Joint Power Control and User Association in Downlink Heterogeneous Networks

by

David Yiwei Ding

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

Abstract

Joint Power Control and User Association in Downlink Heterogeneous Networks

David Yiwei Ding

Master of Applied Science

Graduate Department of Electrical and Computer Engineering

University of Toronto

2016

As 5G aims to provide 1000-fold network capacity to mobile networks [1], heterogeneous networks (HetNet) have become a promising technology to deliver this ambitious goal. This thesis proposes a novel heuristic algorithm to maximize network utility of a downlink SISO HetNet jointly over user association and BS power control. This thesis has the following main contributions. First, the proposed problem formulation allows for investigation of a trade-off between network utility and network power consumption in the HetNet. Through finding the appropriate balance in the trade-off, the proposed algorithm forms a non-greedy heuristic to determine the appropriate activation of BSs under arbitrary total network consumption constraints. Greedy heuristics as motivated by those in literature are further used to compare with the proposed heuristic. Furthermore, this thesis considers the division of the downlink spectrum into multiple sub-bands and optimizes user association and BS powers independently over each sub-band. This provides an interesting alternative to the existing "soft frequency reuse" (SFR) scheme widely employed in mobile networks to maintain high network capacity while ensuring coverage for cell-edge users. Rate performances under various sub-band divisions are compared with the SFR scheme to examine the impact on user rates by considering a two-tier HetNet in a regular hexagonal deployment, where pico BSs are placed at known locations near the edges of macro-cells.

To my mother and brother

Acknowledgements

It has been an honour and privilege to have the opportunity to work in the Communications Group at The Edward S. Rogers Sr. Department of Electrical And Computer Engineering of University of Toronto with amazing and inspiring people. I would like to first and foremost thank my supervisor, Prof. Wei Yu, for his critical guidance and support on this work. It has been of great pleasure to have him as my supervisor as not only did I benefit from his direct input along the way, but also from his passion and insights in the fields related to my work.

My deepest gratitude is also towards Prof. Ben Liang, Prof. Elvino Sousa, and Prof. Raymond Kwong for taking their time serving on my MASc. committee and providing invaluable advice on my thesis.

I would also like to thank Ya-Feng Liu from the Chinese Academy of Sciences for his very helpful insights on solving difficult optimization problems related to this thesis. Also, special thanks to Kaiming Shen, Binbin Dai, and Liang Liu for the direct help they provided on my thesis. I would also like to thank everyone else from the communications lab for the wonderful time spent with them, which includes Caiyi Zhu, Pratik Patel, Alex Daniels, Peter Thompson, Foad Sohrabi, Arvin Ayoughi, Ahmad Khan, and Zhilin Chen.

I would like to thank my loved ones for being with me throughout these two years. I thank my mother for supporting me and taking a lot off my shoulders in these two years and my brother been there for me.

Finally, I would like to acknowledge the Natural Sciences and Engineering Research Council (NSERC), Huawei Technologies Canada Co., Ltd, Ontario Graduate Scholarship, Rogers Scholarship, and ECE Fellowship for financially supporting me throughout my MASc. studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

The next-generation cellular network, popularly coined as "5G", needs to be designed in a way that can accommodate for the explosive growth of mobile traffic data. In 2015, global mobile traffic grew by 74 percent as compared to 2014, and will increase by nearly eight-fold, exceeding 30.6 exabytes by 2020 [2]. Recently, heterogeneous network (HetNet) has become a promising technology to accommodate for the exponential growing demand for mobile data [3, 4]. HetNet is a wireless network architecture where access nodes, separated by tiers, are densely deployed to provide high rates to users anywhere in the network. Tiers of a HetNet refer to different types of access nodes and the cell coverage they provide, and typically, they include macro/pico/femto base stations (BSs). Pico BSs, for example, have lower transmit power than macro BSs, and thus pico BSs are typically deployed near macro-cell edges to provide coverage extension to cell-edge users. Thus, HetNets in principle can shorten the distance between transmitters and receivers to provide high rate data to users in the entire network. An illustration of a HetNet is shown in figure [1] 1.1 and shows the deployment topology of the access nodes on different tiers.

[1] http://mwrf.com/test-amp-measurement/lte-aeicic-test-solution-solves-hetnet-interference [5]

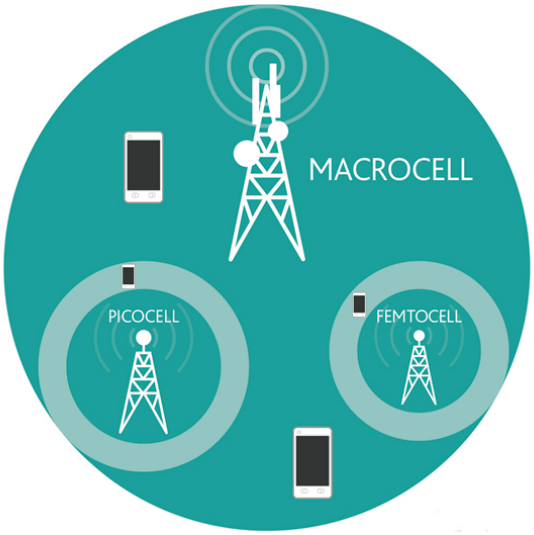


Figure 1.1: Heterogeneous Network Topology

A significant bottleneck in HetNet is the interference created from the dense deployment of BSs. In particular, pico BS and femto BS users suffer in rate because of the high interference from nearby macro BSs. It has been noted that even with targeted deployment of pico BSs in high-traffic zones and hotspots, most users will still detect the strongest downlink signal from macro BSs [6]. This causes the under-utilisation of lower-tier BSs and overloading of high-tier BSs as users associate with the BS providing the strongest downlink signal. Furthermore, as traffic loads of macro BSs increase, so would the BSs' transmit powers, which would produce even more interference to the surrounding low-tier BSs [7]. This creates a positive feedback loop that severely undermines the effectiveness of low-tier BSs. Thus, a crucial design decision in HetNet is to raise the effectiveness of network heterogeneity.

As 5G aims to have 1000-fold capacity [1], a design paradigm shift in recent years that is drawing popularity is the idea of full frequency reuse (FFR) [8]. Under FFR, mobile networks have cell re-use factor of 1, meaning that each cell shares the entire spectrum for radio transmission. The increased bandwidth results in increased rates, but deteriorates rates of cell-edge users. Dense deployment of low-tier access nodes near the cell-edge in a HetNet is a promising solution to increase cell-edge rates on two aspects. First,

as aforementioned, the low-tier access points is closer to cell-edge users than the macro BSs. Second, the transmit powers of these pico BSs is low, which reduces interference to other pico BSs nearby. However, proper management of BS powers and user association is still needed to encourage users to be served on the lower tiers of the network while not suffering high intra-cell and inter-cell interference from nearby macro BSs transmitting on the same spectrum.

Apart from increased rates for cell-edge users, another benefit of a properly managed HetNet is the reduction of network consumption power. Base stations in a cellular network account for 80% of the total network energy consumption [9, 10], with a modern macro BS (e.g. a typical UMTS BS) typically consuming 800-1500W [11]. Offloading traffic to low-tier BSs therefore has the added benefit of reducing network consumption power by turning off macro BSs, in addition to improved rates of cell-edge users due to reduced macro BS interference.

Therefore, there is significant interest on interference reduction as well as traffic offloading from the macro-tier to lower tiers in HetNet. Through appropriately associating users to the BSs on the right tier, and controlling which BSs are operating with how much power, not only will all users enjoy high rates due to the proximity to a BS with little interference, but the macro-tier will also experience less traffic load with potential energy savings if the macro BS is not serving any users and thus can be shut off entirely.

1.2 Related Works

Enhancing the network capacity of HetNets has seen extensive work in literature, where the main goal is to off-load traffic data from macro BSs to lower tier access points. Traditional association of users to BSs providing the maximum signal-to-interference-and-noise ratio (SINR), also known as max-SINR association, as mentioned in [6, 7], under-utilises the lower tiers of the HetNet and often overloads the macro-tier BSs. A

common idea of pushing users to low-tier BSs is cell range expansion (CRE), which is considered in [6, 12–15]. Here a biasing factor is applied to the SINR of low-tier BSs to encourage more users to associate with them instead of macro BSs. Such idea is simple to implement, and robust against location variations of BSs and users [6]. However, a significant drawback of such method is that users associated with low-tier BSs still suffer from high interference of nearby macro BSs [14]. A variation on the CRE idea is to explicitly force users that are associated with high-tier BSs to re-associate with low-tier BSs. There are many metrics to evaluate the "best" low-tier BS for the macro-user to re-associate. For example, in [16], the authors considered a green HetNet where some pico BSs have stored energy off-line from battery supplies, and selects macro-users closest to a pico-BS with most off-line energy to re-associate. Game and auction theory has also been widely applied to encourage users to associate with lower tier BSs for effectiveness in mobile traffic loading in HetNets. In [17], the authors considered the association of users to HetNet access points as an "evolutionary game". Here this is a more user-centric approach to user association where users select BS based on the choices of other users. If many other users are associating with a particular nearby BS, then the current user will make a decision taking into account the loading of that BS as well as the rate achieved on other BSs. In [18], the BSs adjust the transmit powers based on how much each users "bid" to be served by that BS.

An interesting technique to reduce interference in HetNet is to first balance the loading of users on different tiers, then run power control on the BSs to reduce the network interference. In [6, 7], the authors have considered using "cost-benefit" analysis, where the cost represents BS loading, to determine the proper loading of each BS, and in turn encouraging some users to move away from the heavily-loaded macro BSs that would have resulted from max-SINR association. Ye, et al. [6] proposed a distributed user association algorithm based on the primal-dual decomposition of a centralized network utility maximization problem. Shen, et al. [7] developed a distributed pricing-based

association algorithm that is based on dual coordinate descent (DCD) method applied on the dual of the original utility maximization problem. In both cases, the dual of the utility maximization problem is the loading of each BS. For fixed association, Ye, et al. showed that proportional fairness scheduling is the optimal scheduling scheme. As a result, the dual of the primal utility maximization problem is the loading of the BS. The idea is to determine the "price" of each BS based on its loading, and users then associate with BSs based on the cost-benefit analysis. The cost is how heavily-loaded the BS is, and the benefit is the amount of instantaneous rate it can provide to the user. This is effective in pushing certain users to lower tiers of the HetNet when macro BSs' loads become too high. However, the cost of BS loading is not well-defined when BSs have zero loads. In this case, it is not clear what the optimal price for the loading of this BS is. In [6, 7], authors considered the cases when there are many more users than BSs in the network, such that the optimal solution sees each BS having to serve at least one user. However, a tractable solution when the number of users in the network is few is unclear from existing literature.

A prevailing method of controlling interference in HetNet and promote high rates for cell-edge users is the employment of soft frequency reuse (SFR). In recent years, full frequency reuse (FRone) has been proposed to increase the bandwidth for radio transmission in mobile networks, thereby increasing the capacity. In order to reduce the high interference experienced by cell-edge users from FRone, SFR has been a popular solution [19–24]. The idea is to partition the cell into cell-centre and cell-edge regions, and allocate a portion of the spectrum for cell-edge users that is distinct from cell-centre users. Cell-edge users across the cells have a reuse factor of greater than 1. For the cell-centre users, apart from the universal portion of the spectrum allocated to them, they can also access cell-edge users' bands from neighbouring cells, albeit at lower transmit powers to reduce interference. In SFR for HetNet, pico BSs are allocated a portion of the spectrum for cell-edge users with some reuse factor greater than 1, and macro BSs can

access the spectrum of neightbouring pico BSs [24]. The state-of-art in user association and BS power control under the SFR scheme remains to be the heuristic greedy algorithm, both in downlink [21] as well as uplink [22]. A main challenge in SFR is to balance the trade-off between improvement in rate and cell coverage for cell-edge users and network capacity due to reduced bandwidth from spectrum division [19]. In [20], the authors provided a heuristic simulation capturing the performance of SFR scheme under various network parameters. In [19], the authors provided an analytical framework in evaluating the trade-off between coverage and network capacity.

Iterative updating of association and power control remains popular in solving network utility maximization problems. The main drawback is that BSs that were initially associated with users would not be able to turn off, and similarly BSs that were off would not be able to turn on. Shutting off BSs pre-maturely could potentially impact the network capacity. In order to address this issue, several joint optimization methods have been proposed. One such method is the minimization of the network power subject to rate constraints [11, 25] or network coverage [26]. Such methods indirectly reduces the interference of the HetNet by turning off high-powered BSs. Other methods include directly maximizing the sum rate of the network under the multiple-input-multiple-output (MIMO) setting. Such method involves maximizing the weighted sum rate of the network as function of the beamformers, while using the norm of beamformers as a regulator term [4, 25, 27, 28], or constraining the L0-norm of the beamformers to obtain a sparse beamforming solution [29]. Another interesting idea is to balance the tradeoff between network power and user rates. Such idea is explored in [30], where authors considered the maximization of sum network rate subject to user SINR constraints and with trade-off to minimize network power, using a technique called Benders' Decomposition. These ideas are applied to maximize the (weighted) sum rate of the network. However, a more appropriate network performance model invoking log-utility has not been analysed, to the best knowledge of the author, in a joint user association and BS power control optimization.

Log-utility objective promotes cell-edge users to attain higher rates by associating with low-tier access nodes and controlling the power of macro-tier access nodes, which raises network heterogeneity. Furthermore, in order to fully account for the intricate balance between user association and BS power control, these two factors must be optimized jointly. This is a difficult problem in general since all the known problem formulations are non-convex [30]. Currently, the state-of-art is a set of algorithms based on greedy BS selection as a way to push users to BSs of lower tiers and at the same time adjusting for BS powers. Heurestic power level adjustments [31] and iterative BS turn off [9, 11, 27] are some of the prevailing methods to jointly optimize for user association and BS power control. These methods involve defining a heuristic score based network utility [9], or energy usage [11, 27], and iterative activates (or deactivates) BSs that best optimizes for the objective until no such optimizations can be made.

1.3 Thesis Overview

This thesis presents a formulation of network utility maximization problem with possible BS consumption trade-off and investigates the patterns in user association and BS power that result from the optimization. Detailed models on network utility as well as BS consumption powers are presented in chapter 2. User association and BS power control are jointly optimized with a heuristic-based hybrid gradient-"pseudo Newton (pN)" projection method, described in detail in chapter 3. Throughout the simulation, we consider a two-tiered HetNet arranged in a 19-cell hexagonal wraparound layout having three sectors per cell, with three pico BSs per macro BS and one macro BS per sector.

Under the model of a regular network, with pico BSs placed at known locations near a macro-cell edge for simplicity, we simulate the maximization of network utility with potential power consumption trade-off using the proposed algorithm and investigate the following:

1. Can cell-edge users in a HetNet experience higher rates with the division of downlink spectrum into multiple sub-bands, and how does this compare with existing SFR scheme?

2. In a network where there is a total BS consumption power limit, can the HetNet provide the same network utility with tighter consumption power limits? In other words, does the interference reduction of some BSs turning off compensate for the reduced number of active BSs?

3. How to select which BSs to remain active under a total network consumption limit and how to further associate with these BSs?

For the investigation of multiple sub-band in the network, we focus exclusively on the maximization of network log utility, subject to per-BS power spectral density (PSD) and user association constraints. For the investigation of BS turn-off and its impact on network utility, we introduce the power consumption trade-off and vary the weight of the trade-off to analyse network performance. The results, based on the proposed heuristic algorithm and the simulation model, yielded the following:

1. Having the bandwidth divided into multiple sub-bands improve the rates for cell-edge users as compared to that of single-band, and this effect is more pronounced as the number of users in the network increases. Multi-band operation with four sub-bands yielded 16.4%, 22.1%, and 48.6% rate improvements for the 10th percentile users, compared to the single-band operation for 57, 114, and 228 users in the network, respectively. Furthermore, the multi-band set-up allowed flexibility for users to potentially occupy the whole spectrum, where needed, to increase network capacity. In SFR, under the simulation model, each BS can at best access a fraction of the total downlink bandwidth, hence negatively affecting network capacity.

2. Reducing network consumption power accordingly decreases network utility, as turning BSs off do not compensate for the reduced number of active BSs available

to serve all the users. Results showed decreasing trend in network sum rate when the network consumption limit is reduced. This suggests that turning off BSs and pushing the affected users to other BSs increases the loading for those remaining active BSs, and that the deterioration of rates due to the increased loading is not properly compensated by the reduced interference from BS turn-off.

3. Through finding the appropriate trade-off between network utility and network power consumption, a non-greedy heuristic for finding the appropriate set of active BSs is possible. This is done through direct joint optimization of user association and BS power control. For context, the utility-power trade-off curve for the proposed algorithm is compared with the trade-off curve from greedy heuristics as motivated by those presented in section 1.2.

1.4 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 introduces the problem background. Chapter 3 discusses the problem formulation and proposes a hybrid gradient-pN projection algorithm to solve the problem. Chapter 4 presents simulation results of the proposed algorithm and discusses the implications of the results. Finally, Chapter 5 concludes the thesis.

1.5 Notations

Bold-faced lower-case letters represent vectors (e.g. $\mathbf{p}$). Alternatively, regular letters with sub/super scripts enclosed in square brackets also represent vectors, but in the form that illustrates the indexing of each element of the vector. For example, $[x^f_{k,b}]$ represents the vector $\mathbf{x}$, where each element of the vector $\mathbf{x}$ is indexed by three variables: k, b, and f. Regular lower-case letters without enclosing square brackets represent scalars, indexed

by the corresponding sub/superscripts. The L_p norm of a vector is denoted via $|| \bullet ||_p$. Furthermore, $(\bullet)^T$ represents the transpose of the vector. $\mathbb{R}^N$ and $\mathbb{C}^N$ represent sets of real and complex numbers in N dimensions, respectively.

Chapter 2

System Background

Consider a downlink, single-input-single-output (SISO), HetNet of K users and B BSs. The downlink transmission channel is considered to be a flat-fading, time-invariant interference channel. In order to investigate the effects of frequency re-use on cell-edge users in the HetNet, we allow the downlink bandwidth, W, to be divided into F sub-bands of equal bandwidth. Over each sub-band, BSs can serve multiple users. Note that it is permissible to set $F = 1$, in which case no such division exists and the network operates under reuse factor of 1 (universal frequency reuse). The ability to consider different values of F allows for comparison of performance between the multi-band ($F > 1$) and the single-band ($F = 1$) system, and helps to answer the question of whether HetNets provide higher rates to cell-edge users from dividing the entire downlink band into multiple sub-bands than without the sub-bands division. Denote k, b, and f as indices over the set of K users, B BSs, and F sub-bands of transmission, respectively.

The channel gain between user k and BS b over sub-band f is denoted as $g_{k,b}^{f} \triangleq |h_{k,b}^{f}|^2$. Here $h_{k,b}^{f} \in \mathbb{C}$ is the corresponding complex channel coefficient, and is determined from both large-scale fading and small-scale fading. Large-scale fading is caused by the attenuation of transmitted signals as they propagate from BSs to users, and is dependent on the distance between the transmitter and the receiver, as well as shadowing effects.

The collective large-scale fading effect is called *path loss* [32]. Small-scale fading, on the other hand, is dependent on the specific signals transmitted to the receiver. The different arrival times and angles of signal components causes various constructive/destructive interference and thus attenuates the overall transmitted signal accordingly. The overall effect of the two types of fading yields the following equation for $h^f_{k,b}$:

$$h^f_{k,b} = \sqrt{\xi_{k,b}}\phi^f_{k,b} \tag{2.1}$$

In an FDMA system, on each sub-band f, different signals are transmitted, which gives rise to a distinct small-scale fading coefficient $\phi^f_{k,b}$ (assuming the channel is flat-fading). In this thesis, we are assuming that independent signals are transmitted on each sub-band, and without loss of generality, the signals have unit power. Thus, assuming the signals are modulated using quadrature amplitude modulation (QAM), the transmitted signal can be modelled as independent, circularly symmetric complex Gaussian random variables [33], with mean $\mathbf{0}$ and covariance matrix $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$

Large-scale fading is denoted as $\xi_{k,b}$, and is only a function of the distance and the environment between the transmitter and the receiver. We are assuming that such information is available and readily-accessed from a central server. In practice, there is extensive work in literature on estimating channel based on long-term statistics, and many well-known path loss models have been standardized [34] and used throughout the field of wireless communications.

User association and BS power control occur at much larger time-scales than those for individual signal transmissions, and as such they cannot account for fast-fading. As a result, the channel gains used for these network operations are evaluated using *long-term* channel statistics. Thus, the gain between user k and BS b on a particular sub-band using the long-term channel statistics is the same for all other sub-bands, and is only a function of the large-scale fading. Thus, during user association and BS power control,

$g^f_{k,b} = \xi_{k,b} \ \ \forall f$. However, in simulations, actual user achievable rates are still evaluated with fast-fading over many samples of signal transmissions to obtain an average result, which is not necessarily equivalent to the result obtained ignoring small-scale fading.

2.1 Network Utility Model

We denote $r^f_{k,b}$ as the instantaneous rate that user k can achieve with BS b on sub-band f. Denote $\mathbf{p} = [p^1_1, \cdots, p^F_1, \cdots, p^f_b, \cdots, p^F_B]^T$ as the stacked vector of $[p^f_b]$, where p^f_b denotes the downlink transmission PSD of BS b on frequency band f. Let σ^2 be the variance of the receiver noise, and $g^f_{k_b}$ be the long-term channel gain from BS b to user k on sub-band f. Using the Shannon-Hartley equation [33] applied to the downlink SISO network case, we have the following expression for $r^f_{k,b}$:

$$r^f_{k,b} \triangleq \frac{W}{F} \log_2(1 + \gamma^f_{k,b}) \tag{2.2}$$

$$\gamma^f_{k,b} \triangleq \frac{g^f_{k,b} p^f_b}{\sigma^2 + \sum_{l \neq b} g^f_{k,l} p^f_l} \tag{2.3}$$

is the receiver SINR expression for user k from BS b on sub-band f.

Now, the long-term user rates is the average of many user rates achieved after scheduling on the BSs that serve the user. Denote a resource block (RB) as a time slot over a sub-band. In a TDMA system, on each RB for each BS, only one user can occupy the RB to receive the downlink signal. However, over time the rate user k obtains from BS b on sub-band f is the average of the fraction of time slots on sub-band f that user was served by BS b. This motivates the definition of a "resource-sharing" variable, $x^f_{k,b} \in (0, 1)$ that describes the *fraction* of time user k is served by BS b on sub-band f. In our model, we allow multiple users to be served by the same BS on the same sub-band, over a fraction of the total allotted time resource on that sub-band for that BS. Similarly, in an FDMA

system, $x_{k,b}^f$ can also be considered as the "resource-sharing" variable in denoting the fraction of spectrum BS b is allocating for user k on sub-band f.

Denote $\mathbf{x} = [x_{1,1}^1, \cdots, x_{1,1}^F, \cdots, x_{1_B}^F, \cdots, x_{k,b}^f, \cdots, x_{K,B}^F]^T$ as the stacked vector of the resource-sharing variables $[x_{k,b}^f]$ (note that if $x_{k,b}^f = 0$, then it implies that user k is not served by BS b on slot f). The long-term rate user k achieves with BS b over all sub-bands is:

$$s_{k,b} = \sum_{f=1}^{F} x_{k,b}^f r_{k,b}^f \tag{2.4}$$

It follows that the total user rate for user k attained over all frequency bands and by all BSs is thus:

$$R_k \triangleq \sum_{b=1}^{B} s_{k,b} \tag{2.5}$$

to which the network utility comes about as $\sum_{k=1}^{K} \log(R_k)$. The choice of log-utility comes about in the HetNet as an alpha-fairness criterion with $\alpha = 1$, and is used to promote fairness to users with lower rates in the network. This is particularly important in HetNet, as cell-edge users are encouraged to take advantage the deployment of low-tier access points to attain higher rates.

2.2 Base Station Power Model

In order to examine the trade-off between network log utility and total BS consumption power, we present a tractable model describing BS power consumption. As noted in [9,27, 35, 36], the BS power consumption model involves two parts that are closely dependent of each other. The first part is the total BS *transmit power* of BS. The second part represents the BS *on-power*. The on-power is the amount of constant power a BS uses when it is transmitting signal to users on any of its RBs in the downlink. For example,

in [27], the BS power consumption model is as follows:

$$\begin{cases} \eta P_t + P_a, & \text{if } \ P_t > 0 \\ P_s, & \text{otherwise} \end{cases} \tag{2.6}$$

Here, η represents power amplifier efficiency [27]. As the BS transmits power $P_t > 0$, it incurs an on-power of P_a in the active mode. When $P_t = 0$, it incurs an on-power in the sleep mode, P_s.

Without loss of generality, we can consider a simplified power model, as shown in (2.7), such that the on-power is considered 0 when the BS is not transmitting on any of its RBs, and incurs a fixed on-power otherwise. Such simplification is advantageous in that the power model can now be formulated in terms of norms. Specifically, L0-norm is a special norm that is introduced to describe whether a vector is a zero vector or not. For some $\mathbf{x} \in \mathbb{R}^N$, $||\mathbf{x}||_0$ is an integer that represents the number of *non-zero* elements in $\mathbf{x}$. There are many techniques in literature that deal with the non-differentiability of the L0-norm, such as in [37], and they have been applied to solve for similar network problems such as in [29].

$$Q_b(\mathbf{p}_b) \triangleq \sum_{f=1}^{F} p_b^f + \psi_b ||\mathbf{p}_b||_0 \tag{2.7}$$

Here, $Q_b(\mathbf{p}_b)$ represents the total power consumed by BS b. The variable $\mathbf{p}_b$ is the stacked vector of $[p_b^f]$ for some b. If BS b is transmitting at least some power on at least one of the F sub-bands, then it incurs a constant on-power of ψ_b. Otherwise, it is zero. This is reflected by the second term of (2.7) with the use of the L0-norm.

The total network consumption power is defined accordingly in (2.8) as the sum of the total consumption powers of all the BSs in the network.

$$\Theta \triangleq \sum_{b=1}^{B} Q_b \tag{2.8}$$

Chapter 3

Problem Formulation

Here, the thesis presents the user association and BS power control problem for the downlink, SISO HetNet with the network utility and BS power consumption models describe in chapter 2. The problem is formulated as (3.1). This problem formulation allows us to investigate the various aspects of HetNet operations summarized in the thesis overview in section 1.3.

$$\begin{aligned}
\underset{\mathbf{x},\mathbf{p}}{\text{maximize}} \quad & \sum_{k=1}^{K} \log(R_k) - \rho \sum_{b=1}^{B} Q_b(\mathbf{p}_b) && \text{(3.1a)} \\
\text{subject to} \quad & 0 \leq p_b^f \leq \bar{p}_b^f \quad \forall b,\ f && \text{(3.1b)} \\
& \sum_{k=1}^{K} x_{k,b}^f \leq 1 \quad \forall b,\ f && \text{(3.1c)} \\
& x_{k,b}^f \geq 0 \quad \forall k,\ b,\ f && \text{(3.1d)}
\end{aligned}$$

The objective function, (3.1a), represents a trade-off between network log-utility in the first term and the total BS power consumption in the second term. The parameter ρ controls the amount of weight applied to each of the trade-off factors. Note that when $\rho = 0$, the formulated problem reduces to a network utility maximization problem. Constraint (3.1b) is the per BS, per sub-band PSD constraint. Constraints (3.1c) and

(3.1d) describe the resource constraint on each BS b for each sub-band f, as the total fraction of time all the users spend on any RB cannot exceed one.

The first aspect of the problem formulation is that users can be potentially served by more than one BS. This idea is motivated in [6], who also proposed to relax the single-BS association association constraints in order to reduce the complexity of the problem. This technique, called "Fractional User Association (FUA)", turns the combinatorial nature of user association constraints into one which is affine through the relaxation. Although it is more difficult to implement multiple-BS association than single-BS association in a practical system [6], FUA offers an upper bound on the performance of user association. The results showed that a by using a rounding technique, whereby each user is assigned to the BS providing the highest rate among all serving BSs for that user, FUA's upper bound is quite tight compared to directly optimizing for single-BS association. Therefore, for the problem formulation in (3.1), we adopt a similar rounding technique. The single-BS association solution is obtained from the solution to problem (3.1) as follows:

Denote $\mathbf{x}^*$ as the optimization solution of problem (3.1) for user association. The rounding technique we use to obtain solution $\mathbf{y}^*$, that is, single-BS association solution, is:

$$y_{k,b}^{f*} = \begin{cases} x_{k,b}^{f*}, & \text{If } b = \arg\max_{l} \sum_f x_{k,l}^{f*} r_{k,l}^f \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

This allows us to compare single-BS and multiple-BS association results with power control.

Furthermore, the proposed problem formulation also uses the resource-sharing variable, $\mathbf{x}$, to represent both user association and scheduling. In a SISO downlink network where the channel is static during user association, proportional fairness scheduling is the optimal resource allocation scheme for fixed association [6]. This motivated problem

formulation of the user association problem where the network utility function is in the form of a ratio of the instantaneous rates to the BS loads [6, 7]. Since in these formulations, BS loading is an optimization variable, solutions often omit the cases where BSs have zero loads, which makes the denominator equal to zero. The proposed problem formulation addresses this issue, as if a BS is idle, then its corresponding association variable, $x_{k,b}^{f} \quad \forall\ k,\ f$ is simply 0. Performance comparisons of the proposed problem formulation and that proposed in [7] is compared for cases when the number of users in the network is around the same as the number of BSs, in section 4.3, to investigate the effect of zero-loaded BSs in the network on network performance.

The proposed problem formulation also allows for comparison between single-band operations and multi-band operations by setting different values of F. For $F = 1$, this is the single-band operation, and for $F > 1$, this is the multi-band operation with F sub-bands in the downlink spectrum.

The trade-off parameter ρ can be changed to reflect various total network power consumption constraints. With this parameter, we can investigate whether we can achieve the same level of network utility with less BS consumption power through plotting the network utility (or sum rate) achieved as solution to problem 3.1 against different consumption limits. Furthermore, we can plot the same curve with some existing state-of-art algorithms using greedy selection to turn off BSs, and compare the performances to determine if the proposed algorithm can achieve higher network utility with the same consumption limit compared to the greedy algorithm.

3.1 Proposed Algorithm

Note that the objective function of problem (3.1) is non-differentiable due to the presence of the L0-norm. However, a popular technique is to approximate the L0-norm by a mixed L1-L2 norm [27]. It is a common way of inducing *group sparsity* on a vector in groups of a

particular sub-dimension. Here we approximate $||\mathbf{p}_b||_0$ by $w_b||\mathbf{p}_b||_2$, i.e. we are promoting group sparsity on $\mathbf{p}$ over groups of f by the mixed L1-L2 norm where the L2-norm is weighted. The choice of the re-weighted L2-norm, combining with the non-negativity of $\mathbf{p}$, makes the power model differentiable and resolves the integer nature of the L0-norm. The mixed L1-L2 norm has seen application in machine learning as the group least-absolute selection and shrinkage operator, and has been used with similiar effects on promoting sparse beamformers in MIMO systems [25].

Applying the mixed norm technique, the following problem is an approximation to the main problem in (3.1).

$$\max_{\mathbf{x},\mathbf{p}} \quad \sum_{k=1}^{K} \log(R_k) - \rho \sum_{b=1}^{B} \left(\sum_{f=1}^{F} p_b^f + \psi_b w_b ||\mathbf{p}_b||_2 \right) \tag{3.3a}$$

$$\text{s.t.} \quad 0 \leq p_b^f \leq \bar{p}_b^f \quad \forall b,\ f \tag{3.3b}$$

$$\sum_{k=1}^{K} x_{k,b}^f \leq 1 \quad \forall b,\ f \tag{3.3c}$$

$$x_{k,b}^f \geq 0 \quad \forall k,\ b,\ f \tag{3.3d}$$

The weights on L2-norm are updated as follows using techniques in [37]:

$$w_b = \frac{1}{||\mathbf{p}_b||_2 + \tau} \tag{3.4}$$

where $\tau > 0$ is a small regularization variable.

Now that the objective function in problem 3.3 is differentiable over the feasible set, we propose a hybrid gradient-pN projection algorithm to derive a local optimum solution jointly in $\mathbf{x}$ and $\mathbf{p}$ for some fixed ρ. Then we perform bi-section search on ρ to determine the optimal trade-off parameter given $\bar{P}$. The idea is to update jointly the optimization variables $(\mathbf{x}, \mathbf{p})$, and in each step, project the updated variables onto the constraint set.

From a projection perspective, this is practical. The variables $\mathbf{x}$ and $\mathbf{p}$ are separated in the constraints and thus can be projected separately. The projection onto the constraint (3.3b) regarding $\mathbf{p}$ is a simple box-type projection, and projection onto constraints (3.3c) and (3.3d) regarding $\mathbf{x}$ is a projection onto the probability simplex, which is a convex problem with many existing techniques proposed, such as in [38].

For the direction of updating the variables, popular methods are gradient descent and Newton's method. Gradient descent algorithm suffer from slowness in convergence of the objective function, while Newton's method requires higher complexity in calculating the inverse of the Hessian of the objective function [39]. Given that there are $(K+1)\times B\times F$ optimization variables in $(\mathbf{x},\mathbf{p})$, both existing algorithms are quite impractical to update the variables. As such, this thesis proposes a hybrid gradient-pN projection algorithm that gives fairly fast convergence rates without requiring high complexity in computing the inverse of the Hessian. Such method is motivated from [7, 40] and heuristic results on pN-projection have shown that the results are close to the Newton's method but with much reduced complexity. We apply the hybrid gradient-pN projection method to update $(\mathbf{x},\mathbf{p})$. The reasoning for the choice is as follows. As analysed later in section 3.2, computing the first derivative of the objective function with respect to $x^f_{k,b}$ scales linearly in both K and F, while the first derivative of the objective function with respect to p^f_b scales quadratically in K. Therefore, the convergence rate bottleneck is in $\mathbf{p}$, and pN-projection is used to invoke faster convergence. In summary, pN-projection is used on updating $\mathbf{p}$, while $\mathbf{x}$ is jointly updated using gradient projection. Actual algorithm showed fairly quick convergence.

The pN-projection method is as follows. Denote $f(\mathbf{x},\mathbf{p})$ as the objective function in (3.3a). The pN update direction with respect to each optimization variable is equal to the ratio of the partial first derivative of f with respect to the variable to the absolute value of the partial second derivative of f with respect to the variable. That is, let y be an element in the optimization variable vector $[\mathbf{x}^T,\mathbf{p}^T]^T$. The update direction for y,

denoted as Δy, using pN method is:

$$\Delta y = \frac{\partial f}{\partial y} \Bigg/ \left| \frac{\partial^2 f}{\partial y^2} \right| \tag{3.5}$$

while the gradient projection update direction is simply the first derivative of f with respect to y.

Equation (3.5) is essentially a simplified Newton's method update step, where instead of having the update direction equal to $-(\nabla^2 f)^{-1}\nabla f$ as per Newton's method, the proposed method computes only the diagonal entries of the Hessian of f with respect to each variable. Due to the non-concavity of objective function (3.3a), in order to ensure an incremental updating direction, we drop the minus sign in front of the Newton's update step and take the absolute value of the diagonal entries of the simplified "Hessian" inverse. Doing so ensures that for each individual variable, the updating direction is the same as the first derivative with respect to that variable, and thus ensures the incremental increase of the objective function [7].

The update direction with respect to $x_{k,b}^f$ is:

$$\Delta x_{k,b}^f = \frac{r_{k,b}^f}{\sum_{b=1}^{B} \sum_{f=1}^{F} x_{k,b}^f r_{k,b}^f} \tag{3.6}$$

The update direction with respect to p_b^f is:

$$\Delta p_b^f = \frac{\partial f}{\partial p_b^f} \Bigg/ \left| \frac{\partial^2 f}{\partial (p_b^f)^2} \right| \tag{3.7}$$

With:

$$\begin{cases} \frac{\partial f}{\partial p_b^f} = \left\{ \sum_{k=1}^{K} \frac{\left(\frac{x_{k,b}^f}{(1+\gamma_{k,b}^f)}\right)\left(\frac{\gamma_{k,b}^f}{p_b^f}\right) - g_{k,b}^f \sum_{l\neq b}\left[\left(\frac{x_{k,l}^f}{(1+\gamma_{k,l}^f)}\right)\left(\frac{(\gamma_{k,l}^f)^2}{g_{k,l}^f p_l^f}\right)\right]}{R_k \log(2)} \right\} - \rho\left(1 + \psi_b w_b \frac{p_b^f}{||\mathbf{p}_b||_2}\right) \\ \frac{\partial^2 f}{\partial (p_b^f)^2} = \left\{ \sum_{k=1}^{K} \frac{AR_k - B}{R_k^2 \log(2)} \right\} - \rho\left(\psi_b w_b \frac{||\mathbf{p}_b||_2^2 - (p_b^f)^2}{||\mathbf{p}_b||_2^3}\right) \end{cases} \tag{3.8}$$

A and B are defined as follows:

$$\begin{cases} A = -\left(\frac{x_{k,b}^f}{(1+\gamma_{k,b}^f)}\frac{\gamma_{k,b}^f}{p_b^f}\right)^2 - (g_{k,b}^f)^2 \sum_{l\neq b}\left(\frac{x_{k,l}^f}{(1+\gamma_{k,l}^f)}\frac{(\gamma_{k,l}^f)^2}{g_{k,l}^f p_l^f}\right)^2 + 2(g_{k,b}^f)^2 \sum_{l\neq b}\left(\frac{(\gamma_{k,l}^f)^3}{(g_{k,l}^f p_l^f)^2}\right)\left(\frac{x_{k,l}^f}{(1+\gamma_{k,l}^f)}\right)^2 \\ B = \left(\frac{\partial f}{\partial p_b^f}\right)^2 \end{cases} \tag{3.9}$$

Where R_k represents the total long term rates for user k as defined in (2.5), and $\gamma_{k,b}^f$ denotes the receiver SINR between user k and BS b over sub-band f, as defined in (2.3).

The proposed algorithm updates each $x_{k,b}^f$ through:

$$x_{k,b}^f(t+1) = \mathbf{P}_{\Omega_{b,f}}\left(x_{k,b}^f(t) + \alpha_{\text{nt}} \Delta x_{k,b}^f\right) \tag{3.10}$$

Where $\mathbf{P}_{\Omega}(.)$ denotes the projection onto the set of probability simplex, Ω. Here, the probability simplex defined by $\Omega_{b,f}$ is simply:

$$\Omega_{b,f} = \begin{cases} \sum_k x_{k,b}^f \leq 1 \\ x_{k,b}^f \geq 0 \quad \forall k \end{cases} \tag{3.11}$$

The proposed algorithm updates all p_b^f through:

$$p_b^f(t+1) = \left[p_b^f(t) + \alpha_{\text{nt}} \Delta p_b^f\right]_0^{\bar{p}_b^f} \tag{3.12}$$

The update step size, α_{nt}, can be determined via backtracking line search [39].

The proposed algorithm to solve the main problem (3.1) is described in alg 1.

Algorithm 1 Joint Optimization of User Association and BS Power Control

INPUT: Channel Gains, Network Power Consumption Limit $\bar{P}$
OUTPUT: Optimal user association and scheduling $\mathbf{x}^*$ and optimal BS downlink transmit power $\mathbf{p}^*$

1: Pick large starting ρ
2: **repeat**
3: Initialize $\mathbf{x}$, $\mathbf{p}$ to feasible values
4: Initialize $\mathbf{w} = \mathbf{1}$
5: **repeat**
6: Solve problem (3.3) for fixed $\mathbf{w} = [w_b]$ and ρ using hybrid gradient-pN projection
7: Update w_b for all b according to (3.4)
8: **until** Maximum iterations reached, or convergence in (3.3a)
9: Compute the total network power consumption Θ using equation (2.8). If $\Theta < \bar{P}$, record down the network utility. Keep track of the best network utility achieved with $\Theta < \bar{P}$
10: Perform bi-section search update on ρ
11: **until** Maximum iterations reached, or network utility cannot be improved further, or $\Theta = \bar{P}$
12: If needed, apply (3.2) to obtain $\mathbf{y}^*$, the single-BS association solution from the multi-BS solution $\mathbf{x}^*$

3.2 Algorithm Complexity

The proposed joint update of $\mathbf{x}$ and $\mathbf{p}$ in algorithm 1 is dependent on three factors: number of users in the network (K), number of BSs (B), and the number of sub-bands (F). In this section we present a time-complexity analysis of the algorithm in terms of K and F, so that we can obtain a *per-BS* complexity in user association and BS power control. This is useful as the algorithm can potentially be distributed over each BS, and in simulations, the number of BSs is fixed. From the proposed algorithm, the main operations are taking the gradient of the objective function and the projection, thus these

two operations affect the time complexity of the algorithm.

Projection with respect to $\mathbf{x}$ for each (b, f) has complexity of $\mathcal{O}(K \log K)$ using the projection algorithm in [38]. Thus, the total complexity for each BS is $\mathcal{O}(F(K \log K))$. Projection with respect to $\mathbf{p}$ is just a box-type constraint with complexity $\mathcal{O}(F)$. Thus, the overall per-BS time complexity of the projection step of the algorithm is $\mathcal{O}(F(K \log K))$.

For the gradient operation, note that the gradient of the main objective function in (3.1a) with respect to $\mathbf{x}$ linearly scales with K and F. The gradient with respect to $x_{k,b}^f$ is simply $\frac{r_{k,b}^f}{R_k}$. The bottleneck is the calculation of R_k for each k. From (2.5), it is easy to see that for each k the calculation of R_k scales linearly with respect to F for each BS. Thus, the overall complexity of taking the gradient with respect to $\mathbf{x}$ is $\mathcal{O}(KF)$. With respect to $\mathbf{p}$, however, is less straightforward. The operation still scales linearly with respect to F, since BS transmit power on a sub-band is independent of other sub-bands. However, the gradient operation scales quadratically in K since for each BS and for each user k, the gradient is a function of both how much the BS contributes to the rate of user k and how much interference it causes to all other users $k' \neq k$. Thus, the overall gradient complexity is $\mathcal{O}(KF + K^2F) = \mathcal{O}(K^2F)$.

Since the gradient operation is the dominant operation, the overall time complexity of the proposed algorithm for each BS is thus $\mathcal{O}(K^2F)$.

Chapter 4

Simulation Results

We consider a two-tiered HetNet containing 19 macro-cells with 3 sectors, with 1 macro BS at the cell center in each sector, and 3 pico BSs in each sector near the cell-edge for a total of 9 pico BSs in the sector. For simplicity, the pico BSs are placed at known locations near the cell edge. The complete set of simulation parameters is described in table 4.1. The cell layout and path loss models are taken from the standard 3GPP documents [34], and the different BS on-powers are taken from [41]. Figure 4.1 shows a sample realization of the layout with 57 users in the network, with one user distributed uniformly in each sector.

In sections 4.1, 4.2, 4.3, and 4.4, we focus exclusively on network utility maximization in order to investigate the effects of the proposed algorithm on user association and BS power control under various settings, such as multi-BS vs single-BS or multi-band vs single-band. Here, ρ is set to be 0 and we denote such case as the network operating under *maximum network consumption limit*, since any number of BSs can be active as needed to maximize the network log-utility. In section 4.5, when BS turn off is introduced, we then vary ρ to investigate network performance under certain network consumption limits.

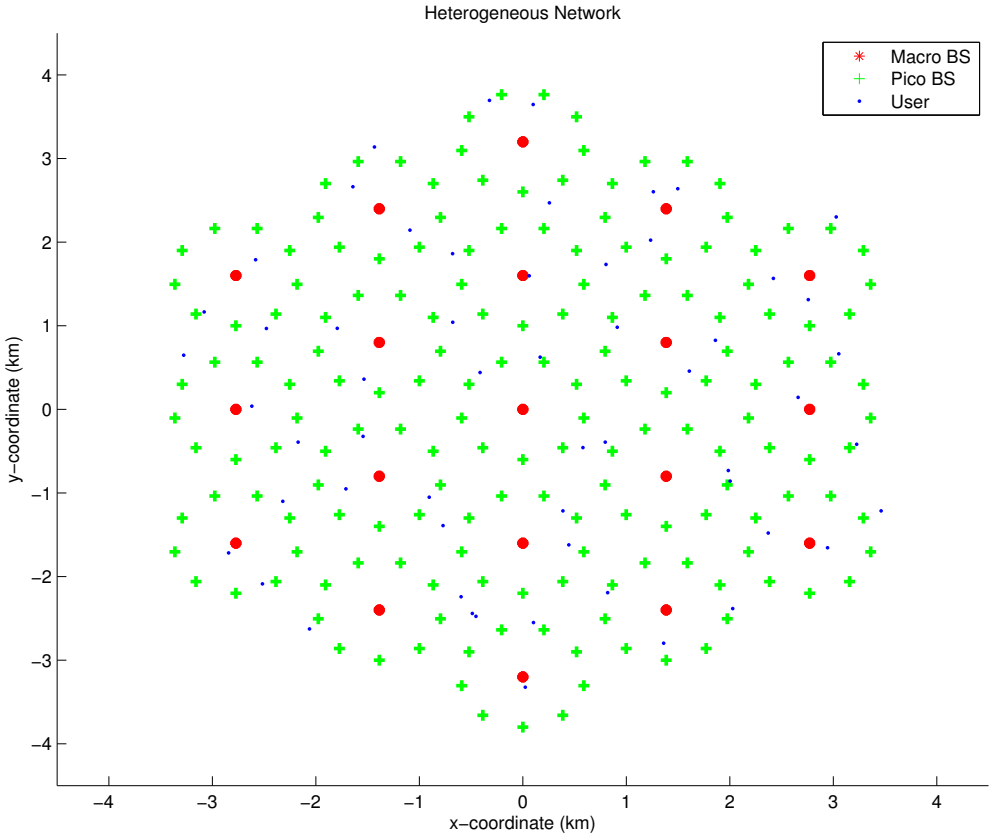


Figure 4.1: Simulation Set-up Example

4.1 Joint vs. Iterative Algorithm

Here we first note that jointly updating user association and BS power control is advantageous over iterative update. The main reason is that a common starting point for the iterative update of $\mathbf{x}$ and $\mathbf{p}$ is to have all the BSs transmitting at maximum power [7]. Such initial point encourages users to be associated with macro BSs and the idle pico BSs are shut off in the subsequent power update. This not only under-utilises pico BSs and reduces cell-edge users' rates, but also overloads macro BSs and thus negatively affects cell-center rates.

Figure 4.2 illustrates the overall rate gain on all percentiles for the joint update over iterative update of user association and BS power, operating under maximum power

Parameter	**Value**
Cell Layout	hexagonal
Tiers	2
Number of Macro BS	57
Number of Pico BS	171
Macro BS Density	1 per cell
Pico BS Density	3 per cell
Sectorization factor	3
Cell Re-Use Factor	1
Macro-Cell Radius	0.8 km
Pico-Cell Radius	0.2 km
Macro BS max transmit PSD	-27 dBm/Hz
Pico BS max transmit PSD	-47 dBm/Hz
Macro BS On-Power	1450 W
Pico BS On-Power	21.32 W
Channel Bandwidth	10 MHz
Noise PSD	-169 dBm/Hz
Noise figure	8 dB
Macro BS Pathloss	$128.1 + 37.6 \log_{10}(d)$ (dB)
Pico BS Pathloss	$140.7 + 36.7 \log_{10}(d)$ (dB)
Shadowing	8 dB
Fast Fading Coefficient	$\sim \mathcal{CN}(0, 1)$

Table 4.1: Simulation Parameters

consumption limit. There is a consistent 0.2 bps/Hz gain for the joint update compared to the iterative update, as expected.

4.2 Single-BS vs Multi-BS

Here we examine the solutions of the joint algorithm to see if users only effectively associate with one BS even though they are allowed to be served by multiple BSs. When we restrict each user to be served by only one BS, we apply the rounding technique discussed in chapter 3 to obtain the single-BS solution. Note that from the proposed problem formulation, it is possible for a user to be served by multiple BSs on the same sub-band. We assume that there is no cooperation among the BSs and the downlink

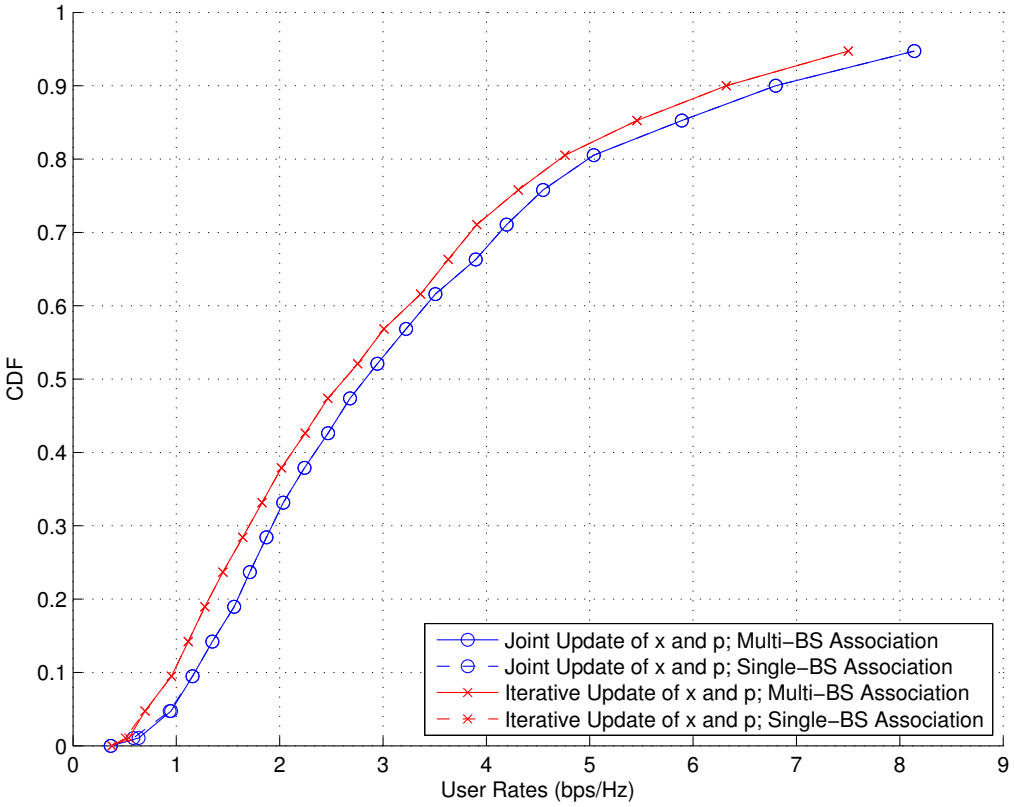


Figure 4.2: User CDF for $K = 57$, $F = 1$

spectrum acts strictly as an interference channel.

As shown in figures 4.3 and 4.4, there is no performance difference between the original problem solution and the solution restricting each user to a single BS. Such result shows that in practice, users should be served by one BS that is providing the highest long-term rates as determined jointly from user association and BS power control. This further shows that if a BS is idle, having to serve a user who is already been served by another BS adds interference, which is not compensated properly from the additional user rates. Also, users served by multiple BSs will likely overload BSs and thus negatively impact network utility. It is interesting to note that there is a small gain in rates for cell-edge users, due to that user opportunistically acquiring some additional rates from nearby BSs that are already serving many users. In such cases, there is no additional interference

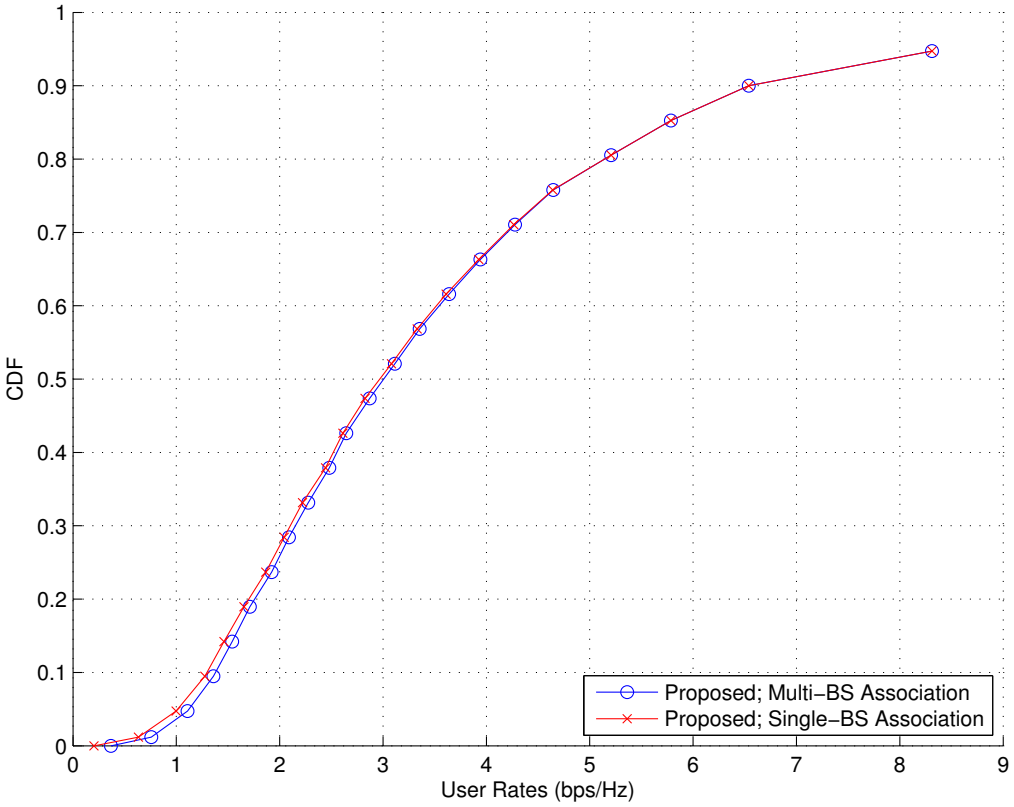


Figure 4.3: User CDF for $K = 57$, $F = 4$

from extra BSs since they are already active, and since those BSs are already serving many users, the rate degradation is not as significant as if the BSs were originally serving just one user. Nonetheless, the rate gain is not significant. The main result here shows that there is a tractable way of solving user association problems without the need to restrict each user explicitly to be served by a single BS, as users tend to be served by a single BS even if opportunities present otherwise. This makes user association constraints affine and non-integer.

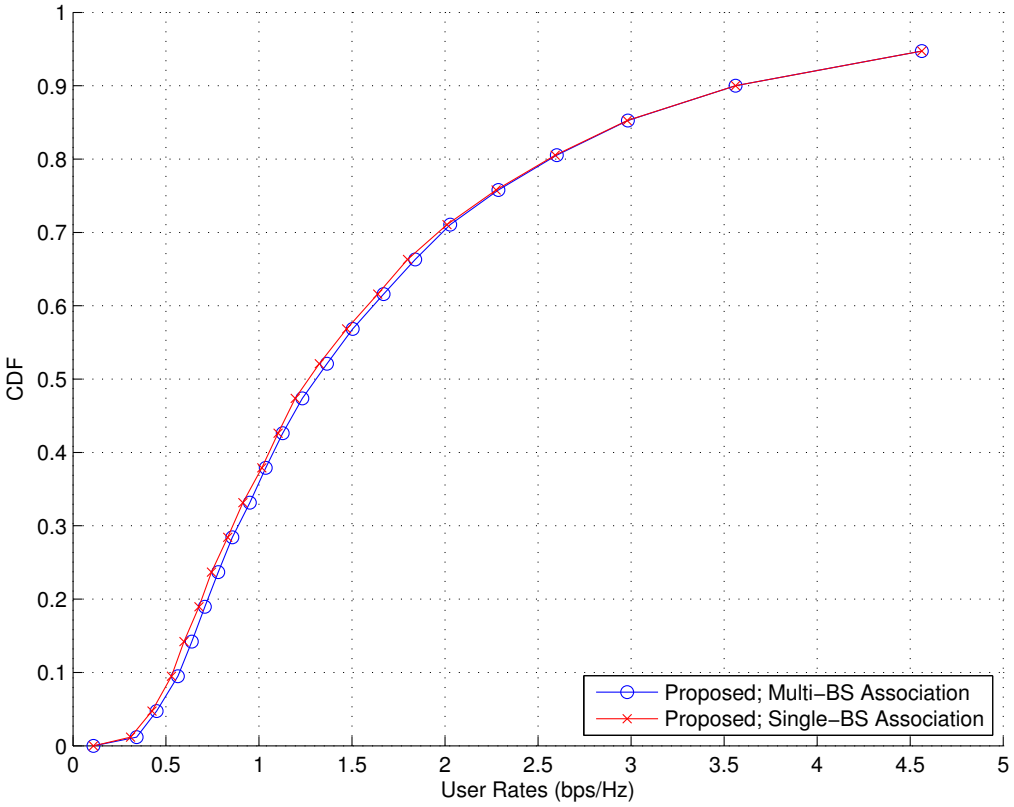


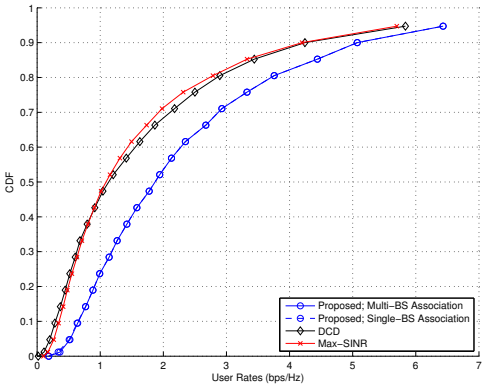
Figure 4.4: User CDF for $K = 228$, $F = 4$

4.3 Single-Band Operations

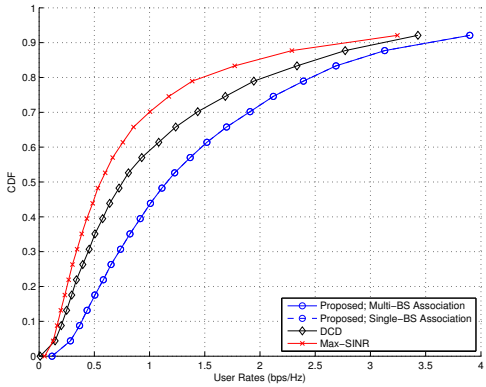
Here we compare the performance of the joint algorithm in alg 1, under the single-band set-up, operating at maximum network power consumption limit, against existing state-of-art user association and BS power control algorithms under the same operating conditions. In order to quantify single-band operations, we solve problem (3.1) setting $F = 1$.

Figure 4.5 shows the user CDF curves of the proposed algorithm 1, along with the dual-coordinate descent (DCD) algorithm with power control in [7] and the max-SINR algorithm with power control for various users in the network. As expected, since the proposed joint algorithm factors in the loading of BSs, 10th percentile users obtained 60% higher rates than max-SINR for each case of number of users in the network. In

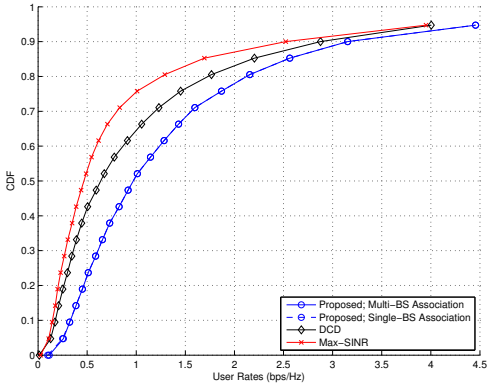
Figure 4.5: User CDF curves under single-band operations



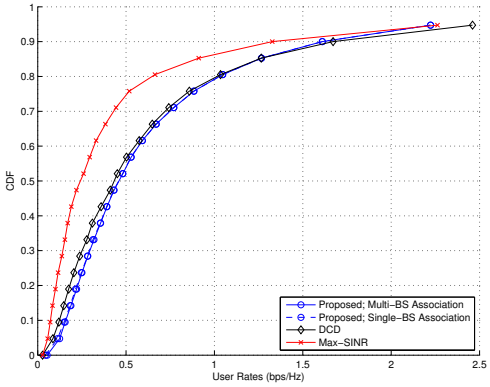
(a) User CDF for $K = 114$, $F = 1$



(b) User CDF for $K = 228$, $F = 1$



(c) User CDF for $K = 285$, $F = 1$



(d) User CDF for $K = 570$, $F = 1$

addition, the proposed algorithm outperforms DCD when the number of users in the network is small. This is due to the fact that some BSs have zero load as the number of users is less or not much greater than number of BSs. As such, the dual update of the loading "cost" for each BS is not appropriately assigned for those idle BSs. The proposed algorithm deals with this issue through the resource-sharing variable $[x^f_{k,b}]$, which simply takes the value of 0 when user k is not served by BS b on sub-band f. This explanation is justified through the DCD algorithm's performance approaching that of the proposed joint algorithm as the number of users in the network increases. When there are 570 users in the network, DCD performance is much closer to the proposed than with the case of 114 users since there are more than twice the number of BSs than users, and so most BSs are serving at least one user.

4.4 Multi-Band vs. Single-Band

In this section, we examine the effect of FDMA on network utility. It is expected that cell-edge users can obtain higher rates from FDMA, since macro BSs can shut off certain bands and allow pico BSs to serve cell-edge users on other bands without interference. In order to quantify FDMA effects, we solve problem (3.1) setting $F > 1$, i.e. we divide the downlink spectrum into multiple sub-bands of equal bandwidth. For the simulation, we investigated the cases for $F = 2$ and $F = 4$.

Furthermore, drawing from Zhuang *et al.* [24]'s model of SFR on HetNet, we add SFR scheme into the comparison as well. Here, we are assuming that the cell-edge spectrum has reuse factor of 3. That is, the entire downlink spectrum is evenly divided into $F = 4$ sub-bands. The macro BSs in all cells have exclusive access to one of the sub-bands. The remaining three sub-bands are allocated to the cell-edge users with each sub-band serving as one frequency reuse channel. Thus, each pico BS in a cell is transmitting on one of the three cell-edge sub-bands. The SFR scheme would reduce inter-cell interference between

pico BSs due to pico BSs in the neighbouring cells transmit on different sub-bands. In order to increase capacity of macro-cell users, macro BSs can transmit on all the cell-edge sub-bands except for the one assigned to pico BSs of the same cell.

Figures 4.6, 4.7, and 4.8 show the user CDF curves for network with 57, 114, and 228 users in the network, respectively, comparing the network rates achieved with different number of sub-bands under the maximum power consumption limit. Note that since the association and power control are done without channel data involving fast-fading and only with long-term statistics, users are not optimized to be served opportunistically on favourable sub-bands. Thus, here the performance gain is purely from the fact that cell-edge users are served on sub-bands with reduced interference, as nearby macro BSs can shut off some sub-bands. In 4.8, we also compare the result with the SFR scheme having reuse factor of 3 for the cell-edge spectrum.

Numerical data on the improvement of cell-edge rates for multi-band over single-band is presented in table 4.2.

Table 4.2: Percent Improvement of 10th Percentile Rates For Multi-band Over $F = 1$ Case

	F = 4	**F = 2**
K = 57	16.4	19.4
K = 114	22.1	32.5
K = 228	48.6	33.8

Fixing the number of sub-bands for FDMA, the results show that the cell-edge user rates improve further with more users in the network. This is due to the fact that when the number of users in the network is large, there are less idle BSs, which makes the impact of interference reduction through active BSs shutting off some of the sub-bands more effective. When most BSs are already idle, i.e. all of their sub-bands are shut-off, users would be more inclined to be served on all sub-bands of a BS since interference from these idle BSs are removed. Furthermore, when the number of users in the network is significantly high, increasing the number of sub-bands in the network would yield a

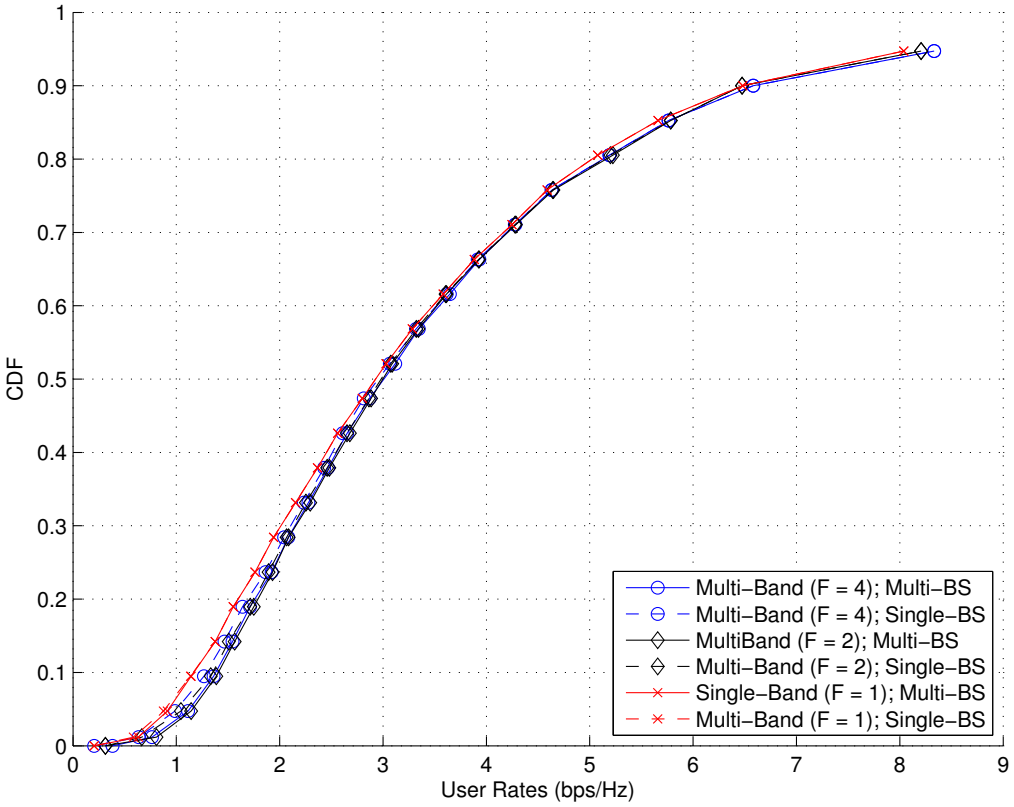


Figure 4.6: User CDF for $K = 57$

greater improvement for cell-edge users due to BSs having more flexibility to serve users on particular bands where nearby BSs are idle on those bands. This is seen as when $K = 228$, having four sub-bands yielded nearly 50% rate improvement, which is significantly higher than 34% provided with only two sub-bands. The trade-off here, however, is the overhead in managing the sub-bands, as the complexity of user association and BS power control increases proportionally to F as discussed in section 3.2. Therefore, it is beneficial to perform FDMA in the network to benefit the cell-edge users. In particular, when the number of users in the network is high, greater number of sub-bands yields higher rates. However, as number of sub-bands increase, computational complexity trade-off arises.

It is interesting to note that under the simulation model, SFR does not have significant improvement of cell-edge rates as compared to single-band. Furthermore, the proposed

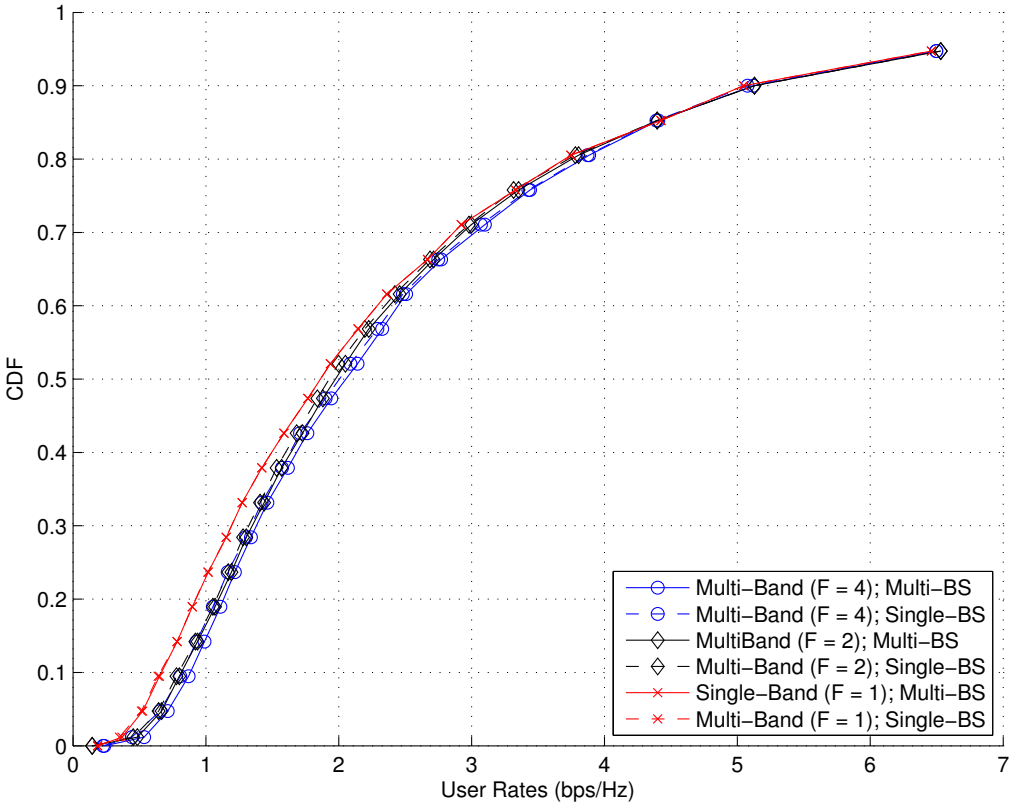


Figure 4.7: User CDF for $K = 114$

algorithm with any number of sub-bands provided higher user rates in higher percentiles than those from SFR. The reasoning is that under the assumption that the bandwidth is evenly allocated to cell-centre and cell-edge users, the reduced available bandwidth negatively impacts network rates. For cell-edge users, the reduced interference from SFR is not compensated by the fact that only 1/4 of the total available bandwidth is assigned to the pico BSs. Similar, for macro BSs, at most 3/4 of the total bandwidth is available for transmission. Under the multi-band set-up, although there is possibility for BSs to shut off certain bands to reduce interference for cell-edge users, it is also possible for BSs to serve users on all sub-bands, where appropriate, to increase network capacity. This flexibility allows for a greater efficiency in bandwidth usage that results in higher user rates. If the bandwidth is allocated in an optimal way, it is possible for cell-edge users

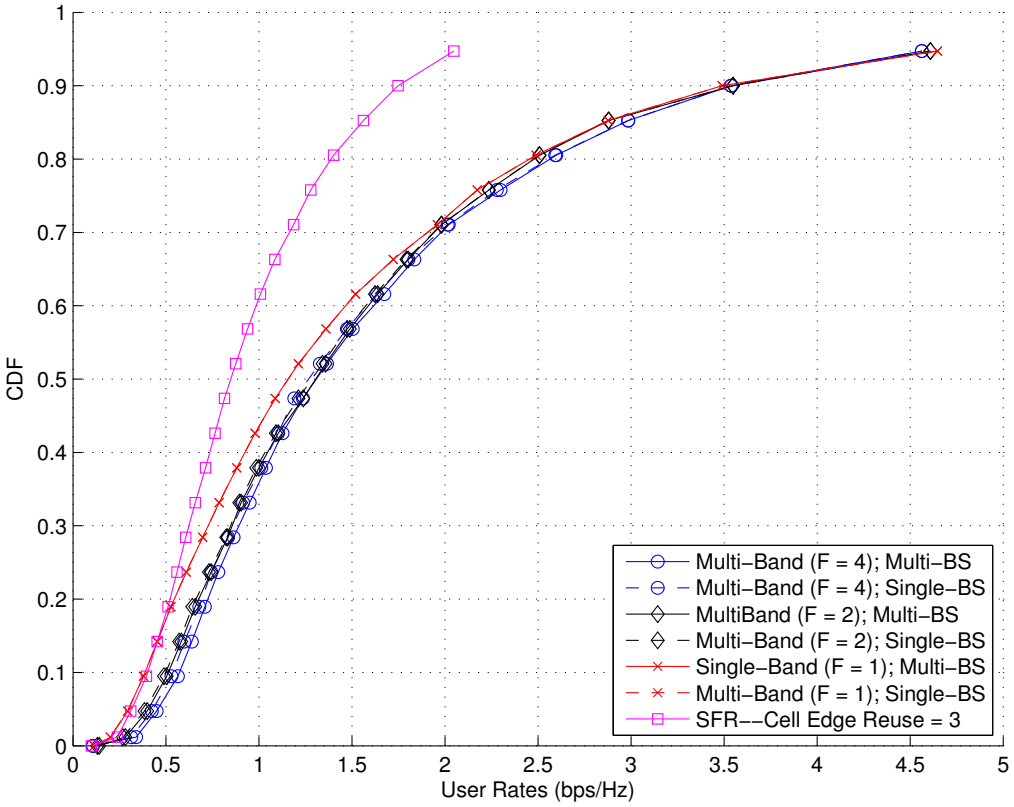


Figure 4.8: User CDF for $K = 228$

to experience reduced interference while maintain high network capacity. This pertains to the trade-off between cell coverage and network capacity and as seen in section 1.2, has been well-studied in literature. This result provides an alternative heuristic analysis of FDMA effects in a HetNet.

4.5 BS Turn Off

In this section, we introduce the network consumption power trade-off by setting a network consumption limit $\bar{P}$, and perform bi-section search on the system parameter ρ to determine the network utility given this limit. We also compared the proposed algorithm with heuristic benchmarks. The benchmark used for turning off BSs is motivated by

state-of-art greedy selection, as such in [9,11,27], such that in each iteration, if the total network consumption power is greater than $\bar{P}$, a least-loaded BS is selected and turned off, and the affected users are re-associated to nearby active BSs.

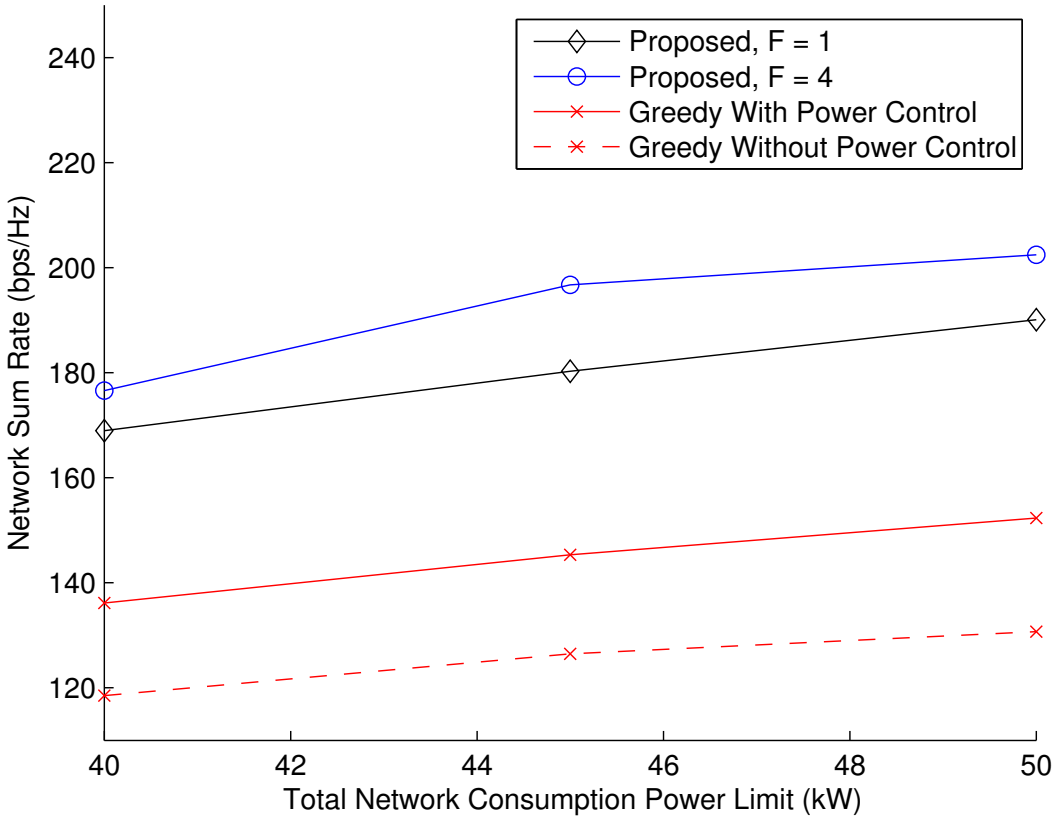


Figure 4.9: Network Sum Rate vs. $\bar{P}$ for $K = 57$

Figure 4.9 shows the trade-off curves of the proposed algorithm along with the greedy benchmark, one with BS power control, and the other simply having BSs either transmitting at max PSD or off (i.e. no power control). Figure 4.10 shows the CDF curves for the various algorithms at $\bar{P} = 50{,}000$ W. The 100% rate gain for 10th percentile users using the proposed algorithm under single-band, as compared to the greedy benchmark with power control at $\bar{P} = 50{,}000$ W further illustrates the fact that prematurely turning off BSs can negatively impact network utility. Furthermore, the presence of multiple sub-bands add more flexibility for users to be served by nearby pico BSs under the power

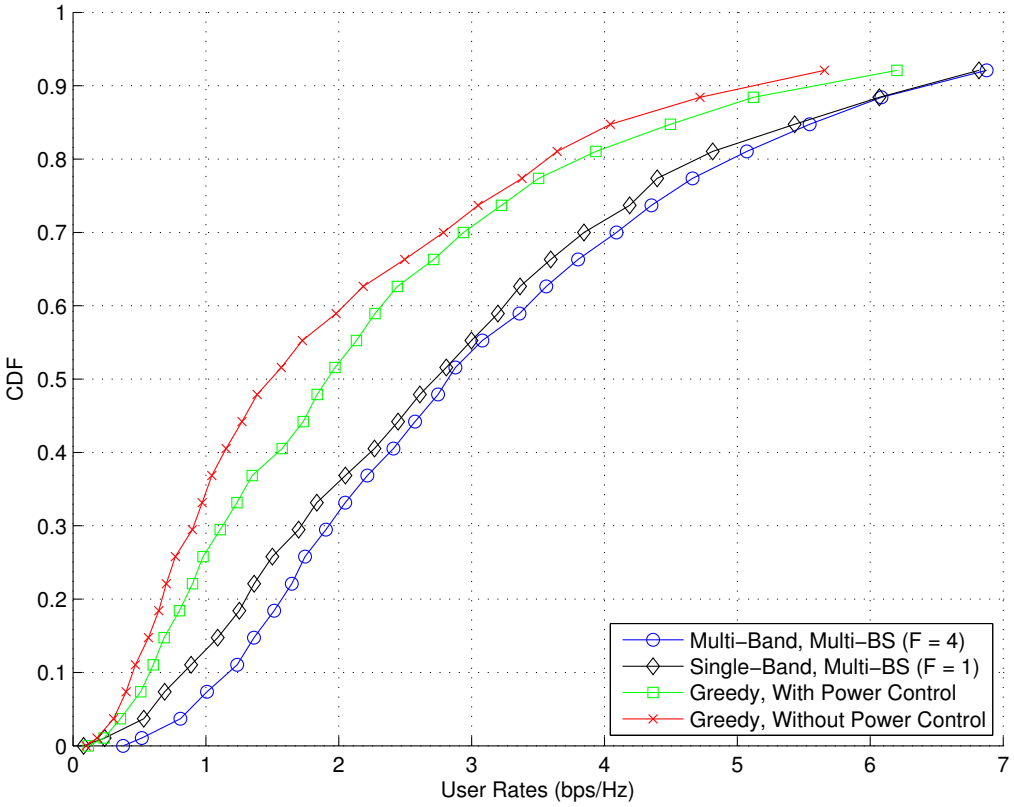


Figure 4.10: User CDF Curves For $\bar{P}$ = 50,000 W, $K = 57$

consumption limit with reduced interference from nearby macro BSs, which attributes to a further 40% rate increase. If many pico BSs are turned off, the pico tier will be severely under-utilised, such that even subsequent heuristics used to further re-associate users to balance loading will not improve the rates. Joint optimization encourages efficient employment of network heterogeneity to improve user rates and balance tier loading.

From 4.9, we also see that it is not possible for the network to expense less power and still maintain the same level of network performance. Although the trade-off curve is concave, the sum rate still decreased as the consumption limit decreased. This shows that by forcing the turn off of some BSs, it increases the loading of the remaining BSs, which results in lower rate for the users that was not compensated by the decrease in interference from the BS turn off.

Chapter 5

Conclusion

In this thesis, we proposed a non-greedy heuristic joint optimization algorithm to maximize the trade-off between network log-utility and network consumption power. When the consideration is only on network maximization, simulation results, based on the regular deployment of access nodes in a two-tiered HetNet, showed that there is no benefit in associating users to multiple BSs due to BS resource constraints. Furthermore, cell-edge user rates are improved when the downlink spectrum is divided into multiple sub-bands due to the ability of BSs to turn off on certain bands to reduce interference and encourage users to moved to lower tiers of HetNet. Furthermore, since each BS can potentially serve users over the entire downlink spectrum under the multi-band setup, this has greater potential to increase network capacity than the SFR scheme considered, as SFR restricts the amount of bandwidth each BS can access to serve its users. In addition, in the single-band setting, the proposed problem formulation accounts for BSs with zero loads and thus outperforms DCD algorithms in [7] and subgradient algorithms in [6] when the number of users in the network is small. When there is a total network consumption limit, the proposed algorithm is a non-greedy heuristic for finding the optimal trade-off between network utility and total consumption power. Results of the propose algorithm showed higher network utility for fixed consumption limit than that from state-of-art

greedy algorithms that were widely discussed in literature. This is due to the fact that these greedy algorithms tend to select BSs that maximize instantaneous user rates and turn off lower tier BSs pre-maturely, and in turn deteriorates rates for cell-edge users who must now be forced to associate with a far macro BS. Joint consideration of user association and BS power control increases the effectiveness of network heterogeneity and therefore increases the network capacity of HetNets. The results of joint optimization with network consumption trade-off also revealed that the network utility cannot be maintained at lower consumption powers by forcing certain BSs off, as the reduced interference does not account for the increased load of remaining active BSs.

Bibliography

[1] Huawei, "5g: A technology vision," Tech. Rep., 2013. [Online]. Available: http://www.huawei.com/5gwhitepaper/

[2] "Cisco visual networking index: Global mobile data traffic forecast update, 2015–2020," Cisco, White Paper 1460442270491203, February 2016.

[3] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, June 2011.

[4] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 226–240, February 2013.

[5] P. Whytock. (2013, September) Lte-aeicic test solution solves hetnet interference. [Online]. Available: http://mwrf.com/test-amp-measurement/lte-aeicic-test-solution-solves-hetnet-interference

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

[7] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1100–1113, June 2014.

[8] H. Sari, S. Sezginer, and E. Vivier, "Full frequency reuse in mobile wimax and lte networks with sectored cells," in *2009 IEEE Mobile WiMAX Symposium (MWS '09)*, July 2009, pp. 42–45.

[9] A. Abbasi and M. Ghaderi, "Energy cost reduction in cellular networks through dynamic base station activation," in *2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2014, pp. 363–371.

[10] F. Chen and X. Qin, "Cooperative virtual cell clustering for green cellular networks," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, May 2012, pp. 1–5.

[11] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1525–1536, September 2011.

[12] D. Lopez-Perez, X. Chu, and İ. Guvenc, "On the expanded region of picocells in heterogeneous networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 281–294, June 2012.

[13] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous lte-a cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1479–1489, December 2010.

[14] D.-H. Sung, J. S. Baras, and C. Zhu, "Coordinated scheduling and power control for downlink cross-tier interference mitigation in heterogeneous cellular networks," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 3809–3813.

[15] K. Okino, T. Nakayama, C. Yamazaki, H. Sato, and Y. Kusano, "Pico cell range expansion with interference mitigation toward lte-advanced heterogeneous networks," in *2011 IEEE International Conference on Communications Workshops (ICC)*, June 2011, pp. 1–5.

[16] Q. Kong and B. Wang, "User association for green heterogeneous cellular networks with hybrid energy supplies," in *2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2014, pp. 1–6.

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

[18] M. Chiang and J. Bell, "Balancing supply and demand of bandwidth in wireless cellular networks: utility maximization over powers and rates," in *2004 INFOCOM. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, March 2004, pp. 2800–2811 vol.4.

[19] T. D. Novlan and J. G. Andrews, "Analytical evaluation of uplink fractional frequency reuse," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 2098–2108, May 2013.

[20] Y. Yu, E. Dutkiewicz, X. Huang, M. Mueck, and G. Fang, "Performance analysis of soft frequency reuse for inter-cell interference coordination in lte networks," in *2010 International Symposium on Communications and Information Technologies (ISCIT)*, Oct 2010, pp. 504–509.

[21] M. Qian, W. Hardjawana, Y. Li, B. Vucetic, X. Yang, and J. Shi, "Adaptive soft frequency reuse scheme for wireless cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 118–131, Jan 2015.

[22] X. Mao, A. Maaref, and K. H. Teo, "Adaptive soft frequency reuse for inter-cell interference coordination in sc-fdma based 3gpp lte uplinks," in *2008 IEEE Global Telecommunications Conference*, Nov 2008, pp. 1–6.

[23] S. Kumar, S. Kalyani, and K. Giridhar, "Optimal design parameters for coverage probability in fractional frequency reuse and soft frequency reuse," *IET Communications*, vol. 9, no. 10, pp. 1324–1331, 2015.

[24] H. Zhuang and T. Ohtsuki, "A model based on poisson point process for analyzing mimo heterogeneous networks utilizing fractional frequency reuse," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6839–6850, Dec 2014.

[25] W. C. Liao, M. Hong, Y. F. Liu, and Z. Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3939–3952, Aug 2014.

[26] C. Y. Chang, W. Liao, H. Y. Hsieh, and D. S. Shiu, "On optimal cell activation for coverage preservation in green cellular networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 11, pp. 2580–2591, Nov 2014.

[27] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud-ran," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[28] W. C. Liao, M. Hong, and Z. Q. Luo, "Base station activation and linear transceiver design for utility maximization in heterogeneous networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 4419–4423.

[29] B. Dai and W. Yu, "Sparse beamforming design for network mimo system with per-base-station backhaul constraints," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2014, pp. 294–298.

[30] L. P. Qian, Y. J. A. Zhang, Y. Wu, and J. Chen, "Joint base station association and power control via benders' decomposition," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1651–1665, April 2013.

[31] P.-H. Chiang, P.-H. Huang, S.-S. Sun, W. Liao, and W.-T. Chen, "Joint power control and user association for traffic offloading in heterogeneous networks," in *2014 IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 4424–4429.

[32] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[33] J. G. Proakis, *Digital communications*, ser. McGraw-Hill series in electrical and computer engineering. New York: McGraw-Hill, 1995. [Online]. Available: http://opac.inria.fr/record=b1089299

[34] 3GPP, "Further advancements for e-utra physical layer aspects," Tech. Rep. TS36.814 v9.0.0.

[35] S. Ping, A. Aijaz, O. Holland, and A. H. Aghvami, "Green cellular access network operation through dynamic spectrum and traffic load management," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 2791–2796.

[36] K. Dufková, M. Bjelica, B. Moon, L. Kencl, and J. Y. L. Boudéc, "Energy savings for cellular network with evaluation of impact on data traffic performance," in *2010 European Wireless Conference (EW)*, April 2010, pp. 916–923.

[37] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008. [Online]. Available: http://dx.doi.org/10.1007/s00041-008-9045-x

[38] W. Wang and M. Á. Carreira-Perpiñán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," *CoRR*, vol. abs/1309.1541, 2013. [Online]. Available: http://arxiv.org/abs/1309.1541

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

[40] W. Yu, T. Kwon, and C. Shin, "Multicell coordination via joint scheduling, beamforming and power spectrum adaptation," in *2011 IEEE INFOCOM*, April 2011, pp. 2570–2578.

[41] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *2010 Future Network Mobile Summit*, June 2010, pp. 1–8.