Resource Management in Next Generation Cellular Networks

by

Yigit Ozcan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Halim Yanikomeroglu Professor, Dept. of Systems and Computer Engineering, Carleton University
Supervisor:	Catherine Rosenberg Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal Member:	Patrick Mitran Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal Member:	Amir Keyvan Khandani Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal-External Member:	Henry Wolkowicz Professor, Dept. of Combinatorics and Optimization, University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Fifth generation of cellular networks brings new challenges to the network operators as new applications create new demands. In this thesis, we will study different topics on cellular networks, explain the challenges of each topic, and propose solutions to tackle these challenges. The topics we consider are: i) uplink scheduling in multi-cell OFDMA networks, ii) downlink scheduling in multi-cell OFDMA networks, iii) full-duplex communications in cellular networks, and iv) cellular networks with intra-cellular traffic.

We begin our study with uplink scheduling in 5G networks as its importance has increased in recent years and the related literature is relatively scarce. Scheduling on the uplink is a challenging task mostly due to power and interference management. In practical scenarios, each cell schedules its own users independently from the other cells. In this case, the interference that is received from the neighboring cells cannot be known since the schedules of the other cells are not known. Therefore, interference has to be estimated in order to estimate the rate of each user. When this estimation is not done properly, it can cause resource losses or under-utilization as we show in this thesis. To avoid this problem, all the cells could be scheduled simultaneously using a cloud radio access network (C-RAN) and hence we can take the exact interference into account while scheduling. Formulating the optimal multi-cell scheduler is straightforward, but it is a very large integer problem that cannot be solved easily and fast. We transform it into a more tractable upper bounding problem and solve it with an iterative algorithm. However, it is still not fast enough to be used in real time. Hence, we focus on improving the existing uplink schedulers by proposing practical solutions for the case when there is no C-RAN and for the case when a C-RAN is present. We also propose a soft frequency reuse (SFR) based scheduler that performs much better than the existing schedulers.

We then perform a similar study for downlink scheduling that carries the majority of the cellular traffic today. While downlink scheduling is easier than the uplink due to simpler interference management, it is still not trivial to achieve performance comparable to the maximum achievable performance using a practical scheduler. The main contribution of this study is to show that a well-tuned SFR-based local scheduler can perform almost as well as the centralized scheduler and hence a centralized scheme for downlink might not be necessary.

We next consider a cellular network where full-duplex communications (FDC) are enabled at the base stations. The coexistence of uplink and downlink transmissions in co-channel cells create new sources of interference that have to be taken into account when studying the performance of FDC. When doing so, traffic asymmetry (TA), i.e., the fact

that the traffic is in general much larger on the downlink than on the uplink, should also be considered. We will show that ignoring TA biases the results in favor of FDC. We compute the performance gain that an FDC-enabled multi-cell OFDMA network has over a regular time division duplex (TDD) system considering all sources of interference and TA by formulating a multi-cell centralized scheduling problem. We use it to analyze the impact of each new source of interference as well as of TA. Our conclusion is that, FDC does not improve performance enough in a multi-cell system to warrant its added complexity in an urban setting when TA and the interference have realistic values. The verdict is slightly better in a rural setting. Furthermore, we show that heterogeneous networks can be a better choice than the homogeneous networks to deploy FDC when it is adjusted well.

We finalize our study with device-to-device (D2D) communications. With the advent of smart phones, there are many new applications that create local (intra-cellular) traffic among the users in the same network. Most work in the literature focuses on the possibility to utilize a direct link between those users to by-pass the base station. Such transmissions are called D2D mode. However, implementing D2D is not easy due to difficult interference management and not knowing the required channel gains. We face a major problem while studying D2D. We need a clear benchmark to evaluate the performance of D2D mode. To this end, we focus on designing a type-aware scheduler (a scheduler that has the information on the type of traffic, which can be downlink, uplink, or intra-cellular) in a case where direct communication between users is not enabled. This scheduler can be seen as the benchmark against D2D mode. We show that performance gain can be obtained by jointly scheduling the uplink and downlink with respect to the case where the scheduler is blind to the types. We show for a homogeneous network that when the traffic types are known to a scheduler, a significant performance gain can be achieved compared to the case where the traffic types are not known. We also analyze heterogeneous networks that consist of macro cells and small cells and show that large performance gain can be obtained by performing type-aware user association jointly with user scheduling.

The main contributions of the thesis are i) to analyze the performance of existing schedulers and see if they perform well enough compared to the maximum achievable performance both on the uplink and the downlink, ii) to propose enhancements or new schedulers when the existing schedulers do not perform well enough, iii) to illustrate when FDC deployment can be useful under which scenarios, and iv) to show the importance of information of traffic types when users with different types of traffic exist.

Acknowledgements

I would like to thank Prof. Catherine Rosenberg for her continuous support both in my academic and personal life not only in Waterloo, but also in Rennes and Barcelona. I would definitely not be able to complete this thesis without her support and supervision. I am very grateful and proud to be supervised by her.

I would also like to thank the committee members of my PhD advisory committee, Prof. Halim Yanikomeroglu, Prof. Patrick Mitran, Prof. Amir Khandani, and Prof. Henry Wolkowicz for their time to review my thesis and their valuable comments that improved the quality of this thesis. I also thank Dr. Fabrice Guillemin for his collaboration on many topics.

My wife, Vicdan, has been my real motivation during my PhD years. She has always been with me, held my hands during my hardest times, and kept me motivated to finish my degree. Feeling her presence is the best thing on earth. Thank you for always being with me. And of course Melissa, my little princess. Whatever I write here will be insufficient to really express my feelings for you. I love you so much. You are the prettiest. I hope I can be a father you are proud of.

I would also like to thank my mother who never stopped loving and supporting me. I feel lucky every day for having a lovely mother like you. Words are never enough to tell how much I love you. Of course, I thank my sister Nazli, who is the funniest girl in the world. We grew up like twins and we were together at every aspect of our lives. Thank you for your never ending support and being the best sister possible.

There are no words that can express how much I miss my father. Every single day is a new challenge without your support and guidance. One of the biggest goals in my life is to become worthy of being your son. Thank you for being my father and everything you had done for me.

My grandfather has been the person who has always encouraged me and led the way in my academic life. I specially thank him for being the best person on earth and also for his full support. I would also like to thank my lovely grandmother, whom I already missed a lot.

I would also like to thank Ugur and Ayse Melen for their support since the first day I arrived to Canada. Without their help, it would be extremely difficult for me to adjust to the new life here. Finally, I would like to thank all my relatives and my friends from Turkey and Canada for their support during my PhD studies.

Dedication

To my family...

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BS	Base Stations
C-RAN	Cloud Radio Access Network
CA	Channel Allocation
CDF	Cumulative Distribution Function
CLA	Coordinated Link Adaptation
CSI	Channel State Information
D2D	Device-to-Device
DL	Downlink
FDC	Full-Duplex Communications
FDD	Frequency Division Duplex
GM	Geometric Mean
HetNets	Heterogeneous Networks
IBI	Inter-BS Interference
IC	Intra-Cellular
ICIC	Inter-Cell Interference Coordination
IoT	Internet of Things
IUI	Inter-User Interference
LBM	Local Benchmark
MAC	Medium Access Control

MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MINLP	Mixed Integer Non-Linear Programming
OD	Orthogonal Deployment
PF	Proportional Fairness
PPP	Poisson Point Process
PRB	Physical Resource Blocks
PSD	Partially Shared Deployment
QoE	Quality of Experience
QoS	Quality of Service
RR	Round Robin
RRM	Radio Resource Management
SC	Self-Interference Cancellation
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCF	Small Cell First
SFR	Soft Frequency Reuse
SI	Self Interference
SINR	Signal to Interference plus Noise Ratio
TA	Traffic Asymmetry
TAS	Type Aware Scheduler
TBS	Type Blind Scheduler
TDD	Time Division Duplex
UA	User Association
UL	Uplink

Chapter 1

Introduction

In this thesis, we will focus on resource management problems in cellular networks for different scenarios. This chapter gives a brief overview of cellular networks, the problems we will consider, and the challenges associated to these problems. We will also explain our contributions and the related messages in the following.

1.1 Overview

A typical cellular network consists of many Base Stations (BS) and users associated to these BSs. A cellular network operates on a licensed frequency band. All transmissions are traditionally from a BS to users or from a user to its BS. The frequency band is divided into smaller units called *subchannels*. Time is also slotted and each slot is called a *subframe*. A set of T subframes is called a frame (typically, a subframe is 1 ms, $T = 10$). Each BS serves its users on the wireless resources (subchannels) allocated to that BS by the operator. Wireless transmissions are usually half-duplex, i.e., the BS either transmits or receives at a given time. A user sees two types of traffic, the uplink traffic that it generates and transmits to its BS, and the downlink traffic that it receives from the BSs.

Next, we will introduce some terminology regarding cellular networks.

Orthogonal Frequency Division Multiple Access (OFDMA)

OFDMA is the physical layer technology used in the forth and fifth generations of cellular networks. Although it will be explained later in Chapter 2, we will briefly describe it here.

OFDMA is made of Physical Resource Blocks (PRB). A PRB consists of one subchannel and one subframe. It is the smallest unit that can be allocated to a user for either a downlink or an uplink transmission. Note that each transmission in a cell creates interference to the neighboring cells. In a traditional cellular network, there is only one transmission per PRB either on the downlink or the uplink. All devices are half duplex, i.e., they cannot transmit and receive on the same PRB. Typically, part of a frame is dedicated for uplink transmissions and while the rest is for the downlink transmissions. This is called Time Division Duplex (TDD). It is also possible to dedicate a set of subchannels to uplink transmissions while the other subchannels are for downlink transmissions. This is called Frequency Division Duplex (FDD). In the following, we will assume that the network operates in TDD mode except when otherwise stated.

There are some important Radio Resource Management (RRM) processes performed by the BSs that critically impact the network performance, namely user scheduling, user association, channel allocation, and interference management. While we will study all of these processes in this thesis, our main focus will be on user scheduling and interference management. Next, we will explain those processes and their related challenges.

User Scheduling and Interference Management

User scheduling [3] is one of the most critical processes in a cellular network. It is responsible for:

- allocating the BS resources, i.e., the cell PRBs, in an efficient and fair manner.
- determining the power used on each PRB by the BS on the downlink and by each user on the uplink. As the power budget is limited for both the users and the BSs, using the power efficiently while managing the interference is very important.
- choosing a Modulation and Coding Scheme (MCS) on each PRB. This depends on the interference and hence the Signal to Interference plus Noise Ratio (SINR) on each PRB.

These three missions have to be completed very fast, e.g., in a few milli-seconds, as user scheduling is a process that is called every frame. In a practical cellular network, each BS performs these processes locally and independently from other BSs. However, with the advent of Cloud Radio Access Network (C-RAN), which will be explained shortly, it might be possible to perform these operations centrally for all the BSs. In the rest of this thesis, we will consider Proportional Fairness (PF) as our fairness criterion. We will describe it in more details in Chapter 3.

Interference management is another process that is crucial to network performance and it is closely related to user scheduling. Each transmission in a cell creates interference at the receivers in the co-channel neighboring cells. This interference depends on the transmitters and the power used. It crucially impacts the system performance since it directly affects the SINR and hence the rate each user gets. Therefore, user scheduling is very important as which user or BS transmits in each cell on a given PRB with which power directly impacts the interference created to other cells. In a practical system, since each BS schedules its own users, it might not be easy for a BS to estimate the interference correctly.

A scheduler relies on computation of SINRs on each PRB to make decisions, which means it needs an estimate of the interference on each PRB. These estimates depend what other BSs are scheduling on these PRBs. So, we have a loop in the sense that the BSs need to know the interference to schedule, and the schedule creates interference. We have also an inter-dependence between BSs.

Channel Allocation (CA) and User Association (UA)

Channel allocation and user association are two other network processes that have significant impact on system performance. User association simply determines which user is associated to which BS. This is very important for load balancing as simple user association schemes, such as those connecting a user to the BS corresponding to the strongest signal received, might sometimes result in poor performance, especially for a heterogeneous network configuration, which will be explained later in this chapter. Although user association is not the main focus of this thesis, we will study it in two chapters.

Channel allocation determines which subchannels are available to which BS. For example, the network might employ a reuse factor r among the BSs, where each BS can only use a subset comprised of $1/r$ of the subchannels. This is very important mainly because it affects the interference and hence the performance of the network. Similar to user association, channel allocation will not be our main focus, but we will study it briefly especially in the context of heterogeneous networks in this thesis.

C-RAN

In a typical cellular network, each BS is responsible for most of the RRM processes, such as scheduling, of its own cell. Hence, there is no coordination among the BSs. However, a recent technology, called C-RAN, enables the centralization and coordination of some processes among a set of BSs.

A C-RAN provides a mean to centralize. However, to use the power of centralization, one might need more information, i.e., more channel gains between the users and the BSs, and this might not be easy. Hence, adding C-RAN brings potential benefits, but new

challenges: 1) the problems are larger as it considers many cells together, 2) there is a need for more information.

The question of which RRM processes to push back to the C-RAN is open for debate. User scheduling is often considered as a strong candidate for this.

Homogeneous Networks and Heterogeneous Networks:

Homogeneous networks are cellular networks, where all the BSs have similar features, such as power budget, antenna gains, and channel models between them and the users, all of which will be explained later in this thesis in detail. We refer to the BSs of a homogeneous networks as macro BSs.

Heterogeneous Networks (HetNets) have recently emerged as a new solution to improve the coverage and throughput performance of the homogeneous networks. It is simply based on deploying low-cost low-powered BSs to decrease the load of the macro BSs. Examples of such low power BSs include pico base stations, relay nodes, and femto base stations. In the following, we will refer to these BSs as *small cells*.

While HetNets bring the opportunity to increase the network performance, they pose new challenges. For example, while a UA policy that associates each user to the BS from which the user receives the strongest signal works quite well for a homogeneous network, it is not the case for a HetNet since macro BSs have higher power and yield better channel gains [4].

In the remaining of the thesis, we will mostly focus on homogeneous networks, but we will extend some of our results to HetNets.

Full-Duplex Communications (FDC)

Unlike the traditional cellular networks that only allow a single transmission on a PRB, FDC allows one downlink and one uplink transmission on a PRB, thanks to the recent advances in self-interference cancellation. A priori, both BSs and users can be FDC-enabled, however it might be more realistic to assume that only the BSs are FDC-enabled. FDC brings additional complexity to the network operation as interference management becomes more complicated. More details on FDC will be given later in this chapter.

Traffic Asymmetry

Today's cellular networks are dominated by downlink traffic [5] even if uplink traffic has also increased recently in volume. Therefore, the ratio of the uplink to downlink traffic volume should be taken into account while operating a network. For example, a TDD system should be parametrized with this in mind.

Traffic Types

As discussed before, a classical cellular traffic has two types of traffic, which are the uplink and downlink traffic. With the advent of recent applications on the smart phones such as photo sharing, the users start exchanging data with users in the same cellular network in close proximity to them. We define this type of traffic as Intra-Cellular (IC). The users might be in the same cell or in adjacent cells. We will study the RRM processes in a network with IC traffic in Chapter 6.

Device-to-Device (D2D) Communications

As the intra-cellular traffic has emerged recently, it introduced the opportunity to improve the network capacity by utilizing direct links between the users. In a traditional network, every transmission occurs between a user and a BS. However, when two users have intra-cellular traffic, since they are geographically close to each other, it might be beneficial to by-pass the BS and enable direct transmission between these users. Such transmissions are called D2D communications. Although it can theoretically improve the performance, it bring many challenges such as interference management, intra-cellular traffic discovery, etc.

1.2 Research Questions and Contributions

In this section, we explain the research questions we try to answer in this thesis together with our contributions and messages. We focus on four main problems, all of which are somehow related to each other.

The main challenge in next generation cellular networks is to meet the ever increasing demand of higher data rates using limited wireless resources. Fig. 1.1 shows the overall mobile data traffic prediction for the next years by Cisco [2]. In order to meet this demand, the RRM processes, which are mentioned above, have to be done in the most efficient way. In the following, we will focus on scheduling in four different scenarios, which are multi-cell uplink scheduling, multi-cell downlink scheduling, a network with full-duplex communications enabled, and a network when there is intra-cellular traffic.

1.2.1 Uplink Scheduling in Multi-Cell OFDMA Networks

While most of the RRM studies in the literature focus on the downlink because it carries the majority of the cellular traffic, the importance of uplink has increased in recent years with the advent of Internet of Things (IoT) and new smart phone applications. It is also pointed

Figure 1.1: Overall Mobile Data Traffic Predictions by Cisco [2]

out in [6] that applications such as Netflix create more and more uplink traffic. Unlike 4G networks, where Single Carrier Frequency Division Multiple Access (SC-FDMA) is used on the uplink, 5G uses OFDMA as the uplink multiple access technology. We conduct a detailed study on uplink scheduling in multi-cell OFDMA networks since it will be the new frontier in 5G networks. As discussed before, interference management is critical for scheduling, hence the scheduler needs to either compute or estimate the interference so that it can compute/estimate the SINR and hence chooses an appropriate MCS. However, the interferers are other users transmitting in other cells and since the neighboring cells' schedules are not known to a BS when it computes its own schedule, the interference has to be estimated and this is not easy. The BS might not be able to decode some PRBs if the interference is under-estimated yielding PRB losses. Some of the PRBs might not be fully utilized due to an over-estimation of the interference (i.e., a higher rate could have been obtained with a different MCS).

Our research questions are: How do existing local schedulers perform compared to the maximum achievable performance? How can we improve the performance of practical schedulers?

Since a BS cannot know other cells' schedules beforehand and interference is directly linked to these schedules, the best performance can be obtained only when all cells are scheduled simultaneously in a centralized fashion as exact interference can be taken into account.

Our contributions on the uplink scheduling can be summarized as follows:

- We begin our study with two existing practical local schedulers and explore their gooput/loss performance, where goodput is defined as the rate seen by a user after possible PRB losses.

- We formulate the centralized multi-cell scheduling problem. Solving the centralized scheduling problem is very hard since it is a very large sized Mixed Integer Non-Linear Programming (MINLP) problem. To solve it, we propose a two-step transformation to transform it into an upper bounding signomial programming problem that can be solved using the iterative algorithm proposed in [7]. We then extract a feasible solution to the original scheduling problem from the result of the upper bounding problem. The feasible solution serves as a lower bound to the optimal solution of the original problem. We will show that the two bounds are close to each other. The results show that the above two local schedulers perform much worse than the centralized scheduler .

- We propose two enhancements that improve the performance of the local schedulers, one when there is no C-RAN and one when there is a C-RAN. The first one is a loss-aware MCS selection that takes possible losses into account, and the second one is called Coordinated Link Adaptation (CLA) and adjust the MCS to avoid all losses. Note that avoiding losses can be critical in 5G networks as losses create jitter and hence degrade the user Quality of Experience (QoE). There enhancements reduce the performance gap with the centralized scheduler significantly, but there is still room for improvement.

- We finally propose a complete suite of schemes, i.e., scheduler and a power allocation scheme, based on Soft Frequency Reuse (SFR) and parametrize it in a robust centralized fashion. It outperforms the benchmark schedulers and perform only 17% worse than the centralized scheduler.

The messages are: Interference estimation in a practical uplink scheduler is critical and when it is not done carefully, the system performance degrades. While a centralized scheduler can perform better than any other scheduler, it is not possible to implement it in real time. Having a C-RAN allows us to provide better enhancements. The proposed enhancements improve the performance of practical schedulers drastically, i.e., more than 50%, we still need a new scheduler to get a reasonable performance, i.e., a performance close to the one of the centralized scheduler.

1.2.2 Downlink Scheduling in Multi-Cell OFDMA Networks

We then study downlink scheduling in a similar fashion. Designing practical downlink schedulers is easier since interference management is simpler as the transmitters in the neighboring cells are always the BSs. However, it is not trivial to design a scheduler that performs well enough.

Our main research question is: What is the maximum achievable performance on the downlink and how do practical schedulers perform compared to it? Similar to the uplink scheduling problem, the best performance can be obtained when all cells are scheduled centrally. To this end, we again study the centralized scheduling problem and compare its performance with practical schedulers. We show that the simple Round Robin (RR) scheduler performs significantly worse than the centralized scheduler and then we will focus on designing an SFR-based scheduler. In this chapter, we also study HetNets. Our main contributions are to show that a well-parametrized SFR scheduler performs well enough so that we might not need centralization via a C-RAN and we also study centralized scheduling in HetNets.

> The main message of this study is that a well-parametrized SFR-based local downlink scheduler performs very well, hence centralizing downlink scheduling might not be necessary.

1.2.3 Full-Duplex Communications in Cellular Networks

After studying uplink and downlink scheduling separately, we focus on how to combine the uplink and downlink to further improve their performances. While traditionally wireless devices are half-duplex, full-duplex communication (FDC) is being considered as a possible 5G technology. Specifically, a BS would be able to transmit and receive at the same time on a given PRB. This is enabled by the advances in self interference cancellation capability (at the BSs) [8]. While it is a promising approach to improve performance, it introduces many new challenges.

New types of interference are introduced in FDC mode such as Self Interference (SI), Inter-BS Interference (IBI), and Inter-User Interference (IUI), all of which will be discussed in detail in Chapter 5. Furthermore, practical scheduling becomes more difficult to design as we need to have more channel gain information since new types of interferences require user-user or BS-BS channel gains, which are not easy to collect, and interference management becomes much more challenging due to the new types of interferences. Therefore,

before deploying an FDC-enabled cellular network, we need to ensure that the gain is significant to warrant its added complexity. In the following, we will focus on a cellular network, where only the BSs are FDC-enabled, and we only consider a centralized multi-cell scheduler both for homogeneous and heterogeneous networks to explore the potential gain of FDC by considering each new type of interference as well as the uplink/downlink traffic asymmetry, which is ignored in the literature.

Our contributions are:

- We formulate a multi-cell joint scheduling problem for FDC-enabled multi-cell networks *that includes all types of interference* and takes traffic asymmetry into account. This problem allocates power and resource blocks jointly for the uplink and the downlink. We also formulate a similar problem for TDD. Using these problems, we study the impact of the different types of interference on the performance gain of FDC over TDD as a function of the traffic asymmetry and draw engineering insights on the potential of FDC in cellular networks for different scenarios (urban and rural) and for different values of Self-Interference Cancellation (SC) and IBI parameters as well as the uplink/downlink traffic asymmetry. In particular, we show that ignoring some of the new sources of interference (for example by analyzing full duplex in a single cell) biases the results in favor of FDC and we *quantify* this.
- We study FDC in HetNets and show that FDC might bring more benefits in HetNets than in homogeneous networks, when the other RRM operations such as CA and UA are performed appropriately.
- We also show how the results produced by models not taking traffic asymmetry into account are biased and favor FDC.

The main message of this study is that the performance of FDC strictly depends on the network parameters and FDC deployment might not always bring a large enough gain to warrant its added complexity. It can be beneficial in the following conditions: i) equal amount of uplink and downlink traffic, ii) rural scenario, iii) HetNet deployment, and iv) very crowded networks.

1.2.4 Cellular Networks with Intra-Cellular Traffic

Many new applications, such as file or photo sharing, create local, i.e., intra-cellular (IC), traffic among cellular users. In a traditional cellular network, all data transfer is done

through the BS (cellular mode). D2D communication has emerged as a component of 5G networks that allows users to by-pass the BS and communicate directly with each other (D2D mode) [9]. However, it is not easy to implement the D2D mode in practical systems since it requires a very careful interference management and the knowledge of channel gains between the users, which is hard to obtain. In the following, we consider three traffic types that are uplink, downlink and IC traffic, where IC traffic is defined as the traffic between two users in close proximity.

We focus on a cellular network with intra-cellular traffic in cellular mode, i.e., when all communications are done between the BSs and the users. Note that a system with D2D mode would need access to information on the *type* of the traffic (flows) and would treat traffic differently based on its type. In the following, we consider different types of traffic, such as uplink, downlink, and intra-cellular. Our main research question is: How would we operate a system in cellular mode if we had access to the traffic type information? We say that the network is *type-aware* when it knows all type of traffic in the system, whereas we call it *type-blind* when it does not the types.

In summary, our contributions can be described as follows:

- We revisit the concept of fairness in the case of multiple flows per user and propose a single metric that measures (device) fairness and efficiency at the same time.
- We formulate and study the coupling of the uplink and the downlink schedulers in the type-aware case for an homogeneous system in the cellular mode. We compare, for different mixes of flows, the gains in performance with respect to the type-blind case, where the schedulers are decoupled.
- We consider a HetNet configuration, for which we formulate and study a type-aware joint user scheduling and user association problem. We compare the gains in performance with respect to the type-blind scheme for different mixes of flows.

The two main messages of the study are that i) important network processes, such as user scheduling and association, should be performed with the knowledge of the type of flows; ii) the uplink and downlink should be jointly scheduled to obtain the best performance when there is IC traffic in the system.

1.3 Outline

The rest of the thesis is organized as follows. In Chapter 2, we explain our system model and the notation commonly used in this thesis. Note that each chapter needs some extra notation for their system model and we explain them in their respective chapters. We give the related literature at each chapter separately. We study uplink scheduling in Chapter 3 by focusing on multi-cell centralized scheduler, practical local schedulers, and their performance improvement. We perform a similar study on downlink scheduling in Chapter 4. We study the performance of full duplex communications in cellular network in Chapter 5. In chapter 6, we study cellular network operations in the existence of intra-cellular traffic and finally conclude the thesis in Chapter 7.

Chapter 2

System Model

In this chapter, we will define the system model we use commonly in this thesis. Note that in some of the following chapters, we need to introduce more details on the system model about the specific problems we study in these respective chapters.

2.1 Network model

We consider a multi-cell OFDMA network that consists of K BSs (set of BSs is $\mathcal{K}$), all of which are using the rM subchannels licensed to the system, where r is the reuse factor between the BSs. Hence, each BS is allowed to use M subchannels. Note that except Chapter 6, where we study intra-cellular traffic, we use a reuse factor $r = 1$ among the BSs. We consider a wrap-around model in order to avoid any border effects.

User scheduling is the process, performed every frame made of T subframes, that allocates the M subchannels to its associated users. A PRB consists of one subchannel and one subframe and it is the smallest scheduling unit. We denote the set of users with $\mathcal{U}$, the set of users associated to BS k with $\mathcal{U}_k$, the set of subchannels with $\mathcal{M}$, and the set of subframes in a frame with $\mathcal{T}$. In some problems, the $\mathcal{U}_k$'s are given, in some others, they are not, i.e., user association is part of the problem.

We consider a full buffer traffic model, i.e., users and BSs always have data to transmit. To simplify the notation, we assume that the channels are flat within a frame time.

2.2 Power, SINR, and rate models

Each user has a power budget P_U and each macro BS has a power budget P_{BS} to be spent in a given subframe over multiple subchannels.

Let $P_{u,k}^{c,t}$ (resp. $p_{u,k}^{c,t}$) denote the power used by BS k to transmit to user $u \in \mathcal{U}_k$ on the downlink on PRB $\{c,t\}$, corresponding to subchannel c and subframe t (resp. used by user u to transmit to BS k on the uplink).

On the downlink, the total interference seen at user $u \in \mathcal{U}_k$ on PRB $\{c,t\}$ is denoted by $I_{u,k}^{c,t}$ and on the uplink the total interference seen at BS k is denoted by $Q_{u,k}^{c,t}$. Note that in a classical cellular network, $I_{u,k}^{c,t}$ includes all the downlink transmissions in the neighboring cells and $Q_{u,k}^{c,t}$ includes all the uplink transmissions in the neighboring cells. However, for the full duplex scenario we will study in Chapter 5, they both include all the uplink and downlink transmissions in the neighboring cells as well as the current cell as will be discussed later.

We denote the channel gain between a user u and a BS k with $G_{u,k}$. We consider symmetric channel gains, i.e., the channel gain between a user and a BS is the same on the uplink and on the downlink. For this kind of channel gains, we use a channel model validated by 3GPP [10]. Numerical details will be given in the following chapters.

The SINR on the downlink, $S_{u,k}^{c,t}$, (resp. on the uplink, $s_{u,k}^{c,t}$) when BS k transmits to (resp. receives from) user $u \in \mathcal{U}_k$ on PRB $\{c,t\}$ is computed as:

$$S_{u,k}^{c,t} = \frac{P_{u,k}^{c,t} \times G_{u,k}}{\mu^{DL} + I_{u,k}^{c,t}}, \tag{2.1}$$

$$s_{u,k}^{c,t} = \frac{p_{u,k}^{c,t} \times G_{u,k}}{\mu^{UL} + Q_{u,k}^{c,t}}, \tag{2.2}$$

where μ^{DL} (resp. μ^{UL}) is the thermal noise on the downlink (resp. uplink) on one PRB.

An adaptive MCS is used and the set of MCS is characterized by the discrete rate function $f(.)$ that maps the SINR to data rates. It is a piece-wise constant function with L steps [1]. Specifically if the SINR on a PRB is in range $(-\infty, \eta_1]$, the rate is zero, if in range $(\eta_j, \eta_{j+1}]$ (for $j \in \{1, \ldots, L-1\}$), then the rate is ϑ_j (see Table 2.1). Hence, if the SINR on a PRB, say $S_{u,k}^{c,t}$ on PRB $\{c,t\}$, has been estimated or computed correctly, the MCS to use should be so that the rate seen by the user on that PRB is $f(S_{u,k}^{c,t})$.

We use the binary variables $x_{u,k}^{c,t,l}$ and $y_{u,k}^{c,t,l}$ to denote if PRB $\{c,t\}$ of cell k is allocated

Table 2.1: Modulation and Coding Schemes [1]

SINR (dB)	-6.5	-4	-2.6	-1	1	3	6.6	10	11.4	11.8	13	13.8	15.6	16.8	17.6
Rate (kb/s)	25.2	38.6	63.8	100.8	147.8	198	248.6	321	404.9	458.6	557.8	655	759.4	860	932.4

to user u using MCS level l on the downlink and uplink, respectively.

2.3 Performance Metric

In the rest of thesis, we will consider proportional fairness as our fairness criterion as it considers efficiency and fairness simultaneously. Therefore, we will use Geometric Mean (GM) of the user throughputs as our performance metric. Specifically, for a network with N users, we will aim at maximizing the following metric:

$$\Gamma(N) = (\prod_{i=1}^{N} \lambda_i)^{1/N}. \tag{2.3}$$

In this case, the higher the GM is, the more fair and efficient the scheduler is.

2.4 Scenarios

In the following, we will consider mainly two scenarios. In the first scenario, there is no coordination among the BSs, hence scheduling is done independently at each BS. In the second scenario, there is a C-RAN in the network that enables the coordination of the scheduling in each cell. A C-RAN is assumed to control a set of cells simultaneously and have access to information, which is not normally available to BSs, such as the channel gains between each user and each BS. However, a practical scheme in that case needs to be fast to compute and not too expensive in terms of information. Most of the centralized problems that we formulate cannot be solved fast enough to be practical. Note that we define a practical scheme as fast scheme that does not require too much information that are difficult to obtain, such as all channel gains in the network.

2.5 Settings

We will consider two different settings, namely a static (snapshot) one and a dynamic one for our performance evaluation. In the first one, we evaluate the performance of a network for several global realizations. A global or local realization ω is characterized by the random deployment of a set of users and the channel gains between each user and each BS. We usually assume that user association is given in this setting. Then, for a given user association, a local realization in cell k is characterized by the set of users associated to BS k and the channel gains between these users and BS k.

The second setting is a dynamic setting where fixed users get powered on, get an association, get served, and then get powered off. We study the dynamic setting for uplink scheduling in Chapter 3. We assume that users get powered on according to a Poisson Point Process (PPP) and depart from the system after completing the upload of a fixed sized file. We analyze the system performance in terms of the average delay experienced by the users. We simulate the dynamic setting with a five percent confidence interval.

Chapter 3

Uplink Scheduling in Multi-Cell OFDMA Networks

3.1 Introduction

The last decade has witnessed a tremendous growth in cellular traffic, which has resulted in a necessity to perform resource management processes in a more efficient way [2]. Such processes include user scheduling, channel allocation, and user association. Most of the work so far has focused on improving the downlink performance since today's applications create a very large amount of downlink traffic [11]. However, more and more applications also need an efficient uplink (e.g., Netflix is creating a large amount of *return* traffic [6]) and with the advent of IoT that allows millions of devices to collect and transmit data [12], the importance of efficient resource management processes on the uplink increases significantly. With the emergence of 5G networks, uplink transmissions will be using OFDMA unlike in LTE-A where SC-FDMA is used. In this chapter, we focus on uplink scheduling in a multi-cell OFDMA network. Note that this is the main chapter of this thesis.

An uplink scheduler decides, in each cell, which user transmits on each PRB and with which transmit power in order to be fair and efficient. A scheduler also assigns an MCS to each {user,PRB} pair. In the following, we only consider proportional fair schedulers [3]. A scheduler also assigns an MCS to each {user, PRB} pair. This assignment requires the computation or estimation of the SINR on each PRB since the signal can only be decoded if the SINR is above a threshold dependent on the MCS. A scheduler needs to compute a new PRB to user assignment along with power and MCS every few milli-seconds, i.e., every frame, and hence the process has to be fast.

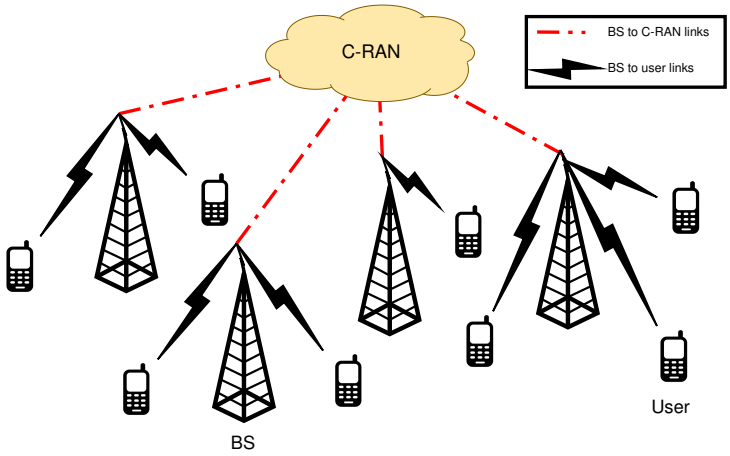


Figure 3.1: An example C-RAN architecture

Typically, each BS performs these processes independently from other BSs. However, the introduction of C-RAN, allows the centralization and coordination of certain processes across many BSs [13]. A sample C-RAN architecture is shown in Fig. 3.1. The decision of which processes are pushed back to the C-RAN is open for debate. User scheduling is often considered as a strong candidate for this.

In this chapter, we focus on a multi-cell OFDMA-based network with no C-RAN (Scenario 1) and with C-RAN (Scenario 2). In the first scenario, there is no coordination between the BSs. Therefore, the scheduler has to estimate the interference and SINR based on measured information. When this is not done perfectly, it could lead to PRB losses, if the BS cannot decode the signal because of an under-estimation of the interference, or to under-utilization, if the interference is over-estimated. We define *goodput* as the effective throughput each user sees after possible PRB losses. Note that since we consider proportional fairness, our performance metric will be the GM of user goodputs.

In the second scenario, there is a C-RAN in the network, in which the scheduling can be done in a centralized fashion. In this case, the C-RAN requires more channel gain information than the BSs can have access to in scenario 1. Then, the challenge is to perform the much larger scheduling task (since it is for all BSs) fast.

We are now ready to state our research questions: are the existing uplink schedulers efficient enough? If not, can we design efficient practical schedulers for those two scenarios?

We start with the study of two practical local benchmark schedulers for the first scenario, one of which is a simple RR scheduler and the other one is proposed in [14]. We

study their goodput/loss performance over many snapshots. We show that interference estimation is critical to their performances and that losses are numerous. To the best of our knowledge, our work is the first to explore the goodput/loss trade-off for an uplink scheduler (Contribution 1). We present the benchmark schedulers together with their goodput/loss performance in Section 3.3[1].

Then, we check how these schedulers perform compared to the best achievable performance. To this end, we consider a fully centralized scheduler that schedules all the cells simultaneously assuming full information (i.e., the channel gains for all $\{i, j\}$ pairs where i is a user and j is a BS). Solving the centralized scheduling problem is very hard since it is a very large sized MINLP problem. To solve it, we propose a two-step transformation to transform it (Contribution 2) into an upper bounding signomial programming problem that can be solved using the iterative algorithm proposed in [7]. We then extract a feasible solution to the original scheduling problem from the result of the upper bounding problem and show that the difference between the feasible solution and the upper bound in terms of GM goodput is very low, hence the upper bound is very tight[2].

We do not believe that such a scheduler could be realized in a C-RAN because i) getting all the cross-channel information, i.e., channel gain with users and other BSs than their own, is not practical and ii) the problem is very large and complex and cannot be solved quickly. However, this centralized scheduling problem is very useful because it provides *an upper bound for the performance of practical scheduling schemes.* We formulate and solve the centralized scheduling problem and then compare its performance with the benchmark scheduler in Section 3.4.

We show that the performance of the benchmark schedulers is significantly worse than of the centralized scheduler (around 55% worse). Then, we focus on means to improve their performance. To do so, we propose two *simple to implement practical* schemes, one for each aforementioned scenario.

The first scheme (Contribution 3), called *Loss-Aware (LA)*, is a *data-driven* MCS selection method and is used in each BS. It replaces the simple MCS selection method, which selects the MCS assuming that SINR estimate is exact, by a method that considers possible PRB losses. It selects the appropriate MCS based on a statistical interference estimation method, i.e., the MCS is selected using the probability distribution of previously measured interference. This method takes into account possible PRB losses, and hence maximizes the *expected* throughput. Note that avoiding losses can be critical in 5G networks as losses

[1]Some of these results are presented in our work [15]

[2]Some of these results are presented in our work [16]. Our work [17] is under review for publication and it is an expanded version of [16].

can create jitter and hence degrade the user QoE.

For the second scenario, we propose that the C-RAN runs a local scheduler per BS (this can be done in parallel and hence fast) and uses on top of it a simple and fast coordination scheme (Contribution 4). In the following, we will refer to this C-RAN based scheduling scheme as the *coordinated link adaptation (CLA)* scheme. CLA reduces the PRB losses significantly even when only partial channel state information (CSI) is available. With CLA, the C-RAN computes the real interference on each PRB at each BS once it knows all the schedules and adjust the MCS on each PRB to avoid losses.

We present and evaluate the performance of the benchmark schedulers along with the proposed schemes, for each scenario, in a dynamic setting in Section 3.5. The schemes significantly improve the delay and loss performance of the existing schedulers with little added complexity. However, when we compare their performance with the centralized scheduler in the snapshot model, we show that the performance gap is still significant (around 30%), which means that a better scheduler is needed.

To this end, we propose to extend SFR to the uplink (Contribution 5). One of the challenges of using SFR on the uplink or on the downlink is the difficulty to parametrize it in a robust fashion. We show how to do so in in Section 3.6. Another challenge is to design a practical scheduler that takes full advantage of SFR. We propose one that uses our LA scheme and show that it outperforms significantly the benchmark schedulers in a dynamic setting in both scenarios. We end the chapter by showing that this scheduler along with CLA reduces the gap with the upper bound, i.e., the centralized scheduler, to only 17% in the second scenario.

Altogether, we believe we provided efficient solutions for the uplink scheduling problem for the two scenarios we considered.

3.2 Related Work

Authors of [18] focus on maximizing the weighted sum-rate of the users. Note that maximizing the sum-rate ignores fairness among the users. Max-min fairness is used in [19] to maximize the performance of the worst user in an OFDMA uplink. Energy efficiency on the uplink is the main focus of [20], where the authors develop low complexity algorithms. However, all these three papers use a single cell scenario and neglect the inter-cell interference. It has been shown in [15] that neglecting inter-cell interference can lead to zero GM goodput since the SINR is grossly overestimated.

It has been shown in [21] that uplink interference depends on the type of scheduler and has a significant effect on the system performance. Therefore, interference management should be taken into account while scheduling the users. Uplink interference estimation is performed in [22] and [23]; however, there is no discussion on scheduling in these papers.

System-wide uplink scheduling is studied in [24] for a simple TDMA-like system that maximizes the sum-rate using a distributed iterative algorithm considering the interference. Heuristic algorithms are proposed in [25] and [26] for more realistic systems; however, there is no discussion on optimality of the schedulers, whereas there is a big amount of data communication overhead among the BSs.

Local scheduling schemes for multi-cell networks are studied in [27], [28], and [14]. The system model has been simplified in [27] by assuming a single modulation scheme and the scheduler maximizes the sum-rate, which results in unfairness. A probabilistic approach is adopted in [28], where transmitting on a PRB is decided based on a probability in order to mitigate inter-cell interference. However, this scheme also uses $\log_2(1+SINR)$ as the rate function. Note that PRB losses are not taken into account when $\log_2(1+SINR)$ is used because different MCS levels are not considered. In real networks, a certain MCS is selected and a PRB might not be decoded at the BS if the SINR is below the threshold for that MCS level. This results in losses and it has a significant impact on system performance [15]. In [14], a local uplink scheduler is proposed that we use as our local benchmark scheduler. It will be described in details in Section 3.3.

Inter-Cell Interference Coordination (ICIC) for the uplink has been studied in [29]. It is based on maximizing the signal to leakage plus noise ratio, where leakage is the interference created to the other cells by one cell. An iterative algorithm is used where the interference created to the other cells on each resource block is limited and this limit is computed iteratively. This scheme requires a high amount of computation and data exchange between the cells and it does not consider possible resource losses. To mitigate the interference problem in Multiple Input Multiple Output (MIMO) systems, a water-filling algorithm is proposed in [30]. Both papers use $\log_2(1+SINR)$ as the rate function; which gives different results than when a realistic piece-wise constant function is used [14].

3.3 Practical Benchmark Schedulers

In this section, we present the two practical benchmark schedulers and evaluate their performances for the first scenario where there is no C-RAN in the system in both static and dynamic settings. In both schedulers, the BSs schedule the users associated to them

locally, independently from the other BSs. In this case, a local scheduler cannot know the exact interference at its BS on each PRB since the transmitters and the power they use on each PRB in the neighboring cells are unknown. This is because a BS cannot know the schedule of another BS beforehand. Therefore, the scheduler has to estimate the interference on each PRB. This can possibly yield under or over-estimation the SINR, which leads to either under-utilization or resource losses.

3.3.1 Round Robin (RR) Scheduler

RR scheduler is the simplest scheduling implementation, which assigns the PRBs one by one to the users in a round-robin fashion (starting from PRB $\{1, 1\}$ then $\{2, 1\}$ and so on) and allocate for a given user, equal power on all subchannels allocated to this user within a subframe, i.e., if user u gets 3 PRBs in subframe t, then the power per PRB would be $P_U/3$. This is the default scheduler in NS-3 [31]. The only remaining job is to allocate an MCS for each PRB and user pair given the allocated power on that PRB. The problem is that to do so the scheduler needs an estimate of the interference. In its simplest form, the scheduler would use the same estimate I^{est} for all PRBs which we assume in the following. Note that I^{est} is an input to the scheduler. If the estimate is not well chosen, there might be losses (because the PRB cannot be decoded correctly) or under-utilization (because a higher MCS could have been used).

In the following, we will show how different values of I^{est} yield different goodput performance.

3.3.2 Local Benchmark (LBM) Scheduler

Next, we present the Local Benchmark (LBM) scheduler originally proposed in [14] in a cell l. LBM allocates a set of subchannels to a user for the duration of the frame. Note that it is different from RR since LBM operates at the subchannel level and RR operates at the PRB level. Similar to the RR scheduler, it uses a fixed estimation I^{est} of the interference on every subchannel, which is an input parameter of the scheduler. Its goal is to give m_i channels to user i such that $\sum_{i \in \mathcal{U}_l} m_i \leq M$ in order to maximize a PF objective function, i.e., $\sum_{i \in \mathcal{U}_l} \log(\lambda_i)$ where λ_i is the throughput that user i would receive if the interference estimate was exact (i.e., estimated throughout).

Since each BS schedules its users separately, we will only focus on one cell, which we call cell l. Let a realization ω_l be characterized by the set $\mathcal{U}_l$ of users associated with cell l with their channel gains to BS l. The operation of the scheduler is based on allocating m_i

subchannels to user i and sharing the power budget of user i, P_U, equally among these m_i channels so as to maximize the proportional fair objective.

Let $R_i^m(I^{est})$ be the estimated throughput of user i when I^{est} is used for interference estimation and user i receives m subchannels. It can be computed as follows:

$$R_i^m(I^{est}) = mf(\frac{(P_U/m)G_{i,l}}{\mu^{UL} + I^{est}}) \tag{3.1}$$

Note that $R_i^m(I^{est})$ can be computed for all users i and all number of subchannels m beforehand once I^{est} is given. We will use the binary variable v_i^m to denote if user i receives m subchannels. Then, given a realization ω_l, a per-channel interference estimation I^{est}, and estimated throughputs $R_i^m(I^{est})$, the corresponding problem $\mathbb{P}^{BM}(\omega_l, I^{est})$ is:

$$\mathbb{P}^{BM}(\omega_l, I^{est}) : \max_{v_i^m \in \{0,1\}, \lambda_i \geq 0} \sum_{i \in \mathcal{U}_l} \log(\lambda_i) \tag{3.2}$$

$$s.t.\ \lambda_i = \sum_{m \in \{1,\dots,M\}} v_i^m R_i^m(I^{est})\ ,\ \forall i \in \mathcal{U}_l \tag{3.3}$$

$$\sum_{i \in \mathcal{U}_l, m \in \{1,\dots,M\}} v_i^m m \leq M \tag{3.4}$$

$$\sum_{m \in \{1,\dots,M\}} v_i^m \leq 1\ ,\ \forall i \in \mathcal{U}_l \tag{3.5}$$

where λ_i is the throughput that user i would receive if the interference estimate was exact (i.e., estimated throughout). Constraint (3.3) determines the estimated throughput of user i depending on how many subchannels are assigned to user i. Constraint (3.4) enforces that the total number of subchannels assigned to the users cannot exceed M.

$\mathbb{P}^{BM}(\omega_l, I^{est})$ is a non-linear integer programming problem due to its non-linear objective function and integer variables. So, it cannot be solved fast. A heuristic algorithm is proposed in [14] that performs quasi-optimally and can be solved fast.

Most of the times, solving $\mathbb{P}^{BM}(\omega_l, I^{est})$ allocates more subchannels to the users with better channel conditions. Therefore, they allocate lower power to their subchannels. This can be considered as a simple power control mechanism that tries to make the received signal strength of each user at the BS similar to each other.

As discussed above, λ_i is the estimated throughput. The goodput seen by a user might be different if the real interference on some of its PRBs is different from I^{est}. When the

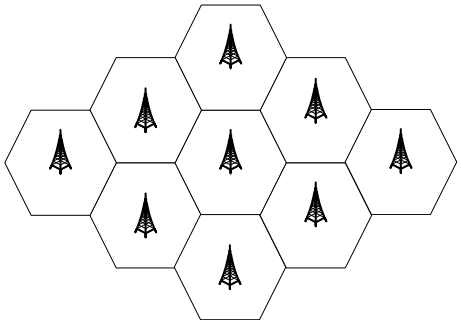


Figure 3.2: Deployment of the 9 BSs.

interference is underestimated, the PRB cannot be decoded at the BS due to low SINR and hence is lost.

3.3.3 Performance Evaluation of the Benchmark Schedulers in a Static Setting

We consider the 9 cell network represented in Fig. 3.2 with a wrap-around model to avoid border effects[3]. The total number of subchannels M is set to 45 and number of subframes T is set to 10. We set P_U to 24 dBm. The piece-wise constant rate function $f(.)$ is given in Table 2.1.

We use one of the channel models recommended by 3GPP [10] to compute the channel gain in dB between user u and BS k:

$$G_{u,k}^{dB} = A_u + A_k - PL(d) - \zeta - \nu, \tag{3.6}$$

where $PL(d)$ represents the path loss computed with the following formula: $128.1 + 37.6 \times \log_{10}(d)$ dB, where d is the distance between u and k. A_u and A_k represent the antenna gains of the users and BSs, which are 0 dBi and 15 dBi, respectively. ζ, a random variable, is the log-normal shadowing of 8 dB standard deviation and ν is the penetration loss of 20 dB. Note that whenever we use $G_{u,k}$ earlier, it had no units and it can be obtained from $G_{u,k}^{dB}$ as $G_{u,k} = 10^{G_{u,k}^{dB}/10}$. Finally, we associate each user to the BS that yields the highest channel gain.

[3]The hexagonal shape of the cells is to be taken symbolically. It does not represent the exact geometrical shape of a coverage area.

We consider a global realization ω characterized by a set of users $\mathcal{U}$ partitioned in subsets $\mathcal{U}_j$ ($j \in \{1,.,9\}$) corresponding to a user association, and the set of channel gains for each pair $\{i,j\}$ where i is a user and j is a BS. Given a value for I^{est}, we will obtain the goodput for each user in $\mathcal{U}$ for one of the two benchmarks in two steps for a given frame:

- Step 1: We schedule each BS $j \in \{1,.,9\}$ using the benchmark under consideration and I^{est}. In this step, each scheduler has an estimate of the goodput in that frame for each user.
- Step 2: We can now compute for each PRB in each cell the real interference since we know which user is transmitting with which power. We can then determine for each PRB if it can be decoded or not depending if the SINR is greater or not than the threshold for the MCS used in that PRB. If the PRB is not decodeable, it is assumed lost and does not count for the goodput. We now have the true value of the goodput for each user in that frame.

For each ω, we repeat the above process for 100 frames (in each frame, we randomize the channel allocation) and compute the estimated and exact goodput of each user averaged over all the frames.

Since we use proportional fairness, the system performance metric is the geometric mean of the user goodputs given as $\Gamma(N) = (\Pi_{i=1}^{N}\lambda_i)^{1/N}$, where N is the number of users in realization ω (i.e., $N = |\mathcal{U}|$) and λ_i is the goodput of user i.

Fig. 3.3 shows the results in terms of the estimated GM goodput and the exact GM goodput for the two benchmarks for different values of I^{est}.

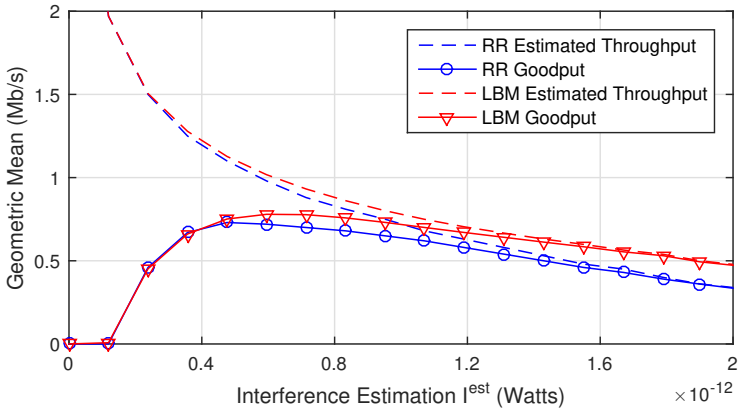


Figure 3.3: GM estimated throughput and GM goodput comparison of the local benchmark schedulers

The figure unveils several insights. First of all, the value of the interference estimate plays an important role in the performance of the two schedulers. Second, there is a significant difference between the estimated GM throughput and the real GM goodput when the interference is under-estimated which corresponds to the left part of the graph. In this case, the PRB loss rate is very high, e.g., around 80% when I^{est} is zero. On the other hand, when the interference is over-estimated (in the right part of the graph) then the estimated GM throughput is almost identical to the GM goodput as the PRB loss rate drops below 1%.. Thirdly, under estimating or over estimating the interference yields bad performance. The maximum GM goodput is achieved when the interference estimate is I^*, which is different for the two schedulers. The PRB loss rate at I^* is 20% for both schedulers. Therefore, good performance can only be obtained at a cost of high losses, which create retransmissions and high jitter. Finally, LBM performs slightly better than the RR scheduler. The maximum value of GM goodput for LBM is 1.07% the maximum value of RR.

We then investigate the goodput/loss trade-off. Since RR and LBM yield similar trade-off curves, we will only illustrate the results for LBM in Fig. 3.4 when the average number of users per cell is set to 10.

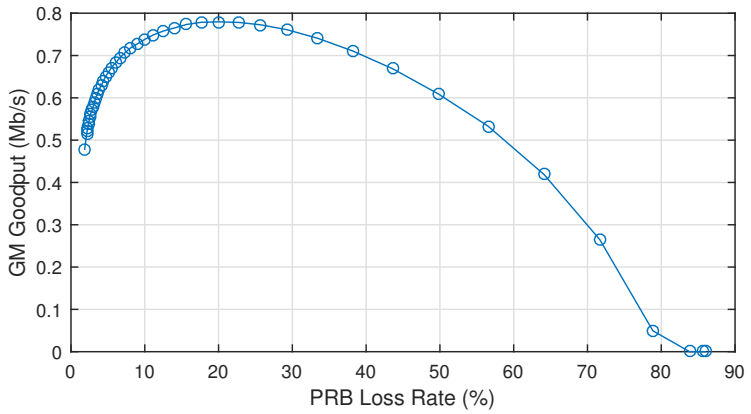


Figure 3.4: GM goodput/PRB loss rate trade-off performance of LBM

It is seen that when PRB loss rate is minimized, the GM goodput is very low compared to the maximum achieved at I^*.

3.3.4 Performance of the Benchmark Schedulers in a Dynamic Setting

In a dynamic setting, a local scheduler would use a data-driven estimator for the interference. A simple one is to collect the interference values in the previous frame and use the average of these values as an estimate for the next frame. This is the one, we will use for the baseline benchmarks. Note that the estimation changes from frame to frame and is different from a cell to another.

To evaluate the performance of the two benchmark schedulers with this interference estimation in a dynamic setting, we consider a scenario where users arrive in the 9-cell network according to a Poisson Point Process with an arrival rate Λ. Each user is associated to the BS that yields the highest channel gain. The users leave the system after completing the upload of a fixed sized file of $F = 10$ Mb. Fig. 3.6 shows the average delay spent in the system as a function of Λ. Recall that when a PRB cannot be decoded, it is considered lost. We do not model retransmissions, instead we assume optimistically that the file is received when a total number of bits equal to F has been received correctly at the receiver. All the simulations are done with a five percent confidence interval.

LBM in a dynamic setting continues to perform better than RR. For example, the arrival rate that yields an average delay less than 10 seconds is around 4.5 users per second for RR, and around 5 users per second for the benchmark scheduler. At these arrival rates, the RR scheduler yields a PRB loss of 27%, and the benchmark scheduler yields a PRB loss rate of 24%. Note that these numbers are quite high and one reason for these high loss ratios is that the MCS selection does not consider possible PRB losses due to bad interference estimation, i.e., it takes the interference estimate at face value.

So far, we have shown how different interference estimates lead to different results and how the benchmarks perform with respect to each other in both static and dynamic settings. The real question is: *How do the two local schedulers perform compared to the maximum achievable performance?* Recall that we can achieve the best performance when we schedule all the cells simultaneously with the knowledge of all channel gain information in the system, since interference can then be taken into account exactly. In the next section, we will formulate and solve the centralized multi-cell scheduling problem and compare its performance with the benchmark schedulers. Note that we do not propose it as a practical scheme due to its complexity, but we will use it to compute offline an upper bound on the performance of practical schedulers.

3.4 Multi-Cell Centralized Scheduler

3.4.1 Formulation

The most important feature of the system-wide user scheduling problem for the uplink of an OFDMA cellular network is that it considers the interference in the formulation as opposed to using an estimate input as in a local scheduling problem. In the problem formulation, we use the binary variable $y_{u,k}^{c,t,l}$ that is equal to 1 if PRB $\{c,t\}$ of BS k is allocated to $u \in \mathcal{U}_k$ using MCS level l. All the variables are continuous and non-negative except when otherwise indicated. The scheduling problem is about allocating PRBs to users (via the binary variables $y_{u,k}^{c,t,l}$) and the corresponding transmit powers (via the continuous variables $p_{u,k}^{c,t}$).

Recall that the SINR to rate mapping function $f(s_{u,k}^{c,t})$, where $s_{u,k}^{c,t}$ is the SINR of user $u \in \mathcal{U}_k$ on PRB $\{c,t\}$, can be written as:

$$\texttt{if} \quad \eta_l \leq s_{u,k}^{c,t} < \eta_{l+1}, \texttt{then} \quad r_{u,k}^{c,t} = \vartheta_l \quad \forall l \in \mathcal{L} \tag{3.7}$$

To include this discrete function in our optimization problem, we use the following constraint:

$$s_{u,k}^{c,t} \geq y_{u,k}^{c,t,l}\eta_l,\ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, l \in \mathcal{L} \tag{3.8}$$

Note that when the SINR $s_{u,k}^{c,t}$ on a PRB $\{c,t\}$ is lower than the decoding threshold η_l, $y_{u,k}^{c,t,l}$ is set to zero. Otherwise, it can be either one or zero. Since the objective is to maximize the goodput, the maximum MCS level that satisfies $s_{u,k}^{c,t} \geq \eta_l$ is selected on each PRB.

Then, for a given global realization ω, we formulate the system-wide problem $\mathbb{P}_{\mathbb{S}}(\omega)$ (given in Equations (3.9)-(3.18)) that maximizes the proportional fair objective function over all users in the system, which is the GM of the user goodputs.

$$\mathbb{P}_{\mathbb{S}}(\omega): \max_{(p_{u,k}^{c,t}),(\lambda_u^k),(y_{u,k}^{c,t,l}),(Q_k^{c,t}),(r_{u,k}^{c,t})} \sum_{u\in\mathcal{U}_k}\sum_{k\in\mathcal{K}} \log(\lambda_u^k) \tag{3.9}$$

$$s.t.\ \ \lambda_u^k = \frac{1}{T}\sum_{c\in\mathcal{M}}\sum_{t\in\mathcal{T}} r_{u,k}^{c,t},\ \ \forall u \in \mathcal{U}_k,\ k \in \mathcal{K} \tag{3.10}$$

$$r_{u,k}^{c,t} = \sum_{l\in\mathcal{L}} y_{u,k}^{c,t,l}\vartheta_l,\ \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.11}$$

$$s_{u,k}^{c,t} \geq y_{u,k}^{c,t,l}\eta_l,\ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, l \in \mathcal{L} \tag{3.12}$$

$$s_{u,k}^{c,t} = \frac{p_{u,k}^{c,t}G_{u,k}}{\mu^{UL} + Q_k^{c,t}},\ \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.13}$$

$$Q_k^{c,t} = \sum_{j\in\mathcal{K}, j\neq k}\sum_{v\in\mathcal{U}_j} P_{v,j}^{c,t}G_{v,k},\ \ \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.14}$$

$$\sum_{c\in\mathcal{M}} p_{u,k}^{c,t} \leq P_U,\ \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, t \in \mathcal{T} \tag{3.15}$$

$$p_{u,k}^{c,t} \leq \sum_{l\in\mathcal{L}} y_{u,k}^{c,t,l}P_U,\ \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.16}$$

$$y_{u,k}^{c,t,l} \in \{0,1\},\ \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.17}$$

$$\sum_{u\in\mathcal{U}_k}\sum_{l\in\mathcal{L}} y_{u,k}^{c,t,l} \leq 1,\ \ \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.18}$$

The throughput of a user is the sum of the rates seen at each PRB (constraint (3.10)). The rate on each PRB is defined by constraint (3.11) by picking the appropriate MCS level as only one MCS level is selected on each PRB by each user, which is enforced by

constraint (3.18). Constraint (3.12) defines which MCS level is selected on each PRB as explained above. Constraint (3.13) defines the SINR on each PRB and constraint (3.14) defines the interference seen by cell k on PRB $\{c, t\}$. Constraint (3.15) ensures that the total power used by user u cannot exceed P_U at a given time. A user cannot allocate power to a PRB if that PRB is not allocated to that user, which is ensured by constraint (3.16). Constraint (3.17) enforces that the variable that indicates if a PRB is assigned to a user is a binary variable while Constraint (3.18) enforces that a PRB can be allocated to only one user using one MCS level.

The main issue with this optimization problem is its computational complexity. It is an MINLP problem: it includes a large number of integer variables $y_{u,k}^{c,t,l}$ and its objective function is non-linear. Hence, it cannot be solved using a commercial solver unless the network size is very small.

Next, we propose a method to eliminate the integer variables through two non-trivial transformations to obtain a signomial programming problem [32] that upper bounds $\mathbb{P}_{\mathbb{S}}(\omega)$.

3.4.2 Transformations Yielding The Upper Bound Signomial Programming Problem

First Transformation

Recall that the discrete rate function $f(\gamma)$ is given in Equation (3.7) and to include it in the optimization problem, we have used the constraints (3.11), (3.12), and (3.18). We propose to replace $f(\gamma)$ by an upper-bounded function, $g(\gamma)$, made up of H linear functions $a_h \times \gamma + b_h$ (see Fig. 3.5). This upper bound can be made tight by selecting H and the $\{a_h, b_h\}$ properly. Let $\mathcal{H}$ be the set of linear functions. We can then replace constraints (3.11), (3.12), and (3.18) with the H linear constraints:

$$r_{u,k}^{c,t} \leq a_h \times s_{u,k}^{c,t} + b_h, \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, h \in \mathcal{H} \tag{3.19}$$

since maximizing the objective function ensures that the maximum feasible rate will be chosen. We can also replace the binary variables $y_{u,k}^{c,t,l}$ by $y_{u,k}^{c,t}$ and hence eliminate many binary variables.

Second Transformation

Binary variables are used in $\mathbb{P}_{\mathbb{S}}(\omega)$ for PRB allocation since a PRB can be allocated to only one user by a BS. We can eliminate those variables altogether if we add the following

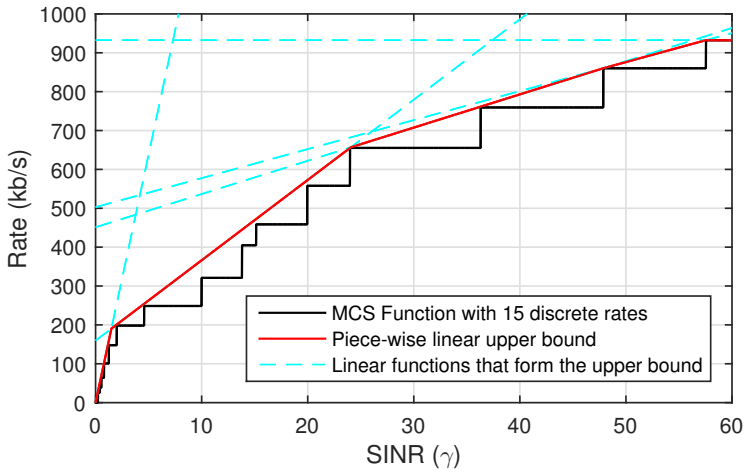


Figure 3.5: MCS function $f(\gamma)$ with 15 discrete rates and the piece-wise linear upper bound $g(\gamma)$ for $H = 5$

two constraints to the problem $\forall u, v \in \mathcal{U}_k, u \neq v, \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}$:

$$p_{u,k}^{c,t} p_{v,k}^{c,t} \leq \sigma, \qquad r_{u,k}^{c,t} r_{v,k}^{c,t} \leq \sigma \tag{3.20}$$

where σ is a very small positive constant. In this case, even though multiple users of a cell can technically use the same PRB, we force the product of the power allocated by those users (and their rates) on that PRB to be close to zero so that only one user actually uses that PRB. Although using one of those constraints would be enough to satisfy this requirement, having both constraints makes the problem more robust.

For a given local realization ω, the sets $\mathcal{U}_k$'s, $\mathcal{H}$, and σ, we formulate $\mathbb{P}_{\mathbb{UB}}(\omega)$, which finds an upper bound for $\mathbb{P}_{\mathbb{S}}(\omega)$:

$$\mathbb{P}_{\mathbb{UB}}(\omega): \max_{(p^{c,t}_{u,k}),(\lambda^k_u),(s^{c,t}_{u,k}),(r^{c,t}_{u,k})} \prod_{u\in\mathcal{U}_k}\prod_{k\in\mathcal{K}} \lambda^k_u \tag{3.21}$$

$$s.t.\ \ p^{c,t}_{u,k}p^{c,t}_{v,k} \le \sigma, \quad \forall u,v \in \mathcal{U}_k, u \neq v, \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.22}$$

$$r^{c,t}_{u,k}r^{c,t}_{v,k} \le \sigma, \quad \forall u,v \in \mathcal{U}_k, u \neq v, \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.23}$$

$$\lambda^k_u = \frac{1}{T}\sum_{c\in\mathcal{M}}\sum_{t\in\mathcal{T}} r^{c,t}_{u,k}, \quad \forall u \in \mathcal{U}_k,\ k \in \mathcal{K} \tag{3.24}$$

$$r^{c,t}_{u,k} \le a_h s^{c,t}_{u,k} + b_h, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, h \in \mathcal{H} \tag{3.25}$$

$$s^{c,t}_{u,k} = \frac{p^{c,t}_{u,k}G_{u,k}}{\mu^{UL} + \sum_{l\in K}\sum_{v\in\mathcal{U}_l, v\neq u} P^{c,t}_{v,l}G_{v,k}}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{3.26}$$

$$\sum_{c\in\mathcal{M}} p^{c,t}_{u,k} \le P_U, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, t \in \mathcal{T} \tag{3.27}$$

$\mathbb{P}_{\mathbb{UB}}(\omega)$ is obtained from $\mathbb{P}_{\mathbb{S}}(\omega)$ by replacing constraints (3.11), (3.12), and (3.18) with constraint (6.4) and adding constraints (3.22) and (3.23) so that there are no binary variables left in the problem. The rest of the constraints are simply rewritten versions of the constraints in $\mathbb{P}_{\mathbb{S}}(\omega)$. Note that maximizing $\prod_{u\in\mathcal{U}_k}\prod_{k\in\mathcal{K}}\lambda^k_u$ is the same as maximizing $\sum_{u\in\mathcal{U}_k}\sum_{k\in\mathcal{K}}\log(\lambda^k_u)$. $\mathbb{P}_{\mathbb{UB}}(\omega)$ is a non-convex problem that includes non-linear constraints. However, $\mathbb{P}_{\mathbb{UB}}(\omega)$ can be further transformed in order to belong to the class of *signomial programming* problems and as such it can be solved by converting the problem into a series of geometric programming problems[4] [7].

To convert $\mathbb{P}_{\mathbb{UB}}(\omega)$ into an equivalent signomial program, we first replace the equality constraints with inequality constraint ($\le$) in constraints (3.24) and (3.26) (this does not change the optimal point of the problem since it is a maximization problem). We then modify the constraints so that the variables are all positive (constraints that were omitted for brevity in $\mathbb{P}_{\mathbb{UB}}(\omega)$) by stating that all the variables are greater than a very small positive number ρ. We also arrange the other constraints and the objective function to highlight the typical structure of a signomial problem. Then, we obtain the signomial problem $\mathbb{P}^*_{\mathbb{UB}}(\omega)$ that upper bounds the original problem $\mathbb{P}_{\mathbb{S}}(\omega)$:

[4] A geometric program can be solved optimally by transforming it into an equivalent convex problem [32].

$$\mathbb{P}^*_{\mathbb{UB}}(\omega): \quad \min_{(p^{c,t}_{u,k}),(\lambda^k_u),(s^{c,t}_{u,k}),(r^{c,t}_{u,k})} \prod_{u\in\mathcal{U}_k}\prod_{k\in\mathcal{K}}(\lambda^k_u)^{-1} \tag{3.28}$$

$$s.t. \ \frac{T\lambda^k_u}{\sum_{c\in\mathcal{M}}\sum_{t\in\mathcal{T}} r^{c,t}_{u,k}} \leq 1, \quad \forall u\in\mathcal{U}_k, \ k\in\mathcal{K} \tag{3.29}$$

$$\frac{r^{c,t}_{u,k}}{a_h s^{c,t}_{u,k}+b_h} \leq 1, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T}, h\in\mathcal{H} \tag{3.30}$$

$$\sum_{l\in K}\sum_{v\in\mathcal{U}_l, v\neq u}\frac{G_{v,k}}{G_{u,k}}p^{c,t}_{v,l}s^{c,t}_{u,k}(p^{c,t}_{u,k})^{-1}+\frac{\mu^{UL}}{G_{u,k}}s^{c,t}_{u,k}(p^{c,t}_{u,k})^{-1}\leq 1, \tag{3.31}$$
$$\forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T}$$

$$\sum_{c\in\mathcal{M}}\frac{p^{c,t}_{u,k}}{P_U}\leq 1, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, t\in\mathcal{T} \tag{3.32}$$

$$\frac{1}{\sigma}p^{c,t}_{u,k}p^{c,t}_{v,k}\leq 1, \quad \forall u,v\in\mathcal{U}_k, u\neq v, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{3.33}$$

$$\frac{1}{\sigma}r^{c,t}_{u,k}r^{c,t}_{v,k}\leq 1, \quad \forall u,v\in\mathcal{U}_k, u\neq v, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{3.34}$$

$$p^{c,t}_{u,k}\geq\rho, r^{c,t}_{u,k}\geq\rho, s^{c,t}_{u,k}\geq\rho, \lambda_{u,k}\geq\rho, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{3.35}$$

An Iterative Method to Solve the Upper Bound Problem

In this section, we briefly explain the method of solving $\mathbb{P}^*_{\mathbb{UB}}(\omega)$ [7]. It is an iterative algorithm in which we convert the problem into a geometric programming problem at each iteration. Note that $\mathbb{P}^*_{\mathbb{UB}}(\omega)$ is not a geometric program because of the first two constraints ((3.29) and (3.30)). Inequality constraints of a geometric programming problem should be of the form:

$$f(x)\leq 1, \tag{3.36}$$

where $f(x)$ is a posynomial[5].

The main idea behind the algorithm in [7] is to modify these constraints such that the program becomes a geometric program. Note that the denominators of the left hand side of those two constraints are posynomials. If they can be approximated as monomials, then $\mathbb{P}^*_{\mathbb{UB}}(\omega)$ becomes a geometric programming problem.

[5]A posynomial is a sum of monomials, of the form $g(x)=\sum_{k=1}^{K} d_k x_1^{a_k^{(1)}} x_2^{a_k^{(2)}}\dots x_n^{a_k^{(n)}}$.

It is shown in [32] that the best monomial approximation around a point $x = y > 0$, say $\bar{h}(x, y)$, for a posynomial $h(x) = \sum_i g_i(x)$, where the $g_i(x)$'s are monomials, is:

$$\bar{h}(x, y) = \prod_i \left(\frac{g_i(x)}{\alpha_i(y)}\right)^{\alpha_i(y)}, \tag{3.37}$$

where $\alpha_i(y)$ is equal to:

$$\alpha_i(y) = \frac{g_i(y)}{h(y)} \tag{3.38}$$

Hence, at a given point y, the constraints (3.29) and (3.30) can be approximated using equations (3.37) and (3.38). By doing so, we obtain the problem $\mathbb{P}^{**}_{\mathbb{UB}}(\omega, y)$, which is a geometric program that can be solved by techniques presented in [32].

The purpose of the iterative algorithm is to find y. The point, around which the approximation is made, changes in each iteration. The overall algorithm is explained below.

Algorithm 1 Solving $\mathbb{P}^{*}_{\mathbb{UB}}(\omega)$ iteratively

1: Find a feasible initial solution for $\mathbb{P}^{*}_{\mathbb{UB}}(\omega)$. Let that vector be $s^{(0)}$.
2: At the t^{th} iteration, compute the monomial approximations for the first two constraints of $\mathbb{P}^{*}_{\mathbb{UB}}(\omega)$ using equations (3.37) and (3.38), with $y = s^{(t-1)}$, and obtain $\mathbb{P}^{**}_{\mathbb{UB}}(\omega, s^{(t-1)})$.
3: Solve $\mathbb{P}^{**}_{\mathbb{UB}}(\omega, s^{(t-1)})$ with the new constraints by converting it into a convex problem using a log transformation as explained in [32] and obtain $s^{(t)}$
4: **if** $||s^{(t-1)} - s^{(t)}|| < \varepsilon$ **then**
5: Algorithm terminates
6: **else**
7: Go to step 2
8: **end if**

3.4.3 Deriving a Feasible Solution to $\mathbb{P}_{\mathbb{S}}(\omega)$

We can derive a feasible solution for $\mathbb{P}_{\mathbb{S}}(\omega)$ out of the solution for $\mathbb{P}^{*}_{\mathbb{UB}}(\omega)$. Note that this feasible solution serves as a lower bound to $\mathbb{P}_{\mathbb{S}}(\omega)$. To obtain it, we use the values of $p^{c,t}_{u,k}$ we obtain by solving $\mathbb{P}^{*}_{\mathbb{UB}}(\omega)$ and compute a feasible PRB allocation using $\forall u \in \mathcal{U}_k$:

$$y^{c,t,l}_{u,k} = \begin{cases} 1, & \text{if } \underset{i \in \mathcal{U}_k}{\arg\max}\ p^{c,t}_{i,k} = u \ \ \& \ \ \eta_l \leq s^{c,t}_{u,k} < \eta_{l+1} \\ 0, & \text{otherwise} \end{cases} \tag{3.39}$$

and the corresponding value for $p_{u,k}^{c,t}$ if $y_{u,k}^{c,t,l} = 1$. This simply means that a PRB is allocated to the user that uses the highest power on that PRB and the appropriate MCS level is also selected depending on the SINR of that user on that PRB. We will show numerically that the gap between the lower and the upper bounds is small and hence the upper bound is tight.

3.4.4 Numerical Results

In this section, we report the results illustrating the tightness of the upper bound and the performance of the benchmark schedulers compared to the centralized scheduler.

We use the same network configuration as before. Each user is associated to the BS that yields the highest channel gain.

We consider the following five linear functions to upper bound the original rate function (in kb/s) as shown in Fig. 3.5:

$$r_{u,k}^{c,t} \leq 127.68 \times s_{u,k}^{c,t} \tag{3.40}$$

$$r_{u,k}^{c,t} \leq 20.66 \times s_{u,k}^{c,t} + 159.6 \tag{3.41}$$

$$r_{u,k}^{c,t} \leq 8.544 \times s_{u,k}^{c,t} + 451.21 \tag{3.42}$$

$$r_{u,k}^{c,t} \leq 7.465 \times s_{u,k}^{c,t} + 502.86 \tag{3.43}$$

$$r_{u,k}^{c,t} \leq 940.8 \tag{3.44}$$

We use Bonmin solver [33] to solve the optimization problems.

Tightness of the Upper Bound

To show the tightness of the upper bound we obtain by solving $\mathbb{P}_{\mathbb{UB}}^{*}(\omega)$, for different values of the number of users in the system, N, we create 100 realizations for each value of N and show the average difference in the GM goodput for the upper bound and the feasible solution in Table 3.1. The difference is around 7% and since the optimal solution to $\mathbb{P}_{\mathbb{S}}(\omega)$ is between the upper bound and the feasible solution, the upper bound we obtain can be considered tight.

Table 3.1: Comparison of the upper bound and the feasible solution as a function of the total number of users N

N: **Total number of users in the system**	45	90	135
Avg. Difference (%)	7.27%	7.42%	7.59%

Performance comparison of the benchmark schedulers and the centralized scheduler

To compare the performance of the benchmark schedulers with the centralized scheduler, we compute the per realization ratio (in percentage) of the best GM goodput of the local scheduler (i.e., the one obtained for I^*) to the GM goodput of the centralized scheduler. Table 3.2 report these ratios averaged over 100 realizations as a function of N, the total number of users in the system. In the table, a value of 43.9 means that the performance of the local scheduler is equal to 43.9% of the centralized system on average. Note that I^* depends on the number of users N in the system. Therefore, we use the best I^* for each value of N in the table (see Fig. 3.3 for more details on I^*).

Table 3.2: Averaged performance ratio of the local schedulers to the centralized scheduler

N: **Total number of users in the system**	**45**	**90**	**135**
RR	39.9	41.2	42.1
LBM	43.9	43.6	44.4

Clearly, there is a very significant performance gap between the performance of the benchmark schedulers and the centralized scheduler. This is mainly due to the fact that the benchmark schedulers have no information about the neighboring cells' schedules, therefore the interference needs to be estimated and this cannot be done accurately. Hence, many PRBs are lost or under-utilized. Also, the centralized scheduler computes the best power allocation, whereas the local schedulers do the power allocation with simpler schemes. In the next section, we will propose enhancements to the schedulers to decrease this performance gap for the two scenarios presented in the introduction.

3.5 Scheduler Enhancements

In this section, we are going to propose enhancements to the benchmark schedulers to minimize the loss rate and then evaluate their performances in a dynamic setting.

3.5.1 The Loss-Aware (LA) Scheme

We begin with the first scenario, where each cell schedules its own users and there is no C-RAN in the network. The LA scheme is a data-driven MCS selection scheme. It collects all the interference measurements during the previous frame[6] (typically there are MT such measurements, i.e., 450 in our case) and obtains a Cumulative Distribution Function (CDF) of that interference. Then, while choosing an MCS for each PRB, it can also consider possible PRB losses due to under-estimating the interference by maximizing the expected PRB rates after possible losses. The details are given next.

To begin with, we need to know the power used on a subchannel by each user in order to estimate the SINR. Since the users allocate equal power to the subchannels allocated to them in RR and LBM, we know the power used on a subchannel by each user i when m_i subchannels are allocated to that user, which is P_U/m_i. Let the SINR of user i of BS k when i uses a power $P_i^{c,t}$ on PRB $\{c,t\}$ be:

$$s_i^{c,t} = \frac{p_i^{c,t} \times G_{i,k}^{c,t}}{\mu^{UL} + Q_k}, \tag{3.45}$$

where Q_k is a *random variable* representing the interference on that PRB which is characterized by its CDF $F_k(.)$. The PRB can be decoded at the z^{th} MCS level only if $\eta_z \leq s_i^{c,t}$, which translates into the interference Q_k being below the threshold:

$$\delta_z(i,k)^{c,t} = \frac{p_i^{c,t} \times G_{i,k}^{c,t}}{\eta_z} - \mu^{UL}, \tag{3.46}$$

the probability of which can be computed given the CDF. Then, the expected rate user $i \in U_k$ can get using MCS level z on that PRB is:

$$E_{i,k}(z) = F_k(\delta_z(i,k)^{c,t}) \times \vartheta_z, \tag{3.47}$$

where ϑ_z is the rate corresponding to the z^{th} MCS.

If we select the MCS z^* maximizing this expectation:

$$E_{i,k}(z^*) = \max_{z \in \mathcal{L}} \; E_{i,k}(z), \tag{3.48}$$

[6]It can also collect the measurements for a few frames instead of just one. Based on results that are not reported here, collecting data for more frames has little impact on system performance while bringing more complexity.

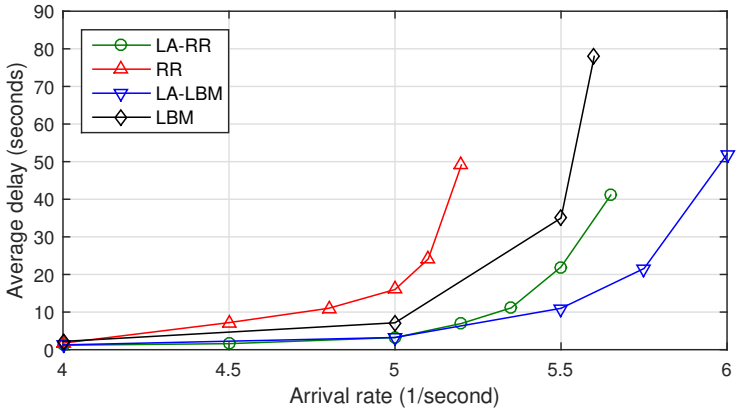


Figure 3.6: Scenario 1: Average delay performance of the RR and local benchmark scheduler for constant interference estimation and with LA

we choose the MCS level that gives the *maximum expected rate*. This MCS selection takes into account possible PRB losses since the probability of loss increases as z increases.

Once the maximum expected rate of each user i for each PRB is computed, we can solve the same scheduling problem for the benchmark scheduler. Note that the same CDF is used on all PRBs.

Fig. 3.6 shows the delay performance of the RR and local benchmark scheduler when they use the LA scheme.

The maximum arrival rate supported by each scheduler improves by around 10% when using the LA scheme. Furthermore, the losses decrease to around 15% from 27% for RR and to 14% from 24% for LBM, which is the main decrease for the decrease in average delay. Therefore, the LA scheme improves the performance significantly. Recall that we do not simulate retransmissions. In a real system where retransmissions are triggered after a certain number of subframes, the losses would have a more serious impact. Hence, lower losses would further improve the performance gain of our LA scheme over the benchmark MCS selection scheme in a real system.

3.5.2 Coordinated Link Adaptation

In the previous system, each BS uses limited information to compute its local schedule, i.e., an interference estimate and the channel gains from all its users. Consider now the

case where the system has a C-RAN (scenario 2). In that case, the scheduling can be done in a coordinated fashion, however, it has to be done fast.

We propose the following two-step process that has the advantage of being simple and fast:

- Step 1: Each cell is scheduled independently (i.e., in parallel) in the C-RAN using a practical local scheduler (e.g., RR or LBM).
- Step 2: The C-RAN coordinates the scheduling using a fairly simple and fast process to minimize losses.

In an ideal case, the C-RAN would know the channel gains between each user and each BS. Then, after the first step, once the scheduling is completed for each cell, the C-RAN could compute the real interference on each PRB for each cell since it knows all the transmitters on that PRB along with the power allocation and the channel gains. Indeed, on PRB $h = \{c, t\}$ in cell l, the interference can be computed by the C-RAN as:

$$Q_{l,h} = \sum_{k \in \mathcal{K}, k \neq l} p^h_{u^h_k, k} \times G_{u^h_k, l} \tag{3.49}$$

where u^h_k is the transmitting user in cell k on PRB h.

Hence, in the second step, the C-RAN computes the exact SINR on each PRB for each cell and can *correct* the MCS, i.e., pick the best MCS to maximize the data rate. Note that this is a very simple level of coordination that does not require any complex computations. It does not change the PRB allocations in terms of users and power. It just adjusts the MCSs through coordination. We call this *coordinated link adaptation (CLA)*.

In the following, we first study the ideal case (with complete information on the channel gains) and then the case with partial information.

The ideal case: since we have full channel gain information, there will be *no PRB losses* after Step 2 since the real SINRs are known.

Fig. 3.7 shows the performance of the benchmark schedulers with and without CLA in the case of scenario 2. Note that when there is no CLA, we assume that the schedulers are using the LA scheme, while when there is CLA, the LA scheme is not used in Step 1 since it does not bring any gain. CLA does improve the delay performance drastically. The maximum arrival rate supported by both schedulers is improved by more than 40%.

Next, we consider the case where not all channel gains are available at the C-RAN.

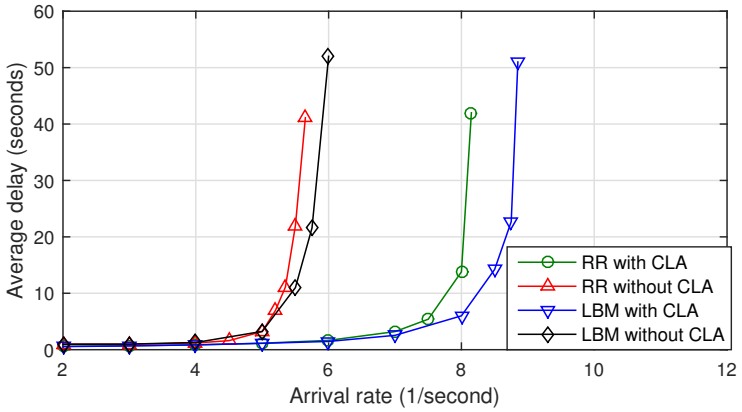


Figure 3.7: Scenario 2: Average delay performance of the RR and local benchmark scheduler with and without CLA (ideal case)

The non-ideal case: In practice, the C-RAN does not have access to full channel gain information. Instead, each user measures its channel gains to some other BSs than the one it is associated with for handover purposes. Let q be the number of neighboring BSs each user i monitors for Channel State Information (CSI) (i.e., the q neighboring BSs yielding the highest gains to that user). Let the set of such neighbors be $\mathcal{N}_i(q)$. Hence, with respect to Eq. (3.49) the C-RAN has access to $G_{u_k^h,l}$ if $l \in \mathcal{N}_{u_k^h}(q)$, otherwise $G_{u_k^h,l}$ is unknown.

While computing the interference, the C-RAN uses the real value of $G_{u_k^h,l}$ if $l \in \mathcal{N}_{u_k^h}(q)$. Otherwise, it uses a *pre-computed estimate* of $G_{u_k^h,l}$. For simplicity, we choose to select a value that does not depend on the PRB or on the transmitting user in cell k. We call it $c(k,l)$ and we compute it offline as the median of the channel gains of all possible locations in the Voronoi cell around BS k. Specifically, for each point in that Voronoi cell, we compute the channel gain between that point and the BS in cell l using Equation (3.6).

In Fig. 3.8, we show how this heuristic performs compared to the ideal case of full information for the cases of $q = 2$ and $q = 3$ when using LBM as the scheduler. Clearly, even when the channel gain information is limited, we can improve the performance of the uplink scheduler by using CLA. For example, when we check the maximum arrival rate that yields a delay of 30 seconds, it is 5.5 users/second without CLA, 7.25 users/second with CLA and $q = 2$, 7.75 users/second with CLA and $q = 3$, and 8.75 users/second with CLA in the ideal case.

In summary, the two-step scheduling process we propose is simple and efficient and works well even when the cross-channel information is limited.

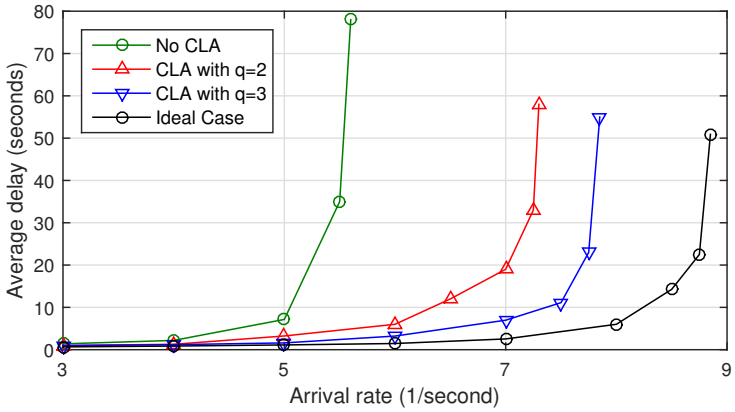


Figure 3.8: Scenario 2: Average delay performance of LBM with and without CLA for the ideal and non-ideal cases

3.5.3 Closing the Gap?

In this section, we have proposed two schemes to enhance the benchmark scheduler, one for scenario 1 and one for scenario 2. The one proposed for scenario 2 has drastically improved the performance, however we still need to check how much the gap with the upper bound reduced. We first note that the gap is clearly still very large in the case of Scenario 1 since CLA cannot be used.

To answer this question in the context of scenario 2, we will compare the performance of the centralized scheduler with the benchmark schedulers when they use CLA. To do so, we revert to the snapshot model. Table 3.3 shows the average performance ratio of the benchmark schedulers to the centralized scheduler for different values of N, the total number of users in the system. Note that the results with CLA show the ideal case where all channel gain information is present at the C-RAN.

Table 3.3: Averaged performance ratio of the local schedulers with CLA to the centralized scheduler

Total number of users in the system	**45**	**90**	**135**
RR	66.4	68.1	68.9
LBM	68.5	70.5	70.9

Table 3.3 shows that the schedulers perform much better when using CLA than when

they do not (recall that the case without CLA was reported in Table 3.2). However, the difference with the centralized scheduler is still significant, i.e., it is around 30%. Therefore, although the proposed LA and CLA schemes can improve the performance of simple benchmark schedulers significantly and are very easy to implement, there is still room for improvement. In the next section, we will propose a new uplink scheduler that is inspired by SFR and uses the LA scheme.

3.6 An SFR-based Practical Scheduler

In the previous sections, we have focused on how to improve the performance of existing local uplink schedulers. The local benchmark schedulers assume that each user transmits at full power and shares its power equally at a given time over all the allocated PRBs. We now propose an SFR scheme that is adapted to the uplink along with its scheduler.

3.6.1 Adapting SFR to the Uplink

The local schedulers we have studied so far use simple power allocation schemes. We will use SFR as the power allocation scheme for our proposed local scheduler that we call SFR-S in the following. The main idea behind SFR on the downlink is to restrict a priori, using a power-map, the power that can be used on different sets of channels such that neighboring BSs do not use the same power on the same set of channels. We propose to extend this notion of the power-map to the uplink, recognizing that contrarily to the downlink, each user brings its own power on the uplink. In the following, we will 1) define the concept of power-map on the uplink, 2) show how to parametrize this power-map in a robust fashion, i.e., this power-map has to work well on a large set of global realizations, 3) propose a practical scheduler to work on that power-map.

Structure of the power-map: We fix the power values used on each sub-channel by each BS. We call such a power allocation scheme a *power-map* that will be made public, i.e., each BS knows the power-map of all the BSs. The structure of our power-map is characterized by the following parameters: $\{m, n, p, P_1, \ldots, P_n\}$. To begin with, we use a coloring scheme[7] to partition the set of cells $\mathcal{K}$ into m colored subsets $\{\mathcal{K}_1, \ldots, \mathcal{K}_m\}$, so that the cells that create high interference to each other are put in different subsets. Similarly, we partition the set of subchannels $\mathcal{M}$ into m equal subsets $\{\mathcal{M}_1, \ldots, \mathcal{M}_m\}$. We call each of these subsets a sub-band. Each cell set $\mathcal{K}_j$ uses the same sub-band as its primary (a

[7]Note that the reuse factor is one and the coloring scheme is used to allocate power on sub-bands.

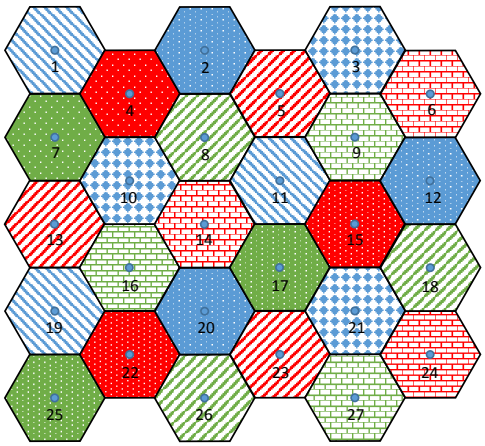


Figure 3.9: Coloring of the cells when $m = n = 3$. The cells with the same color use the same primary band. The cells with the same color and pattern use the exact same power-map.

cell uses most of its power on its primary sub-band) and the other sub-bands as secondary. We denote the subchannels belonging to the primary (resp. secondary) sub-band(s) of cell k with $\mathcal{M}_P(k)$ (resp. $\mathcal{M}_S(k)$). The cells use a very low per subchannel power p on their secondary sub-bands. By doing so, we decrease the impact of high interference received from the first tier neighboring cells since they will use different primary sub-bands.

Now, we consider the colored set $\mathcal{K}_j$ along with its primary sub-band. We further partition $\mathcal{K}_j$ into n subsets $\{\mathcal{K}_{j,1}, \ldots, \mathcal{K}_{j,n}\}$. We also partition its primary sub-band into n equal sub-bands $\{\mathcal{M}_{j,1}, \ldots, \mathcal{M}_{j,n}\}$. A cell j in subset $\mathcal{K}_{j,k}$ uses a per subchannel power $P_{k(mod\ n)}$. Therefore, the cells in $\{\mathcal{K}_{j,1}, \ldots, \mathcal{K}_{j,n}\}$ uses n different per subchannel powers on their primary sub-band with a circular shift. The reason for this second level of division is to balance the interference among the cells in the same colored set to improve the performance. An example power-map with $m = 3$ and $n = 3$ is shown in Fig. 3.10 for the 27-cell example network given in Fig. 3.9 where the cells of a same color belong to the same $\mathcal{K}_j$ (i.e., have the same primary band) and within a color, the ones exhibiting the same pattern belong to the same $\mathcal{K}_{j,k}$, i.e., uses exactly the same powers in the primary band while cells with different patterns see shifted powers in the primary band.

Each primary sub-band consists of $S = \frac{M}{m*n}$ subchannels and we will show how to select the values of the power levels $P_1, \ldots, P_n$ and the secondary sub-band power value p in the following.

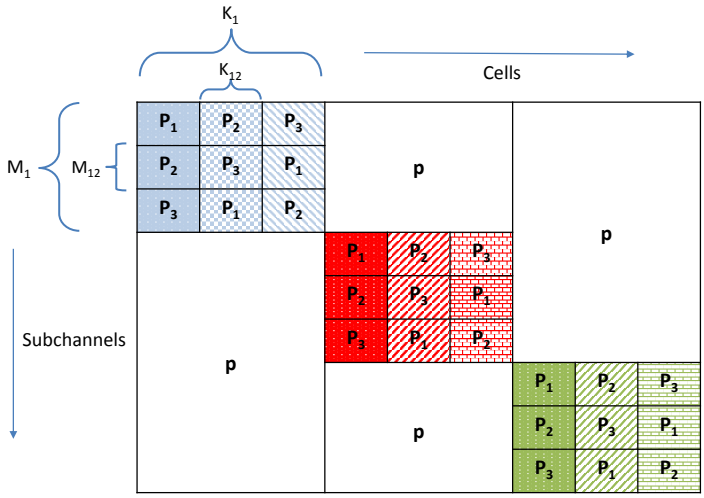


Figure 3.10: The power-map when $m = n = 3$ with the color scheme in Fig. 3.9

Once a power-map structure is decided, i.e., as it was done to obtain the structures in Fig. 3.9 and Fig. 3.10, the only parameters left to choose are the power values.

3.6.2 SFR Parametrization

Our aim here is to find a robust power-map with the structure described above, i.e., one that maximizes the average GM goodput over a large number of realizations (with different number of users). So in the following, we assume that m and n are given as well as the complete structure, i.e., the sets $\mathcal{K}_j$'s, $\mathcal{M}_j$'s and $\mathcal{M}_{j,k}$'s, and we look for values of $p, P_1, ..., P_n$ that are *robust*. This is what we call parametrizing the power-map, To do so, we will consider a snapshot model and we use the optimization problem $\mathbb{P}_{\mathbb{S}}(\omega)$, which was formulated in Section 3.4, with some slight modifications.

The first modification is done in order to maximize the average GM goodput over a set of global calibration realizations Ω_c. The reason we call them calibration realizations is that we only use them to parametrize the power map, but we do not make the performance evaluation with these realizations, i.e., we test the power-map with a different set of realizations to show its robustness. To this end, we maximize the following objective function:

$$\mathcal{Z}(\Omega_c) = \frac{\sum_{\omega \in \Omega_c} \left(\prod_{u \in \mathcal{U}_k(\omega)} \prod_{k \in \mathcal{K}} \lambda_u^k(\omega) \right)^{\frac{1}{|\mathcal{U}(\omega)|}}}{|\Omega_c|} \tag{3.50}$$

This function is basically the arithmetic mean of the GM goodput over the calibration realizations $\omega \in \Omega_c$.

The second modification is on the power constraints. Using the given power-map structure, we know which of the power levels $\{p, P_1, ..., P_n\}$ is used on which subchannel in each cell. Let $P_k(c)$ denote the power level used on sub-channel c in cell k in our power-map. Then, we add the following constraints:

$$\sum_{u \in \mathcal{U}_k} p_{u,k}^{c,t}(\omega) \leq P_k(c), \ \forall c \in \mathcal{M}, k \in \mathcal{K}, t \in \mathcal{T}, \omega \in \Omega_c \tag{3.51}$$

which denotes that the total power used by any user in realization ω cannot exceed $P_k(c)$ on subchannel c of cell k. Note that $P_k(c)$ corresponds to one of the power levels P_1, P_2, P_3, or p depending on the cell k and subchannel index c. For example, if we consider the power-map in Fig. 3.10, the first subchannel corresponds to P_1 for the cells in cell set $K_{1,1}$, P_2 for the cells in cell set $K_{1,2}$, P_3 for the cells in cell set $K_{1,3}$, and p for the remaining cells.

Finally, we formulate the following problem that maximizes the average GM goodput over the calibration realizations in Ω_c given the power-map structure presented above.

$$\mathbb{P}_{\mathbb{SFR}}(\omega): \max_{(p_{u,k}^{c,t}(\omega), \lambda_u^k(\omega), y_{u,k}^{c,t}(\omega), Q_{u,k}^{c,t}(\omega), r_{u,k}^{c,t}(\omega), p, P_j)} \mathcal{Z}(\Omega_c) \tag{3.52}$$

$$s.t. \ \text{constraints (3.10-3.18) and (3.51)} \tag{3.53}$$

This is basically the same problem as $\mathbb{P}_{\mathbb{S}}(\omega)$ but over many realizations and with additional power constraints. This problem can be solved with the method discussed in Section 3.4. However, it is computationally very expensive to solve it that way since we consider many realizations at the same time. To tackle this, we use a brute-force search method. We assume the power per subchannel values for the sub-bands P_1, P_2, P_3, p can take values in the form of P_U/k, where k is an integer. Then, we try all possible values of k on each sub-band and solve all the realizations separately with these power values and get the average GM goodput over all calibration realizations, for each power value combination. Then, we pick the power-map that gives the largest average GM goodput.

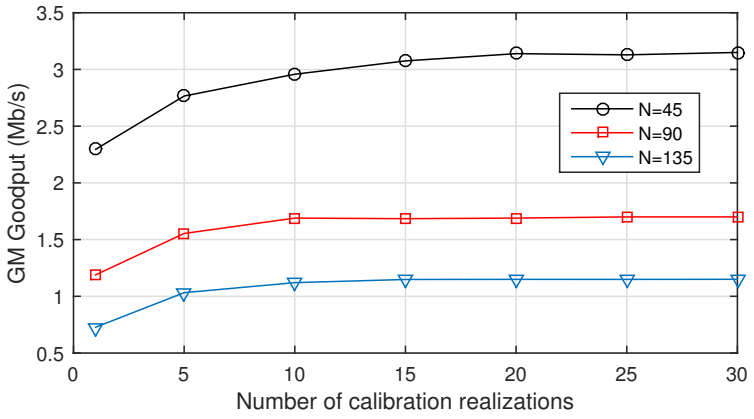


Figure 3.11: Performance of the power-maps obtained with different number of calibration realizations. Each power-map is tested with a separate 100 test realizations.

Note that we only try very high values of k for the secondary sub-bands since they use low power by construction.

We consider the same 9-cell network that we have used before. In this case, we use the power-map depicted in Fig. 3.10. We consider a number H of calibration realizations, each with a number of users selected at random between 45 to 135, which corresponds to 5 to 15 average number of users per cell. Then, we solve $\mathbb{P}_{\mathbb{SFR}}(\Omega_c)$ with the brute-force search method.

We have tried different values of H and for each of them we obtain a different robust power-map, i.e., values for $\{P_1, P_2, P_3, p\}$. We test these different power-maps on 100 test realizations corresponding to different values of the total number of users in the system. Fig. 3.11 shows the results in terms of the GM goodput averaged over the 100 test realizations as a function of H. Clearly, using a calibration set of size larger than $H = 20$ does not provide an increase in performance.

The power values we obtain for $H = 20$ are as follows:

$$\{P_1, P_2, P_3, p\} = \{P_U, 0.17P_U, 0.14P_U, 0.0034P_U\} \tag{3.54}$$

3.6.3 The SFR-S Scheduler

We now propose SFR-S, the online scheduler that goes hand to hand with the power-map computed above. One important difference between SFR-S and the previous benchmark scheduler is that the BSs now collect measurements for each sub-band and create a CDF per sub-band. Then, the rate of each user on each subchannel of each sub-band can be computed.

SFR-S keeps track of what each user i has received since the last event (arrival or departure in the cell). Let $\tau_i(t)$ be the amount of bits sent by i since the last event till t. SFR-S aims at being fair for the group of users currently in the system. Therefore, we reset the variables $\tau_i(t)$ to a small non zero value for all users i in cell k after an arrival or departure event in that cell. In a subframe t, SFR-S (see Algorithm 2) starts by allocating the PRBs with the lowest power. A PRB $\{c, t\}$ is allocated to the user i that sees the highest $R_i^c/\tau_i(t)$ and then the corresponding $\tau_i(t)$ is updated accordingly.

Algorithm 2 Scheduling algorithm for the proposed scheduler in cell l at subframe t

1: $\tau_i(t)$ is the total data uploaded by user i until t.
2: Set c to 1.
3: Sort the subchannels according to power per subchannel in an ascending order
4: Set $P_r(i) = P_U$, which is the remaining power of user i
5: **while** $c \leq M$ **do**
6: $\quad i^* = \underset{i \in \mathcal{U}_l : P_r(i) \geq P_c}{\operatorname{argmax}} \; R_i^c/\tau_i(t)$
7: $\quad$ Allocate PRB $\{c, t\}$ to user i^*
8: $\quad \tau_i^*(t) = \tau_i^*(t) + R_{i^*}^c, \quad c = c + 1$
9: $\quad P_r(i^*) = P_r(i^*) - P_c$
10: **end while**
11: After each arrival or departure at cell l, set $\tau_i(t)$ to δ for all the remaining users in the cell.

3.6.4 Numerical Results

Next, we compare the delay performance of RR, LBM, and SFR-S in a dynamic setting where each user uploads 10 Mb of data first for scenario 1 and then schenario 2. Fig. 3.12 shows the results for scenario 1. Note that the LA scheme is used in all schedulers in order to make a fair comparison.

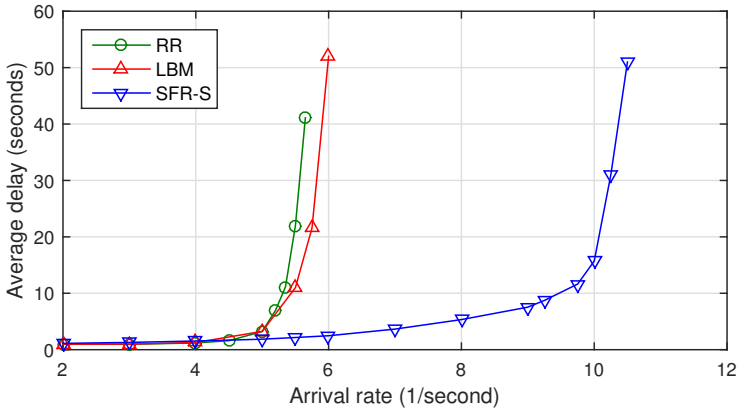


Figure 3.12: Scenario 1: Average delay performance of the RR, local benchmark scheduler and SFR-S when the users upload 10 Mb

Clearly, SFR-S drastically outperforms the two benchmark schedulers in terms of delay. It can support an arrival rate of 10 users/second, while the benchmark scheduler can only support up to 6 users/second. For a target average delay of 10 seconds, the arrival rate is around 5 users/second for RR scheduler, 5.5 users/second for LBM and 9.5 users/second for SFR-S. In terms of losses, SFR-S also performs slightly better than the benchmark schedulers with around 11% PRB loss rate while this number is around 15% for the benchmarks.

Next, we show, in Fig. 3.13, how the performance of RR, LBM, and SFR-S can be further improved when CLA is used in the ideal case (scenario 2). Clearly CLA significantly improves the delay performance of all schedulers. Note that even without CLA, SFR-S outperforms the other two schedulers when they use CLA.

3.6.5 Closing the Gap?

In this section, we have proposed a complete suite of schemes based on SFR that outperforms drastically the benchmark (even when it is enhanced with our proposed schemes). Have we finally reduced the gap with the upper bound enough?

To answer this question, we go back to the snapshot model to compare the performance of fully centralized scheduler with the performance of SFR-S for both aforementioned scenarios. Fig. 3.14 shows the results for the centralized scheduler and for SFR-S without

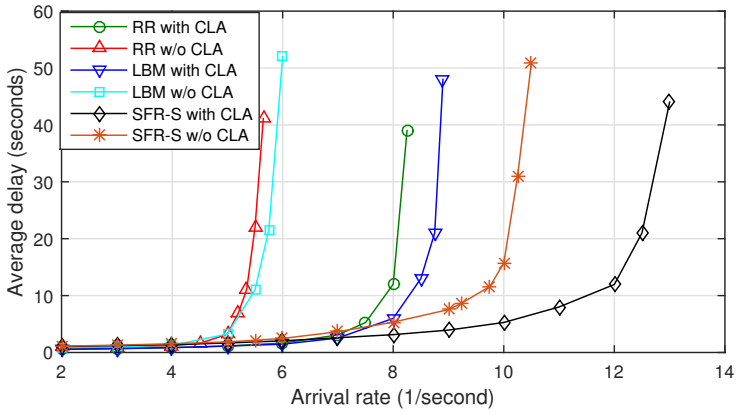


Figure 3.13: Scenario 2: Average delay performance of RR, LBM and SFR-S with and without CLA when the users upload 10 Mb

CLA and with (the ideal case of) CLA as a function of N, the total number of users in the system.

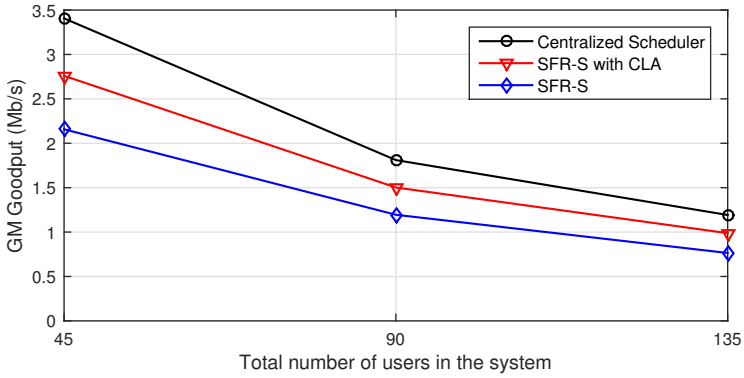


Figure 3.14: GM Goodput for the centralized scheduler and SFR-S with and without CLA as a function of the number of users for the ideal case

The difference between SFR-S with CLA and the centralized scheduler is about 17% (it was more than 30% for the benchmark scheduler with CLA). Since SFR-S and CLA are very fast algorithms than can easily be implemented in real time, we think that SFR-S with CLA is a promising scheme for the uplink.

3.7 Conclusion

In this chapter, we have focused on uplink scheduling. We first have studied a local benchmark scheduler and explored its goodput and loss performance. Then, we have showed how it performs compared to the maximum achievable performance. For that, we have formulated and solved a centralized scheduling problem. To improve the performance of the benchmark scheduler, we have proposed two schemes, one for the scenario without C-RAN and one for the scenario with C-RAN. We have shown that a CDF-based interference estimation coupled with a smart MCS selection improves the delay and loss performance in the first scenario while CLA does even better in the case of scenario 2. Although the proposed solutions are very simple and they have improved the performance of the benchmark scheduler significantly, there is still a performance gap compared to the centralized scheduler. Then, we have proposed a new scheduler that is inspired by SFR, which we parametrize in a robust fashion. We have shown that the proposed scheduler performs remarkably well both in dynamic and static settings and performs only 17% worse than a fully centralized scheduler with negligible complexity compared to the centralized scheduler.

Chapter 4

Downlink Scheduling in a Multi-Cell OFDMA Network: From Full Base Station Coordination to Practical Schemes

4.1 Introduction

In the previous chapter, we have studied uplink scheduling in detail. While the importance of uplink scheduling has increased recently, most of the cellular traffic is still downlink [5]. As a result, most of the work in the literature focuses on the downlink rather than the uplink and the literature on downlink scheduling is very abundant. However, we considered it useful to revisit downlink scheduling with a similar methodology and research questions as we had for the uplink scheduling. In this chapter, we will perform a similar analysis for downlink scheduling, where we first formulate and study a centralized scheduler and then revisit practical schedulers.

There are some crucial differences between user scheduling on the downlink and the uplink. First of all, the power management is different since the BSs are the only sources of power on the downlink, whereas each user brings its own power budget on the uplink. Second, the interference management is different. The interferers in the neighboring cells are not known beforehand on the uplink since the schedules of the neighboring BSs are unknown. However, the interferers are always the neighboring BSs on the downlink. Hence, interference management is simpler as long as we assume that the BSs always transmit on

their subchannels, which is a reasonable assumption with the full buffer traffic model. For example, if we assume that the power per subchannel used by each BS is known beforehand, we can compute the exact interference at each user on the downlink, while it is not possible on the uplink. We assume as in the previous chapter that the channel gains are known perfectly and hence under this assumption, SINRs can be computed at each user exactly (because MCSs can be selected using the default selection method) and there will not be PRB losses.

In this chapter, we will focus on research questions similar to the one we had for uplink scheduling: Do the existing practical schedulers perform well enough compared to the maximum achievable performance? If not, can we design better practical schedulers?

Similar to the uplink, the best performance on the downlink can be obtained only when all the cells are scheduled simultaneously in a C-RAN. The problem formulation for the centralized downlink scheduler is very similar to the one for the uplink scheduler except some minor differences in the problem constraints. In the following, we will first formulate the scheduling problem and then solve it with the exact same method used for the centralized uplink scheduling problem. We will first consider a homogeneous network, and then a HetNet, that includes small cells together with the macro BSs. The purpose of deploying small cells is to decrease the load of the macro BSs and hence to improve the coverage and throughput performance of the cellular networks. HetNets bring additional challenges such as channel allocation (e.g., should channels be different for the macro and the small cells? How many channels should be allocated to the small cells?) and user association. Such processes are rathe simpler processes in homogeneous networks. In the following, we will show the impact of different types of channel allocation and user association methods together with user scheduling in HetNets.

Due to its computational complexity and the need to acquire cross-channel information, we again believe that the centralized scheduler is not easy to implement and hence we will use it to obtain an offline upper bound. Our focus will be on practical local schedulers. As our benchmark, we consider a local scheduler that uses equal power on its subchannels and allocate equal time to each user [34]. It is commonly used by the network operators in today's networks. We show that its performance is significantly worse than the performance of the centralized scheduler. Therefore, we focus on finding a practical local scheduler that is simple to implement and that performs better.

To this end, we focus on SFR-based schemes. SFR relies on a very loose coordination among the BSs, where each BS agrees to use a different pre-defined power on different sets of subchannels. This allows the BSs to serve their *bad* users, i.e., users with poor channel conditions, on the subchannels, which use high power and receive low interference, to

improve fairness and efficiency. While SFR is often suggested to improve the performance on the downlink, it is not trivial to define a power allocation structure, i.e., a power-map, and also to parametrize it.

Our next step is to extract, in the homogeneous case, from these feasible solutions a practical scheme based on a well-parametrized soft frequency reuse (SFR) along with its local scheduler (Contribution 1). Similar to the previous chapter, our aim is to create a per BS power-map that each BS has to follow while scheduling locally its users. We first define the structure of our power-map, which is a more generalized version of the commonly proposed 2-power 3-band SFR power-maps, and then we perform a robust parametrization using a brute-force search in order to find a robust power-map that works well over many different realizations with different number of users. We show that the practical local scheme we propose performs significantly better than a benchmark based on round-robin scheduling, which is the most commonly used scheduler in today's networks, and does only 20% worse than the centralized case. Since the centralized problem takes a very long time to solve, it would have to be significantly approximated to be viable as a practical solution which would probably reduce its performance and hence, it might not be worth it to try to implement scheduling in a C-RAN for the downlink.

Finally, we consider centralized scheduling in a HetNet (Contribution 2). A HetNet brings additional challenges due to heterogeneity between the BSs. For example, channel allocation between the small cells and the macro base stations need to be done carefully. We apply our proposed method to find an upper bound for the global problem (assuming scheduling is done centrally for all BSs) and compare this upper bound with practical scheduling and channel allocation schemes that are inspired by the practical scheme proposed for the homogeneous case[1].

4.2 Related Work

Downlink user scheduling in OFDMA networks has been widely studied in the literature. While some of the earlier work such as [36] considers a single cell model, the inter-cell interference has a significant impact on the system performance [37] since it directly affects the rate each user sees. Therefore, interference management is a crucial part of the user scheduling process.

Interference management can be implemented in different ways. The maximum performance can be obtained only when scheduling is done centrally; however, solving the

[1]Some of these results are presented in our work [35]

centralized global optimization problem is very difficult in that case since it is a very large scaled MINLP. Most of the centralized schedulers, such as [38] and [39], are designed for simpler schemes than OFDMA, e.g., TDMA or CDMA.

To decrease the computational complexity and the amount of data exchanged between the cells, a semi-distributed user scheduling is performed in [40], which allows each BS to collaborate with other cells while scheduling by exchanging important data to balance the interference on the resource blocks. More specifically, each user sends some information to the BS that it is associated to including the channel state information and the two neighboring BSs that create the most dominant interference for that user. A central entity receives information from all the BSs and then the BSs agree on using lower power on certain PRBs. Although this decreases the complexity compared to a fully centralized scheduler [41], this is still very complex in terms of communication overhead; therefore simpler coordination mechanisms are proposed in [42] and [43]. In both cases, scheduling is done locally at each BS. However, the BSs still exchange a large amount of information such as the channel state information of their users.

Typically, each BS schedules its own users independently from the neighboring BSs in practical systems. This eliminates the necessity for the BSs to communicate with each other. Moreover, it decreases the size of the scheduling problem drastically and hence it can be performed very fast. In a simple implementation, the BSs allocate equal power to all its subchannels and serve its users to maximize a PF objective. It has been shown in [44] that the optimal scheduler for this case allocates equal time to each user, i.e., in round-robin (RR) fashion. In the following, we will call this scheduler RR and consider it as our benchmark and show that its performance is significantly worse than the centralized scheduler.

One way to perform interference management in a practical system is the static interference coordination, which is also called frequency reuse [41]. It allows the cells to schedule their users locally without any fine-grained coordination. The main idea behind frequency reuse is to predetermine the power used on each subchannel at each cell so that the interference management becomes simpler and the users with worse channel conditions can be served with the subchannels that use higher power with lower interference. It is much easier to implement than the centralized schedulers and it has been shown that some special reuse schemes such as SFR perform well especially for the edge users [45].

Strict frequency reuse directly divides the set of subchannels into some orthogonal subsets so that each cell can only use one of these subsets to reduce interference. On the other hand, SFR allows all the cells to use all the subchannels, but it limits the power that can be used on some of the subchannels so that non-edge users are served on the

subchannels with high power and low interference. It is a more efficient technique than the strict frequency reuse as it uses the whole bandwidth at each BS [46]. Most of the previous studies do not take full advantage of the SFR concept, by limiting the number of sub-bands to three at most, and the number of power levels to two (i.e., high or low). In this chapter, we adopt a modified version of SFR, where all BSs use the entirety of the available bandwidth, and are not limited to two power levels only. By dividing the frequency band into a number of sub-bands, each BS can transmit on each sub-band with a different power level, adding more flexibility to the traditional SFR scheme.

Authors of [47] and [48] proposed adaptive SFR schemes that dynamically adapt to changing traffic load and user distribution among neighboring cells. Their schemes are based on a continuous exchange of information between neighboring BSs. Such adaptive techniques further raise questions on communication overhead, and on reactivity in adapting the reuse scheme to the actual network state, in real-time. To avoid this problem, we will focus on finding a *robust* power-map that works well for different realizations and load distributions.

In a HetNet that includes small cells together with the macro base stations, scheduling has to be considered jointly with channel allocation and user association [49]. The difficulty of channel allocation and user association processes in HetNets arises from the heterogeneity of the BSs, which does not exist in homogeneous networks. Although there is a large literature dedicated to HetNets including [34], [50], [51], [52], [53], [54], [55], the problem of providing the maximum performance in terms of channel allocation, user scheduling, and user association via full coordination has not been solved.

4.3 The Homogeneous Case

4.3.1 The Global Centralized Scheduling Problem Formulation

In this section, we formulate the system-wide user scheduling problem for the downlink of a homogeneous cellular network. For a given realization ω (i.e., the channel gains $\{G_{u,k}\}$) and the sets $\mathcal{U}_k$'s, we formulate the following system-wide problem $\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega)$ that maximizes the proportional fair objective function over all users in the system:

$$\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega): \max_{(P_{u,k}^{c,t}),(\lambda_u^k),(x_{u,k}^{c,t,l}),(I_{u,k}^{c,t}),(R_{u,k}^{c,t})} \sum_{u\in\mathcal{U}_k}\sum_{k\in\mathcal{K}} \log(\lambda_u^k) \tag{4.1}$$

$$\lambda_u^k = \frac{1}{T}\sum_{c\in\mathcal{M}}\sum_{t\in\mathcal{T}} R_{u,k}^{c,t}, \quad \forall u\in\mathcal{U}_k,\ k\in\mathcal{K} \tag{4.2}$$

$$R_{u,k}^{c,t} = \sum_{l\in\mathcal{L}} x_{u,k}^{c,t,l}\vartheta_l, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{4.3}$$

$$S_{u,k}^{c,t} \geq x_{u,k}^{c,t,l}\eta_l,\ \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T}, l\in\mathcal{L} \tag{4.4}$$

$$S_{u,k}^{c,t} = \frac{P_{u,k}^{c,t}G_{u,k}}{\mu^{DL}+I_{u,k}^{c,t}}, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{4.5}$$

$$I_{u,k}^{c,t} = \sum_{j\in\mathcal{K}, j\neq k}\sum_{v\in\mathcal{U}_j} P_{v,j}^{c,t}G_{u,j}, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{4.6}$$

$$\sum_{u\in\mathcal{U}_k}\sum_{c\in\mathcal{M}} P_{u,k}^{c,t} \leq P_{BS}, \quad \forall k\in\mathcal{K}, t\in\mathcal{T} \tag{4.7}$$

$$P_{u,k}^{c,t} \leq \sum_{l\in\mathcal{L}} x_{u,k}^{c,t,l}P_{BS}, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{4.8}$$

$$x_{u,k}^{c,t,l} \in \{0,1\}, \quad \forall u\in\mathcal{U}_k, k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{4.9}$$

$$\sum_{u\in\mathcal{U}_k}\sum_{l\in\mathcal{L}} x_{u,k}^{c,t,l} \leq 1, \quad \forall k\in\mathcal{K}, c\in\mathcal{M}, t\in\mathcal{T} \tag{4.10}$$

Note that the problem formulation is almost the same as the centralized scheduling problem for the uplink. The only differences are constraint (4.6) since the interference is measured at each user on the downlink while it is measured at each BS on the uplink and constraint (4.7) since the BS has a single power budget on the downlink while each user has its own power budget on the uplink. As a result, we can solve $\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega)$ with the exact same transformations and the iterative algorithm we used to solve the centralized uplink scheduling problem.

4.4 Numerical Results on the Centralized Problem

In this section, we first provide results for the global problem to show the tight gap between solutions of the upper bounding signomial problem and the feasible solutions to $\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega)$ obtained from these solutions.

We then develop practical schemes in the next section using insights from the feasible solutions to $\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega)$ and compare their performance with the global optimal and the RR benchmark that is commonly used in the literature.

For the numerical results, we consider the exact same homogeneous system as in Chapter 3.

4.4.1 Tightness of the Upper Bound

Recall that we first transform the original centralized scheduling problem $\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega)$ to an upper bounding problem that we solve. We can then extract a feasible solution from the result of the upper bounding problem. Table 4.1 shows results on the tightness of the bound for different values of N. For each value of N, 100 realizations are computed and the table shows the arithmetic average of the relative difference in GM as well as the maximum relative difference over these 100 realizations. These results illustrate the tightness of the upper bounds. For example, when $N = 15$, the average difference is below 4%.

Table 4.1: Comparison of the upper bounds and the feasible solutions as a function of N

N: Avg. number of users per cell	5	10	15
Avg. Difference (%)	3.51%	3.62%	3.87%
Max. Difference (%)	6.67%	6.79%	6.94%

Note that the difference between the upper bounds and the feasible solutions is lower in the case of the downlink than the uplink.

4.4.2 Comparing the Upper Bound with the RR Benchmark

In the following, we will consider practical local schedulers, where each BS schedules its own users independently from the other BSs. It is important to note that the exact interference can be computed on the downlink with pre-computed power values since the transmitters in each cell are known and not changing. Then, we can compute the rate each user sees and schedule the users accordingly.

The first practical scheme we consider is based on each BS agreeing *beforehand* to allocate equal power to all its subchannels. Clearly this agreement can be seen as a rough level of BS-coordination. In that case, the scheduler is local and has been shown to be equivalent to a round-robin [34].

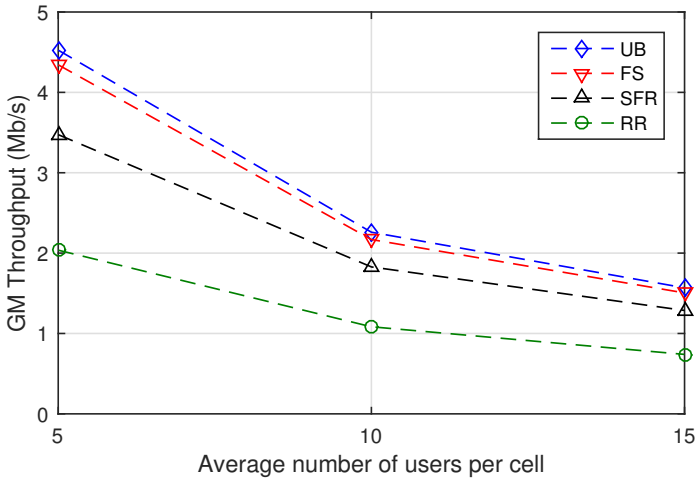


Figure 4.1: GM Throughput comparison as a function of average number of users N per cell

Fig. 4.1 shows results for this scheduler (labeled RR), for the upper bound (labeled UB) and for the feasible solutions obtained from the upper bounds (labeled FS). The fourth curve in that figure (labeled SFR) will be discussed in the next section. Specifically, it shows the arithmetic mean of GM throughput over 100 realizations as a function of N. The relative difference in the geometric mean is around 120% in average between the upper bound and the equal power scheduler. Clearly, the RR benchmark performs poorly when compared to the best achievable performance.

4.5 Design of Efficient and Practical Schemes

We now focus on designing SFR-based scheduling schemes that are practical, i.e., rely on local scheduling and perform better than round-robin. Note that we do not focus on finding the optimal SFR scheduler but our aim is to show the existence of an SFR scheme that performs very well when compared to the optimal scheduler. We explain next how we design a practical scheme.

We start by examining feasible solutions to the centralized scheduling problem for many realizations. We see that good users are given a lot of subchannels each with very low power and there are in fact rare cases where some subchannels are not used. This is clear in the

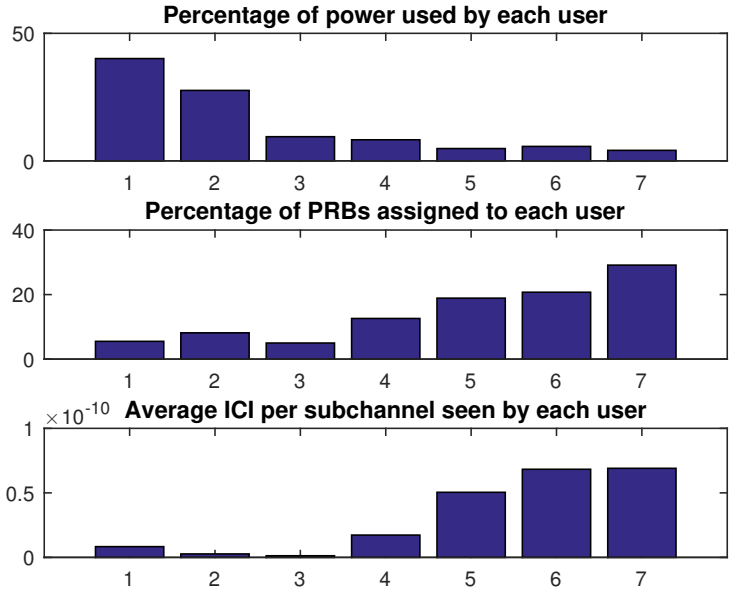


Figure 4.2: Results of one sample realization for one cell. Users are sorted w.r.t their channel gains in an ascending order

realization represented in Fig. 4.2. In this figure, we compare the amount of power, the number of PRBs, and the interference each user sees. There are 7 users in one cell and the users are sorted w.r.t. their channel gains in an ascending order. Clearly, user #7 sees a great channel and gets very little power and many PRBs. On the other hand, user #1 sees a bad channel, gets very high power and very few PRBs.

Based on these insights, we design an SFR-based power-map structure [45], a way to parametrize it, and its associated scheduler. SFR simply divides the subchannels into sets of subchannels, we call them sub-bands, which are used with different powers by different BSs. Our goal is to propose a well parametrized robust SFR scheme as well as the local scheduler that goes along. We first present how we characterize an SFR scheme, then the structure of our SFR scheme, followed by the presentation of the local scheduler that goes with it and finally we show how to parametrize the scheme.

How To Characterize an SFR Scheme: While there are many ways to describe an SFR scheme, we have chosen to characterize it by the following elements:

- The number B of channel sub-bands partitioning the available spectrum made of M channels. We assume that the sub-bands are all of the same size which means that M is divisible by B. Variations can be designed if this is not the case.

- The set of B power levels P_j ($P_j \geq P_{j+1}$). P_j is a per-channel power. The same power level is used on each subchannel of a sub-band for a given cell.

- The possible mappings between sub-bands and power levels, typically there are B mappings that are shifted versions of $(P_1, P_2, \ldots, P_B)$ where this B-uple has to be interpreted as P_1 for sub-band 1, P_2 for sub-band 2 and so on. A shifted version of this B-uple is for example $(P_2, P_3, \ldots, P_B, P_1)$.

- The partitioning of the set $\mathcal{K}$ of BSs into B CELL SETS ($\mathcal{C}_j$), where all the BSs in a CELL SET use the same mapping between power and sub-bands.

Hence, **parametrizing an SFR scheme means selecting values for B, the (P_j)'s and selecting the partition ($\mathcal{C}_j$)'s.**

Let's now discuss one example that we used as a building block to our scheme. It uses reuse-1 with $B = 3$. In that case, we used a partition corresponding to a 3-color scheme. We have tried all the values of P_1 and P_2 (note that $P_1 \geq P_2 \geq P_3$ and $P_{BS} = \frac{M(P_1+P_2+P_3)}{B}$ where P_{BS} is the transmit power of a BS). We generated for different values of $|\mathcal{U}|$, the number of users in the system, 100 realizations. For each realization, we computed the geometric mean (GM) for each power pair (P_1,P_2) and then selected for each $N = \frac{|\mathcal{U}|}{|\mathcal{K}|}$, the pair $\pi(N)$ that gives the highest arithmetic mean of the GMs. The best choice of power levels was such that $P_1 >> P_2 = P_3 \neq 0$. Roughly, each BS transmits with high power on one third of the channels to its bad users and transmits on two thirds of the channels at very low power to its good users. This scheme does better than a scheme based on reuse-1 and equal power by about 35%.

The structure: We are now ready to describe the proposed scheme, which uses $B = 9$[2]. Its structure is represented in Fig. 4.3 (we assume that M is divisible by 9). Note that we have reduced the number of power levels to four and that $6P_4 + P_1 + P_2 + P_3 = P_{BS}\frac{B}{M}$ and $P_1 > P_2 > P_3 > P_4$).

Clearly, now for a given system, we need to map the BSs to the 9 sets to populate the CELL SETS and select the values for P_1 to P_4 Once this is done offline, the BSs

[2]We tried lower values for B such as 3 and 6, however their performances were worse than the one we obtained with $B = 9$

	SB 1	SB 2	SB 3	SB 4	SB 5	SB 6	SB 7	SB 8	SB 9
CELL SET 1	P_1	P_2	P_3	P_4	P_4	P_4	P_4	P_4	P_4
CELL SET 2	P_2	P_3	P_1	P_4	P_4	P_4	P_4	P_4	P_4
CELL SET 3	P_3	P_1	P_2	P_4	P_4	P_4	P_4	P_4	P_4
CELL SET 4	P_4	P_4	P_4	P_1	P_2	P_3	P_4	P_4	P_4
CELL SET 5	P_4	P_4	P_4	P_2	P_3	P_1	P_4	P_4	P_4
CELL SET 6	P_4	P_4	P_4	P_3	P_1	P_2	P_4	P_4	P_4
CELL SET 7	P_4	P_4	P_4	P_4	P_4	P_4	P_1	P_2	P_3
CELL SET 8	P_4	P_4	P_4	P_4	P_4	P_4	P_2	P_3	P_1
CELL SET 9	P_4	P_4	P_4	P_4	P_4	P_4	P_3	P_1	P_2

Figure 4.3: Power allocation for the heuristic schedulers for reuse-1

will schedule locally their users with the power-map corresponding to their CELL SET, knowing the power used on each sub-band by other BSs. For example, if BS j belongs to CELL SET 6, it will use power P_4 on sub-bands $SB1$, $SB2$, $SB3$, $SB7$, $SB8$, $SB9$, power P_1 in $SB5$, power P_2 in $SB6$ and P_3 in $SB4$. Hence, as for the benchmark which used a very special case of power-map, there is an offline process during which the BSs agree on which power level to use on which channel. This can be seen as a rough-grained BS coordination. Then each BS acts independently and locally.

Note that this power-map structure is very similar to the one we have presented in the previous chapter for uplink scheduling (see Fig. 3.9 and Fig. 3.10). Before explaining how we have selected the power levels and the partition of $\mathcal{K}$, we present the local scheduler that is used along with the SFR scheme.

The local scheduler: Note that when we know the power used in each cell, i.e., for a given value of (P_1, P_2, P_3, P_4), we can compute the exact interference on each subchannel for a user, and we can compute the rate each user associated with a BS gets on each sub-band. Then, each cell can schedule their users independently. Assume R_u^s is the per-subchannel rate of user $u \in \mathcal{U}_l$ on sub-band s. Then, the following problem maximizes the proportional fair objective function in cell l. Given our SFR pattern (represented by Fig. 4.3, the CELL SETS and (P_1, P_2, P_3, P_4)), a given cell l and its CELL SET, its set of associated users $\mathcal{U}_l$, the (R_u^s)'s (which define the local realization ω_l), the scheduler can be formulated as

follows:

$$\max_{(x_u^{c,t}),(\lambda_u \geq 0)} \sum_{u \in \mathcal{U}_l} \log(\lambda_u) \tag{4.11}$$

$$s.t.\ \lambda_u = \sum_{s \in \{1,\ldots,9\}} \sum_{(c,t) \in s} x_u^{c,t} \frac{R_u^s}{T}\ ,\ \forall u \in \mathcal{U}_l \tag{4.12}$$

$$\sum_{u \in \mathcal{U}_l} x_u^{c,t} \leq 1,\ \forall (c,t) \in \mathcal{MXT} \tag{4.13}$$

$$x_u^{c,t} \in \{0,1\}\ \forall (c,t) \in \mathcal{MXT} \tag{4.14}$$

This is a small MINLP problem that can be replaced by a simpler convex problem by noting that:

$$\lambda_u = \sum_{s \in \{1,\ldots,9\}} \sum_{(c,t) \in s} x_u^{c,t} \frac{R_u^s}{T} = \sum_{s \in \{1,\ldots,9\}} \frac{R_u^s}{T} \sum_{(c,t) \in s} x_u^{c,t}$$

yielding the following relaxed version that allocates all of a sub-band s to user u for a proportion of time α_u^s.

$$\mathbb{P}_{\mathbb{SFR}}^{\mathbb{DL}}(\omega_l) : \max_{\alpha_u^s \geq 0, \lambda_u \geq 0} \sum_{u \in \mathcal{U}_l} \log(\lambda_u) \tag{4.15}$$

$$s.t.\ \lambda_u = \sum_{s \in \{1,\ldots,9\}} \alpha_u^s R_u^s \bar{s}\ ,\ \forall u \in \mathcal{U}_l \tag{4.16}$$

$$\sum_{u \in \mathcal{U}_l} \alpha_u^s \leq 1\ ,\ \forall s \in \{1,\ldots,9\} \tag{4.17}$$

where $\bar{s}$ is the size of a sub-band in number of subchannels ($\bar{s} = \frac{M}{B}$).

Note that $\mathbb{P}_{\mathbb{SFR}}^{\mathbb{DL}}(\omega_l)$ is a very small sized convex optimization problem that can be solved very fast using any commercial solver. However, it might still not be possible to use such a solver in a BS. Hence, we will next present an online SFR scheduler that does not require to solve the optimization problem above.

Online version of the SFR-based scheduler: Let's consider cell l. The scheduler is called at the beginning of each subframe t, whose duration is δ. Given the set of sub-bands $s \in \{1,\ldots,9\}$, the rate R_u^s for each user $u \in \mathcal{U}_l$ on each sub-band s, and the total amount of data, $\tau_u(t)$, user $u \in \mathcal{U}_l$ has downloaded so far, the following algorithm allocates a full sub-band at each subframe t to one user.

Algorithm 3 Online version of the SFR scheduler in cell l

1: At the beginning of each subframe t, set s to 1.
2: **while** $s \leq 9$ **do**
3: $\quad u^* = \underset{u \in \mathcal{U}_l}{\operatorname{argmax}}\ R_u^s / \tau_u(t)$
4: $\quad$ Allocate sub-band s to user u^* in subframe t
5: $\quad \tau_u^*(t) = \tau_u^*(t) + R_{u^*}^s \times \delta, \quad s = s + 1$
6: **end while**

This algorithm, which is very fast, gives results almost identical as the ones of $\mathbb{P}_{\mathbb{SFR}}^{\mathbb{DL}}(\omega_l)$ irrespective of the sub-band allocation order. For example, the difference in GM throughput is around 1% with 100 realizations when the average number of users per cell is 10.

Finalizing the parametrization of the robust SFR scheme: We allocate the cells to the CELL SETS as represented in Fig. 3.2 and use the same brute force approach to determine the power levels of a robust power-map (see Table 4.2).

Table 4.2: Power Levels for SFR

Sub-bands	$(MP_1)/9$	$(MP_2)/9$	$(MP_3)/9$	$(2MP_4)/3$
Power Values	$0.644P$	$0.201P$	$0.121P$	$0.034P$

We are now ready to revisit Fig. 4.1 where the black curve labeled SFR represents the performance of the SFR scheme. We find out that the proposed SFR scheme performs much better than the equal power scheduler. The difference between the upper bound and the SFR scheme is only around 20%, which indicates that even when reuse-1 is employed between the cells, an excellent performance can be obtained without full (fine-grained) coordination among cells. Remember that a full-grained coordination is complex to perform and would require a real-time solution that would offer a degraded performance w.r.t. to the upper bound and hence might not bring much gain when compared to the very practical SFR scheme.

4.6 The Heterogeneous Case

The methodology we have proposed above to compute an upper bound can be used for different network configurations. In this section, we consider a heterogeneous network (HetNet) that consists of small cells as well as macro BSs. HetNets bring additional

challenges in terms of channel allocation between the small cells and the macro cells, user association, and also user scheduling.

For simplicity, we assume that the user association is given. The fully centralized problem can be written as a reuse-1 problem where the scheduling of all BSs (macro BS and small cells) is done at once. This problem is a simple extension of problem $\mathbb{P}_{\mathbb{S}}^{\mathbb{DL}}(\omega)$, which can be transformed and solved as in the homogeneous case. Next, we present practical schemes based on channel allocation and scheduling methods and compare their performance with the upper bound.

An important process in a HetNet is the channel allocation, which allocates the channels between the small cells and the macro base stations. We study two allocation schemes. In the first one, k subchannels are allocated to the small cells and the remaining $M-k$ subchannels to the macro base stations. In this case, the small cells do not cause any interference to the macro base stations and vice versa. Reuse-1 is used among macro BSs as well as among small cells. This scheme is called *Orthogonal Deployment (OD)*. The second channel allocation scheme we consider is the so called *Partially Shared Deployment (PSD)*, which also allocates k subchannels to the small cells which are used with reuse-1 and $M-k$ to the macro BSs which are also used with reuse-1; however, contrarily to the OD scheme, the macro cells can also transmit on the k subchannels dedicated to small cells but with a total power budget P' for these channels, while using the remaining power $P-P'$ in the remaining $M-k$ subchannels. It has been shown in [34] that PSD outperforms OD when round-robin scheduling is used (i.e., equal power).

For a given channel allocation and user association, each base station (small cell or macro cell) can schedule its own users locally. We will study the following four practical schemes (for a given user association):

- **PSD-SFR:** This scheme uses PSD as the channel allocation method. The small cells employ equal power scheduler, while the macro base stations serve their users with the scheduler based on SFR developed previously.
- **PSD-RR:** PSD is used for channel allocation and all cells use equal power scheduler to serve their users.
- **OD-SFR:** It is the same as **PSD-SFR**, except OD is used for channel allocation.
- **OD-RR:** It is the same as **PSD-RR**, except OD is used for channel allocation.

In the numerical results, we employ the same simple user association method for all schemes. It is proposed in [34] and is called *Small Cell First (SCF)*. In SCF, a user

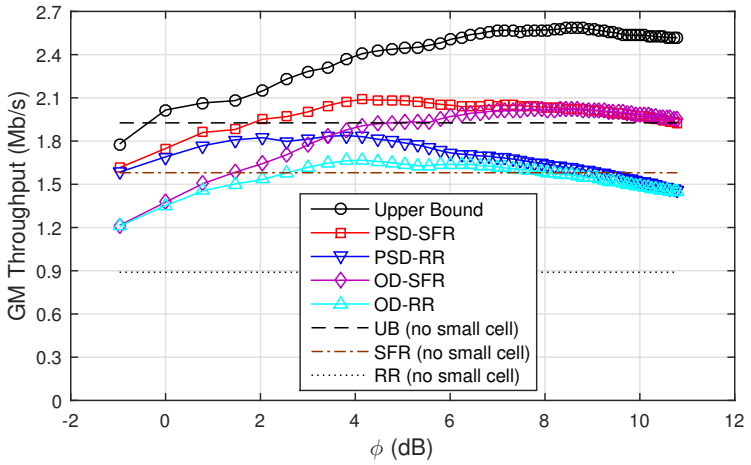


Figure 4.4: Average (over 100 realizations) GM throughput comparison as a function of Φ, the user association parameter for the small cells, for different schemes

associates with the small cell that yields the largest received SINR as long as this SINR is greater than a certain threshold Φ. If not, the user associates to the base station (macro BS or small cell) yielding the best received SINR. Note that computing the SINR from any small cell is easy since they use equal power on their k channels. For the macro BSs, we compute the SINRs they produce on a user by assuming that they also use equal power.

To evaluate the performance of the four schemes described above, we perform computations on the cellular network in Fig. 3.2 with reuse-1; however, there are two small cells in each macro cell at a distance of 230 meters left and right of the macro base station. We use the system parameters described in [34] and set the power budget of the small cells to 30 dBm. We assume that there are 100 users in the system distributed uniformly in the area. For PSD, we use $P' = 2mW$ as suggested in [56]. We compare the performance of the four schemes and the upper bound for different values of Φ. Note that while computing the upper bound, we do not consider any channel allocation method and we assume every BS can use every subchannel for the given user association.

Fig. 4.4 shows the results of our computations as a function of the SCF threshold Φ. For each value of Φ and for each of the four practical schemes described above, we pick the value of k, the number of subchannels allocated to the small cells, that yields the highest average GM throughput. The figure contains 8 plots, 3 where there are no small cells (for comparison) and 5 with small cells. The first remark is that we can improve

the performance significantly with small cells. The second is that the PSD performs much better than OD. The third is that the best value of Φ is different for the upper bound and the 4 practical schemes. The best of the practical scheme is **PSD-SFR** since it offer the maximum average GM. Finally, the upper bound is only 21% better than the maximum of **PSD-SFR**. Therefore, a reasonably good performance can be obtained in HetNets by using the practical scheme **PSD-SFR** as long as we use the right values for k and Φ. Note that the difference between the upper bound and the feasible solution is between 3 and 4% and it is not included in the figure.

4.7 Conclusion

We have revisited user scheduling in OFDMA cellular networks to understand better how much downlink scheduling would benefit from being done in C-RAN. Similar to the previous chapter, we have formulated and solved the centralized multi-cell user scheduling problem. While, we are able to obtain numerical results for reasonably sized network, the complexity is much too high for a process that is called upon every few tens of milli-seconds. However, these results can be used to evaluate the goodness of practical schemes. In both the homogeneous and the heterogeneous cases, we have showed that solutions based on round-robin are performing very badly compared to the upper bound and hence we designed practical (i.e., local) schemes, based on SFR along with its scheduler, and showed that they are only 20% away from the upper bound. Hence, there might not be a need to use a C-RAN for downlink scheduling.

Chapter 5

How Useful is Full Duplex in Cellular Networks?

5.1 Introduction

The use of full duplex communications (FDC) in cellular networks has recently become a hot topic and it is expected to be an important component of 5G networks [8]. FDC is made possible thanks to careful interference management at the BSs [57] to allow simultaneous downlink and uplink transmissions on the same PRB. The coexistence of uplink and downlink transmissions in co-channel cells create new sources of interference (i.e., sources that are not present in a TDD system) and their impact has to be studied carefully. Another important characteristic of any multi-cell OFDMA system is its *Traffic Asymmetry (TA)*, i.e., the fact that the traffic is in general much larger on the downlink than on the uplink. For example, today's cellular traffic is dominated by downlink traffic [5], e.g., in many networks, it is reported that more than 70% of the traffic is downlink. This characteristic has not been taken into account in the evaluation of FDC so far.

Operating an FDC-enabled system is going to be very complex due to the need for careful interference management and for jointly scheduling the uplink and the downlink. Many recent papers have focused on how to operate such a system without questioning if the gain in performance makes the added complexity worth it. We focus here on the following research question: What is the performance gain that an FDC-enabled multi-cell OFDMA network has over a TDD network when taking into account all sources of interference as well as the traffic asymmetry? By answering this question, we will help operators decide if it makes sense to implement FDC based on their network configuration

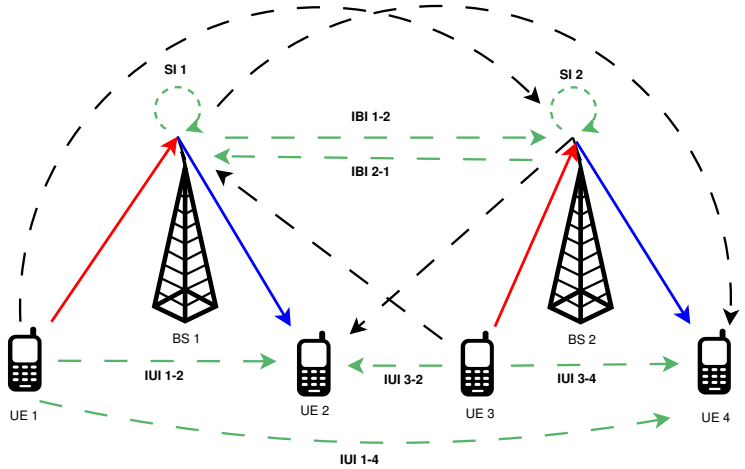


Figure 5.1: Sources of Interference: An example FDC scenario in a two-cell system where straight lines show data transmissions and dashed lines show the interference (all on the same subchannel). The green dashed lines are new types of interference (i.e., not present in a TDD system)

and traffic profile. In the following, we focus on systems where FDC is only enabled at the BSs.

The coexistence of uplink and downlink transmissions in co-channel cells in an FDC-enabled system, create many new sources of interference that have to be taken into account and studied carefully. Specifically, there are three new types of interference (i.e. not present in TDD systems):

- *Self-interference (SI)* at each BS since the transmission of a BS interferes with the simultaneous reception at the same BS;
- *Inter-BS interference (IBI)* between the BSs since each BS transmission is interfering with the reception of the other ones;
- *Inter-user interference (IUI)* between users in the same cell and in different cells (every transmitting user interferes with every receiving user on the same subchannel).

An example full-duplex scenario is shown in Fig. 5.1, where only two co-channel BSs are represented. In the figure, BS 1 and BS 2 are working in FDC mode and the red lines show uplink transmissions from some users and the blue lines show downlink transmissions

towards other users[1] on the same subchannel. The dashed lines show the interference created to the other receivers. The new sources of interference are shown in green while the ones in black are types of interference already present in a traditional TDD system.

The interference present in a TDD system is modeled using path loss models between a BS and a user given by the standardization bodies (e.g., 3GPP). Among the new sources of interference, self interference has been extensively studied and is easily modeled [58]. A BS has limited capability in terms of SI cancellation, therefore there is a residual self interference and its value has a significant impact on the system performance as will be seen in this chapter (Section 5.6) where we show that it is the most dominant source of interference. There also exist models for IUI that were developed in the context of device-to-device (D2D) communications [59]. IBI, on the other hand, has not received enough attention so far. IBI is related to the inter-BS path loss model, which depends on various factors such as antenna height, direction, and being in line of sight or not. In the following, *we evaluate the performance of an FDC-enabled multi-cell system for both homogeneous and heterogeneous networks using a generic inter-BS path loss model as well as existing IUI and SI models.*

Regarding traffic asymmetry, as mentioned earlier, this is a characteristic of a cellular network that has not been taken into account when evaluating FDC. Typically, an operator would know what the average ratio of uplink to downlink throughput is (this is what we call traffic asymmetry in the following) in its network. This would influence the way it selects the parameter β (the fraction of time slots allocated to downlink) if the network is TDD, i.e., it would allocate more or less resources to the uplink based on the ratio. A specific value of TA would yield a constraint coupling the uplink throughput to the downlink throughput and this coupling would mean that the performance of the uplink and the downlink cannot be computed independently. This is true for both a TDD system (our benchmark) and an FDC-enabled system.

Specifically, we analyze the performance of a multi-cell OFDMA network both for FDC and TDD, when all sources of interference are taken into account assuming that scheduling can be done centrally for all cells, i.e., using a *hypothetical* centralized scheduler that has access to all the channel state information. *We are not proposing to use centralized schedulers for operating the systems.* Instead, we are using centralized scheduling in this *offline* study to compute the performance of both systems under *comparable conditions.* By performing the scheduling centrally, we can manage the interference optimally and thus, we obtain upper bounds on the performance of practical systems. We expect the upper bound to be tighter for TDD than FDC since it will be difficult in practice to design joint

[1]Recall that users operate in half-duplex mode.

uplink/downlink scheduling for FDC and to accurately estimate the new types of interference and hence, practical FDC solutions will probably yield much worse performance than the centralized scheduler. In that sense, this study that compares upper bounds gives a slight advantage to FDC. However, we will show that even under these conditions, the performance gains of FDC over TDD are not very significant. We also show that the FDC gain might be higher in HetNets because of the lower power budget of the small cells, which leads to lower interference values.

Most of the existing literature on FDC considers simple system models (e.g., a single cell network), or ignore some sources of interference or over-simplify some important processes such as scheduling and power control. On the other hand, we take all these into account and consider TA on top of it.

The main conclusion we reach is that the performance gain of FDC over a TDD system significantly depends on the network configuration, traffic asymmetry, path loss models, SC capability, etc. The gain is higher in rural settings but even there, it is far from the doubling of capacity as it is claimed in the literature for a realistic value of TA. Moreover, the gain of FDC is higher in HetNets, when the channel allocation and user association processes are done efficiently[2].

5.2 Related Work

We review the recent papers on FDC in terms of the system model they use and the research question(s) they pose. We first discuss the system models.

System models

Many papers in the literature, such as [61], [62], [63], [64], consider a single cell model that neglects the inter-cell interference and more importantly the IBI caused by the full-duplex communications. This is unrealistic.

Since FDC is based on self interference cancellation, an important issue to consider is the residual self interference since it has been shown in previous studies, such as [58], [65], that a BS cannot completely cancel its own interference. Some of the earlier papers, such as [66], [67], [68], and [69], consider perfect interference cancellation.

[2]Some of these results are presented in our work [60], which is in preparation for submission to a journal.

Scheduling is the process where the BSs allocate resources (resource blocks and power) to the users both on the uplink and the downlink. Careful scheduling calls for power management and fairness [14]. Scheduling is easier in a TDD system because the uplink and the downlink can be processed independently. Several papers ignore scheduling altogether or fairness or power management. In a single channel case, power management is simpler but real systems are OFDMA-based.

Table 5.1 summarizes the main characteristics of the system model of the recent papers on FDC, namely whether:

- The system is multi-cell or not. Note that a single cell system cannot take into account important sources of interference.
- Residual self interference is considered or not.
- Traffic asymmetry is taken into account.
- Power management is considered.
- Scheduling is done and if proportional fairness which is the de-facto standard is considered.

As can be seen, our work is the first that has all the required characteristics.

Research questions

We now discuss the research questions addressed by the recent papers. We classify these papers in two categories, i.e., those addressing performance analysis and those addressing operational problems such as scheduling or power management.

The first category includes papers that focus on FDC from any information theory or a stochastic geometry perspective. There have been many studies on FDC based on information theory [80], [81], followed more recently by physical layer studies [58], [65], [82], which focus mostly on sophisticated interference cancellation techniques that allow FDC to be implemented in real systems. [66] and [69] are information theoretical studies that do not focus on scheduling but achievable rates in a single-channel network.

Stochastic geometry is used in [74], [77], [78], and [79] in a multi-cell setting. However, user scheduling is not included in any of these papers. Similarly, [76] considers a multi-cell network without scheduling and also neglects power control on the downlink. Authors

Table 5.1: Main characteristics of the models in the relevant literature (the options in italic are the most realistic)

Ref.	**Setting**	**SI Cancellation**	**Traf. Asym.**	**Power Management**	**Sched./Fairness**
[66]	Single Cell	Perfect	No	Yes (single channel)	No
[69]	Single Cell	Perfect	No	Yes (single channel)	No
[61]	Single Cell	*Residual SI*	No	*Yes (OFDMA)*	Sum-rate
[62]	Single Cell	*Residual SI*	No	*Yes (OFDMA)*	Sum-rate
[63]	Single Cell	*Residual SI*	No	Only on Downlink	Sum-rate
[64]	Single Cell	*Residual SI*	No	*Yes (OFDMA)*	Weighted Sum-rate
[70]	Single Cell	*Residual SI*	No	*Yes (OFDMA)*	Sum-rate
[71]	*Multi Small-Cell*	*Residual SI*	No	*Yes (OFDMA)*	Sum-rate
[72]	Single Cell	*Residual SI*	No	*Yes (OFDMA)*	Max-min
[67]	Single Cell	Perfect	No	No Power Control	No
[24]	*Multi Macro-Cell*	*Residual SI*	No	No Power Control	Weighted Sum-rate
[68]	*Multi Macro-Cell*	Perfect	No	Yes (single channel)	Weighted Sum-rate
[73]	*Multi Macro-Cell*	*Residual SI*	No	Yes (single channel)	Sum-rate
[74]	*Multi Macro-Cell*	*Residual SI*	No	Only on uplink	No
[75]	*Multi Macro-Cell*	*Residual SI*	No	No power control	Round-robin
[8]	*Multi Small-Cell*	*Residual SI*	No	Yes (single channel)	*Proportional fair*
[76]	*Multi Macro-Cell*	*Residual SI*	No	Only on uplink	No
[77]	*Multi Macro-Cell*	*Residual SI*	No	No power control	No
[78]	*Multi Macro-Cell*	*Residual SI*	No	No power control	No
[79]	*Multi Macro-Cell*	*Residual SI*	No	*Yes (OFDMA)*	Round-robin
Our work	*Multi Macro-Cell*	*Residual SI*	*Yes*	*Yes (OFDMA)*	*Proportional fair*

of [79] focus on FDC and D2D with a simple round-robin scheduler. [75] is a study on user association in an FDC-enabled network. It neglects power control and uses a round-robin scheduler. Authors of [68] analyze the spectral efficiency of a single channel multi-cell FDC-enabled network with a stochastic geometry approach.

In the second category are more recent papers on FDC that focus on Medium Access Control (MAC) layer and scheduling.

A game theoretical scheduling approach is used in [70] to maximize the sum-rate throughput. Note that maximizing the sum-rate ignores fairness among the users and zero resource might be assigned to users with poor channel conditions. Dual composition method is used in [71] for scheduling with imperfect channel estimation. The model used in the paper is a two-tiered macro cell, where FDC is allowed only at the small cells. Their approach is to limit the interference created to the other small cells while scheduling. This paper also considers sum-rate maximization. Moreover, considering the lower power budget and lower channel gain of the small cells, the impact of IBI would be more significant in a macro cell network. Max-min fairness is considered in [72] with a user pairing solution, however a single-cell model is used.

A multi-cell system is considered in [83]. A sub-optimal heuristic scheduler is proposed to maximize the weighted sum-rate. However, there is no power control, which is

a very crucial part of FDC. Similarly, [63] deals with the channel allocation problem using matching theory. However, it fixes the transmit power of the uplink users to simplify the model. Authors of [75], [77], [78] also deal with a multi-cell network without power control. Without a detailed power control, it has been shown in [84] that the uplink performance can be significantly deteriorated due to high interference received from the downlink transmissions.

A multi-cell system is considered in [73] for a single channel TDMA-like system, which does not utilize the multi-channel nature of OFDMA. Authors of [8] present some simulation results on a heterogeneous network where FDC is enabled at the small cells. It considers a single channel network with little details on the scheduling method. Note that a multi-channel system is more complex since power management among channels is also important.

This thesis evaluates the performance of an FDC enabled multi-cell network while considering all sources of interference as well as traffic asymmetry. We also quantify the effects of each type of interference on the system performance.

5.3 System Model

We use the same model described in Chapter 2. We will start with a homogeneous network and then extend our results to the HetNets. However, there is some differences especially in the interference models.

As explained previously, FDC introduces new types of interferences that do not exist in a TDD system. The only interfering transmissions for TDD on the uplink (resp. downlink) are the uplink (resp. downlink) transmissions in the neighboring cells. This interference can be computed on a per PRB basis, using the channel gains between the BSs and the users and the transmit power on that subchannel.

The three new types of interference that occur due to FDC are self interference (SI), inter-user interference (IUI), and inter-BS interference (IBI). Next, we explain the model we use for each type of interference.

- **SI**: Recall that users operate in half-duplex mode. Hence, only BSs have the capability to transmit and receive on the same PRB. In doing so, self interference is created on the uplink (due to the downlink transmission on the same subchannel). The BSs can cancel this interference partially using self-cancellation and the degree to which it can do so is represented by the parameter $\mathcal{C}$ (in dB). Specifically, the self-interference

after SC (the residual self interference) is equivalent to an extra noise term equal to (in dBm) $\bar{P} - \mathcal{C}$ if the BS transmits with power $\bar{P}$ on that PRB. Note that the residual self interference can be modeled as an additional noise [62,85] but also using a Rician model [86]. In this chapter, we adopt the first approach, i.e., we consider the residual self interference as an additional noise.

- **IUI**: IUI occurs due to the uplink transmissions of other users (within and outside the cell) while a user is receiving a downlink transmission from its BS. To compute IUI, we need the channel gains between the users (as well as the transmit powers). We represent the channel gain between users u and v with $L_{u,v}$. Although this type of interference does not occur in a TDD system, it has been studied before in the context of D2D communications. Therefore, we will use an existing path loss model between the users (the one given in [59]) to compute the channel gains.

- **IBI**: IBI is the interference between BSs. We need the channel gains between the BSs in order to compute IBI. Let $H_{j,k}$ be the channel gain between BSs j and k. Note that $H_{j,k}$ would certainly depend on many network parameters such as antenna heights, directions, and the fact that the two BSs are in line of sight or not. Our objective in this chapter is *not* to propose a new model but to evaluate the impact of a generic path loss model on the performance. Therefore, we will consider a path loss model characterized by its parameters $\xi = (a_{IBI}, b_{IBI})$, which can be written as $PL(d) = a_{IBI} + b_{IBI} \times \log(d)$, where d is the inter-BS distance. We will focus on how much the pair (a_{IBI},b_{IBI}) affects the performance of FDC.

Recall that on the downlink, the total interference seen at user $u \in \mathcal{U}_k$ on PRB $\{c, t\}$ is denoted by $I_{u,k}^{c,t}$ and on the uplink the total interference seen at BS k is denoted by $Q_{u,k}^{c,t}$. In FDC mode, $I_{u,k}^{c,t}$ includes all the uplink and downlink transmissions in the neighboring cells as well as the uplink transmission in the current cell. In TDD mode, it only includes the downlink transmissions in the neighboring cells. For FDC, the interference $Q_{u,k}^{c,t}$ includes all the uplink and downlink transmissions in the neighboring cells and also the residual interference due to the downlink transmission of BS k, while in the TDD mode, it only includes the uplink transmissions in the neighboring cells.

One important parameter we introduce in this chapter is β that determines the fraction of downlink subframes in the TDD system. Then, the remaining $1 - \beta$ fraction of the subframes are allocated to uplink transmissions.

5.4 System-wide User Scheduling Problem Formulation for FDC

In this section, we present the problem formulation for FDC and show how we solve it. Our aim is to be proportionally fair among all the users in the system. Therefore, our performance will be the GM of overall user throughputs, i.e., the sum of uplink and downlink throughputs.

5.4.1 Key variables and constraints

We start by defining several key variables along with their constraints. Note that all the variables in the formulation are non-negative. The main variables of the problem are the power allocated on the PRBs.

We first define the downlink throughput seen by user $u \in \mathcal{U}_k$. It is the sum of its downlink rates on each PRB (see Equation (5.1)). For all PRBs not allocated to that user on the downlink, the corresponding power will be zero and hence the rate will be zero. The same will be true for the uplink. The downlink rate on each PRB is computed using the SINR on that PRB and the SINR-to-rate mapping function $f(\gamma)$, which we realize using Constraints (5.2) and (5.3) similar to our formulation in centralized downlink scheduler. Constraint (5.3) picks the best MCS level depending on the SINR on that PRB and constraint (5.2) gets the rate of that MCS level. The downlink SINR on each PRB is defined by Equation (5.4). The interference seen at each user on each PRB is defined by Equation (5.5). The first term in the interference formula is due to the downlink transmissions in the neighboring cells, which is the only interference term for a TDD system. The second term is the IUI, which includes all the uplink transmissions in all cells including cell k.

$$\lambda_{u,k}^{DL} = \sum_{c \in \mathcal{M}} \sum_{t \in \mathcal{T}} R_{u,k}^{c,t}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K} \tag{5.1}$$

$$R_{u,k}^{c,t} = \sum_{l \in \mathcal{L}} x_{u,k}^{c,t,l} \vartheta_l, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.2}$$

$$S_{u,k}^{c,t} \geq x_{u,k}^{c,t,l} \eta_l, \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, l \in \mathcal{L} \tag{5.3}$$

$$S_{u,k}^{c,t} = \frac{P_{u,k}^{c,t} \times G_{u,k}}{\mu^{DL} + I_{u,k}^{c,t}}, \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, l \in \mathcal{L} \tag{5.4}$$

$$I_{u,k}^{c,t} = \sum_{j \in \mathcal{K}, j \neq k} \sum_{v \in \mathcal{U}_j} P_{v,j}^{c,t} G_{u,j} + \sum_{j \in \mathcal{K}} \sum_{v \in \mathcal{U}_j, v \neq u} p_{v,j}^{c,t} L_{v,u} \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.5}$$

Next, we define the uplink throughput of each user in a similar fashion using the following equations:

$$\lambda_{u,k}^{UL} = \sum_{c \in \mathcal{M}} \sum_{t \in \mathcal{T}} r_{u,k}^{c,t}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K} \tag{5.6}$$

$$r_{u,k}^{c,t} = \sum_{l \in \mathcal{L}} y_{u,k}^{c,t,l} \vartheta_l, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.7}$$

$$s_{u,k}^{c,t} \geq y_{u,k}^{c,t,l} \eta_l, \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T}, l \in \mathcal{L} \tag{5.8}$$

$$s_{u,k}^{c,t} = \frac{p_{u,k}^{c,t} \times G_{u,k}}{\mu^{UL} + Q_{u,k}^{c,t}}, \ \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.9}$$

$$Q_{u,k}^{c,t} = \sum_{j \in \mathcal{K}, j \neq k} \sum_{v \in \mathcal{U}_j} P_{v,j}^{c,t} H_{k,j}(\xi) + \sum_{j \in \mathcal{K}, j \neq k} \sum_{v \in \mathcal{U}_j} p_{v,j}^{c,t} G_{v,k}$$

$$+ \frac{1}{\mathcal{C}} \sum_{v \in \mathcal{U}_k, v \neq u} P_{v,k}^{c,t}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.10}$$

Equations (5.6-5.9) are similar to Equations (5.1-5.4). We define the uplink interference on each PRB by Equation (5.10). In this case, the interference includes the uplink and downlink transmissions in the neighboring cells as well as the downlink transmission in cell k. The first term of Equation (5.10) represents the IBI, where $H_{k,j}(\xi)$ is the inter-BS channel gain characterized by the parameters $\xi = (a_{IBI}, b_{IBI})$. The second term is the inter-cell interference received from the uplink transmission in the neighboring cells, which also exists in TDD. The constant $\mathcal{C}$ is the self interference cancellation parameter introduced earlier. The higher $\mathcal{C}$ is, the lower the residual interference becomes. Note that in the equation, power and $\mathcal{C}$ are not defined in dBm and dB, respectively, but power is defined in Watts and $\mathcal{C}$ does not have a unit since it is just a coefficient.

Constraints (5.11-5.15) enforce the PRB allocation constraints.

$$P_{u,k}^{c,t} \leq \sum_{l \in \mathcal{L}} x_{u,k}^{c,t,l} P_{BS}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.11}$$

$$p_{u,k}^{c,t} \leq \sum_{l \in \mathcal{L}} y_{u,k}^{c,t,l} P_U, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.12}$$

$$\sum_{u \in \mathcal{U}_k} \sum_{l \in \mathcal{L}} x_{u,k}^{c,t,l} \leq 1, \quad \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.13}$$

$$\sum_{u \in \mathcal{U}_k} \sum_{l \in \mathcal{L}} y_{u,k}^{c,t,l} \leq 1, \quad \forall k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.14}$$

$$\sum_{l \in \mathcal{L}} (x_{u,k}^{c,t,l} + y_{u,k}^{c,t,l}) \leq 1, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, c \in \mathcal{M}, t \in \mathcal{T} \tag{5.15}$$

Constraint (5.11) enforces that a BS k cannot allocate power to serve u on PRB $\{c, t\}$ if that PRB is not allocated to that user on the downlink. Constraint (5.12) is the same for the uplink. Constraints (5.13) and (5.14) enforce that a PRB can be allocated to only one user on the downlink and the uplink, respectively in a given cell. Constraint (5.15) indicates that a user cannot transmit and receive at the same time. Note that there is nothing that prevents a PRB to be simultaneously allocated for a downlink and an uplink transmissions as long as it is to different users.

Finally, we have the following two power budget constraints, which ensure that the BSs and users cannot use more power than their power budget at a given time.

$$\sum_{u \in \mathcal{U}_k} \sum_{c \in \mathcal{M}} P_{u,k}^{c,t} \leq P_{BS}, \quad \forall k \in \mathcal{K}, t \in \mathcal{T} \tag{5.16}$$

$$\sum_{c \in \mathcal{M}} p_{u,k}^{c,t} \leq P_U, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K}, t \in \mathcal{T} \tag{5.17}$$

Next, we introduce the *downlink/uplink traffic asymmetry* constraint. Let $\theta_{u,k} \geq 0$ be the targeted ratio of uplink to downlink throughput of user $u \in \mathcal{U}_k$. This is a parameter that is network specific and can be measured by the operators. We can then add the following constraint in our formulation:

$$\lambda_{u,k}^{UL} = \theta_{u,k} \times \lambda_{u,k}^{DL}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K} \tag{5.18}$$

We define Θ as the vector consisting of $\theta_{u,k}$ values of all the users in the system, i.e., $\Theta = \{\theta_{u,k}\}$.

Finally, we define the overall throughput $\lambda_{u,k}$ of user u in cell k as the sum of the

throughput it sees on the uplink and the downlink, i.e.,

$$\lambda_{u,k} = \lambda_{u,k}^{UL} + \lambda_{u,k}^{DL}, \quad \forall u \in \mathcal{U}_k, k \in \mathcal{K} \tag{5.19}$$

Our objective is to maximize the GM Γ of the overall throughputs of all the users in the system, by computing the power allocations and the PRB allocations jointly for the uplink and the downlink, where:

$$\Gamma = \Big(\prod_{k \in \mathcal{K}} \prod_{u \in \mathcal{U}_k} \lambda_{u,k} \Big)^{1/N}, \tag{5.20}$$

where N is the total number of users in the system.

5.4.2 Problem formulation for FDC

We are now ready to formulate our centralized scheduling problem $\mathcal{P}_{FDC}(\Theta)$ that maximizes the GM of the overall rate each user sees. Its input parameters are $\Theta = \{\theta_{u,k}\}$, the realization ω, inter-BS path loss model parameters $\xi = (a_{IBI}, b_{IBI})$, and SC value $\mathcal{C}$ at the BSs. Its variables are $\{x_{u,k}^{c,t,l}, y_{u,k}^{c,t,l}, P_{u,k}^{c,t}, p_{u,k}^{c,t}, S_{u,k}^{c,t}, s_{u,k}^{c,t}, I_{u,k}^{c,t}, Q_{u,k}^{c,t}, R_{u,k}^{c,t}, r_{u,k}^{c,t}, \lambda_{u,k}^{DL}, \lambda_{u,k}^{UL}, \lambda_{u,k}\}$.

$$\begin{aligned} \mathcal{P}_{FDC}(\Theta): \quad & \max \Gamma \\ & \text{s.t.} \quad \text{constraints (5.1-5.18)} \end{aligned} \tag{5.21}$$

In the following, we will use the same solution technique that we have used in the previous two chapters to solve $\mathcal{P}_{FDC}(\Theta)$. As an aside, we will also show in Section 5.6.4 that when this problem is solved without constraint (5.18), i.e., without enforcing the traffic asymmetry, it gives a similar amount of throughput to the uplink and downlink, which is not appropriate for today's networks, where downlink traffic represents at least 70% of the total traffic.

5.4.3 Transformation and solution technique for $\mathcal{P}_{FDC}(\Theta)$

Similar to the uplink and downlink centralized scheduling problems, $\mathcal{P}_{FDC}(\Theta)$ is an MINLP with large number of binary variables and non-linear constrains. We perform a similar

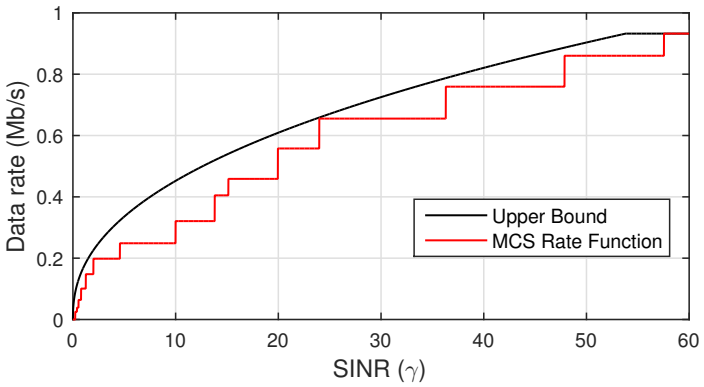


Figure 5.2: The original MCS rate function $f(\gamma)$ and the upper bound we constitute

transformation to obtain a signomial programming problem, however it is now slightly different since both uplink and downlink are included in the problem.

The first transformation is done in order to eliminate the discrete rate function similar to the one we have done for the centralized uplink and downlink scheduling problems in Chapters 3 and 4, respectively. We replace the rate function $f(\gamma)$ with a tight upper bound $g(\gamma)$, which is slightly different from the one we have used before. For the MCS function given in [1], we select $g(\gamma)$ as:

$$g(\gamma) = \min(a\gamma^{0.43}, R_Z) \tag{5.22}$$

where R_Z is the highest rate achievable by $f(\gamma)$ and $a = 0.168$. The tightness of the upper bound is illustrated in Fig. 5.2 for the MCS function $f(\gamma)$ given in Table 2.1 and the upper bound $g(\gamma)$ in Equation (5.22). Although this upper bound is slightly less tight than the one we have used in the previous chapters, it is much simpler to use for the much larger sized centralized scheduling problem for FDC.

We can then replace the binary variables $x_{u,k}^{c,t,l}$ by $x_{u,k}^{c,t}$ and $y_{u,k}^{c,t,l}$ by $y_{u,k}^{c,t}$ and hence eliminate many binary variables.

The second transformation is similar to the one we had in the previous chapters as we replace constraints (5.11-5.15) with the following three constraints:

$$P_{u,k}^{c,t} P_{v,k}^{c,t} \leq \sigma, \quad \forall \{u \neq v\} \in \mathcal{U}_k, c \in \mathcal{M}, k \in \mathcal{K}, t \in \mathcal{T}, \tag{5.23}$$

$$p_{u,k}^{c,t} p_{v,k}^{c,t} \leq \sigma, \quad \forall \{u \neq v\} \in \mathcal{U}_k, c \in \mathcal{M}, k \in \mathcal{K}, t \in \mathcal{T}, \tag{5.24}$$

$$P_{u,k}^{c,t} p_{u,k}^{c,t} \leq \sigma, \quad \forall u \in \mathcal{U}_k, c \in \mathcal{M}, k \in \mathcal{K}, t \in \mathcal{T}, \tag{5.25}$$

where σ is a very small positive number. Constraint (5.23) ensures that a PRB can be allocated for only one downlink transmission in a given cell. Similarly, constraint (5.24) ensures that only one user can use a PRB for an uplink transmission in a given cell. Finally, constraint (5.25) ensures that a user cannot use a PRB for a downlink and an uplink transmission simultaneously.

With these two modifications, we obtain a new problem that finds a tight upper bound to the original problem $\mathcal{P}_{FDC}(\Theta)$. With some further modifications, we can obtain a signomial problem [32]. We first replace the equality signs in constraints (2.1), (2.2), (5.1), (5.5), (5.6), (5.10), and (5.19) with $\leq$ sign, which does not change the optimal point of the problem since this is a maximization problem. We also add a constraint to make all the variables strictly positive by enforcing that all the variables are greater than a very small positive constant. Then, the problem becomes a signomial programming problem that finds an upper bound to the original problem $\mathcal{P}_{FDC}(\Theta)$. We call the new problem $\mathcal{P}_{UB}(\Theta)$. In order to solve the resultant signomial programming problem $\mathcal{P}_{UB}(\Theta)$, we again use the method proposed in [7].

5.4.4 Deriving a feasible solution to $\mathcal{P}_{FDC}(\Theta)$

Note that the problem we solve provides an upper bound to the original scheduling problem $\mathcal{P}_{FDC}(\Theta)$. Using the results of that upper bound problem, we can derive a feasible solution to the original problem as follows:

$$x_{u,k}^{c,t,l} = 1 \ \texttt{if} \ \ \underset{i \in \mathcal{U}_k}{\operatorname{argmax}} \{P_{i,k}^{c,t}\} = u \ \ \& \ \ \eta_l \leq S_{u,k}^{c,t} < \eta_{l+1} \ \ (= 0 \ \texttt{otherwise})$$

$$y_{u,k}^{c,t,l} = 1 \ \texttt{if} \ \ \underset{i \in \mathcal{U}_k}{\operatorname{argmax}} \{p_{i,k}^{c,t}\} = u \ \ \& \ \ \eta_l \leq s_{u,k}^{c,t} < \eta_{l+1} \ \ (= 0 \ \texttt{otherwise})$$

$$x_{u,k}^{c,t,l} = 0 \ \texttt{if} \ \ P_{u,k}^{c,t} < p_{u,k}^{c,t} \ \ (y_{u,k}^{c,t,l} = 0 \ \texttt{otherwise})$$

Then, we set $P_{u,k}^{c,t}$ (resp. $p_{u,k}^{c,t}$) to zero if $x_{u,k}^{c,t,l} = 0$, $\forall l \in \mathcal{L}$ (resp. $y_{u,k}^{c,t,l} = 0$, $\forall l \in \mathcal{L}$). Note that this yields a feasible solution since a PRB is allocated for only one downlink transmission within a cell thanks to the first equation. The second equation ensures that a PRB can be used for only one uplink transmission within a cell. A PRB can be used for one uplink and one downlink transmissions simultaneously, but we prevent the same user to transmit simultaneously on the uplink and downlink using the last equation. Then, we can compute the GM throughput of the feasible solution using the power values we

obtain and the real rate function $f(\gamma)$ instead of the upper bounding function $g(\gamma)$. We will show later that the difference between the feasible solution, which is a lower bound to the original problem, and the upper bound is low (around 5%), therefore the upper bound we obtain is a tight bound.

5.5 Problem Formulation for TDD

We now formulate the centralized scheduling problem for TDD. Recall that β is the proportion of time spent on the downlink. It can take one of the following values $\{0, 1/T, 2/T, \ldots, 1\}$ where T is the number of subframes within a frame. In this case, $\beta \times T$ of the subframes are allocated to the downlink and $(1-\beta) \times T$ subframes are allocated to the uplink. In the downlink subframes, no uplink transmissions can occur and hence $y_{u,k}^{c,t,l}$ should be equal to zero for the downlink subframes and similarly, $x_{u,k}^{c,t,l}$ should be equal to zero for the uplink subframes. The interference includes only the uplink transmissions in the neighboring cells during the uplink subframes (i.e., Constraint (5.10) does not contain the first term in the right hand-side) and only the downlink transmissions in the neighboring cells during the downlink subframes (i.e., Constraint (5.5) has only the first term in the right hand-side). Keeping the variables and parameters the same as above, for a given β and a given Θ, we can formulate $\mathcal{P}_{TDD}(\Theta, \beta)$ that schedules together uplink and downlink under TDD, as follows:

$$\begin{aligned}
\mathcal{P}_{TDD}(\Theta, \beta): \quad & \max \Gamma && (5.26)\\
\text{s.t.} \quad & \text{constraints (5.1-5.18)} \\
& y_{u,k}^{c,t,l} = 0, \quad \forall t \in \{1 \ldots \beta T\} && (5.27)\\
& x_{u,k}^{c,t,l} = 0, \quad \forall t \in \{(\beta T + 1) \ldots T\} && (5.28)
\end{aligned}$$

In this formulation, the first βT subframes are allocated to the downlink and the rest of the subframes are allocated to the uplink. Note that this problem gives us a solution for a given value of $\{\Theta\}$ and β. Since $\mathcal{P}_{TDD}(\Theta, \beta)$ has the same structure than $\mathcal{P}_{FDC}(\Theta)$, it can be solved with the method explained in the previous section. Note that the uplink and the downlink are coupled by the objective function and Constraint (5.18) only.

5.6 Numerical Results

In this section, we evaluate the performance of FDC and TDD for various parameters including the interference parameters $\xi = (a_{IBI}, b_{IBI})$ and $\mathcal{C}$ as well as the traffic asymmetry parameter Θ. To simplify our discussion, we assume the same value of $\theta_{u,k}$ for all the users, i.e., we use $\theta_{u,k} = \theta$ to denote the traffic asymmetry for all the users in the network.

We consider a cellular network composed of 3 BSs and study an urban and a rural setting. The main differences between the two scenarios are the inter-BS distance and the channel models. For the urban setting, the inter-BS distance is 500 meters while it is 1732 meters for the rural setting [14]. Each user is associated to the BS yielding the highest channel gain. We set the total number of users in the system to 30, corresponding to an average of 10 users per cell, unless otherwise stated. The total number of subchannels M is set to 30. The number of subframes T in a frame is set to 10.

The path loss between the BS and the users for the urban setting is computed with the formula explained in previous chapters. For the rural setting, we use $117.5953 + 38.6334 \times \log_{10}(d)$ to compute the path loss between a user and a BS. We also apply fading to all channel gain computations.

For **IUI**, we use a path loss between two users separated by a distance of d equal to $148 + 40 \times \log_{10}(d)$ dB [59]. We use $\mathcal{C} = 110$ dB as the default value for the self-interference cancellation parameter [87]. However, we will later use different values for $\mathcal{C}$ to see its effects on FDC performance. For **IBI**, we use as a default the free space loss model, which occurs when the antennas are in line of sight without any obstacles. We use the following formula to compute the path loss: $128.1 + 20 \times \log_{10}(d)$. Note that this is a conservative model since the path loss might be higher in real life due to obstacles and multi-path fading and this would yield less interference. Other path loss models will be studied later in this chapter.

The results are given as an average of 100 realizations.

5.6.1 Tightness of the upper bound

Recall that we solve, for each realization, a signomial optimization problem that finds an upper bound to the original scheduling problem $\mathcal{P}_{FDC}(\Theta)$ and we derive a feasible solution to each problem from the results of the upper bound problem as explained in the previous section. The differences between the GM throughput (i.e., the objective function) of the upper bound problem and the feasible solution are reported in Table 5.2 for the urban

setting for different values of θ when we use $\mathcal{C} = 110$ dB, [87] and the free space loss model to compute IBI.

Table 5.2: GM throughput (Mb/s) of the upper bound and the feasible solution for $\mathcal{P}_{FDC}(\Theta)$ (urban setting, free space inter-BS path loss and $\mathcal{C} = 110$ dB, $N = 30$)

	$\theta = 0.2$	$\theta = 0.5$	$\theta = 1$
Upper Bound	1.9755	2.3951	2.7915
Feasible Solution	1.8639	2.2642	2.6144

The difference between the upper bound and the feasible solution is about 5% which shows that the upper bound we obtain is tight. In the following, we will show results based on the upper bound.

5.6.2 Performance comparison of FDC and TDD

Next, we compare the performance of FDC and TDD for different values of traffic asymmetry, inter-BS path loss model, SC capability, number of users, and network setting (i.e., urban or rural). We also evaluate the performance of FDC in HetNets. Then, to quantify which type of interference has the dominant effect on FDC performance, we compare different scenarios where we disable some of the interference sources. We also investigate the impact of number of users on FDC performance. Finally, we show how the centralized scheduler performs if we do not enforce the traffic asymmetry between the uplink and the downlink.

Impact of θ on FDC performance

We first evaluate the performance of FDC as a function of θ, which determines the per user uplink to downlink throughput ratio. Fig. 5.3 shows the FDC gains over TDD for the urban and rural settings with 110 dB of SC and the free space loss model for inter-BS channels. Note that for TDD, we choose for each θ the best β, i.e., the one that maximizes the GM of the overall throughput averaged over 100 realizations.

It is clear from Fig. 5.3 that the gain of FDC over TDD strongly depends on the value of θ and increases with θ. This is because if $\theta = 0$, FDC and TDD are the same and the larger θ the more uplink traffic there is and hence the more there is to gain with FDC (note that $\theta = 1$ means that there is as much uplink traffic as downlink traffic). Note that

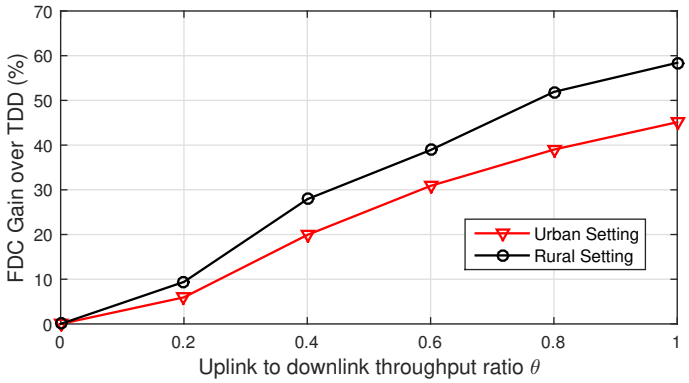


Figure 5.3: Performance gain of FDC over TDD as a function of UL/DL GM throughput ratio θ with free space path loss and $\mathcal{C} = 110$ dB

the range of interest for θ is around the lower values, i.e., (0-0.6), based on real network measurements [5,11] (for example $\theta = 0.6$ would mean that the downlink traffic is 62.5% of the total traffic). In this range, the gain of FDC is at most 30% for the urban setting and 40% for the rural setting. These results are obtained with a conservative IBI (i.e., a free space loss model) and a reasonable $\mathcal{C}$. We will study next how these two parameters affect our conclusions. As mentioned in the introduction, implementing FDC will be complex and hence the performance of a practical FDC scheme might be significantly lower than the performance of our centralized ideal FDC. In that context, a gain of 30% does not seem very large. Therefore, deploying FDC for the urban setting might not be that beneficial. On the other hand, the rural setting can be a better candidate for FDC since it yields a higher gain.

In any case, for the practical range of θ, FDC is far from doubling the data rates. The highest gain is 60% and it can be achieved when uplink and downlink have equal weight which is not the case in today's networks that are dominated by downlink traffic [5, 11].

We now investigate how these conclusions are impacted by the IBI model and by SC.

Impact of inter-BS path loss on FDC performance

We analyze the impact of inter-BS path loss on FDC performance, assuming that the BSs have a SC of 110 dB, i.e., $\mathcal{C}$ =110 dB. We begin with the urban setting. We perform the computations for $\theta = 0.2$ and 0.6.

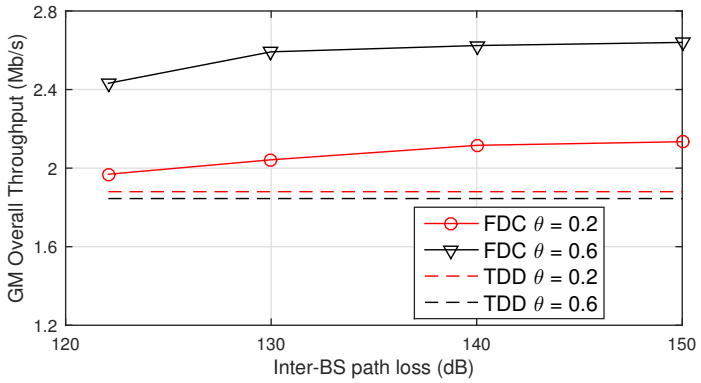


Figure 5.4: Performance of FDC and TDD for the urban setting as a function of inter-BS path loss for θ=0.2 and 0.6 with $\mathcal{C} = 110$ dB and $N = 30$

We use the path loss model, ξ, defined as $PL(d) = a_{IBI} + b_{IBI} \times \log(d)$, where d is the inter-BS distance. Note that d is the same for each BS pair in our system. Fig. 5.4 shows the computed GM overall throughput for different values of inter-BS path loss values computed for different values of the parameters a_{IBI} and b_{IBI}. The first point in the graph (around 120 dB) corresponds to the free space loss model. Note that we only consider path loss values that are yielding less interference than the free space loss model because free space yields the highest possible channel gain between two BSs since it does not consider any obstacles or multi-path fading.

It is clear that a better IBI (i.e., higher inter-BS path loss) yields better gains for FDC over TDD, indeed, the gain for the free space loss model is 29% for $\theta = 0.6$ and 7% for $\theta = 0.2$ while it is 42% for $\theta = 0.6$ and 15% for $\theta = 0.2$ for an inter-BS path loss of 150 dB, which is very optimistic.

Similar results are given in Fig. 5.5 for the rural setting, where again the first point represents the free space loss model. The overall gain of FDC over TDD is significantly higher for the rural setting. For example, for the free space loss model, the gain over TDD is around 29% for the urban case whereas it is 45% for the rural case when $\theta = 0.6$. However, as the path loss increases the amount of performance increase for FDC is lower for rural settings. This is because the BSs are further away from each other and the effect of IBI is less pronounced for the rural case than the urban case.

Remark: To verify that our results hold for bigger networks, we performed some computations for a 5 cell network. The performance gains of FDC over TDD are similar to the

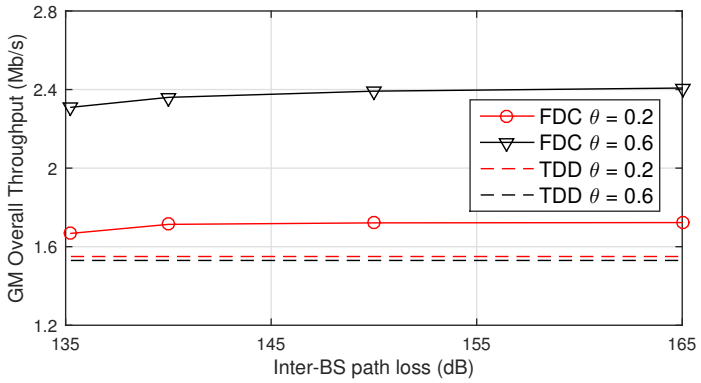


Figure 5.5: Performance of FDC and TDD for the rural setting as a function of inter-BS path loss for θ=0.2 and 0.6 with $\mathcal{C} = 110$ dB

3 cell network.

Impact of SC capability at the BSs

Next, we focus on the impact of the self-interference cancellation capability of the BSs on system performance. We include this effect in our formulation via the self interference cancellation parameter $\mathcal{C}$ (see Constraint (5.10)). In Fig. 5.6, we compare the performance gain of FDC over TDD for $\theta = 0.6$ for both the urban and rural settings when IBI is modeled using the free space loss model.

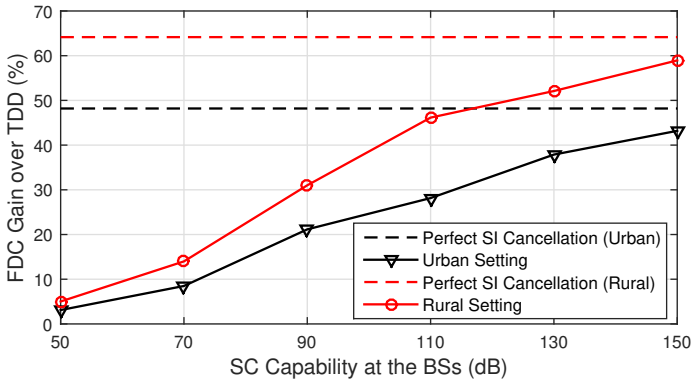


Figure 5.6: Performance gain of FDC over TDD as a function of self interference cancellation when $\theta = 0.6$ and $N = 30$ for urban and rural setting

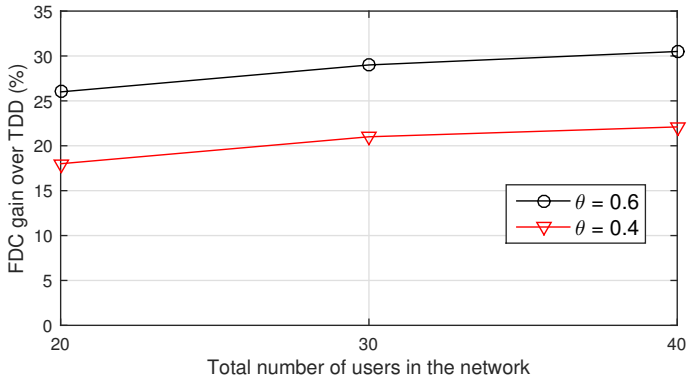


Figure 5.7: Performance gain of FDC over TDD as a function of total number of users in the network for the urban setting with $\mathcal{C} = 110$ dB and the free space inter-BS path loss

The capability to cancel self interference greatly impacts the system performance. For both rural and urban settings, when the BSs can only cancel 50 dB self interference, FDC performs not better than TDD. A good performance can be achieved only when the BSs can cancel high level of self interference. Clearly, the gain is higher for the rural case as explained before. For the urban setting, to see a gain of 40% over TDD, we need a SC of 130 dB for the free space loss model for the urban setting and 110 dB for the rural setting. According to [87], a realistic value of SC is 110 dB, which yields around 30% gain

for the urban setting. Note that a SC value of 150 dB would yield results very similar to the perfect SI cancellation case for both settings.

Impact of the number of users on FDC performance

Next, we present the FDC performance for different numbers of users. Fig. 5.7 shows the results for the urban setting as a function of the total number of users in the network.

It is seen that increasing the number of users in the network slightly improves the gain of FDC over TDD. This can be explained by the fact that with more users, the number of opportunities to pair uplink and downlink users to share the same PRB increases. Therefore, a better FDC gain can be achieved in a more crowded network.

5.6.3 Quantifying the effects of each interference type

In this section, we assume that we can disable certain types of interference so that we can understand how much they contribute to the FDC performance. We consider the 9 FDC schemes shown in Table 5.3. In the table, a cross mark corresponds to disabling that type of interference.

Table 5.3: List of FDC Variations

	IBI	IUI	SC			IBI	IUI	SC
Scheme 1	✗	✗	∞		Scheme 6	✗	✓	110 dB
Scheme 2	✗	✓	∞		Scheme 7	✓	✗	110 dB
Scheme 3	✓	✗	∞		Scheme 8	✓	✓	110 dB
Scheme 4	✓	✓	∞		Scheme 9	✓	✓	50 dB
Scheme 5	✗	✗	110 dB					

Scheme 1 is the most optimistic one as it does not consider any IBI, IUI, or SI. Therefore, it is an upper bound for all other schemes. Among all the nine schemes, Scheme 8 is the most realistic one as it considers IBI and IUI while having a realistic SC capability at the BSs. For the given 9 FDC schemes, we perform computations for 100 realizations and compute the average gain of each scheme over TDD. The results are given in Fig. 5.8.

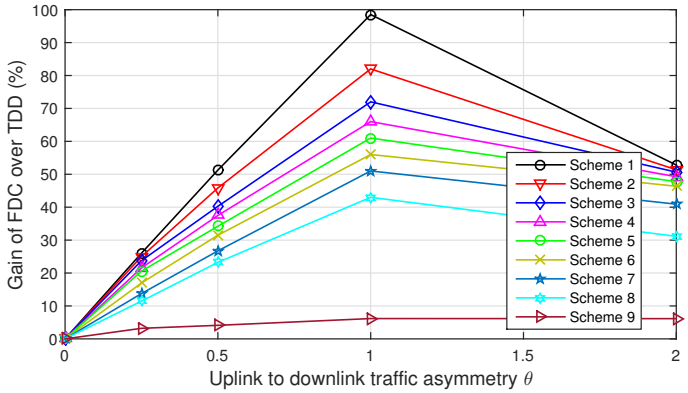


Figure 5.8: Performance gain of FDC over TDD for the urban setting for $N = 30$ as a function of θ for the scenarios in Table 5.3

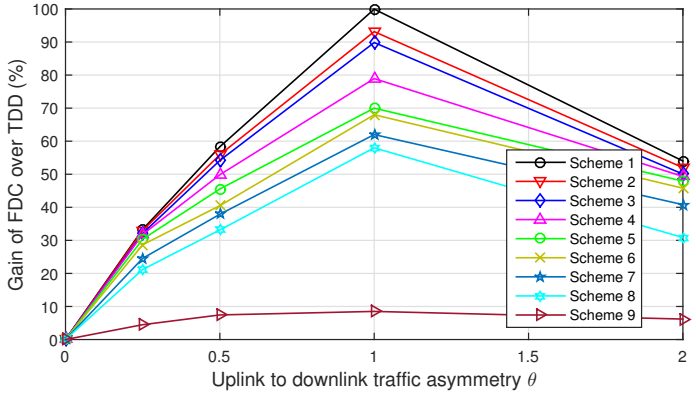


Figure 5.9: Performance gain of FDC over TDD for the rural setting as a function of θ for the scenarios in Table 5.3

Fig. 5.8 shows that the most dominant source of interference is self interference and the amount of interference cancellation $\mathcal{C}$ has a great impact on the FDC performance. When we compare Schemes 4, 8, and 9, where $\mathcal{C}$ is infinity, 110 dB, and 50 dB, respectively, we see that there is a huge difference between the gain of the three schemes. Among the new sources of interference, IUI has the least effect since adding IUI reduces the gain the least. However, its impact is still not negligible as can be seen by comparing Schemes 1 and 2.

The most realistic scheme is Scheme 8 and its reduction with respect to Scheme 1 is very significant. As noted before FDC brings high gain in GM overall throughput when uplink and downlink traffic have similar throughput performance. However, today's networks are mostly dominated by downlink traffic. Therefore, the region we are interested in in these graphs is actually the left half, where the GM downlink throughput is higher than the uplink throughput. In this case, the gains are lower than for the case of equal downlink and uplink throughput.

We report similar results for the rural setting in Fig. 5.9. Note that the gain is a little higher compared to the urban case, but the dominant type of interference remains the self interference.

Note that all studies that are based on single cell, ignore IBI and part of IUI and hence correspond at best to Scheme 2 which is highly optimistic.

5.6.4 Performance of FDC without traffic asymmetry

Next, we present the performance of FDC when the traffic asymmetry between the uplink and downlink is not enforced as we did with Constraint (5.18). It is important to note that almost all the papers in the literature do not take the asymmetry into account.

We perform some computations with and without Constraint (5.18) for the network configuration described earlier. When we do not enforce a traffic asymmetry, each user sees a different ratio between its uplink and downlink throughputs. Moreover, the average uplink to downlink throughput ratio across all users is around 0.8, which is unrealistically high considering today's applications that are downlink dominated. For example, the gain of FDC over TDD is expected to be around 29% when $\theta = 0.6$, however it predicts a gain of 42% gain when we do not enforce the traffic asymmetry.

To illustrate this, we plot in Fig. 5.10 the CDF of the traffic asymmetry that is obtained from the solution of the problem without the constraint that enforces that asymmetry. The CDF is computed over 100 realizations for the urban setting.

Clearly, it gives a higher throughput to the uplink than the downlink for almost 20% of the users, which is not a good way to allocate resources. In doing so the problem without asymmetry constraint is biased in favor of FDC since the gain is higher when the uplink to downlink throughput ratio is close to one. Hence recent papers have over estimated the gain of FDC over TDD by ignoring the traffic asymmetry.

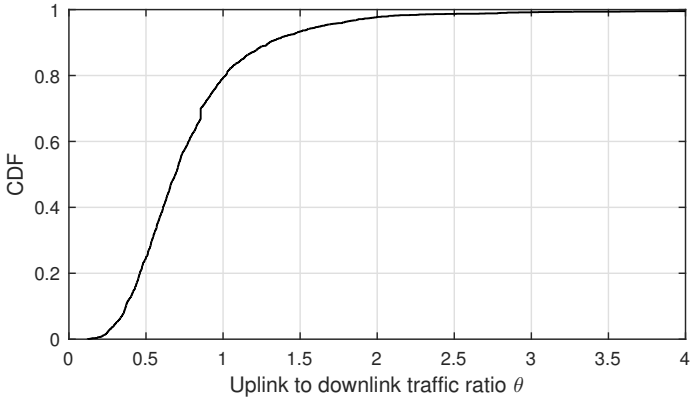


Figure 5.10: CDF of the uplink to downlink throughput ratio without Constraint (5.18) for the urban setting with free space loss model, $N = 30$, and $\mathcal{C} = 110$ dB

5.6.5 Performance of FDC in heterogeneous networks

Finally, we evaluate the performance of FDC in the context of HetNets. As explained in previous chapters, HetNets include low-powered small cells. Since their power budget is comparable to the power budget of the users, small cells might be a better candidate to implement FDC.

To evaluate the performance of FDC in HetNets, we consider the same 3-cell network as before, however in this case, there are two small cells installed at 230 meters right and left of each macro BS. Similar to our study in Chapter 4, we perform channel allocation and user association separately from the user scheduling problem.

For channel allocation, we use the simple orthogonal deployment (OD) method, where k subchannels are allocated to the small cells and the remaining $M - k$ subchannels are allocated to the macro BSs. In this case, the macro BSs and the small cells do not create interference to each other. Hence, for a given user association and for a given k, we can consider the macro BSs and the small cells as two separate networks.

We use the same user association (UA) method as in Chapter 4.6, i.e., small cell first. For this method, just for UA purposes, we compute the SINR per subchannel each user receives from each BS *assuming that the BSs allocate equal power to their subchannels.* Then, a user associates to the small cell if the per subchannel SINR that user receives from that small cell is greater than a threshold ϕ. Otherwise, it associates to the cell that yields the highest SINR.

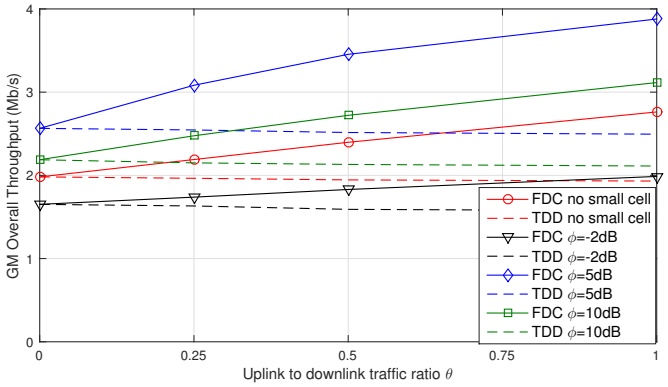


Figure 5.11: Performance of FDC and TDD in HetNets as a function of traffic asymmetry θ for the urban setting with free space inter-BS path loss model, $N = 30$, and 110 dB SC capability

Then, for a given realization, user association (i.e., ϕ), channel allocation (i.e., k), and traffic asymmetry Θ, we solve the centralized scheduling problem $\mathcal{P}_{FDC}(\Theta)$. Similar to what we have done for the homogeneous case, we assume the same traffic asymmetry θ for all users in order to simplify our computations.

The results are given in Fig. 5.11 as a function of the uplink/downlink traffic asymmetry θ for different values of ϕ, i.e., the user association parameter. Note that for each point, we pick the value of k, i.e., the number of subchannels allocated to the small cells, that maximizes the average GM overall throughput of the users.

Clearly, a HetNet deployment requires a good parametrization of the user association scheme since TDD with small cells performs worse when $\phi = -2$ dB than TDD without small cells. Moreover, the performance gain of FDC also depends on user association. However, when UA is done properly, e.g., when $\phi = 5$ dB, the gain increases significantly. For example, when $\theta = 0.5$, the FDC gain over TDD is around 23% without small cells, whereas it is around 39%, with small cells and when $\phi = 5$ dB.

As a result, the gain of FDC is more significant in HetNets due to the smaller power budget of the small cells.

5.7 Conclusion

In this chapter, we study the performance of FDC in multi-cell OFDMA networks. We present an offline study that compares the performance of FDC with a TDD network. We show that the inter-BS path loss model and the self interference cancellation capability of the BSs play a crucial role on system performance. Furthermore, we show the importance of the mix of traffic (i.e., the parameter $\theta_{u,k}$ in our model). Based on the many numerical results obtained for different scenarios, we conclude that the gain of FDC over TDD significantly depends on the network setting and network parameters and it is rarely close to 100% (i.e., a doubling of the throughputs), in fact it might not always be high enough to warrant the very high complexity of a real-time online FDC scheduler. The four scenarios, where FDC can bring more gain are: i) when uplink and downlink traffic have equal weights in volume, ii) rural setting, iii) HetNets, and iv) crowded networks.

Chapter 6

A Benchmark for D2D in Cellular Networks: The Importance of Information

In this section, we will present a benchmark for D2D communications. This study was conducted at the beginning of my Ph.D. studies, hence we use simpler schemes such as reuse-3 among the BSs and simple schedulers for the uplink and the downlink.

6.1 Introduction

A recent trend in cellular networks is the so called D2D communications that enable cellular users, in close proximity to each other, to exchange data directly without using the BS as a relay node [9]. This topic has emerged mostly due to the invent of new smart phone applications that create local, i.e., intra-cellular, traffic among cellular users. Such applications include photo sharing, video sharing, and instant messaging. Although it has been a hot topic in the last few years, many challenges remain to implement D2D communications. For example, interference management is complicated by D2D communications, detecting that a traffic is IC is also hard, and it is not easy to obtain the channel gains between the cellular users and hence to decide if an IC traffic should use D2D communications or be sent via BS(s). In this chapter, we focus on IC traffic when D2D communications are not allowed, i.e., the network operates in a classic cellular mode. We show that there is a significant performance gain if the network processes, such as user scheduling and user

association, are performed with the knowledge of traffic types even when the direct links between users are not utilized. However, this requires the joint operation of the uplink and downlink.

The reason we study this is because the benchmark against the D2D case (i.e., where direct communications are allowed) should be type-aware since it is unfair to assume that this information is available for the D2D case and is not available for the case where direct communications are not allowed.

In a conventional cellular network, uplink and downlink scheduling are performed separately. There are different challenges for uplink scheduling and downlink scheduling. For example, in a multi-cell system, inter-cell interference plays a crucial role in system performance and dealing with interference is not easy especially on the uplink. The power budget on the downlink of a cell comes from a single source (the BS) while it comes from different sources (the users) on the uplink. Apart from those challenges, we show that there is also a need to know the traffic types in order to improve the performance of the cellular network.

Let f_{ij} be the *unidirectional* traffic between any two nodes i and j. If i and j cellular are users belonging to the same cellular network, we say the flow is IC, if i is a cellular user and j is a node outside the cellular network, we call it an uplink flow, and if j is a cellular user and i is a node outside the cellular network, we call it a downlink flow. In a conventional cellular system, an IC flow f_{ij} would use two radio links, one of which is from i to its BS (uplink) and the other one is from the BS of j to j (downlink). Since these two radio links are coupled, we should, if at all possible, couple their scheduling to avoid congestion. Indeed, in a conventional system, if i is very close to its own BS, it might be allocated a high rate and this could be a problem on the downlink of j especially if j is far from its BS. In practical systems, this problem would translate into buffer overflows. The BSs for i and j might be different or might be the same. To illustrate the gain of a type-aware solution, we will restrict our study to a single macro cell (while considering interference from the rest of the cells) and only call IC flows the flows for which i and j share the same BS. We can generalize this in a C-RAN based system [13] to the case where i and j are on different BSs if the BSs are connected to the same C-RAN and scheduling is performed in the C-RAN.

We show that the performance can be significantly improved when the uplink and downlink of an IC flow are jointly scheduled at the BS (this requires the knowledge of the type of a flow). We call this type of schedulers *type-aware* and we call the schedulers that do not take the type of flows into account and hence schedule the uplink and the downlink

independently *type-blind*[1].

In a cellular network, each user has typically multiple flows, possibly of different types. Therefore, the issue of fairness among users or flows arises. In this chapter, we define the concept of *device fairness* that ensures fairness at the device-level irrespective of the number of flows each user has.

We will explain the type-aware scheduler in an homogeneous context and then focus on HetNets, which consist of macro base stations and small cells [49]. In that case, user association, the process of associating each user to either a macro BS or one of the small cells, is critical. We show that when UA is type-aware, system performance can be further improved.

The two main messages of the chapter are that i) important network processes, such as user scheduling and association, should be performed with the knowledge of the type of flows; ii) the uplink and downlink should be jointly scheduled to obtain the best performance when there is IC traffic in the system. This is easy to do when the source and the destination of a flow are associated to the same BS or to different BSs linked to the same C-RAN.

6.2 Related Work

In this section, we explain the work related to our study. We will focus on the papers that are related to mode selection, channel allocation, and user scheduling in D2D communications.

Earlier works addressed simple scenarios including one cellular user and a pair of D2D users in a single cell. Authors of [89] proposed a simple network coding for this scenario. In [90], a mode selection and a power adjustment mechanism are proposed. In both papers, interference is neglected and a very simple channel allocation scheme is implemented. Similarly in [91], a closed-form solution for the rate of a D2D pair is obtained without considering the interference. However, the assumptions and the system model are very simple and performance maximization is done only for the D2D pair while neglecting the performance of the other users.

When spatial reuse between the cellular users (i.e., users with downlink and uplink traffic only) and D2D users (i.e., users with IC traffic only) is utilized, one important question is whether uplink or downlink resources, or both should be shared between the cellular

[1]Some of these results are presented in our work [88]

and D2D users. Most of the time, uplink resources are shared since uplink transmission causes less harm to D2D users due to lower power budget of the users compared to the BSs. In [92], uplink resources are shared with D2D users. However, user scheduling and mode selection are not included in the sense that each cellular user uses one subchannel and D2D users share these subchannels with the cellular users. On the other hand, downlink resources are shared with D2D users in [93] with a similar approach. Both papers consider a single-cell environment which is not a realistic model.

Some of the previous work focused on analytical modeling of D2D communications. In [94], system throughput equations are derived using stochastic geometry. However a very simple channel allocation model is assumed. Authors of [95] extended D2D communication framework to heterogeneous networks and made a similar analysis using stochastic geometry. Game theoretical approaches are used in [92] and [96]. Combinatorial auction is used in [92] for energy efficient channel allocation. Coalition formation game is employed in [96] for channel allocation and interference management.

Authors of [97] consider a relatively more detailed system by modeling the problem as a weighted sum-rate maximization problem. Both uplink and downlink resources are available for D2D users to share with cellular users. Furthermore, mode selection, user scheduling, and power management are also included in the problem. Since the problem is a very large sized nonlinear integer problem, they proposed a heuristic solution to approximate the optimal solution. To this end, it is one of the most detailed formulations. However, similar to the previous examples, this paper considers a single-cell environment. Therefore, inter-cell interference is neglected. Inter-cell interference is central to our work.

One important issue to consider is *fairness*. Since there are users with different number of flows in the system, the question of how to ensure fairness among those users arises. In [98], IC traffic always uses direct links and they share the resources of the cellular users while the cellular users have higher priority meaning that Quality of Service (QoS) requirements of the cellular users are always satisfied, whereas it is not a necessity for the D2D users. Different fairness criteria are defined in [99]. Authors discuss fairness among cellular users, D2D users and they propose a new fairness called *group fairness*. However, only one type of flow per user is considered. Similarly, authors of [100] focuses on max-min fairness for the energy consumption among only users with IC traffic. A game theoretical approach is used in [101] to ensure fairness only among users with IC traffic. In our study, we generalize the notion of fairness in the sense that each user can have multiple types of traffic and we define a utility function that enables us to make a fair comparison among such users. Moreover, no prior work considers a benchmark (i.e., when a network with IC traffic operates in traditional mode without utilizing the direct links between users) that takes into account the traffic types and fairness among users.

6.3 System Details

6.3.1 Overall Network Model

We use the multi-cell system model described in Chapter 2, where each cell has one macro BS equipped with an omni-directional antenna. Additionally, for the HetNet configuration we consider in Section 6.7, there are also two small cells in each cell. In that case, each small cell has a power budget P_S. We focus on the macro cell in the middle (that we call cell 0) while taking into account the interference received from the other cells.

On the downlink, we assume that a BS allocates all its power equally to its allocated subchannels and we know the interferers which are the other BSs transmitting on the same channels as the BS in cell 0. Then, if we assume that the channel gains between i and all these BSs are known exactly, we can compute exactly the interference seen at each user device and hence the SINR.

On the uplink, calculating the interference seen by the BS is not as easy as calculating it for the downlink since we do not know how the power is used and who the interferers in the other cells are (it depends on the schedule in the cells). Hence, we have to estimate the interference and consider decoding errors, due to a bad estimation of the interference.

The final parameter that we have is $0 < \beta < 1$, which determines how much of the scheduling frame time is allocated to the downlink (β) and the uplink ($(1 - \beta)$). Selecting β might not be so straightforward when IC traffic is present.

6.3.2 The Flows

In our system, a user i in cell 0 may have up to four types of flows:

- Uplink (UL): Uplink flow to the Internet or to a user in another cell
- Downlink (DL): Downlink flow from the Internet or from a user in another cell
- IC_o: From i to a device in cell 0
- IC_d: From a device in cell 0 to i

Hence, we categorize a flow from i to a device outside cell 0 as an uplink flow and a flow from a device outside cell 0 to i as a downlink flow. We assume that each user has, at

most, one flow of the first three types (it can have multiple flows of the fourth type). Let $\zeta(i)$ be the destination of the IC_o flow of user i (if any) and let $z(i)$ be the set of sources of the IC_d flows of user i (if any). We define the throughput λ_i^{UL}, λ_i^{DL}, $\lambda_i^{IC_o}$, and $\lambda_i^{IC_d}$ for the UL, DL, IC_o and IC_d flows of user i, respectively. An IC flow contains two hops: the hop between the source (either i or $u \in z(i)$) and the BS and the hop between the BS and destination (either i or $\zeta(i)$).

We assume that the binary matrix X, whose dimension is $4\text{x}N$, showing whether a user has a given type of flow or not, is given. More specifically:

- $X(1,i) = 1$ if device i has a UL flow
- $X(2,i) = 1$ if device i has a DL flow
- $X(3,i) = 1$ if device i has a IC_o flow
- $X(4,i) = 1$ if device i has IC_d flow(s)

6.4 A Single Metric for Fairness and Efficiency

Since each user might have a different number of flows, it is important to decide how to define fairness among those users. A network operator can offer fairness to flows irrespective of the devices on which they are or it can offer fairness to devices without considering the number of flows each device has. We focus on a *device fairness*, where each device is treated as a single entity rather than considering each flow separately. This is because offering flow fairness might cause unfairness among the users with different numbers of flows by assigning higher weights to the users with higher numbers of flows. To this end, we define a utility function for each user that is the GM of the throughput of each flow of types 1, 2, and 3. The reason why we do not include type 4 flows, i.e., IC_d flows, is that we do not want to double count flows. This will become clearer when we define the objective function across all users in the cell. Let $F(i)$ be the set of flows of types 1, 2, and 3 of user i where $1 \leq \|F(i)\| \leq 3$. The utility φ_i of user i is defined as:

$$\varphi_i = \sqrt[|F(i)|]{\prod_{j \in F(i)} \lambda_i^j}, \tag{6.1}$$

where $F(i) = \{j \in \{1,2,3\} \mid X(j,i) = 1\}$. If we use an arithmetic mean instead of a geometric mean in Eq. (6.1), we might assign zero resource to some IC flows.

Note that this metric is defined for a user. In the following, our purpose will be to maximize the GM of the user utilities to be efficient and fair among the users.

6.5 Type-Blind Scheduling

In this section, we explain how user scheduling is performed in a conventional homogeneous cellular network. We consider the homogeneous network shown in Fig. 6.1[2]. Typically, user scheduling is performed separately for the uplink and the downlink and the scheduler does not know if a flow is IC or not, i.e., it is *type-blind* and in this case the scheduler considers it two separate flows, one uplink and one downlink. We explain its operation in this subsection. It allocates all the resources to the downlink for a time fraction β and then on the uplink for the remaining frame time.

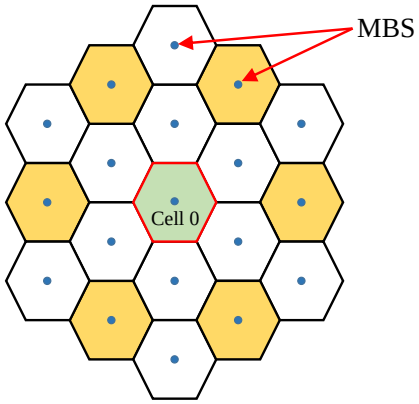


Figure 6.1: Homogeneous network configuration with a reuse factor (r) of 3. The cells interfering with cell 0 are shown in yellow.

6.5.1 Downlink Scheduler

We consider a simple downlink scheduler [34], where a BS allocates equal power to all its available subchannels and allocates all subchannels to a user at a given time. Then, the users are time scheduled. With these assumptions along with the one stating that the channel gains are known perfectly, we can compute the exact interference (and the SINR) seen at each user and hence, knowing the rate function $f(.)$, we can pick the best MCS, i.e., the one that yields the maximum achievable rate for each user. For a given realization ω, where a realization corresponds to the random deployment of the users within cell 0

[2]Reuse-3 is used because this study is chronologically the oldest study in this thesis. We have used reuse-1 in the remaining of this thesis.

and their corresponding channel gains, let r_i be the rate user i sees (over all the channels) when it is scheduled and $\mathcal{U}_0$ be the set of users associated to BS_0. Then, assuming the full buffer case, the following problem maximizes the proportional fair objective function over the downlink:

$$\mathbb{P}_{\mathbb{DL}}(\omega): \quad \max_{\alpha_i \geq 0, \lambda_i} \sum_{i \in \mathcal{U}_0} \log(\lambda_i) \tag{6.2}$$

$$s.t. \quad \lambda_i = \alpha_i r_i \ , \ \forall i \in \mathcal{U}_0 \tag{6.3}$$

$$\sum_{i \in \mathcal{U}_0} \alpha_i \leq \beta \tag{6.4}$$

where λ_i is the throughput user i sees on the downlink and α_i is the fraction of time user i is scheduled. Note that $\mathbb{P}_{\mathbb{DL}}(\omega)$ is a convex problem and it was previously shown that the optimal scheduler allocates equal time to each user [34], i.e., $\lambda_i = \frac{\beta}{|\mathcal{U}_0|} r_i$.

6.5.2 Uplink Scheduler

As discussed before, scheduling on the uplink is more challenging. First, the interference cannot be known exactly since the transmitters and the power they allocate on each subchannel in the neighboring cells are unknown. Furthermore, power allocation is not as simple as on the downlink since there are multiple possible transmitters in a cell.

The scheduler proposed in [14] allocates $m_i \geq 1$ subchannels to user i for the duration of a frame, where m_i is an integer. The power budget of user i is divided equally between these m_i channels. The scheduler uses the same interference estimate I^{est} on all subchannels. This scheduler is not flexible enough for the type-aware scheduler because it might be necessary to allocate a user less than T PRBs (i.e., less than one subchannel during the whole frame). Hence, we propose a scheduler that can be seen as a more flexible version of the one in [14]. We also use the same interference estimate I^{est} on all subchannels.

We assume the subchannels are organized into blocks of different sizes and that a user can only be allocated one block at a time for transmission. If a user is allocated a block of $1 \leq k \leq M$ subchannels at a given time, we assume that its power budget is shared equally among the k subchannels. Let the number of blocks of size k be t_k. Our uplink scheduler computes for every frame the values of t_k (since the realization can change from one frame to another) and allocates a block of size k to user i for a fraction of time ξ_i^k.

An example of a power allocation and a schedule with 5 subchannels and 4 users (U1 to U4) is given in Figure 6.2. There are three subchannel blocks, shown as B_1, B_2, and B_3,

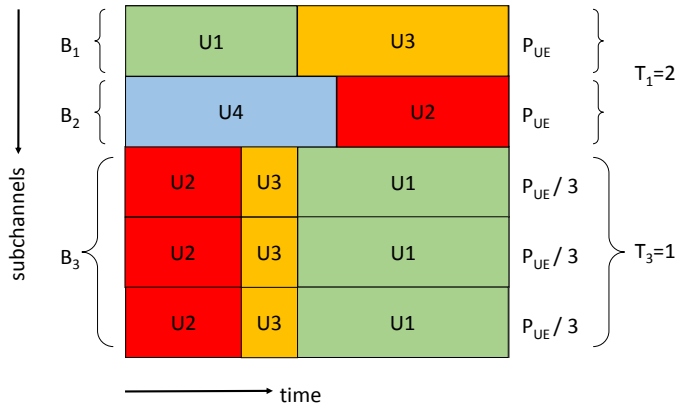


Figure 6.2: An example scheduling on an example power allocation scheme

which are also variables of the scheduler. The first two blocks have one subchannel each and hence are used with P_U whereas the third block has 3 subchannels and hence is used with $P_U/3$. Therefore T_1 is equal to 2 and T_3 is equal to 1. This figure shows a feasible schedule since a user uses only one block at a given time with all its power allocated to the subchannels of the block.

Let $R_i^k(I^{est})$ be the rate seen by user i on a subchannel block of size k when the interference estimate is I^{est}. This can be computed by first computing the SINR with equation (2.2) with I^{est} as the interference estimate and then mapping this SINR to a data rate using the rate function $f(.)$ and k. Specifically, for a realization ω, given I^{est}, $R_i^k(I^{est})$ and $\mathcal{U}_0$, the uplink scheduler solves the problem $\mathbb{P}_{\mathbb{UL}}(\omega, I^{est})$.

$$\mathbb{P}_{\mathbb{UL}}(\omega, I^{est}) : \max_{(\xi_i^k),(t_k),(\lambda_i((I^{est}))} \sum_{i \in \mathcal{U}_0} \log(\lambda_i(I^{est})) \tag{6.5}$$

$$s.t. \quad \lambda_i(I^{est}) = \sum_{k \in \{1...M\}} \xi_i^k R_i^k(I^{est}) \ , \ \forall i \in \mathcal{U}_0 \tag{6.6}$$

$$\sum_{k \in \{1...M\}} \xi_i^k \leq (1-\beta), \ \forall i \in \mathcal{U}_0 \tag{6.7}$$

$$\sum_{i \in \mathcal{U}_0} \xi_i^k \leq t_k(1-\beta), \ \forall k \in \{1..M\} \tag{6.8}$$

$$\sum_{k \in \{1...M\}} k t_k \leq M \tag{6.9}$$

$$t_k \in \mathbb{Z}^+, \ \xi_i^k \geq 0, \ \forall k \in \{1..M\}, \ \forall i \in \mathcal{U}_0 \tag{6.10}$$

The estimated throughput $\lambda_i(I^{est})$ of user i is defined as the sum of rates it sees on each block (constraint (6.6)). Constraint (6.7) ensures that the total time a user is scheduled cannot exceed the uplink frame duration. Constraint (6.8) ensures that the total time users are scheduled on blocks of size k cannot exceed $t_k(1-\beta)$. Constraint (6.9) enforces that the total number of subchannels allocated to the blocks cannot exceed the total number of subchannels M.

A crucial part of this scheduler is the computation of the $R_i^k(I^{est})$'s. If we use an optimistic interference estimate (a small value for I^{est}), we might see many losses since the real interference might be higher. Recall from previous chapters that we define the *goodput* seen by a user as the effective rate this user sees after considering PRB losses. For a low value of I^{est}, we show that the GM estimated by solving $\mathbb{P}_{\mathbb{UL}}(\omega, I^{est})$ is different from the goodput GM. Next, we evaluate the performance of the uplink scheduler with the same methodology as we evaluated the performance of local uplink schedulers in Chapter 3.

We consider a snapshot scenario in which we create a global realization made of $N = 10$ users per cell. We schedule each cell locally using the same interference estimate I^{est} and obtain the estimated GM for cell 0 as the value of the local objective function. The resultant schedule is mapped to the PRBs for each cell. We can then compute the goodput GM since we now have the real interference values (once we know the scheduling in each cell, we know the exact interference). We perform this simulation for multiple time slots with different PRB allocation and take the time average. The decoding rule is as follows for a given PRB: If the real SINR is higher than the threshold of the MCS, the user gets the rate of that MCS from the PRB. Otherwise, the PRB cannot be decoded and we consider it lost. We repeat this for 100 realizations and take the average GM goodput. Fig. 6.3

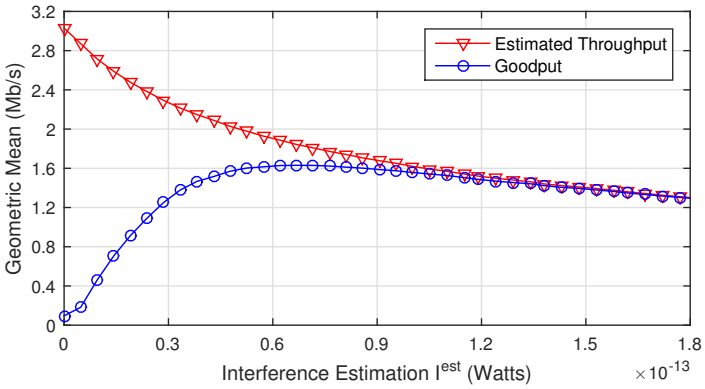


Figure 6.3: Comparison of GM estimated throughput and GM goodput as a function of I^{est} for the uplink scheduler

shows the results.

We can see that the GM throughput estimated by $\mathbb{P}_{\mathbb{UL}}(\omega, I^{est})$ is significantly lower than the GM goodput for low values of I^{est}. However, the GM estimated throughput and the GM goodput overlap after some point. In the following, we will select the lowest value of I^{est} that yields a difference of less than 0.5% between the two curves, which corresponds to $I^{est} = 1.6 \times 10^{-13}$ Watts. Then, we will assume that the estimated throughput is the same with the goodput of each user.

6.5.3 Operation of the type-blind benchmark scheduler

Now, we explain how the benchmark scheduler works using $\mathbb{P}_{\mathbb{DL}}(\omega)$ and $\mathbb{P}_{\mathbb{UL}}(\omega, I^{est})$ defined above. It does not consider the type of flows and schedule the users on the uplink and downlink separately.

The uplink scheduling is performed for all the users who have an uplink flow or who are the source of an IC flow. Then, the throughput is equally divided for each user among the flows that use uplink transmission. Then, downlink is scheduled for all the users who have a downlink flow or who are the destination of an IC flow(s). In this case, when the throughput of the first hop an IC flow is higher than the throughput of the second hop of the IC flow, then the uplink resources are wasted. Moreover, the scheduler does not consider the utility metric per user since it does not know the flow types, hence it results in an unfair schedule.

6.6 Type-Aware Scheduling

6.6.1 Formulation

Recall that an IC flow (within a cell) uses simultaneously the uplink and the downlink of that cell. If its uplink and downlink scheduling are not coupled as in the type-blind case, it is possible that the flow receives a higher goodput on the uplink than on the downlink and this would create a buffer overflow at the BS. To avoid overflows and wastage, we have to constrain the goodput seen by an IC flow on the uplink to be equal to the goodput seen on the downlink. This couples the scheduling on the uplink and the downlink and makes the computations of the schedules more difficult. We show next how to do it and what can be gained in terms of performance by doing it. We formulate the problem for the optimal type-aware scheduler. Our aim is to be proportionally fair in the utilities (Eq. (6.1)) of the users. Since we focus on device fairness, each flow of each user is not treated as a separate entity, but we consider a user as a single entity in our scheduling problem. To avoid double-counting, we do not include IC_d flows in the computation of the utility of a user since each IC_o flow for i is an IC_d flow for another user. We consider only the first 3 rows of the matrix X.

For a user i with an IC_o flow, i.e., an intra-cellular flow originated in i, the IC_o throughput λ_i^{IC} is defined as the minimum of two different equations, one corresponding to the throughput on the uplink hop and the other to the throughput on the downlink hop. Both throughputs must be equal to each other to avoid wastage or overflow. We do rate matching to avoid possible overflow at the downlink buffers, which we do not model. The throughput of the uplink (resp. downlink) flow of user i is denoted as λ_i^{UL} (resp. λ_i^{DL}). If there is no flow of this type, the throughput is zero.

We use the uplink and downlink schedulers defined in the previous section. However, we need to extend the notation since each user might have different types of flows. Previously, we used α_i for the fraction of time user i is served on the downlink. Now, we define α_i^{DL} and α_i^{IC} for the fraction of time user i is served on the downlink for its DL flow and IC_o flow, respectively. Similarly, we extend the notation of ξ_i^k to $\xi_i^{k,UL}$ and $\xi_i^{k,IC}$ as the fraction of time user i uses subchannel block k on the uplink for its UL and IC_o flows, respectively.

For a realization ω, we formulate $\mathbb{P}_{\mathbb{OPT}}(\omega, I^{est})$, given $\{(F(i)),\ X,\ (\zeta(i)),\ \beta,\ I^{est},\ (R_i^k(I^{est})),\ (r_i)\}$. The variables are $\{(\lambda_i^{IC}),\ (\lambda_i^{DL}),\ (\lambda_i^{UL}),\ (\alpha_i^{DL}),\ (\alpha_i^{IC}),\ (\xi_i^{k,IC}),\ (\xi_i^{k,UL}),\ (\varphi_i)$, and $(t_k)\}$, which are all non-negative.

$$\mathbb{P}_{\mathbb{OPT}}(\omega, I^{est})): \quad \max \sum_{i \in \mathcal{U}_0} \log(\varphi_i) \tag{6.11}$$

$$s.t. \quad \varphi_i = \sqrt[|F(i)|]{\prod_{j \in F(i)} \lambda_i^j} \ , \ \forall i \in \mathcal{U}_0, \tag{6.12}$$

$$\lambda_i^{UL} = \sum_{k \in \{1...M\}} \xi_i^{k,UL} R_i^k(I^{est}) \ , \ \forall i \in \mathcal{U}_0 \tag{6.13}$$

$$\lambda_i^{IC} = \sum_{k \in \{1...M\}} \xi_i^{k,IC} R_i^k(I^{est}) \ , \ \forall i \in \mathcal{U}_0 \tag{6.14}$$

$$\lambda_i^{DL} = \alpha_i^{DL} r_i \ , \ \forall i \in \mathcal{U}_0 \tag{6.15}$$

$$\lambda_i^{IC} = \alpha_i^{IC} r_{\zeta(i)} \ , \ \forall i \in \mathcal{U}_0 \tag{6.16}$$

$$\sum_{i \in \mathcal{U}_0} (\xi_i^{k,UL} + \xi_i^{k,IC}) \leq t_k(1-\beta) \ , \ \forall k \tag{6.17}$$

$$\sum_{k \in \{1...M\}} (\xi_i^{k,UL} + \xi_i^{k,IC}) \leq (1-\beta) \ , \ \forall i \in \mathcal{U}_0 \tag{6.18}$$

$$\sum_{k \in \{1...M\}} k t_k \leq M \tag{6.19}$$

$$\sum_{i \in \mathcal{U}_0} (\alpha_i^{DL} + \alpha_i^{IC}) \leq \beta \tag{6.20}$$

$$\xi_i^{k,UL} \leq X(1,i) \ , \ \forall i \in \mathcal{U}_0 \ , \ \forall k \in \{1...M\} \tag{6.21}$$

$$\xi_i^{k,IC} \leq X(3,i) \ , \ \forall i \in \mathcal{U}_0 \ , \ \forall k \in \{1...M\} \tag{6.22}$$

$$\alpha_i^{DL} \leq X(2,i) \ , \ \forall i \in \mathcal{U}_0 \tag{6.23}$$

$$\alpha_i^{IC} \leq X(3,i) \ , \ \forall i \in \mathcal{U}_0 \tag{6.24}$$

$$t_k \in \mathbb{Z}^+ \ , \ \forall k \in \{1...M\} \tag{6.25}$$

The throughput of the UL flow of user i is defined by constraint (6.13) and of the DL flow by constraint (6.15). The throughput of the IC_o flow of user i is defined by (6.14) and (6.16). Constraints (6.21-6.24) ensure that a device does not get resources if it does not have a flow that uses that type of resources. Constraints (6.17) and (6.20) ensure that the time users are scheduled on a subchannel cannot exceed the total uplink and downlink subframe time, respectively. Since the objective function is the GM of user utilities and each utility is the GM of flow throughputs, we guarantee that no flow gets a zero rate.

$\mathbb{P}_{\mathbb{OPT}}(\omega, I^{est})$ is an NP-hard problem since it is a mixed integer program. For reasonable size problems, it can be solved by commercial solvers such as Bonmin [33]. Its high computational complexity is not a problem for our offline study, which is focused on showing the gain by jointly scheduling the uplink and downlink when the types of the flows are

known.

6.6.2 Numerical Results

Now, we compare the performance of the Type Aware Scheduler (TAS) with the Type Blind Scheduler (TBS). We consider the 19 cell system shown in Fig. 6.1 and focus on cell 0. We only consider the six other cells that use the same set of subchannels as cell 0. The power budgets and the channel models are the same with the ones used in the previous chapters. The number of users in $\mathcal{U}_0$ is set to 10. We assume there are the same number of users in each of the six cells.

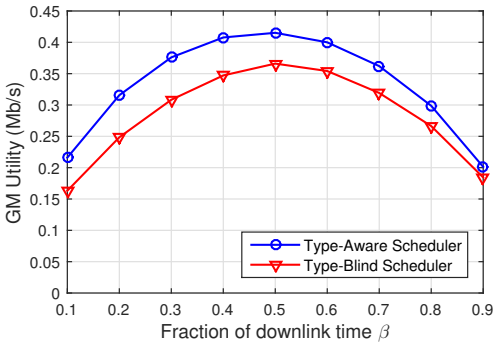
We consider a snapshot scenario in which we create a global realization made of $N = 10$ users per cell. The destination node $\zeta(i)$ of the IC_o flow of each user i is selected randomly. We select I^{est} so that losses are negligible as discussed previously. Our performance metric is the GM of the φ_i of the users. We consider two scenarios:

- **Scenario 1:** 10 users in cell 0 with all three types of flows, i.e., $X(1,i) = X(2,i) = X(3,i) = 1$,
- **Scenario 2:** D (≤ 10) users with an IC_o flow, i.e. $X(3,i) = 1$, and $\{10 - D\}$ users with UL and DL flows and no IC_o flow, i.e., $X(1,i) = X(2,i) = 1$.

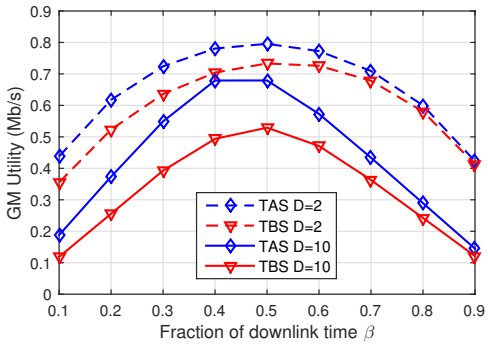
We start with the first scenario. Fig. 6.4a shows the performance difference between the two schedulers. Both schedulers achieve their peak performance when β is 0.5, i.e., equal time for the uplink and downlink. In that case, there is a 14% gain for the TAS.

Next, we consider the second scenario. Note that the two schedulers perform exactly the same if there is no IC traffic in the network. Fig. 6.4b shows the performance of the two schedulers when D is 2 and 10. The difference in GM utility is 9% for $D = 2$, whereas it reaches 28% for $D = 10$. To examine this further, we plot Fig. 6.4c which shows the performance of the two schedulers as D increases when $\beta = 0.5$. The gain obtained with the TAS increases with the amount of IC traffic.

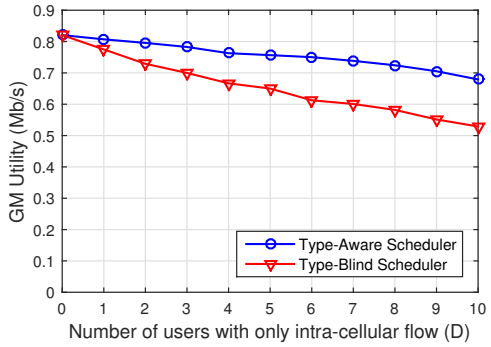
By avoiding wastage and overflows, the TAS can do much better that the type-blind one. Hence, the researchers who study the performance gain of D2D communications should use the TAS as their benchmark since this is what can be achieved with a well designed scheduler when direct communications is not enabled. There are two main reasons for this difference. First, the type-blind scheduler does not limit the throughput of the uplink hop of an IC flow if the downlink hop has worse channel characteristics. This avoids some of



(a) Scenario 1: GM utility vs β



(b) Scenario 2: GM utility vs β



(c) Scenario 2: GM utility vs D with β=0.5

Figure 6.4: GM comparison for the type-aware and type-blind schedulers for different scenarios in the homogeneous case

the data to be transmitted on the second hop. Since the type-blind scheduler does not know the flow types, it also cannot share the resources in an efficient way between the flows.

6.7 Joint User Association and User Scheduling in Heterogeneous Networks

We now examine the effects of type knowledge on the UA process of an HetNet. We consider the cellular network shown in Fig. 6.1 with two small cells added to each cell at a distance of 230 meters left and right of the MBS. We focus on cell 0 but now the users in $\mathcal{U}_0$ have the choice to associate with the MBS or one of the two small cells. An IC flow remains a flow between any two users in $\mathcal{U}_0$, irrespective of their association.

We consider an orthogonal deployment, where c subchannels are allocated to the small cells and $M - c$ to the MBS. For downlink scheduling, we assume each BS (macro BS and small cell) allocates equal power to its subchannels and serve one user at a time. For the uplink, we adapt $\mathbb{P}_{\mathbb{UL}}(\omega, I^{est})$ to the HetNet case. For the interference estimation, we consider two estimates, one for the MBS and one for the small cells. These two estimates are independent of each other due to the orthogonal deployment. Similar to the homogeneous case, we consider a conservative estimation in order to avoid PRB losses.

UA is a critical process that associates each user to a single BS. Furthermore, the best performance can be obtained only when it is jointly performed with user scheduling [34]. Here, we compare the performance of joint type-aware UA/scheduling and type-blind UA/scheduling. For the type-blind UA, we use the optimal UA for the downlink, which can be found by solving the integer program described in [34]. Once the UA is given, the user scheduling can be performed independently at each BS for the type-blind scheme as explained in Section 6.6.

For the type-aware case, we perform the UA and scheduling jointly while coupling the uplink and downlink. We assume that the BSs of a macro-cell are coordinated using a C-RAN [13]. The problem formulation is not given here for brevity. Briefly, we add binary variables to show to which BS a user is associated, and solve $\mathbb{P}_{\mathbb{OPT}}(\omega, I^{est})$ with the new UA and HetNet constraints.

We consider Scenario 2, which was explained in the previous section with $\beta = 0.5$. We use the system parameters described in [34] and set P_S to 30 dBm. The performance difference of the type-aware UA and type-blind UA is given in Fig. 6.5 as a function

of c, the number of subchannels allocated to the small cells. In the figure, type-aware scheme (TAS) corresponds to the case where UA and scheduling are done jointly with the knowledge of flow types, whereas type-blind scheme (TBS) corresponds to the case where UA and scheduling are done without the knowledge of the flow types.

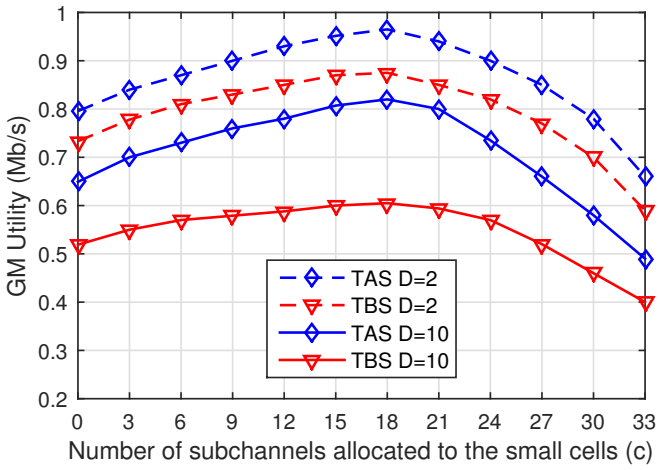


Figure 6.5: GM comparison for the type-aware scheme (TAS) and type-blind scheme (TBS) as a function of c for a HetNet

We consider two cases where D, the number of users with IC flows, is 2 and 10. Clearly, when the UA and scheduling are performed with the knowledge of the type of traffic, the performance is much better. The difference increases as D increases. The maximum performance is achieved for both schemes when $c = 18$ and in that case, the GM difference is 11% and 36% for $D = 2$ and $D = 10$, respectively.

6.8 Conclusion

We analyze cellular networks with IC traffic in this chapter. We show that when the scheduler has more information about the traffic types of each user, the performance can be improved significantly. Essentially, user scheduling should be performed jointly on the uplink and downlink for IC traffic to avoid resource losses caused by bottlenecks. We also claim that a fair benchmark to use to evaluate the performance gains of D2D direct communications should be type-aware since any solution involving D2D direct communications

would require the knowledge of the type of the traffic. We also show that UA in a Hetnet should be type-aware.

Chapter 7

Conclusion

7.1 Summary

In this thesis, we studied various aspects of OFDMA cellular networks. We basically focused on four problems that are uplink scheduling, downlink scheduling, full-duplex communications, and intra-cellular traffic,.

We began our study with uplink scheduling in 5G networks by considering a multi-cell OFDMA network since uplink will be the new frontier in 5G networks and it uses a different technology, i.e., OFDMA, than the one used in LTE, i.e., SC-FDMA. It is a very challenging problem since inter-cell interference has a crucial impact on the system performance and is not easy to estimate. We studied a centralized scheduler, which turned out to be a very large sized MINLP problem. We transformed it into a signomial problem and solve it with an iterative algorithm. Since it is not feasible for a real system due to implementation complexity and not being able to have all the channel gains in the system, we focused on practical local schedulers. We showed the trade-off between the throughput and PRB losses for practical local uplink schedulers. Since their performance is much worse than the centralized scheduler, we proposed two schemes to improve their performance that are loss-aware MCS selection for the case without C-RAN and coordinated link adaptation for the case with C-RAN. Although these schemes significantly improves the performance of the practical benchmark schedulers, there is still a big gap with the performance of the centralized scheduler. Then, we proposed an SFR-based scheduler that we parametrized in a robust fashion. Its performance is close to the centralized scheduler when coordinated link adaptation is used, hence it is a more feasible scheduler when we consider its simple implementation.

We then had a similar analysis for downlink scheduling. Although downlink scheduling is simpler than uplink scheduling due to simpler interference management, finding the optimal system-wide scheduler is still as challenging as the case of uplink. We showed that the RR scheduler that is commonly used by today's network operators performs very poorly compared to the centralized scheduler. We then studied SFR-based schedulers and showed that a well-parametrized scheduler can perform quite well, hence it questions the necessity of coordination among the BSs.

After that, we studied full-duplex communications in 5G cellular networks. We used a similar centralized scheduler to see the potential gain of FDC. We showed that the gain significantly depends on the network configuration as well as uplink/downlink traffic asymmetry and it might not be sufficient enough to warrant the complexity of FDC implementation in a real network. We showed which network scenarios are more promising for FDC deployment.

Finally, we focused on cellular networks in the existence of intra-cellular traffic. We showed how we can ensure fairness among users with multiple types of flows and also how scheduling is performed differently when we have the knowledge of flow types compared to a type-blind scheduler. The main message was to show the importance of flow types in cellular networks. It is clear that the performance can be significantly increased with that knowledge even without utilizing the direct links between the users.

7.2 Future Research Directions

In our study, we have considered proportional fairness throughout this thesis. We will extend this to the concept of generalized α-fairness, from which many different fairness criteria such as PF or max-min can be derived. We will apply this to uplink and downlink scheduling first and then to D2D and FDC-enabled networks.

Although our main focus was uplink scheduling in this thesis, we have not considered the problem of uplink scheduling in HetNets. We will extend our uplink study to HetNets to find practical schemes.

Backhaul limitation is one other aspect that we will consider as future work. It has been shown before that limited backhaul significantly affects the scheduling process on the downlink [102]. We will study its impact on the uplink as well as FDC-enabled networks and cellular networks with intra-cellular traffic.

One important future work will be to study intra-cellular traffic in D2D mode, where direct links between the users are enabled. For this case, we first study the multi-cell

centralized scheduling problem to see how much gain is possible due to D2D mode. Then, we will focus on practical schemes that are implemented fast.

Finally, we will focus on network slicing, which will be a very important component of 5G networks.

References

[1] D. Lopez-Perez, A. Ladanyi, A. Juttner, H. Rivano, and J. Zhang, "Optimization method for the joint allocation of modulation schemes, coding rates, resource blocks and power in self-organizing LTE networks," in *2011 IEEE INFOCOM*, April 2011, pp. 111–115.

[2] Cisco White Paper, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper," Cisco, Tech. Rep., 2017.

[3] T. Bu, L. Li, and R. Ramjee, "Generalized Proportional Fair Scheduling in Third Generation Wireless Data Networks," in *Proceedings IEEE INFOCOM 2006*, April 2006, pp. 1–12.

[4] A. Shaverdian, J. Ghimire, and C. Rosenberg, "Simple and efficient network-aware user association rules for heterogeneous networks," *Computer Networks*, vol. 156, pp. 20 – 32, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128618304183

[5] B. Yang, W. Guo, Y. Jin, and S. Wang, "Smartphone data usage: downlink and uplink asymmetry," *Electronics Letters*, vol. 52, no. 3, pp. 243–245, 2016.

[6] V. K. Adhikari, M. Varvello, V. Hilt, M. Steiner, and Z. Zhang, "Unreeling Netflix: Understanding and improving multi-CDN movie delivery," in *2012 IEEE INFOCOM*, March 2012, pp. 1620–1628.

[7] G. Xu, "Global optimization of signomial geometric programming problems," *European Journal of Operational Research*, vol. 233, no. 3, pp. 500 – 510, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221713008394

[8] R. Li, Y. Chen, G. Y. Li, and G. Liu, "Full-Duplex Cellular Networks," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 184–191, April 2017.

[9] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, March 2012.

[10] 3rd Generation Partnership Project, "Further advancements for E-UTRA physical layer aspects (Release 9)," 3GPP, Tech. Rep. TR 36.814, 2010.

[11] International Telecommunication Union, "IMT traffic estimates for the years 2020 to 2030," ITU, Tech. Rep. ITU-R M.2370-0, 07 2015.

[12] A. Al-Fuqaha *et al.*, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[13] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks: A Technology Overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[14] A. Kumar and C. Rosenberg, "Energy and Throughput Trade-Offs in Cellular Networks Using Base Station Switching," *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, Feb 2016.

[15] Y. Ozcan and C. Rosenberg, "Efficient Loss-Aware uplink scheduling," in *2018 IEEE Wireless Communications and Networking Conference (WCNC) (IEEE WCNC 2018)*, Barcelona, Spain, Apr. 2018.

[16] Y. Özcan, C. Rosenberg, and F. M. Guillemin, "Uplink scheduling in Multi-Cell OFDMA networks with and without coordination," in *2019 IEEE Wireless Communications and Networking Conference (WCNC) (IEEE WCNC 2019)*, Marrakech, Morocco, Apr. 2019.

[17] Y. Ozcan and C. Rosenberg, "Uplink Scheduling in Multi-Cell OFDMA Networks: The Cases With and Without C-RAN," *Submitted to IEEE Transactions on Mobile Computing*, 2019.

[18] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 226–234, February 2009.

[19] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1507–1511, May 2008.

[20] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-Complexity Energy-Efficient Scheduling for Uplink OFDMA," *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 112–120, January 2012.

[21] D. C. Dimitrova, G. Heijenk, J. L. van den Berg, and S. Yankov, "Scheduler-dependent inter-cell interference and its impact on lte uplink performance at flow level," in *Wired/Wireless Internet Communications: 9th IFIP TC 6 International Conference, WWIC*. Springer Berlin Heidelberg, 2011, pp. 285–296.

[22] F. Z. Kaddour, E. Vivier, M. Pischella, and P. Martins, "A New Method for Inter-Cell Interference Estimation in Uplink SC-FDMA Networks," in *2012 IEEE VTC*, May 2012, pp. 1–5.

[23] Q. Li, M. Zhou, Y. Wu, S. Feng, and P. Zhang, "Precise interference estimation for the uplink of LTE heterogeneous networks," in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 1235–1240.

[24] K. Shen and W. Yu, "A coordinated uplink scheduling and power control algorithm for multicell networks," in *49th Asilomar Conference on Signals, Systems and Computers*, Nov 2015, pp. 1305–1309.

[25] J. Niu, D. Lee, T. Su, G. Y. Li, Z. Tang, and Y. Fu, "Multi-cell cooperative scheduling for uplink SC-FDMA systems," in *2013 IEEE PIMRC*, Sept 2013, pp. 1582–1586.

[26] P. Frank, A. Müller, H. Droste, and J. Speidel, "Cooperative interference-aware joint scheduling for the 3GPP LTE uplink," in *IEEE PIMRC*, Sept 2010, pp. 2216–2221.

[27] M. Moretti and A. Todini, "A Resource Allocator for the Uplink of Multi-Cell OFDMA Systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2807–2812, August 2007.

[28] E. Yaacoub and Z. Dawy, "Proportional fair scheduling with probabilistic interference avoidance in the uplink of multicell OFDMA systems," in *2010 IEEE Globecom Workshops*, Dec 2010, pp. 1202–1206.

[29] M. B. S. A. Kafafy and K. M. F. Elsayed, "Autonomous uplink intercell interference coordination in LTE systems with adaptively-tuned interference limits," in *IEEE WCNC*, 2014, pp. 1815–1820.

[30] C. Shi, M. Honig, S. Nagaraj, and P. Fleming, "Uplink distributed power and receiver optimization across multiple cells," in *IEEE WCNC*, 2012, pp. 1608–1612.

[31] "The Network Simulator NS-3," https://www.nsnam.org/.

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

[33] P. Bonami, J. Forrest, C. Laird, F. M. J. Lee, and A. Wächter, "Bonmin: Basic Open-source Nonlinear Mixed INteger programming (July 2006)." [Online]. Available: http://www.coin-or.org/Bonmin

[34] D. Fooladivanda and C. Rosenberg, "Joint Resource Allocation and User Association for Heterogeneous Wireless Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 248–257, January 2013.

[35] Y. Özcan and C. Rosenberg, "Revisiting downlink scheduling in a Multi-Cell OFDMA network: From full base station coordination to practical schemes," in *2019 Wireless Days (WD) (WD'19)*, Manchester, United Kingdom (Great Britain), Apr. 2019.

[36] X. Zhang, E. Zhou, R. Zhu, S. Liu, and W. Wang, "Adaptive multiuser radio resource allocation for OFDMA systems," in *GLOBECOM '05. IEEE Global Telecommunications Conference, 2005.*, vol. 6, Dec 2005, pp. 5 pp.–3850.

[37] L. Venturino, N. Prasad, and X. Wang, "Coordinated Scheduling and Power Allocation in Downlink Multi-cell OFDMA Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2835–2848, July 2009.

[38] A. Gjendemsjo, D. Gesbert, G. E. Oien, and S. G. Kiani, "Optimal Power Allocation and Scheduling for Two-Cell Capacity Maximization," in *WiOpt 2006*, April 2006, pp. 1–6.

[39] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *IEEE INFOCOM 2003*, vol. 1, March 2003, pp. 786–796 vol.1.

[40] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1414–1425, April 2010.

[41] A. S. Hamza, S. S. Khalifa, H. S. Hamza, and K. Elsayed, "A Survey on Inter-Cell Interference Coordination Techniques in OFDMA-Based Cellular Networks," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1642–1670, 2013.

[42] I. G. Fraimis, V. D. Papoutsis, and S. A. Kotsopoulos, "A Decentralized Subchannel Allocation Scheme with Inter-Cell Interference Coordination (ICIC) for Multi-Cell OFDMA Systems," in *IEEE GLOBECOM 2010*, Dec 2010, pp. 1–5.

[43] A. Triki and L. Nuaymi, "Intercell Interference Coordination Algorithms in OFDMA Wireless Systems," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–6.

[44] L. Li, M. Pal, and Y. R. Yang, "Proportional fairness in multi-rate wireless lans," in *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, April 2008, pp. 1004–1012.

[45] G. Giambene and T. A. Yahiya, "LTE planning for Soft Frequency Reuse," in *2013 IFIP Wireless Days (WD)*, Nov 2013, pp. 1–7.

[46] M. Maqbool, P. Godlewski, M. Coupechoux, and J.-M. Kélif, "Analytical performance evaluation of various frequency reuse and scheduling schemes in cellular ofdma networks," *Performance Evaluation*, vol. 67, no. 4, pp. 318 – 337, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166531609001011

[47] 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 *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, Nov 2008, pp. 1–6.

[48] 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.

[49] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, June 2011.

[50] Y. Jin and L. Qiu, "Joint User Association and Interference Coordination in Heterogeneous Cellular Networks," *IEEE Communications Letters*, vol. 17, no. 12, pp. 2296–2299, December 2013.

[51] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 22–30, June 2011.

[52] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi, "Lte-advanced: Heterogeneous networks," in *2010 European Wireless Conference (EW)*, April 2010, pp. 978–982.

[53] K. I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L. G. U. Garcia, and Y. Wang, "Carrier aggregation for lte-advanced: functionality and performance aspects," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 89–95, June 2011.

[54] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and sinr coverage analysis for decoupled uplink-downlink biased cell associations in hetnets," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5360–5373, Oct 2015.

[55] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier hetnets with large-scale antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3251–3268, May 2016.

[56] A. Shaverdian, B. S. Krishnan, and C. Rosenberg, "Almost blank subframes versus partially shared deployment in heterogeneous networks," in *2016 IEEE PIMRC*, Sept 2016, pp. 1–6.

[57] S. Goyal, P. Liu, S. S. Panwar, R. A. Difazio, R. Yang, and E. Bala, "Full duplex cellular systems: Will doubling interference prevent doubling capacity?" *IEEE Communications Magazine*, vol. 53, no. 5, 2015.

[58] A. K. Khandani, "Two-way (true full-duplex) wireless," in *2013 13th Canadian Workshop on Information Theory*, June 2013, pp. 33–38.

[59] D. Wang and X. Wang, "Effective Interference Cancellation Schemes for Device-to-Device Multicast Uplink Period Underlaying Cellular Networks," *Wireless Personal Communications*, vol. 75, no. 4, Apr 2014.

[60] Y. Ozcan and C. Rosenberg, "How Useful is Full Duplex in Cellular Networks? The Impact of Interference and Traffic Asymmetry in a Multi-Cell OFDMA Network," *In Preparation*, 2019.

[61] P. Tehrani, F. Lahouti, and M. Zorzi, "Resource allocation in OFDMA networks with half-duplex and imperfect full-duplex users," in *IEEE International Conference on Communications (ICC)*, May 2016.

[62] C. Nam, C. Joo, and S. Bahk, "Joint Subcarrier Assignment and Power Allocation in Full-Duplex OFDMA Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3108–3119, June 2015.

[63] B. Di, S. Bayat, L. Song, and Y. Li, "Radio resource allocation for full-duplex OFDMA networks using matching theory," in *IEEE Conference on Computer Communications Workshops*, April 2014, pp. 197–198.

[64] A. C. Cirik, K. Rikkinen, and M. Latva-aho, "Joint Subcarrier and Power Allocation for Sum-Rate Maximization in OFDMA Full-Duplex Systems," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[65] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving Single Channel, Full Duplex Wireless Communication," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '10. ACM, 2010, pp. 1–12.

[66] A. Sahai, S. Diggavi, and A. Sabharwal, "On degrees-of-freedom of full-duplex uplink/downlink channel," in *2013 IEEE Information Theory Workshop (ITW)*, Sept 2013, pp. 1–5.

[67] C. Pradhan and G. R. Murthy, "Full-duplex communication for future wireless networks: Dynamic resource block allocation approach," *Physical Communication*, vol. 19, pp. 61 – 69, 2016.

[68] S. Goyal, P. Liu, S. Hua, and S. Panwar, "Analyzing a full-duplex cellular system," in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, March 2013, pp. 1–6.

[69] J. Bai, S. Diggavi, and A. Sabharwal, "On degrees-of-freedom of multi-user MIMO full-duplex network," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 864–868.

[70] M. Al-Imari, M. Ghoraishi, P. Xiao, and R. Tafazolli, "Game Theory Based Radio Resource Allocation for Full-Duplex Systems," in *IEEE Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[71] H. Xiao and Z. Feng, "A Novel Fractional Frequency Reuse Architecture and Interference Coordination Scheme for Multi-Cell OFDMA Networks," in *2010 IEEE 71st Vehicular Technology Conference*, May 2010, pp. 1–5.

[72] J. M. B. da Silva, G. Fodor, and C. Fischione, "Spectral efficient and fair user pairing for full-duplex communication in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7578–7593, Nov 2016.

[73] S. Goyal, P. Liu, and S. S. Panwar, "User Selection and Power Allocation in Full-Duplex Multicell Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2408–2422, March 2017.

[74] A. AlAmmouri, H. ElSawy, and M. Alouini, "Harvesting full-duplex rate gains in cellular networks with half-duplex user terminals," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–7.

[75] S. Sekander, H. Tabassum, and E. Hossain, "Decoupled Uplink-Downlink User Association in Multi-Tier Full-Duplex Cellular Networks: A Two-Sided Matching Game," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2778–2791, Oct 2017.

[76] A. Shojaeifard, K. Wong, M. D. Renzo, G. Zheng, K. A. Hamdi, and J. Tang, "Massive MIMO-Enabled Full-Duplex Cellular Networks," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4734–4750, Nov 2017.

[77] J. Bai and A. Sabharwal, "Asymptotic Analysis of MIMO Multi-Cell Full-Duplex Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2168–2180, April 2017.

[78] I. Atzeni and M. Kountouris, "Full-Duplex MIMO Small-Cell Networks With Interference Cancellation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8362–8376, Dec 2017.

[79] K. S. Ali, H. ElSawy, and M. Alouini, "Modeling Cellular Networks With Full-Duplex D2D Communication: A Stochastic Geometry Approach," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4409–4424, Oct 2016.

[80] X. Chen, M. Suzuki, N. Miki, and N. Nagai, "Simultaneous estimation of echo path and channel responses using full-duplex transmitted training data sequences," *IEEE Transactions on Information Theory*, vol. 41, no. 5, pp. 1409–1417, Sept 1995.

[81] V. R. Cadambe and S. A. Jafar, "Degrees of Freedom of Wireless Networks With Relays, Feedback, Cooperation, and Full Duplex Operation," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2334–2344, May 2009.

[82] M. Sakai, H. Lin, and K. Yamashita, "Self-interference cancellation in full-duplex wireless with iq imbalance," *Physical Communication*, vol. 18, pp. 2 – 14, 2016.

[83] X. Shen, X. Cheng, L. Yang, M. Ma, and B. Jiao, "On the design of the scheduling algorithm for the full duplexing wireless cellular network," in *2013 Proceedings IEEE GLOBECOM*, Dec 2013, pp. 4970–4975.

[84] H. ElSawy, A. AlAmmouri, O. Amin, and M. S. Alouini, "Can Uplink Transmissions Survive in Full-duplex Cellular Environments?" in *22th European Wireless Conference*, May 2016.

[85] N. H. Mahmood, I. S. Ansari, G. Berardinelli, P. Mogensen, and K. A. Qaraqe, "Analysing self interference cancellation in full duplex radios," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[86] A. Shojaeifard, K. Wong, M. D. Renzo, G. Zheng, K. A. Hamdi, and J. Tang, "Self-Interference in Full-Duplex Multi-User MIMO Channels," *IEEE Communications Letters*, vol. 21, no. 4, pp. 841–844, April 2017.

[87] D. Bharadia, E. McMilin, and S. Katti, "Full Duplex Radios," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 375–386, Aug. 2013.

[88] Y. Ozcan, C. Rosenberg, and F. Guillemin, "A benchmark for d2d in cellular networks: The importance of information," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–7.

[89] Y. Pei and Y. C. Liang, "Resource allocation for device-to-device communications overlaying two-way cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3611–3621, July 2013.

[90] C. H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2752–2763, August 2011.

[91] S. Shalmashi, E. Björnson, S. B. Slimane, and M. Debbah, "Closed-form optimality characterization of network-assisted device-to-device communications," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2014, pp. 508–513.

[92] F. Wang, C. Xu, L. Song, and Z. Han, "Energy-efficient resource allocation for device-to-device underlay communication," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 2082–2092, April 2015.

[93] D. Zhu, J. Wang, A. L. Swindlehurst, and C. Zhao, "Downlink resource reuse for device-to-device communications underlaying cellular networks," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 531–534, May 2014.

[94] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6727–6740, Dec 2014.

[95] H. Sun, M. Wildemeersch, M. Sheng, and T. Q. S. Quek, "D2d enhanced heterogeneous cellular networks with dynamic tdd," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4204–4218, Aug 2015.

[96] Y. Li, D. Jin, J. Yuan, and Z. Han, "Coalitional games for resource allocation in the device-to-device uplink underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3965–3977, July 2014.

[97] M.-H. Han, B.-G. Kim, and J.-W. Lee, "Opportunistic resource scheduling for d2d communication in ofdma networks," *Computer Networks*, vol. 73, pp. 319 – 334, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128614002916

[98] W. Zhao and S. Wang, "Resource allocation for device-to-device communication underlaying cellular networks: An alternating optimization method," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1398–1401, Aug 2015.

[99] P. C. Nguyen and B. D. Rao, "Fair Scheduling Policies Exploiting Multiuser Diversity in Cellular Systems With Device-to-Device Communications," *IEEE Trans. on Wireless Communications*, vol. 14, no. 9, pp. 4757–4771, Sept 2015.

[100] K. Yang, J. Wu, X. Gao, X. Bu, and S. Guo, "Energy-efficient power control for device-to-device communications with max-min fairness," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016, pp. 1–5.

[101] F. Boabang, H.-H. Nguyen, Q.-V. Pham, and W.-J. Hwang, "Network-assisted distributed fairness-aware interference coordination for device-to-device communication underlaid cellular networks," *Mobile Information Systems*, vol. 2017, 2017, cited By 3. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013293356&doi=10.1155%2f2017%2f1821084&partnerID=40&md5=92759c46aa45233e344c4d065a044c58

[102] J. Ghimire and C. Rosenberg, "Revisiting scheduling in heterogeneous networks when the backhaul is limited," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2039–2051, Oct 2015.