaul connection. As shown in Fig. 1.1, a typical HetNet involves a macrocell overlaid by densely deployed low power BSs as picocells, small cells (femtocells), as well as relay stations and device-to-device (D2D) connections. These low-power BSs can be either operator deployed or user deployed, and may coexist in the same proximity as the macrocell, potentially sharing the same spectrum i.e. using a co-channel deployment. Deploying such low power BSs has the benefits of offloading the macrocells, improving indoor coverage and cell-edge user performance, and boosting spectral efficiency per unit area via spatial reuse. They can be deployed with relatively low network overhead, and have high potential for reducing the energy consumption of future wireless networks [2].

In Fig. 1.1, the macrocell is foreseen as the wide area layer that is responsible for providing universal coverage with a guaranteed minimum data rate requirement, maximum tolerable delay, and outage constraints. They are operator deployed with

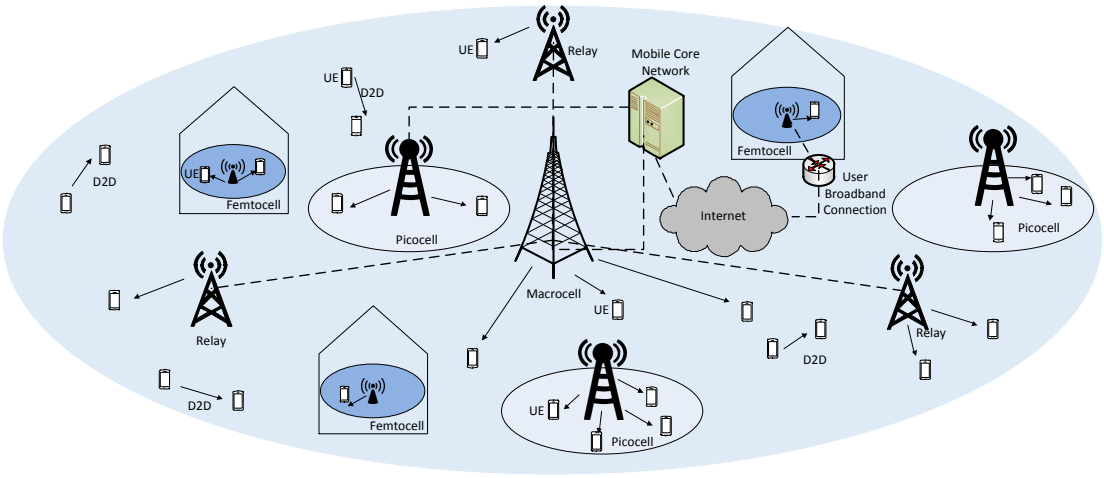

Figure 1.1: A multi-tier HetNet composed of macrocells, picocells, femtocells, relays, and D2D links. Arrows indicate wireless links, whereas the dashed lines denote the backhaul connections.

large transmit power and rely on dedicated backhaul connection to the core network. Picocells are low-power operator-deployed cell towers with the same backhaul and access features as macrocells. They use lower transmission powers than the macrocell and are utilized mainly for capacity and coverage enhancements in hotspots with insufficient macrocell penetration. Relays are usually operator-deployed access points that aim at enhancing signal level at cell edge user equipments (UEs) whose channel gains with the macrocell experience deep fading and who have poor coverage areas in the existing networks. They have similar transmit power as picocells. Femtocells, also known as home BSs or home evolved node Bs (eNBs), are low-cost low-power user deployed access points, offloading data traffic using consumers broadband connection (digital subscriber line (DSL), cable, or fiber) as a backhaul, and serving a dozen active users, mostly in homes or enterprises. They have much lower transmit powers. Finally, D2D are links that enable UEs that are in close proximity to directly communicate

by reusing the cellular resources rather than using the uplink and downlink resources in the cellular mode when communicating through the BS.

1.1.2 Macrocell Underlaid with Small Cells

Of all the additional tiers deployed, I will focus in this report on the small cells tier as shown in Fig. 1.2. Small cells are low-cost low-power BSs, which have similar functionalities as macrocell BSs, but with much smaller form factor. They are mainly deployed to provide localized coverage and capacity at households or in hot-spot areas such as city centres and transport hubs. Small cell BSs also use the same interfaces as macrocell BSs, and thus can be easily integrated, coexist and cooperate with the existing macro-cellular networks. However, in contrast to existing macrocell BSs, which often can only be deployed within a few hundred meters from their ideal location due to site acquisition issues, small cell BSs can be placed much closer to their ideal positions given their reduced size. As a result, they can be deployed in strategic locations to leverage current infrastructure, while taking UE densities, traffic demands and radio propagation conditions into account. Small cells have two options for deployment:

- Outdoors on street objects (e.g., lamp posts, bus shelters and buildings sides) to provide service to the surrounding streets and the lower floors of buildings.
- Indoors (will be referred to as home small cells or femtocells [3]) in public spaces and highly demanding areas as well as in the middle floors of high buildings to provide service to its middle and high floors and those of neighboring buildings.

Remark 1.1.1. *There are other technologies, like WiFi, that are capable of providing indoor wireless coverage. However, it is foreseen that WiFi and small cells will*

complement each other, rather compete with each other. For example, WiFi systems are known to access the unlicensed spectrum. Moreover, a large number of WiFi users transmit simultaneously on the same unlicensed band, which yields dramatically poor throughputs. In such situations, it is usually beneficial to make use of an efficiently-managed small cell network, operating over the licensed spectrum, to improve the performance. On the other hand, in situations where small cells are hindered by cross-tier and co-tier interferences, some traffic can be offloaded to the WiFi network to help relieve congestion on small cell networks [4].

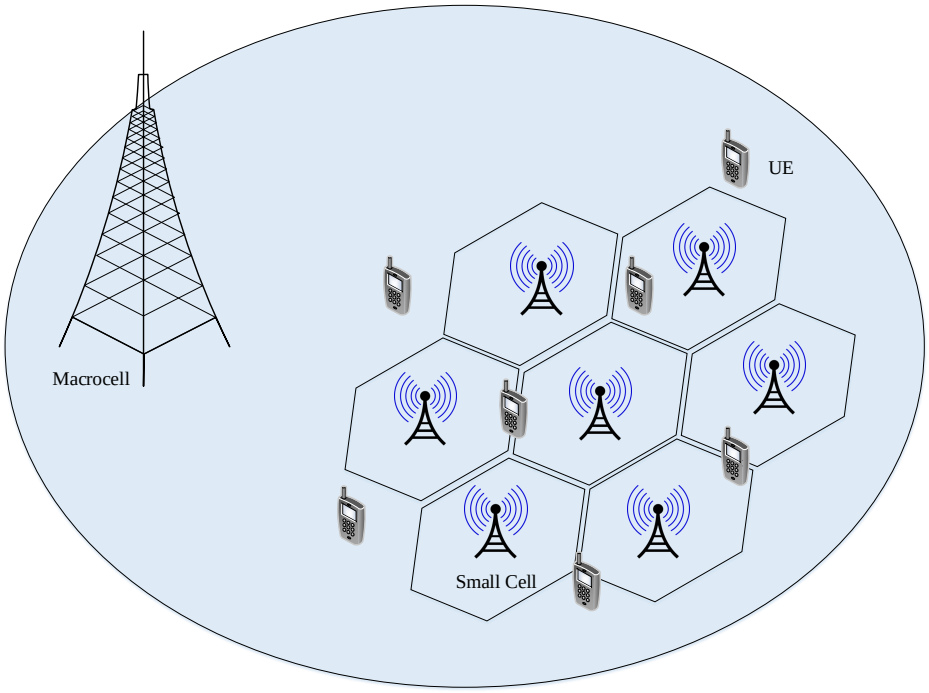

Figure 1.2: Network topology under study in this thesis.

1.1.3 Challenges in the Deployment of Small Cells

Small cells can have a very high dynamic nature. Since home small cells are randomly deployed by customers in an unplanned manner, they can be shut down at any moment. They can be removed from one place and installed in another. Hence, it is a challenging problem for operators to dynamically manage radio resources allocated to small cells. Even for outdoor operator deployed small cells, equipping small cells with self organizing capabilities is beneficial in order to make sure they are aware of the surrounding environment. Hence, distributed optimization techniques for interference mitigation are crucial. In addition, the deployment of small cells is expected to face numerous challenges. These challenges will be of more importance when the deployments of small cells become denser in urban environments [3]. The major challenges in the deployment of small cells include the following:

- **Access Modes**

Small cells can offer services to a small number of UEs. Hence, it is important to determine which UE is granted access to the small cell. In this context, three access modes are defined for small cells: closed access, open access and hybrid access. In closed access mode, only a set of registered UEs are allowed access to the small cell. On the other hand, in the open access mode, every UE can access the small cell and benefit from its services. It is foreseen that, customers would favor the closed access mode for their deployed small cells. The reason is quite simple as it is the customer who purchased the small cell and it is the one who would pay for the backhaul and electricity. Hence, a customer would not want its small cell resources to be shared with others. Open access mode, on other hand, is more favored by wireless operators, specially in public areas like airports, shopping malls and universities to provide good coverage to the users in that area. An access mode that acts as a compromise between

the two access schemes is the hybrid access mode. In it, additional UEs are granted access to the small cell beside its set of registered UEs, so long as the quality-of-service (QoS) of the registered UEs can be maintained. Therefore, the appropriate access mode should be carefully selected after a thorough analysis.

- **Management and Operation**

Small cells can be user deployed, thus, can be turned on and off at any time. Hence small cells deployment is completely dynamic and random. Moreover, the number and positions of small cells can continuously vary within a macrocell. This makes the classical centralized solutions for network management and operation impractical. Hence, small cells need to be intelligent enough to autonomously integrate into a radio access network. In other words, distributed solutions for small cells operation with some coordination and message passing with a central entity are more viable. Deploying small cells with distributed solutions for management is deemed beneficial in terms of achieving reduction in the operational expenditure (OPEX) by removing any human intervention in the operational tasks.

- **Interference Management**

This is one of the most important challenges for dense deployment of small cells. In order to use frequency spectrum as efficiently as possible, mobile operators prefer co-channel deployment of small cells to sufficiently increase the overall capacity, but this causes the problem of interference. As many small cells and macrocells are accessing the same spectrum simultaneously in a universal frequency re-use fashion, this can deteriorate each others' performance severely.

In HetNets with small cells deployments, two types of interference can be identified: co-tier interference and cross-tier interference. Co-tier interference refers to the

interference caused by network elements belonging to the same tier. In a network with small cells deployed, it refers to the interference at a small cell due to transmissions from other small cells. Cross-tier interference, on the other hand, is caused by network elements belonging to different tier. In this case, it will be interference caused to a small cell by transmissions from a macrocell and vice versa. Most of the existing small cell deployments are configured to transmit on a dedicated carrier different from that of the macrocells. While this avoids cross-tier interference, it also limits the available radio spectrum that each cell can access, and is less efficient than cochannel deployments, in which small cells and macrocells share the same frequency bands. However, while cochannel operation provides better frequency utilisation, the additional cross-tier interference can result in coverage and handover issues for mobile UEs. Hence, interference mitigation techniques should be devised in order to efficiently manage the outlined interference types. Different interference mitigation techniques employed in this thesis are outlined as follows:

1. Power and Sub-channel allocation

   In a single cell OFDMA system, power and sub-channel allocation refer to the allocation of sub-channels to different UEs and simultaneously determining the power level on each allocated sub-channel according to the channel conditions. Since, no two UEs are allocated the same sub-channel, interference is totally eliminated in a single cell system. In a multi-cell system, however, interference exists as sub-channels are re-used in the different cells. Globally performing power and sub-channels allocation in a multi-cell system is a challenging problem. In this thesis, clustering is used as a compromise between the level of interference and the network performance. In clustering, small cells can be grouped into cooperative groups. Within each group, interference can be mini-

mized through proper sub-channel allocation. On the other hand, out of cluster interference towards other clusters, as well as other network tiers, can be minimized by adapting the transmit power.

2. Coordinated Multipoint Transmission (CoMP)

   When small cells are equipped with multiple antennas, the availability of additional spatial dimensions allows the possibility of coordinating beamforming vectors across the small cells, further improving the overall performance. In other words, instead of considering transmission from other small cells as a source of interference, it is taken into account as an extra means to enhance the overall system performance. This is referred to as coordinated multipoint transmission (CoMP) [5]. The underlying concept of CoMP is quite simple: the coordinated small cells no longer adjust their parameters independently of each other, but instead coordinate the precoding or decoding processes, relying on the availability of channel state information (CSI) and the amount of information signaling over the backhaul links among the small cells.

   According to the extent of coordination among the cells, different forms of CoMP transmission schemes exist such as: dynamic selection, dynamic blanking, joint transmission, and coordinated scheduling/beamforming [6]. Those different forms differ in the type of information that is required to be shared between the different small cells, whether it is the UE data or the CSI data. In my work, I employ the joint transmission (JT) or network multiple-input multiple-output (MIMO) CoMP scheme. In it, very tight coordination among the small cells is assumed to perform JT. User data is exchanged among the coordinated small cells such that the multiple cells are simultaneously transmitting/receiving data signals to/from the UEs within the coordinated area of

multiple cells.

- **Backhauling**

The term backhaul network refers to the intermediate network that includes the links between the radio access network and the core network. In order to reap the benefits resulting from rolling out a large number of small cells and allow anytime, anywhere wireless broadband connectivity through wireless technologies, operators must still face the challenge of backhauling the traffic from the small cells to the core network in a cost effective manner. Moreover, since small cells are deployed in larger numbers and should incur a much lower cost than macrocells, the cost per small cell backhaul connection has to be significantly lower than that per macrocell backhaul connection. There are various solutions for implementing the small cells backhaul network that are either wired or wireless [7]. Wired solutions for small cells backhaul connections include fibre and Digital subscriber line (DSL). Wireless backhaul solutions, on the other hand, include sub-6 GHz point to multi-point (PtMP), microwave PtMP, microwave point to point (PtP), and millimetre wave PtP. Since each small cell in a dense urban scenario will meet different environmental conditions, the appropriate small cell backhaul solution in such cases will be comprised of a mix of the available backhaul options. In addition, the major criteria in the assessment of the viability of a certain backhaul solution include cost, the required capacity, line of sight (LOS) availability, network topology, and carrier frequency.

- **Mobility Management and Handover**

Handovers are essential in order to provide a seamless uniform service when users move in or out of the cell coverage. However, indoor small cells are mainly intended for providing indoor coverage for users and, apparently, no specific mobility management

is needed. However, with dense small cell deployment, there would be an urgent need for mobility management and handover techniques. This is a key challenge, since a small cell can have a large number of neighbours and these neighbours are created on a dynamic ad hoc basis, making it difficult to constantly keep track of neighbouring small cells. The communication with large number of neighbouring small cells for handover would also be difficult due to limited radio resources. In addition, handover in small cells is highly dependent upon the employed access mode. The number of handovers is very large in the case of open access, while it is significantly lower in closed and hybrid access modes. Hence, efficient mobility management techniques are required that take into consideration the access mode employed by the small cell, together with its method of connection to the mobile core network.

1.2 Resource Allocation in Macrocell Small Cell Networks

Resource allocation in wireless networks is used to assign the available resources in an efficient way to satisfy certain requirements subject to physical and regulatory constraints. To accomplish this, RA algorithms are proposed to exploit the variations in wireless channels by adaptively distributing scarce communication resources to either maximize or minimize some network performance metrics. Particularly, in an OFDMA environment with small cells underlaying an existing macrocell, the available resources are usually, power and sub-channels. Requirements are QoS constraints like minimum data rates and maximum delay tolerances and constraints are maximum transmit powers for the macrocell and the small cells and protection constraints like maximum tolerable interference levels. It is evident that the design of efficient RA mechanisms (either centralized or decentralized) is a key challenge in multi-tier HetNets. In the subsequent chapters of this thesis, as shown in Fig. 1.3, I investigate

different issues related to RA in macrocell small cell networks (the network topology shown in Fig. 1.2).

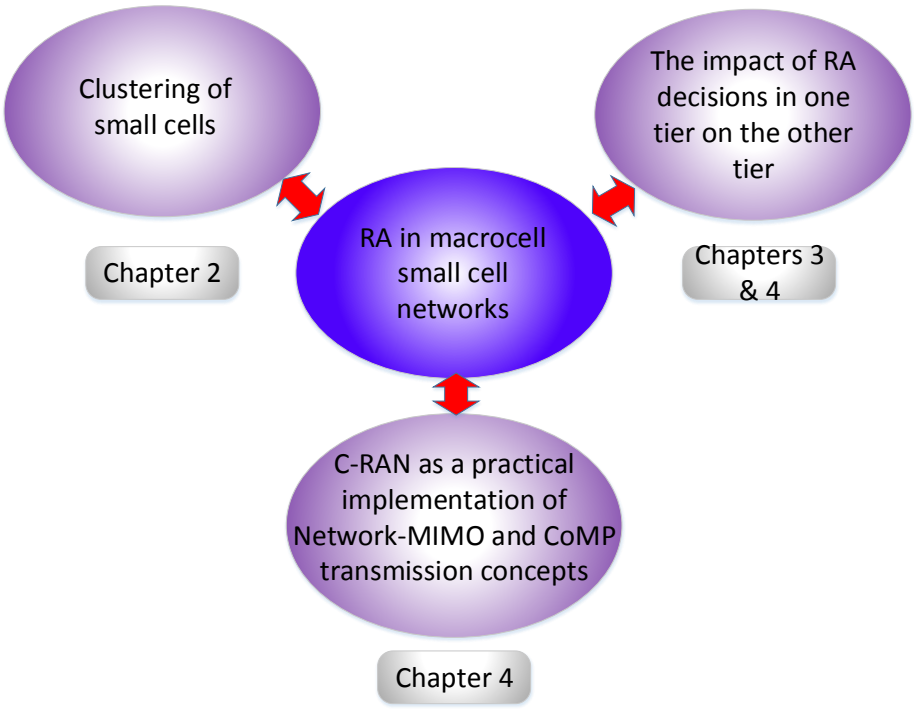

Figure 1.3: Resource allocation issues in macrocell small cell networks.

First of all, according to Section 1.1.3, co-tier interference among small cells is a critical issue. Efficient RA schemes are required to handle co-tier interference properly. Centralized solutions for RA can reach the optimal solution, but they require global information. Distributed solutions, on the other hand, are more computationally efficient, but may not give good quality solutions as the centralized ones. Hence, a solution that gives a compromise between the solution quality and the computational efforts is required. In Chapter 2, I propose a semi-distributed (hierarchical) interference management scheme based on joint clustering and resource allocation for small cells. The problem is formulated as a mixed integer non-linear program (MINLP).

The solution is obtained by dividing the problem into two sub-problems, where the related tasks are shared between the femto gateway (FGW) and small cells. The FGW is responsible for clustering, where correlation clustering is used as a method for small cells grouping. In this context, a low complexity approach for solving the clustering problem is used based on semi-definite programming (SDP). In addition, an algorithm is proposed to reduce the search range for the best cluster configuration. For a given cluster configuration, within each cluster, one small cell is elected as a cluster head (**CH**) that is responsible for resource allocation among the small cells in that cluster. The **CH** performs sub-channel and power allocation in two steps iteratively, where a low-complexity heuristic is proposed for the sub-channel allocation phase.

Second, cross-tier interference was shown to be as important as the co-tier one. Hence, RA results in one network tier has an impact on the performance of UEs belonging to other network tiers. This motivates us to ask the following fundamental question. If one network tier can satisfy the QoS requirement of its UEs using different RA schemes, is there one scheme of them that is favorable from the point of view of the other network tiers ?. In this context, in Chapter 3, I present a joint sub-channel and power allocation framework for downlink transmission in an OFDMA-based cellular network composed of a macrocell underlaid by small cells. In this framework, the RA problems for both the macrocell and small cells are formulated as optimization problems. For the macrocell, I formulate an RA problem that is aware of the existence of the small cell tier. In this problem, the macrocell performs RA to satisfy the data rate requirements of macro user equipments (MUEs) while maximizing the tolerable interference from the small cell tier on its allocated sub-channels. Although the RA problem for the macrocell is shown to be a MINLP, I prove that the macrocell can

solve another alternate optimization problem that will yield the optimal solution with reduced complexity. For the small cells, following the same idea of tier-awareness, I formulate an optimization problem that accounts for both RA and admission control (AC) and aims at maximizing the number of admitted users while simultaneously minimizing the consumed bandwidth. Similar to the macrocell optimization problem, the small cell problem is shown to be an MINLP. I obtain a sub-optimal solution to the MINLP problem relying on convex relaxation. In addition, I employ the dual decomposition technique to have a distributed solution for the small cell tier.

Finally, Cloud Radio Access Network (C-RAN) has emerged as a novel mobile network architecture which can address a number of challenges the operators face while trying to support growing end users' needs. The main idea behind C-RAN is to pool the Baseband Units (BBUs) from multiple base stations into centralized BBU pool for statistical multiplexing gain, while shifting the burden to the high-speed wireline transmission of In-phase and Quadrature (IQ) data. Thus, small cells can be deployed in a C-RAN architecture and reap the claimed benefits that C-RAN can offer. However, problems remain due to the limited bandwidth of the fronthaul links connecting the small cells to the BBU pool. Moreover, cross-tier interference between the macrocell and the small cells must still be handled properly. In this context, in Chapter 4, I present a joint RA and AC framework for an OFDMA-based downlink cellular network composed of a macrocell underlaid by a C-RAN of small cells. In this framework, the RA problems for both the macrocell and small cells are formulated as optimization problems. In particular, the macrocell, being aware of the existence of the small cells, maximizes the sum of the interference levels it can tolerate subject to the macrocell power budget and the QoS constraints of MUEs. On the other hand, the small cells minimize the total downlink transmit power subject to their power budget,

QoS requirements of small cell UEs (SUEs), interference thresholds for MUEs, and fronthaul constraints. Moreover, AC is considered in the resource allocation problem for the small cells to account for the case where it is not possible to support all SUEs. Besides, to allow for the existence of other network tiers, small cells have a constraint on the number of sub-channels that can be allocated. Both optimization problems are shown to be MINLPs for which, lower complexity algorithms are proposed that are based on the framework of successive convex approximation (SCA).

1.3 Contributions

The contributions of our studies are three folded and can be summarized as follows:

- Clustering and Resource Allocation for Dense Femtocells in a Two-Tier Cellular OFDMA Network

    1. I propose a framework for clustering, sub-channel and power allocation in a two-tier macrocell small cell network. These processes are performed in a hierarchical fashion as follows. The FGW is responsible for the clustering sub-problem whereas, within each cluster, one small cell elected as a **CH** is responsible for sub-channel and power allocation.
    2. I formulate the clustering sub-problem as a correlation clustering problem which is solved by using an SDP-based algorithm, to avoid the exponential complexity associated with obtaining the optimal cluster by exhaustive search. In addition, I analyze the complexity of solving this correlation clustering problem.
    3. In the correlation clustering formulation, there is a penalty term. By going through the predefined range for the penalty term, different cluster

configurations can be obtained. Hence, to reduce the complexity further, I propose another algorithm that reduces the search range of the penalty term.

4. After the FAPs are organized into disjoint clusters, each **CH** performs sub-channel and power allocation with interference constraints to neighboring FUEs and MUEs as well as minimum rate requirements. For this, I propose a low-complexity scheme for sub-channel allocation.

- Tier-Aware Resource Allocation in OFDMA Macrocell-Small Cell Networks

1. I develop a complete framework for tier-aware resource allocation in an OFDMA-based two-tier macrocell-small cell network with new objectives, which are different from the traditional sum-power or sum-rate objectives.
2. For the macrocell tier, I formulate a resource allocation problem that is aware of the existence of the small cell tier and show that it is an MINLP.
3. I prove that the macrocell can solve another alternate optimization problem that yields the optimal solution for the MINLP with polynomial time complexity.
4. I compare the proposed method for the macrocell RA problem to the traditional "minimize the total sum-power" problem and show that the proposed method outperforms the traditional one in terms of the average number of admitted SUEs.
5. For the small cell tier, I formulate a joint resource allocation and admission control problem that aims at maximizing the number of admitted SUEs and minimizing their bandwidth consumption to accommodate additional tiers, and show that it is an MINLP.

6. I offer an upper bound solution to the MINLP through convex relaxation and propose a solution to the convex relaxation that can be implemented in a distributed fashion using dual decomposition.

- Resource Allocation for an OFDMA Cloud-RAN of Small Cells Underlaying a Macrocell

  1. I develop a complete framework for downlink radio resource allocation in an OFDMA-based two-tier cellular network where a macrocell is underlaid with a C-RAN of small cells.
  2. For the macrocell, I formulate a resource allocation problem that is aware of the existence of the small cell tier. Specifically, the macrocell aims at maximizing the sum of tolerable interference levels for the MUEs subject to the macrocell power budget constraint and the QoS requirements of the MUEs. In addition, I relax the simplifying assumptions made in [8].
  3. I show that the macrocell RA problem is an MINLP. To gain further insights into the behavior of the macrocell RA problem with the proposed objective function, I investigate the single MUE macrocell problem and show how to obtain its optimal solution, with polynomial time complexity, despite its non-convexity.
  4. Motivated by the single MUE macrocell problem, I propose a low complexity solution for the multi-MUE macrocell RA problem that relies on the framework of SCA.
  5. In the C-RAN tier, I formulate the RA problem whose objective is to minimize the total downlink transmit power subject to QoS constraints for SUEs, MUEs' interference thresholds, small cells power budget, and

fronthaul capacity constraints as an optimization problem.

6. I show that the C-RAN RA problem is an MINLP. Hence, I offer a low complexity solution based upon the SCA approach. In the SCA approach, each approximating convex problem can be cast as a second order cone program (SOCP) [9]. Moreover, I incorporate AC, jointly with RA, to deal with infeasibility issues in the C-RAN RA problem in case it is not possible to support all SUEs with their QoS requirements.

1.4 Organization

The rest of this thesis is organized as follows. In Chapter 2, I study the idea of clustering small cells into coordinated groups and its impact on the small cell tier performance. In Chapter 3, I investigate the impact of different macrocell RA policies on the small cell tier and propose a distributed solution for small cell RA problem. In Chapter 4, I investigate the impact of different macrocell RA policies on the small cell tier which is deployed in a C-RAN architecture and solve a more general macrocell RA problem than the one in Chapter 3. Finally, in Chapter 5, I summarize the work done, highlight the main contributions, and discuss possible future research directions.

Chapter 2

Clustering and Resource Allocation for Dense Femtocells in a Two-Tier Cellular OFDMA Network

2.1 Introduction

In this Chapter, I formulate the problem of joint clustering, sub-channel and power allocation as an optimization problem. This problem, however, is NP-hard and cannot be solved by a central entity such as a femto gateway (FGW) in a practical system. Therefore, it is divided into two sub-problems, namely, the clustering sub-problem and the sub-channel and power allocation sub-problem. The tasks are divided between the FGW and the femtoce access point (FAPs) in a hierarchical semi-distributed fashion. The FGW will be responsible for the clustering sub-problem. After the FAPs are divided into clusters (i.e., disjoint groups), one FAP, in each cluster, is elected as a cluster head (**CH**) and performs sub-channel and power allocation in its cluster. Now, the clustering subproblem can be solved through exhaustive search by

trying all possible cluster configurations, given the layout of femtocells and channel gains. However, the number of possible cluster configurations grows exponentially with the number of femtocells. Hence, another approach is proposed for forming clusters using the concepts of correlation clustering and penalty term. Since the correlation clustering problem is an NP-hard problem, it is solved by formulating the problem as a semi-definite program (SDP). The SDP is considered as a relaxation to an integer linear program (ILP), where an integer solution to the relaxed SDP problem is recovered back through randomized rounding. Now, using the penalty term, which has a predefined range, different cluster configurations are obtained. To further reduce the computational burden on the FGW in searching through the range of the penalty term, an algorithm is proposed which eliminates the requirement of going through the entire range of the penalty term.

After the FAPs are organized into disjoint clusters, each **CH** performs sub-channel and power allocation within its cluster. Joint sub-channel and power allocation is performed in two phases. In the first phase, for a given power allocation, sub-channel allocation is performed. Since the sub-channel allocation problem turns out to be an ILP, a sub-optimal scheme for sub-channel allocation is proposed. Given the sub-channel allocation, the power allocation is performed. The entire process of sub-channel and power allocation is performed iteratively until convergence. The sub-channel and power allocation is performed for each possible cluster configuration and the cluster configuration yielding the highest data rate is the optimal one. It is worth mentioning that the problem of optimal clustering and the tradeoff between co-tier interference and the share in the available spectrum, which depends on the cluster size, as analyzed in this Chapter, have not been investigated in the existing literature.

The rest of this Chapter is organized as follows. Section 2.3 describes the system model and assumptions. Section 2.4 outlines the joint clustering, sub-channel and power allocation problem which turns out to be an MINLP. In Section 2.5, the clustering sub-problem is discussed, together with the proposed low-complexity algorithms for obtaining the best cluster configuration. Then, Section 2.6 discusses the sub-channel and power allocation sub-problems, where, sub-channel and power allocation are done iteratively in two steps and a low-complexity sub-channel allocation heuristic is proposed. In Section 2.7, numerical results for different scenarios and topologies demonstrate the performance gains due to the proposed schemes. Finally, Section 2.8 concludes the Chapter.

2.2 Related Work

Several works in the literature have considered the problem of resource allocation in two-tier cellular wireless networks. In [10], a downlink power control method was proposed to mitigate interference in a macrocell-femtocell network. The QoS for macro UEs (MUEs) was guaranteed by limiting interference from femtocells to nearby MUEs. A centralized and a distributed approach for power control were proposed and compared. However, no frequency allocation was considered. In [11], a joint power control and resource allocation scheme for a co-channel femtocell-macrocell network was proposed with QoS guarantees for both MUE and femto UE (FUE). Both centralized and distributed approaches were proposed. In [12], a power loading and resource allocation scheme was proposed for a femtocell network, where interference constraints were considered to protect the MUEs. An iterative water filling approach was considered to satisfy the above target. In [13], a joint power and sub-channel allocation scheme was proposed to maximize system capacity for dense indoor mobile

communication systems. The authors assumed femtocells to be densely deployed and proposed a centralized resource allocation framework. The authors proved that the optimal power allocation in such an environment has a special form known as binary power allocation, where for every sub-channel, only a single femtocell loads power. In [14], the authors proposed a scheme for sub-channel and power allocation in a femtocell-macrocell network with spectrum sharing. However, the authors neglected co-tier interference and considered cross-tier interference only. Another scheme for interference-aware resource allocation in a co-channel deployment of a two-tier femtocell network was studied in [15], where the sub-channel and power allocation problem was modeled as a non-cooperative game. The MUEs were protected by pricing transmissions of femto users. However, no protection is done for co-channel femto users. In addition, unlimited amount of information is allowed to be transmitted between MBS and FAPs. The above works did not consider the idea of clustering by placing the FAPs into disjoint groups.

There have been works in the literature which have considered the idea of clustering femtocells into coordinated groups. In [16], the authors proposed a dynamic clustering-based sub-band allocation scheme in a dense femtocell environment. However, the authors only considered frequency allocation. In [17], the authors proposed a cooperative scheme for femtocell network, where cooperation was modeled as a coalitional game. In their work femtocells cooperated by forming coalitions. In a coalition, interference was eliminated among femtocells through interference alignment. However, split spectrum operation was assumed, where femtocells had their dedicated spectrum which reduced spectrum efficiency. In [18], the authors proposed a scheme for macro and femtocell interference mitigation, where cross-tier interference was minimized through cognitive and sensing capabilities by employing both overlay

and underlay modes of operation. To mitigate co-tier interference, femtocells were grouped into clusters, where a cluster was a group of co-channel FAPs that used the same sub-channel. However, there was no QoS guarantee for the FAPs. In addition, the number of clusters was assumed to be known and fixed a priori (i.e., static) and hence the flexibility of a dynamic clustering process was not exploited.

The authors in [19] proposed a joint power control and resource allocation algorithm in an OFDMA femtocell network, where femtceolls were grouped into disjoint clusters. However, the authors only considered orthogonal channel assignment (split spectrum) between femtocells and macrocells; hence, no cross-tier interference existed. In [20], the authors proposed a clustering scheme known as Similarity-Based clustering for cognitive FAPs in a shared spectrum environment. After FAPs were organized into disjoint clusters, cluster heads performed resource allocation. However, clustering and resource allocation were performed separately. In addition, no sense of optimality was considered in [17–20] when dealing with the optimal cluster size. The authors in [21] and [22] studied the problem of sub-channel and power allocation in a two-tier clustered femtocell network with universal frequency reuse. However, only equal-sized clusters were assumed. Another femtocell clustering scheme was proposed in [23], where the number of clusters was known a priori and no sub-channel allocation was considered. In [24], another work for sub-channel and power allocation with clustering of femtocells was considered. In their network model, both the co-tier and cross-tier interferences were considered. However, the entire frequency band was divided into three parts: one part dedicated for femtocells, another was dedicated for macrocells, and the last part was available for both femtocells and macrocells. This approach generally is not suitable in a highly dynamic environment and leads to low spectral efficiency.

Different from the clustering schemes which assume a split spectrum operation with dedicated spectrum for femtocells, I assume a more general universal frequency reuse approach which is known to be more flexible and offers a higher spectral efficiency. Note also that, although clustering of femtocells has been investigated in the literature, the trade-off between the perceived interference and the available bandwidth due to clustering and the problem of optimizing the cluster size have not been studied.

2.3 System Model and Assumptions

2.3.1 System Model

Fig. 4.1 shows the topology under consideration in this Chapter, where femtocells are deployed in a dense manner to cover an indoor area. In such an environment, channels between femto user equipments (FUEs) and their FAPs generally experience good propagation conditions. However, signals received from outdoor macrocells are highly attenuated. All user equipments, i.e., the FUEs and macro user equipments (MUEs) exist indoor, however, the MUEs are served by their outdoor MBSs. We denote by $\mathcal{S}$ the set of femtocells, where $S = |\mathcal{S}|$, and by m the macrocell. Let $\mathcal{N}$ denote the set of indices of the sub-channels in the system and $N = |\mathcal{N}|$.

We assume that the channel states of the sub-carriers are the same within a sub-channel n of bandwidth Δf. We denote a user belonging to femtocell s by k_s and a user belonging to the macrocell B by k_B. We define $g^n_{k_i,j}$ as the channel gain between user k_i and base station j on sub-channel n. Traditionally, resource allocation is performed based on the observed instantaneous channel gains. However, this assumption has been claimed impractical [25]. Also, performing resource allocation

based on instantaneous channel gains (which consider small-scale fading) will lead to extensive signaling and huge amount of information transfer in the femtocell backhaul. In addition, in a practical system, clustering cannot be performed at the time scale of small-scale fading. Therefore, in our model, similar to [20], average channel gains[1] (which consider path-loss and log-normal shadowing) are used for clustering and resource allocation.

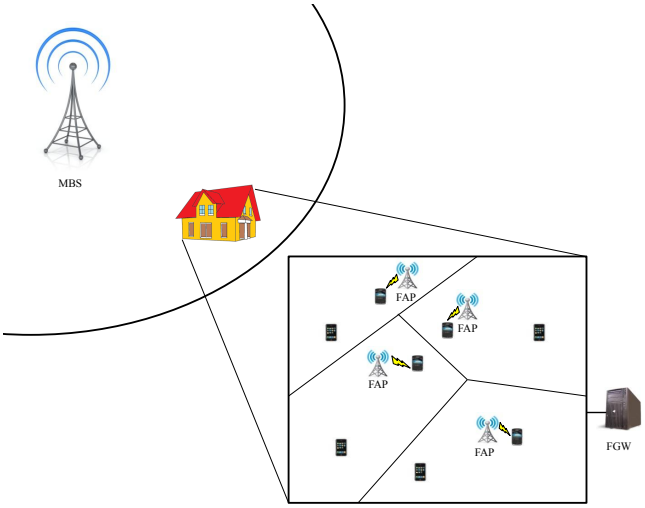

Figure 2.1: Network topology under consideration.

Since the distance between an FUE k_i and its serving FAP i is very small, channel gain between a femtocell j and an FUE k_i, served by another femtocell i, can be approximated by the channel gain between the two femtocells i, j, i.e., $g^n_{k_i,j} \approx g^n_{i,j}$ [18]. Following this assumption, we can deduce that channel gains are symmetric, i.e., for two FAPs i, j and two FUEs k_i, k_j, we have $g^n_{k_j,i} = g^n_{k_i,j}$ since the channel gain between the two FAPs i and j is the same as the channel gain between the two FAPs j and i. The unit power SINR of an FUE k_s served by femtocell s on sub-channel n can be

[1]Although for a certain link, different sub-channels have the same average channel gains, they experience different interference levels. Consequently, sub-channel allocation in a two-tier OFDMA network under co-channel deployment is a non-trivial problem.

defined as follows:

$$\gamma^n_{k_s,s} = \frac{g^n_{k_s,s}}{\sum_{j\neq s, j\in\mathcal{S}} P^n_{k_j,j} g^n_{k_s,j} + P^n_{k_B,B} g^n_{k_s,m} + N_o} \tag{2.1}$$

where $P^n_{k_i,i}$ is the power assigned to the link between user k_i and base station i on sub-channel n and N_o is the noise power. We assume closed subscriber group (CSG) femtocells, where access to each femtocell is restricted to registered UEs only, and both the macrocells and femtocells share the same frequency spectrum.

Femtocell networks use cell-specific reference signals and unique cell-ids. All FUEs are capable of receiving the cell-specific reference signals and identifying the interference source. In addition, femtocells are connected to the mobile core network, using the user's broadband connection (Digital Subscriber Line DSL or cable television), via an intermediate entity called the Femto Gateway (FGW) or Home eNB Gateway (HeNB GW) [26]. The interface between a femtocell and an FGW is referred to as the S1 interface. The FGW provides concentration and aggregation functionalities to a group of femtocells. Hence, an FGW appears to the mobile core network as an eNB and to a femtocell as a mobile core network element. The FGW can obtain all necessary information about channel gains between femtocells through the S1 interface [24], based on which, the FGW can propose different clustering configurations. However, due to the limitations and the unpredicted delay on the S1 interface (being based on the user's broadband connection), the FGW is not able to obtain instantaneous channel gain information from the femtocells. Communication between femtocells is possible, thanks to the X2 interface [26]. However, information such as the average channel gains are only exchanged. Information about user scheduling and sub-channel allocation are not exchanged between neighboring femtocells.

2.4 Joint Clustering, Sub-channel and Power Allocation for Femtocell Network

The femtocell network is divided into disjoint clusters. The idea behind clustering is to divide the joint sub-channel and power allocation problem into smaller sub-problems. We define $\mathcal{C}$ as the set of clusters of femtocells. A cluster $c_l \in \mathcal{C}$ is the l^{th} set of femtocells such that $c_l \subseteq \mathcal{S}$, $\forall l \in \{1, 2, ..., |\mathcal{C}|\}$, $\bigcup_{l=1}^{|\mathcal{C}|} c_l = \mathcal{S}$, and $\bigcap_{l=1}^{|\mathcal{C}|} c_l = \varnothing$. Note that the entire set of sub-channels $\mathcal{N}$ is available to each cluster and within a cluster, no two femtocells transmit simultaneously on the same sub-channel. Therefore, there is no co-tier interference within a cluster. For very small cluster sizes, with one extreme being no clustering, the share of each femtocell in the available spectrum is high; however, the co-tier interference could be significant in this case. On the other hand, for large cluster sizes, co-tier interference among neighboring femotcells is minimized. However, the share of sub-channels for each femtocell would be small. This suggests that cluster size is an important parameter to give a compromise between the share in the available spectrum and the co-tier interference.

We define the following optimization problem for joint clustering, sub-channel and

power allocation:

$$\max_{\Gamma^n_{k_s,s},P^n_{k_s,s}} \sum_{l=1}^{|\mathcal{C}|}\sum_{s\in c_l}\sum_{n=1}^{N}\Gamma^n_{k_s,s}\Delta f\log_2\left(1+P^n_{k_s,s}\gamma^n_{k_s,s}\right) \tag{2.2}$$

subject to

$$\text{C1}: \sum_{n=1}^{N}\Gamma^n_{k_s,s}\Delta f\log_2\left(1+P^n_{k_s,s}\gamma^n_{k_s,s}\right)\geq R_s, \forall s$$

$$\text{C2}: \sum_{n=1}^{N}\Gamma^n_{k_s,s}P^n_{k_s,s}\leq P_{s,max}, \forall s$$

$$\text{C3}: \Gamma^n_{k_s,s}P^n_{k_s,s}g^n_{k_B,s}\leq \xi^n_{k_B}, \forall n, s$$

$$\text{C4}: \Gamma^n_{k_s,s}P^n_{k_s,s}g^n_{k_j,s}\leq \xi^n_{k_j}, \forall n, s\in c_l, j\notin c_l$$

$$\text{C5}: \sum_{s\in c_l}\Gamma^n_{k_s,s}=1, \forall n$$

$$\text{C6}: \bigcup_{l=1}^{|\mathcal{C}|}c_l=\mathcal{S},\ \text{C7}: \bigcap_{l=1}^{|\mathcal{C}|}c_l=\varnothing$$

$$\text{C8}: |c_l|\leq \upsilon, \forall l\in\{1,2,...,|\mathcal{C}|\}$$

$$\text{C9}: P^n_{k_s,s}\geq 0, \forall n, s$$

$$\text{C10}: \Gamma^n_{k_s,s}\in\{0,1\}, \forall n, s$$

where $\Gamma^n_{k_s,s}\in\{0,1\}$ is an indicator that takes the value of 1 if sub-channel n is allocated to the link between user k_s and femtocell s and 0 otherwise, $P^n_{k_s,s}\geq 0$ is the power assigned to the link between them.

In the optimization problem (2.2), the objective it to maximize the sum-rate of the femtocell clusters subject to a data rate requirement R_s and total power budget $P_{s,max}$ for each femtocell s as indicated in C1 and C2, respectively. We have interference constraints for neighboring MUEs k_B served by macrocell B and FUEs k_j served by femtocells j in a neighboring cluster as given in C3 and C4, respectively. C5 is the

exclusion constraint indicating that, in a cluster c_l, sub-channel n can be used in one femtocell only. Constraints C6 and C7 indicate that the whole set of clusters $\mathcal{C}$ form the set of femtocells $\mathcal{S}$ and that the set of clusters are disjoint. Finally, C8 limits the maximum cluster size to υ. In our work, the maximum cluster size will be limited by the number of available sub-channels, i.e., $\upsilon = N$.

This problem is an MINLP whose solution is intractable. It includes both continuous and discrete variables. In addition, solving problem (2.2) requires a centralized mode of operation which is too complex for a practical solution. Hence, to solve this problem, we propose to divide it into two sub-problems, i.e., the clustering sub-problem and the sub-channel and power allocation sub-problem. First, the FGW gathers information about average channel gains among all the FAPs. Based on the range of the penalty term (to be discussed in Section 2.5.2), the FGW performs the clustering phase and obtains a group of candidate cluster configurations. The FGW sends this clustering information to the FAPs through the S1 interface (wired backhaul). Within each cluster, one femtocell takes the role of a **CH**[2] and performs sub-channel and power allocation for each candidate cluster configuration[3]. Then it forwards the average achievable data rate to the FGW. The cluster configuration yielding the highest average data rate for all FAPs is the best cluster configuration. It is worth mentioning that the best cluster configuration is channel gain-dependent, but since it is done based on path-loss and large-scale fading, the obtained cluster configuration remains stable for long intervals of time. Fig. 2.2 summarizes the overall clustering and resource management operation.

[2]The election process of the **CH**s is out of the scope of this Chapter.

[3]A **CH** performs sub-channel and power allocation within the corresponding cluster c_l. By setting the maximum cluster size to a small value, the computational burden on the **CH** can be limited.

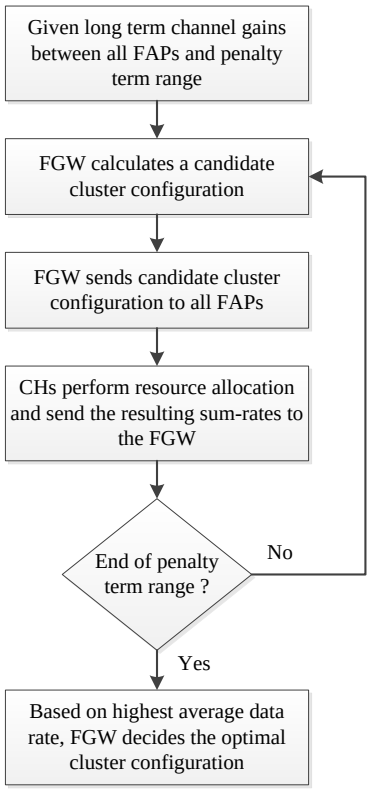

Figure 2.2: Overall operation of clustering and resource allocation.

2.5 Clustering Sub-problem

2.5.1 Optimal Clustering

Optimal clustering can be obtained by an exhaustive search. For a given number of femtocells, all possible clustering configurations for the femtocells are tried. For a given clustering configuration, sub-channel and power allocation is performed. The cluster configuration yielding the highest sum-rate is the optimal cluster configuration.

For S FAPs, the number of possible ways to cluster them is given by the Stirling

Number of the Second Kind [27]:

$$\sum_{i=1}^{S} \frac{1}{i!} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} j^S \approx \mathcal{O}\left(S^S\right). \tag{2.3}$$

It is clear that the number of possible cluster configurations (Bell Number) grows exponentially with the number of FAPs. Therefore, searching for the optimal cluster configuration by exhaustive search is prohibitive.

2.5.2 Correlation Clustering

In the original correlation clustering problem, we are given an undirected graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $(i, j) \in \mathcal{E}$ is the set of edges between each two pair of vertices. Each edge (i, j) is given two non-negative weights $w^+_{i,j}$ and $w^-_{i,j}$. The target is to cluster the vertices into sets of similar vertices, where the degree to which they are similar is given by $w^+_{i,j}$ and the degree to which they are different is given by $w^-_{i,j}$. In other words, the target is to find a partition that maximizes the total $w^+_{i,j}$ weight of edges inside the sets of partitions plus the total $w^-_{i,j}$ weight of edges between the sets of partitions [28]. One benefit of correlation clustering is that it does not necessitate the specification of the number of clusters. However, correlation clustering problems are generally NP-hard.

In the context of femtocell networks, two femtocells $i, j \in \mathcal{S}$ are highly similar, i.e., have high $w^+_{i,j}$, if they are severely interfering to each other. Two femtocells are severely interfering to each other if they have good channel gain $g^n_{i,j}$ between each other. So, we can make $w^+_{i,j}$ proportional to the channel gain between the two femtocells i, j by setting $w^+_{i,j} = g^n_{i,j}$. In this way, $w^+_{i,j}$ represents the motive why two femtocells would be in the same cluster. Two severely interfering femtocells,

when placed in the same cluster, will have lower co-tier interference. Hence, their sum-rate will increase. However, as the number of femtocells in a cluster increase, although femtocells enjoy lower co-tier interference, the share in the available number of sub-channels for each femtocell decreases, which will cause the sum-rate to decrease. Hence, to be inline with the objective of the optimization formulation in (2.2) which is the sum-rate maximization, we need another parameter to indicate why two femtocells would not favor to be in the same cluster. This term will be $w^-_{i,j}$ and will be referred to as the penalty term. It is referred to as a penalty because, when two femtocells join the same cluster, they have a benefit of reduced co-tier interference at the penalty of reduced available bandwidth. In our work, $w^-_{i,j}$ will be the same for all femtocells $i, j \in \mathcal{S}$, i.e., $w^-_{i,j} = w^-_n$. In this way, w^-_n would be the parameter used to represent the trade-off between the level of co-tier interference and the available bandwidth to use. Hence, by varying the penalty term w^-_n in the range ($\min_{i,j} w^+_{i,j} + \Delta \leq w^-_n \leq \max_{i,j} w^+_{i,j} + \Delta$), where, $\Delta > 0$ is a very small number, the FGW can obtain different cluster configurations.

If $w^-_n = \min_{i,j} w^+_{i,j} + \Delta$, the resulting configuration will be to put all femtocells in a single cluster and if $w^-_n = \max_{i,j} w^+_{i,j} + \Delta$, the resulting configuration will be to have each femtocell in a single cluster. Define $\mathbf{W}^+ \in \mathbb{S}^{S \times S}_+$, where $[\mathbf{W}^+]_{i,j} = w^+_{i,j}$. We have $w^-_n = w^+_{i,j} + \Delta$, where $1 \leq i < j \leq S$. Define $\mathbf{w}^- \in \mathbb{R}_+^{\frac{S^2-S}{2}}$, where $\mathbf{w}^{-T} = \left(w^-_1, ..., w^-_n, ..., w^-_{\frac{S^2-S}{2}}\right)$. The elements of the vector $\mathbf{w}^-$ constitute the range of the penalty term. Hence, by trying all the values in $\mathbf{w}^-$, different cluster configurations are possible.

Thus we have the following problem to be solved $\forall w_n^- \in \mathbf{w}^-$:

$$\max_{x_{i,j}} \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} w_{i,j}^+ x_{i,j} + w_n^- \left(1 - x_{i,j}\right) \tag{2.4}$$

subject to

$$\text{C1} : x_{i,i} = 1, \forall i \in \mathcal{S}$$

$$\text{C2} : x_{i,j} = x_{j,i}, \forall i, j \in \mathcal{S}$$

$$\text{C3} : x_{i,j} + x_{j,k} - x_{i,k} \leq 1, \forall i, j, k \in \mathcal{S} : k > i, j \neq i, k$$

$$\text{C4} : \sum_{j \in \mathcal{S}} x_{i,j} \leq \upsilon, \forall i \in \mathcal{S}$$

$$\text{C5} : x_{i,j} \in \{0, 1\}, \forall i, j \in \mathcal{S}.$$

In the optimization problem (2.4), C1 indicates that a femtocell is in the same cluster with itself. C2 is a symmetry condition which specifies that it is the same to say that i and j are in the same cluster or j and i are in the same cluster. C3 is the triangular condition which states that if femtocells i, j are in the same cluster and femtocells j, k are in the same cluster, then i, k must be in the same cluster as well. C4 is a limit on the maximum cluster size to be υ. Finally, C5 states that $x_{i,j}$ is a binary decision variable which takes a value of 1 if i, j are in the same cluster and 0, otherwise. The following two problems arise with this approach:

- The problem (2.4) is an ILP, which is NP-hard. It can be solved optimally using Brach and Bound (BnB). However, BnB has an exponential complexity, hence it is not suitable as a practical solution. Therefore, another approach based on SDP is used to solve the problem.

- Solving problem (2.4) for all $w_n^- \in \mathbf{w}^-$ requires solving $\frac{S^2 - S}{2} \approx \mathcal{O}\left(S^2\right)$ problems. This represents a heavy computational burden on the FGW. Hence, an

algorithm is proposed that reduces the computational burden by eliminating the requirement of going through the entire range of the penalty term.

Sections 2.5.2 and 2.5.2 discuss the approaches that will be followed to address the above mentioned problems.

SDP-based correlation clustering algorithm

A lower complexity solution to the correlation clustering problem can be obtained by using SDP. SDP is a generalization of linear programming with an additional semi-definite constraint. The benefit of the SDP approach is that, though a relaxation to the original NP-hard problem, it can be solved to within an additive error of ϵ in time that is polynomial in the size of the input and $\log \frac{1}{\epsilon}$. The SDP formulation for the correlation clustering problem can be written as follows:

$$\max_{x_{i,j}} \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} w_{i,j}^{+} x_{i,j} + w_{n}^{-} \left(1 - x_{i,j}\right) \tag{2.5}$$

subject to

$$\text{C1} : x_{i,i} = 1, \forall i \in \mathcal{S}$$

$$\text{C2} : x_{i,j} + x_{j,k} - x_{i,k} \leq 1, \forall i, j, k \in \mathcal{S} : k > i, j \neq i, k$$

$$\text{C3} : \sum_{j \in \mathcal{S}} x_{i,j} \leq \upsilon, \forall i \in \mathcal{S}$$

$$\text{C4} : x_{i,j} \geq 0, \forall i, j \in \mathcal{S}$$

$$\text{C5} : X = (x_{i,j}) \succeq 0.$$

In (2.5), C4 indicates that the binary variable $x_{i,j}$ is relaxed and C5 is the semi-definite constraint restricting the matrix $\mathbf{X} \in \mathbb{S}_{+}^{S \times S}$ of the variables $x_{i,j}$ to be Positive Semi-Definite (PSD). A symmetric matrix $\mathbf{X}$, that is PSD as well, can be written as

$\mathbf{X} = \mathbf{V^T V} = \mathbf{Q D^{\frac{1}{2}}} \left(\mathbf{Q D^{\frac{1}{2}}}\right)^T$, where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_S] \in \mathbb{R}^{S \times S}$ such that $\mathbf{v}_i \in \mathbb{R}^S$, $\mathbf{Q}$ is a matrix of the eigen vectors of $\mathbf{X}$, and $\mathbf{D}$ is a diagonal matrix whose elements are the eigen values of $\mathbf{X}$. It is worth mentioning that an element $[\mathbf{X}]_{i,j} = \mathbf{v}_i.\mathbf{v}_j$. Each vector $\mathbf{v_i}$ corresponds to a certain femtocell $i \in \mathcal{S}$. To solve (2.5), we use CVX, a software package for specifying and solving convex programs [29], [30].

After solving the SDP formulation, obtaining $\mathbf{X}$ and putting the resulting solution into vector format $\mathbf{X} = \mathbf{V^T V}$, we have a relaxed solution that is an upper bound to the original clustering problem (2.4). This relaxed solution will be rounded to obtain an integer solution that is feasible to (2.4). One way to accomplish this is through Randomized Rounding [31]. This can be done by generating a number L of unit norm random vectors $\mathbf{r}_i^T = (r_{i1}, ..., r_{iF}), \mathbf{r}_i \in \mathbb{R}^S$, $i = 1, 2, ..., L$, where each component of $\mathbf{r}_i$ is drawn from $\mathcal{N}(0, 1)$, the normal distribution with mean 0 and variance 1. The number of vectors L must satisfy that $2^L \geq S$. The L random vectors give rise to 2^L clusters $c_1, c_2, ..., c_{2^L}$. The mapping of femtocells into clusters is then done as follows:

$$
\begin{aligned}
c_1 &= \{i \in \mathcal{S} : \mathbf{r}_1.\mathbf{v}_i \geq 0, ..., \mathbf{r}_{L-1}.\mathbf{v}_i \geq 0, \mathbf{r}_L.\mathbf{v}_i \geq 0\} \\
c_2 &= \{i \in \mathcal{S} : \mathbf{r}_1.\mathbf{v}_i \geq 0, ..., \mathbf{r}_{L-1}.\mathbf{v}_i \geq 0, \mathbf{r}_L.\mathbf{v}_i < 0\} \\
c_3 &= \{i \in \mathcal{S} : \mathbf{r}_1.\mathbf{v}_i \geq 0, ..., \mathbf{r}_{L-1}.\mathbf{v}_i < 0, \mathbf{r}_L.\mathbf{v}_i \geq 0\} \\
&... \\
c_{2^L} &= \{i \in \mathcal{S} : \mathbf{r}_1.\mathbf{v}_i < 0, ..., \mathbf{r}_{L-1}.\mathbf{v}_i < 0, \mathbf{r}_L.\mathbf{v}_i < 0\}.
\end{aligned}
\tag{2.6}
$$

In this way, the FAPs are partitioned into clusters. Denote the objective function value of (2.4) by Z_{OPT} and denote the objective function value of (2.5), after performing the Randomized Rounding technique, by Z_{RR}. Given a constant $\alpha < 1$ that is

function of the number of vectors L, it has been shown in [28] that $Z_{RR} \geq \alpha Z_{OPT}$. In other words, if we solve the formulation in (2.5) and then obtain a feasible solution to (2.4) by using the Randomized Rounding technique, we obtain an $\alpha-$approximation algorithm where, the resulting solution Z_{RR} is within a factor of α of the value of the optimal solution Z_{OPT}. The chances that the solution Z_{RR} becomes very close to the optimal solution Z_{OPT} increase as the entire procedure is repeated independently, according to **Algorithm 1**, t number of times [32]. As will be shown in the **Appendix**, for a suitably selected value of t, the complexity of **Algorithm 1** is mainly contributed by step 3.

Algorithm 1 : SDP-based correlation clustering algorithm

1: **Input:** $\mathbf{W}^+$ and w_n^-
2: **Output:** Candidate Cluster Configuration for given w_n^-
3: Solve SDP problem (2.5) $\rightarrow \mathbf{X}$
4: Rewrite $\mathbf{X} = \mathbf{V}^T\mathbf{V}$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_S]$
5: **for** $i = 1 : t$ **do**
6: Generate L random vectors $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_L$ such that $2^L \geq S$
7: Map femtocells into clusters according to (2.6)
8: Translate the mapping into a solution X_{int}
9: Calculate objective function value (2.4) using X_{int}
10: **end for**
11: Choose a feasible clustering solution that has the highest objective function value in (2.4) $\rightarrow$ Candidate Cluster Configuration

To estimate the complexity of **Algorithm 1**, we need to calculate the complexity of the following:

- Step 3 involves solving (2.5), which is an SDP. A standard SDP problem can

be written as [33]:

$$\max_{X} \operatorname{Tr}(CX) = \sum_{i=1}^{p} \sum_{j=1}^{p} c_{i,j} x_{i,j} \tag{2.7}$$

subject to

$$\operatorname{Tr}(A_i X) \trianglerighteq_i b_i, \forall i = 1, ..., r$$

$$X \succeq 0$$

where $C, X, A_i \in \mathbb{S}_+^{p \times p}$ and $b_i \in \mathbb{R}$. Also, " $\trianglerighteq_i$ " can represent either " $\geq$ ", " $=$ " or " $\leq$ " for each i. Most of the convex optimization toolboxes use the interior-point algorithm. Hence, the worst-case complexity of solving (2.5) is $\mathcal{O}(\max\{r, p\}^4 p^{1/2} \log(1/\epsilon))$. By analyzing (2.5), we obtain the following: $\mathbf{X} \in \mathbb{S}_+^{S \times S}$, hence $p = S$. We have S equality constraints for C1, $\binom{S}{2}(S-2) = S(S-1)(S-2)/2!$ inequality constraints for C2, S inequality constraints for C3, and $S(S+1)/2$ inequality constraints for C4 by exploiting the symmetry of X. Hence, $r = \mathcal{O}(S^3)$. The overall complexity of solving (2.5) then becomes $\mathcal{O}((S^3)^4 S^{1/2} \log(1/\epsilon))$. It is worth mentioning that this complexity is pessimistic as it does not take into account the sparsity in the data matrices A_i which can decrease the complexity significantly and speed up the solution process [34], [35].

- Step 4 involves an eigen decomposition for the matrix X which has a complexity $\mathcal{O}(S^3)$ [36].

- Finally, steps 5-10 have a complexity of $\mathcal{O}(t)$.

With an appropriate value for t, the major complexity for this algorithm, therefore, comes from step 3.

Algorithm to reduce the range of the penalty term

Solving the correlation clustering problem for each value $w_n^- \in \mathbf{w}^-$ requires solving $\frac{S^2-S}{2} \approx \mathcal{O}(S^2)$ problems. These problems share the same structure and the set of constraints but differ only in the coefficient values in the objective function. This type of analysis lies under the topic of parametric and post-optimality analysis in integer linear problems [37] and [38]. However, solving $\mathcal{O}(S^2)$ problems poses a computation burden on the FGW.

When solving those problems, it is observed that for some weights $w_L^-, w_H^- \in \mathbf{w}^-$, the same solution in terms of the integer variables can be obtained. According to this observation, a heuristic rule, proposed in [39], states that if the same integer variables are obtained for solving the problem at w_L^- and w_H^-, then it can be assumed that the same integer variables will be optimal for all $w_n^- \in \left[w_L^-, w_H^-\right]$. Hence, it is not necessary to solve for all values w_n^- between w_L^- and w_H^-. Based on that heuristic rule, we propose **Algorithm 2**. The idea of this algorithm is to have two weights w_L^- and w_H^- each time and to compare the resulting integer variables at the two weights. If they are the same, then it is not necessary to inspect the weights in between w_L^- and w_H^-. If they have different solutions, then the interval between w_L^- and w_H^- is reduced and so on. In the algorithm, a matrix $SolMatrix$ is maintained to store the solution corresponding to each weight w_n^- in order not to solve for the same weight more than once. We also have the variables $SolL$ and $SolH$, which are used to store the resulting solution of (2.5) corresponding to w_L^- and w_H^-, respectively. In addition, the variable $HopRatio$ is an optimization variable that controls the interval length between w_L^- and w_H^-. Its optimal value for the least number of necessary computations is obtained from the numerical results. The algorithm terminates when all the elements of $\mathbf{w}^-$ are

inspected. The efficiency of the algorithm is tested by using the following measure:

$$\text{Reduction in computation } (\%) = \left[\frac{\left(\frac{S^2-S}{2} - counter \right)}{\left(\frac{S^2-S}{2} \right)} \right] \times 100 \tag{2.8}$$

where *counter* is a variable that keeps track of how many optimization problems have been solved. Thus, the larger the computation reduction, the better the algorithm efficiency.

Algorithm 2 : Penalty term range reduction algorithm

Input: $\mathbf{W}^+$, $\mathbf{w}^-$ and *HopRatio*
Output: Candidate Cluster Configuration $\forall w_n^- \in \mathbf{w}^-$
Initialize: $counter = 0$
IndexL=1, w_L^-=$\mathbf{w}^-$(*IndexL*), *IndexEnd*=length($\mathbf{w}^-$)
For given w_L^-, solve SDP Problem (2.5) using **Algorithm 1** → *SolL* → Update *SolMatrix*
Increment *counter*
$IndexH = \left\lfloor \frac{(IndexEnd - IndexL)}{HopRatio} \right\rfloor + IndexL$
while all points in $\mathbf{w}^-$ are not solved yet **do**
    w_H^-=$\mathbf{w}^-$(*IndexH*)
    if Current w_H^- is solved **then**
        Get *SolH* from *SolMatrix*
    else
        For given w_H^-, solve SDP Problem (2.5) using **Algorithm 1** → *SolH* → Update *SolMatrix*
        Increment *counter*
    end if
    if *SolH*=*SolL* & $IndexL \neq IndexH$ **then**
        Mark all points between IndexL and IndexH as solved
        $IndexL = IndexH$
        $IndexH = \left\lfloor \frac{(IndexEnd - IndexH)}{HopRatio} \right\rfloor + IndexH$
    else
        if $IndexH = IndexL + 1$ **then**
            Find the first unsolved point and quit if no points are found
            Take the preceding point as IndexL with the corresponding solution as *SolL*
            $IndexH = \left\lfloor \frac{(IndexEnd - IndexL)}{HopRatio} \right\rfloor + IndexL$
        else if $IndexH = IndexL$ **then**
            $IndexH = IndexL + 1$
        else
            $IndexH = \left\lfloor \frac{(IndexH - IndexL)}{HopRatio} \right\rfloor + IndexL$
        end if
    end if
end while

2.6 Sub-channel and Power Allocation

After obtaining the list of candidate cluster configurations, the FGW sends these configurations one by one to the femtocells through the wired backhaul. The femtocells are organized into disjoint groups c_l. Within each cluster c_l, a FAP will be elected as a **CH** and will be responsible for sub-channel and power allocation for all FAPs in the cluster. For each cluster c_l, we require to maximize the sum-rate of all femtocells within the cluster, given that there are interference constraints for FUEs in neighboring clusters as well as MUEs. In addition, we have a data rate requirement for each femtocell. Thus, the **CH** within each cluster c_l solves the following optimization problem:

$$\max_{\Gamma^n_{k_s,s}, P^n_{k_s,s}} \sum_{s \in c_l} \sum_{n=1}^{N} \Gamma^n_{k_s,s} \Delta f \log_2 \left(1 + P^n_{k_s,s} \gamma^n_{k_s,s}\right) \tag{2.9}$$

subject to

$$\text{C1} : \sum_{n=1}^{N} \Gamma^n_{k_s,s} \Delta f \log_2 \left(1 + P^n_{k_s,s} \gamma^n_{k_s,s}\right) \geq R_s, \forall s$$

$$\text{C2} : \sum_{n=1}^{N} \Gamma^n_{k_s,s} P^n_{k_s,s} \leq P_{s,max}, \forall s$$

$$\text{C3} : \Gamma^n_{k_s,s} P^n_{k_s,s} g^n_{k_B,s} \leq \zeta^n_{k_B}, \forall n, s$$

$$\text{C4} : \Gamma^n_{k_s,s} P^n_{k_s,s} g^n_{k_j,s} \leq \zeta^n_{k_j}, \forall n, s \in c_l, j \notin c_l$$

$$\text{C5} : \sum_{s \in c_l} \Gamma^n_{k_s,s} = 1, \forall n.$$

C3 and C4 are interference constraints for neighboring MUEs k_B served by macrocell B and FUEs k_j served by femtocells j in a neighboring cluster, respectively. The reference user concept [40] is applied, where the interference constraints are for the MUEs and FUEs having the strongest channel gain to the target femtocell s. It is

worth mentioning that those UEs might not be actually co-channel ones. Hence, although this is a pessimistic approach that can lead to a reduction in the achievable data rates, the actual co-channel UEs are automatically protected. Each FAP keeps a list of reference users and sends this information to its **CH**. A FAP i can determine a reference FUE k_j based on the channel gain to its serving FAP j. Since the distances between a FAP and its FUE is small, the channel gain between a FAP i and an FUE k_j in another FAP j is approximated by the channel gain between the two FAPs i, j which can easily be determined due to relatively fixed FAP positions. For the reference MUE, an approach similar to that in [40] can be followed, where the authors discussed methods to estimate the channel gain between a FAP i and an MUE k_B served by a macrocell B. The interference constraints in C3 and C4 can be further simplified to a single constraint by picking the one of them having the higher channel gain to the target femtocell s. In this way, the interference constraints can be rewritten as

$$\Gamma^n_{k_s,s} P^n_{k_s,s} \leq \min\left(\frac{\zeta^n_{k_B}}{g^n_{k_B,s}}, \frac{\zeta^n_{k_j}}{g^n_{k_j,s}}\right) = \frac{\zeta^n_k}{g^n_{k,s}}, \forall n, s \in c_l. \tag{2.10}$$

A user k now is either a neighboring MUE or an FUE whichever has higher channel gain to the target femtocell s. Since problem (2.9) is an MINLP, joint sub-channel and power allocation is computationally intractable. Therefore, we follow the traditional method of solving the problem in two phases iteratively as follows [40], [41], [42]:

- *Phase 1*: For a given power allocation, the **CH** performs sub-channel allocation. For each femtocell on each sub-channel, $P^n_{k_s,s}$ is initialized as the minimum of either $\frac{P_{s,max}}{N}$ or $\frac{\zeta^n_k}{g^n_{k,s}}$. The idea is to keep power as uniform as possible and at the same time not to violate the interference constraints.

- *Phase 2*: Given the resulting sub-channel allocation, the **CH** performs power allocation.

2.6.1 Sub-channel Allocation

Sub-channel allocation by branch and bound

For a given power allocation, we have an ILP that can be optimally solved by using the BnB technique. BnB is guaranteed to find the optimal sub-channel allocation but its complexity in the worst-case is as high as that of exhaustive search.

Sub-channel allocation by heuristic

In this section, we propose a low-complexity heuristic scheme for sub-channel allocation. If we reformulate the objective function and data rate constraint in (2.9) to $\Delta f \log_2\left(1+\Gamma^n_{k_s,s}P^n_{k_s,s}\gamma^n_{k_s,s}\right)$, we have the expressions that are equivalent to the original ones for binary values of $\Gamma^n_{k_s,s}$ [43]. If we further relax $\Gamma^n_{k_s,s}$ to take any value $[0,1]$, we have a non-linear program that is convex in $\Gamma^n_{k_s,s}$ [44].

By applying the KKT conditions, we have the following formula for sub-channel allocation among femtocells:

$$\Gamma^n_{k_s,s} = \left[\frac{(1+\lambda_{1,s})\,\Delta f}{\ln 2\left(\lambda_{2,s}P^n_{k_s,s}+\lambda^n_{3,s}P^n_{k_s,s}g^n_{k,s}-\lambda^n_4\right)} - \frac{1}{P^n_{k_s,s}\gamma^n_{k_s,s}}\right]^+ \tag{2.11}$$

where $\lambda_{1,s}$, $\lambda_{2,s}$, $\lambda^n_{3,s}$, and λ^n_4 are the Lagrange multipliers associated with the data rate constraints, total power budget, interference constraints, and exclusion constraints, respectively. Therefore, we can deduce that a sub-cahannel n is better be allocated to a user k_s in femtocell s having good SINR value $P^n_{k_s,s}\gamma^n_{k_s,s}$ and having

low channel gain $g^n_{k,s}$ to a reference user (hence, low interference to this user).

Based on this rule, we propose **Algorithm 3** by constructing the following ratio for each femtocell s on sub-channel n: $\frac{P^n_{k_s,s}\gamma^n_{k_s,s}}{g^n_{k,s}}$. A sub-channel n is allocated to the link between user k_s and femtocell s based on the following criterion:

$$k_s(n) = \arg\max_s \left(\frac{P^n_{k_s,s}\gamma^n_{k_s,s}}{g^n_{k,s}} \right) \tag{2.12}$$

where $k_s(n)$ denotes sub-channel n assigned to FUE k_s. This approach maximizes the cluster sum-rate; however, we may have some femtocells with no sub-channels allocated, leading to an infeasible solution to the original problem. Hence, an additional step is required to fix this issue. Given the list of unsatisfied femtocells (i.e., femtocells with no sub-channels allocated) and satisfied femtocells (i.e., femtocells with allocated sub-channels), the **CH** picks up the best sub-channel (for an unsatisfied femtocell) from the most satisfied femtocell and reallocates it to the unsatisfied femtocell. The procedure continues until all unsatisfied femtocells are allocated a sub-channel. Based on the fact that the FUE is close to its FAP (and hence has a good channel condition), the power allocation step can satisfy its data rate requirement later. In the algorithm, we denote by $\mathcal{N}_s$ the set of sub-channels allocated to a femtocell s defined as $\mathcal{N}_s = \left\{n \in \mathcal{N} : \Gamma^n_{k_s,s} = 1\right\}$. The matrix $\mathbf{\Gamma}$ is a binary matrix of dimension $c_l \times N$ that holds the sub-channel allocation indicators for all femtocells $s \in c_l$. The sets $\mathcal{S}$ and $\mathbf{U}$ are the sets of satisfied femtocells and unsatisfied femtocells, respectively, defined as follows: $\mathcal{S} = \left\{s \in c_l, \exists n \in \mathcal{N} : \Gamma^n_{k_s,s} = 1\right\}$ and $\mathbf{U} = \left\{s \in c_l : \Gamma^n_{k_s,s} = 0, \forall n \in \mathcal{N}\right\}$.

The complexity of this algorithm is as follows: steps $2-4$ have a complexity of $\mathcal{O}(FN)$ and steps $6-10$ have a complexity of $\mathcal{O}(S^2 + SN)$. Hence, the overall complexity is of $\mathcal{O}(S^2 + SN)$.

Algorithm 3 : Heuristic sub-channel allocation for c_l

Input: Ratio $\left(\frac{P^n_{k_s,s}\gamma^n_{k_s,s}}{g^n_{k,s}}\right) \forall s \in c_l, n$
for $n = 1 : N$ **do**
    Allocate sub-channel n to user k_s in femtocell s according to (2.12)
end for
Put all unsatisfied femtocells in set $\mathbf{U}$, satisfied femtocells in set $\mathcal{S}$, and define $\mathbf{\Gamma}$ as the sub-channel allocation matrix
while $\mathbf{U} \neq \emptyset$ **do**
    Find femtocell $s' \in \mathcal{S}$ with maximum number of allocated sub-channels
    CH allocates sub-channel n to femtocell $s \in \mathbf{U}$ such that $n = \arg\max_{n \in \mathbf{N_{s'}}} \left(\frac{P^n_{k_s,s}\gamma^n_{k_s,s}}{g^n_{k,s}}\right)$
    Set $\Gamma^n_{k_s,s} = 1$, $\Gamma^n_{k_{s'},s'} = 0$ and update $\mathbf{\Gamma}, \mathbf{U}$, and $\mathcal{S}$
end while

2.6.2 Power Allocation

Given the sub-channel allocation, the **CH**s can perform power allocation in an optimal manner since it is a non-linear convex[4] problem that can be efficiently solved by using the interior point method [44] with a computational complexity of $\mathcal{O}(N^3)$ [43].

2.6.3 Iterative Sub-channel and Power Allocation Algorithm

Algorithm 4 summarizes the iterative sub-channel and power allocation steps. Given the current interference measurements $Interference^{(0)}$ reported by the FUEs, the data rate requirements and the candidate cluster configuration, the **CH**s perform sub-channel and power allocation iteratively. In each iteration, the **CH**s perform resource allocation simultaneously relying on interferences generated by the other FAPs in the previous iteration [45]. This approach was shown to converge faster than the sequential operation [46]. In some cluster c_l in the z^{th} iteration, each FAP reports the interference measurements by its FUE in the $(z-1)^{th}$ iteration to its **CH**. For

[4]We render the power allocation problem convex by considering the interference term in the SINR expression to be due to other FAPs transmissions from the previous iteration.

all FAPs, the power is initialized as the minimum of $\left(\frac{P_{s,max}}{N}, \frac{\zeta_k^n}{g_{k,s}^n}\right)$. The sub-channel allocation is done for the given power initialization. Power is then allocated knowing the resulting sub-channel allocation. In the z^{th} iteration, knowing the resulting sub-channel and power allocation and the interference measured in the $(z-1)^{th}$ iteration, the data rate can be estimated. The sub-channel and power allocation iterations proceed until the data rate converges or the maximum number of iterations is reached.

Algorithm 4 : Sub-channel and power allocation

1: **Input:** $Interference^{(0)}$, data rate requirement, and candidate cluster configuration
2: **while** diff $\geq \epsilon$ and $z <$max_iteration **do**
3: All **CH**s do simultaneously:
4: Initialize $P_{k_s,s}^n$ as min $\left(\frac{P_{s,max}}{N}, \frac{\zeta_k^n}{g_{k,s}^n}\right) \forall n, s$
5: Do sub-channel Allocation $\rightarrow$ $new\_ch\_allocation^{(z)}$
6: For given $new\_ch\_allocation^{(z)}$, do power allocation $\rightarrow$ $new\_power\_allocation^{(z)}$
7: For given $new\_ch\_allocation^{(z)}$, $new\_power\_allocation^{(z)}$ and $Interference^{(z-1)}$, estimate $DR^{(z)}$ $\forall s$
8: Perform Interference measurements $\rightarrow$ $Interference^{(z)}$
9: diff $=\mid DR^{(z)} - DR^{(z-1)} \mid$
10: Increment z
11: **end while**

2.7 Performance Evaluation

2.7.1 Parameters

We evaluate the system performance with the proposed algorithms through extensive simulations under various topologies and scenarios. In our numerical results, each femtocell has 1 FUE closely located to its FAP. We consider an indoor area with a dense deployment of femtocells (in terms of density per square metre). 4 sub-channels are available for the femtocell and macrocell network. 4 MUEs exist indoor and are

assumed to be served by their macrocell. The macrocell transmits with uniform power on the 4 sub-channels. To model the propagation environment, the channel models from [47] are used. The channel gains include path-loss and shadowing. For path-loss between a femtocell and its FUE, $PL = 38.46 + 20 \log R$ and for path-loss between a femtocell and a general UE, $PL = 38.46 + 20 \log R + qL_{iw}$, where R is the distance between a FAP and the FUE and qL_{iw} accounts for losses due to walls. For path-loss between the macrocell and a general UE existing indoor, $PL = 15.3 + 37.6 \log R + L_{ow}$, where L_{ow} accounts for losses due to inner and outer walls.

Table 2.1 shows the parameters used for the numerical results. The worst-case initial interference ($Interference^{(0)}$) is assumed, where all FAPs are assumed to be transmitting on all sub-channels with uniform power. A small value for the data rate requirement is used to guarantee the feasibility of (2.9) especially since the worst-case interference scenario is assumed in the numerical results. The data rate constraint is imposed to guarantee that no FAP ends up with no sub-channels allocated. However, it will be shown in the numerical results that the average achievable data rate per FAP is in the order of Mbps. The numerical results are obtained and averaged for different realizations, where in each realization some or all of the followings are varied: positions of FAPs and UEs and channel gains. The channel realizations leading to infeasible optimization problems are excluded from the numerical results. We use "*fmincon*" function in MATLAB [48] to solve the power allocation problem.

The following measures are used to evaluate the performance of the proposed algorithms:

- the average data rate achieved in each femtocell is defined as: $\frac{\sum_{s=1}^{S} \sum_{n=1}^{N} \Gamma_{k_s,s}^{n} \Delta f \log_2\left(1+P_{k_s,s}^{n} \gamma_{k_s,s}^{n}\right)}{S}$.
- the average femtocell transmit power is defined as: $\frac{\sum_{s=1}^{S} \sum_{n=1}^{N} P_{k_s,s}^{n}}{S}$.

- the average interference at an FUE k_s on a sub-channel n due to downlink femtocell transmissions $j \neq s$ is defined as: $\frac{\sum_{s=1}^{S}\left(\sum_{n=1}^{N}\Gamma_{k_s,s}^{n}\left(\sum_{j=1,j\neq s}^{S}P_{k_j,j}^{n}g_{k_s,j}^{n}\right)\right)}{FN}$. The idea is to estimate the average interference to an FUE k_s on its allocated sub-channels.

For comparison, an uncoordinated scheme is used as a benchmark, where a femtocell has access to the entire spectrum but no transmission coordination exists with neighboring femtocells. In addition, the closest scheme in spirit to our clustering scheme is the similarity-based clustering in [20]. In this scheme, clustering and resource allocation are performed separately. Also, there is no notion of optimality when specifying the cluster size. The performance results for this scheme are shown to demonstrate the efficacy of our proposed framework.

Table 2.1: Parameters

Parameter	Value
Carrier frequency	2.0 GHz
Available sub-channels	4
Sub-channel bandwidth, Δf	180 KHz
No. of FUEs per femtocell	1
Distance between FUE and FAP	1 m
Data rate requirement, R_s	10^2 bps
Noise power density, N_o	-174 dBm/Hz
Macrocell radius	200 m
Distance between indoor building and MBS	100 m
Number of MUEs	4
Standard deviation of shadowing between macrocell and indoor UE	10 dB
Standard deviation of shadowing between femtocell and its FUE	4 dB
Standard deviation of shadowing between femtocell and another UE	8 dB
Outdoor wall loss, L_{ow}	30 dB

2.7.2 Numerical Results for a Small scenario

We have 4 femtocells existing in an area of dimensions $30m \times 30m$. In this scenario, the number of FAPs is equal to the number of available sub-channels. The idea here is to verify the fact that by sacrificing a part of the spectrum, an advantage in terms of increased data rate can be achieved due to the decrease in co-tier interference. The positions of the FAPs and UEs are fixed; however, the channel gains are varied. The numerical results are obtained by taking the average of 25 different channel gain realizations. The sub-channel and power allocation is done iteratively according to **Algorithm 4**, wherein the sub-channel allocation is done by using BnB. The target here is to study the performance of correlation clustering using both BnB and the low-complexity SDP approach using **Algorithm 1** are compared to the optimal clustering obtained by exhaustive search.

Fig. 2.3 shows the variation in femtocell data rate with the interference threshold. We have $P_{s,max} = 30$ mW, $P_{macro} = 20$ W, $L_{ow} = 30$ dB, and $qL_{iw} = 5$ dB. We observe that clustering offers a higher average data rate in comparison to the uncoordinated scheme. Also, correlation clustering has a performance that is close to the optimal solution and better than the uncoordinated approach. Correlation clustering reduces the search space for the optimal cluster configuration with the drawback of the possibility of missing the optimal cluster configuration. It is also clear that, with enough number of iterations for randomized rounding, correlation clustering using **Algorithm 1** gives the same performance as that achieved by the BnB method. The average data rate achieved with clustering using the similarity-based technique from [20] is shown as well. It is observed that the performance of this scheme can be even worse than that of the uncoordinated scheme.

Fig. 2.4 shows that clustering allows FAPs to transmit more power, hence increas-

ing the achievable data rate as FAPs now do not have to avoid causing interference to nearby FUEs in neighboring femtocells because they are possibly now in the same cluster.

Fig. 2.5 shows that in the clustered schemes the FUEs suffer lower interference due to transmissions from other FAPs. Since in an uncoordinated scheme a generic FAP is using the entire spectrum, the corresponding FUE experiences interference in the entire band, and hence suffers from a higher interference.

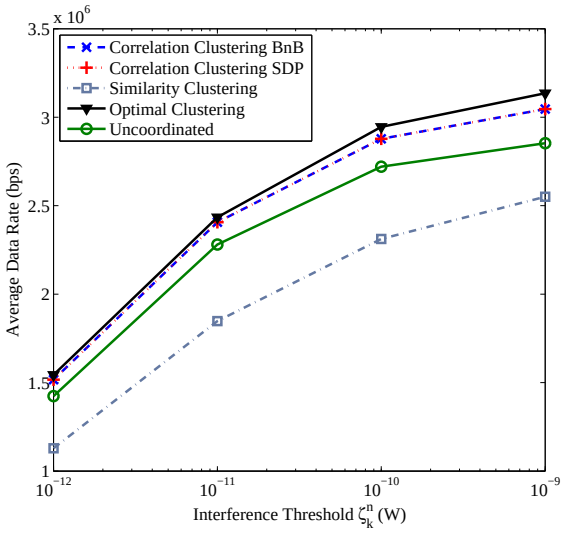

Figure 2.3: Average data rate vs. interference threshold ζ_k^n.

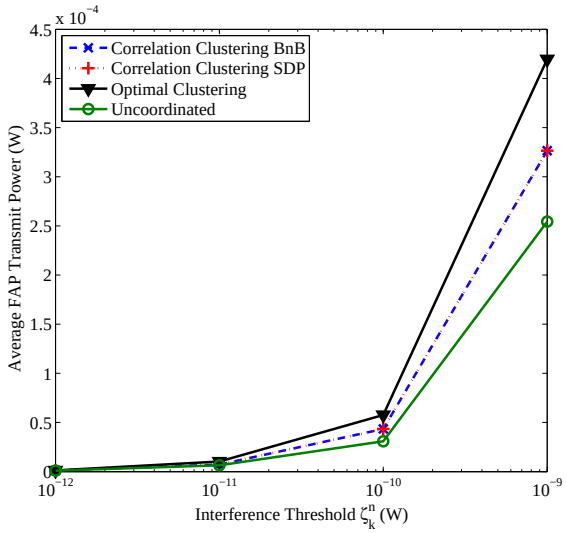

Figure 2.4: Average transmit power vs. interference threshold ζ_k^n.

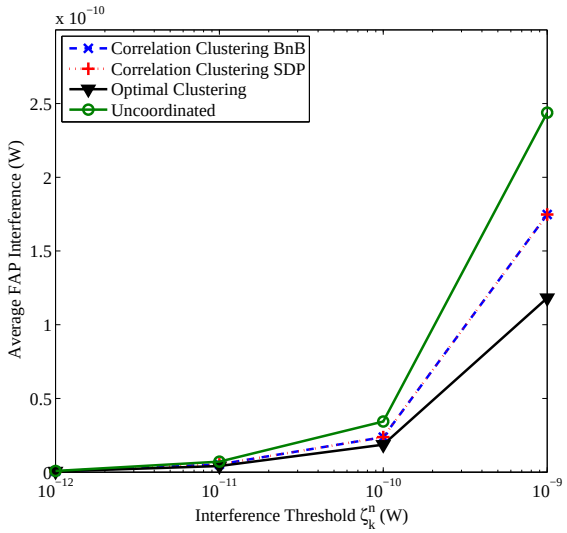

Figure 2.5: Average FAP interference vs. interference threshold ζ_k^n.

Fig. 2.6 shows the variation in femtocell data rate with the inner wall losses. We have $P_{s,max} = 30$ mW, $P_{macro} = 20$ W, $L_{ow} = 30$ dB, and $\zeta_k^n = 10^{-10}$ W. In Fig.

2.6, we notice that as the inner wall losses increase, the co-tier interference decreases, hence, the achieved data rate increases as well. We also notice that the gain due to clustering, relative to the uncoordinated scheme, decreases as the inner wall losses increase since each femtocell starts to be more isolated. When a femtocell is well isolated, it is more beneficial to act in an uncoordinated manner and to make use of the full spectrum. Also, similarity-based clustering, which does not consider the trade-off between interference and bandwidth, is observed to perform poorly compared to the other schemes.

Fig. 2.7 shows that an FUE in the uncoordinated scheme suffers from the highest co-tier interference when compared to the clustering-based schemes. As the inner wall loss increases, the average co-tier interference at an FUE decreases for the uncoordinated scheme. The average FAP interference for the clustering schemes starts to increase at high inner wall loss because, in this case, each FAP becomes well isolated and each FAP can achieve the best performance when it performs resource allocation independently (i.e., in an uncoordinated manner).

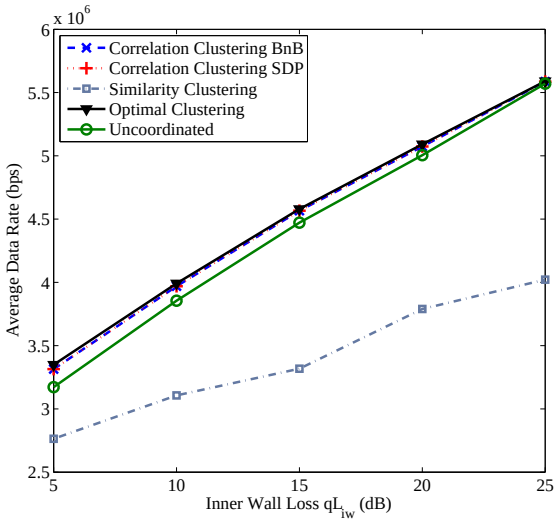

Figure 2.6: Average data rate vs. inner wall loss qL_{iw}.

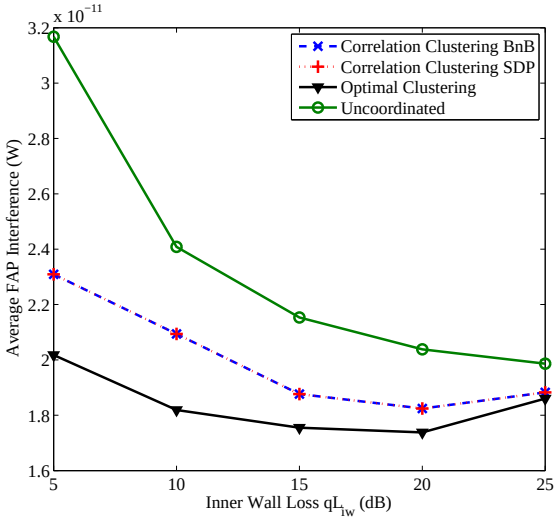

Figure 2.7: Average FAP interference vs. inner wall loss qL_{iw}.

Fig. 2.8 shows the variation in fetmocell data rate with the macro cell power. We have $P_{s,max} = 30$ mW, $L_{ow} = 30$ dB, $qL_{iw} = 5$ dB, and $\zeta_k^n = 10^{-10}$ W. It is clear from

Fig. 2.8 that as the macrocell power increases, the cross-tier interference increases and hence, the achieved data rate decreases. Although the cross-tier interference becomes more dominant, clustering is still beneficial.

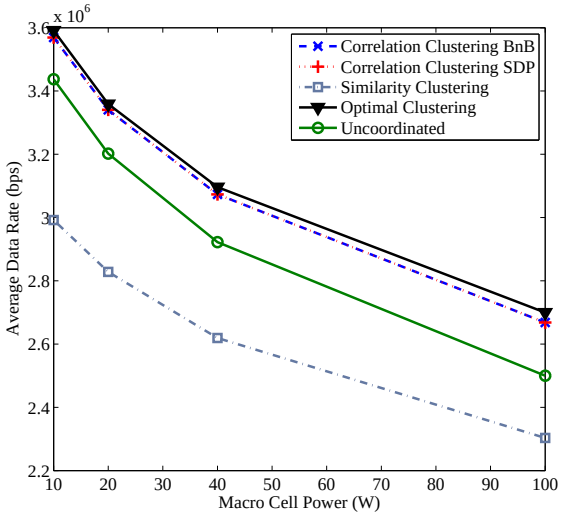

Figure 2.8: Average data rate vs. macro cell transmission power.

2.7.3 Numerical Results for a Big Scenario

In this scenario, we are interested now in the case, where the number of femtocells is larger than the number of available sub-channels. The sub-channel and power allocation is done iteratively according to **Algorithm 4**, wherein sub-channel allocation is done by using BnB. Clustering is done sub-optimally according to **Algorithm 1**. Fig. 2.9 shows the average femtocell data rate vs. the number of femtocells for two indoor area sizes, $(30m \times 30m)$ and $(70m \times 70m)$. We have $P_{s,max} = 30$ mW, $P_{macro} = 20$ W, $L_{ow} = 30$ dB, $qL_{iw} = 5$ dB and $\zeta_k^n = 10^{-10}$ W. The results are obtained for this scenario by taking the average of 100 realizations, where in each realization, the positions of the FAPs and UEs as well as the channel gains are varied. We can no-

tice that as the number of femtocells increases, for fixed number of sub-channels, the co-tier interference increases. Hence, the average achievable data rate for a femtocell decreases. We also notice the gain in average data rate due to clustering relative to the uncoordinated scheme. However, the gain due to clustering decreases as the dimensions of the indoor area increases. This is due to the fact that as the dimensions of the indoor area increase, the femtocells are generally further located from each other, and hence, have lower co-tier interference between each other.

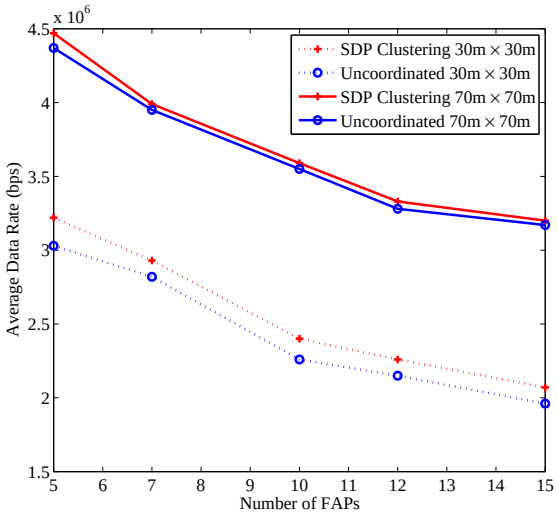

Figure 2.9: Average data rate vs. the number of femtocells.

Performance of penalty term range reduction algorithm

Fig. 2.10 shows the percentage reduction in computation vs. the *HopRatio* for 3 network sizes of 10, 15, and 25 FAPs. The results are obtained and averaged over 100 realizations, where in each realization, the positions of the FAPs as well as the channel gains are varied. As shown in the figure, it is not necessary to go through the entire range of the penalty term to obtain the different cluster configurations.

Also, it is clear that the optimal *HopRatio* is 2. At this value, a 60.5% reduction in computation is possible for a femtocell network of 25 FAPs.

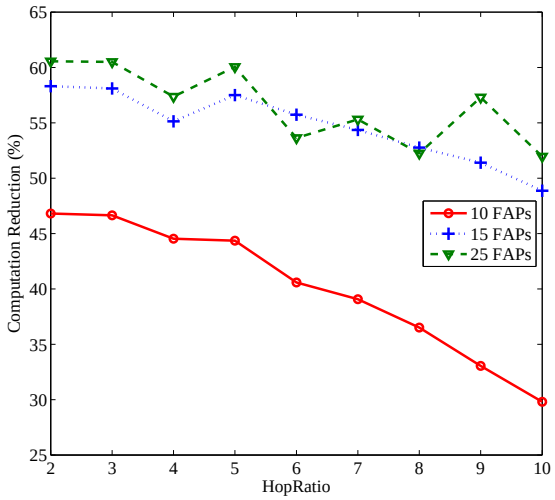

Figure 2.10: Computation reduction percentage vs. the *HopRatio*.

Sub-optimal clustering, sub-channel and power allocation

We study the femtocell network performance with the proposed framework for clustering, sub-channel and power allocation. In this case, clustering is done using **Algorithm 2**. In addition, sub-channel allocation is done both optimally using BnB and sub-optimally using **Algorithm 3**. Fig. 2.11 shows the variation in average data rate in a femtocell with the transmit power of the FAP for 10 FAPs deployed in an area of dimensions $50m \times 50m$. We have $P_{macro} = 20$ W, $L_{ow} = 30$ dB, $qL_{iw} = 5$ dB, and $\zeta_k^n = 5 \times 10^{-12}$ W. The results are obtained and averaged over 25 realizations, where in each realization, the positions of the FAPs and UEs as well as the channel gains are varied. We notice that increasing the FAP transmit power increases the average data rate in a femtocell at the beginning, where the system is power-limited. After

this, the increase in achievable data rate decreases with the FAP transmit power, where the system now is interference threshold-limited. We notice also that allowing clustering is still beneficial compared to the uncoordinated scheme, even when sub-channel and power allocation are performed sub-optimally. We also notice that the heuristic sub-channel allocation scheme yields almost the same performance as the optimal sub-channel allocation using BnB.

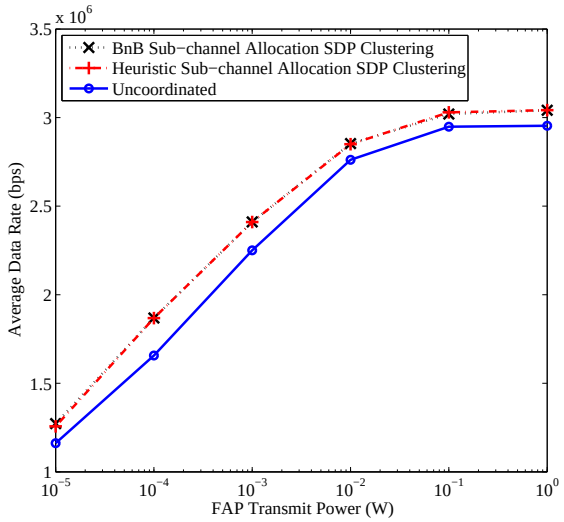

Figure 2.11: FAP average data rate vs. FAP transmit power.

2.8 Conclusion

I have proposed a clustering, sub-channel and power allocation framework to be implemented in a semi-distributed fashion in a two-tier OFDMA cellular network. In this context, the FGW will be responsible for the clustering phase, and then the **CH** (elected from the femtocell group) will be responsible for the sub-channel and power allocation phase. To accomplish this, low-complexity algorithms have been proposed for the clustering, sub-channel and power allocation sub-problems. Through exten-

sive numerical studies, it has been observed that for resource allocation, allowing for coordination among femtocells through clustering is beneficial in comparison to the uncoordinated scheme. Also, the proposed correlation clustering approach, which considers the trade-off between the bandwidth and interference, offers a performance that is very close to that of the optimal clustering; however, with much reduced complexity.

Chapter 3

Tier-Aware Resource Allocation in OFDMA Macrocell-Small Cell Networks

3.1 Introduction

In this Chapter, I formulate the RA problem for a two-tier OFDMA wireless network composed of a macrocell overlaid by small cells. The objective of the macrocell is to allocate resources to its macro UEs (MUEs) to satisfy their data rate requirements. In addition, knowing about the existence of small cells, the macrocell allocates the radio resources (i.e., sub-channel and power) to its MUEs in a way that can sustain the highest interference level from the small cells. For this reason, I formulate an optimization problem for the macrocell with an objective that is different from those in the traditional RA problems. The macrocell maximizes the sum of the interference levels it will be capable of tolerating from the small cells tier. Now, since small cells create dead zones around them in the downlink, the MUEs should be protected

against transmissions from the small cells [49], [50]. Hence, knowing about the maximum allowable interference levels for MUEs, the small cells perform RA by solving an optimization problem whose objective function combines both the AC and the consumed bandwidth (i.e., number of allocated sub-channels). The objective of the small cell tier is to admit as many small cell UEs (SUEs) as possible at their target data rates and consume the minimum amount of bandwidth. Again, this follows the same notion of tier-awareness by leaving as much bandwidth as possible for other network tiers (e.g., for device-to-device (D2D) communication). For this, an optimization problem is formulated for the small cell tier with the aforementioned objective, given the QoS requirements of SUEs and the interference constraints for the MUEs. Dual decomposition is used to have a decentralized RA and AC problem by decomposing the optimization problem into sub-problems for each small cell to solve. For this, only local channel gain information is used along with some coordination with the Home eNB Gateway (HeNB-GW) [26].

The rest of this Chapter is organized as follows. Section 3.2 reviews the related work. Section 3.3 presents the system model and assumptions for this work. In Section 3.4, the optimization problems are formulated for both the macrocell tier and the small cell tier, followed by the use of dual decomposition to have a decentralized operation. Numerical results are discussed in Section 3.5 and finally Section 3.6 concludes the work done.

3.2 Related Work

The RA problem in OFDMA multi-tier cellular networks has been extensively studied in the literature. The authors in [40] studied the RA problem in a multi-tier cellular network to maximize the sum-throughput subject to simple power budget and sub-

channel allocation constraints. However, no QoS constraints were imposed. In [51], the RA problem in a femtocell network was modeled, with interference constraints for MUEs, in order to achieve fairness among femtocells. No QoS constraints, however, were imposed for femtocell users. In [52], the RA problem in a two-tier macrocell-femtocell OFDMA network was modeled as a Stackelberg game, where the macrocell acts as the leader and the femtocells act as the followers. However, no interference constraints for MUEs were considered. Also, no QoS constraints were imposed for femtocells. Reference [53] studied the RA problem in a two-tier network composed of macrocells and femtocells which aimed at maximizing the sum-throughput of femtocells subject to total sum-rate constraint for the macrocell. Nevertheless, no QoS constraints were imposed for femtocells. The authors in [54] studied the RA problem with QoS and interference constraints in a two-tier cellular network and used clustering as a technique to reduce overall complexity.

In the above works, either no QoS constraints were imposed or the RA problems with QoS constraints were assumed feasible. In other words, admission control (AC) [55], which is a technique to deal with infeasibility when it is not possible to support all UEs with their target QoS requirements was not studied. The authors in [56] proposed a distributed self-organizing RA scheme for a femtocell only network, with the aim of minimizing the total transmit power subject to QoS constraints. It was shown that minimizing the transmit power (which results in reduced interference) does not necessarily reduce throughput.

Several works in the literature have considered the AC problem. For cellular cognitive radio networks, [57] studied the problem of admission and power control to admit the maximum number of secondary links and maximize their sum-throughput subject to QoS requirements and interference constraints for primary links. However,

power control was done centrally. The authors in [58] considered the problem of admission and power control, where the primary users are guaranteed a premium service rate and the secondary users are admitted (as many as possible) so long as the primary users are not affected. In [59], the authors proposed a joint rate and power allocation scheme with explicit interference protection for primary users and QoS constraints for secondary users, where admission control was performed centrally. However, [57–59] only considered single-channel systems. The authors in [60] studied the problem of joint rate and power allocation with admission control in an OFDMA-based cognitive radio network subject to QoS requirements for secondary users and interference constraints for primary users. However, resource allocation and admission control were performed centrally. In addition, channels were randomly allocated to secondary users.

In relay networks, [61] studied the problem of power allocation in amplify and forward wireless relay systems for different objectives, where admission control was employed as a first step preceding power control. However, only one channel was considered. In addition, power and admission control were done centrally. In [62], a joint bandwidth and power allocation for wireless multi-user networks with admission control in relay networks was proposed for different system objectives. Unequal chunks of bandwidths were allocated. However, the resource allocation was performed centrally.

For a two-tier small cell network, [63] studied joint admission and power control. Small cells are admitted into the network so long as the QoS of macrocell users is not compromised. Admission and power control were performed in a distributed fashion. However, only a single channel system was considered. Reference [64] proposed a distributed admission control mechanism for load balancing among sub-carriers with

multiple QoS classes. In addition, small cells mitigate co-tier and cross-tier interferences using slot allocation of different traffic streams among different sub-carriers. However, no power allocation was performed.

We notice that none of the quoted works considers the interaction between the different network tiers and the consequences of RA decisions of one tier on the other one. In addition, it is desirable to have an RA and AC scheme that is implementable in a distributed fashion in a dense multi-tier OFDMA network.

3.3 System Model, Assumptions, and Resource Allocation Framework

3.3.1 System Model and Assumptions

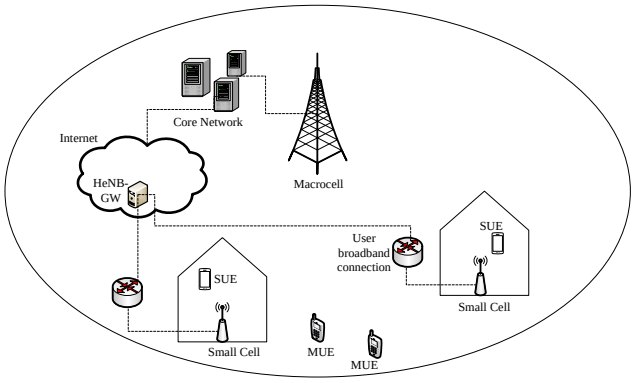

Figure 3.1: Network topology under consideration. Dashed lines indicate back-haul connections.

We consider the downlink of a two-tier cellular network, where a single macrocell, referred to by the index B and with coverage radius R_B, is overlaid with S small cells, as shown in Fig. 4.1. Denote by $\mathcal{S}$ the set of small cells, where $S = |\mathcal{S}|$. A

closed-access scheme is assumed for all small cells, where access to a small cell is restricted only to the registered SUEs. All small cells are connected to the mobile core network. For example, femtocells can connect to the core network by using the DSL or CATV modems via an intermediate entity called the Femto Gateway (FGW) or HeNB-GW [26] which can take part in the resource allocation operation for femtocells.

We denote by $\mathcal{M}$ the set of MUEs served by the macrocell B with $M = |\mathcal{M}|$. Each MUE m has a data rate requirement of R_m. In addition, denote by $\mathcal{F}$ the set of SUEs in the system with $F = |\mathcal{F}|$. Each SUE f has a data rate requirement of R_f. We refer to the set of SUEs associated to small cell s by $\mathcal{F}_s$. We assume that all UEs are already associated with their BSs and that this association remains fixed during the runtime of the resource allocation process. We have $\bigcup_{s=1}^{S} \mathcal{F}_s = \mathcal{F}$ and $\bigcap_{s=1}^{S} \mathcal{F}_s = \varnothing$. All MUEs exist outdoor and all SUEs exist indoor. We have an OFDMA system, where we denote by $\mathcal{N}$ the set of available sub-channels with $N = |\mathcal{N}|$ and Δf is the bandwidth of a sub-channel n. Cochannel deployment is assumed, where the macrocell and all the small cells have access to the same set of sub-channels $\mathcal{N}$ [51–53]. This is foreseen to increase spectral efficiency. Hence, any UE (be it an MUE or an SUE) is allowed to share the same sub-channel. Let $\Gamma_{i,j}^{n}$ be the sub-channel allocation indicator where, $\Gamma_{i,j}^{n} = 1$ if sub-channel n is allocated to UE j served by BS i and 0 otherwise.

The UEs are capable of using two modes of sub-channel allocation, namely, the exclusive mode and the time sharing mode. For the exclusive mode, in a given transmission frame, sub-channel n is used by one UE only. In the time sharing mode, a sub-channel n is allocated to a certain UE a portion of the time. In this way, multiple UEs can time share a sub-channel n in a given transmission frame [65].

Denote by $P_{i,j}^n$ and $g_{i,j}^n$ the allocated power to and the channel gain of the link between BS i and UE j on sub-channel n. Channel gains account for path-loss, log-normal shadowing, and fast fading. The channel gains are assumed to remain static during the resource allocation process. The worst-case received SINR $\gamma_{B,m}^n$ of an MUE m served by macrocell B on a sub-channel n is defined as:

$$\gamma_{B,m}^n = \frac{P_{B,m}^n g_{B,m}^n}{I_m^n + N_o} \tag{3.1}$$

where I_m^n is the maximum tolerable interference level at MUE m on sub-channel n and N_o is the noise power. According to (4.2), the following constraint holds for small cell transmission powers on sub-channel n:

$$\Gamma_{B,m}^n \left(\sum_{s=1}^{S} \sum_{f \in \mathcal{F}_s} \Gamma_{s,f}^n P_{s,f}^n g_{s,m}^n \right) \leq \Gamma_{B,m}^n I_m^n \tag{3.2}$$

where the constraint is active only if sub-channel n is allocated to MUE m, i.e., $\Gamma_{B,m}^n = 1$. Similarly, we can define the received SINR $\gamma_{s,f}^n$ of an SUE f served by small cell s on a sub-channel n as:

$$\gamma_{s,f}^n = \frac{P_{s,f}^n g_{s,f}^n}{\sum_{m=1}^{M} \Gamma_{B,m}^n P_{B,m}^n g_{B,f}^n + N_o}. \tag{3.3}$$

In the denominator of (3.3), the first term represents cross-tier interference from the macrocell, whereas the second term encompasses both co-tier interference from other small cells and the noise power[1].

[1] A similar approach was followed in [66] to account for co-tier interference in densely deployed small cells. In other words, co-tier interference can be accounted for by a rise in the noise power level. This is acceptable due to the wall penetration loss and the relatively low transmission powers of indoor small cell base stations.

3.3.2 Tier-Aware Resource Allocation Framework

Fig. 3.2 describes the RA framework proposed in this Chapter in a given RA time slot. Given the rate requirements for the MUEs, the macrocell starts by allocating resources to its MUEs and specifies the maximum tolerable interference levels on each allocated sub-channel. Those RA results remain fixed throughout the current RA time slot. The macrocell then sends those RA results to the HeNB-GW which broadcasts them to the small cells. The small cells then perform RA and AC for their SUEs. For the resulting resource allocation for small cells, the MUEs perform interference measurements and report them to the macrocell BS. The macrocell BS then updates the HeNB-GW and the iterations repeat until the interference thresholds for all the MUEs are met and the RA and AC converge for all small cells. Then, in a new RA time slot, the whole operation repeats. Note that the iterative behavior of the small cells RA and AC takes place due to the distributed nature of RA and AC in small cells. This iterative behavior, however, does not take place if RA and AC in small cells are performed centrally.

The awareness of the macrocell about the small cell tier is reflected in the way the radio resources are allocated in the macrocell. The macrocell allocates resources to its MUEs in a way that can tolerate the maximum interference possible from the samll cell tier. Note, however, that the minimum rate constraints of all MUEs must be satisfied in the sense that the rate requirement for none of the MUEs is compromised for admitting new SUEs. On the other hand, the awareness of the small cell tier about the existence of other tiers is reflected in the fact that the resource allocation in the small cell tier satisfies the rate requirements of the SUEs using the minimum amount of bandwidth resources.

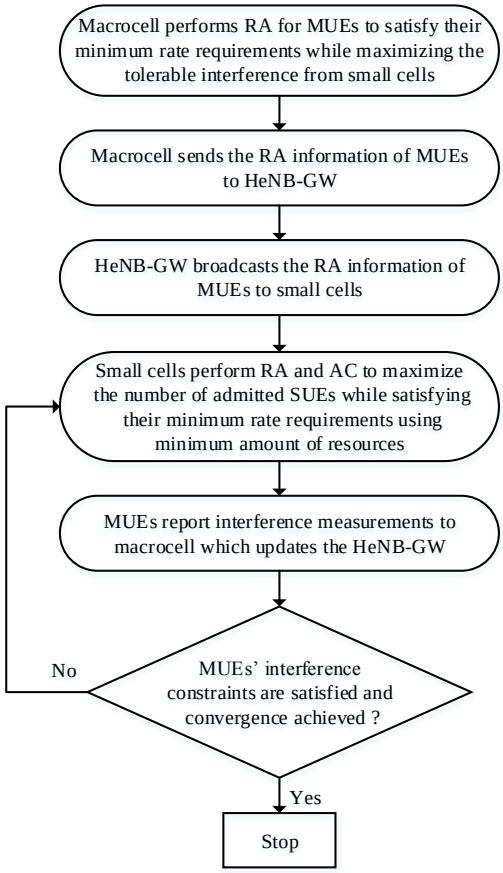

Figure 3.2: The RA framework for the macrocell and the small cells.

3.4 Problem Formulations for Resource Allocation

3.4.1 Problem Formulation for Macrocell

The macrocell is responsible for providing the basic coverage for the MUEs [67]. Hence, the target of the macrocell is to allocate the resources to its MUEs to satisfy their data rate requirements. As will be shown in this section, this task will be accomplished by using different strategies. One strategy would be for the macrocell to maximize the sum of the interference levels it can tolerate from the small cell tier. The

other strategy would be for the macrocell to minimize the total transmission power. As was mentioned previously, we foresee that those different strategies, employed by the macrocell, will impact the performance of small cells differently.

Maximize the sum of tolerable interference levels

One way of performing RA in the macrocell is to allocate resources to the MUEs in a way that maximizes the sum of the maximum tolerable interference levels on the allocated sub-channels. The motivation behind this objective is to allow the maximum possible freedom for the small cell tier in using the sub-channels. In this context, uniform transmit power is assumed on the allocated sub-channels in the macrocell, i.e., $P_{B,m}^n = \frac{P_{B,max}}{N_{ac}}$, where $P_{B,max}$ is the total macrocell power and $N_{ac} \leq N$ is the number of allocated sub-channels. Although the optimal power allocation usually has a water-filling type of solution, it was reported in [68–70] that equal power allocation gives close to the optimal performance, specially at high SINR[2]. Denote by $\mathcal{N}_m$ the set of sub-channels allocated to MUE m. Hence, we can define the following optimization

[2]We will show, subsequently, that each MUE will end up being allocated one sub-channel only. Hence, the high SINR requirement for guaranteeing close to the optimal performance under equal power allocation will be achieved through the enforced data rate requirement.

problem:

$$\max_{\{I_m^n\}} \sum_{m=1}^{M} \sum_{n=1}^{N} I_m^n$$

subject to

$$\text{C1} : \sum_{n \in \mathcal{N}_m} \Delta f \log_2 \left(1 + \frac{P_{B,m}^n g_{B,m}^n}{I_m^n + N_o}\right) \geq R_m, \forall m \in \mathcal{M}$$

$$\text{C2} : \mathcal{N}_i \bigcap \mathcal{N}_j = \varnothing, \ \forall i, j \in \mathcal{M}, i \neq j$$

$$\text{C3} : \bigcup_{m=1}^{M} \mathcal{N}_m \subseteq \{1, 2, ..., N\}$$

$$\text{C4} : I_m^n \leq I_{max}, \forall m \in \mathcal{M}, n \in \mathcal{N}$$

$$\text{C5} : I_m^n \geq 0, \forall m \in \mathcal{M}, n \in \mathcal{N} \tag{3.4}$$

where the objective is to maximize the sum of the tolerable interference levels I_m^n for all MUEs m on all sub-channels n. C1 is the data rate constraint for each MUE m. C2 and C3 indicate that the sets of sub-channels allocated to the MUEs are disjoint (the OFDMA constraint) and constitute the entire set of sub-channels $\mathcal{N}$. C4 is a constraint that sets an upper bound for I_{max} on I_m^n. Recall that if sub-channel n is not allocated to MUE m, then MUE m will not have any restrictions on the level of interference on that sub-channel. In other words, the maximum tolerable interference by MUE m on sub-channel n should ideally be $I_m^n = \infty$. However, to have a finite value for the objective function of (3.4), instead of ∞, we use I_{max}, where I_{max} is an arbitrarily large value. Hence, $I_m^n = I_{max}$ indicates that sub-channel n is not used by MUE m as it has no restrictions on the level of interference on that sub-channel and the achievable rate on that sub-channel will be virtually zero. Finally, C5 indicates that I_m^n should be positive.

In general, (3.4) is an MINLP whose feasible set is non-convex due to C1 and the combinatorial nature of sub-channel allocation. Besides, $P_{B,m}^n$ is unknown as the number of allocated sub-channels N_{ac} is not known yet. However, by carefully inspecting (3.4), some interesting features can be revealed which lead to the possibility of obtaining the optimal solution of (3.4) with polynomial time complexity. We shall assume that (3.4) is always feasible and that in the extreme case, an MUE can have its rate requirement satisfied with one sub-channel only. The last assumption is possible thanks to the fact that the macrocell in our model has a control on the allowable interference level I_m^n on the allocated sub-channel. We shall start first by observing the features for a single MUE system and then will show how to solve (3.4) in a multi-MUE case. The following Lemmas reveal some of the interesting features of (3.4).

Lemma 3.4.1. *At optimality, all data rate constraints C1 hold with equality.*

Proof. Since the objective function in (3.4) is monotonically increasing in I_m^n and C1 is monotonically decreasing in I_m^n, C1 must hold with equality at optimality for all MUEs. □

Lemma 3.4.2. *At optimality, an MUE m is assigned a single sub-channel i with $I_m^i < I_{max}$.*

Proof. To establish this result, we assume first that $P_{B,m}^n = \frac{P_{B,max}}{N}$. Furthermore, we assume that for an MUE m, at optimality, $I_m^n < I_{max}$, $\forall n \in \mathcal{N}$ with an objective function value $\text{Obj}_m = \sum_{n=1}^{N} I_m^n$ for MUE m. However, according to Lemma 3.4.1, the objective function is monotonically increasing in I_m^n, whereas the constraint C1 is monotonically decreasing in I_m^n. Besides, I_{max} has an arbitrarily large value. Therefore, we can decrease the value of I_m^i on a certain sub-channel $i \in \mathcal{N}$ and increase the values

of all other $I_m^j, j \in \mathcal{N}, j \neq i$. In this way, we end up with $I_m^j = I_{max}$, for $j \in \mathcal{N}, j \neq i$. Meanwhile, I_m^i reaches a value I_m^{i*} such that the rate constraint for MUE m is met with equality resulting in a new objective function value $\text{Obj}'_m = ((|\mathcal{N}| - 1)\, I_{max} + I_m^{i*})$. By comparing Obj'_m and Obj_m, we find that we have $(|\mathcal{N}| - 1)$ variables I_m^n which have reached the upper bound value I_{max}, whereas one variable has decreased to I_m^{i*}. Hence, we can deduce that Obj'_m is clearly greater than Obj_m. Hence, the initial assumption of optimality is contradicted. □

Recall that, a value of I_{max} for a certain I_m^n means that sub-channel n is not actually allocated to MUE m. Hence, for the macrocell to maximize the sum of tolerable interference levels, the macrocell will try to set as many I_m^n to I_{max} as possible. In other words, the macrocell will try to render as many sub-channels as possible unallocated. This leads to the fact that a system with N sub-channels and M MUEs, where $M \leq N$, will end up with $N_{ac} = M$ sub-channels only allocated to the M MUEs which leads to a minimal use of the available system bandwidth. Hence, $P_{B,m}^n$ can be further adjusted to $P_{B,m}^n = \frac{P_{B,max}}{N_{ac}} = \frac{P_{B,max}}{M}$ and I_m^{i*} is adjusted accordingly.

Lemma 3.4.3. *The allocated sub-channel i for MUE m is the one with the highest channel gain $g_{B,m}^i, i \in \mathcal{N}$.*

Proof. According to Lemma 3.4.2, at optimality, $\text{Obj}'_m = ((|\mathcal{N}| - 1)\, I_{max} + I_m^{i*})$ for MUE m, where I_m^{i*} is selected such that the achieved data rate on sub-channel i is equal to R_m. Hence, from the rate constraint formula, $I_m^{i*} = \left(\frac{P_{B,m}^i g_{B,m}^i}{2^{R_m/\Delta f} - 1} - N_o\right)$, it is clear that I_m^{i*} is directly proportional to $g_{B,m}^i$. Therefore, to maximize Obj'_m, we need to maximize I_m^{i*}. Hence, MUE m should be allocated sub-channel i such that $i = \arg\max_{n \in \mathcal{N}} g_{B,m}^n$. □

Now, in a multi-MUE system, by using the criterion derived in Lemma 3.4.3 for sub-channel allocation, we might end up with several MUEs assigned the same sub-channel. Besides, different MUEs might have different rate requirements R_m. Hence, we cannot rely on the direct proportionality to the sub-channel gains only for sub-channel allocation. Generally, according to (3.4), a sub-channel n should be allocated to the MUE that is capable of tolerating a higher interference level I_m^n. Since I_m^{i*} was shown to be directly proportional to $\frac{g_{B,m}^i}{2^{R_m/\Delta f}-1}$, we define the following reward for each MUE m on a sub-channel n:

$$\mathcal{X}_m^n = \frac{g_{B,m}^n}{2^{R_m/\Delta f}-1}. \tag{3.5}$$

In order to maximize the sum of tolerable interference, sub-channel allocation can be done then according to the following Theorem.

Theorem 3.4.1. *To maximize the sum of tolerable interference levels, the macrocell can solve the following alternate optimization problem:*

$$\begin{aligned}
&\max_{\{\Gamma_{B,m}^n\}} \sum_{m=1}^{M}\sum_{n=1}^{N} \Gamma_{B,m}^n \mathcal{X}_m^n \\
&\text{subject to} \\
&C1: \sum_{n=1}^{N} \Gamma_{B,m}^n = 1,\ \forall m \in \mathcal{M} \\
&C2: \sum_{m=1}^{M} \Gamma_{B,m}^n \leq 1,\ \forall n \in \mathcal{N} \\
&C3: \Gamma_{B,m}^n \in \{0,1\},\ \forall m \in \mathcal{M}, n \in \mathcal{N}
\end{aligned} \tag{3.6}$$

where the objective in (3.6) is to maximize the sum of the allocated rewards defined in (3.5), which has been shown to be proportional to the maximum tolerable

interference level. C1 restricts the number of allocated sub-channels to any MUE m to one sub-channel only, whereas C2 restricts sub-channel n to be allocated to at most one MUE. Then for each MUE m with allocated sub-channel n, the maximum tolerable interference level is given by: $I_m^n = \left(\frac{P_{B,m}^n g_{B,m}^n}{2^{R_m/\Delta f}-1} - N_o\right)$, where $P_{B,m}^n = \frac{P_{B,max}}{M}$. All the remaining sub-channels will have a value of I_{max} for the maximum tolerable interference level. Hence, (3.4) is solved optimally.

Proof. We have shown that I_m^{i*} is directly proportional to $\frac{g_{B,m}^i}{2^{R_m/\Delta f}-1}$. Moreover, according to Lemmas 3.4.2 and 3.4.3, at optimality, each MUE will have only one sub-channel which is the one with the highest reward. Besides, no two MUEs can have the same sub-channel allocated. Hence, we can define the optimization problem in (3.6), whose objective function is directly proportional to I_m^{i*}. In addition, it allocates a single sub-channel only (the one with the highest reward) to each MUE, while respecting the OFDMA constraint. □

Note that problem in (3.6) is the well-known assignment problem [71]. In this case, the Hungarian algorithm can be used to solve (3.6) for global optimality with a runtime complexity[3] of $\mathcal{O}(N^3)$ [75].

In this way, the macrocell has allocated sub-channels to its MUEs in a way that satisfies their data rate requirements and that can tolerate the maximum possible interference from the small cell tier.

Minimize the total sum-power

An alternate strategy that can be employed by the macrocell while performing RA is to minimize the total sum-power, given the data rate requirements of the MUEs. This

[3]The complexity of $\mathcal{O}(N^3)$ is acceptable for practical OFDMA-based resource allocation algorithms to be deployed in an on-line manner [51], [72], [73], [74].

problem has been studied extensively in the literature [76]. However, the formulation developed in [76] does not account for the maximum tolerable interference level I_m^n. Hence, we include it here with the required modification to determine the maximum tolerable interference level I_m^n. Throughout the rest of this work, we will refer to this RA method as the *Traditional* method. We have, thus, the following optimization problem:

$$\begin{aligned}
&\min_{\{P_{B,m}^n\}} \sum_{m=1}^{M}\sum_{n=1}^{N} P_{B,m}^n \\
&\text{subject to} \\
&\text{C1}: \sum_{n\in\mathcal{N}_m} \Delta f \log_2\left(1+\frac{P_{B,m}^n g_{B,m}^n}{I_m^n+N_o}\right) \geq R_m, \forall m \in \mathcal{M} \\
&\text{C2}: \mathcal{N}_i \bigcap \mathcal{N}_j = \varnothing,\ \forall i,j \in \mathcal{M}, i \neq j \\
&\text{C3}: \bigcup_{m=1}^{M} \mathcal{N}_m \subseteq \{1,2,...,N\} \\
&\text{C4}: P_{B,m}^n \geq 0, \forall m \in \mathcal{M}, n \in \mathcal{N}.
\end{aligned} \tag{3.7}$$

In (3.7), given the maximum tolerable interference level on each allocated sub-channel I_m^n, the macrocell seeks a power and sub-channel allocation solution that minimizes the sum-power. The problem in (3.7) can be shown to be strongly NP-hard [77]. In other words, it cannot be solved by a pseudo-polynomial time algorithm and finding its global optimal solution is generally NP-hard. This NP-hardness is attributed to the non-convexity of the optimization problem in (3.7) due to the combinatorial nature of sub-channel allocation. Hence, solving this problem by using Lagrange dual decomposition will, generally, result in a non-zero duality gap between

its primal and dual solutions. However, in a multi-carrier OFDMA system, according to [76] and [78], this duality gap virtually vanishes as the number of sub-channels goes to infinity. This implies that, in an OFDMA system with a sufficiently large number of sub-channels N (N should be greater than 8 [76]), Lagrange dual decomposition can be used to solve the problem in (3.7) efficiently in the dual domain.

Remark 3.4.1. *The reason for choosing the resource allocation solution based on the formulation in (3.7) as the baseline is the following. With this solution, at optimality, all MUEs have their rate requirements satisfied with equality. This is also the case for the solution obtained from the formulation in (3.4). In other words, with both RA strategies, the MUEs achieve the same performance. The difference however lies in the way the resources are allocated, which subsequently impacts the performance of the small cell tier.*

In (3.7), the same value of I_m^n is assumed $\forall m \in \mathcal{M}, n \in \mathcal{N}$, i.e., $I_m^n = I_{th}$. For a fair comparison between (3.4) and (3.7), the macrocell adjusts the maximum tolerable interference level I_{th} such that $\sum_{m=1}^{M}\sum_{n=1}^{N} P_{B,m}^n = P_{B,max}$. This can be accomplished by using the bisection method according to **Algorithm 5** as given below, where $I_{th,H} > I_{th,L}$.

Algorithm 5 Bisection method to find optimal I_{th}

1: Macrocell initializes $I_{th,L}$, $I_{th,H}$, and δ
2: **while** $|\sum_{m=1}^{M}\sum_{n\in\mathcal{N}_m} P_{B,m}^n - P_{B,max}| > \delta$ **do**
3: $I_{th,M} = (I_{th,L} + I_{th,H})/2$
4: Macrocell solves the optimization problem in (3.7)
5: **if** $\sum_{m=1}^{M}\sum_{n\in\mathcal{N}_m} P_{B,m}^n > P_{B,max}$ **then**
6: $I_{th,H} = I_{th,M}$
7: **else if** $\sum_{m=1}^{M}\sum_{n\in\mathcal{N}_m} P_{B,m}^n < P_{B,max}$ **then**
8: $I_{th,L} = I_{th,M}$
9: **end if**
10: **end while**

Assuming that $I_{th,L}$ is initialized to 0, the bisection method usually requires $\mathcal{O}(\ln(\frac{I_{th,H}}{\delta}))$ iterations to converge. Besides, the complexity of each iteration is of $\mathcal{O}(NM^3)$ [76]. After **Algorithm 5** terminates, $I_{th,M}$ gives the optimal value of I_{th}. The optimization problem in (3.7) ends up with the power and sub-channel allocation to the MUEs with a uniform maximum tolerable interference level I_{th} on all allocated sub-channels. In general, as will be shown in the numerical results, (3.7) leads to a higher number of allocated sub-channels to the MUEs than (3.4) does. It is of interest to study the effect of the two different RA methods on the small cell tier.

3.4.2 Problem Formulation for Small Cells

Due to the small distance and the good channel conditions between small cells and SUEs, small cells are capable of serving registered SUEs with higher data rates than the macrocell. However, this should not be at the cost of QoS degradation at MUEs as they are served by the macrocell and provided with basic coverage at possibly lower rates [53]. Hence, given the maximum tolerable interference levels on each allocated sub-channel for the MUEs, each small cell now tries to admit as many SUEs as possible at their target data rate by using the minimum possible bandwidth. Again, the idea here is to leave as much bandwidth as possible for the other network tiers (e.g., for D2D communication).

Centralized operation

To accomplish the aforementioned requirements, we define the optimization problem in (3.8), where the objective is to maximize the number of admitted SUEs while minimizing the number of allocated sub-channels. We have the admission control variable $y_{s,f}$ which takes the value of 1 if SUE f is admitted in small cell s and

0 otherwise. Recall from Section 3.3.1 that the set $\mathcal{F}_s$ denotes the set of SUEs already associated with small cell s. Yet, SUEs in the set $\mathcal{F}_s$ can still have their rate requirements satisfied or not. This will be indicated through the variable $y_{s,f}$. By controlling the weighting factor $\epsilon \in [0,1]$, admission control can be given higher priority over the number of used sub-channels. Note that the formulation (3.8) is in spirit of the formulation in [58]. However, the objective in [58] was to maximize the number of admitted UEs while minimizing the transmission power. In our work, we have a different objective. Moreover, the work in [58] was done in the context of single channel systems.

$$\max_{\left\{\Gamma^n_{s,f}, P^n_{s,f}, y_{s,f}\right\}} (1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} y_{s,f} - \epsilon\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\Gamma^n_{s,f}$$

subject to

$$\text{C1}: \sum_{n=1}^{N}\Delta f \log_2\left(1+\frac{P^n_{s,f}g^n_{s,f}}{\sum_{m=1}^{M}\Gamma^n_{B,m}P^n_{B,m}g^n_{B,f}+N_o}\right) \geq y_{s,f}R_f,\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s$$

$$\text{C2}: \sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N} P^n_{s,f} \leq P_{s,max},\ \forall s\in\mathcal{S}$$

$$\text{C3}: \Gamma^n_{B,m}\left(\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} P^n_{s,f}g^n_{s,m}\right) \leq \Gamma^n_{B,m}I^n_m,\ \forall n\in\mathcal{N}$$

$$\text{C4}: P^n_{s,f} \leq \Gamma^n_{s,f}P_{s,max}\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}$$

$$\text{C5}: \sum_{f\in\mathcal{F}_s}\Gamma^n_{s,f} \leq 1,\ \forall s\in\mathcal{S}, n\in\mathcal{N}$$

$$\text{C6}: P^n_{s,f} \geq 0,\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}$$

$$\text{C7}: \Gamma^n_{s,f} \in \{0,1\},\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}$$

$$\text{C8}: y_{s,f} \in \{0,1\},\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}. \tag{3.8}$$

In (3.8), C1 is a data rate constraint for an SUE f which is active only if SUE f is admitted, i.e., $y_{s,f} = 1$. C2 is the power budget constraint for each small cell s restricting the total transmission power of small cell s to be less than or equal to $P_{s,max}$. C3 is a constraint on the maximum cross-tier interference[4] introduced to MUE m using sub-channel n. This is imperative as the MUEs have a strictly higher priority in accessing the underlying frequency bands than the SUEs. C4 ensures that if sub-channel n is not allocated to SUE f, its corresponding transmit power $P^n_{s,f} = 0$. C5 constrains sub-channel n to be allocated to at most one SUE f in small cell s. C6 ensures that the power $P^n_{s,f}$ should be positive, and finally, C7 and C8 indicate that $\Gamma^n_{s,f}$ and $y_{s,f}$ are binary variables. One benefit of the optimization problem formulation in (3.8) is that it is always feasible. To see this, a trivial feasible solution of (3.8) is $\Gamma^n_{s,f} = 0, P^n_{s,f} = 0$ and $y_{s,f} = 0,\ \forall s \in \mathcal{S}, f \in \mathcal{F}_s, n \in \mathcal{N}$.

Proposition 3.4.1. *By choosing $\epsilon < \frac{1}{1+SN}$, (3.8) admits the maximum number of SUEs while consuming the minimum number of sub-channels.*

Proof. Let $\left(\Gamma^{n*}_{s,f}, P^{n*}_{s,f}, y^*_{s,f}\right),\ \forall s \in \mathcal{S}, f \in \mathcal{F}_s, n \in \mathcal{N}$ denote an optimal solution of (3.8). Let $\left(\Gamma^{\hat{n}}_{s,f}, P^{\hat{n}}_{s,f}, y^{\hat{}}_{s,f}\right), \forall s \in \mathcal{S}, f \in \mathcal{F}_s, n \in \mathcal{N}$ be a feasible solution that admits one more SUE than the optimal solution, i.e., $\sum_{s=1}^{S} \sum_{f \in \mathcal{F}_s} y^{\hat{}}_{s,f} = \sum_{s=1}^{S} \sum_{f \in \mathcal{F}_s} y^*_{s,f} + 1$.

[4]Note that this constraint is known as the interference temperature constraint in cognitive radio networks and it has also been used extensively in two-tier macrocell-small cell networks [15], [54], [66], [79], [80].

The objective of the feasible solution can be written as:

$$
\begin{aligned}
&(1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} \hat{y_{s,f}} - \epsilon\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\hat{\Gamma_{s,f}^{n}} \overset{(1)}{\geq} \\
&(1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} y_{s,f}^{*} + (1-\epsilon) - \epsilon SN \overset{(2)}{\geq} \\
&(1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} y_{s,f}^{*} \overset{(3)}{\geq} \\
&(1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} y_{s,f}^{*} - \epsilon\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\Gamma_{s,f}^{n*}.
\end{aligned}
$$

The first inequality holds due to the fact that $\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\hat{\Gamma_{s,f}^{n}}$ is upper bounded by SN when all sub-channels in all small cells are allocated. The second inequality holds by setting $(1-\epsilon)-\epsilon SN > 0$. Hence, we have $\epsilon < \frac{1}{1+SN}$. The last inequality holds due to the non-negativity of $\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\Gamma_{s,f}^{n*}$. In this way, the value of the objective function for the feasible solution is higher than the optimal one, which contradicts the optimality of $\left(\Gamma_{s,f}^{n*}, P_{s,f}^{n*}, y_{s,f}^{*}\right)$. Thus, there is no other solution that admits a higher number of SUEs under the constraint in (3.8). Given the optimum value for the admission control variable $y_{s,f}^{*}$, (3.8) reduces to a feasible sub-channel and power allocation problem with respect to the variables $\Gamma_{s,f}^{n}$ and $P_{s,f}^{n}$ that aims at minimizing the number of used sub-channels subject to the given constraints. □

The problem in (3.8) is an MINLP whose feasible set is non-convex due to the combinatorial nature of sub-channel allocation and admission control. However, for small-sized problems, I use OPTI [81], which is a MATLAB toolbox to construct and solve linear, nonlinear, continuous, and discrete optimization problems, to obtain the optimal solution. Obtaining the optimal solution, however, for larger problems is

intractable. Another approach that can render the problem in (3.8) more tractable is to have a convex reformulation of (3.8) by relaxing the constraints C7 and C8 and allowing $\Gamma_{s,f}^{n}$ and $y_{s,f}$ to take any value in the range $[0, 1]$. Thus, $\Gamma_{s,f}^{n}$ is now a time sharing factor that indicates the portion of time sub-channel n is allocated to SUE f [41], [42], whereas $y_{s,f}$ indicates the ratio of the achieved data rate for SUE f. Hence, we define the convex optimization problem in (3.9), where $\tilde{P}_{s,f}^{n}$ can be related to $P_{s,f}^{n}$ in (3.8) as $\tilde{P}_{s,f}^{n} = \Gamma_{s,f}^{n} P_{s,f}^{n}$ to denote the actual transmit power [65]. Now, the problem in (3.9) is a convex optimization problem with a linear objective function and convex feasible set. This means that its global solution can be efficiently obtained, for example, by a general interior-point method [44]. Such a method usually converges within a few tens of iterations, where the complexity is of $\mathcal{O}((SFN)^3)$ per iteration.

$$\max_{\left\{\Gamma^n_{s,f},\tilde{P}^n_{s,f},y_{s,f}\right\}} (1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} y_{s,f} - \epsilon\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\Gamma^n_{s,f}$$

subject to

$$\text{C1}: \sum_{n=1}^{N}\Gamma^n_{s,f}\Delta f \log_2\left(1+\frac{\left(\tilde{P}^n_{s,f}g^n_{s,f}/\Gamma^n_{s,f}\right)}{\sum_{m=1}^{M}\Gamma^n_{B,m}P^n_{B,m}g^n_{B,f}+N_o}\right) \geq y_{s,f}R_f,\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s$$

$$\text{C2}: \sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\tilde{P}^n_{s,f} \leq P_{s,max},\ \forall s\in\mathcal{S}$$

$$\text{C3}: \Gamma^n_{B,m}\left(\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\tilde{P}^n_{s,f}g^n_{s,m}\right) \leq \Gamma^n_{B,m}I^n_m,\ \forall n\in\mathcal{N}$$

$$\text{C4}: \sum_{f\in\mathcal{F}_s}\Gamma^n_{s,f} \leq 1,\ \forall s\in\mathcal{S}, n\in\mathcal{N}$$

$$\text{C5}: \tilde{P}^n_{s,f} \geq 0,\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}$$

$$\text{C6}: \Gamma^n_{s,f}\in(0,1],\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}$$

$$\text{C7}: y_{s,f}\in[0,1]\,,\ \forall s\in\mathcal{S}, f\in\mathcal{F}_s, n\in\mathcal{N}.$$

(3.9)

It is worth mentioning that the solution of the problem in (3.9) gives an upper bound to the optimal solution of the problem in (3.8). Yet, it reveals insights into the behavior of the solution of the problem in (3.8). Note that the time sharing in sub-channel allocation can be implemented in practice [82]. Therefore, we will keep $\Gamma^n_{s,f}$ relaxed. On the other hand, in order to obtain a binary solution for the relaxed admission control variable $y_{s,f}$, iterative user removal algorithms are needed [55], which can be accomplished using Algorithm 6 described below.

The idea of **Algorithm** 6 is to iteratively solve the problem in (3.9) and each time set the value of the minimum non-zero $y_{s,f}$ to zero. The Algorithm terminates

Algorithm 6 Joint RA and AC algorithm with one-by-one removal

1: Solve problem (3.9)
2: **if** All the variables $y_{s,f}$ are binary **then**
3: Terminate
4: **else**
5: $(s^*, f^*) = \arg \min_{s \in \mathcal{S}, f \in \mathcal{F}} y_{s,f}$ such that $y_{s,f} \neq 0$
6: Set $y_{s^*,f^*} = 0$
7: Go to step 1
8: **end if**

when all $y_{s,f}$ are binary, i.e., either 0 or 1. Using **Algorithm** 6, the problem in (3.9) will be solved at most $\mathcal{O}(F)$ times.

Note that the solution of (3.9) necessitates the existence of a central node with global channel state information which can be, for example, the HeNB-GW. However, in a densely deployed small cell network, centralized solutions might not be practical. Hence, it is foreseen that having a decentralized solution with some coordination with a central node will be a more viable option.

Distributed operation

To fulfill the requirement of having a decentralized solution for (3.9), we use the dual decomposition method [83]. For this purpose, we define the following partial Lagrangian function of the primal problem in (3.9) formed by dualizing the constraint C3:

$$\mathcal{L}\left(\Gamma_{s,f}^n, \tilde{P}_{s,f}^n, y_{s,f}, \boldsymbol{\eta}\right) = (1-\epsilon)\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s} y_{s,f} - \epsilon\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\sum_{n=1}^{N}\Gamma_{s,f}^n + \sum_{n=1}^{N}\eta^n\left(\Gamma_{B,m}^n I_m^n - \Gamma_{B,m}^n\left(\sum_{s=1}^{S}\sum_{f\in\mathcal{F}_s}\tilde{P}_{s,f}^n g_{s,m}^n\right)\right) \tag{3.10}$$

where $\boldsymbol{\eta}$ is the Lagrange multiplier vector (with elements η^n) associated with the cross-tier interference constraint C3. Then the Lagrange dual function is represented as

$$g(\boldsymbol{\eta}) = \max_{\left\{\Gamma^n_{s,f}, \tilde{P}^n_{s,f}, y_{s,f}\right\}} \mathcal{L}\left(\Gamma^n_{s,f}, \tilde{P}^n_{s,f}, y_{s,f}, \boldsymbol{\eta}\right)$$

subject to

$$\text{C1},\ \text{C2},\ \text{C4} - \text{C7}. \tag{3.11}$$

From (3.10), the maximization of $\mathcal{L}$ can be decomposed into S independent optimization problems for each small cell s as follows:

$$g_s(\boldsymbol{\eta}) = \max_{\left\{\Gamma^n_{s,f}, \tilde{P}^n_{s,f}, y_{s,f}\right\}} (1-\epsilon) \sum_{f \in \mathcal{F}_s} y_{s,f} - \epsilon \sum_{f \in \mathcal{F}_s} \sum_{n=1}^{N} \Gamma^n_{s,f} - \sum_{f \in \mathcal{F}_s} \sum_{n=1}^{N} \eta^n \Gamma^n_{B,m} \tilde{P}^n_{s,f} g^n_{s,m}$$

subject to

$$\text{C1},\ \text{C2},\ \text{C4} - \text{C7}, \quad \forall s \in S. \tag{3.12}$$

Note that probelm in (3.12) is still a convex optimization problem which can be solved by each small cell s using the interior point method with a complexity of $\mathcal{O}((FN)^3)$ per iteration.

Thus, the Lagrange dual function is

$$g(\boldsymbol{\eta}) = \sum_{s=1}^{S} g_s(\boldsymbol{\eta}) + \sum_{n=1}^{N} \eta^n \Gamma^n_{B,m} I^n_m. \tag{3.13}$$

Then, the dual problem is given by:

$$\min_{\boldsymbol{\eta}\geq 0} g(\boldsymbol{\eta}). \tag{3.14}$$

In order to solve the dual problem, $\boldsymbol{\eta}$ can be updated efficiently using the ellipsoid method [84], which is known to converge in $\mathcal{O}(N^2)$ iterations. A sub-gradient $\boldsymbol{d}$ of this problem required for the ellipsoid method is derived in the following proposition.

Proposition 3.4.2. *For the optimization problem in (3.9) with a dual objective $g(\boldsymbol{\eta})$ defined in (3.11), the following choice of d^n is a sub-gradient for $g(\boldsymbol{\eta})$:*

$$d^n = \Gamma^n_{B,m} I^n_m - \Gamma^n_{B,m} \left(\sum_{s=1}^{S} \sum_{f\in\mathcal{F}_s} \tilde{P}^{n*}_{s,f} g^n_{s,m} \right) \tag{3.15}$$

where d^n is an element of $\boldsymbol{d}$ and $\Gamma^{n*}_{s,f}, \tilde{P}^{n*}_{s,f}$, and $y^*_{s,f}$ optimize the maximization problem in the definition of $g(\boldsymbol{\eta})$.

Proof. For any $\boldsymbol{\xi} \geq 0$,

$$\begin{aligned} g(\boldsymbol{\xi}) &\geq \mathcal{L}(\Gamma^{n*}_{s,f}, \tilde{P}^{n*}_{s,f}, y^*_{s,f}, \boldsymbol{\xi}) \\ &= g(\boldsymbol{\eta}) + \sum_{n=1}^{N} (\xi^n - \eta^n) \left[\Gamma^n_{B,m} I^n_m - \Gamma^n_{B,m} \left(\sum_{s=1}^{S} \sum_{f\in\mathcal{F}_s} \tilde{P}^{n*}_{s,f} g^n_{s,m} \right) \right] \end{aligned}$$

□

Algorithm 7 gives the practical implementation of the distributed joint RA and AC operation for the small cells. After the macrocell has performed RA for its MUEs, it sends the sub-channel allocation information for its MUEs and the initialized multiplier $\boldsymbol{\eta}$ to the HeNB-GW. For a given $\boldsymbol{\eta}$, all small cells solve their optimization problem in (3.12) simultaneously. For the given resource allocation in the small cells,

the MUEs estimate the resulting interference levels and send them to the macrocell which updates the multiplier values using the ellipsoid method. The macrocell then informs the updated multiplier values to the HeNB-GW, which broadcasts them to the small cells, and the entire operation repeats. Note that the small cells can obtain the channel gains $g^n_{s,m}$ relying on the techniques proposed in [40].

Finally, the remaining issue is to obtain a feasible primal solution to (3.9) based on the resulting solution from the Lagrangian dual in (3.14). It was reported in [83] and [85] that the iterations of the dual decomposition method are, in general, infeasible with respect to (3.9). This infeasibility, however, is not severe as large constraint violations usually get penalized. Hence, using a simple procedure, one can recover a primal feasible solution that serves as a lower bound for the optimal solution of (3.9). Suppose that the reported interference level by an MUE m allocated a sub-channel n was found to be:

$$\sum_{s=1}^{S} \sum_{f \in \mathcal{F}_s} \tilde{P}^n_{s,f} g^n_{s,m} = \alpha^n I^n_m, \ \alpha^n > 1. \tag{3.16}$$

A straightforward way to recover feasibility is for the HeNB-GW to instruct all small cells transmitting on sub-channel n to scale down their transmission powers by the factor α^n. For the updated power values, the entire problem is solved to obtain the updated values of sub-channel allocation and admission control variables. The gap between the lower bound offered by this procedure and the upper bound offered by (3.14), referred to as the duality gap, diminishes with iterations. Convergence to the optimal solution is guaranteed since the primal optimization problem in (3.9) is convex.

Algorithm 7 Distributed joint RA and AC algorithm

1: Macrocell initializes $\boldsymbol{\eta}$, L_{max}, sends sub-channel allocation information $\Gamma^n_{B,m}$ and $\boldsymbol{\eta}$ to HeNB-GW and sets iteration counter $l = 1$
2: HeNB-GW broadcasts $\Gamma^n_{B,m}$ and $\boldsymbol{\eta}$ values to all small cells
3: **repeat**
4: All small cells solve (3.12) in parallel
5: All MUEs estimate interference levels on allocated sub-channels and report them to the macrocell
6: Macrocell evaluates the sub-gradient (3.15) and updates $\boldsymbol{\eta}$ using the ellipsoid method
7: Macrocell sends updated $\boldsymbol{\eta}$ to HeNB-GW
8: HeNB-GW broadcasts updated $\boldsymbol{\eta}$ to all small cells
9: Macrocell sets $l = l + 1$
10: **until** Convergence or $l = L_{max}$

3.5 Numerical Results and Discussions

3.5.1 Parameters

We evaluate the system performance through extensive simulations under various topologies and scenarios. We have a macrocell located at the origin with radius 300 m. A hotspot of small cells exists at a distance of 100 m from the macrocell. Since MUEs that are far away from the hotspot of small cells will not be affected by the small cells transmissions, we focus our study on those MUEs that will be affected by small cell transmissions. Hence, all MUEs exist randomly outdoor in this hotspot and are served by the macrocell. Each small cell has 2 indoor SUEs located randomly on a circular disc around the small cell with an inner radius of 3 m and an outer radius of 10 m [47]. The macrocell has a total power budget of $P_{B,max} = 20$ W.

To model the propagation environment, the channel models from [47] are used. The channel gains include path-loss, log-normal shadowing, and multipath Rayleigh fading. The path-loss between a small cell and its served SUE, $PL = 38.46 + 20 \log R$ and the path-loss between a small cell and the outdoor MUEs, $PL = \max(38.46 +$

$20 \log R, 15.3 + 37.6 \log R) + L_{ow}$, where R is the distance between a small cell and the UE and L_{ow} accounts for losses due to walls. For path-loss between the macrocell and an SUE existing indoor, $PL = 15.3 + 37.6 \log R + L_{ow}$ and for path-loss between the macrocell and its MUE, $PL = 15.3 + 37.6 \log R$. We have the following values for the standard deviation of log-normal shadowing: 4 dB for shadowing between SUE and its small cell, 8 dB for shadowing between MUE and small cell and 10 dB for shadowing between macrocell and SUE or MUE. Multipath fading is assumed to follow a Rayleigh distribution, where Rayleigh fading channel gains are modeled as i.i.d. unit-mean exponentially distributed random variables. We assume $\Delta f = 180$ KHz, $\epsilon = \frac{0.9}{1+SN}$, and noise power, $N_o = 10^{-13}$ W. I_{max} is set to any arbitrary large number. All the rate requirements in the numerical results are specified in terms of spectral efficiency (bps/Hz).

In the numerical results, the following performance metrics are used:

- Average percentage of admitted SUEs $= \frac{\sum_{s=1}^{S} \sum_{f \in \mathcal{F}_s} y_{s,f}}{F} \times 100$.
- Average percentage of channel usage $= \frac{\sum_{s=1}^{S} \sum_{f \in \mathcal{F}_s} \sum_{n=1}^{N} \Gamma_{s,f}^{n}}{SN} \times 100$.

3.5.2 Numerical Results

Comparison between the proposed and the traditional RA methods for macrocell

In this section, we compare the two RA schemes in the macrocell, namely, the proposed formulation in (3.4) and the traditional formulation in (3.7). Fig. 3.3 shows the channel gain realizations for a snapshot of 3 MUEs with 10 sub-channels.

Figs. 3.4 and 3.5 compare the two RA results for the given snapshot with $R_m = 5$ bps/Hz. Each figure shows the allocated power $P_{B,m}^{n}$ by the macrocell and the maxi-

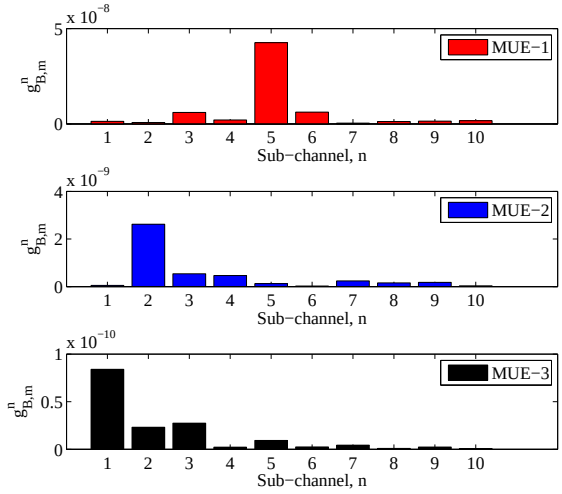

Figure 3.3: Channel gains $g^n_{B,m}$ for MUEs $\{1, 2, 3\}$.

mum tolerable interference level I^n_m on allocated sub-channel n to MUE m. No values for power $P^n_{B,m}$ on the x-axis indicate unallocated sub-channel with the corresponding value for I^n_m set to I_{max}, which means that this sub-channel can be used by the small cell tier unconditionally. For further clarification, Table 3.1 shows the absolute values of $P^n_{B,m}$ and I^n_m.

It is clear from Fig. 3.4 that most of the sub-channels are allocated to the MUEs (9 sub-channels out of 10 are allocated to the 3 MUEs), when using the traditional scheme for RA. We notice also that the macrocell favors good sub-channels as they require less transmit power to achieve the rate requirements for the MUEs, leading at the end to minimum transmit power requirements.

Fig. 3.5, on the other hand, shows that the 3 MUEs require only 3 sub-channels to achieve their rate requirements, as was proved before, using the proposed scheme. Again, the macrocell allocates the best sub-channels to the MUEs. From Figs. 3.4 and 3.5 and Table 3.1, we can notice that the entire power budget of macro BS, $P_{B,max}$, is

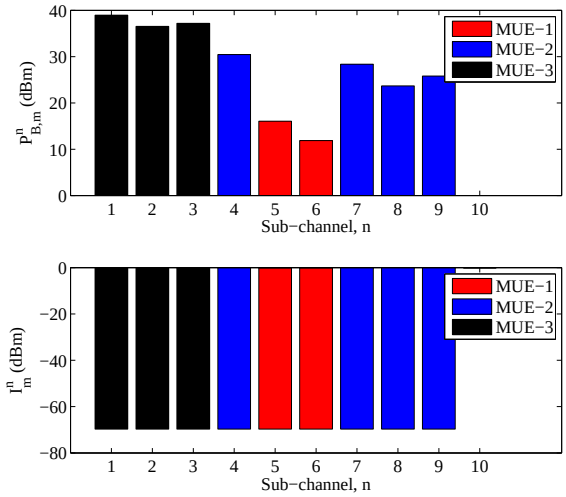

Figure 3.4: Allocated power $P^n_{B,m}$ and maximum tolerable interference level I^n_m for MUEs $\{1, 2, 3\}$ using the traditional scheme.

used in both cases. It is worth mentioning that when we use the traditional scheme for macrocell resource allocation, it does not necessarily mean that it will consume less power than the proposed scheme, since the maximum tolerable interference level I^n_m is adjusted according to **Algorithm 5** by the macrocell to use the entire power budget. It rather means that, given the maximum tolerable interference levels, the resulting sub-channel and power allocation for the traditional scheme will consume the minimum power and any other allocation will consume a higher power.

Now, for the maximum tolerable interference levels I^n_m, it is obvious from Figs. 3.4 and 3.5 and Table 3.1 that the proposed scheme can sustain higher interference levels from the small cell tier. It is of interest to compare the effect of the two different RA schemes for the macrocell on the small cell tier.

Fig. 3.6 compares the average percentage of admitted SUEs when the macrocell performs RA according to the traditional method based on (3.7) and the proposed method based on (3.4) with two different wall loss L_{ow} scenarios. We have the follow-

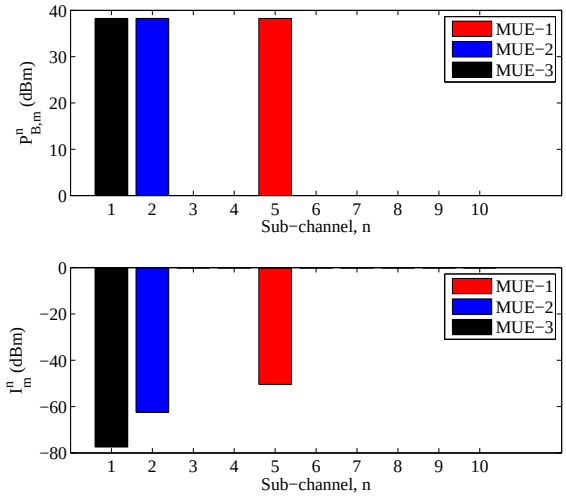

Figure 3.5: Allocated power $P^n_{B,m}$ and maximum tolerable interference level I^n_m for MUEs $\{1, 2, 3\}$ using the proposed scheme.

Table 3.1: Absolute Values of $P^n_{B,m}$ and I^n_m for the Traditional and Proposed Macrocell RA Schemes

			Sub-channel#									
			1	2	3	4	5	6	7	8	9	10
Traditional	$P^n_{B,m}$ (W)	MUE-1	0	0	0	0	0.0402	0.0153	0	0	0	0
		MUE-2	0	0	0	1.1097	0	0	0.6855	0.2333	0.3805	0
		MUE-3	7.8381	4.4890	5.2142	0	0	0	0	0	0	0
	I^n_m (W) $\times 10^{-10}$	MUE-1	I_{max}	I_{max}	I_{max}	I_{max}	1.0709	1.0709	I_{max}	I_{max}	I_{max}	I_{max}
		MUE-2	I_{max}	I_{max}	I_{max}	1.0709	I_{max}	I_{max}	1.0709	1.0709	1.0709	I_{max}
		MUE-3	1.0709	1.0709	1.0709	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}
Proposed	$P^n_{B,m}$ (W)	MUE-1	0	0	0	0	6.6667	0	0	0	0	0
		MUE-2	0	6.6667	0	0	0	0	0	0	0	0
		MUE-3	6.6667	0	0	0	0	0	0	0	0	0
	I^n_m (W) $\times 10^{-10}$	MUE-1	I_{max}	I_{max}	I_{max}	I_{max}	91.604	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}
		MUE-2	I_{max}	5.6298	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}
		MUE-3	0.1796	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}	I_{max}

ing scenario: 2 small cells located at $(-10, -100), (10, -100)$ in a square area hotspot of dimensions 20×20 m^2, 10 sub-channels, $P_{s,max} = 30$ mW, $R_f = 50$ bps/Hz, and $R_m = 5$ bps/Hz. Numerical results are obtained and averaged for 50 different realizations, where in each realization, the UE positions and the channel gains are varied. The small cell problem is solved centrally by using the convex formulation in (3.9). It is clear from the figure that the proposed RA method for the macrocell outperforms the traditional one. When the macrocell performs RA according to the proposed method, it consumes the minimum bandwidth, and therefore, frees as many sub-

channels as possible for the small cells. On the other hand, the traditional method consumes more bandwidth than the proposed one, hence, the small cells have more interference constraints to abide by. We also notice that as the wall losses increase, the small cells tend to be more isolated and the impact of resource allocation in the macrocell on the small cell performance becomes lower.

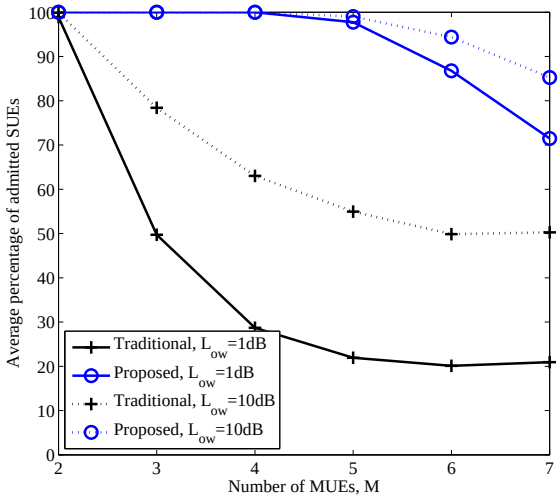

Figure 3.6: Average percentage of admitted SUEs vs. number of MUEs M when the macrocell employs both the proposed and the traditional methods for RA with different wall loss scenarios.

Comparison between the different formulations for the RA problem for small cells

Fig. 3.7 compares the objective function values for the MINLP formulation in (3.8), the centralized convex formulation in (3.9), the centralized convex formulation in (3.9) with one-by-one removal by using Algorithm 6, and the distributed formulation in (3.10) for a snapshot of the following scenario: 2 small cells located at $(-10, -100), (10, -100)$ in a square area hotspot of dimensions 20×20 m^2, 3 sub-

channels, 3 MUEs, $P_{s,max} = 30$ mW, $L_{ow} = 1$ dB, and $R_f = 5$ bps/Hz. As was stated previously, the convex formulation provides an upper bound for the solution of the MINLP formulation. Also, we notice that the centralized and distributed formulations have the same solution due to the convexity of the centralized formulation in (3.9). The centralized formulation with one-by-one removal exhibits a performance that lies in-between the MINLP one (3.8) and the centralized one (3.9). This can be attributed to the fact that in the MINLP formulation (3.8), $\Gamma_{s,f}^n$ and $y_{s,f}$ are binary. On the other hand, the centralized formulation in (3.9) has $\Gamma_{s,f}^n$ and $y_{s,f}$ relaxed. Since the centralized formulation with one-by-one removal has $\Gamma_{s,f}^n$ relaxed but $y_{s,f}$ binary, its performance will lie, generally, in between the other formulations.

It is worth mentioning that the convex formulation exhibits a behavior similar to the MINLP formulation. Hence, solving the convex formulation reveals insights into the behavior of the solution of the MINLP formulation. We also notice that as R_m increases, the interference constraints for the MUEs become tighter. Hence, the average number of admitted SUEs decreases. Since the objective function in our formulation gives higher priority to admission control, the value of objective function decreases with increasing R_m.

Figs. 3.8 and 3.9 show the average percentage of admitted SUEs and channel usage in a small cell vs. R_m for the same scenario considered in Fig. 3.7. As was discussed in Fig. 3.7, as the rate requirements for the MUEs increase, they have tighter interference constraints. Hence, the percentage of admitted SUEs generally decreases. We notice in Fig. 3.8 that, initially, the average percentage of admitted SUEs is almost constant. This is due to the increased number of used sub-channels as shown in Fig. 3.9. As the MUEs' rate requirements increase further, the increase in the number of used sub-channels is not enough to accommodate the rate requirements

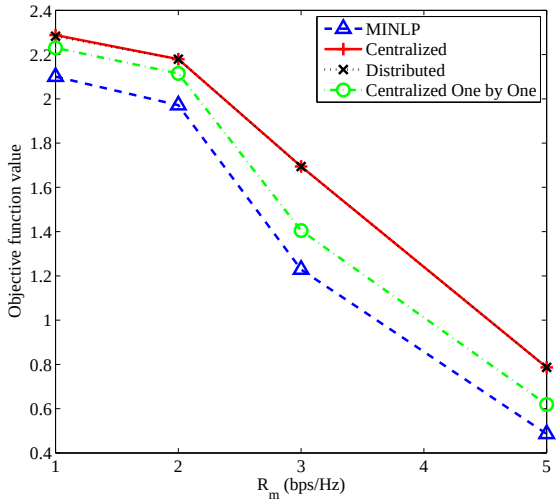

Figure 3.7: The values of objective function for different formulations vs. R_m.

of the SUEs, hence, the average percentage of admitted SUEs decreases.

Convergence behavior

Using the same scenario described for the previous figure, Fig. 3.10 shows the convergence behavior of **Algorithm 7**, where the upper bound refers to (3.14) and the lower bound refers to the feasible objective obtained by the procedure at the end of Section 3.4.2. In the figure, the best lower bound is obtained by keeping track of the best primal feasible objective resulting through iterations. It is clear that **Algorithm 7** converges to the optimal solution of the problem in (3.9) within a few iterations.

Average percentage of admitted SUEs vs. R_f

In this scenario, we have the following setup: 5 small cells located at $(-20,-100),(-20,-140),(20,-140),(20,-100),(0,-120)$ in a square area hotspot of dimensions 40×40 m^2, 5 sub-channels, 5 MUEs, $P_{s,max} = 30$ mW, $L_{ow} = 1$ dB

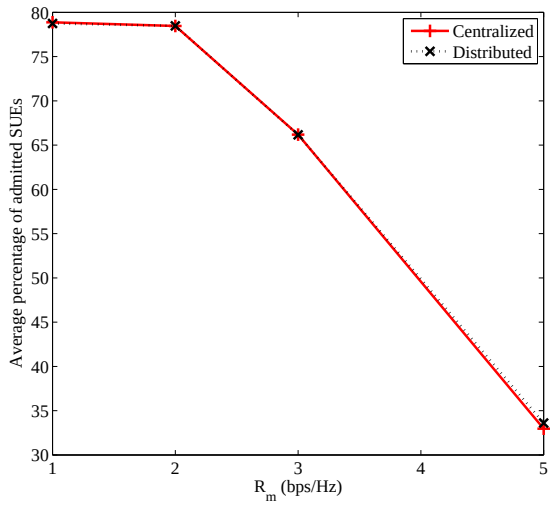

Figure 3.8: Average percentage of admitted SUEs vs. R_m.

and $R_m = 4$ bps/Hz. Numerical results are obtained and averaged for 50 different realizations, where in each realization, the UE positions and channel gains are varied. Fig. 3.11 shows the average percentage of admitted SUEs vs. R_f. We notice that, generally, as the rate requirement increases, more SUEs are in outage. We also notice that the distributed scheme converges approximately to the same solution as the centralized solution.

Average percentage of admitted SUEs vs. $P_{s,max}$

We have the same setup as the one for the previous figure except for $R_f = 10$ bps/Hz. Fig. 3.12 shows the average percentage of admitted SUEs vs. $P_{s,max}$. We notice that as the maximum transmit power of the small cells increases, the average number of admitted SUEs increases. This rate of increase, however, is not fixed as the system is limited by the interference constraints for MUEs.

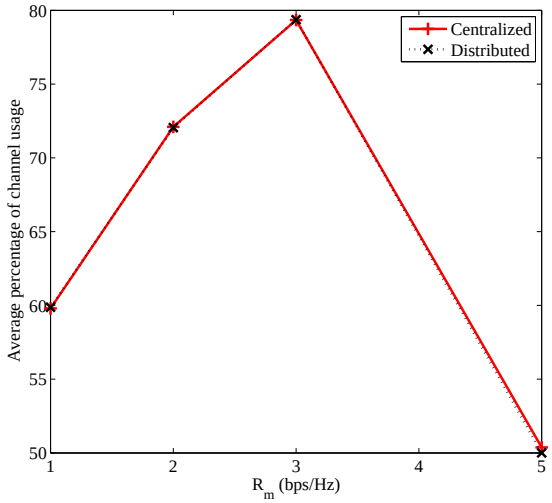

Figure 3.9: Average percentage of channel usage vs. R_m.

3.5.3 Summary of Major Observations

The major observations from the numerical analysis can be summarized as follows:

- In a multi-tier network, it is critical to consider the impact of RA decisions in one tier on the other one. For the macrocell network, as different RA schemes are used to achieve the same rate requirements for the MUEs, they affect the performance of the small cell tier differently.

- The proposed problem formulation for resource allocation in the macrocell leads to a minimal use of the system bandwidth which allows to admit a higher number of SUEs when compared to the traditional scheme that minimizes the sum-power.

- For a given macrocell RA policy, increasing the rate requirements for the MUEs degrades the performance of small cell tier in terms of the average number of admitted SUEs.

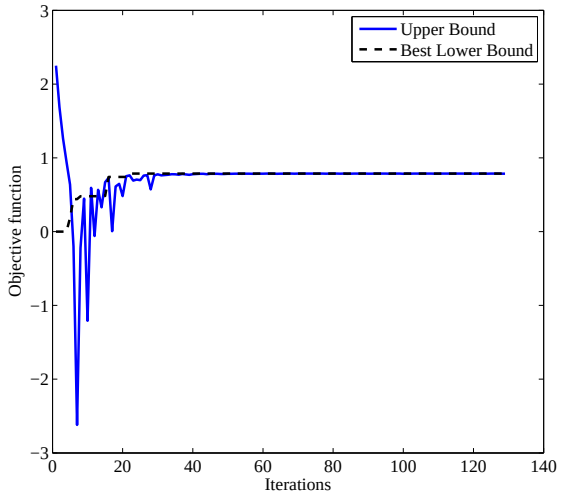

Figure 3.10: Convergence of **Algorithm 7**.

- By exploiting the time sharing property (i.e., the UEs time share the sub-channels when served by the small cells), a convex optimization formulation can be developed for the RA and AC problem for the small cells. This convex formulation enables us to solve the problem efficiently in a distributed fashion. The distributed algorithm for resource allocation for the small cell tier converges to the same solution as the centralized solution.

- If the deployment of the small cells is such that they are well isolated, resource allocation in the macrocell might have very little effect on the performance of small cells.

3.6 Conclusion

We have proposed a complete framework for resource allocation and admission control in a two-tier OFDMA cellular network. Different optimization problems with new objectives have been formulated for the macrocell tier and the small cell tier. The

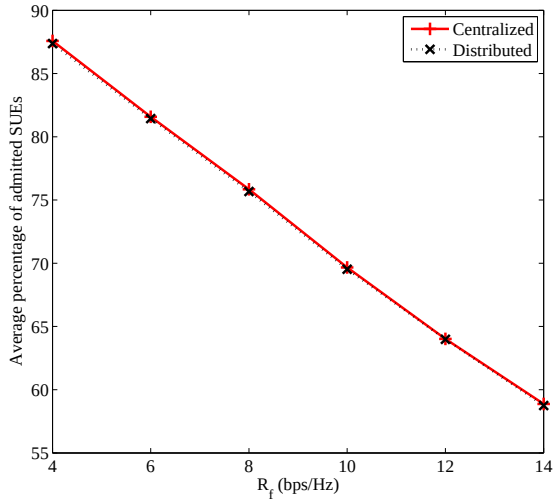

Figure 3.11: Average percentage of admitted SUEs vs. R_f.

macrocell tier aims at allocating resources to its MUEs in a way that can tolerate the maximum possible interference from the small cell tier. This problem has been shown to be an MINLP. However, we have proved that the macrocell can solve an alternate optimization problem that yields the optimal solution in polynomial time. Now, given the interference constraints for the MUEs, the small cells perform resource allocation and admission control with the objective of maximizing the number of admitted SUEs and serving them with the minimum possible bandwidth. This problem has also been shown to be an MINLP. A convex relaxation has been used to study the behavior of the MINLP formulation. Since centralized solutions for resource allocation are not practical for dense networks, a distributed solution for resource allocation and admission control has been proposed using dual decomposition technique and has been shown to converge to the same solution as the centralized one. Numerical results have shown the efficacy of the proposed tier-aware resource allocation methods.

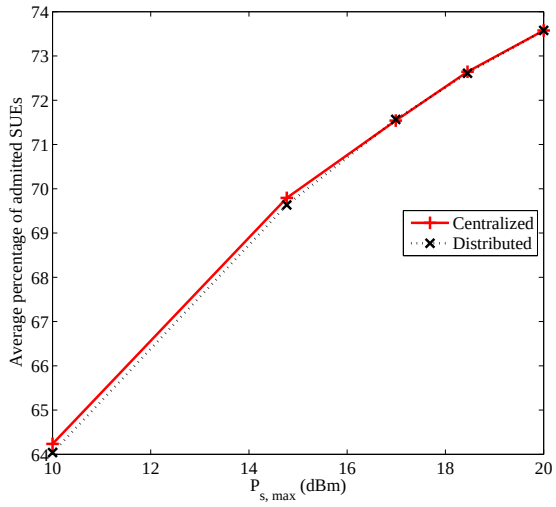

Figure 3.12: Average percentage of admitted SUEs vs. $P_{s,max}$.

Chapter 4

Resource Allocation for an OFDMA Cloud-RAN of Small Cells Underlaying a Macrocell

4.1 Introduction

It is anticipated that by 2020, wireless communication systems will have to support more than 1000 times todays traffic volume [1, 49]. Besides, there will be an unprecedented number of mobile devices with a much wider range of use-case characteristics and diverse QoS requirements. Hence, the fifth-generation (5G) wireless technologies are being sought. In this context, cloud radio access network (C-RAN) has been considered as a typical realization of a mobile network which is capable of supporting soft and green technologies in 5G networks in year 2020 horizon [86]. More specifically, C-RAN has been proposed as a platform for the practical implementation of network multiple-input multiple-output (MIMO) and coordinated multi-point (CoMP) transmission concepts [5]. C-RAN can manage the inter-cell interference, increase

network capacity and energy efficiency, and reduce both the network capital expenditure (CAPEX) and operating expense (OPEX) [87]. To achieve these benefits in a multi-tier 5G network, small cells can be deployed in a C-RAN architecture.

In this Chapter, I study the RA problem in a two-tier OFDMA wireless network composed of a macrocell overlaid by small cells deployed in a C-RAN architecture. The objective of the macrocell is to allocate resources to its MUEs to satisfy their data rate requirements, in a way that can tolerate the highest possible interference from the small cells. In other words, the macrocell objective is to maximize the sum of the interference levels it can tolerate. This is unlike the traditional "min sum-power" and "max sum-rate" objective functions [76]. I advocated the use of such objective function in the previous Chapter, as well as in a previous work [8], and argued that such objective function brings up the notion of "tier-awareness", where the macrocell performs resource allocation while taking into consideration the impact of its RA decisions on the small cells performance. However, some simplifying assumptions were made in [8], which I relax in this Chapter. I show that this problem is an MINLP problem whose optimal solution is intractable. However, to gain further insights into the behavior of the RA problem with the proposed objective function, I investigate the macrocell RA problem for a single MUE. In that case, I show that I can obtain the optimal solution, with polynomial time complexity, despite the non-convexity of the single MUE RA problem. Motivated by the observations from the single-MUE macrocell RA problem, I propose a lower complexity algorithm for the multi-MUE case that relies on the framework of successive convex approximation (SCA) presented in [88] and [89]. The general idea of SCA approach is to approximate the original non-convex problem by another convex one, which can be efficiently solved. Then, by updating the variables involved, I iteratively solve the approximating convex problem

until convergence. Upon convergence, I end up with a feasible solution to the original non-convex problem that satisfies its necessary KKT optimality conditions.

For the C-RAN of small cells, the objective is to minimize the total downlink transmit power while satisfying the QoS requirements of SUEs. However, one of the limiting factors for C-RAN performance is the limited capacity of the fronthaul links connecting the small cells to the cloud [87]. Besides, MUEs must usually be protected against small cells transmissions. Moreover, to incorporate the idea of "tier-awareness" in the C-RAN RA problem, I impose a limit on the the number of sub-channels that can be allocated to the SUEs. This brings additional constraints for the RA problem for the small cells in terms of fronthaul capacity constraints, interference thresholds for the MUEs, and the number of allocated sub-channels. In addition, AC is considered in case the C-RAN optimization problem becomes infeasbile [55]. I show that the RA problem for the C-RAN is an MINLP for which I propose a low-complexity, yet efficient, solution based on the framework of SCA.

Numerical results confirm the different impacts of the different RA policies employed at the macrocell on the small cells' performance. Hence, it is important to carefully select the objective function employed in the macrocell RA problem. Moreover, I investigate the effect of the different parameters of the C-RAN RA problem and their impact on the overall C-RAN performance.

The rest of this Chapter is organized as follows. Section 4.2 reviews the related work and highlights the contribution of this Chapter. Section 4.3 presents the system model and assumptions. In Section 4.4, I present the optimization problem formulation for the macrocell RA problem and in Section 4.5, I present the optimization problem formulation for the C-RAN RA problem. Numerical results are discussed in Section 4.6 and, finally, Section 4.7 concludes the Chapter.

4.2 Related Work

In [90], the authors formulated the joint remote radio head (RRH) selection and power consumption minimization (in both the RRHs and the fronthaul links), subject to user QoS requirements and RRH power budget, as a group sparse beamforming problem. The authors in [91] studied a similar problem and designed a joint downlink and uplink user association and beamforming scheme. However, the works in [90] and [91] assumed unlimited capacity fronthaul links. The authors in [92] studied the problem of minimizing the amount of backhaul user data transfer in a multi-cell CoMP network subject to users' QoS constraints and per-base station (BS) power constraints. The authors in [93] aimed at optimizing the end-to-end TCP throughput performance of Mobile Cloud Computing (MCC) users in a C-RAN network through topology configuration and rate allocation. In their work, the impact of the limited C-RAN backhaul capacity is, mainly, on the accuracy of the obtained channel state information (CSI). The authors in [94] studied the delay-aware cooperative beamforming problem in a C-RAN network, where the impact of the limited backhaul capacity was on the cooperating small cells' cluster size. However, [92–94] did not constrain the capacity consumption of individual backhaul links. In other words, they did not impose explicit backhaul capacity constraints on the backhaul links. The authors in [95] studied the problem of AC and total power minimization in a C-RAN subject to fronthaul constraints, power budget constraints, and user QoS requirements. The authors in [96] aimed at maximizing the downlink weighted sum-rate in a C-RAN network, with per-BS power budget and backhaul constraints, using different clustering configurations. The authors in [97] designed a user-specific clustering scheme which aimed at maximizing the average throughput in a C-RAN network while considering the limitations on the existing backhaul links. The authors in [98] minimized the total

network transmit power in a C-RAN subject to users' QoS for both secure communication and efficient wireless power transfer, limited backhaul capacity, and power budget constraints. However, the authors in [95–98] only considered single channel single-tier systems. Moreover, [96–98] did not consider AC to deal with infeasibility problems in case it is not possible to satisfy the QoS of all served UEs. The authors in [99] aimed at maximizing the system throughput by jointly designing a wireless power control and a practical fronthaul quantization scheme, based upon uniform scalar quantization, in an OFDMA C-RAN system. In [100], the authors studied the CoMP transmission in a C-RAN network and targeted at maximizing the users' achievable rate through optimizing the distributed compression rate used at each coordinating node. In [101], the problem of maximizing the weighted sum-rate in a C-RAN network, subject to power and backhaul capacity constraints, through the joint design of precoding and backhaul (multivariate) compression strategies was studied. The authors in [102] reduced the load over the fronthaul links in an OFDMA C-RAN network by employing the techniques of distributed compressive sensing, where distributed fronthaul compression is used at the RRHs, whereas recovery techniques are deployed at the BBU pools.

Different from the cited works, I propose to study the "Tier-Aware" RA problem in a multi-channel (OFDMA) and multi-tier system with a macrocell that is overlaid with C-RAN of small cells. In particular, I formulate the RA problem for both the macrocell tier and the C-RAN tier and show the impact of different RA policies in the macrocell tier on the performance of the C-RAN of small cells tier. Moreover, the C-RAN RA problem accounts for SUEs' QoS requirements, limited capacity fronthaul links, and interference thresholds for the MUEs. Besides, the C-RAN RA problem takes into consideration AC in case it is not possible to support all SUEs at their

target QoS requirements.

4.3 System Model and Assumptions

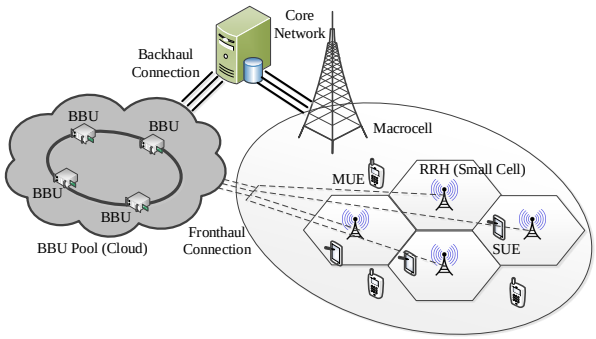

Figure 4.1: Two-tier network with small cells deployed in a C-RAN architecture within the coverage area of a macrocell.

We consider the downlink of a two-tier cellular network where, a single macrocell, referred to by the index B, is overlaid with S small cells. Denote by $\mathcal{S}$ the set of small cells where, $S = |\mathcal{S}|$. The small cells are deployed using a C-RAN architecture as shown in Fig. 4.1. A C-RAN is composed of a centralized processor that is referred to as the baseband unit (BBU) pool or the cloud. Small cells or RRHs[1] are connected to the BBU pool through limited capacity fronthaul links to enable centralized processing, collaborative transmission, and real-time cloud computing. Small cells are responsible for forwarding RF signals to UEs while leaving the baseband processing and general RA tasks to the BBU pool. Hence, small cells can be relatively simple. Such a two-tier setup is referred to as "C-RAN for capacity boosting", where small cells are deployed in a C-RAN architecture within an underlying macrocell in order

[1]Throughout the rest of this Chapter, the terms "small cells" and "RRHs" will be used interchangeably.

to boost network capacity [87]. For simplicity, we assume the macrocell B to be equipped with a single antenna, whereas each small cell is equipped with Ω antennas[2] [103]. All UEs are equipped with single antenna. Denote by $\mathcal{M}$ the set of MUEs with $M = |\mathcal{M}|$. In addition, denote by $\mathcal{F}$ the set of SUEs in the system with $F = |\mathcal{F}|$. To limit computational complexity, an SUE f is constrained to be served by a cluster $\mathcal{S}_f = \left\{s_1, s_2, ..., s_{|\mathcal{S}_f|}\right\}$ of nearby small cells, where $|\mathcal{S}_f| < S$ [104].

We have an OFDMA system, where we denote by $\mathcal{N}$ the set of available sub-channels with $N = |\mathcal{N}|$ and Δf is the bandwidth of a sub-channel n. To maximize spectral efficiency, universal frequency reuse is assumed, where the macrocell and all the small cells have access to the same set of sub-channels $\mathcal{N}$ [66]. Denote by Γ_j^n the sub-channel allocation indicator i.e. $\Gamma_j^n = 1$, if sub-channel n is allocated to UE j and takes the value of 0 otherwise. In this context, we define $\mathcal{N}_j = \{n \in \mathcal{N} : \Gamma_j^n = 1\}$, i.e., the set of sub-channels allocated to UE j. We have $N_j = |\mathcal{N}_j|$. Denote by $P_{i,j}^n$ the allocated power to the link between BS i and UE j on sub-channel n. Moreover, denote by $\mathbf{w}_{i,j}^n \in \mathbb{C}^{\Omega \times 1}$ as the precoding vector at small cell i corresponding to the signal transmitted to SUE j on sub-channel n. Define $g_{B,j}^n$ as the channel power gain between macrocell B and UE j on sub-channel n. On the other hand, $\mathbf{h}_{i,j}^n \in \mathbb{C}^{\Omega \times 1}$ denotes the complex channel gain between small cell i and UE j on sub-channel n. Channel gains are time varying and account for path-loss and fast fading.

For an MUE m served by macrocell B on sub-channel n, its received signal can

[2]This can be attributed to the fact that the macrocell tier is assumed to be an already existing one that provides universal coverage, whereas the C-RAN of small cells tier is deployed on top of the macrocell tier as an enhancement layer.

be written as:

$$\vartheta_{B,m}^{n} = \underbrace{H_{B,m}^{n}\sqrt{P_{B,m}^{n} l_{B,m}} u_{m}^{n}}_{\text{desired signal}} + \underbrace{\sum_{f=1}^{F}\left(\sum_{s\in\mathcal{S}_f} \mathbf{h}_{s,m}^{nH}\mathbf{w}_{s,f}^{n}\right) u_{f}^{n}}_{\text{cross-tier interference signal}} + \omega_{m}^{n} \tag{4.1}$$

where u_j^n is the transmitted information symbol with unit variance for UE j on sub-channel n, $l_{B,j}$ is the path-loss between macrocell B and UE j, $H_{B,j}^n$ is the fast fading coefficient between macrocell B and UE j on sub-channel n, and ω_j^n is the additive white Gaussian noise (AWGN) in sub-channel n at UE j with zero mean and variance N_o. In (4.1), cross-tier interference is due to the transmissions of small cells to their SUEs on sub-channel n.

In the subsequent sections, the cross-tier interference signal power in (4.1) will be replaced by the maximum interference I_m^n that MUE m can tolerate on sub-channel n. In this context, the received SINR $\gamma_{B,m}^n$ of an MUE m served by macrocell B on a sub-channel n can be expressed as follows:

$$\gamma_{B,m}^{n} = \frac{P_{B,m}^{n} g_{B,m}^{n}}{I_{m}^{n} + N_o} \tag{4.2}$$

where $g_{B,j}^n \triangleq \left|H_{B,j}^n\right|^2 l_{B,j}$ and I_m^n is the maximum tolerable interference level at MUE m on sub-channel n. According to (4.2), the following constraint holds for the transmission powers of small cells using sub-channel n:

$$\Gamma_{m}^{n}\left(\sum_{f=1}^{F}\left|\sum_{s\in\mathcal{S}_f}\mathbf{h}_{s,m}^{nH}\mathbf{w}_{s,f}^{n}\right|^{2}\right) \leq \Gamma_{m}^{n} I_{m}^{n} \tag{4.3}$$

where the constraint is active only if sub-channel n is allocated to MUE m, i.e., $\Gamma_m^n = 1$. Hence, (4.2) can be referred to as the worst-case SINR assuming the inequality in

(4.3) holds with equality[3].

For an SUE f served by a cluster of RRHs $\mathcal{S}_f$ on sub-channel n, its received signal can be written as:

$$\vartheta^n_{\mathcal{S}_f,f} = \underbrace{\left(\sum_{s\in\mathcal{S}_f} \mathbf{h}^{nH}_{s,f}\mathbf{w}^n_{s,f}\right) u^n_f}_{\text{desired signal}} + \underbrace{\sum_{i=1,i\neq f}^{F}\left(\sum_{l\in\mathcal{S}_i} \mathbf{h}^{nH}_{l,f}\mathbf{w}^n_{l,i}\right) u^n_i}_{\text{co-tier interference signal}} + \underbrace{\sum_{m=1}^{M}\Gamma^n_m H^n_{B,f}\sqrt{P^n_{B,m}l_{B,f}}u^n_m}_{\text{cross-tier interference signal}} + \omega^n_f \tag{4.4}$$

The interference is due to transmissions of small cells to other SUEs and macrocell transmissions to its MUEs. Hence, its SINR can be defined as:

$$\gamma^n_{\mathcal{S}_f,f} = \frac{\left|\sum_{s\in\mathcal{S}_f} \mathbf{h}^{nH}_{s,f}\mathbf{w}^n_{s,f}\right|^2}{\sum_{i=1,i\neq f}^{F}\left|\sum_{l\in\mathcal{S}_i} \mathbf{h}^{nH}_{l,f}\mathbf{w}^n_{l,i}\right|^2 + \sum_{m=1}^{M}\Gamma^n_m P^n_{B,m}g^n_{B,f} + N_o} \tag{4.5}$$

4.4 Formulation of RA Problem for Macrocell

We develop the mathematical formulation for the RA framework for the macrocell. The macrocell allocates sub-channels to its MUEs and specifies, on each allocated sub-channel, the suitable power level and the maximum tolerable interference level. The objective of the macrocell is to allocate resources to its MUEs in a way that can sustain the highest possible interference levels. This problem has been studied before in our previous work [8]. However, uniform power allocation was assumed in [8].

[3]Note that (4.3) is also known as the interference temperature constraint in cognitive radio networks. It has been used, as well, extensively in two tier macrocell small cell networks to limit the cross-tier interference introduced to MUEs, who have a strictly higher priority in accessing the underlying frequency bands [66].

Moreover, it was assumed that the QoS requirements for the MUEs can be satisfied using a single sub-channel only. In the current work, we relax those simplifying assumptions. The formulated optimization problem can be described mathematically as follows:

$$\max_{\left\{\Gamma_m^n, I_m^n, P_{B,m}^n\right\}} \sum_{m=1}^{M}\sum_{n=1}^{N} I_m^n$$

subject to

$$\text{C1}: \sum_{n=1}^{N} \Delta f \log_2\left(1+\frac{P_{B,m}^n g_{B,m}^n}{I_m^n + N_o}\right) \geq R_m,\ \forall m \in \mathcal{M}$$

$$\text{C2}: \sum_{m=1}^{M}\sum_{n=1}^{N} P_{B,m}^n \leq P_{B,max}$$

$$\text{C3}: \sum_{m=1}^{M} \Gamma_m^n \leq 1,\ \forall n \in \mathcal{N}$$

$$\text{C4}: P_{B,m}^n \leq \Gamma_m^n P_{B,max},\ \forall m \in \mathcal{M}, n \in \mathcal{N}$$

$$\text{C5}: I_m^n \geq (1-\Gamma_m^n)\, I_{max},\ \forall m \in \mathcal{M}, n \in \mathcal{N}$$

$$\text{C6}: \Gamma_m^n \in \{0,1\}, P_{B,m}^n \geq 0, I_m^n \geq I_{min}, I_m^n \leq I_{max},\ \forall m \in \mathcal{M}, n \in \mathcal{N} \tag{4.6}$$

In (4.6), the objective is to maximize the sum of the tolerable interference levels. C1 is a minimum data rate requirement for each MUE m. C2 is a power budget constraint for the macrocell B. C3 dictates that sub-channel n is allocated to at most one MUE. C4 enforces the power level $P_{B,m}^n$ to zero if sub-channel n is not allocated to MUE m. Similarly, C5 enforces the maximum tolerable interference level I_m^n to I_{max} if sub-channel n is not allocated to MUE m. Finally, C6 indicates that Γ_m^n is a binary variable, whereas $P_{B,m}^n$ is a continuous non-negative variable. Moreover, C6 sets lower and upper bounds on the values of I_m^n. Note that the upper bound value I_{max} is a relatively large value which is imposed to guarantee a finite value for the

objective function. In other words, a sub-channel n that is not allocated to an MUE m will have $I_m^n = I_{max}$. From C5 and C6, one can observe that the formulation in (4.6) will try to set as many Γ_m^n as possible to zero in order to yield $I_m^n = I_{max}$. In other words, this formulation tends to minimize the number of used sub-channels as was already discussed in [8] and [105].

The optimization problem in (4.6) is a mixed integer non-linear problem whose feasible set is non-convex and is NP-hard to solve in general. The non-convexity is attributed to two reasons. The first one is the combinatorial nature of sub-channel allocation. The second one is due to the constraint C1 being neither jointly concave nor convex in the variables $P_{B,m}^n$ and I_m^n. In the following sub-section, we examine the RA problem for a single MUE in order to see how the system behaves according to the proposed objective of maximizing the sum of tolerable interference levels. Hence, no sub-channel allocation indicator will exist for the single MUE scenario. Then, we will discuss how to solve the problem for the mutli-MUE case.

4.4.1 Macrocell RA Problem With Single MUE

In this subsection, we analyze the RA problem for a single MUE m based on the objective of maximizing the sum of tolerable interference levels. We have the following

optimization problem:

$$\max_{\left\{I_m^n, P_{B,m}^n\right\}} \sum_{n=1}^{N} I_m^n$$

subject to

$$\text{C1}: \sum_{n=1}^{N} \Delta f \log_2 \left(1 + \frac{P_{B,m}^n g_{B,m}^n}{I_m^n + N_o}\right) \geq R_m$$

$$\text{C2}: \sum_{n=1}^{N} P_{B,m}^n \leq P_{B,max}$$

$$\text{C3}: P_{B,m}^n \geq 0, I_m^n \geq I_{min}, I_m^n \leq I_{max}, n \in \mathcal{N} \tag{4.7}$$

Define the following Lagrange dual function:

$$\mathcal{L}\left(\{I_m^n, P_{B,m}^n\}, \alpha, \lambda\right) = \sum_{n=1}^{N} I_m^n + \alpha \left(\sum_{n=1}^{N} \Delta f \log_2 \left(1 + \frac{P_{B,m}^n g_{B,m}^n}{I_m^n + N_o}\right) - R_m\right) + \lambda \left(P_{B,max} - \sum_{n=1}^{N} P_{B,m}^n\right) \tag{4.8}$$

where $\alpha \geq 0$ and $\lambda \geq 0$ are the Lagrange multipliers for the constraints C1 and C2, respectively. Denote by I_m^{n*} and $P_{B,m}^{n*}$ the optimal solution for problem (4.7). Applying the KKT conditions, we have the following necessary conditions for the

optimality of I_m^{n*} and $P_{B,m}^{n*}$:

$$\frac{\partial \mathcal{L}\left(\cdots\right)}{\partial P_{B,m}^{n*}} \begin{cases} < 0, & P_{B,m}^{n*} = 0 \\ = 0, & 0 < P_{B,m}^{n*} < P_{B,max}\ , \forall n \in \mathcal{N} \\ > 0, & P_{B,m}^{n*} = P_{B,max} \end{cases} \tag{4.9}$$

$$\frac{\partial \mathcal{L}\left(\cdots\right)}{\partial I_m^{n*}} \begin{cases} < 0, & I_m^{n*} = I_{min} \\ = 0, & I_{min} < I_m^{n*} < I_{max}\ , \forall n \in \mathcal{N} \\ > 0, & I_m^{n*} = I_{max} \end{cases} \tag{4.10}$$

$$\alpha \left(\sum_{n=1}^{N} \Delta f \log_2 \left(1 + \frac{P_{B,m}^{n*} g_{B,m}^{n}}{I_m^{n*} + N_o} \right) - R_m \right) = 0 \tag{4.11}$$

$$\lambda \left(P_{B,max} - \sum_{n=1}^{N} P_{B,m}^{n*} \right) = 0 \tag{4.12}$$

Those conditions are necessary but not sufficient for the optimality of I_m^{n*} and $P_{B,m}^{n*}$ owing to the non-convexity of the problem (4.7) [44]. However, we will propose an algorithm that will solve the equations resulting from the KKT conditions and show that it yields the optimal solution.

Proposition 4.4.1. *At optimality, the rate constraint C1 and the total power budget constraint C2 in (4.7) hold with equality.*

Proof. To establish this result, assume that, at optimality, the constraints C1 and C2 in (4.7) are inactive and that the optimal solution is $\{I_m^{n\wedge}, P_{B,m}^{n\wedge}\}$. However, we observe that the objective function is monotonically increasing in I_m^n, whereas the constraint C1 is monotonically decreasing in I_m^n. Hence, for a given power allocation $\{P_{B,m}^{n\wedge}\}$, $\{I_m^{n\wedge}\}$ can be increased to $\{I_m^{n\vee}\}$ until the constraint C1 holds with equality. On the other hand, on an allocated sub-channel n, $I_m^{n\vee}$ is allowed to increase further only if $P_{B,m}^{n\wedge}$ is allowed to increase. Since the available power budget $P_{B,max}$ is finite,

$I_m^{n\vee}$ can be increased further until constraint C2 holds with equality. This results in a strictly larger objective function which contradicts the assumption that the solution $\{I_m^{n\wedge}, P_{B,m}^{n\wedge}\}$ is optimal. □

Proposition 4.4.2. *On an allocated sub-channel n with $0 < P_{B,m}^{n*} < P_{B,max}$ and $I_{min} < I_m^{n*} < I_{max}$, $P_{B,m}^{n*}$ and I_m^{n*} can be expressed as follows:*

$$P_{B,m}^{n*} = \left(\frac{\alpha \Delta f}{\lambda \ln 2} - \frac{I_m^{n*} + N_o}{g_{B,m}^n} \right) \tag{4.13}$$

$$I_m^{n*} = \frac{-P_{B,m}^{n*} g_{B,m}^n}{2} - N_o + \frac{\sqrt{\left(P_{B,m}^{n*} g_{B,m}^n\right)^2 + \left(4\alpha \Delta f P_{B,m}^{n*} g_{B,m}^n / \ln 2\right)}}{2} \tag{4.14}$$

and the following relation holds:

$$\frac{P_{B,m}^{n*}}{I_m^{n*} + N_o} = \frac{1}{\lambda} \tag{4.15}$$

Proof. This can be proved by setting both differentiations of $\mathcal{L}\left(\{I_m^n, P_{B,m}^n\}, \alpha, \lambda\right)$ with respect to I_m^{n*} and $P_{B,m}^{n*}$ to zero. From the two resulting equations, (4.13)-(4.15) can be deduced, keeping in mind that I_m^{n*} cannot be negative. □

Proposition 4.4.3. *Let the set of sub-channels $\mathcal{N} = \{1, 2, ..., N\}$ be sorted such that $g_{B,m}^1 \geq g_{B,m}^2 \geq ... \geq g_{B,m}^N$ and assume that it is possible to serve MUE m by the first $K \leq N$ best sub-channels. On the set $\mathcal{K} = \{1, 2, ..., K\}$ of sub-channels, we will have $I_m^K \geq I_{min}$, whereas $I_m^n = I_{min}, n = 1, 2, ..., K-1$. On the other hand, $I_m^n = I_{max}, n = K+1, K+2, ..., N$.*

Proof. To see this, assume at the beginning that the data requirement R_m is low such that it is possible to serve MUE m by one sub-channel only which will be sub-channel 1 in our case (the sub-channel with the highest gain). According to *Proposition* 4.4.1,

constraints C1 and C2 should hold with equality. Hence, $P_{B,m}^1 = P_{B,max}$ and I_m^1 can be determined from constraint C1. For all other $n = 2, 3, ..., N$, $P_{B,m}^n = 0$ and $I_m^n = I_{max}$ with an objective function value $(N-1)\, I_{max} + I_m^1$. As the rate requirement R_m gradually increases, I_m^1 will be the only variable affected to satisfy constraint C1, whereas all other $I_m^n, n = 2, 3, ..., N$ will remain equal to I_{max}. Sub-channel $n = 2$ will not have a value of $P_{B,m}^n$ that is greater than zero and a value of I_m^n that is less than I_{max} until I_m^1 reaches I_{min}. Similar arguments apply for the case of multiple allocated sub-channels. □

According to the previous discussions, problem (4.7) can be solved optimally by using **Algorithm** 8. **Algorithm** 8 seeks a solution at which the constraint C1 holds with equality and the whole power budget in C2 is consumed.

Algorithm 8 Optimal interference and power allocation algorithm

1: Sort the set of available sub-channels $\mathcal{N} = \{1, 2, ..., N\}$ such that $g_{B,m}^1 \geq g_{B,m}^2 \geq ... \geq g_{B,m}^N$
2: Initialize $k = 1$
3: Assume $P_{B,m}^k = P_{B,max}$ and find I_m^k using constraint C1
4: **if** $I_m^k \geq I_{min}$ **then**
5: Set $P_{B,m}^n = 0$ and $I_m^n = I_{max}, \forall n = 2, 3, ..., N$
6: Terminate as the optimal solution has been obtained
7: **else**
8: Increment k
9: **repeat**
10: Set $I_m^n = I_{min}, \forall n = 1, 2, ..., k-1$
11: Solve the following system of non-linear equations in the unknowns α and λ using (4.13)-(4.15):

$$\sum_{n=1}^{k} P_{B,m}^n = P_{B,max}$$
$$\sum_{n=1}^{k-1} \log_2\left(1 + \frac{P_{B,m}^n g_{B,m}^n}{I_{min} + N_o}\right) + \log_2\left(1 + \frac{g_{B,m}^k}{\lambda}\right) = \frac{R_m}{\Delta f}$$

12: **until** A feasible solution is obtained or else increment k
13: **end if**
14: Set $I_m^n = I_{max}$ and $P_{B,m}^n = 0,\ \forall n = k+1, ..., N$
15: Undo the sub-channels sorting

The idea of **Algorithm** 8 is based upon hypothesis testing. First, it makes the assumption that only one sub-channel is needed (which is the best sub-channel) and checks whether there is a feasible solution at which the constraints hold with equality. If not, the hypothesis is rejected and a new hypothesis is formed by including the second best sub-channel and so forth. The total number of hypotheses is N which is the total number of available sub-channels. This implies a linear complexity in the number of sub-channels. On the other hand, step 1 requires the sub-channels to be sorted. The complexity of the best sorting algorithm is $\mathcal{O}\left(N \log_2 N\right)$ [106].

4.4.2 Macrocell RA Problem With Multiple MUEs

In the single-MUE resource allocation case, a greedy approach was adopted, where the sub-channels with the highest gains are assigned first. This approach leads to the optimal allocation in the sense of maximizing the sum of tolerable interference levels. The RA problem becomes much more difficult in the case of multiple MUEs due to the coupling among MUEs in terms of the requirement of the orthogonality of the allocated sub-channels and the share that each MUE receives in the available power budget $P_{B,max}$. A more tractable approach that has been followed in the literature is to separate the RA problem into two steps [40], [41], [42]. The first step is to perform sub-channel allocation for given power and interference allocation. Then for the resulting sub-channel allocation, power and interference allocation can be performed.

Sub-channel allocation in the macrocell

In this section, we propose a suboptimal sub-channel allocation algorithm. In this algorithm, uniform power distribution and $I_m^n = I_{min}$ are assumed across all sub-channels. We define $R_{ach,m}$ as the achieved data rate for MUE m, $\mathcal{A}$ as the set

of available sub-channels for allocation, N_{ac} as the total number of allocated sub-channels, and $\mathcal{M}_{cand}$ as the set of candidate MUEs for sub-channel allocation. The steps are given in **Algorithm** 9.

Algorithm 9 Sub-channel allocation in the macrocell

1: Initialize $N_{ac} = N$
2: **repeat**
3: Initialize $R_{ach,m} = 0$, $\mathcal{N}_m = \phi$, $P^n_{B,m} = P_{unif} = \frac{P_{B,max}}{N_{ac}}$, $\forall n \in \mathcal{N}, m \in \mathcal{M}$ and $\mathcal{A} = \mathcal{N}$
4: **Phase 1** (Initialize $\mathcal{M}_{cand} = \mathcal{M}$)
5: **repeat**
6: $n^*, m^* = \underset{n \in \mathcal{A},\ m \in \mathcal{M}_{cand}}{\arg\max}\ g^n_{B,m}$
7: Set $\Gamma^{n^*}_{m^*} = 1$
8: Update $\mathcal{N}_{m^*} = \mathcal{N}_{m^*} \cup \{n^*\}$, $\mathcal{A} = \mathcal{A} - \{n^*\}$, $R_{ach,m^*} = R_{ach,m^*} + \Delta f \log_2\left(1 + \frac{P_{unif} g^{n^*}_{B,m^*}}{I_{min} + N_o}\right)$ and $\mathcal{M}_{cand} = \mathcal{M}_{cand} - \{m^*\}$
9: **until** $\mathcal{M}_{cand} = \varnothing$
10: **Phase 2**
11: **repeat**
12: $m^* = \underset{m \in \mathcal{M}}{\arg\min} R_{ach,m}$
13: $n^* = \underset{n \in \mathcal{A}}{\arg\max} g^n_{B,m^*}$
14: Set $\Gamma^{n^*}_{m^*} = 1$
15: Update $\mathcal{N}_{m^*} = \mathcal{N}_{m^*} \cup \{n^*\}$, $\mathcal{A} = \mathcal{A} - \{n^*\}$ and $R_{ach,m^*} = R_{ach,m^*} + \Delta f \log_2\left(1 + \frac{P_{unif} g^{n^*}_{B,m^*}}{I_{min} + N_o}\right)$
16: **until** No MUE m has $R_{ach,m} < R_m$
17: Update N_{ac}
18: **until** No change in the total number of allocated sub-channels N_{ac}

In **Algorithm** 9, it is assumed at the beginning that all sub-channels are allocated, i.e., $N_{ac} = N$. **Phase 1** ensures that each MUE $m \in \mathcal{M}$ is allocated one good sub-channel. Note that maximizing the sum of tolerable interference levels leads to a minimal use of the available bandwidth [8, 105]. Hence, **Phase 2** allocates enough number of sub-channels for each MUE just to satisfy its data rate requirements. In step 17, the total number of allocated sub-channels N_{ac} is updated, based on which, the iterations may repeat. **Algorithm** 9 terminates when there is no change in the number of allocated sub-channels. Steps 5-9 have a complexity that is of $\mathcal{O}(NM^2)$. Steps 11-16, on the other hand, have a complexity that is of $\mathcal{O}(N(N+M))$. The

overall algorithm will repeat at most $\mathcal{O}(N)$ times.

Power and maximum tolerable interference levels specification in the macrocell

For the given sub-channel allocation, the problem of power and maximum tolerable interference level allocation can be reformulated as shown in (4.16).

This problem is still non-convex owing to the constraint C1. One way to deal with such type of problems is by the use of an SCA approach [88, 89]. SCA approaches are capable of obtaining solutions that satisfy the KKT conditions of the original non-convex problem. The general idea of an SCA algorithm is to approximate the original non-convex problem by a series of convex approximations. Therefore, starting at some initial point, we solve an approximate convex problem and use the solution of that approximate problem as an initial point for a new convex problem. The entire procedure repeats until it converges to a solution that satisfies the KKT conditions of the original non-convex problem. Similar approaches have been adopted in [107] and [108].

$$\max_{\left\{I_m^n, P_{B,m}^n\right\}} \sum_{m=1}^{M} \sum_{n \in \mathcal{N}_m} I_m^n$$

subject to

$$\text{C1}: \sum_{n \in \mathcal{N}_m} \Delta f \log_2 \left(1 + \frac{P_{B,m}^n g_{B,m}^n}{I_m^n + N_o}\right) \geq R_m, \ \forall m \in \mathcal{M}$$

$$\text{C2}: \sum_{m=1}^{M} \sum_{n \in \mathcal{N}_m} P_{B,m}^n \leq P_{B,max}$$

$$\text{C3}: P_{B,m}^n \geq 0, I_m^n \geq I_{min}, I_m^n \leq I_{max}, \ \forall m \in \mathcal{M}, n \in \mathcal{N} \tag{4.16}$$

Problem (4.16) is said to have a difference of convex (d.c.) structure [107] as constraint C1 can be written as the difference of two concave functions as follows:

$$\Delta f \log_2\left(1+\frac{P_{B,m}^n g_{B,m}^n}{I_m^n+N_o}\right)=\Delta f\left\{\log_2\left(I_m^n+N_o+P_{B,m}^n g_{B,m}^n\right)-\log_2\left(I_m^n+N_o\right)\right\} \tag{4.17}$$

It is obvious from (4.17) that the term $\log_2\left(I_m^n+N_o\right)$ is concave in I_m^n, which is the reason for non-convexity of the problem. Thus, by using an SCA approach, in the z^{th} iteration, the term $\log_2\left(I_m^n+N_o\right)$ will be replaced by an approximating convex one. Let $g\left(I_m^n\right)=\log_2\left(I_m^n+N_o\right)$ and $\bar{g}\left(I_m^n, I_m^{n,(z-1)}\right)=a_m^{n,(z)}\left(I_m^n+N_o\right)+b_m^{n,(z)}$ be the approximating convex one around the point $I_m^{n,(z-1)}$, where $I_m^{n,(z-1)}$ is the optimal solution of the approximated problem in the $(z-1)^{th}$ iteration. According to [88], $\bar{g}\left(I_m^n, I_m^{n,(z-1)}\right)$ must have the following properties:

$$\begin{aligned}
&g\left(I_m^n\right)\leq \bar{g}\left(I_m^n, I_m^{n,(z-1)}\right),\ \forall I_m^n\\
&g\left(I_m^{n,(z-1)}\right)=\bar{g}\left(I_m^{n,(z-1)}, I_m^{n,(z-1)}\right)\\
&\partial g\left(I_m^{n,(z-1)}\right)/\partial I_m^n=\partial \bar{g}\left(I_m^{n,(z-1)}, I_m^{n,(z-1)}\right)/\partial I_m^n
\end{aligned} \tag{4.18}$$

According to the properties in (4.18), the approximation constants $a_m^{n,(z)}$ and $b_m^{n,(z)}$ are chosen as follows:

$$\begin{aligned}
a_m^{n,(z)}&=\frac{1}{\ln 2\left(I_m^{n,(z-1)}+N_o\right)}\\
b_m^{n,(z)}&=\log_2\left(I_m^{n,(z-1)}+N_o\right)-\frac{1}{\ln 2}
\end{aligned} \tag{4.19}$$

Therefore, in the z^{th} iteration, constraint C1 in (4.16) can be re-written as:

$$\sum_{n\in\mathcal{N}_m} \Delta f \left\{\log_2\left(I_m^n + N_o + P_{B,m}^n g_{B,m}^n\right) - \bar{g}\left(I_m^n, I_m^{n,(z-1)}\right)\right\} \geq R_m,\ \forall m \in \mathcal{M} \quad (4.20)$$

Note that, with constraint C1 in (4.16) expressed as shown in (4.20), problem (4.16) now becomes a convex optimization problem that can be efficiently solved by the interior point method [44]. The general SCA algorithm for solving (4.16) is shown in **Algorithm** 10.

Proposition 4.4.4. *In **Algorithm** 10, the iterates produce a sequence of non-decreasing objective function values that is guaranteed to converge. Moreover, convergence takes place to a solution that satisfies the KKT conditions of (4.16).*

Proof. The convergence of **Algorithm** 10 is guaranteed as follows. In the z^{th} iteration, the optimal solution $\{I_m^{n,(z)}\}$ and $\{P_{B,m}^{n,(z)}\}$ are feasible for the approximated problem in the $(z+1)^{th}$ iteration. Thus, the objective function of the approximated problem in the $(z+1)^{th}$ iteration is greater than or equal to that in the z^{th} iteration. In other words, **Algorithm** 10 produces a sequence of non-decreasing objective function values. In addition, the problem is bounded above due to the imposed constraints. Hence, **Algorithm** 10 converges to some local optimum solution of (4.16). From the analysis in [88] and [89], this solution can be shown to satisfy the KKT conditions of the original problem (4.16). □

Although the resulting solution from **Algorithm** 10 is a local optimum in general, such an SCA approach generally yields solutions that are very close to the global optimum [109].

Algorithm 10 Successive convex approximation algorithm for problem (4.16)

1: Initialize $I_m^{n,(0)} = I_{min},\ \forall n \in \mathcal{N}, m \in \mathcal{M}$ and set $z = 1$
2: **repeat**
3: Form the z^{th} approximated problem by expressing constraint C1 in (4.16) as shown in (4.20)
4: Solve the z^{th} approximated convex problem to obtain $I_m^{n,(z)}$
5: Increment z
6: **until** Convergence

4.5 Formulation of RA Problem for C-RAN

Given the resource allocation results of the macrocell problem, the objective of the RA problem for the C-RAN is to minimize the total downlink transmit power subject to the QoS requirements of SUEs, interference thresholds for MUEs, and the fronthaul capacity constraints. This RA problem can be stated as follows:

$$\min_{\{\Gamma_f^n, \mathbf{w}_{s,f}^n\}} \sum_{s=1}^{S}\sum_{f=1}^{F}\sum_{n=1}^{N} \left\|\mathbf{w}_{s,f}^n\right\|^2$$

subject to

$$\text{C1}: \sum_{n=1}^{N} \Delta f \log_2\left(1+\gamma_{\mathcal{S}_f,f}^n\right) \geq R_f,\ \forall f \in \mathcal{F}$$

$$\text{C2}: \left\|\mathbf{w}_{s,f}^n\right\|^2 \leq \Gamma_f^n P_{s,max}\ , \forall s \in \mathcal{S}, \forall f \in \mathcal{F}, n \in \mathcal{N}$$

$$\text{C3}: \sum_{n=1}^{N} \Gamma_f^n \leq q_f,\ \forall f \in \mathcal{F}$$

$$\text{C4}: \sum_{f=1}^{F}\sum_{n=1}^{N} \left\|\mathbf{w}_{s,f}^n\right\|^2 \leq P_{s,max},\ \forall s \in \mathcal{S}$$

$$\text{C5}: \sum_{f=1}^{F} \Psi_{st}\left(\sum_{n=1}^{N} \left\|\mathbf{w}_{s,f}^n\right\|^2\right) \leq C_{s,max},\ \forall s \in \mathcal{S}$$

$$\text{C6}: \Gamma_m^n \left(\sum_{f=1}^{F}\left|\sum_{s\in\mathcal{S}_f} \mathbf{h}_{s,m}^{nH}\mathbf{w}_{s,f}^n\right|^2\right) \leq \Gamma_m^n I_m^n,\ \forall n \in \mathcal{N}$$

$$\text{C7}: \Gamma_f^n \in \{0,1\},\ \mathbf{w}_{s,f}^n \in \mathbb{C}^{\Omega\times 1}\ , \forall s \in \mathcal{S}, \forall f \in \mathcal{F}, n \in \mathcal{N} \tag{4.21}$$

where the objective in (4.21) is to perform sub-channel and power allocation to minimize the total downlink transmit power. Constraint C1 is the data rate constraint for each SUE f. Constraint C2 enforces the transmission power on an unallocated sub-channel to be zero. Constraint C3 limits the number of allocated sub-channels to SUE f to q_f sub-channels. The idea is to prevent the cloud from greedily allocating all the available sub-channels to its SUEs in order to leave some sub-channels for the other network tiers (e.g., for D2D communication). C4 is an indication of the total power budget available at each RRH s which is $P_{s,max}$. C5 is the fronthaul constraint which limits the number of baseband signals transmitted on the fronthaul link between the cloud and RRH s to $C_{s,max}$, where $\Psi_{st}(x), x \geq 0$ is the step function which is defined as:

$$\Psi_{st}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \end{cases} \tag{4.22}$$

Recall that $\left\|\mathbf{w}_{s,f}^n\right\|^2 = P_{s,f}^n$ denotes the power allocated by RRH s to SUE f on sub-channel n. If $\left\|\mathbf{w}_{s,f}^n\right\|^2 = 0, \forall n \in \mathcal{N}$, this indicates that RRH s does not serve SUE f and that the fronthaul link between the cloud and RRH s does not carry the baseband signal for SUE f. On the other hand, if there is at least one sub-channel n such that $\left\|\mathbf{w}_{s,f}^n\right\|^2 > 0$, then RRH s serves SUE f and the fronthaul link between the cloud and RRH s carries the baseband signal for SUE f. Owing to the finite capacity of the fronthaul link between RRH s and the cloud, there is an upper bound $C_{s,max}$ on the number of baseband signals that can be transmitted from the cloud to an RRH s. C6 puts a limit on the total interference introduced to MUE m by the downlink transmissions of RRHs on sub-channel n, which is active only if sub-channel n is allocated to MUE m, i.e., $\Gamma_m^n = 1$. Finally, C7 indicates that Γ_f^n is a binary variable

whereas $\mathbf{w}_{s,f}^{n}$ is an $\Omega \times 1$ complex vector.

Problem (4.21) is an MINLP non-convex problem whose optimal solution is intractable due to the combinatorial nature of sub-channel allocation, the QoS constraint C1, and the discontinuous step function $\Psi_{st}(x)$ in C5. Moreover, the problem in (4.21) can easily become infeasible when, for instance, the QoS constraints for all the SUEs cannot be achieved simultaneously. Therefore, some users should be dropped by employing admission control [55, 110]. It becomes reasonable then to maximize the number of SUEs that can be admitted at their target QoS, while minimizing the total downlink transmit power. We shall account for infeasibility and how to incorporate admission control in a later section. Now, since it will be prohibitive to jointly perform sub-channel allocation, power allocation and admission control, it will be separated into two phases: a sub-channel allocation phase and an admission control and power allocation phase.

4.5.1 Sub-channel Allocation in the C-RAN

For the sub-channel allocation phase, it will be beneficial to distinguish between sub-channels that are already allocated to MUEs and those sub-channels that are not. Denote the stacked $\Omega|\mathcal{S}_f| \times 1$ channel vector between a UE i and the set of RRHs $\mathcal{S}_f$, serving SUE f, on sub-channel n as $\mathbf{h}_{\mathcal{S}_f,i}^{n} = \left[\mathbf{h}_{s_1,i}^{nH}, \mathbf{h}_{s_2,i}^{nH}, ..., \mathbf{h}_{s_{|\mathcal{S}_f|},i}^{nH}\right]^{H}$. Define the set $\mathcal{N}' = \{n \in \mathcal{N} : \Gamma_m^n = 0,\ \forall m \in \mathcal{M}\}$ as the set of sub-channels that are not allocated to MUEs. The cloud will start by allocating those sub-channels in the set $\mathcal{N}'$ to each SUE f. More specifically, the cloud will sort the sub-channels in the set $\mathcal{N}'$ for each SUE f in a descending order according to the metric: $\left(\left\|\mathbf{h}_{\mathcal{S}_f,f}^{n}\right\|\right)$. In other words, the cloud will allocate the best sub-channels first to each SUE. If those sub-channels in the set $\mathcal{N}'$ are not enough to satisfy the QoS requirement of the SUEs, the cloud

Algorithm 11 Sub-channel allocation in the C-RAN

Given the set $\mathcal{N}'$
for $f = 1 : F$ **do**
  Initialize $\mathcal{N}_f = \varnothing$, $\mathcal{A} = \mathcal{N}'$
  Sort sub-channels in the set $\mathcal{A}$ in descending order according to the metric $\left(\left\|\mathbf{h}^n_{\mathcal{S}_f,f}\right\|\right)$
  if $q_f \leq |\mathcal{A}|$ **then**
    Allocate the first q_f sub-channels in the set $\mathcal{A}$ to SUE f
  else
    Allocate all sub-channels in the set $\mathcal{A}$ to SUE f
  end if
  Update $\mathcal{N}_f$
end for
if All SUEs have their QoS requirement q_f satisfied **then**
  Terminate
else
  Given the set $\mathcal{N}''$
  for $f = 1 : F$ **do**
    Initialize $\mathcal{A} = \mathcal{N}''$
    Sort sub-channels in the set $\mathcal{A}$ in descending order according to the metric $\left(\left\|\mathbf{h}^n_{\mathcal{S}_f,f}\right\| / \left\|\mathbf{h}^n_{\mathcal{S}_f,m}\right\|\right)$
    Allocate enough sub-channels from the set $\mathcal{A}$ to SUE f such that its QoS requirement q_f is satisfied
    Update $\mathcal{N}_f$
  end for
end if

will start allocating the remaining sub-channels in the set $\mathcal{N}'' = \{n : n \in \mathcal{N} - \mathcal{N}'\}$, $(\mathcal{N}' \bigcup \mathcal{N}'' = \mathcal{N}, \mathcal{N}' \bigcap \mathcal{N}'' = \varnothing)$ to the SUEs. In this case, the cloud will sort the sub-channels in the set $\mathcal{N}''$ for each SUE f in a descending order according to the metric: $\left(\left\|\mathbf{h}^n_{\mathcal{S}_f,f}\right\| / \left\|\mathbf{h}^n_{\mathcal{S}_f,m}\right\|\right)$. The idea of this metric is to allocate sub-channels to SUEs that have high gains and that will cause low interference to the MUEs. The overall sub-channel allocation algorithm can be summarized in **Algorithm** 11 the complexity of which is of $\mathcal{O}\left(FN \log_2 N\right)$.

4.5.2 Joint Power Allocation and Admission Control in the C-RAN

Given sub-channel allocation, the C-RAN can perform power allocation and admission control. First, for the set of RRHs $\mathcal{S}_f$ serving SUE f, let us define

$\mathbf{w}^n_{\mathcal{S}_f,f} = \left[\mathbf{w}^{nH}_{s_1,f}, \mathbf{w}^{nH}_{s_2,f}, ..., \mathbf{w}^{nH}_{s_{|\mathcal{S}_f|},f}\right]^H$, where $\mathbf{w}^n_{\mathcal{S}_f,f} \in \mathbb{C}^{\Omega|\mathcal{S}_f|\times 1}$. Second, we introduce the unconstrained real admission control variables y_f that take the value of 0 only if SUE f is admitted. Third, let $\mathbf{Q}_{s,f} = \text{diag}\left(\left[\overbrace{\mathbf{0}_{1\times\Omega}}^{s_1}, ..., \overbrace{\mathbf{1}_{1\times\Omega}}^{s_i}, \overbrace{\mathbf{0}_{1\times\Omega}}^{s_{i+1}}, ..., \overbrace{\mathbf{0}_{1\times\Omega}}^{s_{|\mathcal{S}_f|}}\right]\right)$, if $s_i = s$ and $s \in \mathcal{S}_f$. Hence, $\left\|\mathbf{w}^n_{s,f}\right\|^2 = \left\|\mathbf{Q}_{s,f}\mathbf{w}^n_{\mathcal{S}_f,f}\right\|^2$. In addition, by introducing the additional auxiliary SINR variables μ^n_f, we can rewrite the constraint C1 in (4.21) as follows:

$$\gamma^n_{\mathcal{S}_f,f} \geq \mu^n_f, \ \forall f \in \mathcal{F}, n \in \mathcal{N}_f \tag{4.23}$$

$$\sum_{n\in\mathcal{N}_f} \Delta f \log_2\left(1+\mu^n_f\right) \geq R_f, \ \forall f \in \mathcal{F} \tag{4.24}$$

The equivalence between equations (4.23) and (4.24) and the constraint C1 in (4.21) stems from the fact that the inequalities in (4.23) and (4.24) hold with equality at the optimum solution. Moreover, by exploiting the freedom of choosing the phase of each precoding vector $\mathbf{w}^n_{\mathcal{S}_f,f}$ [110] and from equation (4.5), equation (4.23) can be further re-written as:

$$\mathbf{h}^{nH}_{\mathcal{S}_f,f}\mathbf{w}^n_{\mathcal{S}_f,f} \geq \sqrt{\mu^n_f}\zeta^n_f \tag{4.25}$$

$$\zeta^n_f \geq \sqrt{\sum_{i=1,i\neq f}^{F}\left|\mathbf{h}^{nH}_{\mathcal{S}_i,f}\mathbf{w}^n_{\mathcal{S}_i,i}\right|^2 + \sum_{m=1}^{M}\Gamma^n_m P^n_{B,m} g^n_{B,f} + N_o} \tag{4.26}$$

$$\text{Im}\left(\mathbf{h}^{nH}_{\mathcal{S}_f,f}\mathbf{w}^n_{\mathcal{S}_f,f}\right) = 0 \tag{4.27}$$

where ζ^n_f is an additionally introduced auxiliary variable. Again, the equivalence between equations (4.25)-(4.27) and equation (4.23) can be recognized from the fact that all the inequalities in (4.25)-(4.27) hold with equality at the optimum solution.

Moreover, equation (4.27) forces the imaginary part of $\mathbf{h}_{\mathcal{S}_f,f}^{nH}\mathbf{w}_{\mathcal{S}_f,f}^{n}$ to 0. This can be accomplished by rotating the phase of $\mathbf{w}_{\mathcal{S}_f,f}^{n}$ such that $\mathbf{h}_{\mathcal{S}_f,f}^{nH}\mathbf{w}_{\mathcal{S}_f,f}^{n}$ becomes real valued and positive.

By inspecting (4.25), we can see that it is not a convex constraint due to the term $g\left(\mu_f^n,\zeta_f^n\right)=\sqrt{\mu_f^n}\zeta_f^n$. One approach that was followed in [111] is to replace the non-convex term $g\left(\mu_f^n,\zeta_f^n\right)=\sqrt{\mu_f^n}\zeta_f^n$, in the z^{th} iteration, by its convex upper bound $\bar{g}\left(\mu_f^n,\zeta_f^n,\phi_f^{n(z-1)}\right)=\frac{\phi_f^{n(z-1)}}{2}\left(\zeta_f^n\right)^2+\frac{1}{2\phi_f^{n(z-1)}}\mu_f^n$, for given $\phi_f^{n(z-1)}$, and to iteratively solve the resulting problem by updating $\phi_f^{n(z-1)}$ until convergence. It is straightforward to show that $g\left(\mu_f^n,\zeta_f^n\right)$ and $\bar{g}\left(\mu_f^n,\zeta_f^n,\phi_f^{n(z-1)}\right)$ will satisfy the conditions in (4.18) if we set $\phi_f^{n(z-1)}=\sqrt{\mu_f^n}/\zeta_f^n$. It was shown in [89] that such SCA approach converges to a KKT point of (4.21).

Remark 4.5.1. *A lower complexity solution can be obtained by predefining the SINR variables μ_f^n for each SUE f on each allocated sub-channel n. One way of accomplishing this is by assuming equal SINR across the set of allocated sub-channels $\mathcal{N}_f$. Hence, we can set $\mu_f^n=\mu_f=2^{\left(R_f/\left(\Delta f|\mathcal{N}_f|\right)\right)}-1,\ \forall f\in\mathcal{F},n\in\mathcal{N}_f$. In this way, (4.25) becomes a convex constraint. This approach was followed in [112]. We shall refer to this scheme as "Equal SINR". Note that with either of the approaches, an admitted SUE will have the same performance. However, the two approaches have different impact on the network performance which will be assessed in Section 4.6.*

As for constraint C5 in (4.21), the step function $\Psi_{st}\left(\sum_{n\in\mathcal{N}_f}\left\|\mathbf{Q}_{s,f}\mathbf{w}_{\mathcal{S}_f,f}^{n}\right\|^2\right)$ can be approximated by using the following continuous concave approximation:

$$\Psi_{st}\left(\sum_{n\in\mathcal{N}_f}P_{s,f}^{n}\right)\approx\Xi\left(\mathbf{P}_{s,f}\right)=\frac{\sum_{n\in\mathcal{N}_f}P_{s,f}^{n}}{\sum_{n\in\mathcal{N}_f}P_{s,f}^{n}+\delta}\tag{4.28}$$

where, for ease of presentation, we have let $P_{s,f}^{n}=\left\|\mathbf{Q}_{s,f}\mathbf{w}_{\mathcal{S}_f,f}^{n}\right\|^2$, $\mathbf{P}_{s,f}=$

$\left[P_{s,f}^1, ..., P_{s,f}^n, ..., P_{s,f}^N\right]^T$, and δ is an arbitrary small number that is $\ll 1$. Moreover, by using the 1st order Taylor expansion, the expression in (4.28) can be further approximated around some point of interest $\mathbf{P}_{s,f}^{(z-1)}$ as[4]:

$$\Theta\left(\mathbf{P}_{s,f}, \mathbf{P}_{s,f}^{(z-1)}\right) = \Xi\left(\mathbf{P}_{s,f}^{(z-1)}\right) + \nabla\Xi\left(\mathbf{P}_{s,f}^{(z-1)}\right)\left(\mathbf{P}_{s,f} - \mathbf{P}_{s,f}^{(z-1)}\right) \tag{4.29}$$

Hence, around some point $\mathbf{P}_{s,f}^{(z-1)}$, constraint C5 in (4.21) can be re-written as:

$$\sum_{f=1}^{F} \Theta\left(\mathbf{P}_{s,f}, \mathbf{P}_{s,f}^{(z-1)}\right) \leq C_{s,max}, \ \forall s \in \mathcal{S} \tag{4.30}$$

Thus, given sub-channel allocation, the overall power allocation and admission

[4]Actually, $\mathbf{P}_{s,f}^{(z-1)}$ will be the optimal solution resulting from solving the joint power allocation and admission control problem in the $(z-1)^{th}$ iteration.

control problem in (4.21) in the z^{th} iteration, can be reformulated as follows:

$$\min_{\{\mathbf{w}^n_{\mathcal{S}_f,f},y_f,\mu^n_f,\zeta^n_f\}} D\sum_{f=1}^{F} y_f^2 + \sum_{f=1}^{F}\sum_{n\in\mathcal{N}_f}\left\|\mathbf{w}^n_{\mathcal{S}_f,f}\right\|^2$$

subject to

$$\text{C1}: \mathbf{h}^{nH}_{\mathcal{S}_f,f}\mathbf{w}^n_{\mathcal{S}_f,f} + y_f \geq \frac{\phi_f^{n(z-1)}}{2}\left(\zeta^n_f\right)^2 + \frac{1}{2\phi_f^{n(z-1)}}\mu^n_f,\ \forall f\in\mathcal{F}, n\in\mathcal{N}_f$$

$$\text{C2}: \zeta^n_f \geq \sqrt{\sum_{i=1,i\neq f}^{F}\left|\mathbf{h}^{nH}_{\mathcal{S}_i,f}\mathbf{w}^n_{\mathcal{S}_i,i}\right|^2 + \sum_{m=1}^{M}\Gamma^n_m P^n_{B,m} g^n_{B,f} + N_o},\ \forall f\in\mathcal{F}, n\in\mathcal{N}_f$$

$$\text{C3}: \text{Im}\left(\mathbf{h}^{nH}_{\mathcal{S}_f,f}\mathbf{w}^n_{\mathcal{S}_f,f}\right) = 0,\ \forall f\in\mathcal{F}, n\in\mathcal{N}_f$$

$$\text{C4}: \prod_{n\in\mathcal{N}_f}\left(1+\mu^n_f\right) \geq 2^{R_f/\Delta f},\ \forall f\in\mathcal{F}$$

$$\text{C5}: \sum_{f=1}^{F}\sum_{n\in\mathcal{N}_f}\left\|\mathbf{Q}_{s,f}\mathbf{w}^n_{\mathcal{S}_f,f}\right\|^2 \leq P_{s,max},\ \forall s\in\mathcal{S}$$

$$\text{C6}: \sum_{f=1}^{F}\Theta\left(\mathbf{P}_{s,f},\mathbf{P}^{(z-1)}_{s,f}\right) \leq C_{s,max},\ \forall s\in\mathcal{S}$$

$$\text{C7}: \Gamma^n_m\left(\sum_{f=1}^{F}\left|\mathbf{h}^{nH}_{\mathcal{S}_f,m}\mathbf{w}^n_{\mathcal{S}_f,f}\right|^2\right) \leq \Gamma^n_m I^n_m,\ \forall n\in\mathcal{N}$$

$$\text{C8}: \mathbf{w}^n_{\mathcal{S}_f,f}\in\mathbb{C}^{\Omega|\mathcal{S}_f|\times 1}, \mu^n_f \geq 2^{R_f/\Delta f}-1, \zeta^n_f\geq 0,\ \forall f\in\mathcal{F}, n\in\mathcal{N} \tag{4.31}$$

In (4.31), the objective is to maximize the number of admitted SUEs (the admission control part) while simultaneously minimizing the total downlink transmit power. This is achieved through the incorporation of the admission control variable y_f that takes the value of 0 only if SUE f is admitted. By setting as many y_fs to 0 as possible, the number of admitted SUEs is maximized. In the objective function, D is a large positive constant that gives a higher weight to the admission control part. One benefit of the formulation in (4.31) is that it is always feasible. To see this, one

can choose large enough y_fs to satisfy the constraint C1. It is worth mentioning that if (4.21) is feasible, then in (4.31), $y_f = 0,\ \forall f \in \mathcal{F}$.

Remark 4.5.2. *Note that the incorporation of the admission control variable y_f in (4.31) has been inspired by the work previously done in [110]. However, the work in [110] considered single-channel single-tier systems. Moreover, the work in [110] only considered simple SINR and power budget constraints.*

The problem in (4.31) is a convex optimization problem. In fact, it can be cast as an SOCP [9]. Thus, it can be efficiently solved by numerical solvers such as CVX [113]. To this end, we present **Algorithm** 12 that iteratively solves (4.31) by updating the points of interest $P_{s,f}^{n,(z-1)}$ and $\phi_f^{n,(z-1)}$ until convergence. Upon convergence, if $y_f = 0,\ \forall f \in \mathcal{F}$, then we terminate. Otherwise, find $f^* = \arg\max_f y_f$ such that $y_{f^*} \neq 0$, drop f^*, and rerun **Algorithm** 12. In this way, **Algorithm** 12 will be run at most $\mathcal{O}(\mathcal{F})$ times.

Proposition 4.5.1. *In **Algorithm** 12, the iterates produce a sequence of non-increasing objective function values that is guaranteed to converge. Moreover, convergence takes place to a solution that satisfies the KKT conditions of (4.21).*

Proof. Proposition 4.5.1 can be proved by using arguments similar to those discussed in the proof of *Proposition* 4.4.4. Hence, we omit the proof for brevity. □

Algorithm 12 Successive convex approximation algorithm for problem (4.21)

1: Initialize $P_{s,f}^{n,(0)}$ and $\phi_f^{n,(0)}$, $\forall s \in \mathcal{S},\ f \in \mathcal{F},\ n \in \mathcal{N}_f$ and set $z = 1$
2: **repeat**
3: Solve problem (4.31) to obtain $\mathbf{w}_{\mathcal{S}_f,f}^n$, μ_f^n, and ζ_f^n
4: Update $P_{s,f}^{n,(z)} = \left\| \mathbf{Q}_{s,f} \mathbf{w}_{\mathcal{S}_f,f}^n \right\|^2$ and $\phi_f^{n,(z)} = \sqrt{\mu_f^n}/\zeta_f^n$, $\forall s \in \mathcal{S},\ f \in \mathcal{F},\ n \in \mathcal{N}_f$
5: Increment z
6: **until** Convergence

4.6 Numerical Results and Discussion

4.6.1 Parameters

We evaluate the system performance through extensive simulations under various topologies and scenarios. We consider a two-tier network as shown in Fig. 4.1 with 5 small cells deployed at $(0,0)$, $(3R_{radius}/2, d)$, $(3R_{radius}, 0)$, $(0, -2d)$ and $(3R_{radius}/2, -d)$, where $d = 250\ m$ and $R_{radius} = 2d/\sqrt{3}$, that are within the coverage of a macrocell. The macrocell is deployed at $(2R_{radius}, -2d)$. All small cells are equipped with $\Omega = 4$ antennas. All UEs exist outdoor. We have 2 SUEs/small cell that are randomly deployed inside each small cell such that the distance between them and their closest small cell is either d or $d/2$. Channel gains account for both path-loss and Rayleigh fading. Path loss models from [47] are used. The path-loss between an RRH i and UE j is modeled as $l_{i,j} = 140.7 + 36.7 \log_{10}(d_{i,j})$, where $d_{i,j}$ is the distance (in km) between RRH i and UE j. On the other hand, the path loss between macrocell B and UE j is modeled as $l_{B,j} = 128.1 + 37.6 \log_{10}(d_{B,j})$. In addition, Rayleigh fading is modeled using independent, zero mean, and unit variance complex Gaussian random variables.

Unless otherwise stated, we have $N_o = 10^{-13}$ W, $D = 10^5$, $\delta = 10^{-5}$, $P_{B,max} = 20$ W, $P_{s,max} = 10$ W, $I_{max} = 10^5 N_o$, $I_{min} = N_o$, $R_m = 5$ bps/Hz, $q_f = 4$ sub-channels, $M = 3$ MUEs, and $|\mathcal{S}_f| = 3,\ \forall f \in \mathcal{F}$, meaning that each SUE f picks up the three closest small cells as its serving cluster. All data rate requirements are mentioned in terms of spectral efficiency (bps/Hz). Also, ideal fronthaul links ($C_{s,max} = \infty$) are assumed in all numerical results. However, we will also assess the effect of varying the available fronthaul capacity (i.e., non-ideal backhaul) on the performance of SUEs.

4.6.2 Numerical Results

Macrocell RA Problem With Single MUE

In Fig.4.2, we show the sum of the tolerable interference levels (on the allocated sub-channels only, i.e., those sub-channels with non-zero power) vs. the data rate requirement R_m. We have $N = 10$ sub-channels with $\mathbf{g}_{B,m} = \{g^n_{B,m} : n = 1, 2, ..., N\} = \{5, 4.56, 4.11, 3.67, 3.22, 2.78, 2.33, 1.89, 1.44, 1\}$. Channel gains and the sum of tolerable interference levels are normalized with respect to N_o. Results are obtained by using both our proposed **Algorithm** 8 and "*fmincon*" function in Matlab [48]. When calling this function, we choose "*sqp*" as its solving algorithm. Since "*fmincon*" is not guaranteed to give the optimal solution, owing to the non-convexity of the optimization problem (4.7), it is run a large number of times using different initial points and we keep track of the best feasible solution. In the figure, we also show the number of allocated sub-channels N_m to MUE m for each range of data rates. It is clear from the figure that as R_m increases, the number of allocated sub-channels increases. In addition, for the same number of allocated sub-channels, as R_m increases, the sum of tolerable interference levels decreases. This is justifiable since the interference levels that MUE m can tolerate decrease as R_m increases. Finally, Fig. 4.2 shows that **Algorithm** 8 gives the optimal solution as was discussed before.

Convergence behavior of Algorithm 10

Fig. 4.3 shows the convergence behavior of **Algorithm** 10. We have $N = 10$ sub-channels, and $R_m = 7$ bps/Hz. It is obvious from Fig. 4.3 that **Algorithm** 10 produces a sequence of non-decreasing objective function values and it converges within a reasonable number of iterations.

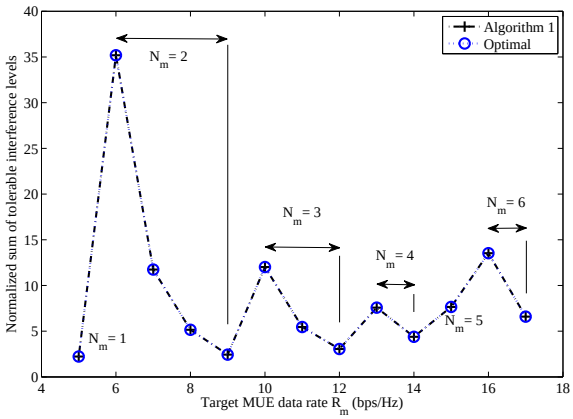

Figure 4.2: Normalized sum of tolerable interference levels on the allocated sub-channels vs. the data rate requirement R_m.

Performance of SUEs under proposed and traditional macrocell RA

Figs. 4.4 and 4.5 show the performance of SUEs (number of admitted SUEs and total downlink transmit power of small cells), when the macrocell performs resource allocation in order to "maximize the sum of tolerable interference" as proposed in this Chapter in Section 4.4 (labeled as "Proposed" in the figures). For the sake of comparison, we show the performance of SUEs when the macrocell performs resource allocation according to the traditional objective of "minimizing the total sum-power" [76] (labeled as "Traditional" in the figures) as was discussed in [8]. For a given macrocell resource allocation result, the performance of SUEs in Figs. 4.4 and 4.5 is obtained by averaging over 50 realizations, where in each realization, the positions of the SUEs are varied. We have $N = 10$ sub-channels available. The observations from Figs. 4.4 and 4.5 can be summarized as follows:

- The average percentage of admitted SUEs is higher when the macrocell performs RA according to the proposed method. This can be explained as follows. When

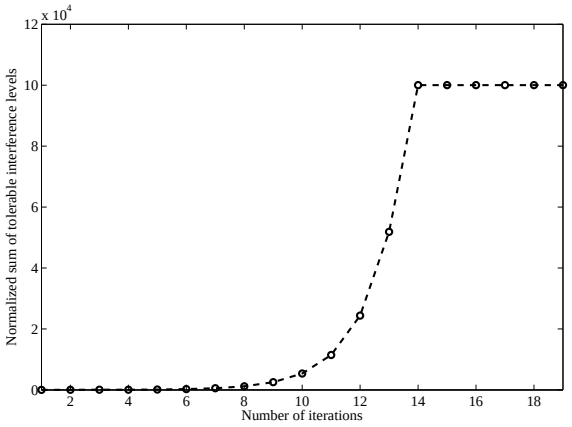

Figure 4.3: Convergence behavior of **Algorithm** 10.

the macrocell performs RA according to the proposed method, the macrocell is freeing as much sub-channels as possible for the small cells to use. This is not the case when the macrocell performs RA according to the traditional "minimize the total sum-power" method.

- For the proposed method for macrocell RA, the average percentage of admitted SUEs decreases, whereas the total downlink transmit power increases, when the data rate requirements for the MUEs R_m increases. At low data rate requirements of MUEs, small number of sub-channels are allocated to the MUEs. Hence, small cells have enough sub-channels to allocate to the SUEs that are free from cross-tier interference from the macrocell. In addition, there are no interference constraints imposed on those free sub-channels. As the data rate requirements of MUEs increase, the number of allocated sub-channels to the MUEs increases, as was shown in Fig. 4.2. Hence, small cells are now obliged to allocate sub-channels to the SUEs that experience cross-tier interference from the macrocell. This decreases the average percentage of admitted SUEs and increases the total downlink transmit power for small cells.

- For the traditional macrocell RA method with the objective of "minimizing the total sum-power", the performance of the SUEs is not very sensitive to the data rate requirement R_m of MUEs. This can be attributed to the fact that this RA method tends to allocate all the available sub-channels to the MUEs. Hence, irrespective of R_m, SUEs experience cross-tier interference from the macrocell across all available sub-channels.

- As the data rate requirement R_f of SUEs increases, the average percentage of admitted SUEs decreases, whereas, the total downlink transmit power for small cells increases.

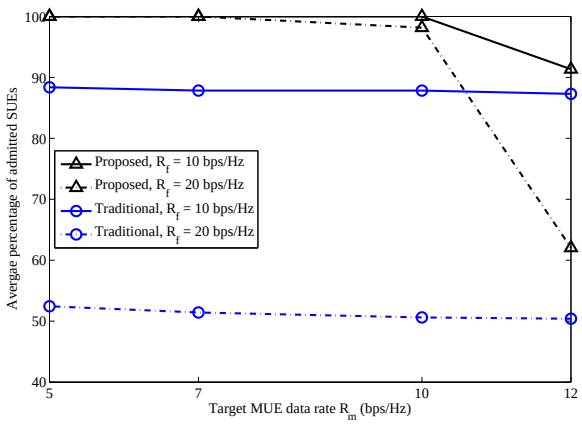

Figure 4.4: Average percentage of admitted SUEs vs. the MUEs data rate requirement R_m for different SUEs data rate requirement R_f.

Performance of SUEs under variable and equal SINR

In Fig. 4.6, we study the percentage of admitted SUEs and the total downlink transmit power vs. the target SUE data rate requirement R_f under variable and equal achieved SINR on the allocated sub-channels. In all the remaining results, we

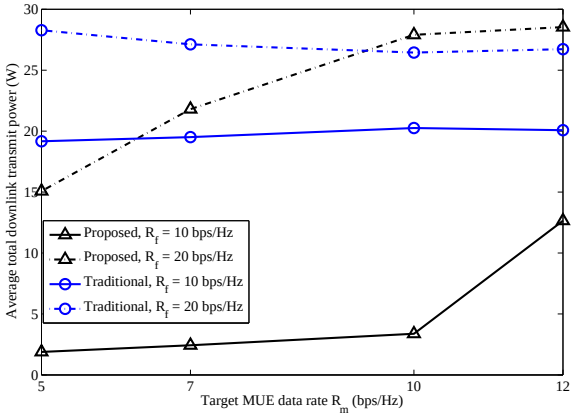

Figure 4.5: Average total downlink transmit power vs. the MUEs data rate requirement R_m for different SUEs data rate requirement R_f.

study the network performance for a snapshot of the setup described in Section 4.6.1. Resource allocation for the MUEs is performed according to (4.6) and we have $N = 5$ sub-channels available.

Note that an admitted SUE, in either cases, will have the same data rate requirement R_f achieved. However, as shown in Fig. 4.6, the impact on the C-RAN network performance is different. Better network performance (higher number of admitted SUEs) can be achieved by adapting the achieved SINR on each allocated sub-channel according to its channel gain and the level of interference experienced on that allocated sub-channel.

Performance of SUEs under different available fronthaul capacity

Figs. 4.7 and 4.8 study the percentage of admitted SUEs and the total downlink transmit power vs. the target SUE data rate requirement R_f under different available fronthaul capacity $C_{s,max}$ at each small cell.

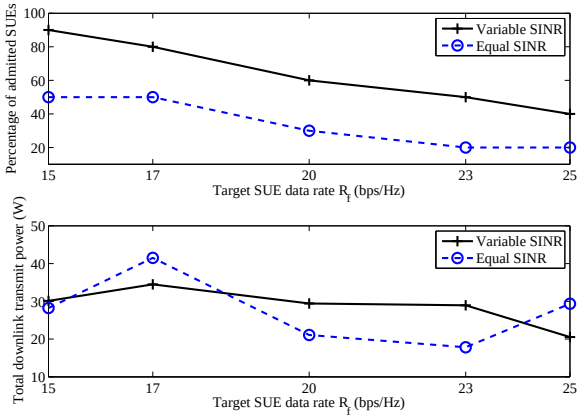

Figure 4.6: Percentage of admitted SUEs and the total downlink transmit power vs. the target SUE data rate requirement R_f under variable and equal achieved SINR.

Each small cell can serve a higher number of SUEs when more fronthaul capacity becomes available. This translates to a higher number of admitted SUEs as shown in Fig. 4.7. In addition, from Fig. 4.8, it is obvious that when the same number of SUEs is admitted, less total downlink transmit power is required when the available fronthaul capacity at each small cell increases.

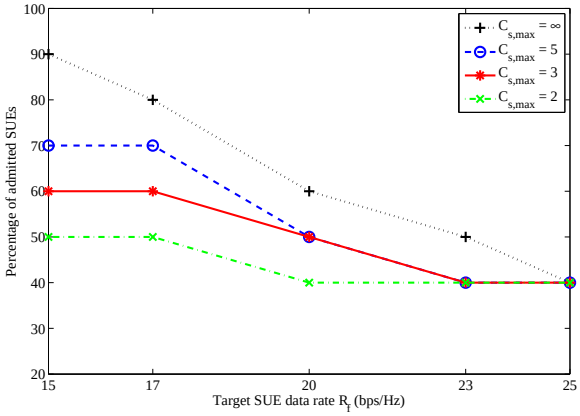

Figure 4.7: Percentage of admitted SUEs vs. the target SUE data rate requirement R_f under different available fronthaul capacity $C_{s,max}$.

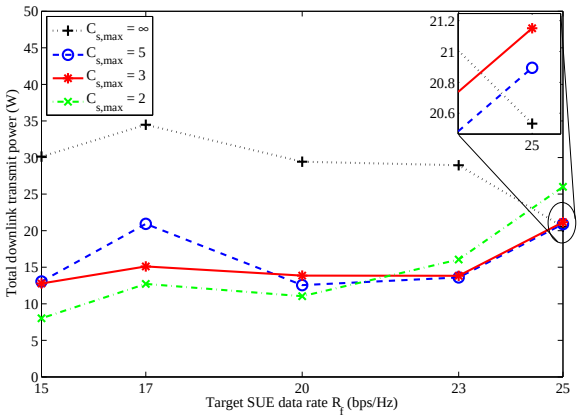

Figure 4.8: Total downlink transmit power vs. the target SUE data rate requirement R_f under different available fronthaul capacity $C_{s,max}$.

Performance of SUEs under different number of allocated sub-channels q_f

Fig. 4.9 studies the percentage of admitted SUEs and the total downlink transmit power vs. the data rate requirement R_f of SUEs for different number of allocated sub-channels q_f.

It is clear from the figure that, the C-RAN can support a higher number of SUEs when they are allocated a larger number of sub-channels. This can be attributed to the fact that allocating more sub-channels to the SUEs provides additional degrees of freedom for the C-RAN to exploit. Therefore, a better network performance is achieved.

Convergence behavior of Algorithm 12

Finally, Fig. 4.10 shows the convergence behavior of **Algorithm** 12 for a feasible scenario of the optimization problem (4.21), where we have $R_f = 10$ bps/Hz.

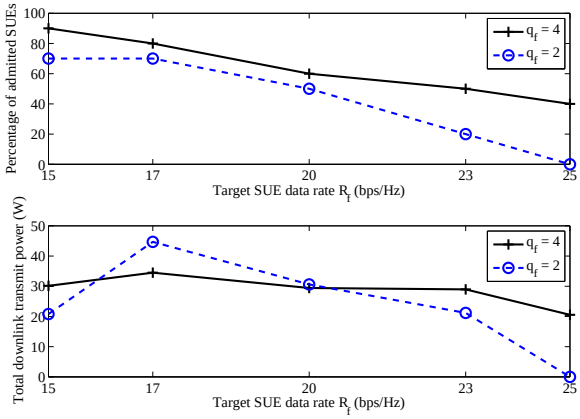

Figure 4.9: Percentage of admitted SUEs and the total downlink transmit power vs. the target SUE data rate requirement R_f under different number of allocated sub-channels q_f.

The upper figure shows the admission control part of the objective function in (4.31), i.e., $\sum_{f=1}^{F} y_f^2$, whereas the lower figure shows the total downlink transmit power part of the objective function. Since the considered scenario is a feasible one, the admission control part converges to 0, i.e., all SUEs are admitted. Also, it is obvious that **Algorithm** 12 produces a sequence of non-increasing objective function values and it converges in a few number of iterations.

4.7 Conclusion

I have proposed a complete framework for the resource allocation and admission control problem in a two-tier OFDMA cellular network that is composed of a macrocell which is overlaid with cloud-RAN of small cells. Different from the traditional objective functions, a new resource allocation problem with new objective function has been proposed for the macrocell where, the macrocell aimed at allocating resources to its MUEs in a way that can tolerate the maximum possible interference from the

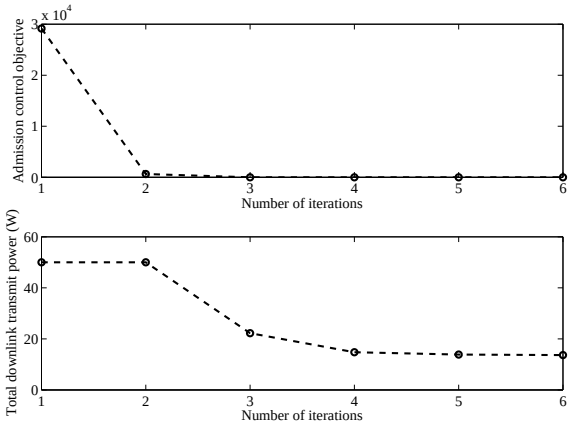

Figure 4.10: Convergence behavior of **Algorithm** 12.

small cell tier. Now, given the interference constraints for the MUEs, the small cells perform resource allocation and admission control with the objective of minimizing the total downlink transmit power subject to QoS and fronthaul constraints. Both problems have been shown to be MINLP and lower complexity algorithms have been proposed for both problems based on the framework of successive convex approximation. Numerical results have confirmed the significance of the new objective for the macrocell resource allocation problem and its positive impact on the performance of SUEs. Moreover, I have investigated the effect of different parameters of the resource allocation problem for small cells on the performance of SUEs.

Chapter 5

Conclusion and Future Work

Global mobile traffic will boom in the years to come, thanks to the increasing popularity of smart mobile devices and the introduction of affordable data plans by cellular operators. In addition, data hungry mobile applications, such as audio and video streaming, social sharing, or cloudbased services, are more and more popular among users. On the other hand, traditional cellular networks are reaching their breaking points, and the conventional cellular grid structure that has been devised to cater to large coverage areas and optimized for homogeneous traffic is facing unprecedented challenges. Hence, base station densification, through the deployment of additional low power nodes within the coverage area of traditional cellular networks, is foreseen to be a viable solution that can significantly boost the overall spectral efficiency and energy efficiency through a full spatial resource reuse.

However, dense deployment of small cells and low power nodes on top of the traditional cellular network poses several challenges that must be addressed through efficient resource allocation schemes. In this thesis, I have developed several resource allocation schemes that can handle co-tier, as well as cross-tier, interference in an efficient manner. Moreover, several practical aspects have been taken into consideration

in the resource allocation schemes. Numerical results have shown that the proposed schemes can provide promising spectrum utilization, enhanced network throughput, and QoS guarantee for UEs under controlled interference with low computational complexity.

I summarize the contributions of this thesis in Section 5.1 and then point out the future research directions in Section 5.2.

5.1 Contributions

Motivated by the exponential growth of mobile traffic demand in wireless networks, I have studied RA in multi-tier HetNets with small cell deployments. Specifically, in Chapter 2, I have investigated the idea of clustering (coordination) of small cells into cooperative groups and its effect on the small cells performance. Clustering has been used as a technique to mitigate co-tier interference and to divide the centralized small cells RA problem into smaller sub-problems. Hence, reducing overall complexity and promoting the use of semi-distributed RA solutions. The main contributions of Chapter 2 can be summarized as follows:

- I have proposed a framework for clustering, sub-channel and power allocation in a two-tier macrocell small cell network. This framework was implemented in a hierarchical fashion where the responsibilities were shared between the FGW and the small cells.

- To avoid the exponential complexity associated with obtaining the optimal cluster, I have formulated the clustering sub-problem as a correlation clustering problem which was solved by using an SDP-based algorithm. In addition, I offered a complexity analysis for the proposed SDP algorithm.

- I have proposed another algorithm to reduce the computation burden and to eliminate the necessity of going through the whole penalty term range for obtaining the optimal cluster configuration.

- I have proposed a heuristic low complexity power and sub-channel allocation algorithm to be executed by the **CH** in each cluster.

Moreover, in Chapter 3, I have investigated the impact of RA decisions of one tier on the other one by proposing RA algorithms that are tier-aware. In addition, owing to the expected high density of small cells deployments, distributed RA solutions have been proposed. The main contributions of Chapter 3 can be summarized as follows:

- I have developed a complete framework for tier-aware resource allocation in an OFDMA-based two-tier macrocell-small cell network with new objectives, coping with the new requirements of multi-tier HetNets.

- I have formulated a tier-aware RA problem for the macrocell and showed that it was an MINLP.

- I have proved that the macrocell could solve another alternate optimization problem that would yield the optimal solution for the MINLP with polynomial time complexity.

- I have shown that proposed macrocell RA problem outperformed the traditional "minimize the total sum-power" problem in terms of its impact on the small cells tier performance.

- For the small cell tier, I have formulated a joint resource allocation and admission control problem that aimed at maximizing the number of admitted SUEs

with minimum bandwidth consumption to accommodate additional tiers, and showed that it was an MINLP.

- I have offered an upper bound solution to the MINLP problem through convex relaxation and proposed a solution to the convex relaxation that was implemented in a distributed fashion using dual decomposition.

Finally, in Chapter 4, I have studied C-RAN as an architecture for deploying small cells that has the potential to face a lot challenges that face BS densification. The main contributions of Chapter 4 can be summarized as follows:

- I have developed a complete framework for downlink radio resoure allocation in an OFDMA-based two-tier cellular network where a macrocell is underlaid with a C-RAN of small cells.

- I have formulated a resource allocation problem for the macrocell that is aware of the existence of the small cell tier and have shown that it was an MINLP. In addition, I relaxed the simplifying assumptions made in [8].

- I have investigated the single MUE macrocell problem and have shown how to obtain its optimal solution, with polynomial time complexity, despite its non-convexity.

- Based on the observations that I got from solving the single MUE macrocell problem, I have proposed a low complexity solution for the multi-MUE macrocell RA problem that relies on the framework of SCA.

- I have formulated an RA problem for the C-RAN of small cells whose objective was to minimize the total downlink transmit power subject to QoS constraints

for SUEs, MUEs' interference thresholds, small cells power budget, and fronthaul capacity constraints as an optimization problem. I have shown that it was an MINLP.

- I have offered a low complexity solution based upon the SCA approach. Moreover, I have incorporated AC, jointly with RA, to deal with infeasibility issues in the C-RAN RA problem in case it was not possible to support all SUEs with their QoS requirements.

5.2 Future Work

In this section, I outline some of the possible future research directions in the context of small cells.

5.2.1 Robust Resource Allocation

In all the proposed resource allocation schemes, perfect information was assumed to be available at all decision making entities. However, given the random, and sometimes erratic, nature of many wireless channels, the actual performance of a wireless network can be severely influenced. Hence, it is desirable and critical to have RA solutions that can withstand uncertainty and are robust to inevitable errors in the obtained information.

5.2.2 Cell Association Schemes for Multi-tier Cellular Networks

In the presented work, it has always been assumed that UEs are already associated with their respective cells, and that this association remains fixed throughout the RA operation. To offer a complete and practical RA framework, it is interesting,

however, to incorporate cell association as a preceding step to the RA operation. Several association schemes already exist in the literature that are simple and analytically tractable. However, it is foreseen that they will be unable to guarantee the optimum performance in multi-tier HetNets unless critical parameters, such as bias values, transmit power of the users in the uplink and BSs in the downlink, resource partitioning, etc. are optimized.

5.2.3 Other Objective Functions and Constraints

In Chapters 3 and 4, it was shown that the new proposed objective function (sum of tolerable interference levels) for the macrocell had a positive impact on the small cells' performance, when compared to the traditional objective functions. This motivates us to look for other objective functions, specially when additional network tiers exist. Moreover, for the C-RAN of small cells' optimization problem in Chapter 4, additional practical constraints might be incorporated in the C-RAN RA problem such as power consumption in the cloud as well as cloud processing capability constraints.

Bibliography

[1] Ericsson, "5G Radio Access, Research and Vision," *white paper*, 2013.

[2] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Communications*, vol. 18, pp. 22–30, June 2011.

[3] A. N. T. Zahir, K. Arshad and K. Moessner, "Interference management in femtocells," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 293–311, 2013.

[4] M. Bennis, M. Simsek, A. Czylwik, W. Saad, S. Valentin, and M. Debbah, "When cellular meets WiFi in wireless small cell networks," *IEEE Communications Magazine*, vol. 51, pp. 44–50, June 2013.

[5] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, pp. 1380–1408, December 2010.

[6] S. Sun, Q. Gao, Y. Peng, Y. Wang, and L. Song, "Interference management through CoMP in 3GPP LTE-advanced networks," *IEEE Wireless Communications*, vol. 20, pp. 59–66, February 2013.

[7] J. Robson, "Small cell backhaul requirements," *NGMN Alliance, White Paper*, 2012.

[8] A. Abdelnasser, E. Hossain, and D. Kim, "Tier-aware resource allocation in OFDMA macrocell-small cell networks," *IEEE Transactions on Communications*, vol. 63, pp. 695–710, March 2015.

[9] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 13, pp. 193 – 228, 1998.

[10] X. Li, L. Qian, and D. Kataria, "Downlink power control in co-channel macrocell femtocell overlay," in *43rd Annual Conference on Information Sciences and Systems (CISS), 2009.*, pp. 383–388, 2009.

[11] G. Cao, D. Yang, X. Ye, and X. Zhang, "A downlink joint power control and resource allocation scheme for co-channel macrocell-femtocell networks," in *IEEE Wireless Communications and Networking Conference (WCNC), 2011*, pp. 281–286, 2011.

[12] F. Cao and Z. Fan, "Power loading and resource allocation for femtocells," in *IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011*, pp. 1–5, 2011.

[13] J. Kim and D.-H. Cho, "A joint power and subchannel allocation scheme maximizing system capacity in indoor dense mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 59, pp. 4340 –4353, nov. 2010.

[14] J. Zhang, Z. Zhang, K. Wu, and A. Huang, "Optimal distributed subchannel, rate and power allocation algorithm in ofdm-based two-tier femtocell networks," in *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring), 2010*, pp. 1–5, 2010.

[15] H. Zhang, X. Chu, W. Zheng, and X. Wen, "Interference-aware resource allocation in co-channel deployment of ofdma femtocells," in *IEEE International Conference on Communications (ICC), 2012*, pp. 4663–4667, 2012.

[16] W. Li, W. Zheng, W. Xiangming, and T. Su, "Dynamic clustering based subband allocation in dense femtocell environments," in *IEEE 75th Vehicular Technology Conference (VTC Spring), 2012*, pp. 1 –5, may 2012.

[17] F. Pantisano, M. Bennis, W. Saad, M. Debbah, and M. Latva-aho, "Interference alignment for cooperative femtocell networks: A game-theoretic approach," *IEEE Transactions on Mobile Computing,*, vol. PP, no. 99, p. 1, 2012.

[18] G. Ning, Q. Yang, S. K. Kwak, and L. Hanzo, "Macro- and femtocell interference mitigation in ofdma wireless systems," in *IEEE Global Communications Conference (GLOBECOM), 2012*, December 2012.

[19] A. Hatoum, R. Langar, N. Ait Saadi, R. Boutaba, and G. Pujolle, "QoS-based Power Control and Resource Allocation in OFDMA Femtocell Networks," in *IEEE GLOBECOM*, pp. 1–7, dec 2012.

[20] K. Hosseini, H. Dahrouj, and R. Adve, "Distributed clustering and interference management in two-tier networks," in *IEEE Global Telecommunications Conference (GLOBECOM), 2012*, pp. 4483–4488, Dec. 2012.

[21] A. Abdelnasser and E. Hossain, "Joint subchannel and power allocation in Two-Tier OFDMA HetNets with clustered femtocells," in *IEEE ICC 2013 - Wireless Networking Symposium (ICC'13 WN)*, (Budapest, Hungary), June 2013.

[22] A. Abdelnasser and E. Hossain, "Subchannel and power allocation schemes for clustered femtocells in Two-Tier OFDMA HetNets," in *IEEE International Conference on Communications 2013: IEEE ICC'13 - The IEEE ICC 2013 2nd International Workshop on Small Cell Wireless Networks (SmallNets) (ICC'13 - IEEE ICC'13 - Workshop SmallNets)*, (Budapest, Hungary), June 2013.

[23] S.-E. Wei, C.-H. Chang, Y.-E. Lin, H.-Y. Hsieh, and H.-J. Su, "Formulating and solving the femtocell deployment problem in two-tier heterogeneous networks," in *IEEE International Conference on Communications (ICC), 2012*, pp. 5053 –5058, June, 2012.

[24] H. Li, X. Xu, D. Hu, X. Tao, P. Zhang, S. Ci, and H. Tang, "Clustering strategy based on graph method and power control for frequency resource management in femtocell and macrocell overlaid system," *Journal of Communications and Networks*, vol. 13, no. 6, pp. 664–677, 2011.

[25] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 46–55, 2002.

[26] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; (Release 11)," TS 36.300, 3rd Generation Partnership Project (3GPP), Sep 2013.

[27] M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,*. Dover Publications, Incorporated, 1974.

[28] D. P. Williamson and D. B. Shmoys, *The Design of Approximation Algorithms*. New York, NY, USA: Cambridge University Press, 1st ed., 2011.

[29] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta."

[30] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control* (V. Blondel, S. Boyd, and H. Kimura, eds.), Lecture Notes in Control and Information Sciences, pp. 95–110, Springer-Verlag Limited, 2008.

[31] V. V. Vazirani, *Approximation algorithms*. New York, NY, USA: Springer-Verlag New York, Inc., 2001.

[32] U. Feige and G. Schechtman, "On the optimality of the random hyperplane rounding technique for max cut," tech. rep., Algorithms, 2000.

[33] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[34] I. Pólik and T. Terlaky, "Interior point methods for nonlinear optimization," in *Nonlinear Optimization*, pp. 215–276, Springer, 2010.

[35] J. F. Sturm, "Implementation of interior point methods for mixed semidefinite and second order cone optimization problems," *Optimization Methods and Software*, vol. 17, no. 6, pp. 1105–1154, 2002.

[36] B. Gärtner and J. Matoušek, *Approximation algorithms and semidefinite programming.* Berlin: Springer, 2012.

[37] A. M. Geoffrion and R. Nauss, "Exceptional paper–parametric and postoptimality analysis in integer linear programming," *Management Science*, vol. 23, no. 5, pp. 453–466, 1977.

[38] L. Jenkins, "Parametric methods in integer linear programming," *Annals of Operations Research*, vol. 27, no. 1, pp. 77–96, 1990.

[39] L. Jenkins, "Parametric mixed integer programming: An application to solid waste management," *Management Science*, vol. 28, no. 11, pp. pp. 1270–1284, 1982.

[40] K. Son, S. Lee, Y. Yi, and S. Chong, "REFIM: A practical interference management in heterogeneous wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 1260 –1272, june 2011.

[41] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser ofdm systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, 2005.

[42] M. Tao, Y.-C. Liang, and F. Zhang, "Resource allocation for delay differentiated traffic in multiuser ofdm systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2190–2201, 2008.

[43] D. Bharadia, G. Bansal, P. Kaligineedi, and V. Bhargava, "Relay and power allocation schemes for ofdm-based cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 2812 –2817, september 2011.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[45] F. Wang, M. Krunz, and S. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 74 –87, feb. 2008.

[46] G. Scutari, D. Palomar, and S. Barbarossa, "Simultaneous iterative water-filling for gaussian frequency-selective interference channels," in *IEEE International Symposium on Information Theory, 2006*, pp. 600–604, 2006.

[47] 3GPP, "Further advancements for E-UTRA physical layer aspects (Release 9)," TR 36.814, 3rd Generation Partnership Project (3GPP), Mar 2010.

[48] MATLAB, *version 8.0.0.783 (R2012b)*. The MathWorks Inc., 2012.

[49] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Communications*, vol. 21, pp. 118–127, June 2014.

[50] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Communications Magazine*, vol. 47, pp. 41–48, September 2009.

[51] V. Nguyen Ha and L. Bao Le, "Fair resource allocation for OFDMA femtocell networks with macrocell protection," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 1388–1401, March 2014.

[52] S. Guruacharya, D. Niyato, D. I. Kim, and E. Hossain, "Hierarchical competition for downlink power allocation in OFDMA femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 1543–1553, April 2013.

[53] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 342–355, January 2014.

[54] A. Abdelnasser, E. Hossain, and D. I. Kim, "Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1628–1641, March 2014.

[55] M. Andersin, Z. Rosberg, and J. Zander, "Gradual removals in cellular pcs with constrained power control and noise," in *Proc. IEEE PIMRC, 1995*, vol. 1, pp. 56–60 vol.1, Sep 1995.

[56] D. Lopez-Perez, X. Chu, A. Vasilakos, and H. Claussen, "Power minimization based resource allocation for interference mitigation in OFDMA femtocell networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 333–344, February 2014.

[57] J. Tadrous, A. Sultan, and M. Nafie, "Admission and power control for spectrum sharing cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 1945–1955, 2011.

[58] I. Mitliagkas, N. Sidiropoulos, and A. Swami, "Joint power and admission control for ad-hoc and cognitive underlay networks: Convex approximation and distributed implementation," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4110–4121, 2011.

[59] L. B. Le and E. Hossain, "Resource allocation for spectrum underlay in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5306–5315, 2008.

[60] W. J. Shin, K. Y. Park, D. I. Kim, and J. W. Kwon, "Large-scale joint rate and power allocation algorithm combined with admission control in cognitive radio networks," *Journal of Communications and Networks*, vol. 11, no. 2, pp. 157–165, 2009.

[61] K. Phan, T. Le-Ngoc, S. Vorobyov, and C. Tellambura, "Power allocation in wireless multi-user relay networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2535–2545, 2009.

[62] X. Gong, S. Vorobyov, and C. Tellambura, "Joint bandwidth and power allocation with admission control in wireless multi-user networks with and without relaying," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1801–1813, 2011.

[63] S. E. Nai, T. Quek, M. Debbah, and A. Huang, "Slow admission and power control for small cell networks via distributed optimization," in *IEEE Wireless Communications and Networking Conference (WCNC), 2013*, pp. 2261–2265, 2013.

[64] S. Namal, K. Ghaboosi, M. Bennis, A. MacKenzie, and M. Latva-aho, "Joint admission control and interference avoidance in self-organized femtocells," in *Conference Record of the 44th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2010*, pp. 1067–1071, 2010.

[65] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser ofdm with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1747–1758, Oct 1999.

[66] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 538–549, April 2012.

[67] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond [accepted from open call]," *IEEE Communications Magazine*, vol. 51, pp. 154–160, July 2013.

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

[69] F. Cao and Z. Fan, "Power loading and resource allocation for femtocells," in *Proc. 2011 73rd IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2011.

[70] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP J. on Wireless Commun. and Networking*, vol. 2, pp. 222–235, 2004.

[71] W. L. Winston, *Operations Research: Applications and Algorithms*. Cengage Learning, 2003.

[72] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," in *Proc. 2000 IEEE Global TeleCommun. Conference (GLOBECOM)*, vol. 1, pp. 103–107 vol.1, 2000.

[73] J.-H. Noh and S.-J. Oh, "Distributed SC-FDMA resource allocation algorithm based on the hungarian method," in *Proc. 2009 70th IEEE Vehicular Technology Conference Fall (VTC 2009-Fall),*, pp. 1–5, Sept 2009.

[74] Y. Tachwali, B. Lo, I. Akyildiz, and R. Agusti, "Multiuser resource allocation optimization using bandwidth-power product in cognitive radio networks," *IEEE J. on Sel. Areas in Commun.*, vol. 31, pp. 451–463, March 2013.

[75] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization : Algorithms and Complexity*. Englewood Cliffs, NJ: Prentice-Hall Inc., July 1998.

[76] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *2006 IEEE International Symposium on Information Theory*, pp. 1394–1398, 2006.

[77] Y.-F. Liu and Y.-H. Dai, "On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems," *IEEE Trans. on Signal Processing*, vol. 62, pp. 583–596, Feb 2014.

[78] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, 2006.

[79] S. Shen and T. Lok, "Dynamic power allocation for downlink interference management in a two-tier OFDMA network," *IEEE Trans. on Vehicular Technology*, vol. 62, pp. 4120–4125, Oct 2013.

[80] H. Zhang, C. Jiang, N. Beaulieu, X. Chu, X. Wen, and M. Tao, "Resource allocation in spectrum-sharing OFDMA femtocells with heterogeneous services," *IEEE Trans. on Commun.*, vol. 62, pp. 2366–2377, July 2014.

[81] J. Currie and D. I. Wilson, "OPTI: Lowering the Barrier Between Open Source Optimizers and the Industrial MATLAB User," in *Foundations of Computer-Aided Process Operations* (N. Sahinidis and J. Pinto, eds.), (Savannah, Georgia, USA), 8–11 January 2012.

[82] X. Zhang, L. Chen, J. Huang, M. Chen, and Y. Zhao, "Distributed and optimal reduced primal-dual algorithm for uplink OFDM resource allocation," in *Proc. 2009 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC*, pp. 4814–4819, Dec 2009.

[83] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods," *Notes for EE364B, Stanford University*, 2007.

[84] S. Boyd and C. Barratt, "Ellipsoid method," *Notes for EE364B, Stanford University, 2008*, 2008.

[85] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 531–552, March 2011.

[86] C. Mobile, "C-RAN: The road towards green RAN," *white paper*, 2011.

[87] 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, pp. 405–426, Firstquarter 2015.

[88] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. pp. 681–683, 1978.

[89] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *Journal of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010.

[90] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 2809–2823, May 2014.

[91] S. Luo, R. Zhang, and T. Lim, "Downlink and uplink energy minimization through user association and beamforming in C-RAN," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 494–508, Jan 2015.

[92] J. Zhao, T. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2762–2775, June 2013.

[93] Y. Cai, F. Yu, and S. BU, "Dynamic operations of cloud radio access networks (C-RAN) for mobile cloud computing systems," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[94] J. Li, M. Peng, A. Cheng, and Y. Yuling, "Delay-aware cooperative multipoint transmission with backhaul limitation in Cloud-RAN," in *Proc. IEEE ICC, 2014*, pp. 665–670, June 2014.

[95] V. N. Ha and L. B. Le, "Joint coordinated beamforming and admission control for fronthaul constrained cloud-RANs," in *Proc. IEEE GLOBECOM, 2014*, pp. 4054–4059, Dec 2014.

[96] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[97] D. Liu, S. Han, C. Yang, and Q. Zhang, "Semi-dynamic user-specific clustering for downlink Cloud Radio Access Network," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[98] D. Ng and R. Schober, "Secure and green swipt in distributed antenna networks with limited backhaul capacity," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2015.

[99] L. Liang, S. Bi, and R. Zhang, "Joint power control and fronthaul rate allocation for throughput maximization in broadband cloud radio access network," *arXiv:1407.3855*.

[100] Y. Qi, M. Imran, A. Quddus, and R. Tafazolli, "Achievable rate optimization for coordinated multi-point transmission (CoMP) in cloud-based RAN architecture," in *Proc. IEEE ICC, 2014*, pp. 4753–4758, June 2014.

[101] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Transactions on Signal Processing,*, vol. 61, pp. 5646–5658, Nov 2013.

[102] X. Rao and V. Lau, "Distributed fronthaul compression and joint signal recovery in Cloud-Ran," *IEEE Transactions on Signal Processing*, vol. 63, pp. 1056–1065, Feb 2015.

[103] M.-L. Ku, L.-C. Wang, and Y.-L. Liu, "Joint antenna beamforming, multiuser scheduling, and power allocation for hierarchical cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 896–909, May 2015.

[104] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous net-

works," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 226–240, February 2013.

[105] A. Abdelnasser and E. Hossain, "Joint resource allocation and admission control in OFDMA-based multi-tier cellular networks," in *Proc. IEEE GLOBECOM, 2014*, pp. 4689–4694, Dec 2014.

[106] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing. Second Edition.* Cambridge Univ. Press, 1992.

[107] J. Papandriopoulos and J. Evans, "Scale: A low-complexity distributed protocol for spectrum balancing in multiuser dsl networks," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3711–3724, 2009.

[108] J. Li, S. Dey, and J. Evans, "Maximal lifetime power and rate allocation for wireless sensor systems with data distortion constraints," *IEEE Transactions on Signal Processing,*, vol. 56, pp. 2076–2090, May 2008.

[109] M. Chiang, C. W. Tan, D. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 2640–2651, July 2007.

[110] E. Matskani, N. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 2682–2693, July 2008.

[111] L. Tran, M. Hanif, A. Tolli, and M. Juntti, "Fast converging algorithm for weighted sum rate maximization in multicell MISO downlink," *IEEE Signal Processing Letters*, vol. 19, pp. 872–875, Dec 2012.

[112] A. Abdelnasser and E. Hossain, "Two-tier OFDMA cellular Cloud-RAN: Joint resource allocation and admission control," in *IEEE Globecom, 2015, accepted*, Dec. 2015.

[113] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1."