Abstract

Resource Allocation in Backhaul Constrained Small Cell Networks

Zhe Cui

Master of Applied Science

Graduate Department of Electrical and Computer Engineering

University of Toronto

2015

Small cell networks have the potential to significantly increase data rates in existing cellular networks by increasing frequency reuse due to reduced transmit powers. Contrary to their large cell counterparts, small cell access points (APs) are connected to the core network via a backhaul link with limited capacity.

In this thesis, we study the joint resource allocation problem of maximizing the weighted sum rate of users in small cell networks where APs are subject to finite backhaul capacity constraints.

We develop a low-complexity iterative algorithm for the case of single antenna (SISO) and multiple antenna (MISO) APs that converges to a locally optimal solution of this non-convex problem. The key innovation is the efficient use of bisection search to satisfy both power and backhaul constraints resulting in decreased complexity. Distributed and semi-distributed variants of the algorithm are provided that allow trade-off between signaling overhead and performance.
Dedication

To my parents Ping & Yunping

and

my brother Jack
Acknowledgements

The thesis could not have been completed without the support, advice, friendship and love for all the people in my life, I am eternally grateful to have such great company.

First, I’d like to thank my supervisor, Professor Raviraj Adve, for his guidance, encouragement and support. He has taught me to be patient when confronting challenges and to always persevere, thinking critically and objectively. His enthusiasm and boundless energy is both inspirational and contagious. This thesis would not have been possible without the tremendous influence that Prof. Adve gave me.

I would like to thank my many friends in the Communications Group for their support these past few years. A special thank you to Ehsan Karamad and Kianoush Hosseini for the helpful discussions and advice with regards to research and academic life. Thank you Masoud Barekatain, Peter Sam Raj, Gagan Goel and Jaya Prakash "JP" Champati for coffee breaks and foosball when I needed it the most. My lab mates: Arin Minasian, Angelia Tian, Max Yuan, Sanam Sadr, Mohammad-Shahin Mahanta and Amir Aghaei, thank you for making BA7114 a fun and productive work environment.

A big thank to my roommate and gym partner Farhan Khan for all the fun times, your friendship made my graduate school experience an enjoyable experience!

My gratitude to my best and oldest friend Bobbie Chang for his help and support since the schoolyard days. Thank you for helping me keep things in perspective and for all the laughs over the years.

To my brother Jack, thanks for all the joy and laughter you bring into my life and making me always look forward to coming home.

Most importantly, I’d like to thank my parents, my Mother Ping and my Father Yunping for their unwavering love and support. I am blessed and grateful for the opportunities that are given to me due to their sacrifices and dedication. Their love gives me the strength and motivation to strive towards my goals and become a better person.
## Contents

1 Introduction ............................. 1  
   1.1 Motivation and Overview ................. 3  
   1.2 Literature Survey ......................... 5  
   1.3 Thesis Contributions ...................... 6  
   1.4 Organization of Thesis .................... 8  
   1.5 Notation ................................ 8  

2 Backhaul Constrained Networks .............. 9  
   2.1 System Model .............................. 9  
   2.2 Problem Formulation ....................... 13  

3 Resource Allocation with Single Antenna Access Points ..................... 15  
   3.1 Lagrangian and KKT Conditions .......... 16  
   3.2 User Scheduling ........................... 19  
   3.3 Low-Complexity Optimization Algorithm .... 23  
   3.4 Distributed and Semi-Distributed Algorithms .... 26  
   3.5 Numerical Results and Discussion ........ 29  
   3.6 Summary ................................ 37  

4 Resource Allocation with Multiple Antenna Access Points .......... 40  
   4.1 Lagrangian and KKT Conditions .......... 41  
   4.2 Low-Complexity Algorithm ................ 42
List of Tables

3.1 Simulation Parameters for Single Antenna AP Network . . . . . . . . . . . . . 30
3.2 Algorithm Convergence Speed . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
4.1 Simulation Parameters for Multiple Antenna AP Network . . . . . . . . . . 48
List of Figures

2.1 Illustration of the system at hand. .................................................. 10

3.1 Weighted Sum Rate versus $B_m^{\text{max}}$. ....................................... 32

3.2 Frequency Reuse Factor versus $B_m^{\text{max}}$ for $SNR_{dB} = 3$ in Standard Deployment. 33

3.3 Percent Backhaul Achieved versus $B_m^{\text{max}}$. .............................. 33

3.4 Percent Backhaul and Power Used versus $B_m^{\text{max}}$ for $SNR_{dB} = 3$ in Dense Deployment. .............................. 34

3.5 Comparison of Percent Backhaul Achieved versus $B_m^{\text{max}}$ for $SNR_{dB} = 3$ in Standard Deployment. .............................. 35

3.6 Cumulative User Rate versus Time Slot for $SNR_{dB} = 3$ with Proportional Fair Scheduling ($T = 30$). .............................. 36

3.7 Empirical CDF of User Rates for $SNR_{dB} = 3$. ................. 37

3.8 Percent of Exact Pricing Objective Achieved in Standard Deployment for $SNR_{dB} = 3$. ................. 38

3.9 Percent of Exact Pricing Objective Achieved in Dense Deployment for $SNR_{dB} = 3$. ................. 38

4.1 Percent Backhaul Achieved versus $B_m^{\text{max}}$ in Standard Deployment. ................. 49

4.2 Percent Backhaul Achieved versus $B_m^{\text{max}}$ in Dense Deployment. ................. 49

4.3 Percent of Exact Pricing Objective Achieved in Standard Deployment for $SNR_{dB} = 3$. ................. 50
4.4 Percent of Exact Pricing Objective Achieved in Dense Deployment for $SNR_{dB} =$ 3

\[ \text{50} \]
Since their inception, wireless communication systems have changed the way we communicate with one another. As technology advanced, wireless networks have transitioned from servicing primarily voice traffic to data traffic. The proliferation of devices such as smartphones, wearable technology and other connected electronics all need to be serviced with more reliability and higher data rates.

Communication, at its core, is the attribution of meaning to the variation of a physical signal. In wireless communication we modulate a baseband signal to a higher radio frequency (RF) for transmission. We divide the RF portion of the electromagnetic spectrum into frequency and time resource slots, each able to be used to service a user. Transmissions to different users occurring on the same resource slot causes interference that degrades the performance to all users scheduled on that slot. Government bodies regulate the usage and licensing of spectrum, auctioning off new spectrum conservatively and fetching high prices. Simply buying more spectrum is not an option for service providers to meet the increasing data demand.

The wireless channel is inherently a challenging medium to communicate over due to its physical properties: a signal is subject to potentially significant distortion due to the fact that the receiver additively combines multipath components that can combine destructively and the channel is time-varying due to the constantly changing physical environment. Ad-
ditionally, signal power is decreased by the objects blocking its path to the receiver (shadow fading) and also due to the freespace path loss where signal power falls off inversely proportional to square of the distance [1].

This path loss is the key idea behind cellular networks, because signal power decreases as distance between transmitter and receiver increases, it can be exploited. A service area is divided into cells each served by a base station. By having adjacent cells use non-overlapping sets of frequency slots, interference can be reduced at the cost of reduced spectral efficiency. In the light of the increased demand for data, current and future wireless standards employ full frequency reuse meaning all resource slots are available for use by all cells. This allows for greater data throughput but only if cells can coordinate to mitigate the interference.

Another way to increase capacity of a cellular network is to reduce the size of the cell: by reducing the distance between transmitter and receiver, better signal quality is possible and cell sizes are reduced, allowing resource slots to be reused more frequently. Since a cell of any size can roughly serve the same amount of users, this method boosts area spectral efficiency making it ideal for dense urban areas where users are concentrated.

Heterogeneous networks (HetNets) are an emerging technology to overlay the traditional network with dense deployments of small cells increasing the frequency reuse. There are many challenges that arise when a network is comprised of different tiers of cells served by access points that vary greatly in transmit power, processing power and connection to the backbone infrastructure. Based on dense deployments of access points, small cell (SC) networks have the potential to meet the rapid growth in demand for mobile data [2, 3]. While initially considered for reliable access indoors, outdoor SC networks are now being considered. These networks enhance the traditional cellular network by decreasing the distance between mobile user equipment (UE) and the access point (AP) and through greater frequency reuse within the SC layer. However, with transmissions from closely-spaced APs that may share frequencies, interference needs to be properly addressed and controlled to reap these benefits [4, 5]. Furthermore, while traditional base-stations (BS) are connected to the core network through large capacity/low delay backhaul links, SC APs usually do
not have that luxury: as an example indoor APs (femtocells) are usually connected to the core network via a broadband connection that has significantly less bandwidth and increased delay \cite{6,7}.

In a parallel development, Coordinated Multipoint (CoMP) has been proposed to increase network-wide data rates by having APs cooperate in their transmissions \cite{8}. There are two main variants of CoMP for the downlink. First, in Joint Scheduling and Beamforming CoMP (JB-CoMP), downlink transmissions are scheduled in adjacent cells to avoid interference and beamformers are designed jointly to reduce interference, the user data only needs to be available at the serving sector. Alternatively, in Joint Transmission CoMP (JT-CoMP) the data to a user is transmitted simultaneously from multiple APs requiring user data to be available at all cooperating APs. The trade-off in using JT-CoMP over JB-CoMP is reduced interference at the cost of less efficient use of backhaul capacity.

### 1.1 Motivation and Overview

Motivated by the need to manage interference and optimize use of resources, this thesis extends the notion of CoMP to dense SC networks. We maximize the weighted sum rate to users in the downlink of orthogonal frequency division multiple access (OFDMA) networks while imposing backhaul capacity constraints on each AP. Orthogonal subchannels are allocated dynamically in an interference-aware manner thereby efficiently utilizing available spectrum without suffering from excessive interference. In considering the use of CoMP for SC networks, the reduced available backhaul becomes a crucial limitation. While, in practice, backhaul connections can be either fixed (wired) or variable (e.g., a wireless backhaul), as in \cite{9} we impose a fixed backhaul capacity limit between the core network and each AP.

Due to the fact that systems that incorporate data sharing such as JT-CoMP require data to be available at all transmitting APs, this overhead is significant when each AP has a very limited backhaul capacity. We consider systems that coordinate at the scheduling and beamforming level while transmitting independent data streams at each AP.
In allocating resources to small cell networks there are several questions that must be answered:

- **User Association**: which set of APs will serve each user?
- **User Scheduling**: which user will be served in a given frequency slot?
- **Spectrum Allocation**: how should subchannels be allocated to scheduled users?
- **Backhaul Allocation**: how much backhaul capacity should be allocated to a given user?
- **Power Control**: how to choose transmit powers in a way that does not cause excessive interference while maximizing rate.
- **Beamforming**: when multiple transmit antenna are used, how to design precoders?

The answers to use questions are typically done independently in stages, in this thesis we propose a method of solving this problem *jointly*. By defining the link $mkn$ as the downlink transmission between AP $m$ and user $k$ on subchannel $n$, we see that when link $mkn$ is active (transmitting at a non-zero data rate), it is interpreted as user $k$ being associated with/scheduled for transmission on AP $m$ and allocated subchannel $n$ as the frequency resource.

Our system model uses features of modern networks: we choose to use an Orthogonal Frequency Division Multiple Access (OFDMA) network. Wherein, we divide the frequency resources into a large number of closely spaced orthogonal subcarriers and assign them to individual users to achieve multiple access. A user may be served by multiple APs provided that they use orthogonal subchannels (in most discussion of SCs, a user is assumed to be associated with a single AP). This feature is currently not implemented in practice however in the context of backhaul constrained networks, however the work in [10] showed that allowing a user to connect to multiple APs simultaneously, on different frequency slots, provides significant gains. In dense SC networks with backhaul constraints this means that multiple APs can coordinate and transmit independent data streams to the intended
user when the backhaul of a specific AP is inadequate to support the entire data stream. Coordinating APs only need to exchange Channel State Information (CSI) information, a relatively small overhead as compared with data sharing \[11\]. Furthermore, this flexibility simplifies the user association problem by allowing a greater degree of freedom over forcing a user to associate with only one AP. This allows users to be served by cooperating APs in orthogonal frequency clusters as opposed to JT-CoMP.

The weighted sum-rate maximization problem with user association is well known to be non-convex and NP-hard in general \[12\], thus difficult to solve. Finding the globally optimum solution can be done by methods such as branch-and-bound \[13\] or by polyblock approximation methods \[14\] but the complexity grows exponentially in the number of variables, making them unsuitable for even medium-scale problems. Lower complexity algorithms, such as successive convex approximation or dual methods \[15–18\], lead to locally optimal solutions and simulations show they perform close to global optimization methods.

1.2 Literature Survey

Constraints stemming from non-ideal backhaul connections have been studied extensively in the CoMP literature but mostly in the context of the cost of signaling overhead that needs to be exchanged between cooperative nodes: in both \[19\] and \[20\] downlink CoMP algorithms were proposed that reduce this overhead. Capacity bounds and a rate/backhaul trade-off in different scenarios were studied in \[21\] that showed that significant performance gains were possible using CoMP techniques even with a strongly limited backhaul. In \[22\], the authors use stochastic geometry to quantify the signaling overhead involved in cooperation. The effect of backhaul delay was investigated in \[23,24\], an issue not considered in this thesis.

Recent works in CoMP with clustering such as \[25,26\] propose algorithms without enforcing an explicit backhaul rate constraint at each AP. In \[27\], the authors adopt a data-sharing JT-CoMP model for the downlink where each AP is constrained by a finite capacity link. Using a reweighted $\ell_1$-norm technique, they proposed a scheme that jointly optimizes a clus-
ter of AP’s beamformers that serve a user sharing the same data. In [28], the authors studied the benefits of switching dynamically between JT-CoMP and JB-CoMP when backhaul is limited. In [29], it was shown that JB-CoMP outperforms JT-CoMP when the backhaul capacity is severely limited by eliminating the need to share user data across sectors. In light of this result, due to the limited backhaul capacity of SC APs, we seek algorithms that do not require data sharing among cooperating SCs.

Optimizing OFDMA systems further requires that subchannels be allocated in some optimal manner. This is a combinatorial problem and finding the optimal solution then requires a computationally prohibitive exhaustive search over all possible spectrum allocations and user schedules. The authors of [30] seek to maximize the weighted sum rate for MIMO-OFDMA APs and users in the downlink. There is no data sharing amongst cooperating APs. To reduce the complexity of frequency resource allocation, they utilize a polynomial complexity algorithm and then uses a two-dimensional nested subgradient search to calculate the optimal dual variables for power and backhaul feasibility to select the appropriate precoder transmit powers. The users are fixed to a given cell ahead of time and are only served by that AP. In [31], cooperating cells seek to optimize sum rate for a system of single antenna APs and users under a JT-CoMP scheme where APs select to link other APs to cooperate for joint transmission. By forming a cooperation link, APs improve the rates their user receives at the cost of expending that rate in the backhaul by sharing it amongst that link. Spectrum allocation, user scheduling, power control and backhaul feasibility portions of the heuristic algorithm are performed separately. The work in [32] extends this to a MIMO setup where APs and users have multiple antennas.

1.3 Thesis Contributions

In this thesis we make three contributions: first, we analyze the weighted-sum user rate maximization problem in the downlink for backhaul constrained OFDMA CoMP SC networks. Our analysis of the system model shows computational complexity can be reduced
through using a specific interference model that allows us to remove the integer constraints associated with user scheduling, cell association and spectrum allocation. However, the solution is feasible in that it meets the integer constraints. Importantly, avoiding the integer constraints allows us to solve large-scale problems in a time-efficient manner.

For single antenna APs where only one user can be scheduled on each subchannel, we develop an algorithm, consistent with [33], that can be applied to large-scale problems. This leads to a far more efficient algorithm, called Improved Iterative Waterfilling (IIWF), based on the dual decomposition method introduced in [18] and in [34] for the multiple antenna AP case.

Second, we propose a computationally efficient algorithm that jointly performs spectrum allocation, user association, power control and rate optimization. The algorithm uses a bisection search to determine the dual variables to satisfy both power and backhaul constraints. Importantly, we show that a single one-dimensional search for the associated Lagrange multiplier is adequate. This allows for a significantly simpler implementation with lower computational complexity over traditional subgradient and augmented Lagrangian approaches. The proposed algorithm solves the equations imposed by the Karush-Kuhn-Tucker (KKT) conditions directly and converges to a locally optimal solution by introducing a link price that accounts for interference caused by using a given link on other links on the same frequency resource. The algorithm has the further advantage that it can be implemented asynchronously at each AP. For the multiple transmit antenna (MISO) case, the algorithm leads to sparse solutions where sparsity is enforced directly by interference and link price instead of other methods such as reweighted $\ell_1$-norm approximation. The system backhaul capacity is explicitly included in the problem formulation rather than other methods of penalizing cluster sizes like the approximate $\ell_2$-norm penalty used in [35].

Third, while the algorithm described above assumes global knowledge of channel state information, we present fully-distributed and semi-distributed variants of the algorithm that allows control of the trade-off between performance and communication overhead.
1.4 Organization of Thesis

In Chapter 2 we introduce the system model and formulate the optimization problem. In Chapter 3 we analyze the optimality conditions for APs with a single transmit antenna and then develop the iterative algorithm with distributed and semi-distributed variants. In Chapter 4 we extend the results from the previous chapter to APs with multiple transmit antennas and develop the corresponding iterative algorithm with distributed and semi-distributed variants. We conclude the thesis and discuss further work in Chapter 5.

1.5 Notation

We use lower case italics for scalars (e.g. $x$), lower case boldface for column vectors (e.g. $\mathbf{x}$) and upper case boldface for matrices (e.g. $\mathbf{X}$). $\mathbf{X}^T$, $\mathbf{X}^H$, $\mathbf{X}^\dagger$, $\|\mathbf{X}\|$ denote the transpose, the Hermitian transpose, the pseudo-inverse and the Frobenius norm respectively for matrix $\mathbf{X} \in \mathbb{C}^{m \times n}$. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. $E(x)$ denotes expectation for random variable $x$. $\mathbf{1}\{x\}$ is the indicator function that is 1 when $x > 0$ and 0 otherwise.
Chapter 2

Backhaul Constrained Networks

2.1 System Model

We consider the downlink of an OFDMA system with \( N \) orthogonal subchannels, \( M \) APs each equipped with \( N_{t,m} \) transmit antennae and \( K \) single antenna users. The access points are subject to a power constraint and a finite capacity backhaul link to the core network. Figure 2.1 illustrates a network with \( M = 2 \) APs, \( K = 3 \) users and \( N = 3 \) orthogonal subchannels.

We index the link between AP \( m \) and user \( k \) on subchannel \( n \) as \( mkn \). We allow each user to be served by multiple APs on orthogonal subchannels by receiving independent data streams. The user data on link \( mkn \) is a complex symbol \( d_{mkn} \) such that:

\[
E[|d_{mkn}|^2] = 1, E[d_{m_1k_1n_1}d_{m_2k_2n_2}^*] = 0.
\]

(2.1)

The data signal transmitted by AP \( m \) on subchannel \( n \) is given by:

\[
x_{mn} = \sum_{k=1}^{K} w_{mkn} d_{mkn} \in \mathbb{C}^{N_{t,m}},
\]

(2.2)

where \( w_{mkn} \in \mathbb{C}^{N_{t,m}} \) is the beamformer vector for link \( mkn \).
The power used on link $mkn$ is

$$P_{mkn} = w_m^H w_{mkn}. \quad (2.3)$$

A link is said to be active when there is a non-zero power $P_{mkn} > 0$ allocated to it. For a given subchannel $n$, user $k$ is both associated with AP $m$ and scheduled for downlink transmission when the link is active. By allowing all APs to transmit to all users on any subchannel, dynamic user association, user scheduling and spectrum allocation can be jointly performed.

The signal received by user $k$ on subchannel $n$ is given by:

$$y_{kn} = \sum_{m'=1}^{M} h_{m'kn}^H x_{m'n} + z_{kn} \in \mathbb{C}, \quad (2.4)$$
where \( h_{mkn} \in \mathbb{C}^{N_t,m} \) is the complex channel and \( z_{kn} \in \mathbb{C} \) is the additive circularly-symmetric Gaussian noise with variance \( \sigma_{kn}^2 \) and zero mean.

We assume a block fading channel model where, on each subchannel, the channel gain is constant. Additionally, we assume perfect and instantaneous channel state information (CSI) is available at a centralized node and ignore the delay and overhead needed to share CSI between coordinating cells. To focus on algorithm development we assume perfect synchronization and ignore issues such as inter carrier interference.

For notational simplicity, we define the gain matrix \( G_{mkn} \) for link \( mkn \) as:

\[
G_{mkn} = h_{mkn} h_{mkn}^H \in \mathbb{C}^{N_t,m \times N_t,m}.
\]  

(2.5)

The interference seen on link \( mkn \) is a sum of the self-interference (intracell interference \( I_{mkn}^{\text{intra}} \) caused to user \( k \) due to transmissions from the same AP to another user, i.e., due to \( P_{mk'n}, \forall k' \neq k \)) and the more common intercell interference \( I_{mkn}^{\text{inter}} \) (i.e., due to \( P_{mk'n}, \forall m' \neq m \)):

\[
I_{mkn} = I_{mkn}^{\text{intra}} + I_{mkn}^{\text{inter}}
\]  

(2.6)

\[
I_{mkn}^{\text{intra}} = \sum_{k' \neq k} w_{mk'n}^H G_{mkn} w_{mk'n}
\]  

(2.7)

\[
I_{mkn}^{\text{inter}} = \sum_{m' \neq m} \sum_{k''} w_{m'k'n}^H G_{m'kn} w_{m'k''n}
\]  

(2.8)

Note that, due to the assumption that data to be transmitted on link \( mkn \) is only available at AP \( m \), in considering link \( mkn \), any power allocated to user \( k \) by an AP \( m' \neq m \) is considered interference. The expression in (2.8), therefore, includes a sum over all users and does not require the constraint \( k'' \neq k \).

The signal-to-interference-plus-noise (SINR) ratio \( \gamma_{mkn} \) for link \( mkn \) is:

\[
\gamma_{mkn} = \frac{w_{mkn}^H G_{mkn} w_{mkn}}{\sigma_{kn}^2 + I_{mkn}}
\]  

(2.9)
The achieved rate on link $mkn$ is a function of the SINR given by:

$$R_{mkn} = \log (1 + \gamma_{mkn}),$$  \hspace{1cm} (2.10)

and the achieved rate for user $k$ is the sum over all contributions from all APs $R_k = \sum_{m,n} R_{mkn}$. We ignore any gap to capacity of practical coding and modulation schemes.

The limited available power and backhaul impose constraints on the powers allocated and rates achieved:

$$N \sum_{n=1}^{N} P_{mkn} \leq P_{m}^{\text{max}}, \quad \forall m.$$  \hspace{1cm} (2.11)

The backhaul constraint for each AP $m$ translates into a sum rate constraint, i.e., the total data rate allocated to all active links $R_{m}^{\text{total}}$ must not exceed the capacity of the backhaul connection $B_{m}^{\text{max}}$:

$$K \sum_{k=1}^{K} N \sum_{n=1}^{N} R_{mkn} \leq B_{m}^{\text{max}}, \quad \forall m.$$  \hspace{1cm} (2.12)

This constraint forces APs to not cause more interference than is needed to achieve full backhaul-utilization.

It is important to note that since the achieved rates are functions of the power allocations, the backhaul constraint in (2.12) is, in fact, another form of a power constraint. We will use this fact later when developing the resource allocation algorithm. By varying an AP’s power budget $P_{m}^{\text{max}}$, backhaul capacity $B_{m}^{\text{max}}$ and transmit antennas $N_{t,m}$, we can model different tiers of the network.
2.2 Problem Formulation

We now formulate our optimization problem:

\[
\text{maximize } \sum_{k=1}^{K} \alpha_k \sum_{m=1}^{M} \sum_{n=1}^{N} R_{mkn} \quad (2.13)
\]

subject to:

\[
\sum_{k=1}^{K} \sum_{n=1}^{N} |w_{mkn}|^2 \leq P_{m}^{\text{max}}, \forall m, \quad (2.14)
\]

\[
\sum_{k=1}^{K} \sum_{n=1}^{N} R_{mkn} \leq B_{m}^{\text{max}}, \forall m, \quad (2.15)
\]

\[
\|w_{mkn}\|^2 \geq 0, \forall m, k, n, \quad (2.16)
\]

Our objective function is the weighted sum-rate (WSR) for all users and we are optimizing over the set of transmit beamformers \(w_{mkn}\) which is represented as \(W\). Here \(\alpha_k \geq 0\) are the weights indicating the priorities of users\(^1\) while (2.14) and (2.15) represent the power and backhaul constraints described earlier. Finally, (2.16) ensures that the magnitude of every transmit beamformer is always non-negative.

It is worth noting that the optimization problem in (2.13)-(2.16) does not include any predefined user-AP associations, user scheduling and spectrum allocations, which would manifest itself as an integer constraint. For example: if we defined \(K_m\) as the set of users associated with AP \(m\), selecting which users are in membership of each set is an integer constraint (e.g., introducing binary variables \(s_{mkn} \in \{0, 1\}\) to determine if user \(k\) is associated and scheduled by AP \(m\) on subchannel \(n\), and introducing the constraint that at most \(N_{t,m}\) non-zero \(s_{mkn} = 1\) or \(\|K_m\| \leq N_{t,m}\) (alternatively \(\sum_k 1 \{s_{mkn}\} \leq N_{t,m}, \forall m, n\)).

Let \(W^n_m\) be the transmit beamforming matrix for AP \(m\) on subchannel \(n\):

\[
W^n_m = [w_{m1n} \cdots w_{mkn}] \in \mathbb{C}^{N_{t,m} \times K} \quad (2.17)
\]

\(^1\)By setting the appropriate weights \(\alpha_k\), we can introduce fairness measures, e.g., proportional fairness where in time slot \(T+1\), we set \(\alpha_k = 1/\bar{R}_k(T)\), where \(\bar{R}_k(T)\) is the average rate allocated to user \(k\) in the previous \(T\) time slots.\(^{36}\)
By optimizing strictly over the set of all transmit beamformers $\mathcal{W}$, we seek to find sparse $W_m$'s which maximize our system objective (2.13) efficiently without resorting to mixed-integer-programs which are intractable for larger problem sizes due to their combinatorial nature (e.g., to find the global optimum of allocating $N$ channels to $K$ users requires evaluating every possible channel-user combination).

Solving the WSR maximization problem in (2.13)-(2.16) efficiently is the core contribution of this thesis. This problem combined with user scheduling, cell association, spectrum allocation and backhaul is well known to be non-convex and NP-hard [12]. We obtain a locally optimal solution by expanding on the Improved Iterative Waterfilling algorithm of [18] and [34] to include the explicit backhaul constraint and jointly perform scheduling, cell association and spectrum allocation.
Chapter 3

Resource Allocation with Single Antenna Access Points

In this chapter we focus on single antenna APs. In this case the beamformers and channels become scalars for convenience, we use $P_{mkn} = \|w_{mkn}\|^2$ to denote the power allocated on link $mkn$ and $G_{mkn} = \|h_{mkn}\|$ to denote the corresponding channel power. A link is active if $P_{mkn} \neq 0$. The optimization variables are, therefore, the powers. In the previous chapter we developed the system model and optimization problem at hand. It is worth noting that the optimization problem in (2.13)-(2.16) does not include any explicit integer constraints. We begin by determining the Karush-Kuhn-Tucker (KKT) conditions for the problem without explicit integer constraints and then show that the optimal conditions of this relaxed problem are satisfy the integer constraint that scheduling needs.
3.1 Lagrangian and KKT Conditions

To develop the KKT conditions, we start with the Lagrangian of the objective function with the power and backhaul constraints:

\[
\mathcal{L}(P, \nu, \lambda) = \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{m=1}^{M} \alpha_k \log \left( 1 + \frac{P_{mkn} G_{mkn}}{I_{mkn} + \sigma^2} \right) + \sum_{m=1}^{M} \nu_m \left( B_{m}^{\text{max}} - \sum_{n=1}^{N} \sum_{k=1}^{K} R_{mkn} \right) + \sum_{m=1}^{M} \lambda_m \left( P_{m}^{\text{max}} - \sum_{n=1}^{N} \sum_{k=1}^{K} P_{mkn} \right),
\]

where \(\nu_m\) denotes the dual variable for the backhaul constraint of AP \(m\), \(\lambda_m\) the dual variable for the power constraint of AP \(m\), with \(\nu\) and \(\lambda\) being, respectively, the corresponding vectors. Similarly \(P = [P_{mkn}]_{m=1,k=1,n=1}^{M,K,N}\) denotes the set of powers being optimized.

By substituting the explicit expression in (2.10) for rate, we can analyze the KKT conditions for local optimality. Taking the partial derivative of the Lagrangian with respect to link power \(P_{mkn}\) and, we have

\[
\frac{\partial \mathcal{L}(P, \nu, \lambda)}{\partial P_{mkn}} = \frac{\alpha_k G_{mkn}}{P_{mkn} G_{mkn} + I_{mkn} + \sigma^2} - \sum_{k'=1,k'\neq k}^{K} \frac{(\alpha_{k'} + \nu_m) G_{mk'n}}{I_{mk'n} + \sigma^2} \cdot \frac{P_{mk'n} G_{mk'n}}{P_{mk'n} G_{mk'n} + I_{mk'n} + \sigma^2} - \sum_{m'=1,m'\neq m}^{M} \sum_{k'=1}^{K} \frac{(\alpha_{k'} - \nu_{m'}) G_{mk'n}}{I_{mk'n} + \sigma^2} \cdot \frac{P_{m'm'k'n} G_{m'm'k'n}}{P_{m'm'k'n} G_{m'm'k'n} + I_{m'm'k'n} + \sigma^2} - \lambda_m = 0
\]

Rearranging this expression and introducing a link price \(t_{mkn}\), we have:

\[
\frac{\alpha_k - \nu_m}{P_{mkn}} + \frac{I_{mkn} + \sigma^2}{G_{mkn}} = \lambda_m + t_{mkn},
\]
where

\[ t_{mkn} = \sum_{k'=1, k' \neq k}^{K} \frac{(\alpha_{k'} - \nu_m) G_{mk'n}}{I_{mk'n} + \sigma^2} \frac{P_{mk'n} G_{mk'n}}{P_{mk'n} + I_{mk'n} + \sigma^2} + \sum_{m'=1, m' \neq m}^{M} \sum_{k'=1}^{K} (\alpha_{k'} + \nu_{m'}) G_{mk'n} \frac{P_{m'k'n} G_{m'k'n}}{P_{m'k'n} + I_{m'k'n} + \sigma^2} \]  

(3.4)

The link price in \( t_{mkn} \) summarizes the effect of the interference caused by using power \( P_{mkn} \), on link \( mkn \), on other users on the same subchannel. The vector of all link prices is represented as \( t \). Now using (3.3), we have an explicit expression for optimal power \( P_{mkn} \):

\[ P_{mkn} = \left( \alpha_k - \nu_m \frac{t_{mkn} - \lambda_m}{G_{mkn}} - \frac{I_{mk'n}}{G_{mk'n}} + \sigma^2 \right)^+, \]  

(3.5)

where \( (x)^+ \) denotes \( \max(x, 0) \). We see that the higher the link price \( t_{mkn} \) is, the less desirable it is to allocate power to link \( mkn \).

The system of equations in (3.5) is combined with the usual complementary slackness conditions:

\[ \lambda_m \left( \frac{P_{m_{\text{max}}} - \sum_{n=1}^{N} \sum_{k=1}^{K} P_{mkn}}{\nu_m} \right) = 0, \lambda_m \geq 0 \]  

(3.6)

\[ \nu_m \left( \frac{B_{m_{\text{max}}} - \sum_{n=1}^{N} \sum_{k=1}^{K} R_{mkn}}{\nu_m} \right) = 0, \nu_m \geq 0 \]  

(3.7)

To solve this KKT system, we must determine powers \( P \) and link prices \( t \) that satisfy (3.4)-(3.7). Finding the optimal dual variables can be done by using a two-dimensional search or other classical constrained optimization techniques such as the subgradient method but this is computationally intensive and might not be practical for large problem sizes.

Our main contribution of an efficient method of finding optimal dual variables will be discussed in the remainder of the section. The key is recognizing that the backhaul constraint is, in effect, also a power constraint. We note that when \( I_{mkn} \) and \( t_{mkn} \) are fixed, \( P_{mkn} \) is monotonic in the dual variables \( \nu_m \) and \( \lambda_m \). In addition, because (2.10) is a strictly
monotonic function in $P_{mkn}$ for fixed $I_{mkn}$, the power and rate expressions are coupled, i.e.,

\[ P_{mkn} = f(\nu_m, \lambda_m), \]  
\[ R_{mkn} = g(P_{mkn}) = g(f(\nu_m, \lambda_m)) \]  

(3.8) \hspace{1cm} (3.9)

where $f(\nu, \lambda)$ is monotonically increasing inversely proportional to $\lambda_m$ and $g(\cdot)$ is also monotonically increasing in $P_{mkn}$.

The fact that the constraints are coupled suggests that we, in fact, need only one of the two Lagrange multipliers. We denote $P_{m}^{\text{total}} = \sum_{k,n} P_{mkn}$ and $R_{m}^{\text{total}} = \sum_{k,n} R_{mkn}$ as the total power and rate sustained by AP $m$. In using an iterative solver, given power and rate tolerances, $\epsilon_1$ and $\epsilon_2$, there are four cases that can occur after each iteration at each AP $m$:

1. a tight power constraint, but slack backhaul constraint: $P_m^{\text{max}} - P_m^{\text{total}} \leq \epsilon_1$ and $B_m^{\text{max}} - R_m^{\text{total}} > \epsilon_2$,

2. a tight backhaul constraint, but slack power constraint: $B_m^{\text{max}} - R_m^{\text{total}} \leq \epsilon_2$ and $P_m^{\text{max}} - P_m^{\text{total}} > \epsilon_1$,

3. both constraints tight: $P_m^{\text{max}} - P_m^{\text{total}} \leq \epsilon_1$ and $B_m^{\text{max}} - R_m^{\text{total}} \leq \epsilon_2$,

4. the iterations have converged, but neither constraint is tight: $P_m^{\text{max}} - P_m^{\text{total}} > \epsilon_1$ and $B_m^{\text{max}} - R_m^{\text{total}} > \epsilon_2$.

We set $\nu_m = 0$ and use bisection search on just $\lambda_m$ (for fixed $I_{mkn}$ and $t_{mkn}$). For example, we set $\nu_m = 0$ and perform waterfilling on $\lambda_m$, for APs $m = 1, \ldots, M$ iteratively. We terminate when AP $m$ satisfies one of of the four cases for some chosen small values of $\epsilon_1, \epsilon_2 > 0$. This is because when maximum power is allocated by AP $m$, the resulting rate is a function of power: if the backhaul constraint is violated, the AP can decrease the transmit power by increasing $\lambda_m$. For convenience we will call these conditions the coupled bisection conditions. Complementary slackness tells us in cases 1 & 2 only one dual variable is non-zero and for case 3 because the constraints are coupled, bisection on either variable will still result in the same achieved rate and power allocation. Case 4 occurs when the
link prices and interference make it optimal to not use all of the available power/backhaul capacity.

### 3.2 User Scheduling

Using a single antenna AP would, in practice, constrain us to serve at most one user on each subchannel \( n \), i.e., for each \( n \), we must choose a \( k^* \) such \( P_{mk^*n_0} > 0 \) and \( P_{mk'n_0} = 0, \forall k' \neq k^* \). This manifests itself as the mixed-integer constraint:

\[
\sum_{k=1}^{K} 1\{P_{mkn}\} \leq 1, \forall m \in \{1, 2, \ldots, M\}, n \in \{1, 2, \ldots, N\}.
\]

We now show that for single antenna APs the solutions to (2.13)-(2.16) satisfy this property without having to enforce it explicitly. This gives us the advantage of solving large problem sizes without resorting to mixed-integer programs. In other words, in our model, as a purely mathematical construct, APs may connect to all users on a single subchannel; however, due to the nature of the problem, all but one of the links are inactive (the corresponding \( P_{mkn} = 0 \)). User association and spectrum allocation are, therefore, achieved by solving the optimization problem.

In this section, we do not consider the limited backhaul capacity of AP because we have shown in the previous section that we can transform the WSR problem with backhaul constraints to an equivalent power allocation using only sum power constraints.

We begin with the case of \( M = N = 1 \) and then generalize the result. For the single AP, single carrier case, the intuitive result is to allocate all available power to the user that achieves the greatest weighted rate. For simplicity, we use \( R_i \) to denote the weighted rate for user \( i \) in this section.

**Lemma 1.** For \( M = N = 1 \) the optimal solution is to allocate all power \( P_{\text{max}} \) to user \( k^* \) where \( k^* = \arg \max_k \left( \alpha_k \log(1 + \frac{G_k P_{\text{max}}}{\sigma_k^2}) \right) \).

**Proof.** We first consider the problem of a single AP with \( K = 2 \) users with channel gains
Chapter 3. Resource Allocation with Single Antenna Access Points

Given that $G_1 > G_2$, equal noise variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and user priority $\alpha_1 > \alpha_2$. Given a sum power constraint of $P_1 + P_2 \leq P_{\text{max}}$. We maximize the sum rate given by:

$$R = \alpha_1 \log \left(1 + \frac{G_1 P_1}{G_1 P_2 + \sigma^2}\right) + \alpha_2 \log \left(1 + \frac{G_2 P_2}{G_2 P_1 + \sigma^2}\right)$$

(3.10)

Without assuming that all the available power is allocated, we denote $P_m = P_1 + P_2$ as the total power allocated. Taking the derivative with respect to $P_1$, we find that there is a single critical point where $\frac{\partial R}{\partial P_1} = 0$:

$$P_1 = \frac{\alpha_2 G_2 (G_1 P_m + \sigma^2) - \alpha_1 G_2 \sigma^2}{G_1 G_2 (\alpha_1 + \alpha_2)}$$

(3.11)

Substituting (3.11) into $\frac{\partial^2 R}{\partial P_1^2}$:

$$\frac{\partial^2 R}{\partial P_1^2} = \frac{A_1 C_1}{(B_1 - C_1 P_1)^2} + \frac{A_2 C_2}{(B_2 + C_2 P_1)^2} \geq 0,$$

(3.12)

where $A_1 = \alpha(G_1^2 P_m + \sigma^2), B_1 = G_1^2 P_m^2 + 2G_1 P_m \sigma^2 + \sigma^4, C_1 = G_1^2 P_m + G_1 \sigma^2, A_2 = \alpha_2(G_2^2 P_m + G_2 \sigma^2), B_2 = G_2 P_m \sigma^2 + \sigma^4$ and $C_2 = G_2^2 P_m + G_2 \sigma^2$.

The result in Equation (3.12) indicates that $R$ is convex in $P_1$, with global minimum at (3.11). For $P_1 \in [0, P_{\text{max}}]$, $R$ therefore has two maxima at $P_1 = P_{\text{max}}$ and $P_1 = 0$. The achieved rates are:

$$R(P_1 = 0, P_2 = P_{\text{max}}) = \alpha_2 \log \left(1 + \frac{G_2 P_{\text{max}}}{\sigma^2}\right)$$

(3.13)

$$R(P_1 = P_{\text{max}}, P_2 = 0) = \alpha_1 \log \left(1 + \frac{G_1 P_{\text{max}}}{\sigma^2}\right).$$

(3.14)

The optimal power allocation is therefore to allocate all power to user 1 if $\alpha_1 \log \left(1 + G_1/\sigma^2\right) > \alpha_2 \log \left(1 + G_2/\sigma^2\right)$.
Extending to $K$-users we have:

$$
\text{maximize} \quad R(P_1, \ldots, P_K) = \sum_{k=1}^{K} \alpha_k \log \left( 1 + \frac{G_k P_k}{I_k + \sigma^2} \right) 
$$
(3.15)

subject to

$$
\sum_{k=1}^{K} P_k \leq P_{\text{max}},
$$
(3.16)

where $I_k = G_k(\sum_{k' \neq k} P_{k'})$ is the interference.

The weighted rate for user $k$ is:

$$
R_k(P_k, I_k) = \alpha_k \log \left( 1 + \frac{G_k P_k}{I_k + \sigma^2} \right).
$$
(3.17)

Without loss of generality, we order the users such that

$$
\alpha_1 \log(1 + \frac{G_1 P_{\text{max}}}{\sigma^2}) > \alpha_2 \log(1 + \frac{G_2 P_{\text{max}}}{\sigma^2}) > \cdots > \alpha_k \log(1 + \frac{G_k P_{\text{max}}}{\sigma^2}),
$$
(3.18)

then for some fixed powers $\{P_k, k = 1, \ldots, K\}$ such that $\sum_{k=1}^{K} P_k = P_{\text{max}}$ and $P_K + P_{K-1} = \bar{P}_{K-1}$, we have:

$$
R(P_1, \ldots, P_{K-1}, P_K) - R(P_1, \ldots, \bar{P}_{K-1}, 0) = R_{K-1}(P_{K-1}, I_{K-1}) + R_{K}(P_K, I_K)
$$

$$
- R_{K-1}(\bar{P}_{K-1}, I_{K-1} - G_{K-1} P_K)
< 0,
$$
(3.20)

where the final result is true since the optimal solution to the 2 user case is

$$
\max_{P_K, P_{K-1}} R_K + R_{K-1} = \alpha_{K-1} \log \left( 1 + \frac{G_{K-1} \bar{P}_{K-1}}{I_{K-1} - G_{K-1} P_K + \sigma^2} \right)
$$
(3.21)

for users $K$ and $K-1$, i.e., reallocating power from a weaker user to a stronger user increases the value of the objective function. Repeating this argument recursively for all users, we
have

\[ P^* = [P_{\text{max}}, 0, \ldots, 0]^T \]  \hspace{1cm} (3.22)

\[ R^*(P^*) = \alpha_1 \log \left( 1 + \frac{G_1 P_{\text{max}}}{\sigma_1^2} \right). \]  \hspace{1cm} (3.23)

with only user \( K = 1 \) being served with no interference and the lemma is proved. \( \square \)

By extending Lemma 1 to the multicarrier scenario with multiple APs, locally optimal solutions will have at most one user scheduled on each subchannel. Using an approach similar to the proof in the lemma, the following theorem is easily proved.

**Theorem 1.** With \( M > 1 \) APs, at an optimal power allocation \( P^* \), Lemma 1 holds for all APs i.e., optimal power allocation requires that at each AP only one user is active on any given subchannel.

**Proof.** Since every subchannel can be treated independently, without loss of generality we assume \( N = 1 \) and drop the index \( n \). We denote \( G_{mk}, P_{mk} \) and \( \sigma_{mk}^2 \) as the gain, power and noise on link \( mk \) respectively. First we fix powers \( P_{m'k} \) for AP \( \forall m' \neq m \). The interference from other APs seen by users served by AP \( m \) is:

\[ I_{mk} = \sum_{m' = 1, m' \neq m}^{M} G_{m'k} \left( \sum_{k' = 1, k' \neq k}^{K} P_{m'k'} \right). \]  \hspace{1cm} (3.24)

Their individual weighted link rates \( R_{mk} \) can be expressed as:

\[ R_{mk} = \alpha_k \log \left( 1 + \frac{G_{mk} P_{mk}}{I_{mk} + \sum_{k' \neq k} G_{mk} P_{mk'} + \sigma_{mk}^2} \right). \]  \hspace{1cm} (3.25)

By absorbing the intercell interference into noise terms \( \sigma_{mk}'^2 \) on each subchannel, our objective function is a sum of objectives for the problem in (3.15):

\[ \sigma_{mk}'^2 = \sigma_{mk}^2 + I_{mk}. \]  \hspace{1cm} (3.26)
The link price $t_{mk}$ for $mk$ can be computed according to (3.4) when $P_{m'k}$ for AP $\forall m' \neq m$ is fixed. From (3.3), we can obtain the appropriate power allocations $P_{mk}$ for AP $m$. When $t_{mk}$ is converged to a KKT point and AP $m$ has more than one non-zero power allocation $P_{mk}$, we have the same form as in Lemma 1 where the sum rate from AP $m$ is maximized when all power is allocated to user $k^*$ according to Lemma 1. Any AP that does not satisfy Lemma 1 can always increase its own rate allocated when the powers of other APs are fixed. Therefore, all APs will satisfy the mixed integer constraint at any optimal point.

Theorem 1 indicates that for any locally optimal solution, the scheduling property holds even without including the mixed-integer constraint explicitly in the optimization problem. Our choice of interference model therefore allows for the efficient computation of user schedules and spectrum allocation without explicitly enforcing integer constraints. It is worth commenting that while this result is intuitive, the form of the power allocation in (3.15) has not been considered in the literature. It is a result of our need to optimize (2.13)-(2.16) that does not include an integer constraint.

### 3.3 Low-Complexity Optimization Algorithm

Having introduced the crucial simplification of having to obtain only one Lagrange multiplier, we are now able to use well-accepted optimization techniques in developing our low-complexity optimization algorithm. Our algorithm for computing locally optimal power $P^*$ takes the form of waterfilling. Fixing link prices $t_{mkn}$ and interference $I_{mkn}$, we perform the coupled bisection on $\lambda_m$ to obtain optimal powers for those fixed $t_{mkn}$ and $I_{mkn}$. Since rate and power are coupled, from the coupled bisection conditions, we meet both the rate and power constraints. We then update the link prices computed from the new powers and repeat until $t_{mkn}$ and the objective (2.13) converge along with the complementary slackness conditions (3.6) and (3.7). We name our algorithm Improved Iterative Waterfilling with Backhaul (IIWFB) with the pseudocode provided in Algorithm 1. The reference in the pseudocode to user-scheduling is detailed below. The pseudocode for the coupled bisection
search is given in Algorithm 2.

Algorithm 1 Improved Iterative Waterfilling with Backhaul
1: Initialize $P_{mkn}$, $t_{mkn}$ $\forall m \in \{1, 2, \ldots, M\}, k \in \{1, 2, \ldots, K\}, n \in \{1, 2, \ldots, N\}$.
2: loop until $t_{mkn}$ and WSR converge
3: loop until $P$ converges
4: for AP $m = 1 \cdots M$ do
5: Calculate $I_{mkn}$ according to (2.6).
6: Obtain $\lambda_m$ via bisection search Algorithm 2.
7: Calculate $P$ using (3.5) including user-scheduling.
8: end for
9: end loop
10: Update $t_{mkn}$ according to (3.4).
11: end loop

Algorithm 2 Coupled Bisection Search on $\lambda_m$
1: Fix $t_{mkn}$ and $I_{mkn}$.
2: Initialize $\lambda_{m,\text{min}}$, $\lambda_{m,\text{max}}$, $\lambda_m$ and $\epsilon$.
3: loop until one of the coupled bisection conditions is satisfied
4: Calculate $P_{mkn}$ from (3.5) and update $P_{m}^{\text{total}}$.
5: Calculate $R_{mkn}$ from (2.10) and update $R_{m}^{\text{total}}$.
6: if $P_{m}^{\text{total}} > P_{m}^{\text{max}}$ or $R_{m}^{\text{total}} > B_{m}^{\text{max}}$ then
7: $\lambda_{m,\text{min}} = \lambda_m$.
8: else
9: $\lambda_{m,\text{max}} = \lambda_m$.
10: end if
11: $\lambda_m = (\lambda_{m,\text{min}} + \lambda_{m,\text{max}})/2$.
12: end loop

The algorithm allows for an asynchronous implementation where individual APs can perform the inner loop as link prices are updated. The APs and users can measure interference on each channel. However, to compute the link prices $t_{mkn}$, all power allocations and channel gains must be known at a central node. This requires the overhead of transmitting $2MKN$ variables whenever the prices need to be updated.

Initializing the algorithm for a feasible power allocation can be done by evening splitting powers across all links for each AP (i.e. $P_{mkn}^{\text{init}} = \frac{P_{m}^{\text{max}}}{NK}$). Given this initial power allocation, there is no guarantee that the allocated rates will be backhaul feasible. However, in practice the initial data rates for each user can be allocated as $R_{mkn}^{\text{init}} = \min\left[\frac{P_{m}^{\text{max}}}{NK}, \log\left(1 + \frac{P_{mkn}^{\text{init}} G_{mkn}}{I_{mkn} + \sigma_{mk}^2}\right)\right]$,
we allocate the minimum of either rates are split evenly between links or the Shannon capacity based on the initial power $P^{\text{init}}_{mkn}$.

The complexity for the bisection search is that it requires at most

$$n_{\text{iteration}} = \frac{\log \epsilon_0 - \log \epsilon}{\log 2}$$  \hspace{1cm} (3.27)

iterations per inner loop to find the correct dual variable where $\epsilon_0 = \lambda_{m,\text{max}} - \lambda_{m,\text{min}}$ is the initial search bracket size.

Convergence for iterative waterfilling algorithms is difficult to prove in general but can be enforced by slowing the update speed of the link prices [18]:

$$t_{mkn} = (1 - \beta) t^{\text{old}}_{mkn} + \beta \hat{t}_{mkn},$$  \hspace{1cm} (3.28)

where $0 < \beta < 1$, $t^{\text{old}}_{mkn}$ represents the link price from the previous iteration and $\hat{t}_{mkn}$ is the link price computed from the current allocated power $P$. In practice, we observe fast convergence in numerical simulations even for larger problem sizes, e.g., $MKN > 1000$, in fewer than 200 outer loop updates.

By using parameter $\beta$ to ensure convergence of the outer loop, the inner loop may not always satisfy the user scheduling requirements due to the fact that $t_{mkn}$ is now a weighted sum of the current and previous price and only at the optimal $t_{mkn}$ will the user scheduling requirement be satisfied. Alternatively, we can set $\beta = 1$ to not slow down the update of link prices which will satisfy the user scheduling property but at the cost of no guarantee on outer loop convergence.

We propose the following heuristic to resolve the scheduling conflicts for given invalid power allocation $P$: for AP $m$ transmitting on subchannel $n$, we select user $k^*$ where $k^* = \arg \max_k (P_{mkn})$ as the user that is scheduled and for users $k \neq k^*$ on $n_0$ the power is set to 0. With the corrected user scheduling we then iterate until the algorithm converges.

\footnote{While setting these users’ power to zero, seems to ‘waste’ available power, in practice these powers are negligibly small.}
Another approach, based on Lemma 1, is a user selection scheme that ensures that the user scheduling requirement is satisfied in all intermediate steps of the algorithm. The rationale is that the link prices $t_{mkn}$ now also affects the desirability to use $mkn$ in addition to $\alpha_k$, $I_{mkn}$ and $G_{mkn}$: fixing $I_{mkn}$ and $t_{mkn}$, AP $m_0$ selects user $k^*$ to be scheduled on subchannel $n_0$ at the beginning of Algorithm 2 where:

$$k^* = \arg \max_k \left( \frac{\alpha_k}{t_{mkn_0}} - \frac{I_{mkn_0}}{G_{mkn_0}} + \sigma^2 \right)^+, \quad (3.29)$$

and in the event of a tie, one of the tied users will be selected randomly to be scheduled. While this approach implies a slightly higher computational cost per iteration, it in fact, speeds convergence and, in our testing, executes faster. We therefore adopt it in our implementation.

This algorithm offers an decentralized implementation where APs perform power allocation and scheduling independently given link prices $t_{mkn}$ and interference can be measured $I_{mkn}$ locally at each AP.

### 3.4 Distributed and Semi-Distributed Algorithms

The numerical results in the next section will show that the algorithm described in Section 3.3 is very effective in optimizing the available resources while meeting the constraints imposed. However, implementing this algorithm requires global knowledge of all channel powers at a central node. In this section we investigate distributed and semi-distributed variants of the IIWFB algorithm. The variants are based on the fact that the algorithm allows for an asynchronous implementation where individual APs can perform the inner loop as link prices are updated.

We assume that each AP has knowledge of the power of its individual channels to all users. Further, we assume that the APs and users can measure interference on each subchannel.

The pricing terms $t$ play a crucial role in solving the KKT system: when the prices
converge, the algorithm converges to a locally optimal solution. However, to utilize the full pricing scheme, all APs require knowledge of allocated transmit powers, $P_{mkn}$, channel gains, $G_{mkn}$, and interference, $I_{mkn}$, of all links. In this section, we investigate how pricing affects the performance of the algorithm and consider the effects of schemes with reduced (the semi-distributed variant) and without exchanging the prices (the distributed variant).

We can model our network as two distinct sets: the set of APs $\mathcal{M}$ and the set of users $\mathcal{K}$. The link $mkn$ represents the connection that requires information transfer between members of $\mathcal{M}$ and $\mathcal{K}$. The algorithm introduced in Section 3.3 allows every AP $m \in \mathcal{M}$ to be connected to every user $k \in \mathcal{K}$ resulting in a fully connected network.

The pricing term in (3.4) can be divided into two parts:

$$t_{mkn} = t^{\text{intra}}_{mkn} + t^{\text{inter}}_{mkn},$$

(3.30)

where we define $t^{\text{intra}}_{mkn}$ and $t^{\text{inter}}_{mkn}$ as

$$t^{\text{intra}}_{mkn} = \sum_{k'=1, k' \neq k}^{K} \frac{(\alpha_{k'} - \nu_{m}) G_{mk'n}}{I_{mk'n} + \sigma^2} \cdot \frac{P_{mk'n} G_{mk'n}}{P_{mk'n} G_{mk'n} + I_{mk'n} + \sigma^2}$$

(3.31)

and

$$t^{\text{inter}}_{mkn} = \sum_{m'=1, m' \neq m}^{M} \sum_{k'=1}^{K} \frac{(\alpha_{k'} + \nu_{m'}) G_{mk'n}}{I_{mk'n} + \sigma^2} \cdot \frac{P_{m'k'n} G_{m'k'n}}{P_{m'k'n} G_{m'k'n} + I_{m'k'n} + \sigma^2}$$

(3.32)

The first term $t^{\text{intra}}_{mkn}$ is the intracell pricing term that reflects the cost of the interference created when allocating power to link $mkn$ to transmissions to other users $k' \neq k$ by AP $m$ on subchannel $n$. One could, therefore, propose a distributed algorithm in that each AP uses $t^{\text{intra}}_{mkn}$ for its link prices. However, this can be further simplified: as shown in Section 3.2, each AP will eventually select a single user to transmit to, thus because only selected user $k^*$ is being served on subchannel $n$ by AP $m$, $P_{mk'n} = 0$, $\forall k' \neq k^*$ and thus $t^{\text{intra}}_{mkn} = 0$, $\forall k' \neq k^*$.

In the scenario where no pricing information ($t^{\text{no pricing}}_{mkn} = t^{\text{intra}}_{mkn}$) is exchanged, APs perform iterative water-filling without coordination. After each iteration, APs see the in-
interference rise on links where other APs are transmitting that causes it to increase its own transmit power until either the power budget or backhaul capacity is expended. The advantage of this approach is that it does not require any exchange of information between APs. This benefit comes at the expense of sum-rate performance and is, therefore, fully distributed.

The second term \( t_{\text{inter}}^{mkn} \) is the intercell pricing term that reflects the cost of the interference created when allocating power to link \( mkn \) to transmissions by other APs \( m' \neq m \) to users. By gathering information about transmissions from other APs, intercell prices will signal to an AP to decrease transmit power when it is causing excessive interference. To compute the exact intercell price, each AP must have the channel gains \( G_{mkn} \) and interference \( I_{mkn} \) on all links that can incur a significant overhead as the problem size increases. To reduce this burden, we start by writing the intercell price as a sum of individual components from the other APs:

\[
F_{mm'n} = \sum_{m'=1, m' \neq m}^M F_{mm'n}, \tag{3.33}
\]

where

\[
F_{mm'n} = \sum_{k'=1}^K \left( \alpha_{kk'} + \nu_{m'n} \right) G_{mkn} I_{m'k'n} + \sigma^2 \left( P_{m'k'n} G_{m'k'n} + I_{m'k'n} + \sigma^2 \right). \tag{3.34}
\]

We note that each AP transmits to at most one user per subchannel, the actual information exchanged per component in \( F_{mm'n} \) is the transmit power \( P_{m'k'n}, G_{m'k'n} \) and interference \( I_{m'k'n} \).

By reducing the links between APs and users in the network, we can seek semi-distributed schemes, which do not require full exchange of channel gains and interference. One heuristic is to approximate the link prices by only exchanging intercell price components for APs that are within some distance \( d_{\text{max}} \) of one another. When \( d_{\text{max}} = 0 \), we have the No Pricing scheme and on the other extreme for \( d_{\text{max}} = \infty \), we have the full cooperation scheme of Section 3.3 with exact prices for each AP \( m \), we form a set of neighbours \( N_m \) that are located within \( d_{\text{max}} \) and approximate the price as:
\[ t_{mkn} = t_{mkn}^{\text{intra}} + \sum_{m' \in \mathcal{N}_m} F_{mm'n}. \] (3.35)

This heuristic captures the intuition behind the pricing terms: users that are desirable for AP \( m \) to transmit to have higher channel gains and other APs \( m' \) that are close to \( m \) can potentially become strong interferers for AP \( m \)'s scheduled user \( k^* \). For APs that are far apart, it is unlikely that both APs will transmit to the same user or interfere significantly with each other. The choice of \( d_{\text{max}} \) provides a control parameter to choose the level of information exchange in this semi-distributed algorithm.

### 3.5 Numerical Results and Discussion

In this section we present simulation results to illustrate the efficacy of our algorithms as described in Sections 3.3 and 3.4. Our simulations consider a single tier small cell network with APs and users uniformly randomly distributed in a square geographical area with side length \( d_{\text{area}} \) in meters. The channel model accounts for path loss and log-normal fading. The path loss, in dB, between the \( k^{th} \) user and \( m^{th} \) AP is given by the 3GPP model [40]:

\[ \gamma_{mk} = \max (15.3 + 37.6 \log_{10} d_{mk}, 36 + 20 \log_{10} d_{mk}) + q_{mk} W + L, \] (3.36)

where \( d_{mk} \) is the distance between user \( k \) and AP \( m \) in meters, \( q_{mk} \) is a random variable representing the total number of internal walls between user \( k \) and AP \( m \), \( W \) is the partition loss of internal walls in dB, \( L \) is the penetration loss of an outdoor wall in dB. The overall channel gain on link \( mkn \) is given by:

\[ G_{mkn}, dB = \gamma_{mk} + 10 \log_{10} |h_{mkn}|^2 + \psi_{mk}, \] (3.37)
Table 3.1: Simulation Parameters for Single Antenna AP Network

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Side Length $d_{\text{area}}$</td>
<td>500 m (Standard), 100 m (Dense)</td>
</tr>
<tr>
<td>Number of AP $M$</td>
<td>3 (Standard), 20 (Dense)</td>
</tr>
<tr>
<td>Number of Users $K$</td>
<td>10 (Standard), 30 (Dense)</td>
</tr>
<tr>
<td>Number of Subchannels $N$</td>
<td>16 (Standard), 25 (Dense)</td>
</tr>
<tr>
<td>System Bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Subchannel Bandwidth</td>
<td>15 kHz</td>
</tr>
<tr>
<td>UE Noise PSD</td>
<td>-174 dBm/Hz</td>
</tr>
<tr>
<td>UE antenna pattern &amp; gain</td>
<td>Omni-directional, 0 dB</td>
</tr>
<tr>
<td>AP antenna pattern &amp; gain</td>
<td>Omni-directional, 0 dB</td>
</tr>
<tr>
<td>AP Transmit Power Budget, $P_{m}^{\text{max}}$</td>
<td>24 dBm</td>
</tr>
<tr>
<td>$\mu_s$</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>10 dB</td>
</tr>
<tr>
<td>Traffic Model</td>
<td>Full buffer</td>
</tr>
<tr>
<td>$\alpha_k$</td>
<td>1 or Proportional fair</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.7</td>
</tr>
<tr>
<td>$\epsilon_1 = \epsilon_2$</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>Maximum IIWFB Iterations</td>
<td>200</td>
</tr>
<tr>
<td>Maximum Subgradient Iterations</td>
<td>2000</td>
</tr>
</tbody>
</table>

where $\psi_{mk} \sim \ln \mathcal{N}(\mu_s, \sigma_s)$ represents the log-normal shadowing and $h_{mkn} \sim \mathcal{CN}(0,1)$ represents the small-scale fading on link $mkn$. The unit variance of the small-scale fading is scaled by the SNR which is set to 3 dB unless otherwise stated.

The parameters common to all simulations are summarized in Table 3.1. For this comparison, all user weights are set to $\alpha_k = 1, \forall k$ so each user has equal priority with one example illustrating a proportional fair assignment. As the table shows, we consider two deployment scenarios:

1. Standard Deployment: $M = 3$, $K = 10$, $N = 16$ and $d_{\text{area}} = 500$ m
2. Dense Deployment: $M = 20$, $K = 30$, $N = 25$ and $d_{\text{area}} = 100$ m

We define a frequency reuse factor as:

$$f_r = \frac{N_{\text{active}}}{N},$$ (3.38)
where $N_{\text{active}}$ is the total number of active links, i.e., all links where $P_{mkn} > 0$.

Our algorithm is benchmarked against two other algorithms: a greedy scheme and the subgradient method \[37\] used in \[30\]. In point of fact, the WSR performance of the sub-gradient method is approximately the same as our IIWFB algorithm. However, our method stands out in its computational efficiency. In the greedy scheme, each subchannel is allocated only once to the link with the best channel gain $G_{mkn}$. Each AP then performs Algorithm 2 to calculate the appropriate water level that satisfies both power and backhaul constraints. This approach has no interference because each subchannel is only used by one link and thus the frequency refuse factor is $f_r = 1/M$. The advantage of our algorithm over the approach used in \[30\] is that instead of using nested two-dimensional subgradient search for the appropriate dual variables, we use the coupled bisection search in Algorithm 2. Subgradient methods are known for their slow convergence rate due to the use of a fixed step size when updating the subgradient whereas the bisection search will converge in a fixed number of iterations (3.27).

We initialize our algorithm with equal power allocation across all links, that is $P_{mkn} = P_{\text{max}} / (KN)$. In our simulations we compared initializing with equal power allocation and random power allocation across links scaled to be power feasible. We found that equal power allocation achieved equal or greater final objective about 63% of the time compared with random power allocation. However, the achieved objectives are within $\pm 0.005\%$ indicating robustness of our algorithm to choice of initialization.

Plots of the WSR versus $B_m^{\text{max}}$ in the standard and dense scenarios (Figure 3.1)\[2\] show that the proposed algorithm has significant gains in spectral efficiency over the greedy scheme. It is worth emphasizing that the dense deployment requires the optimization of $M \times K \times N = 15000$ variables. Established techniques, such as the subgradient method, are unable to effectively solve such large problems due to slow convergence \[38\].

Figure 3.2 shows that there is an increase in spectral efficiency that can be attributed to the frequency reuse achieved by the algorithm. Based on system fading conditions, locations

\[2\] The legend in Figure 3.1 also applies for Figure 3.3
Chapter 3. Resource Allocation with Single Antenna Access Points

Figure 3.1: Weighted Sum Rate versus $B_m^{\text{max}}$.

of APs/users and interference, the algorithm finds the level of frequency reuse for locally optimum weighted sum rate.

In Figure 3.3, we plot the achieved sum rate expressed as a percentage of the total available backhaul in the system $B_{\text{total}} = \sum_{m} B_m^{\text{max}}$. Since the transmit power is fixed at 24dBm, as the backhaul capacity increases, the system transitions from being backhaul constrained to being power constrained (i.e., more power is needed to achieve the total backhaul capacity) that is illustrated for the dense deployment in Figure 3.4. For a system in the backhaul limited regime ($B_m^{\text{max}} < 10$ (bits/s/Hz) in the standard deployment), the proposed algorithm achieves close to the total backhaul available. In the power-limited regime, the algorithm essentially ignores the backhaul constraint and behaves much as the standard improved iterative waterfilling algorithm of [18] (as in case 1 in Section 3.1).

Figure 3.5 compares the performance of the proposed algorithm (IIWFB) with the modified algorithm (IIWFB with Selection) with the selection step in (3.29) and the subgradient method. We see that the performance of the algorithm improves with the addition of the scheduling step and the performance is comparable with 2000 iterations of the subgradient method.
Figure 3.2: Frequency Reuse Factor versus $B_m^{\text{max}}$ for $SNR_{dB} = 3$ in Standard Deployment.

Figure 3.3: Percent Backhaul Achieved versus $B_m^{\text{max}}$. 
As mentioned before, the key benefit of the IIWFB algorithm is its low complexity. Given the various optimization steps involved, it is hard to evaluate the computational complexity of the algorithms at hand. As a proxy, Table II compares the execution times of three schemes in Figure 3.5 on an Intel Xeon E5345 @ 2.33 GHz. The execution time is averaged over 100 realizations of fading channels and spatial environments. As the table shows, the addition of the selection step on average increases the convergence speed of the proposed algorithm in standard deployment. Importantly, both IIWFB variants are two orders of magnitude faster that the subgradient approach. Furthermore, the subgradient approach is unable to optimize the dense scenario due to the problem size. It is also worth noting that the subgradient method requires synchronous updates of $P$ done at a central processor compared with IIWFB where the inner loop can be performed asynchronously and only requires a central processor to update the prices.

The significant increase in spectral efficiency achieved by the IIWFB algorithm over
Figure 3.5: Comparison of Percent Backhaul Achieved versus $B_{m}^{\text{max}}$ for $SNR_{dB} = 3$ in Standard Deployment.

Table 3.2: Algorithm Convergence Speed

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average Simulation Time (Standard)</th>
<th>Average Simulation Time (Dense)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIWFB</td>
<td>5.50 seconds</td>
<td>454.75 seconds</td>
</tr>
<tr>
<td>IIWFB with Scheduling</td>
<td>1.21 seconds</td>
<td>563.26 seconds</td>
</tr>
<tr>
<td>Subgradient Method</td>
<td>264.07 seconds</td>
<td>-</td>
</tr>
</tbody>
</table>

the greedy scheme can be attributed to the frequency reuse achieved by the algorithm - as shown in Figure 3.2. Based on system fading conditions and locations of APs and users, the algorithm automatically finds the optimal level of frequency reuse.

In Figure 3.6 we demonstrate proportional fairness by selecting $\alpha_k = 1/\bar{R}_k(T)$ with $T = 30$ previous time slots by plotting the logarithm of cumulative achieved rates for each user over all time slots, $R_k^{\Sigma}$. In this simulation, we fix the AP and user spatial locations while updating the small-scale fading in every time slot. Here $M = 3$, $K = 5$, $N = 8$ and $d_{area} = 100$ m.

In Figure 3.7 the scheduling performance of the proposed algorithm is compared with
a round-robin (RR) scheduler. Here the backhaul $B_{\text{max}} = \infty$, $M = 10$ APs, $K = 20$ users, $N = 1$ subchannel, $d_{\text{area}} = 100 \text{ m}$ and the round-robin scheduler allocates maximum power to the scheduled user. The empirical cumulative distribution function (CDF) of achieved user rates shows the significant increase in performance of our algorithm (IIWFB) over the round-robin scheduler (RR).

To evaluate the performance of the distributed and semi-distributed schemes proposed in Section 3.4, we compare the achieved objective with that of the algorithm with exact prices in Figure 3.8 and Figure 3.9. The amount of overhead for each scheme is compared as a percentage of the full pricing scheme. For both the standard and dense deployment, at severely limited backhaul capacity all algorithms achieve the total system capacity. Due to the sparseness of APs in the standard deployment, we consider $d_{\text{max}} = 200 \text{ m}$ and $d_{\text{max}} = 400 \text{ m}$ as the distance thresholds for neighbours to exchange pricing factors. Having a larger distance threshold ($d_{\text{max}} = 400 \text{ m}$ versus $d_{\text{max}} = 200 \text{ m}$) only improves the achieved objective slightly at the expense of increased overhead of 84% compared to 36% but shows
significant improvement over no pricing exchange in the standard deployment. For the dense deployment, we reduce the distance thresholds accordingly to \( d_{\text{max}} = 20 \) m and \( d_{\text{max}} = 40 \) m. We observe that the gains achieved for \( d_{\text{max}} = 40 \) m and \( d_{\text{max}} = 20 \) m thresholds are significant over the No Pricing scheme at the cost of 35% and 11% increase in overhead. This significantly improved objective when using a larger threshold can be attributed to the dense deployment: reduced distance between transmitting APs, both an increase in APs and users thus the more accurate prices allow the system to mitigate potential strong interferers.

### 3.6 Summary

In this work we studied joint user scheduling, cell association, spectrum allocation and power control for single antenna APs and users in downlink for a small cell network with backhaul constraints.

We showed that adopting the self-interference model, integer constraints in the scheduling and user association step can be bypassed, which allows the algorithm to scale for larger
Figure 3.8: Percent of Exact Pricing Objective Achieved in Standard Deployment for $SNR_{dB} = 3$.

Figure 3.9: Percent of Exact Pricing Objective Achieved in Dense Deployment for $SNR_{dB} = 3$. 
problem sizes. Furthermore, by writing the backhaul constraint as a power constraint, we avoided a two-dimensional search for the required Lagrange dual variables; this allows for efficient computation of the optimal dual variables via bisection, which allows for efficient computation to achieve locally optimal weighted sum rate utility. Furthermore, the algorithm allows for an asynchronous implementation where only pricing information must be computed and exchanged from a central processor. The dynamic frequency reuse property of this algorithm suggests that given a specific fading environment, the optimal frequency reuse factor and pattern can be computed. Finally, we present a fully-distributed and semi-distributed algorithm that allows control between performance and communication overhead required.
Chapter 4

Resource Allocation with Multiple Antenna Access Points

In Chapter 2, we formulated the optimization problem at hand for the case of multiple transmit antennas. In Chapter 3, we focused on the special case of a single antenna. In this Chapter 4, we determine the KKT conditions for the original problem (2.13)-(2.16) when the number of transmit antennae $N_{t,m} > 1$, to determine the algorithm for maximizing weighted sum-rate in the Multiple-Input-Single-Output (MISO) interference channel. As we will see, our algorithm builds on the algorithm developed in Chapter 3.
4.1 Lagrangian and KKT Conditions

The optimization problem is repeated here for convenience.

\[
\begin{align*}
\text{maximize} & \quad \sum_{k=1}^{K} \alpha_k \sum_{m=1}^{M} \sum_{n=1}^{N} R_{mkn} \\
\text{subject to } & \quad \sum_{k=1}^{K} \sum_{n=1}^{N} |w_{mkn}|^2 \leq P_m^{\text{max}}, \forall m, \quad (2.14) \\
& \quad \sum_{k=1}^{K} \sum_{n=1}^{N} R_{mkn} \leq B_m^{\text{max}}, \forall m, \quad (2.15) \\
& \quad \|w_{mkn}\|^2 \geq 0, \forall m, k, n, \quad (2.16)
\end{align*}
\]

We first dualize (2.13)-(2.16) with respect to the sum power and sum backhaul constraints:

\[
\mathcal{L}(W, \lambda, \nu) = \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{n=1}^{N} \alpha_k \log (1 + \gamma_{mkn}) + \sum_{m=1}^{M} \lambda_m \left( P_m^{\text{max}} - \sum_{k=1}^{K} \sum_{n=1}^{N} \|w_{mkn}\|^2 \right) + \sum_{m=1}^{M} \nu_m \left( B_m^{\text{max}} - \sum_{k=1}^{K} \sum_{n=1}^{N} R_{mkn} \right) \quad (4.1)
\]

where \( \nu_m \) denotes the dual variable for the backhaul constraint of AP \( m \), \( \lambda_m \) the dual variable for the power constraint of AP \( m \), with \( \nu \) and \( \lambda \in \mathbb{R}^M \) being, respectively the corresponding vectors.

Substituting the explicit equations for power and rate, we take the partial derivative
with respect to transmit beamforming vector $w_{mkn}$ to get:

$$\frac{\partial L(W, \lambda, \nu)}{\partial w_{mkn}} = \left( \frac{\alpha_k - \nu_m}{1 + \gamma_{mkn}} \right) \frac{w_{mkn}^H (G_{mkn} + G_{mkn}^H)}{(\sigma^2_{kn} + I_{mkn})^2} - \sum_{k=1, k' \neq k}^K \left( \frac{\alpha_{k'} - \nu_m}{1 + \gamma_{mk'n}} \right) \frac{w_{mk'n}^H G_{mk'n} w_{mk'n}}{(\sigma^2_{k'n} + I_{mk'n})^2} \left( \frac{w_{mkn}^H (G_{mk'n} + G_{mk'n}^H)}{\sigma^2_{k'n} + I_{mk'n}} \right)^2 - 2\lambda_m w_{mkn}. \quad (4.3)$$

We define the link price on $mkn$ as $T_{mkn} \in \mathbb{C}^{N_t \times N_t}$ which summarizes the interference that transmitting on link $mkn$ causes to other links:

$$T_{mkn} = \sum_{k'=1, k' \neq k}^K \left( \frac{\alpha_{k'} - \nu_m}{1 + \gamma_{mk'n}} \right) \frac{w_{mk'n}^H G_{mk'n} w_{mk'n}}{(\sigma^2_{k'n} + I_{mk'n})^2} \left( \frac{w_{mkn}^H (G_{mk'n} + G_{mk'n}^H)}{\sigma^2_{k'n} + I_{mk'n}} \right)^2 + \sum_{m'=1, m' \neq m \neq k' \neq k}^M \sum_{k''=1}^K \left( \frac{\alpha_{k''} - \nu_{m'}}{1 + \gamma_{m'k'\prime n}} \right) \frac{w_{m'k'\prime n}^H G_{m'k'\prime n} w_{m'k'\prime n}}{(\sigma^2_{k'\prime n} + I_{m'k'\prime n})^2} \left( \frac{w_{mkn}^H (G_{m'k'\prime n} + G_{m'k'\prime n}^H)}{\sigma^2_{k'\prime n} + I_{m'k'\prime n}} \right)^2. \quad (4.4)$$

We now set the partial derivative to 0 to see the KKT conditions for local optimality are:

$$\frac{\partial}{\partial \sigma^2_{kn} + I_{mkn} + w_{mkn}^H G_{mkn} w_{mkn}} \left( \frac{(\alpha_k - \nu_m) G_{mkn} w_{mkn}}{(\sigma^2_{kn} + I_{mkn} + w_{mkn}^H G_{mkn} w_{mkn})^2} = (T_{mkn} + \lambda_m I_p) w_{mkn}, \quad (4.5a) \right.$$  

$$\lambda_m \left[ P_m^{\text{max}} - \sum_{m=1}^M \sum_{k=1}^K \sum_{n=1}^N \|w_{mkn}\|^2 \right] = 0, \quad (4.5b)$$

$$\nu_m \left[ B_m^{\text{max}} - \sum_{m=1}^M \sum_{k=1}^K \sum_{n=1}^N R_{mkn} \right] = 0, \quad (4.5c)$$

where (4.5b) and (4.5c) are the usual complementary slackness conditions.

### 4.2 Low-Complexity Algorithm

In this section, we propose an iterative algorithm to solve the KKT conditions (4.5a)-(4.5c). Using an approach similar to that of in [34] we have the following result:
Lemma 2. The optimal beamforming vectors for (2.13)-(2.16) are of the form:

\[ \mathbf{w}_{mkn} = \beta_{mkn} \mathbf{A}_{mkn}^\dagger \mathbf{h}_{mkn} \]  \hspace{1cm} (4.6)

where \( \mathbf{A}_{mkn} = \mathbf{T}_{mkn} + \lambda_m \mathbf{I}_{N_t,m} \), \( \lambda_m \geq 0 \) and \( \beta_{mkn} \geq 0 \) are constants to be determined.

Proof. Let us consider when all APs are power constrained, that is \( \nu_m = 0, \forall \ m = 1 \cdots M \). For the equality to hold in (4.5a), there are four cases for \( \lambda_m \) and \( \mathbf{w}_{mkn} \):

- \( \text{C1 : } \lambda_m \neq 0 \) and \( \mathbf{w}_{mkn} \neq 0 \) \hspace{1cm} (4.7a)
- \( \text{C2 : } \lambda_m = 0 \) and \( \mathbf{w}_{mkn} \neq 0 \) \hspace{1cm} (4.7b)
- \( \text{C3 : } \lambda_m \neq 0 \) and \( \mathbf{w}_{mkn} = 0 \) \hspace{1cm} (4.7c)
- \( \text{C4 : } \lambda_m = 0 \) and \( \mathbf{w}_{mkn} = 0 \) \hspace{1cm} (4.7d)

For (4.7c) and (4.7d), the user does not receive any data on link \( mkn \), we shall examine (4.7a) and (4.7b) in detail. When (4.7a) is true, since \( \mathbf{t}_{mkn} \) is a positive-definite matrix we have:

\[ (\mathbf{T}_{mkn} + \lambda_m \mathbf{I}_{N_t,m}) \mathbf{w}_{mkn} \neq 0. \]  \hspace{1cm} (4.8)

Combined with (4.5a), this means \( \mathbf{h}_{mkn}^H \mathbf{w}_{mkn} \neq 0 \). Therefore, the non-zero beamformer \( \mathbf{w}_{mkn} \) must be of the form:

\[ \mathbf{w}_{mkn} \propto [\mathbf{T}_{mkn} + \lambda_m \mathbf{I}_{N_t,m}]^\dagger \mathbf{h}_{mkn}. \]  \hspace{1cm} (4.9)

Since \( \lambda_m \geq 0 \), AP \( m \) is power constrained, we introduce positive scalar \( \beta_{mkn} \) to ensure that power constraint (2.11) is satisfied.

For case (4.7b), with \( \lambda_m = 0 \) and \( \mathbf{w}_{mkn} \neq 0 \), this implies that for (4.8) to hold, we have two situations:
1. \( h^H_{mkn} w_{mkn} = 0 \) and \( T_{mkn} w_{mkn} = 0 \).

2. \( h_{mkn} \in \text{range}\{T_{mkn}\} \).

The first situation means the transmit beamformer \( w_{mkn} \) is orthogonal to the channel vector \( h_{mkn} \) so the user receives no data from AP \( m \) on subchannel \( n \). We can set \( w_{mkn} = 0 \) without decreasing the objective function and even reduce interference to other links.

The second situation shows that there is a non-zero beamforming vector which satisfies (4.9). This concludes our proof.

Now that the form for optimal \( w_{mkn} \) is known, we seek an efficient method of finding the optimal \( w^*_{mkn}, \lambda^* \) and \( \nu^* \). When \( w_{mkn} \neq 0 \), we can substitute (4.6) into (4.5a) and rearrange the terms to get:

\[
\beta^2_{mkn} = \frac{[(\alpha_k - \nu_m) h^H_{mkn} A^\dagger_{mkn} h_{mkn} - I_{mkn} - \sigma^2_{kn}]}{(h^H_{mkn} A^\dagger_{mkn} h_{mkn})^2},
\]

when interference \( I_{mkn} \) and price \( t_{mkn} \) is fixed, \( \beta_{mkn} \) is a function of \( \lambda_m \). We note that the positivity constraint (2.16) and \( \beta_{mkn} \geq 0 \) naturally enforces sparsity in our solution. Links which are not desirable due to the interference or high prices \( T_{mkn} \) will be dropped leaving a sparse \( W^n_m \). The proof in [34] shows that (4.10) is a non-negative non-increasing function of \( \lambda_m > 0 \) and furthermore, (4.10) is a strictly decreasing function of \( \lambda_m > 0 \) whenever \( \beta_{mkn} > 0 \). Now, we can determine the appropriate \( \lambda_m \) via bisection search to find the appropriate \( \beta_{mkn} \) that satisfies (2.11) by using (4.10) and (4.11).

\[
\sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{n=1}^{N} \|\beta_{mkn} A^\dagger_{mkn} h_{mkn}\|^2 \leq P_{m}^{\text{max}}
\]

(4.11)

Using the same approach as for the single-antenna AP case, when \( I_{mkn} \) is fixed \( R_{mkn} \) is strictly a function of \( w_{mkn} \) which scales with \( \beta_{mkn} \). The coupled bisection search on \( \lambda \) in Algorithm 4 below will find the optimal \( \beta_{mkn} \) that satisfies both (2.11) and (2.12) can be used. We fix interference from other APs \( I_{\text{inter}}_{mkn} \) (2.8) with prices \( T_{mkn} \) and perform bisection.
search on $\lambda_m$. With each new value of $\lambda_m$, we compute the corresponding $\beta_{mkn}$’s and update $\mathbf{w}_{mkn}$ to calculate the sum power $P_{m}^{\text{total}}$ and sum rate $R_{m}^{\text{total}}$. Once we have the sum power and sum rate for a given $\lambda_m$, we adjust it until we satisfy the coupled bisection conditions.

In Algorithm 3 we have our iterative algorithm Improved Iterative Waterfilling with Backhaul for solving (2.13)-(2.16) for the multiple transmit antenna system. We update prices $t_{mkn}$ in the outer loop and compute $\lambda_m$ for fixed $t_{mkn}$ and $I_{mkn}^{\text{inter}}$ using the coupled bisection search Algorithm 4 in the inner loop. In practice, we would set $l_{\text{in}}^{\text{max}}$ and $l_{\text{out}}^{\text{max}}$ as the maximum iterations allowed for the inner and outer loop respectively.

The computational complexity of the algorithm scales with computing the pseudoinverse of $\mathbf{T}_{mkn}$ in the bisection search which scales as $MNKN_{t,m}^3$. To reduce the complexity, we can remove users $k'$ from AP $m$’s user set if after several iterations the associated beamformers $\mathbf{w}_{mk'n}$ remain zero. As is in the single antenna case, convergence conditions for the IIWFB algorithm is not proven but we observe that convergence is always observed in our simulations.

To reduce the communication overhead needed to compute $t_{mkn}$, we can adapt the method used in Section 3.4 to the MISO system. We define the intracell prices $\mathbf{T}_{mkn}^{\text{intra}}$ and intercell prices $\mathbf{T}_{mkn}^{\text{inter}}$:

\[
\mathbf{T}_{mkn}^{\text{intra}} = \sum_{k'=1,k'\neq k}^{K} \frac{(\alpha_{k'} - \nu_m) \gamma_{mk'n} \mathbf{G}_{mk'n}}{\sigma_{k'n}^2 + I_{mkn} + \mathbf{w}_{mk'n} \mathbf{G}_{mk'n} \mathbf{w}_{mk'n}},
\]

\[
\mathbf{T}_{mkn}^{\text{inter}} = \sum_{m'=1,m'\neq m}^{M} \mathbf{F}_{mm'n},
\]

where

\[
\mathbf{F}_{mm'n} = \sum_{k''=1}^{K} \frac{(\alpha_{k''} - \nu_{m'}) \gamma_{m'k'n} \mathbf{G}_{m'k'n}}{\sigma_{k'n}^2 + I_{m'k'n} + \mathbf{w}_{m'k'n} \mathbf{G}_{m'k'n} \mathbf{w}_{m'k'n}} \in \mathbb{C}^{N_{t,m} \times N_{t,m}}.
\]

The algorithm offers a decentralized implementation where each AP performs beamforming and power allocation independently given the prices $\mathbf{T}_{mkn}$ and the interference $I_{mkn}$ can be measured locally at each AP.
Using [3.35] we reduce communication overhead by only exchanging pricing information between neighbours within $d_{\text{max}}$ distance.

**Algorithm 3** Improved Iterative Waterfilling with Backhaul for MISO

1. Initialize $w_{mkn}$, $T_{mkn}$ for $m,k,n$.
2. loop until $T_{mkn}$ and WSR converge
3. loop until $W$ converges
4. for AP $m = 1 \cdots M$ do
5. Calculate $I_{mkn}$ according to (2.6).
6. Obtain $\lambda_m$ via bisection search Algorithm 4.
7. Update $w_{mkn}$ according to new $\lambda_m$ from (4.6).
8. end for
9. end loop
10. Update $T_{mkn}$ according to (3.4).

**Algorithm 4** Coupled Bisection Search on $\lambda_m$ for MISO

1. Fix $T_{mkn}$ and $I_{mkn}^{\text{inter}}$.
2. Initialize $\lambda_m, \lambda_{m,\text{min}}$, $\lambda_{m,\text{max}}$, $\lambda_m$ and $\epsilon$.
3. loop until $I_{mkn}^{\text{intra}}$ converges and one of the coupled bisection conditions (3.1) is satisfied
4. Calculate $\beta_{mkn}$ from (4.10).
5. Update $w_{mkn}$ from (4.6).
6. Calculate $I_{mkn}^{\text{intra}}$ from (2.7).
7. Calculate $P_{mkn}$ from (2.3) and update $P_m^{\text{total}}$.
8. Calculate $R_{mkn}$ from (2.10) and update $R_m^{\text{total}}$.
9. if $P_m^{\text{total}} > P_m^{\text{max}}$ or $R_m^{\text{total}} > B_m^{\text{max}}$ then
10. $\lambda_{m,\text{max}} = \lambda_m$.
11. else
12. $\lambda_{m,\text{min}} = \lambda_m$.
13. end if
14. $\lambda_m = (\lambda_{m,\text{min}} + \lambda_{m,\text{max}})/2$.
15. end loop

### 4.3 Numerical Results and Discussion

In this section we present simulation results for the algorithm presented in this chapter using the same channel model as that was used in 3.5 and the simulation parameters in Table 4.1. User weights $\alpha_k$ are set equally to 1. While we have not derived an equivalent scheduling lemma in the MISO case analog to that of the one for SISO in Section 3.2.
our simulations show that upon convergence each AP will have at most $N_{t,m}$ non-zero beamformers $w_{mkn}$ (e.g. rank $\{W_{m}^{n}\} \leq N_{t,m}$).

1. Standard Deployment: $M = 5$, $K = 10$, $N = 1$, $N_{t,m} = 3$ and $d_{\text{area}} = 100$ m

2. Dense Deployment: $M = 5$, $K = 20$, $N = 1$, $N_{t,m} = 4$ and $d_{\text{area}} = 100$ m

We choose initial beamformers where users are first associated with the AP with the strongest signal. Each AP then schedules users with the semi-orthogonal user scheduling from [41] using semi-orthogonality cutoff $\alpha = 0.2$. The initial beamformers are in-cell zero-forcing beamformers with equal power allocated to each scheduled link, links which are not scheduled are set to 0. We denote this user schedule as $S_{\text{SUS}}$ and $W_{\text{SUS}}$ as the initial beamformer.

In Figure 4.1-Figure 4.2, we see the performance of the proposed algorithm compared under several setups.

1. **IIWFB SUS Init**: initialized with $S_{\text{SUS}}$ and $W_{\text{SUS}}$.

2. **IIWFB SUS Init - Fixed Schedule**: initialized with $S_{\text{SUS}}$ and $W_{\text{SUS}}$. Schedule is fixed to $S_{\text{SUS}}$.

3. **IIWFB SUS No Pricing**: initialized with $S_{\text{SUS}}$ and $W_{\text{SUS}}$. Using $\tilde{t}_{mkn} = t_{\text{intra}}^{mkn}$, no pricing from neighbouring APs.

4. **Per Cell SUS-ZF**: initialized with $S_{\text{SUS}}$ and $W_{\text{SUS}}$. Schedule is fixed to $S_{\text{SUS}}$, each cell performs SUS-ZF and iterates until $I_{mkn}$ converges. Coupled bisection is used to satisfy both power and backhaul constraints during power allocation.

The proposed algorithm (IIWFB SUS Init) achieves a higher WSR when user schedules are not fixed given the same initial beamformers. Performance is worse than Per Cell SUS-ZF when no intercell prices are exchanged.

We observe similar results when evaluating the performance of the semi-distributed and distributed schemes proposed in Section [3.3] Using $d_{\text{max}} = 60$ achieves roughly 95% of the
Table 4.1: Simulation Parameters for Multiple Antenna AP Network

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Side Length $d_{area}$</td>
<td>100 m (Standard), 100 m (Dense)</td>
</tr>
<tr>
<td>Number of AP $M$</td>
<td>5 (Standard), 5 (Dense)</td>
</tr>
<tr>
<td>Number of Users $K$</td>
<td>10 (Standard), 20 (Dense)</td>
</tr>
<tr>
<td>Number of Subchannels $N$</td>
<td>1 (Standard), 1 (Dense)</td>
</tr>
<tr>
<td>System Bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Subchannel Bandwidth</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Noise PSD</td>
<td>-174 dBm/Hz</td>
</tr>
<tr>
<td>UE Noise Figure</td>
<td>9 dB</td>
</tr>
<tr>
<td>UE antenna pattern &amp; gain</td>
<td>Omni-directional, 0 dB</td>
</tr>
<tr>
<td>AP antenna pattern &amp; gain</td>
<td>Omni-directional, 0 dB</td>
</tr>
<tr>
<td>AP Transmit Power Budget, $P_{m}^{\text{max}}$</td>
<td>24 dBm</td>
</tr>
<tr>
<td>$\mu_s$</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>10 dB</td>
</tr>
<tr>
<td>Traffic Model</td>
<td>Full buffer</td>
</tr>
<tr>
<td>$\alpha_k$</td>
<td>1</td>
</tr>
<tr>
<td>$\epsilon_1 = \epsilon_2$</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>Maximum Inner Loop Iterations</td>
<td>10</td>
</tr>
<tr>
<td>Maximum Outer Loop Iterations</td>
<td>50</td>
</tr>
</tbody>
</table>

objective with exact pricing using only 57% of the overhead. The lose in performance is less significant for lower backhaul $B_m^{\text{max}}$ due to the fact that APs can transmit with full power $P_{m}^{\text{total}} = P_{m}^{\text{max}}$ while still approaching the full backhaul usage while for higher backhaul capacity the performance suffers because with incomplete prices APs will allocate power without knowledge of its impact on other links.
Chapter 4. **Resource Allocation with Multiple Antenna Access Points**

Figure 4.1: Percent Backhaul Achieved versus $B_{m}^{\text{max}}$ in Standard Deployment.

Figure 4.2: Percent Backhaul Achieved versus $B_{m}^{\text{max}}$ in Dense Deployment.
Chapter 4. Resource Allocation with Multiple Antenna Access Points

Figure 4.3: Percent of Exact Pricing Objective Achieved in Standard Deployment for $SNR_{dB} = 3$.

Figure 4.4: Percent of Exact Pricing Objective Achieved in Dense Deployment for $SNR_{dB} = 3$. 

- $d_{max} = 30$ (17% overhead)
- $d_{max} = 60$ (57% overhead)
- No Pricing
4.4 Summary

In this chapter we extended the results of Chapter 3 to allow for APs with $N_{t,m} > 1$ by determining the KKT conditions for local optimality. We then proposed the Improved Iterative Waterfilling with Backhaul for MISO systems which allows us find locally optimum solutions to the WSR. The positivity constraint on power allows the algorithm to encourage sparsity naturally and thus perform joint user association, scheduling, subchannel allocation, beamforming and power control. A semi-distributed and distributed scheme is proposed based on the approach used in Section 3.4. Finally, we evaluate the simulation performance of the algorithm illustrating the efficacy and performance trade-offs when using the distributed and semi-distributed variants.
Chapter 5

Conclusion

In this thesis, we studied resource allocation for maximizing the weighted sum-rate of users in backhaul-constrained small cell networks. As shown in the literature, the benefits of cooperation in resource allocation is especially important in small cell networks where interference needs to be mitigated to realize the capacity gains offered. Schemes which can jointly optimize user scheduling, cell association, spectrum allocation and power control can vastly improve performance over performing each step in stages.

In the single antenna case, we showed that we can relax the integer constraint where only one user must be scheduled on a subchannel. In analyzing the KKT conditions of the non-convex problem, we propose an efficient iterative algorithm which uses the notion of link prices to jointly optimize user association, scheduling, spectrum allocation and power control to converge at a locally optimal solution. We introduce an efficient coupled bisection search which can solve for both power and backhaul dual variables with an one-dimensional search which allows each step of the algorithm to have feasible power and backhaul allocations. We then analyze the effect of link prices and develop a distributed and semi-distributed algorithm which allows for a trade-off between performance and communication overhead.

The algorithm proposed for the single antenna case is then extended to APs equipped with multiple transmit antennas. We use the same link pricing approach and coupled bisection to develop an iterative algorithm for the MISO system. Each AP designs sparse
beamformers for users with feasible power and backhaul in each intermediate step. Again, distributed and semi-distributed variations of the algorithm are presented.

The work in this thesis could be extended in multiple ways:

- We studied the optimization of the weighted sum-rate objective and greatly reduced the computational complexity by using the coupled-bisection search. Other applications of the coupled-bisection can be evaluated for alternative objective functions.

- To reduce the overhead of computing link prices, we used distance as a heuristic. Investigation into other heuristics for approximating link prices to further reduce overhead and complexity may be useful.

- In adopting a cooperation model, we allowed APs to jointly serve users without having to share data. While our analysis through numerical simulations show the efficacy of this, a comparison into the performance of carrier aggregation versus JP-CoMP clustering should be studied when APs are under limited backhaul capacities.

- The effect of delay in the backhaul has not been considered, further study to include backhaul delay with the results from this work would provide a more accurate backhaul model.
Bibliography


