Politecnico di Torino

Department of Electronics and Telecommunication

Electrical, Electronics and Communications Engineering Doctorate Program

Gabriel Oliveira Ferreira

A Joint Optimization Approach for Power-Efficient Heterogeneous OFDMA Radio Access Networks

Torino
2025

Gabriel Oliveira Ferreira

A Joint Optimization Approach for Power-Efficient Heterogeneous OFDMA Radio Access Networks

Thesis presented to the Doctorate Program in Electrical, Electronics and Communications Engineering of Politecnico di Torino as part of requirements for obtaining the title of Doctor in Electrical, Electronics and Communications Engineering.

Supervisor: Prof. Dr. Fabrizio Dabbene (CNR)
Co-supervisor: Prof. Dr. Chiara Ravazzi (CNR)
Co-supervisor: Prof. Dr. Giuseppe Calafiore (Politecnico di Torino)

Torino
2025

Gabriel Oliveira Ferreira

A Joint Optimization Approach for Power-Efficient Heterogeneous OFDMA Radio Access Networks

Thesis presented to the Doctorate Program in Electrical, Electronics and Communications Engineering of Politecnico di Torino as part of requirements for obtaining the title of Doctor in Electrical, Electronics and Communications Engineering.

Referees:

Dr. Constantino M. Lagoa	Pennsylvania State University
Dr. Paolo Dini	Centre Tecnològic de Telecomunicacions de Catalunya

Torino
2025

Abstract

Heterogeneous networks have emerged as a popular solution for accommodating the growing number of connected devices and increasing traffic demands in cellular networks. While offering broader coverage, higher capacity, and lower latency, the escalating energy consumption poses sustainability challenges. Additionally, since macro and micro base stations operating at the same frequency are considerably close to each other, a non-optimized network in terms of resource allocation could lead not only to low quality of service from the users perspective (capacity, connection, latency, etc), but also high transmission power expenditures, making the technology not viable from an economic viewpoint. Therefore, this work proposes a joint optimization problem to minimize the transmission power of base stations in OFDMA heterogeneous networks, while respecting users' individual throughput constraints. The decision variables are the association between user equipment and base station, users' individual working bandwidth, and base stations transmission power. The technique is formulated in a format such that an initial highly non-convex problem is posed, after a power function approximation and change of variables followed by a logarithmic transformation, as a Geometric Program, which is known to be convex. The proposed approach is then evaluated in multiple realistic scenarios regarding a real-world neighbourhood from a large European city. With data provided by a leading mobile network operator (including the location of the transmitting base stations, their type, steering direction, transmission technology, etc) and information from Open Street Maps, we are able to reconstruct the environment into a ray tracing simulator allowing us to carefully emulate the electromagnetic waves propagation and estimate the channel gains between each user equipment and all base stations. The results are promising both in terms of optimal value, considerably smaller when compared to literature approaches, and in terms of runtime to obtain the optimal solution. Finally, two work extensions are presented: first the channel gains are considered to be random variables due to motion of objects causing shadowing, and the resulting chance constrained optimization problem is then solved with a robust geometric program; subsequently MPC is applied to the network when one of the base stations is abruptly switched off to simulate a failure environment, so the transmission powers can be controlled while minimizing the users' throughput constraints violation.

Key-words: Heterogeneous networks, geometric program, joint optimization, transmission power, working bandwidth, channel gains, random.

Contents

List of Figures

List of Tables

List of Acronyms and Notation

BS	Base Station.
RB	Resource Block.
UE	User equipment.
EE	Energy Efficiency.
GP	Geometric Programming.
NR	New Radio.
RAN	Radio Access Network.
QoS	Quality of Service.
SIR	Signal to Interference Ratio.
SNR	Signal to Noise Ratio.
SCA	Successive Convex Approximation.
PPF	Power Function Approximation.
RAT	Radio Access Technologies.
MPC	Model Predictive Control.
MNO	Mobile Network Operator.
LTE	Long-Term Evolution.
MIGP	Mixed Integer Geometric Programming.
SINR	Signal to Interference and Noise Ratio.
CDMA	Code Division Multiple Access.
TDMA	Time Division Multiple Access.
FDMA	Frequency Division Multiple Access.
OFDM	Orthogonal Frequency Division Multiplexing.
OFDMA	Orthogonal Frequency Division Multiple Access.
HetNet	Heterogeneous Network.

$[x]$	$\{1,\ldots,x\}$.
x_{ij}	element in row i and column j in a given matrix.
$\lfloor x_{ij} \rfloor$	floor(x_{ij}), i.e. largest integer smaller than x.
$\lceil x_{ij} \rceil$	ceil(x_{ij}), i.e. smallest integer larger than x.
$x \in [0,1]^{n \times N}$	matrix of the elements x_{ij}, with $x_{ij} \in [0,1]$.
$x \in \{0,1\}^{n \times N}$	matrix of the elements x_{ij}, with $x_{ij} \in \{0,1\}$.
$\overline{x}_{ij}$	$1 - x_{ij}$ for binary variables.

Chapter 1

Introduction

The efficiency of Radio Access Networks (RAN) is of ever-growing importance to guarantee the combined performance and sustainability of mobile networks. The increasing user traffic demands, the growing number of connected devices and objects, the emergence of new classes of services with very strict Quality of Service (QoS) requirements, and the surging heterogeneity of applications jointly set very high expectations and standards for the operation of 6G RANs [1]. As a consequence, RAN optimization is a subject that has received substantial attention from the scientific community, and a wide range of problems have been explored in this area, including how to ensure energy savings via spectral efficiency, transmission power minimization, base stations (BS) deployments, and user association, among others.

Optimizing the many variables that characterize modern RANs is especially challenging in ultra-dense heterogeneous networks (HetNet), where macro, micro, and femto cells are jointly deployed within a relatively small geographical area in order to provide high wireless capacity and enhanced QoS to the local end terminals [2]. These dense and hierarchical deployments are today a reality, especially in populated urban centers: Figure 1.1 shows an example of such a RAN configuration observed in a production-grade mobile network, which includes overlapping macro and micro cells providing coverage to the city center of a major European metropolis. In these heterogeneous RAN infrastructures, a lack of optimization of transmission powers, resource allocations, and assignments between users and BSs might lead to high inefficiency, such as high transmission powers and low QoS for users [3, 4].

In heterogeneous cellular infrastructure scenarios, not only the configurations of the RAN is especially complex, but, if not performed properly, it also exposes the mobile network operator to high expenditures in terms of energy costs that risk to make the whole technology not viable from an economic viewpoint. The problem is exacerbated in newer generation of Radio Access Technologies (RAT) that are increasingly demanding

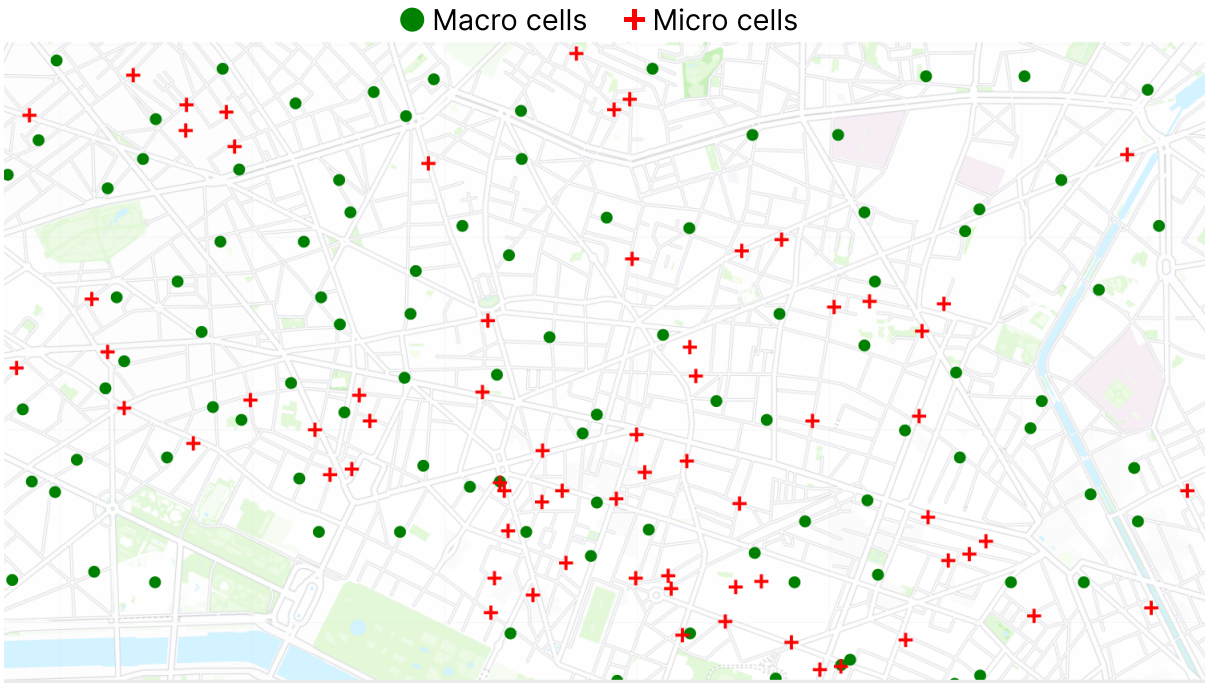


Figure 1.1: Real-world topology of an operational heterogeneous network deployed in a neighborhood of a large European city.

from an energy perspective: for instance, 5G New Radio is known to be two to three times more energy-consuming than its 4G equivalent [5], due to a need to face much higher traffic volumes per user. The problem is such that over 90% of leading mobile network operators have expressed concerns about the rise in energy-induced operating expenses (OPEX) [6]. It is thus unsurprising that many operators and equipment vendors are investing significant effort on developing solutions to make RANs more energy prudent [7].

1.1 Literature Review

The approaches proposed in the literature to tackle the HetNet resource allocation problem can be grouped into three general classes: (i) iterative/sequential-based, (ii) global optimization-based, and (iii) convex optimization-based. With respect to such classification, the solution we propose may be seen as a specific case of the third class, being based on a suitably designed convex approximation of the original problem. We expound on these approaches separately in the next subsections.

1.1.1 Iterative/sequential approaches

Two types of iterative/sequential approaches can be distinguished. In the first one, the non-convex function is linearized and solved approximately around a point of the optimization space. Then, based on the solution, a new point of the optimization space

is selected and this process is repeated until convergence is achieved. The second solves the problem in *ceteris paribus* manner, splitting the problem variables into two subsets of constant and non-constant parameters. Then, a solution to a simplified version of the optimization problem is retrieved considering only the impact of the non-constant parameters. Consequently, the parameters of the two sets are interchanged, *i.e.*, those that were previously constant are now variables and vice versa, and a solution for the new problem is computed. Again, this process is repeated until convergence to a local minimum is achieved.

Along these lines, in [8], the authors studied the problem of joint user association and resource allocation in heterogeneous networks, assuming that the transmission power at all time is the maximum allowed. A relaxation for an NP-Hard optimization problem was developed, where the new convex formulation is used to obtain upper bounds to the network aggregated throughput. The authors then propose an association rule based on users' signal to interference and noise ratio (SINR) levels and prove its efficacy by comparing the optimal values obtained for the relaxed convex problem with the one when the assignments are obtained with the association rule. Through experiment, the authors show that the optimal values are tight and conclude that their association technique is a simple but yet effective rule.

Optimization problem for orthogonal frequency division multiple access (OFDMA) heterogeneous networks is studied in [9] for downlink transmissions. The work proposes a joint optimization for user association, sub-channel allocation, and power allocation. The problem is divided into two: first, the powers are fixed and a solution for users association and sub-channel allocation is obtained with the Hungarian algorithm; subsequently, the variables of the first sub-problem are fixed, and a power allocation problem is solved with the difference of two concave functions approximation method. Then, these problems are solved alternately to obtain a local optimal solution. [10] proposes analytical solutions for association problems subject to backhaul constraints and network topology, allowing each user equipment (UE) at a given location to determine its optimal connecting BS. Subsequently, an iterative algorithm to reach the optimal load of the BS inside the heterogeneous network is derived.

In [11], energy efficiency (EE) resource allocation for OFDMA heterogeneous networks with dense deployment is studied. The proposed algorithms have as constraints minimum throughput for delay sensitive users and fairness for delay tolerant users, in addition to the ones imposed by the OFDMA technology. The resource block and power allocations problems, which are non-convex and with combinatorial characteristics, are tackled with three different approaches, named: mixed-integer programming formulation, time-sharing formulation, and sparsity inducing formulation. All of them lead to non-convex problems,

which is circumvented using successive convex approximation (SCA) methods, obtaining the problem solution through iterative procedures.

The iterative/sequential approaches have two main limitations in our context:

1. The optimal solution of sequential approaches are highly dependent on the initial condition provided to the solver, usually leading to local minimum. More importantly, providing a feasible initial resource allocation for large scale problems can be extremely difficult.

2. The calculation of the gradient (gradient based approach) or the two convex functions to represent the non-convex constraints (DC programming) only work for points that are close enough, i.e inside a validity region, from the initial condition. Therefore, small steps inside the decision variables set are taken. These procedures might be time consuming due to possibly many iterations required. Additionally, defining the validity region size is not a trivial task.

1.1.2 Global optimization techniques

A second class of solution directly tackles the nonconvex optimization problem by means of general global optimization techniques, such as genetic algorithms, particle swarm/ant-colony optimization, Lyapunov drift plus penalty techniques, and Graph Theory based stochastic optimization [12–14]. These strategies have been especially studied in the context of heterogeneous cellular networks. For a better overview on 5G resource allocation for heterogeneous networks the reader may refer to the Survey in [15], where the authors present the taxonomy and strategies applied to such scenarios.

A particular class of solutions we highlight here are those based on neural networks (NNs) approximations. Indeed, NNs, known to be universal approximators of a very large class of functions, have been also applied to resource/power allocation problems. In [16], beamforming coordination and sub-carriers/power allocation problems are addressed with the neural network based algorithm transfer learning multi-agent deep Q-network (TL-MADQN). The algorithm is able to extract knowledge from pre-trained agents (BS or channel gains) and adapt to new networks environments, selecting the most appropriate allocation strategy. In [17], the problem of joint association and resource allocation to maximize the system throughput is formulated as a mixed integer quadratically constrained quadratic program. The authors decompose the problem into two, and the subpart associated to the binary variables is solved with a neural network (NN) approach. The proposed NN is based on encoder-decoder architecture and makes use of the well-known long-short term memory (LSTM) and pointer network mechanism.

Global optimization based approaches generally lead to computationally expensive

solutions. More importantly, they usually fail to provide convergence guarantees (usually converging to local optima), and they may require fine tuning of different hyperparameters, which may be difficult in practice. Also, they usually scale badly with the problem dimensions. Moreover, in the case of NN-based solutions, they require substantial quantity of data during the training phase, which might not be available in real-world contexts.

1.1.3 Convex optimization problems

Recently, a third line of attack to the problem has emerged, inspired by the work of Boyd [18]. These convex optimization-based approaches aim at transforming the original problem by rewriting/approximating it in a convex form. For instance, [18] formulates a convex optimization problem based on geometric programming (GP) to solve the problem of simultaneous routing and power allocation in code-division multiple access (CDMA) wireless data networks. GP is also used in [19], where the problems of minimizing transmitter power subject to outage probability and minimizing outage probability subject to transmission power constraints are written in such a format. In that work, an outage happens when the quality of service of a user, measured in terms of signal to interference ratio (SIR), is less than a given threshold value. Many of the works discussed in the previous topic also apply convex optimization techniques, however solving the problem multiple times through successive approximation methods.

In the authors' opinion, this latter class of approaches represents the most promising solution to our problem for several reasons: i) convex reformulations lead to problems which are easily solvable with available convex optimization tools, ii) they are guaranteed to converge to the global optimum of the approximation/reformulation, and iii) more importantly, they scale polynomially with the problem size.

On the other hand, this approach has not been further developed for some technical reasons: it is in general difficult to extend the original formulation of [18] to more complex setups, due to the fact that the logarithmic approximation $\log_2(1+\gamma) \approx \log_2(\gamma)$ introduced in that paper does not easily allow for generalizations (such as optimizing the amount of bandwidth allocated to each user), and becomes loose for small values of γ. The work we present here exactly addresses and solves these issues: by proposing a novel approximation based on piecewise power functions, we show that we can still obtain a GP formulation, and, by increasing the number of approximating intervals, tightly approximate the solution. Moreover, being based on convex programming, our solution does not require a feasible initial resource allocation nor training data availability.

Finally, we point out that most of the solutions proposed in the literature aim at maximizing user throughput subject to a fixed bound on the total power consumption. As discussed in the introduction, we choose a different approach here, aiming at minimizing

the power consumption under constraints of guaranteed single-user throughput. Indeed, while our approach may be easily adapted to the problem usually considered in the literature, since we formulate both constraints and objective function as convex functions, we feel that the problem formulation discussed here better complies with the increasing request of designing green and sustainable telecommunication networks.

1.2 Goals

The main objective of this work is to contribute to the endeavour of making dense, heterogeneous RAN deployments greener, by proposing and efficiently solving a novel joint optimization problem that allows minimizing transmission power in orthogonal networks composed of different types of cells, while meeting individual users' throughput requirements. Unlike many previous works in the field, our optimization approach does not rely on iterative techniques that, in general, can only guarantee local minima through a computationally demanding search, whilst their convergence highly depends on the initial condition selection. In addition, our problem formulation allows considering multiple variables simultaneously: user association, resource allocation, and transmission power, instead of dividing the optimization problem into smaller ones in a *ceteris paribus* manner.

We achieve the targets above by deriving original piecewise concave approximations of the Shannon-Hartley formula, and consequently applying a variable transformation that enables posing the optimization problem into a mixed-integer geometric program (MIGP) form. A differentiating aspect of our study is also that, unlike previous works that rely on simplistic evaluation scenarios, we demonstrate the effectiveness of the proposed optimization framework in dependable settings where up to 800 users generate demands modeled after mobile network traffic measurements and served by real-world heterogeneous RAN deployments, in the presence of wireless channel characteristics reproduced via a high-performance signal propagation solver.

1.3 The BANYAN project

The work presented in this thesis is part of the Big dAta aNalYtics for radio Access Networks (BANYAN) project. BANYAN was a cooperation between the following academic and industrial institutions:

- Ranplan Wireless Network Design Ltd (United Kingdom).
- Orange TM (France).
- Consiglio Nazionale delle Ricerche- CNR (Italy).

- Politecnico di Torino (Italy).
- IMDEA Institutes (Spain).
- University of Cambridge (United Kingdom).

The main goals of the project were related to open issues in the context of 5G heterogeneous Radio Access Networks planning, listed below.

1. Modelling and forecasting 5G network traffic for different types of services (streaming, download, data transmission, videos, applications, etc) in time and space dimensions.
2. Geo-location and characterizing the patterns of mobile traffic of RANs concerning in-building locations.
3. Development of data-driven algorithms for the allocation of 5G networks resources.
4. Coordination of both indoor and outdoor heterogeneous RANs to respect users requirements in terms of quality of service.
5. Orchestrate the 5G RANs to respect service requirements and dynamics via network slicing (configuration allowing multiple independent networks to use the same physical infrastructure).

The BANYAN project was supported in part by the European Commission through the Horizon 2020 Framework Programme under Grant H2020-MSCA-ITN-2019, and in part by Marie Sklodowska-Curie Actions Innovative Training Networks (MSCAITN)-EID through Big dAta aNalYtics for radio Access Networks under Proposal 860239.

1.4 Published papers

1. Forecasting Network Traffic: A Survey and Tutorial With Open-Source Comparative Evaluation, in IEEE Access [20].
2. A Joint Optimization Approach for Power-Efficient Heterogeneous OFDMA Radio Access Networks, in IEEE Journal on Selected Areas in Communications [21].
3. Power minimization and resource allocation in HetNets with uncertain channel gains, in IEEE Communications Letters [22].

The first paper of the list above is concerned with goal 1 of the BANYAN project, while the second and third ones are related to goals 3 and 4, respectively.

During the Ph.D., the following presentations were performed:

- Poster Day at Politecnico di Torino, Italy 2022.
- Poster Day at Politecnico di Torino, Italy 2023.
- IMDEA summer school on data-driven 5G RAN, Spain 2023.
- AUTOMATICA.it conference, Italy 2023.
- AUTOMATICA.it conference, Italy 2024.
- BANYAN workshop, Sweden 2024.
- Poster Day at Politecnico di Torino, Italy 2024.

1.5 Text organization

For a better comprehension of the many parts of this work, it is divided in different Chapters, described below.

Chapter 2

In this Chapter, the theoretical background required to comprehend the problem formulation and its proposed solution is presented. Specifically, details on *Wireless Communication* such as cellular networks and the corresponding multiple access techniques are explained. Subsequently, we present discussions on *Convex Optimization Problems* focusing on the Geometric Programming formulation: definition, advantages, complexity, and solving methods. Details on Model Predictive Controller main concepts are also briefly presented.

Chapter 3

The basis problem to be solved is presented in this Chapter. The constraints that must be respected and the objective function to be minimized are discussed. The optimization problem at hand, which is highly non-convex, works as a starting point for the subsequently Chapters.

Chapter 4

In Chapter 4, we present the core idea of this work (also used in Chapters 5 and 6): the non-convex constraints are approximated with a piecewise power function followed by a change of variables allowing us to rewrite the constraints as *log-sum-exp*, which are known to be convex functions. We discuss the complexity of the new optimization problem (Mixed-Integer Geometric Program) and prove its efficiency in a highly realistic scenario

obtained with real-world data from a large European city neighbourhood. Comparisons with non-convex optimization techniques are performed.

Chapter 5

Some works in the Literature realize experiments showing that the channel-gains between users equipment and base stations are random quantities respecting a log-normal distribution. Therefore, we extend the problem treated in the previous Chapter to the case where the channel-gains are not deterministic. In summary, the Mixed-Integer Geometric Programming becomes a chance constrained optimization problem and the latter is formulated as a Robust Mixed-Integer Geometric Programming.

Chapter 6

In this Chapter, Model Predictive Control is applied to manage the base stations transmission powers and respect the users equipment throughput demands (within a certain margin), when there is a network failure, i.e. one of the base stations is abruptly switched-off. To do so, we rely on a trajectory prediction strategy and the problem is extended to a time-varying scenario.

Chapter 7

Conclusion and the perspectives of future works are presented in this Chapter.

Chapter 2

Theoretical Background

2.1 Wireless Communication

Many different wireless communications systems are simultaneously operational to allow an effective data transmission between devices, such as: Wireless LANs (Local Area Networks), Broadband Wireless Access, Satellite Networks, and Cellular Networks. Since the latter one has experienced exponential growth due to ever increasing number of connected devices (as smartphones) with their high data transmission demands, the upcoming optimization approaches are then focused on scenarios based on cellular networks.

2.1.1 Cellular Networks

Cellular networks rely on frequency reuse of radio waves propagated by different base stations to provide connection to wireless devices. The idea is to minimize the interference between base stations by placing those operating at the same frequency, equal colors in Figure 2.1, separated from each other. In this manner, the network is composed of many non-overlapping cells.

The Mobile Networks Operators (MNOs) then have a trade-off problem: if the *reuse distance* (distance between the base stations in cells using the same frequency) is too small, the quality of signal is degraded by the intercell interference; conversely, if it is very large, the frequencies sets (scarce resources) are not reused in an optimized manner.

Nowadays, searching for lower cost networks, higher transmission capacity, and low latency applications, the MNOs place small/micro base stations with less transmission power inside the macrocell, leading to so called Heterogeneous Networks. Since the macro and small base stations inside a heterogeneous network operate at the same frequency range, power and resource optimization are of fundamental importance to guarantee that the interference is kept below a threshold limit. As presented in Figure 2.2, once an equipment is connected to a base station, the remaining ones send interference signal.

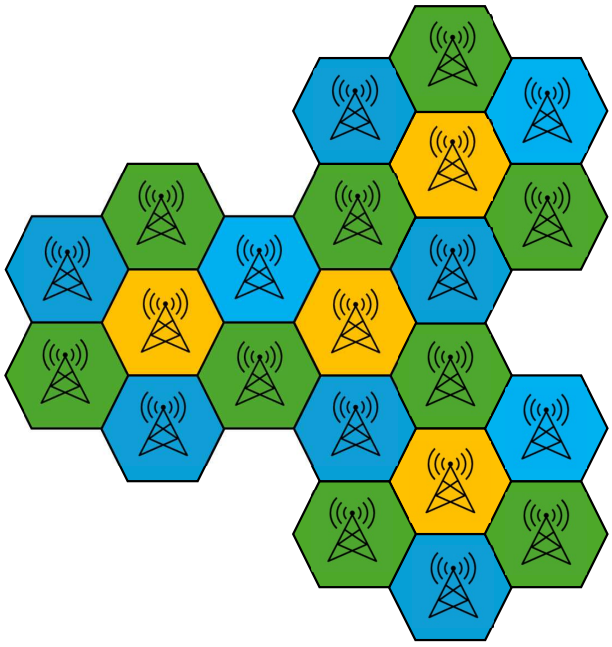


Figure 2.1: Cellular networks with non-overlapping neighbouring cells.

Additionally, in general these small cells are located at the street level and their coverage area of signal propagation are irregular due to shadowing effects caused by buildings, cars, pedestrians, etc.

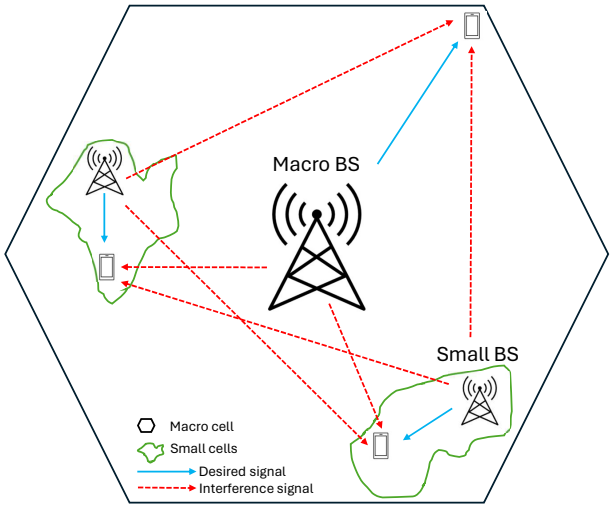


Figure 2.2: Heterogeneous network.

2.1.2 Multiple Access Technologies

As a consequence of the many devices that require connection with a single base station simultaneously, the available signal is then divided along time, frequency, and code dimensions, leading to the multiple access technologies discussed in the sequence. Since this study develops optimization problems for cellular networks based on Orthogonal

Frequency Division Multiple Access, more information about this technology is given. For details beyond the ones in the sequence or other multiple access techniques, the reader may refer to [23, 24].

FDMA and TDMA

The first generation (1G) of mobile networks, mostly developed during the 1980s, was based on analog communication techniques and used Frequency Division Multiple Access (FDMA) as channel access method. In FDMA, the total system bandwidth is divided into multiple frequency channels and each one is assigned to a unique user. To avoid interference, part of the spectrum is not used so adjacent frequency channels are separated (guard band), as presented in Figure 2.3 (a). 1G was able to provided only voice services to the users.

The 2G services, introduced in the 1990s, were based on digital technology. Different from 1G, 2G made available text and packet data services, and used a combination of FDMA and Time Division Multiple Access (TDMA). TDMA allows multiple access by allocating the entire frequency spectrum for a single device at an specif time slot, as in Figure 2.3 (b).

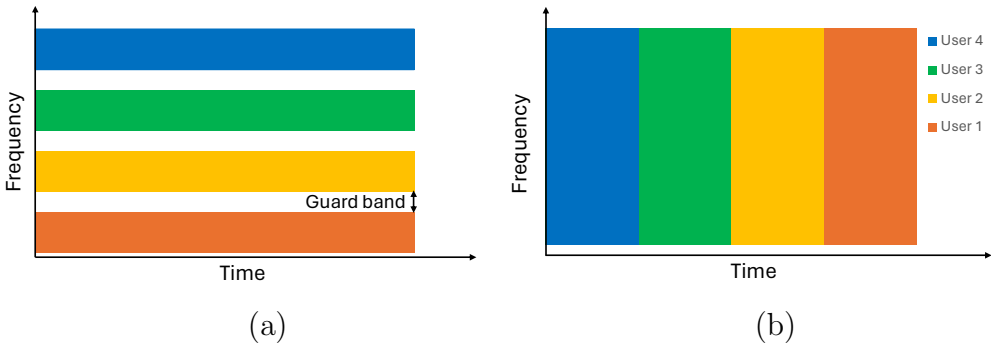


Figure 2.3: (a) FDMA and (b) TDMA multiple access technologies.

One of the most used technology standards for 2G is the Global System for Mobile communications (GSM). GSM allows multiple access by sharing the frequency and time spectrums simultaneously. In summary, the bandwidth is divided as in the FDMA, however each frequency channel is used by multiple users, each at a unique time slot as in TDMA. This is represented in Figure 2.4.

CDMA

CDMA stands for Code Division Multiple Access and it is a technology used in latter 2G but was also basis for the upcoming third generation of mobile devices. CDMA is able to provide connection for more users with less transmission power connection, when compared to the aforementioned techniques. The idea is that the information that needs

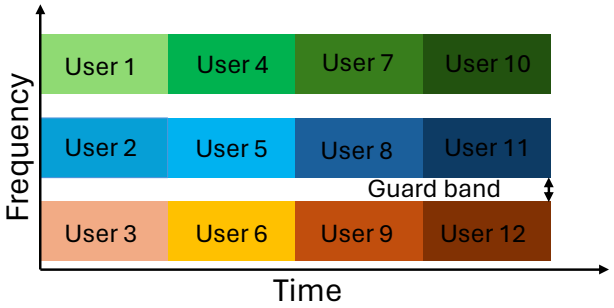


Figure 2.4: Combination of FDMA and TDMA to allow more users connection simultaneously.

to be sent to different users share simultaneously the same frequency and time dimensions, as in Figure 2.5. The information is modulated by spreading codes and each receiver is then able to use this code to retrieve the respective data.

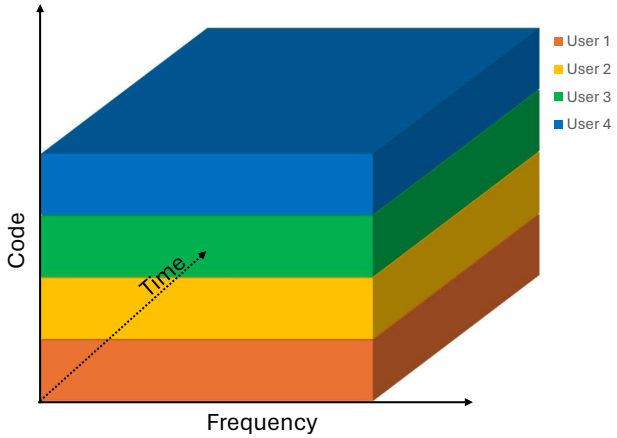


Figure 2.5: CDMA users sharing frequency and time domains simultaneously.

OFDM

A clear disadvantage of FDMA is the requirement of guard-bands to avoid frequency channels interference, preventing part of the spectrum to be used for communication. Instead, the Orthogonal Frequency Division Multiplexing (OFDM) divides the entire bandwidth into sub-carriers that overlap (no guard band required), however are orthogonal, therefore not interfering with each other. As shown in Figure 2.6, at the highest value of each sub-carrier, all others are zero.

OFDM is used by both 4G Long-Term Evolution (LTE) and 5G New Radio (NR) in the wireless communication between users devices and base stations (air interface). Nevertheless, OFDM still has the limitation of, given a time slot, all sub-carriers are assigned to a unique user.

OFDMA

Orthogonal Frequency Division Multiple Access (OFDMA) is the natural evolution of OFDM, since it also devides the total system bandwidth into orthogonal narrow sub-

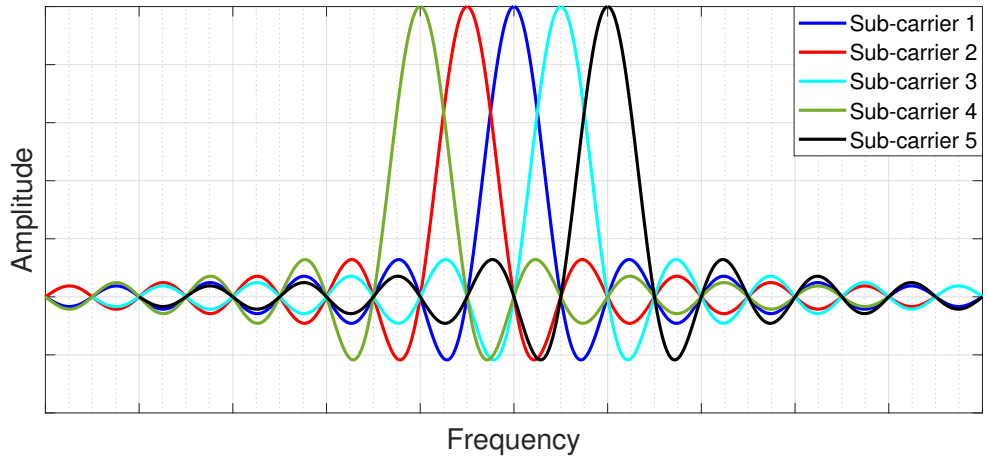
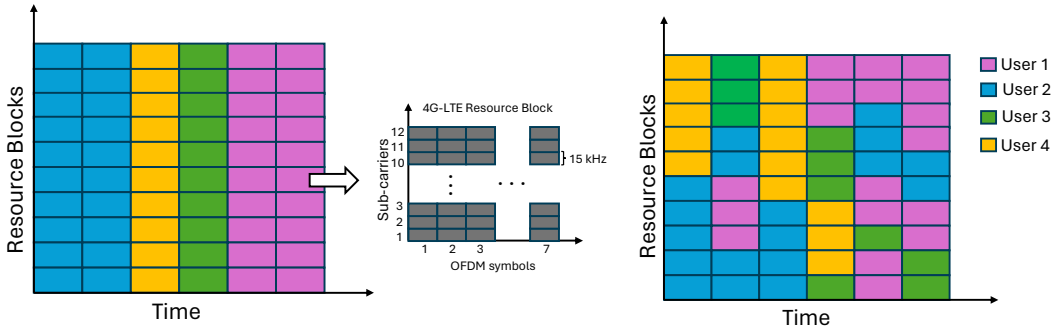


Figure 2.6: OFDM orthogonal sub-carriers.

bands and allows different sets of sub-carriers to be allocated for different users at a single time slot, as exposed in Figure 2.7. It is also used in 4G-LTE and 5G-NR.



(a) (b)

Figure 2.7: (a) OFDM and (b) OFDMA RB allocation.

A Resource Block (RB), the smallest unit that can be allocated to a user, is a set of sub-carriers. In 4G-LTE an RB is a set containing 12 sub-carriers with 15 kHz each and 7 OFDM symbols. Therefore, one can clearly see that the bandwidth allocation in OFDMA cellular networks has an integer nature, since RBs are exclusively assigned to a user at a given time slot.

The maximum data rate transferred (capacity) with wireless communications are intrinsically related to the available bandwidth. According to the Shannon-Hartley Theorem [25], the capacity C (bits/s) of a wireless channel is a function of the channel bandwidth B (Hz) and the received Signal-to-Noise Ratio (SNR: desired signal power (W) divided by the noise power (W)), as

$$C = B \log_2(1 + \text{SNR}). \tag{2.1}$$

Any digital communication with data rate above C has a probability of error bounded

away from zero.

In the context of a heterogeneous network, besides the noise power, one should also consider the interference caused by neighboring small BSs working at the same frequency channels. Therefore, SNR becomes SINR: Signal-to-Interference and Noise Ratio, leading to

$$C = B \log_2(1 + \text{SINR}). \tag{2.2}$$

Specifically in OFDMA applications, the total capacity of a user i connected to a BS j is given by

$$C_i = \sum_{j=1}^{N_{RB_j}} \alpha_{ij} B_{RB_j} \log_2(1 + \text{SINR}_{RB_{ij}}), \tag{2.3}$$

with

$$\alpha_{ij} = \begin{cases} 1, \text{ if } RB_j \text{ is assigned to the user,} \\ 0, \text{ otherwise,} \end{cases}$$

where B_{RB_j}, N_{RB_j}, and SINR_{RB_j} denote the bandwidth of a single RB of BS j, the total number of RBs of BS j, and the SINR of that RB, respectively. Due to the orthogonal characteristics of the sub-carriers, an RB of a BS receives interference just of an RB operating at the same frequency, same color in Figure 2.8, from another BS, which impacts the value of the $\text{SINR}_{RB_{ij}}$.

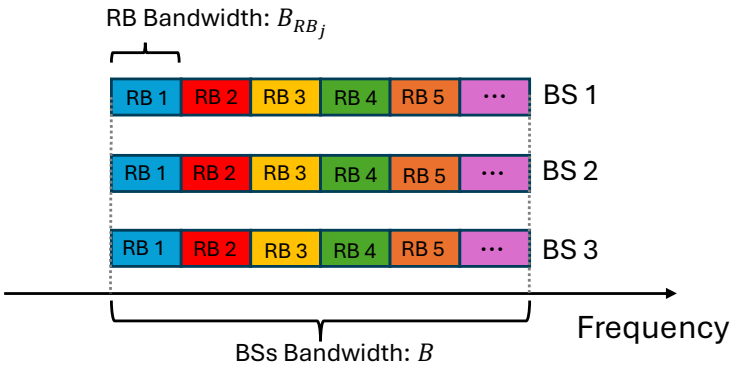


Figure 2.8: RBs from different BSs operating at the same frequency channel (same color), leading to interference.

As it will be further presented, the SINR is a function of the BSs transmission powers and a gain/loss, namely channel gain, between each pair transmitter and receiver.

2.2 Convex Optimization Problems

The optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f_0(x) \\ \text{s.t.: } & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{2.4}$$

is convex if

- the objective (or cost) function $f_0(x)$ is convex;
- the functions $f_i(x)$ are convex $\forall i \in m$;
- the functions $h_i(x)$ are affine $\forall i \in p$.

This class of optimization problems can be equivalently defined as a problem where one wants to minimize a convex function, with the restriction that the decision variables belong to a convex set. In the context of convex optimization problems, any local optimal solution is also globally optimal and can be always found (when feasible).

As discussed in [26], despite the fact of the required special format of the functions $f_i(x)$, many practical problems can be formulated as convex optimization problems, or at least well approximated by one. Electronic circuit design, finance, control systems, signal processing, and modeling are some examples of fields of applications for convex optimization problems. Furthermore, writing a problem in such a format allows it to be solved globally and efficiently, for instance using the interior points method.

"If we can formulate a problem as a convex optimization problem, then we can solve it efficiently, just as we can solve a least-squares problem efficiently. With only a bit of exaggeration, we can say that, if you formulate a practical problem as a convex optimization problem, then you have solved the original problem", (S. Boyd and L. Vanderberghe, 2004, p.8).

Many widely known optimization problems are convex:

- Linear Programs (LP);
- Quadratic Programs (QP);
- Quadratic Constrained Quadratic Programs (QCQP);
- Second Order Cone Models (SOCM);
- Semidefinite Programs (SDP);
- Geometric Programs (GP).

In Chapters 3 and 4, respectively, we present a non-convex optimization problem related to OFDMA cellular networks and formulate it as a convex one, more specifically as a Geometric Program. Therefore, details on GPs will be addressed in the sequence.

2.2.1 Optimality conditions for convex optimization

Proposition 1 *If* (2.4) *is an unconstrained optimization problem, then x is optimal if and only if*

$$\nabla f_0(x_{op}) = 0. \tag{2.5}$$

Proposition 2 *Let* $\mathcal{X}$ *be the convex set satisfying the equality and inequality constraints in* (2.4) *and* x_{op} *be a candidate optimal point. Then,* $x_{op} \in \mathcal{X}$ *is optimal if and only if*

$$\nabla f_0(x_{op})^\top (x - x_{op}) \geq 0, \quad \forall x \in \mathcal{X}. \tag{2.6}$$

Proof: Both propositions can be explained with the first order Taylor-Series expansion considering points in the neighbourhood of x_{op}, i.e. $x - x_{op} = \epsilon \mathrm{v}$ for small positive values of ϵ and v being a directional unit vector, given by

$$\begin{aligned} f_0(x) \simeq & f_0(x_{op}) + \nabla f_0(x_{op})^\top (x - x_{op}), \\ = & f_0(x_{op}) + \epsilon \nabla f_0(x_{op})^\top \mathrm{v}. \end{aligned} \tag{2.7}$$

Since

$$\lim_{\epsilon \longrightarrow 0} \frac{f_0(x_{op} + \epsilon \mathrm{v}) - f_0(x_{op})}{\epsilon} = \nabla f_0(x_{op})^\top \mathrm{v}, \tag{2.8}$$

the inner product

$$\nabla f_0(x_{op})^\top \mathrm{v} = ||\nabla f_0(x_{op})||_2 \, ||\mathrm{v}||_2 \cos\theta \tag{2.9}$$

measures the rate of variation of function $f_0(x)$ around x_{op} along direction v, where θ is the angle between the given vectors. Now let $\mathcal{H}_{++}$ and $\mathcal{H}_{-}$ be the half-spaces defined as

$$\begin{aligned} \mathcal{H}_{++}(x_{op}) = & \nabla f_0(x_{op})^\top (x - x_{op}) > 0, \\ \mathcal{H}_{-}(x_{op}) = & \nabla f_0(x_{op})^\top (x - x_{op}) \leq 0. \end{aligned} \tag{2.10}$$

From (2.7), (2.9), and 2.10, one can clearly notice the following:

- Any point at $\mathcal{H}_{++}(x_{op})$ cannot be optimal, since the function increases on directions forming a positive inner product with the gradient. In these cases, $f_0(x) > f_0(x_{op})$ $\forall\ x \in \mathcal{H}_{++}(x_{op})$.

- The function decreases *locally* (small steps) in $\mathcal{H}_{-}(x_{op})$, since points in that direction form a negative inner product with the gradient, leading to $f_0(x) < f_0(x_{op})$.

- If an optimization problem is unconstrained (as in Proposition 1) and the gradient at a given point is different from 0, this point cannot be optimal since another point forming a negative inner product with the gradient, leading to $f_0(x) < f_0(x_{op})$, can always be found. Therefore, a point is optimal if and only if $\nabla f_0(x_{op}) = 0$.

- For constrained optimization problems, Proposition 2 is equivalent to stating that x_{op} is optimal if all points that satisfy the constraints belong to $\mathcal{H}_{++}(x_{op})$. x_{op} is then in the boundary of the feasible set and no points forming negative inner product with $\nabla f_0(x_{op})$ can satisfy the constraints. Therefore, x_{op} must be the optimal point.

These optimality conditions are used in many algorithms designed to solve convex optimization problems, such as *Gradient Descent* and *Interior-Point Methods*. The latter one is especially important in the context of GPs, as discussed below.

2.2.2 Geometric Programs

In the following Chapters, Geometric Program is used to deal with the optimization problem at hand. [27] presents a detailed and meticulous description/tutorial of this class of optimization problems and discussions beyond the summary presented below can be found there.

Definition

Geometric Programs (or Programming) are optimization problems whose decision variables are positive and appear in constraints and/or objective function as a non-negative linear combination of positive monomials (posynomials), which in general are not convex. When such a case happens, one can rewrite objective function and constraints such that the problem becomes convex. Formally, a Geometric Program has the form

$$\begin{aligned} \min_{\zeta \in \mathbb{R}^n} \quad & f_0(\zeta) \\ \text{s.t.: } & f_i(\zeta) \leq 1, \quad i = 1, \ldots, m \\ & h_i(\zeta) = 1, \quad i = 1, \ldots, p \\ & \zeta > 0, \end{aligned} \tag{2.11}$$

with $f_0(\zeta)$ and $f_i(\zeta)$ $i = 1, \ldots, m$ being posynomial functions and $h_i(\zeta)$ $i = 1, \ldots, p$ monomial functions.

From posynomial/monomial to a convex formulation

The main idea of Geometric Programs is to obtain a non-linear but convex representation of an originally non-convex problem. For instance, consider a vector of variables $\nu = [\nu_1, \nu_2, \nu_3]$ and another of constants $c = [c_1, c_2]$ such that each element of $\nu \in \mathbb{R}^*_+$ and $c \in \mathbb{R}^*_+$. Then

$$\mathrm{f}(\nu) = c_1\, \nu_1^{s_1}\, \nu_2^{s_2}\, \nu_3^{s_3} + c_2\, \nu_1^{s_4}\, \nu_2^{s_5}\, \nu_3^{s_6} \tag{2.12}$$

is not convex. s is a vector containing constants that are not required to be positive, i.e $s = [s_1, \ldots, s_6] \in \mathbb{R}$. The change of variables $\nu_i = e^{\zeta_i}$ leads to the following convex function

$$\mathrm{f}(\zeta) = c_1\, e^{\zeta_1 s_1 + \zeta_2 s_2 + \zeta_3 s_3} \;+\; c_2\, e^{\zeta_1 s_4 + \zeta_2 s_5 + \zeta_3 s_6}. \tag{2.13}$$

Additionally, since monotone transformation of constraints or objective functions lead to an equivalent optimization problem,

$$f(\zeta) = \log(\mathrm{f}(\zeta)) = \log(c_1\, e^{\zeta_1 s_1 + \zeta_2 s_2 + \zeta_3 s_3} \;+\; c_2\, e^{\zeta_1 s_4 + \zeta_2 s_5 + \zeta_3 s_6}) \tag{2.14}$$

is also a convex function, since it is a *log-sum-exp*.

For the specific cases where the objective and constraints functions appear as a monomials instead of positive linear combination of them ($c_2 = 0$ in $\log(\mathrm{f}(\zeta))$), the optimization problem becomes an LP:

$$f(\zeta) = \log(\mathrm{f}(\zeta)) = \log(c_1\, e^{\zeta_1 s_1 + \zeta_2 s_2 + \zeta_3 s_3}) = \log(c_1) + \zeta_1 s_1 \;+\; \zeta_2 s_2 \;+\; \zeta_3 s_3. \tag{2.15}$$

A small example

Consider the following optimization problem

$$\begin{aligned} \min_{\nu \,\in\, \mathbb{R}^2} \quad & \frac{\nu_1}{\nu_2} \\ \text{s.t.: } & 1 \leq \nu_1 \leq 5, \\ & 0 < \nu_2 \leq 5, \\ & \nu_1^3\, \nu_2^1 + \; \nu_1^2 \nu_2^2 \leq 10. \end{aligned}$$

With some mathematical manipulation, it can be written with posynomials, according to

$$\begin{aligned} \min_{\nu \,\in\, \mathbb{R}^2} \quad & \nu_1 \nu_2^{-1} \\ \text{s.t.: } & \nu_1^{-1} \leq 1, \\ & 0.2\, \nu_1 \leq 1, \\ & 0.2\, \nu_2 \leq 1, \\ & \nu_2 > 0, \\ & 0.1\, \nu_1^3\, \nu_2^1 + 0.1\, \nu_1^2 \nu_2^2 \leq 1. \end{aligned}$$

This problem is clearly non-convex, since the objective function and its feasible set are both non-convex as presented in Figure 2.9.

Conversely, with the aforementioned change of variables and logarithmic transforma-

tion, the optimization problem becomes

$$\begin{aligned} \min_{\zeta \in \mathbb{R}^2} \quad & \zeta_1 - \zeta_2 \\ \text{s.t.:} \quad & -\zeta_1 \leq 0, \\ & \zeta_1 + \log 0.2 \leq 0, \\ & \zeta_2 + \log 0.2 \leq 0, \\ & \log\left(0.1\, e^{3\zeta_1 + \zeta_2} + 0.1\, e^{2\zeta_1 + 2\zeta_2}\right) \leq 0. \end{aligned}$$

Since objective and the left side of all inequalities are composed of convex functions, the reformulated optimization problem is convex. This is also illustrated in Figure 2.10. Note that the feasible set is convex and the objective function became a hyperplane (which is also convex).

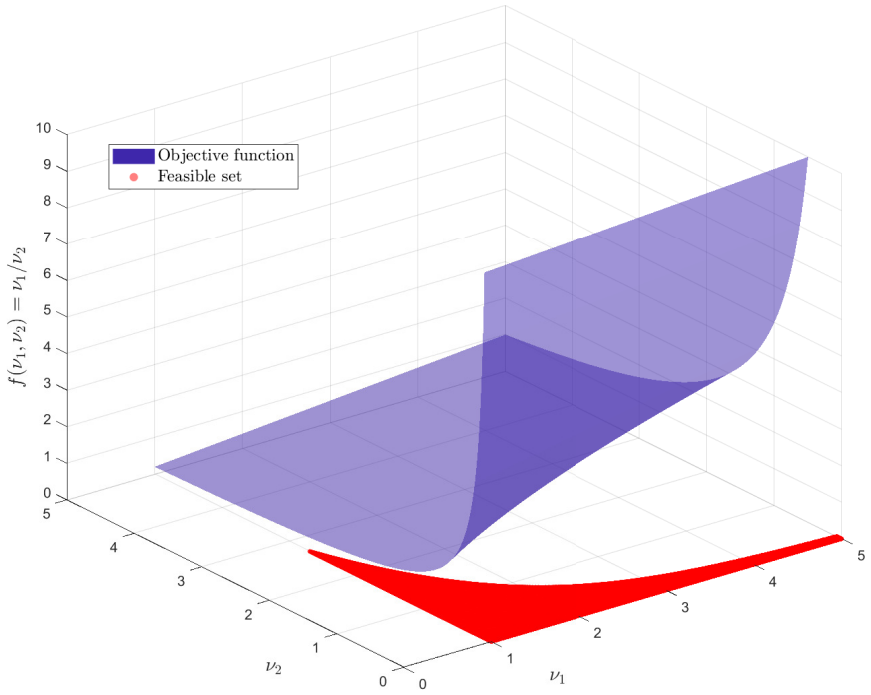


Figure 2.9: Non-convex optimization problem with posynomial formulation.

The optimal solution is $(\zeta_1 = 1.06 \times 10^{-9}, \zeta_2 = 0.9938)$ leading to an optimal value of $\zeta_1 - \zeta_2 = -0.9938$. This result was obtained with CVX [28, 29] (Matlab based modeling system for convex optimization). Then the optimal (and global) solution to the original problem can be then recovered:

- $\nu_1 = e^{\zeta_1} = e^{1.0638 \times 10^{-9}} = 1$;
- $\nu_2 = e^{\zeta_2} = e^{0.9938} = 2.7016$;

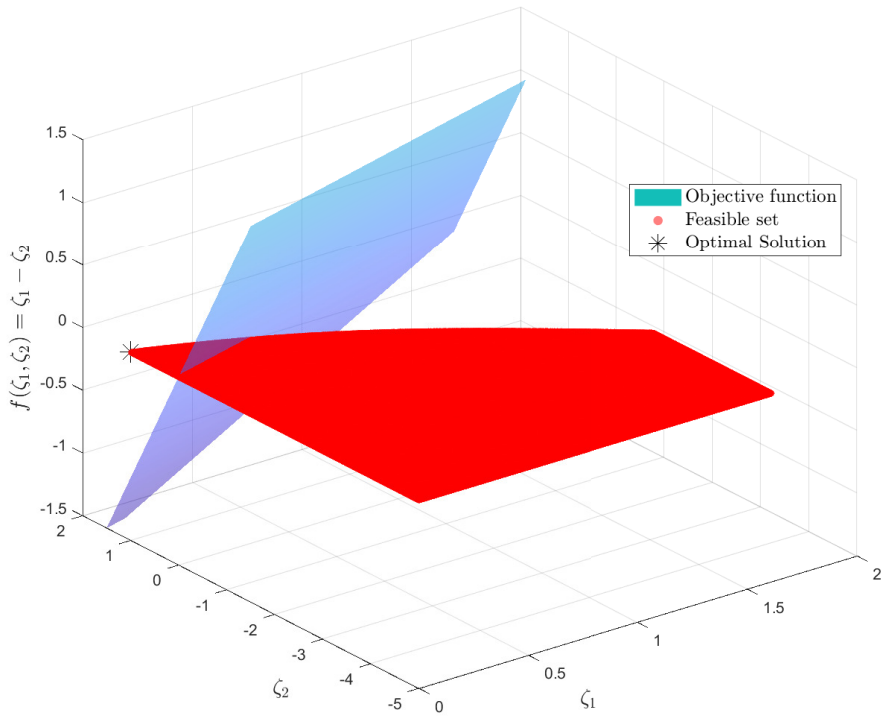


Figure 2.10: Resulting convex optimization problem.

- $\frac{\nu_1}{\nu_2} = 0.3702$.

An important point to notice is that in the non-convex formulation ν_2 is bounded, i.e $0 < \nu_2 \geq 5$, however ζ_2 is unbounded below since any value respects the constraint $e^{\zeta_2} > 0$.

Advantages of GP

The choice of working with GPs is due to their many advantages over non-convex optimization problems, such as:

1. No requirement of parameters tuning;
2. No starting points (initial guess) are required;
3. The global optimal solution can be always found (for feasible problems);
4. Provide fast solutions when solved with the interior-points method.

On the complexity of the GP algorithm

Geometric Programming, once transformed into its log-sum-exp representation, can be globally solved in polynomial time using standard barrier-based interior-point methods for

convex programming. For n_v variables and n_c inequality constraints, ϵ convergence can be reached in a number of outer (centering) iterations essentially proportional to $\log(n_c/\epsilon)$, whereby each outer iteration requires a number of inner (Newton) iterations that depend on various parameters (Lipschitz constants of gradients and Hessian) but scales very mildly as a double log in the precision, i.e. as $\log\log(c/\epsilon)$, with c a given constant. For the Mixed-Integer GP, a locally optimal solution can be found with branch & bound methods, which in a worst-case scenario, increases the computational effort exponentially with the problem size. On the other hand, it should be remarked that the literature presents several (possibly heuristic) methods with low complexity for selecting these integer parameters. Such methods can be surely adapted to the problem at hand. For further details on GP and convex optimization problems, the reader is referred to [26, 27].

Note that each SINR_{RB} in (2.3) is a function of desired, interference and noise powers, which are all positive quantities as required to pose the problem as a GP. In the next Chapters we show the problem to be solved and its conversion to the convex formulation.

2.3 Model Predictive Controller

Model Predictive Controller (MPC) is an optimization problem based control strategy that minimizes an objective function related to a dynamical system, possibly subject to constraints. At each time-step of the optimization, the controller receives the measured (or estimated) plant states and minimizes a cost function over pre-defined prediction and control horizons. The first control action is then applied to the plant and the remaining ones are discard, since the plant predicted output, which is based on its mathematical model, might be different from the real one due to modeling errors (caused by linearization, disturbances, etc). This process is then repeated for the next iterations. The MPC key features are in Figure 2.11 and defined below.

- Controlled variables: plant states/outputs that are required to follow specif references and performance.
- Manipulated variables: variables in the plant that can be manipulated (control inputs), such that the controlled variables reach the desired reference and performance.
- Prediction horizon: number of future samples that the optimizer evaluates to calculate the control actions and minimize the objective function. In general, this value should be large enough to cover the plant transient response. However if too large, it can lead to very high computational cost.
- Control horizon: inside a prediction horizon, it can be defined as the number of different control actions sent to the plant before it becomes constant. When too

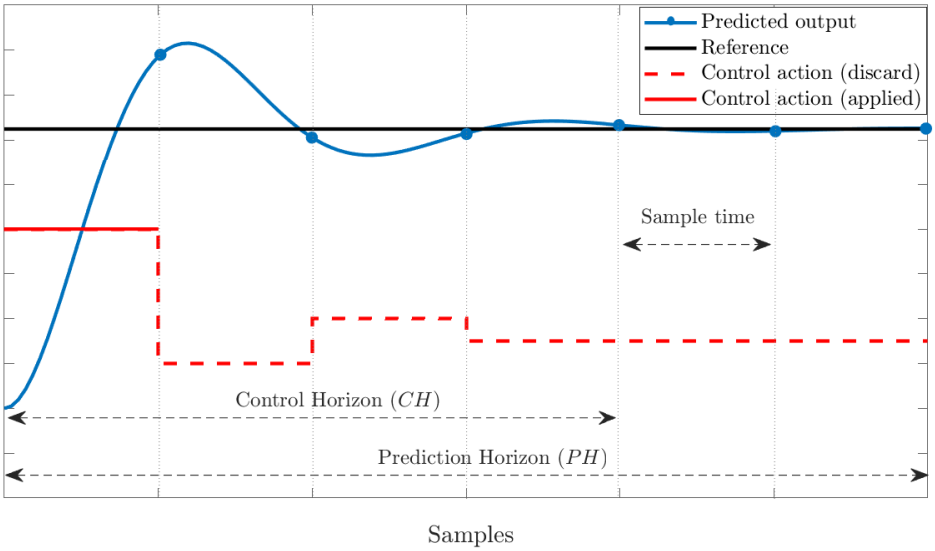


Figure 2.11: One MPC iteration.

large, it also increases the computation cost, however if it is too small, the objective function might not be properly minimized.

- Objective (cost) function: should be minimized by the MPC. For instance, in trajectory following problems, it can be formulated as the difference between predicted output and reference signal. The objective function can also depend on other control parameters such as the manipulated variables.

- Constraints: the plant might have physical limitations, such as bounds to the manipulated variables and their time derivatives, that are then posed as optimization problem constraints. The plant dynamical model is also a problem constraint.

Let $u_k \in \mathbb{R}^{n_u}$ and $x_k \in \mathbb{R}^{n_x}$ be the vectors of manipulated and state variables of a discrete-time system whose state equation is given by

$$x_{k+1} = f(x_k, u_k, k). \tag{2.16}$$

$\mathcal{X}$ and $\mathcal{U}$ are the sets representing the values that state and manipulated variables can assume, respectively. Then, one MPC iteration is defined by an optimization problem of

the form

$$\begin{aligned} \min_{u_0,\ldots,u_{PH-1}} \quad & \sum_{k=0}^{PH-1} \ell(x_k,u_k) \\ \text{s.t.: } & x_{k+1} = f(x_k,u_k,k), \quad k = 0,\ldots,PH-1 \\ & x_k \in \mathcal{X}, \\ & u_k \in \mathcal{U}, \\ & u_{k+1} = u_k, \quad k = CH,\ldots,PH-1, \end{aligned} \tag{2.17}$$

where the cost function $\ell(x_k,u_k)$ can be formulated to allow reference tracking

$$\ell(x_k,u_k) = ||x_k - x_{ref}||_2^2 + ||u_k - u_{ref}||_2^2,$$

with x_{ref} being the states reference and u_{ref} the reference control input or equilibrium point x_{eq} convergence

$$\ell(x_k,u_k) = ||x_k - x_{eq}||_2^2 + ||u_k||_2^2.$$

Note that the formulation above might not be a convex optimization problem depending on the sets $\mathcal{X}$ and $\mathcal{U}$ and function $\ell(x_k,u_k)$. MPC is a topic of many researches nowadays due to their many different formulations/applications and details beyond the ones given here can be seen at [30, 31].

Chapter 3

Problem Formulation

In this Chapter we formulate the problem of resource allocation, users association to BSs, and BSs transmission power minimization subject to individual users throughput constraints. The optimization scenario is a heterogeneous network consisting of N BSs, n users, and OFDMA multiple access technology. The optimization problem obtained in the sequence is highly non-convex, and it serves as basis for the well-posed formulations presented in the next Chapters.

3.1 Optimization Problem

3.1.1 Bandwidth assignment relaxation

Equation (2.2) clearly demonstrates that the download capacity (bits/s) of a user has an integer characteristic, since it depends on the number of allocated RBs. Then, the following problem relaxation is used:

$$\sum_{j=1}^{N_{RB_j}} \alpha_{ij} B_{RB_j} \simeq x_{ij} B_j \tag{3.1}$$

where $\alpha_{ij} \in \{0,1\}$ is a binary variable that indicates if a resource block with bandwidth B_{RB_j} in base station j is allocated to user i. Therefore, the summation in (3.1) is used to calculate the total bandwidth assigned to a user i connected to BS j. $x_{ij} \in [0,1]$ is a continuous variable allocating a percentage of the total bandwidth B_j (Hz) of BS j to user i. We denote by $x \in [0,1]^{n \times N}$ the matrix of the elements x_{ij}, with $x_{ij} \in [0,1]$.

We point out that our formulation considers the case in which the total bandwidth B_j of an antenna is shared equally between a finite number of resource blocks of identical transmission power for the same BS. This is captured by the continuous variables $x_{i,j}$, which shall be interpreted as the percentage of resource blocks of BS j assigned to user i. Of course, in real-world applications x_{ij} are discrete variables. However, if the number of resource block is sufficiently large, this approximation rapidly becomes negligible. Details

on how the sub-carriers are divided in the considered OFDMA setup, and on how to obtain the discrete variables values given the continuous x_{ij}, are later addressed.

3.1.2 The SINR

As previously mentioned, the SINR is the ratio between the received power of the desired signal and the sum of interference and noise powers. Let $\mathcal{P} \doteq [P_1, \ldots, P_N]$, then for a user i connected to a RB in BS j

$$\mathrm{SINR}_{ij}(\mathcal{P}) = \frac{P_j g_{ij}}{\eta^2 + \sum_{k \neq j} P_k g_{ik}}. \tag{3.2}$$

In (3.2), η^2 denotes the noise power, P_j (W) is the transmission power of one RB from the BS to which the user is connected and P_k, $k \neq j$, are the transmission powers of one RB from the remaining BSs. The quantities g_{ij} and g_{ik} represent the channel gains, which are assumed to be known, and quantify the transmission power losses due to

- shadowing: caused by obstacles that are in the line-of-sight between transmitter (BS) and receiver (user equipment), attenuating the signal power due to reflection, scattering, and diffraction of the electromagnetic waves propagated by the transmitter;
- path loss: dissipation of electromagnetic waves propagated by the transmitter, assuming no obstacles between transmitter and receiver.

Theoretically, such values can be calculated by solving Maxwell's equations, however in practice that would require many parameters that might not be available and would demand high computational complexity. In this work, we use a ray tracing simulator, which allows us to carefully estimate the channel gains. This is discussed in Chapter 4.

3.1.3 The throughput capacity

The downlink transmission rate of a user i connected to a BS j can be calculated with the *Shannon-Hartley Theorem* [25]:

$$C_i = x_{ij} B_j \log_2(1 + \mathrm{SINR}_{ij}(\mathcal{P})), \tag{3.3}$$

with C_i in bits/s. Note that we consider all RBs from the same BS transmitting at the same power level. This allows us to use the previous problem relaxation because the SINR of each resource block contributes equally to the throughput capacity of a user, since the channel gains depend on the BS rather than on the RBs.

3.1.4 Constraints

Consider a scenario where each of n users shall be connected to one base station (BS) among N possible. One of the requirements of the problem is that a user must be connected to just one BS. To specify this constraint, we introduce binary variables $z_{ij} \in \{0,1\}$, $i \in [n]$, $j \in [N]$ as

$$z_{ij} = \begin{cases} 1, \text{ if user } i \text{ is connected to BS } j, \\ 0, \text{ otherwise.} \end{cases}$$

$z \in \{0,1\}^{n \times N}$ is the matrix with elements $z_{i,j}$. Then, the constraints of the considered optimization problem may be written as follows.

Each user must be connected to just one BS

$$\sum_{j=1}^{N} z_{ij} = 1, \quad i \in [n]. \tag{3.4}$$

A BS j cannot provide more resources than it has available, that is:

$$\sum_{i=1}^{n} x_{ij} \leq 1, \quad j \in [N]. \tag{3.5}$$

Each user i has a minimum required throughput t_i (bits/s):

$$C_i \geq t_i. \tag{3.6}$$

where $t \in \mathbb{R}^n_{\geq 0}$ is the vector with the elements t_i. Combining (3.3) with equation in (3.4), the constraint (3.6) can be rewritten in the following equivalent form:

$$\sum_{j=1}^{N} \left[x_{ij} B_j \log_2(1 + \mathrm{SINR}_{ij}(\mathcal{P})) \right] z_{ij} \geq t_i, \quad i \in [n]. \tag{3.7}$$

3.1.5 A non-convex optimization problem

The optimization problem that minimizes the BS transmission powers and meets the constraints above is written as

$$\min_{x_{ij},z_{ij},P_j} \quad \sum_{j=1}^{N} P_j \tag{3.8a}$$

s.t.:

$$x_{ij} \in [0,1], \quad i \in [n], \quad j \in [N], \tag{3.8b}$$

$$z_{ij} \in \{0,1\}, \quad i \in [n], \quad j \in [N], \tag{3.8c}$$

$$\sum_{j=1}^{N} z_{ij} = 1, \quad i \in [n], \tag{3.8d}$$

$$\sum_{i=1}^{n} x_{ij} \leq 1, \quad j \in [N], \tag{3.8e}$$

$$\sum_{j=1}^{N} \left[x_{ij} B_j \log_2(1 + \mathrm{SINR}_{ij}(\mathcal{P}))\right] z_{ij} \geq t_i, \quad i \in [n]. \tag{3.8f}$$

Note that the above optimization problem is usually very hard to solve, since it entails a combination of binary and continuous variables. More importantly, constraint (3.8f) is highly non convex, and also contains a product of continuous and binary variables, to deal with the requirement of assigning a user to just one BS. To tackle this last issue, we make use of a standard optimization technique known as the *Big-M Method* [32, 33].

Lemma 1 (Big-M trick) *Let* $P_j \doteq e^{q_j}$ *and define*

$$f(q, x_{ij}) \doteq x_{ij} B_j \log_2(1 + \mathrm{SINR}_{ij}(q)), \tag{3.9}$$

where q *is a vector with the elements* q_j *and with a slight abuse of notation we let* $\mathrm{SINR}(q) \doteq \mathrm{SINR}_{ij}(\mathcal{P}(q))$. *Then, the set of constraints in* (3.4) *and* (3.7) *can be rewritten as follows*

$$f(q, x_{ij}) \geq t_i - M\overline{z}_{ij}, \quad i \in [n], j \in [N], \tag{3.10}$$

$$\sum_{j=1}^{N} \overline{z}_{ij} = N - 1, \quad i \in [n] \tag{3.11}$$

where M *is a sufficiently large constant.*

Proof: The constraint in (3.10) is inactive when $\overline{z}_{ij} = 1$, since $f(q, x_{ij}) \geq -M$ is always satisfied if M is chosen large enough. On the contrary, $\overline{z}_{ij} = 0$ activates the restriction and the user's throughput requirement is respected. Constraint (3.11) forces

each user to be connected to only one BS (user i is connected to j if and only if $\overline{z}_{ij} = 0$). From (3.11), one can write

$$\sum_{j=1}^{N} 1 - z_{ij} = N - 1, \quad \sum_{j=1}^{N} z_{ij} = 1, \quad i \in [n],$$

repeating constraint in (3.4). By multiplying (3.10) with z_{ij} and summing over j, we have

$$\sum_{j=1}^{N} f(q, x_{ij}) z_{ij} \geq t_i \sum_{j=1}^{N} z_{ij} - M \sum_{j=1}^{N} \overline{z}_{ij} z_{ij}, \quad i \in [n],$$
$$\sum_{j=1}^{N} f(q, x_{ij}) z_{ij} \geq t_i.$$

which is the constraint in (3.7). Then 3.10-3.11 $\Rightarrow$ (3.4)-(3.7). Conversely, (3.7) can be written as

$$f(q, x_{ij}) z_{ij} \geq t_i - \sum_{k \neq j}^{N-1} z_{ik}\, f(q, x_{ij}), \;\; j \in [N]. \tag{3.12}$$

Since each user is connected to just one BS, we have that $\sum_{k \neq j}^{N-1} z_{ik} = \bar{z}_{ij}$, leading to

$$f(q, x_{ij}) z_{ij} \geq t_i - \bar{z}_{ij} f(q, x_{ij}), \;\; j \in [N]. \tag{3.13}$$

If constraint in (3.13) is satisfied, the following is also respected:

$$f(q, x_{ij}) z_{ij} \geq t_i - \bar{z}_{ij} M, \;\; j \in [N], \tag{3.14}$$

for $M \gg f(q, x_{ij})$, as

$$\begin{cases} f(q, x_{ij}) \geq t_i, \text{ if } z_{ij} = 1, \\ 0 \geq t_i - M, \text{if } z_{ij} = 0. \end{cases}$$

Note that the constraint is active when $z_{ij} = 1$ and inactive otherwise, exactly like the formulation in (3.10)-(3.11). Then (3.4)-(3.7)$\Rightarrow$ 3.10-3.11. Therefore (3.4)-(3.7)$\Leftrightarrow$ 3.10-3.11, concluding the proof.

Therefore, optimization problem (3.8) is equivalent to

$$\min_{x_{ij}, z_{ij}, q_j} \quad \sum_{j=1}^{N} e^{q_j} \tag{non-convex-OFDMA}$$

s.t.:

$$x_{ij} \in [0,1], \quad i \in [n], \quad j \in [N], \tag{3.15a}$$
$$z_{ij} \in \{0,1\}, \quad i \in [n], \quad j \in [N], \tag{3.15b}$$
$$\sum_{j=1}^{N} \overline{z}_{ij} = N - 1, \quad i \in [n], \tag{3.15c}$$
$$\sum_{i=1}^{n} x_{ij} \leq 1, \quad j \in [N], \tag{3.15d}$$
$$f(q, x_{ij}) \geq t_i - M \overline{z}_{ij}, \quad i \in [n]. \tag{3.15e}$$

3.2 Final considerations

This optimization problem will serve as a starting point for the following Chapters of this work. First, we present a solution to it and subsequently the problem where the transmission powers are controlled with Model Predictive Control. Additionally, the case where the channel gains are considered to be random quantities is also studied.

Chapter 4

Mixed-Integer Geometric Programming

In this Chapter, a novel optimization approach for orthogonal heterogeneous networks is proposed to minimize transmission power while respecting individual users' throughput constraints. The problem is formulated as a Mixed-Integer Geometric Program, and optimizes at once multiple system variables such as user association, working bandwidth, and base stations transmission powers. Crucially, the proposed approach becomes a convex optimization problem when user-base station associations are provided. Evaluations in multiple realistic scenarios from the production mobile network of a major European operator and based on precise channel gains and throughput requirements from measured data validate the effectiveness of the proposed approach. Overall, our original solution paves the road for greener connectivity by reducing the energy footprint of heterogeneous mobile networks, hence fostering more sustainable communication systems.

4.1 Piecewise power function approximation

The solution to optimization problem (3.15) is based on a piecewise power function approximation of the throughput constraint followed by variables transformation, leading to a Mixed-Integer Geometric Program.

A key ingredient to our approach is the introduction of a piecewise power function (PPF) approximation. The proposed PPF approach is a method that allows to approximate a function with a set of m simpler concave and monotonically increasing functions in power form. Here the approach is employed to achieve a lower approximation of $\log_2(1+\gamma)$, where $\gamma > 0$ represents the signal-to-interference-plus-noise ratio (SINR). By dividing the range of γ into intervals, the original function is approximated using a family of m simpler power functions of the form $a\gamma^b$. This lower approximation ensures that the resulting PPF captures the essential behavior of the original one within each interval. More importantly, we show how this introduced approximation allows to reformulate the original problem into a log-sum-exp form, thus providing a computationally efficient and tractable repre-

sentation for optimization and mathematical analysis. In a sense, this approach can be seen as a generalization and extension of the original approximation proposed by Boyd in [18], who proposed to approximate $\log_2(1+\gamma) \approx \log_2(\gamma)$.

Formally, we consider to be given an interval $[0,\tilde{\gamma}]$. Then, we divide this interval into m subintervals with extremes

$$0 = \gamma_1 < \gamma_2 <, \ldots, < \gamma_m < \gamma_{m+1} = \tilde{\gamma},$$

and, for $\ell \in [m]$, find m approximations of the form

$$\varphi_\ell(\gamma) = a_\ell \gamma^{b_\ell}, \quad a_\ell > 0, b_\ell \in (0,1) \tag{4.1}$$

valid in the interval $[\gamma_\ell,\gamma_{\ell+1}]$.

The parameters a_ℓ,b_ℓ of this function are computed by imposing that the function φ_ℓ in (4.1) interpolates $\log_2(1+\gamma)$ at the extremes γ_ℓ,$\gamma_{\ell+1}$. That is, we solve for a_ℓ and b_ℓ the following system of equations

$$\begin{cases} \varphi_\ell(\gamma_\ell) = a_\ell \gamma_\ell^{b_\ell} = \log_2(1+\gamma_\ell) \\ \varphi_\ell(\gamma_{\ell+1}) = a_\ell \gamma_{\ell+1}^{b_\ell} = \log_2(1+\gamma_{\ell+1}). \end{cases} \tag{4.2}$$

System (4.2) consists of a pair of equations in two unknown variables and, as long as $\gamma_{\ell+1} > \gamma_\ell > 0$, admits the unique solution

$$\begin{aligned} b_\ell &= \frac{\log(\log_2(1+\gamma_{\ell+1})) - \log(\log_2(1+\gamma_\ell))}{\log(\gamma_{\ell+1}) - \log(\gamma_\ell)}, \\ a_\ell &= e^{\log(\log_2(1+\gamma_{\ell+1})) - b_\ell \log(\gamma_{\ell+1})}. \end{aligned} \tag{4.3}$$

The following proposition proves some key properties of the introduced power functions.

Proposition 3 *For any φ_ℓ defined as in (4.1), for $\ell \in [m]$, we have*

1. *$\varphi_\ell(0) = 0$;*
2. *φ_ℓ is monotonically increasing in $(0, +\infty)$;*
3. *φ_ℓ is a concave function in $(0, +\infty)$.*

Proof: Point 1 can be directly verified. Regarding points 2-3, the first and second derivatives of $\varphi_\ell(\gamma)$ have the following properties

$$\varphi'_\ell(\gamma) = \frac{a_\ell b_\ell}{\gamma^{1-b_\ell}} > 0, \; \varphi''_\ell(\gamma) = \frac{a_\ell b_\ell (b_\ell - 1)}{\gamma^{2-b_\ell}} < 0.$$

Therefore, the function φ_ℓ is monotonically increasing and concave in $(0, +\infty)$.

Proposition 4 *The function $\varphi_\ell(\gamma)$ defined as in (4.1), on the interval $[\gamma_\ell,\gamma_{\ell+1}]$ with $0 < \gamma_\ell < \gamma_{\ell+1} < \infty$ has the following property*

$$\begin{cases} \varphi_\ell(\gamma) \geq \log_2(1+\gamma), & \gamma \in [0,\gamma_\ell], \\ \varphi_\ell(\gamma) \leq \log_2(1+\gamma), & \gamma \in [\gamma_\ell,\gamma_{\ell+1}], \\ \varphi_\ell(\gamma) > \log_2(1+\gamma), & \gamma \in (\gamma_{\ell+1},\infty). \end{cases} \tag{4.4}$$

Proof: It can be easily verified that 0, γ_ℓ, and $\gamma_{\ell+1}$ are the only contact points between a given $\varphi_\ell(\gamma)$ and $\log_2(1+\gamma)$. Finally, notice that since $\varphi_\ell(0) = \log_2(1+0) = 0$ and

$$\lim_{\gamma\to 0} \varphi'_\ell(\gamma) = +\infty > \lim_{\gamma\to 0} \log_2(1+\gamma)', \tag{4.5}$$

the first inequality in (4.4) is true in a neighborhood of the origin.

Moreover, due to the fact that $\varphi_\ell(\gamma)$ and $\log_2(1+\gamma)$ are monotonically increasing and they intercept at γ_ℓ and $\gamma_{\ell+1}$, second and third inequalities in (4.4) also hold.

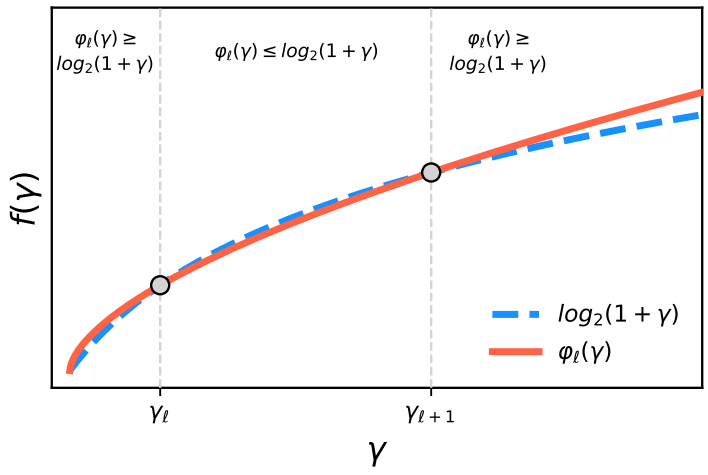
The meaning of Proposition 4 is depicted in Figure 4.1(a): the power-function approximation $\varphi_\ell(\gamma)$ (red line) is guaranteed to stay below the curve $\log_2(1+\gamma)$ (dashed blue line) inside the interval $[\gamma_\ell,\gamma_{\ell+1}]$, while it is always above outside the interval. We also note that a direct consequence of (4.5) is that

$$\varphi_\ell(\gamma) > \log_2(1+\gamma), \text{ as } \gamma \to 0. \tag{4.6}$$

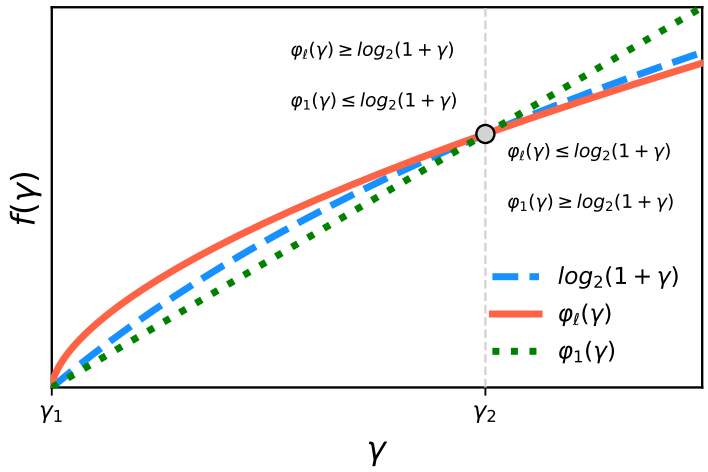
Therefore, the function $\varphi_1(\gamma)$ used to approximate the users' throughput around the origin, where $\gamma_1 = 0$ and $\gamma_2 > 0$, must be linear, i.e, $b_1 = 1$, since we cannot overestimate the throughput capacity of a user (also note that (4.3) would not have a solution for $\gamma_\ell = 0$). The behavior of the linear function $\varphi_1(\gamma)$ is show in Figure 4.1 (b).

Figure 4.2 (a) shows the impact of the hyper-parameter m, i.e. the number of subintervals of the piecewise approximation. It is clear that $m = 2$ leads to a very conservative approximation, since its respective functions strongly underestimate the original one, which could make the problem infeasible. On the contrary, already $m = 5$ generates a far better approximation compared to both $m = 2$ and $\log_2(1+\gamma) \simeq \log_2(\gamma)$. Clearly, the approximation becomes increasingly better as m increases. The impact of the choice of m is further discussed in Remark 1. Figure 4.2 (b) shows the intervals where each of the functions are active. Furthermore, one can see that if the active constraint is respected for a given γ, the inactive ones are simultaneously respected, since they are always overestimated. As a consequence, solving the optimization problem considering only the active function and its respective interval is equivalent to solving for all $\varphi_\ell(\gamma)$ with $\ell \in [m]$ simultaneously regardless the value of γ. Note also that the first function is linear.

Remark 1 (Interval length $\tilde{\gamma}$) *The proposed approach is based on PPF approximations inside a given interval $[0,\tilde{\gamma}]$. The value $\tilde{\gamma}$ represents an upper bound on the SINR, and can*



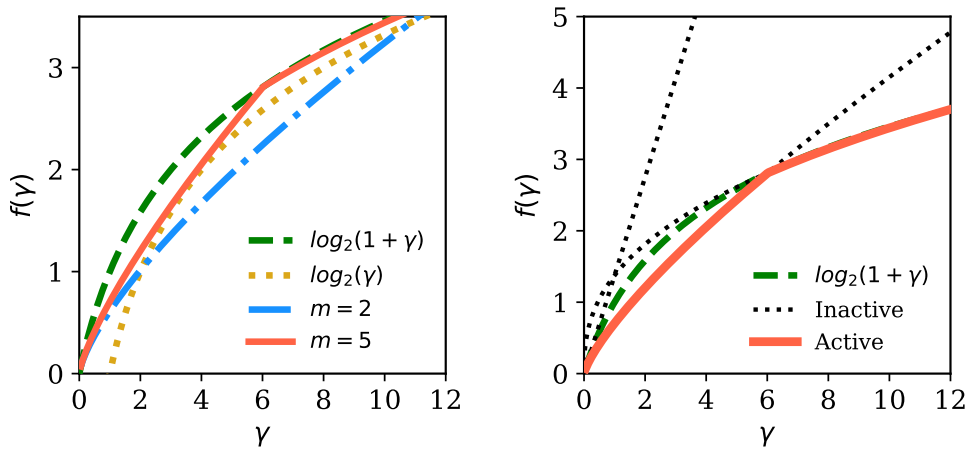
(a)



(b)

Figure 4.1: Visual representation of (a) inequalities in (4.4) and (b) requirement of a linear $\varphi_1(\gamma)$.

always be found given bounds on the individual powers P_j. In practice, $\tilde{\gamma}$ may be considered as a design parameter; to this end, we note that, if $\tilde{\gamma}$ is very small, the problem may become infeasible since the set of possible solutions would be very conservative. Conversely, a large $\tilde{\gamma}$ would require a high value of m, due to the aforementioned reasons. Despite the fact this might not affect the optimal value (considering a proper approximation), it increases the computational cost since each $\ell \in [m]$ represents a constraint in the optimization problem. For each new function approximation, i.e new ℓ, n more constraints must be satisfied, since it is related to all users throughput levels, increasing the computation time. Therefore, $\tilde{\gamma}$ should be set properly, so the problem is feasible and m is not very large. Note that in case of an infeasible problem (which could happen by not respecting the constraints or the interval $[0, \tilde{\gamma}]$), both m and $\tilde{\gamma}$ can be increased. In the simulations sections we show how



(a) (b)

Figure 4.2: (a) Impact of m and (b) constraints active intervals.

m may be chosen to trade-off optimality and computational complexity.

Finally, we point out that the active interval selection for each $\varphi_\ell(\gamma)$ does not need to follow a specific rule. For instance, one could just divide the approximation interval into m ones of equal length.

4.1.1 PPF approximation error

For any $\gamma \in [\gamma_\ell,\gamma_{\ell+1}]$ in (4.4), the error between the original function and its respective approximation is given by

$$\mathrm{erf}_\ell(\gamma) = \log_2(1+\gamma) - a_\ell \gamma^{b_\ell}. \tag{4.7}$$

Therefore, by setting $\mathrm{erf}'(\gamma_\ell) = 0$ it is possible to calculate the point at which the error is maximized, according to

$$\begin{aligned} \mathrm{erf}'_\ell(\gamma) =& \frac{1}{\log(2)(1+\gamma)} - \frac{a_\ell b_\ell}{\gamma^{1-b_\ell}} \\ =& \frac{\gamma^{1-b_\ell} - a_\ell b_\ell \log(2)(1+\gamma)}{\log(2)(1+\gamma_\ell)\gamma^{1-b_\ell}} = 0, \end{aligned} \tag{4.8}$$

which is satisfied if

$$\varrho_\ell(\gamma) = \kappa_\ell \gamma^{1-b_\ell} - \gamma - 1 = 0, \tag{4.9}$$

with

$$\kappa_\ell = \frac{1}{\log(2) a_\ell b_\ell}.$$

Despite the fact (4.9) cannot be solve analytically, in the interval $[\gamma_\ell,\gamma_{\ell+1}]$ it is differentiable and has only one possible solution, as a consequence of inequalities in (4.4). Then,

with a few iterations of the Newton-Raphson numerical method, shown in Algorithm 1, the solution can be easily obtained.

Algorithm 1: Maximum error in PPF

Choose the interval ℓ.

Define starting point:

$^0\gamma \in (\gamma_\ell, \gamma_{\ell+1})$

$^1\gamma \in (^0\gamma + \epsilon, \infty)$

while $|\Delta| > \epsilon$ **do**

$^1\gamma = {}^0\gamma - \frac{\varrho(^0\gamma)}{\varrho'(^0\gamma)}$

$\Delta = {}^1\gamma - {}^0\gamma$

$^0\gamma = {}^1\gamma$

end

$\arg\max \mathbf{erf}_\ell(\gamma) = {}^1\gamma$

Note that

$$\varrho'_\ell(\gamma) = \kappa_\ell(1 - b_\ell)\gamma^{-b_\ell} - 1, \tag{4.10}$$

ϵ is a small number serving as a convergence condition, and $^1\gamma$ should be initiated such that the **while** condition is satisfied at the first iteration. As presented in Table 4.1, the convergence of the algorithm for this application is usually very fast (not more than 5 iterations) considering $\epsilon = 0.0001$, $\tilde{\gamma} = 500$, and

$$^0\gamma = \frac{\gamma_\ell + \gamma_{\ell+1}}{2}.$$

$\max \mathrm{erf}(\gamma_\ell)$ refers to the maximum error with respect to interval $[\gamma_\ell, \gamma_{\ell+1}]$ for $\ell \in [m]$.

Table 4.1: Newton-Raphson method convergence.

m	$\max \mathrm{erf}_\ell(\gamma)$	Iterations
2	[1.5493, 0.3662]	[2, 5]
3	[0.7322, 0.2178, 0.1464]	[2, 4, 5]
4	[0.4217, 0.1431, 0.1038, 0.0786]	[2, 4, 4, 4]
5	[0.2730, 0.1007, 0.0773, 0.0607, 0.0489]	[2, 4, 4, 4, 4]
6	[0.1908, 0.0744, 0.0595, 0.0483, 0.0398, 0.0334]	[2, 3, 4, 4, 4, 4]
7	[0.1407, 0.0571, 0.0471, 0.0393, 0.0330, 0.0281, 0.0242]	[2, 3, 3, 3, 4, 4, 4]

The intervals $[\gamma_\ell, \gamma_{\ell+1}]$ were selected such that $f(\gamma_{\ell+1}) - f(\gamma_\ell)$ have the same value for $\ell \in [m]$, as illustrated in Figure 4.3.

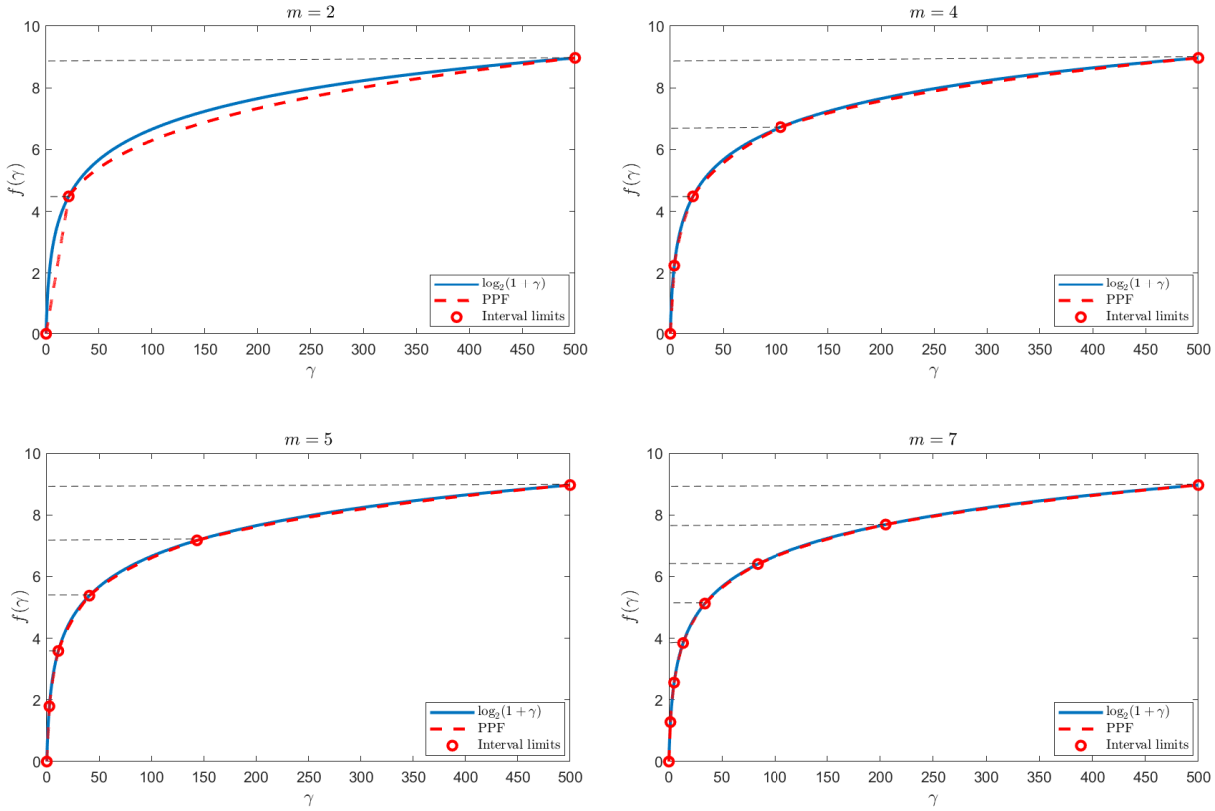


Figure 4.3: Constant values of $f(\gamma_{\ell+1}) - f(\gamma_\ell)$ for all $\ell \in [m]$.

4.1.2 A convex constraint via PPF

The importance of the proposed approximation relies in the fact it allows to reformulate our original problem in convex form. This is shown in the following theorem, which represents the main technical result of the work.

Theorem 1 *Given m subintervals between 0 and $\tilde{\gamma}$, define m approximating functions as follows*

$$\hat{f}_\ell(q, u_{ij}) \doteq \log\left(\frac{\eta^2}{g_{ij}} e^{-q_j - \frac{u_{ij}}{b_\ell}} + \sum_{k \neq j} \frac{g_{ik}}{g_{ij}} e^{q_k - q_j - \frac{u_{ij}}{b_\ell}}\right), \tag{4.11}$$

for $\ell \in [m]$. Then, constraint in (3.7) is implied by the following set of (possibly more conservative) m convex constraints

$$\hat{f}_\ell(q, u_{ij}) \leq \frac{\log\left(\frac{B_j a_\ell}{t_i}\right)}{b_\ell} + M\bar{z}_{ij}, \quad \ell \in [m], i \in [n], j \in [N]. \tag{4.12}$$

Moreover, the conservatism of this approximation vanishes for $m \to \infty$.

Proof: By using the proposed approximation, inequality (3.7) can be written as

$$C_i \simeq x_{ij} B_j a_\ell \left(\frac{P_j g_{ij}}{\eta^2 + \sum_{k \neq j} P_k g_{ik}}\right)^{b_\ell} \geq t_i, \quad \ell \in [m]. \tag{4.13}$$

Then, with some mathematical manipulations and taking the log of both sides, we have

$$a_\ell \left(\frac{\eta^2}{g_{ij}} P_j^{-1} x_{ij}^{-\frac{1}{b_\ell}} + \sum_{k \neq j} \frac{g_{ik}}{g_{ij}} P_k P_j^{-1} x_{ij}^{-\frac{1}{b_\ell}} \right)^{-b_\ell} \geq \frac{t_i}{B_j},$$

$$-\log \left(\frac{\eta^2}{g_{ij}} P_j^{-1} x_{ij}^{-\frac{1}{b_\ell}} + \sum_{k \neq j} \frac{g_{ik}}{g_{ij}} P_k P_j^{-1} x_{ij}^{-\frac{1}{b_\ell}} \right) \geq \frac{\log\left(\frac{t_i}{B_j a_\ell}\right)}{b_\ell}.$$

With the variables transformation $P_j = e^{q_j}$ and $x_{ij} = e^{u_{ij}}$, it is possible to write

$$-\log \left(\frac{\eta^2}{g_{ij}} e^{-q_j - \frac{u_{ij}}{b_\ell}} + \sum_{k \neq j} \frac{g_{ik}}{g_{ij}} e^{q_k - q_j - \frac{u_{ij}}{b_\ell}} \right) \geq \frac{\log\left(\frac{t_i}{B_j a_\ell}\right)}{b_\ell},$$

then we have

$$\hat{f}_\ell(q, u_{ij}) \leq \frac{\log\left(\frac{B_j a_\ell}{t_i}\right)}{b_\ell}, \tag{4.14}$$

which is constraint in equation (4.12) before the Big-M Method. It should be noticed that constraint in (4.14) is a *log-sum-exp* function, which is convex. The step-by-step of this conversion is given in Appendix A.

4.2 MIGP optimization problem

The results of the previous section allow us to formulate our main optimization problem. To this end, let $\tilde{P}_j$ be the maximum transmission power of a RB in BS j, defined as the ratio between total transmission power of a BS and its number of RBs. Hence, a Mixed-Integer Geometric Programming (MIGP) problem that solves optimization problem (3.15) (non-convex-OFDMA) is given by

Non-Convex

$$\min_{x_{ij},z_{ij},q_j} \quad \sum_{j=1}^{N} e^{q_j} \qquad \text{(non-convex-OFDMA)}$$

s.t.:

$$x_{ij} \in [0,1], \quad i \in [n], \quad j \in [N], \tag{3.15a}$$
$$z_{ij} \in \{0,1\}, \quad i \in [n], \quad j \in [N], \tag{3.15b}$$
$$\sum_{j=1}^{N} \bar{z}_{ij} = N-1, \quad i \in [n], \tag{3.15c}$$
$$\sum_{i=1}^{n} x_{ij} \leq 1, \quad j \in [N], \tag{3.15d}$$
$$f(q, x_{ij}) \geq t_i - M\bar{z}_{ij}, \quad i \in [n]. \tag{3.15e}$$

MIGP Reformulation

$$\min_{\bar{z}_{ij},u_{ij},q_j} \quad \sum_{j=1}^{N} e^{q_j} \qquad \text{(OP-MIGP)}$$

s.t.:

$$e^{u_{ij}} \leq 1, \quad i \in [n], j \in [N] \tag{4.15a}$$
$$\bar{z}_{ij} \in \{0,1\}, \quad i \in [n], \quad j \in [N], \tag{4.15b}$$
$$e^{q_j} \leq \tilde{P}_j, \quad j \in [N], \tag{4.15c}$$
$$\sum_{j=1}^{N} \bar{z}_{ij} = N-1, \quad i \in [n], \tag{4.15d}$$
$$\sum_{i=1}^{n} e^{u_{ij}} \leq 1, \quad j \in [N], \tag{4.15e}$$
$$\hat{f}_\ell(q,u_{ij}) \leq \frac{\log\left(\frac{B_j a_\ell}{t_i}\right)}{b_\ell} + M\bar{z}_{ij}, \quad i \in [n], j \in [N], \ell \in [m]. \tag{4.15f}$$

Note that the above problem is a mixed-integer one (since it contains the binary variables $\bar{z}_{ij}$) which solves simultaneously association, assignment, and the power optimization problems. It should be noted that, for given $\bar{z}_{ij}$, i.e. for given association, the problem becomes a classical Geometric Program, i.e. a convex optimization problem for which a global optimal solution can be found in polynomial-time. This is an important remark since, as we will see in the application section, there are many practical situations in which simple assignment rules appear to work rather efficiently. In such cases, our approach hence becomes even faster and more scalable. When $\bar{z}_{ij}$ are problem decision variables, only sub-optimal results are guaranteed. A more detailed discussion on the complexity of the MIGP approach is discussed next.

4.2.1 Sub-carrier assignments given x

In OFDMA networks, the wide bandwidth transmitted signal is divided into many orthogonal narrow band sub-carriers, such that they do not interfere with each other. Sets of these sub-carriers form a resource block, which can be allocated to the users according to their throughput requirements (represented by the variables $x_{ij} = [0,1]$). Therefore, after solving the aforementioned optimization problem, we still need to assign ρ_{ij} RBs to user i connected to BS j given the respective x_{ij} computed via optimization problem (4.15).

Since each RB is allocated exclusively to one user at a given instant, ρ_{ij} should clearly be an integer variable. Ideally, if a user requires a non-integer ρ_{ij}, one could just round such a value to its smallest greater integer (ceil function). In practice this is not possible,

since the number of RBs is limited and the ceil rounding of all ρ_{ij} would lead to more RBs than the BS can provide. Instead, we propose to solve the optimization problem where the users' throughput requirements and the BSs available working bandwidth are slightly increased and decreased, respectively. As a consequence, some of the users would still respect the throughput requirements even if ρ_{ij} is rounded to its greatest smaller integer (floor function) and some RBs could be allocated to those who do not satisfy such a condition. The procedure to obtain ρ_{ij} given x_{ij} is presented in Algorithm 2, with

- δ_1: parameter to increase users' throughput requirements;
- δ_2: parameter to decrease BS working bandwidth;
- B_0: bandwidth of one RB.

Algorithm 2: RB assignemtns given x_{ij}.

Increase/decrease throughput/working bandwidth:
$\hat{t}_i \leftarrow (1+\delta_1)t_i, \quad i \in [n]$
$\sum_{i=1}^{n} e^{u_{ij}} \leq 1-\delta_2, \quad j \in [N]$

Solve MIGP in (4.15)

for $i \in [n]$
  $j \leftarrow \arg\max(X(i,:))$
  $\rho_{ij} \leftarrow \frac{\max(X(i,:))B}{B_0}$
  if $\lfloor \rho_{ij} \rfloor B_0 \log_2(1+\mathrm{SINR}_{ij}(q)) \geq t_i$
    $\rho_{ij} \leftarrow \lfloor \rho_{ij} \rfloor$
  else
    $\rho_{ij} \leftarrow \lceil \rho_{ij} \rceil$
  end if
end for

Note that one needs to certify that the number of RBs assigned to the users does not exceed the maximum available. In case such a constraint is not respected, δ_1 and δ_2 can be increased. Practically, this means that more users respect the throughput requirement with $\rho_{ij} \leftarrow \lfloor \rho_{ij} \rfloor$ and more RBs are available to users that must use the ceil function. Increasing these parameters can lead to higher transmission power consumption, since they are responsible for increasing throughput demands (δ_1) and decreasing the available bandwidth (δ_2). In extreme cases, the optimization problem might even become infeasible. Consequently, they should be as small as possible.

4.3 Application use cases

We demonstrate the applicability of the proposed algorithm in few HetNet deployments, based on the real-world configuration of a production-grade RAN. To this end, we first outline the process we followed to generate faithful HetNet deployment replicas of an MNO in a major European city. Next, we present how MIGP enables energy-efficient resource allocation in a HetNet, benchmarking its performance against other available solutions. Finally, we probe the scalability of MIGP and of the benchmarks, as the complexity of the HetNet, the number of users, and their traffic demands increase.

4.3.1 Realistic scenario generation

We generate our realistic scenarios based on the data obtained from the production network of a leading mobile network operator (MNO) in a European capital city. The data provided includes the location of the transmitting BS and the configuration of their antennas, *i.e.*, the antenna type (macro or micro cell) and steering direction, the technology for that deployment (2G, 3G, 4G, or 5G NSA), and the corresponding statistics about session-level traffic demands. This information allows generating a high-fidelity digital twin of the actual network that enables dependable assessments under realistic conditions.

For each scenario, S, abiding by the operational network topology we choose a set of base stations, BS_j, composed of one or more antennas, and we measure the number of sessions arriving on a 1-minute interval of an average busy work day. This interval represents a moment where the network operates at higher loads, with the number of connected users and their traffic demands reaching their peak values. We do not access individual users data in order to comply to privacy rules of the country where the network is located; instead, our data is aggregated at BS level with 1 minute granularity. This means we cannot simply replicate the session-level values, but need to rely on statistics from the measured values.

For each scenario S, we estimate the number of session arriving at each BS_j during the time interval of a busy work day. From this, we calculate the shares of sessions belonging to each mobile applications, obtaining a heterogeneity of session demands in each scenario, since each mobile service has its own characteristic traffic load. For each expected session arriving at BS_j, we estimate traffic and duration (and thereby throughput) according to the demands of the corresponding mobile service. These demands are calculated based on the statistics collected at BS level across the network over multiple months, which allows us to reliably estimate the session-level traffic probability distribution function of each mobile service, which we use to generate samples of traffic consumed by each session. In

this regard, this methodology not only respects the architecture of a major operational network but also the expected density and spatial distribution of users, while complying with user-privacy rules.

Each generated session is assigned to an individual user placed in a random location, lying within the covered area of s-th HetNet topology, determined initially by a Poisson random point process but with additional conditions to respect the boundaries of the street layout of the topology (as to not allocate too many users unrealistically inside buildings). We note that while looking at the traffic and solving the problem for each BS_j, we consider the full coverage area attributed to all 5G NSA-enabled antennas comprising each BS_j.

To evaluate the received signal distribution within the coverage area of each BS_j we employ a ray tracing simulator which allows us to carefully emulate the electromagnetic wave propagation in the urban fabric [34,35]. That enables us to quantify the interaction of the waves with objects in the propagation environment and to have accurate estimates of the channel gains $g_{i,j}$ required to evaluate equation (3.2). To conduct the site-specific ray tracing simulations for each HetNet topology, we fetch from Open Street Maps information related to the urban environment, such as the city layout and building height, the terrain type, and the foliage, Figure 4.4.

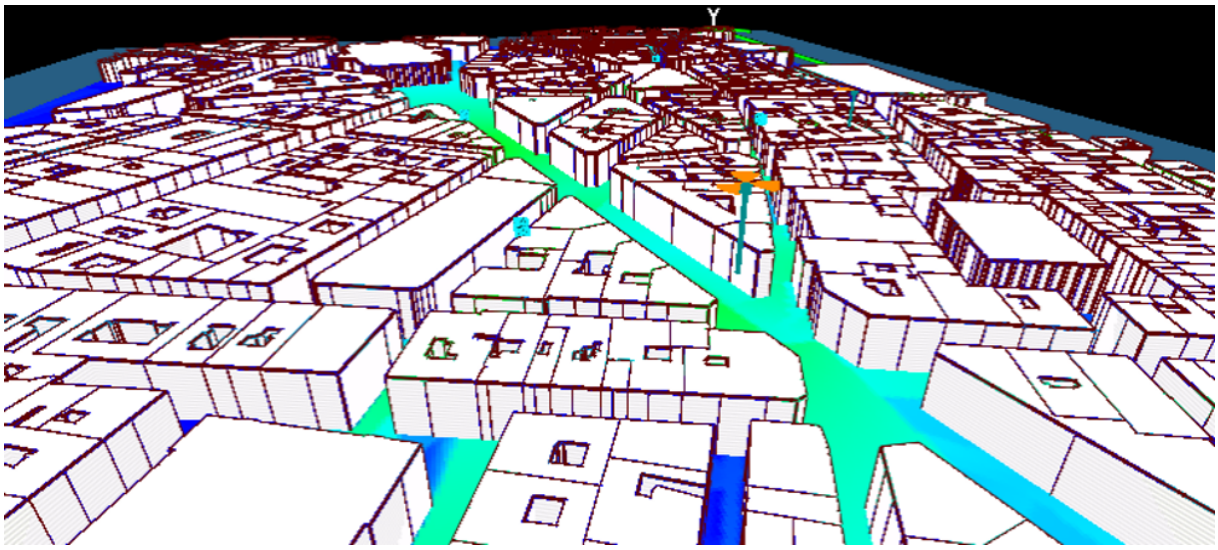


Figure 4.4: Site-specific ray tracing simulations with real-world environment information.

Subsequently, we place antennas and configure their properties according to the specifications provided by the MNO. A HetNet topology with 1 macro (BS_3) and 4 micro BSs is shown in Figure 4.5, depicting how the signal emitted by each BS propagates within the urban fabric, and how the micro BSs compensate for the low signal level of the macro cell at larger distances. It is worth mentioning that macro cell BS_3 is composed of 3 directional antennas, while the remaining micro cells contain a single antenna with an omnidirectional radiation pattern. In total, we consider 5 scenarios to evaluate the optimization algorithms, each one assuming different spatial and network complexities.

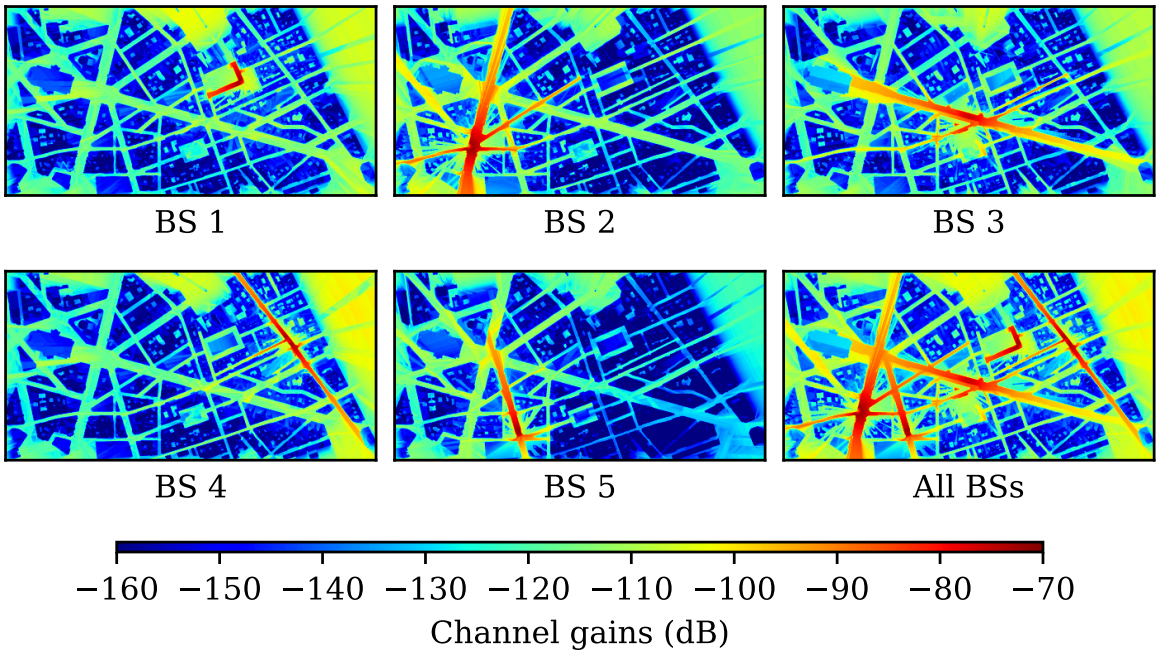


Figure 4.5: Channel gains provided by BSs and the full scenario for S_1.

Scenario S_1 is a *downtown* location that will be used in our comparative analysis in Section 4.3.2 and 4.3.3, containing 1 macro cell and 4 micro cells located in the historical center in the downtown area of the city under consideration. The remaining 4 scenarios are utilized in the scalability analysis in Section 4.3.5, where we expound on the resilience of MIGP at a variety of different HetNet topologies spread throughout the city. Scenarios S_2 and S_3 build on the HetNet operating at the same downtown region of S_1, but we increase the number of micro BS to 7 and 8, respectively, to evaluate the scalability of our solution in denser network configurations. Scenario S_4 represents a less dense HetNet topology located by the *river* side of the city, consisting of 1 macro and 2 micro cells. Finally, scenario S_5 is on an area at the *hill* side of the city, composed of 3 macro and 2 micro cells. All scenarios consider BSs that are 5G NSA-enabled.

The MIGP depends on the hyper-parameters presented in Table 4.2. The utility of each one was discussed on previous sections. The values of a_ℓ and b_ℓ allow a piecewise concave approximation of $\log_2(1+\text{SINR}_{ij}(\mathcal{P}))$ when $\text{SINR}_{ij}(\mathcal{P}) \in [0,\ 513.85]$. Despite the fact there is no closed-form for setting δ_1 and δ_2, we did so by gradually increasing both and checking if the solution to the largest scenario (800 users) could respect the maximum number of RBs (500/BS). Then, the parameters that provided a feasible solution to the largest scenario ($\delta_1 = 0.05$ and $\delta_2 = 0.16$) were also used for all others. $\delta_2 > \delta_1$ because by experimental results, we verified that the δ_2 has more influence on providing a solution inside the physical network limitation on the number of RBs.

Table 4.2: Optimization problem hyper-parameters.

Hyper-parameter	Value
δ_1	0.05
δ_2	0.16
M	10^6
a_ℓ, $\ell \in [m]$	[1.408, 0.7330, 1.3150, 1.9061, 3.1232]
b_ℓ, $\ell \in [m]$	[1, 0.7821, 0.4201, 0.2589, 0.1697]

4.3.2 Qualitative performance analysis

Initially, we demonstrate the applicability of the proposed framework to a single HetNet deployment of scenario S_1, benchmarking the performance of MIGP with other available approaches typically employed for user allocation. To this end, we consider a 5G new radio (NR) access network operating at 3.5 GHz with 500 resource blocks, utilizing an OFDMA technique with 100 MHz of bandwidth. As previously mentioned, S_1 consists of 1 macro and 4 micro BSs that accommodate the throughput requirements of 400 users, bounded between 1.35 Kbps and 18.72 Mbps. Pushing for greener communications, via MIGP we aspire to minimize the total transmission power of the HetNet. The noise power is set to -174 dBm/Hz.

The resulting assignment, Figure 4.6, shows a clear tendency of connecting the users to the BS providing the highest channel gain. We also note on the map a pattern of users being mostly around the outside of buildings, which is amplified by the propagation loss experienced by 3.5MHz communication inside buildings. As previously explained, the working bandwidth variables in x are relaxed to a continuous domain, but in practice we need integer numbers ρ_{ij} representing the number of resource blocks made available to a user. Therefore, after solving the MIGP, such a conversion is done by using Algorithm 2. The number of users n_j connected to BS j, the number of RB used in each BS, and the respective transmission powers are provided in Table 4.3. P_j^{Tot} stands for the total transmission power of BS j (each RB consuming P_j Watts).

Table 4.3: Number of users connected to each BS for S_1.

BS j	n_j	RB_j	P_j (W)	P_j^{Tot} (W)
$j=1$	43	423	0.2894×10^{-3}	0.1224
$j=2$	161	495	0.7117×10^{-3}	0.3523
$j=3$	142	490	0.6535×10^{-3}	0.3202
$j=4$	24	413	0.0038×10^{-3}	1.5694×10^{-3}
$j=5$	30	435	0.2151×10^{-3}	0.0936
$\sum_{j=1}^{N}$	400	2256	0.0019	0.8901

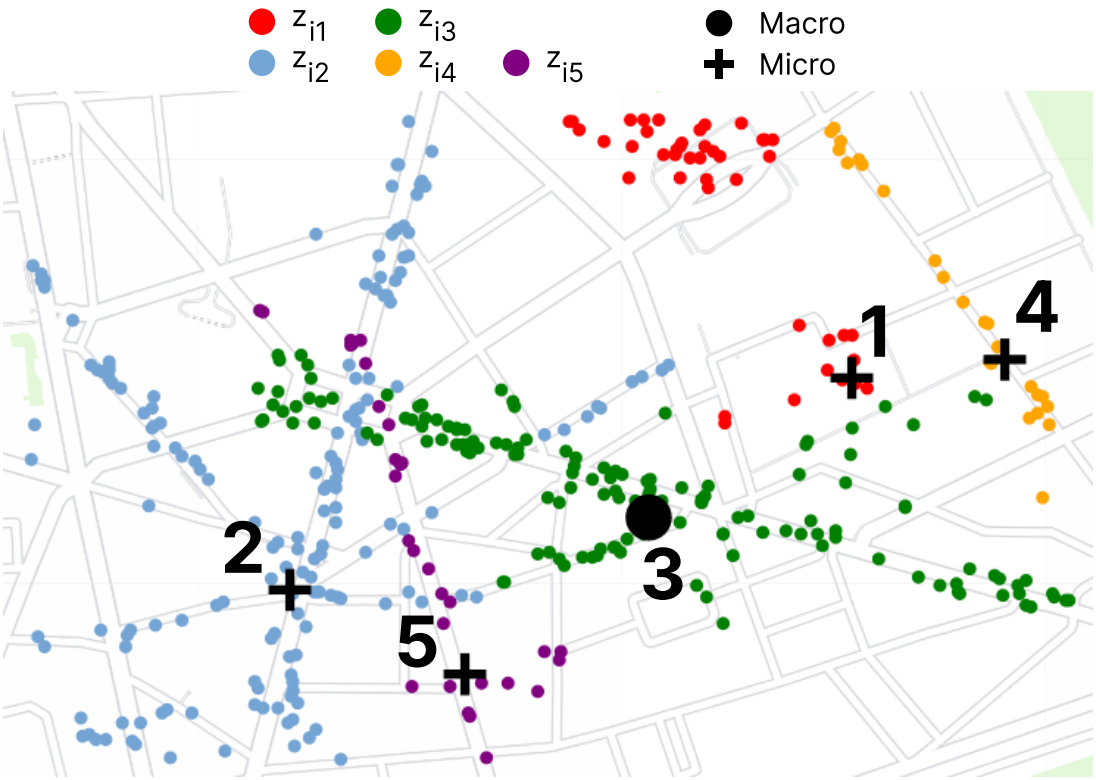


Figure 4.6: Users' assignment after MIGP optimization for S_1.

4.3.3 Comparison with global optimization approaches

Since our proposed approach deals simultaneously with the variables x, z and $\mathcal{P}$, we can compare it with some standard procedures performed in realistic network scenarios. As described in [8], the users' association can be done with simple rules, such as *Received Signal Power* or *Maximum channel gain*, defined below.

- Received Signal Power (RSP): a user i is associated to the base station j that maximizes the product between transmission power and channel gain, i.e., $j = \arg\max_j(g_{ij}P_j)$.
- Maximum channel gain (MCG): a user i is associated to the base station j that maximizes the channel gain between i and j, that is, $j = \arg\max_j(g_{ij})$.

Note that these simple association rules provide a matrix z. Therefore our problem becomes a Geometric Program (GP), which is convex. In Figure 4.7, we used the MCG rule and refer to this situation as (GP + fixed z_{ij}). For S_1, where each base station j have n_j users assigned via association rule, the resources can be equally shared between the UE according to

$$x_{ij} = \begin{cases} \frac{1}{n_j}, \text{ if } z_{ij} = 1, \ \forall i \in [n], \ \forall j \in [N] \\ 0, \text{ otherwise.} \end{cases}$$

This situation is referred as (GP + fixed z_{ij}, x_{ij}). We also compare our approach with its early stage (without Big-M trick, piecewise concave approximation, and variables transformation) given by optimization problem (3.8). Since this is highly non-convex, finding its optimal solution is not trivial and in general just local optimal is achieved. To deal with this problem, we adopt a general global optimization approach. In particular, we compare our MIGP approah with one of the state-of-the-art commercial software for global optimization, currently utilized by main players in network optimization use: the package MIDACO (Mixed Integer Distributed Ant Colony Optimization) solver is able to solve mixed integer non-linear programs with discontinuities, non-convexity, and stochastic noise. More details on MIDACO are given in [36].

In Figure 4.7, the optimal value considering the situations described above is calculated for different numbers of users connected to the network during S_1, keeping the % of sessions across mobile applications consistent to respect traffic dynamics. The corresponding KWh consumption if the power levels are kept constant for our network snapshot are also shown. The number of users associated to each BS is in Figure 4.8.

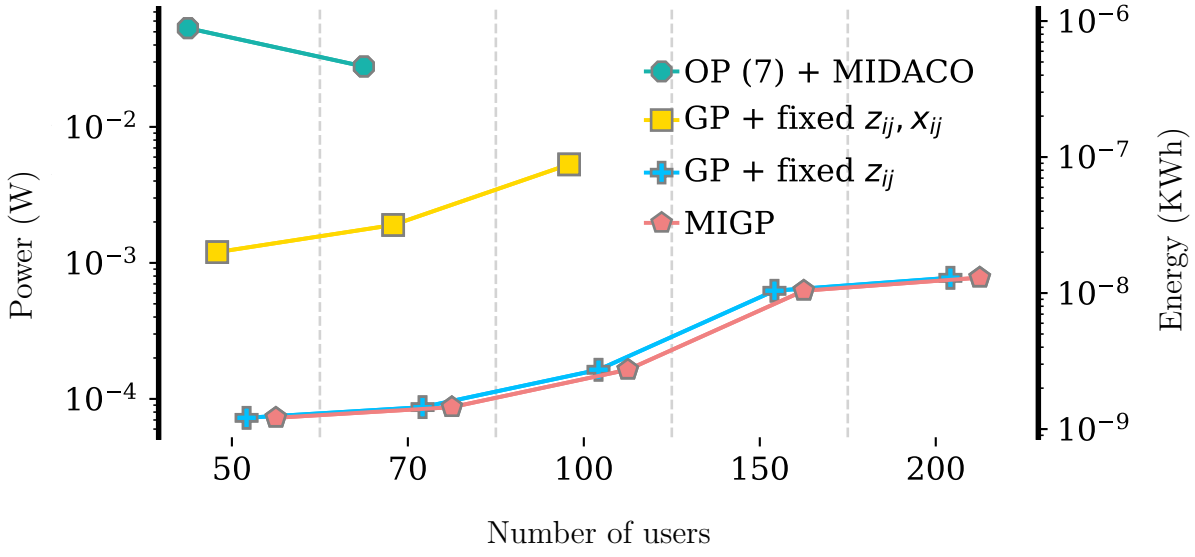


Figure 4.7: Optimal value as a function of number of users in the network for different approaches for S_1.

From these results, a series of keypoints can be highlighted.

1. **GP+fixed z_{ij} × MIGP**: the assignments and the corresponding optimal values obtained with MIGP and (GP + fixed z_{ij}) are almost the same for this network. This happens because the BSs are directional and without large cross-interference, as shown in Figure 4.5. Physically, one could say that the BS deployment is good, which was expected since the data set is from a realistic scenario from a major European city. Regarding the optimization problem point of view, assigning the users according to the MCG rule makes the problem considerably simpler, since it

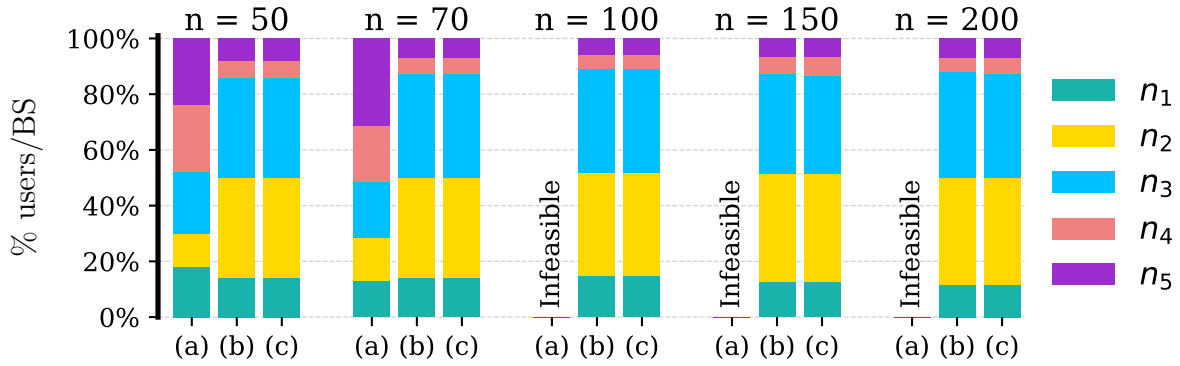


Figure 4.8: Number of users connected to each BS of S_1 for (a) OP(7) + MIDACO, (b) GP + fixed z_{ij} and (c) MIGP.

becomes purely convex (can be solved globally and efficiently) and the number of problem variables is reduced, since $\boldsymbol{z}$ would be known in advance. However, there is no guarantee that using the MCG rule in a network where the cross-interference between BSs is large might be a good option. In such a situation, solving the MIGP is more appropriate.

2. **OP** (3.8)+**MIDACO**: as already discussed, the problem in its most natural format is highly non-convex. We solved it with MIDACO considering that the stop condition is *true* when the optimal value converges and the initial condition is given by the lower bound of each variable. One may notice that the results obtained with both MIGP and (GP + fixed $\boldsymbol{z_{ij}}$) are considerably better in terms of optimal value when compared to (OP (3.8) + MIDACO). Therefore, the piecewise concave approximation followed by the variables transformation are indeed efficient mathematical tools to pose the problem in a more suitable format.

3. **GP+fixed $\boldsymbol{z_{ij}}$,$\boldsymbol{x_{ij}}$**: fixing the working bandwidth $\boldsymbol{x}$ of users considerably degrades the optimal value, since each one has specifics throughput requirements and channel gains. Furthermore, for $\boldsymbol{n \geq 150}$ the problem becomes infeasible, therefore sharing such a resource equally between users is not a good choice and $\boldsymbol{x}$ should remain as problem variable. Note that the approach of fixing $\boldsymbol{x}$ is rather common in the literature [37].

4.3.4 Comparison between GP and sequential approaches

Parallel to the previous analysis and comparison with global optimization methods, we also performed some numerical comparison between our approach and two "classical" iterative solutions: i) a successive linearization approach and gradient descent (GD), and ii) a more sophisticated DC-programming approach. In both cases, to perform a fair

comparison, we considered the case of fixed assignment, that is case 1) above, with fixed z_{ij}.

For the GD case, a classical linearization of the function

$$\log_2\left(1+\frac{P_j g_{ij}}{\eta^2+\sum_{k\neq j} P_k g_{ik}}\right)$$

around the current feasible point is performed. For the second case, we used a difference-of-concave (DC) functions approximation of the form

$$\log_2\left(1+\frac{P_j g_{ij}}{\eta^2+\sum_{k\neq j} P_k g_{ik}}\right) \simeq$$
$$\log_2\left(\eta^2+\sum_{j=1}^{N} P_j g_{ij}\right) - h(\mathcal{P}(0)) - \nabla h(\mathcal{P}(0))^\top(\mathcal{P}-\mathcal{P}(0)),$$

with $h(\mathcal{P}) = \log_2\left(\eta^2+\sum_{k\neq j} P_j g_{ij}\right)$ and $\mathcal{P}(0)$ being the point around which each linearization is calculated.

Due to the limitations of sequential approaches discussed in 1.1.1, we set the constraint $||\mathcal{P}-\mathcal{P}(0)||_\infty \leq 10^{-4}$ for GD_1 and DC_1, and $||\mathcal{P}-\mathcal{P}(0)||_\infty \leq 10^{-5}$ for GD_2 and DC_2. The initial condition for the GD, DC, and GP approaches were determined as:

1. each user is associated to the BS providing the highest channel gain (all techniques);
2. the amount of bandwidth allocated to a user is proportional to the respective throughput demand, according to (GDs, DCs, GP + fixed x_{ij}, z_{ij}):

$$x_{ij} = \frac{t_i}{\sum_{i=1}^{n} t_i}, \;\; j \in [N]$$

3. $P_j = 0.08$ for all $j \in [N]$ (experimentally verified as a feasible initial condition- GDs and DCs).

The obtained results are presented in Table 4.4. Note that our approaches (GP + fixed z_{ij} and GP + fixed x_{ij}, z_{ij}) provide not only dramatically better optimal values, but also requires a substantially smaller runtime. Additionally, the scenarios have just 30 and 40 users because of the complexity in finding initial feasible resource allocation for larger cases, which is necessary for GDs and DCs.

4.3.5 Scalability of MIGP across different network topologies

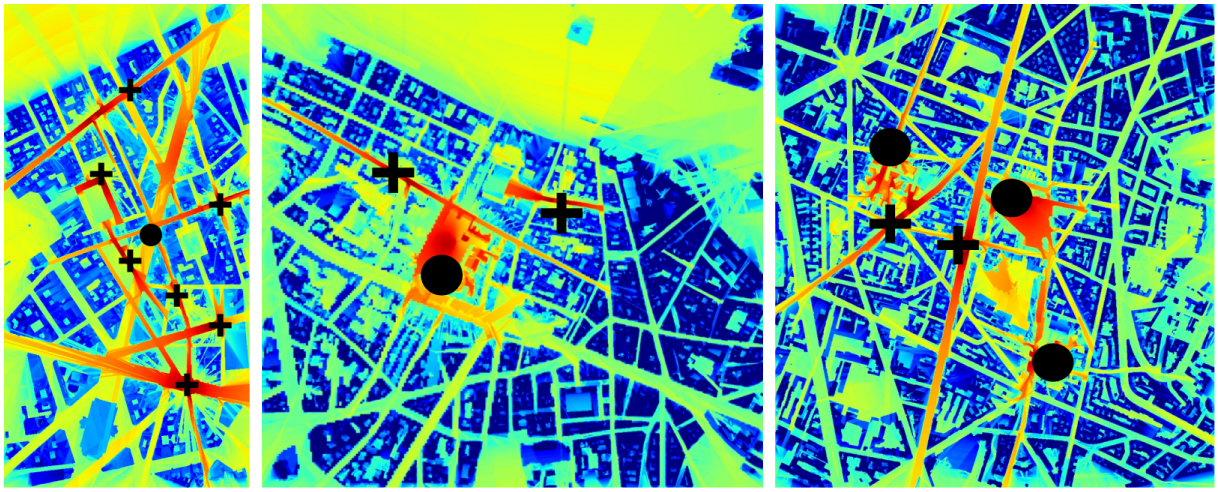
Finally, we test how MIGP performs across different network topologies, which offer variations on the number of BS, density of the deployment, number of users and propagation paths due to the geography of the city. We exemplify the scenarios through Figure

Table 4.4: Comparison between GP and sequential approaches.

Tech.	**n = 30**		**n = 40**	
	t(s)	$\sum_j P_j$	t(s)	$\sum_j P_j$
GD_1	217.9	0.2650	266.7	0.2650
GD_2	2230.5	0.2651	2677.5	0.2651
DC_1	471.7	1×10^{-3}	550.5	1×10^{-3}
DC_2	4701.9	6.52×10^{-4}	5436.8	7.02×10^{-4}
GP + fixed x_{ij}, z_{ij}	6.2	1.68×10^{-4}	8.3	1.98×10^{-4}
GP + fixed z_{ij}	6.1	5.83×10^{-5}	8.2	6.75×10^{-5}

4.9, which shows downtown scenario S_3 (over same area as S_1 and S_2, but with more BS), river scenario S_4 and hill scenario S_5.

Figure 4.10 compares the optimal values obtained with MIGP and OP (3.8)+MIDACO for the 4 network scenarios, i.e., distinct throughput requirements, channel gains, number of BS, and users. In all cases, the MIGP leads to better results.



(a) Downtown (b) River (c) Hills

Figure 4.9: Channel gains and antenna placements for (a) dense downtown deployment of scenario S_3, (b) the less dense river deployment of S_4 and (c) the hill deployment of S_5.

The results above show that MIGP, contrary to OP (3.8)+MIDACO, is able to provide feasible solutions even for scenarios with considerable numbers of BSs and users. Since the association variables are naturally binary, increasing the number of BSs is computationally more expensive than increasing the number of users. However, as previously explained, in environments with good BS deployment (not large cross interference), the MIGP generally associates a user with the BS providing the respective highest channel gain. When using simple association rules, the problem becomes purely convex and can be solved globally

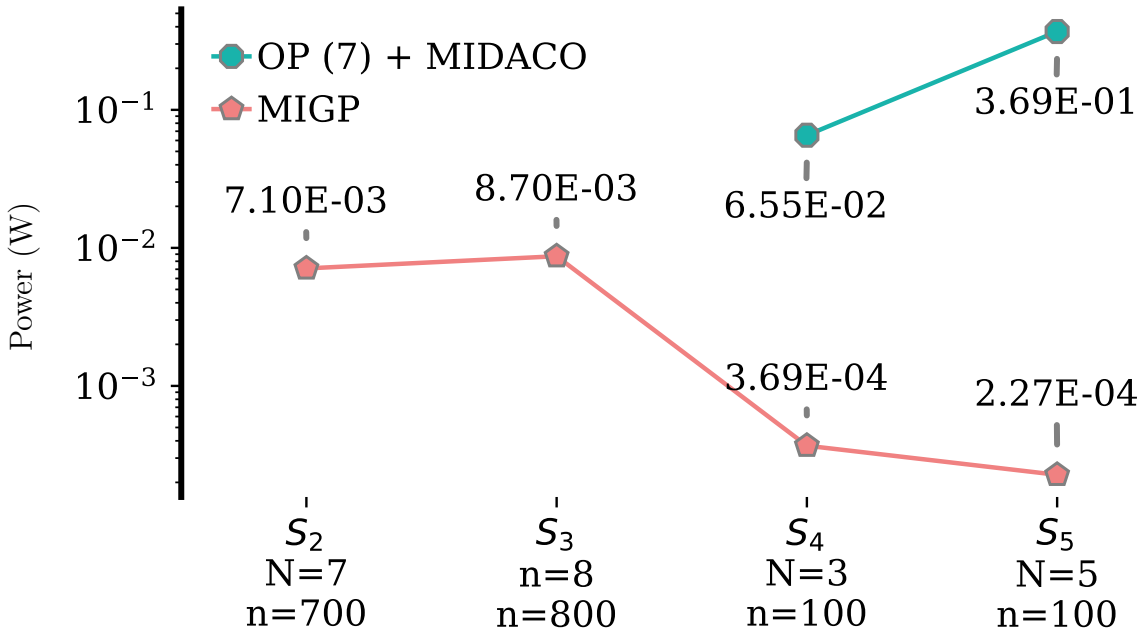


Figure 4.10: Power consumption $\sum_{j=1}^{N} P_j$ for different networks.

and efficiently.

4.3.6 The impact of m on the optimal value and optimization runtime

As previously discussed in Remark 1, the number m of concave power functions used in the PPF approximation has an impact both on the goodness of the approximation, and consequently on the obtained optimal value, and on the optimization problem complexity, and thus on the solution runtime. Therefore, the choice of m should obey a trade-off between accuracy and computational complexity. In particular, we run experiments varying the number of functions m, considering network scenarios with 50, 100, and 200 users. As presented in Figure 4.11, small values of m lead to solutions with lower execution time, however higher values for the objective function. Conversely, for values of m greater than 15, the time to compute a solution increases without yielding significant improvement on the BSs transmission powers.

Analyzing the optimal values plots of Figure 4.11, we observe that the value $m = 5$ may be considered a good trade-off, since it represents the point where the curve changes its slope. More generally, we suggested choosing values of m in the interval 5-10.

4.4 Final considerations

This Chapter introduced an innovative approach to address the challenge of minimizing transmission power in OFDMA heterogeneous networks while ensuring individual users' throughput requirements are met. The proposed method is formulated as a Mixed-Integer

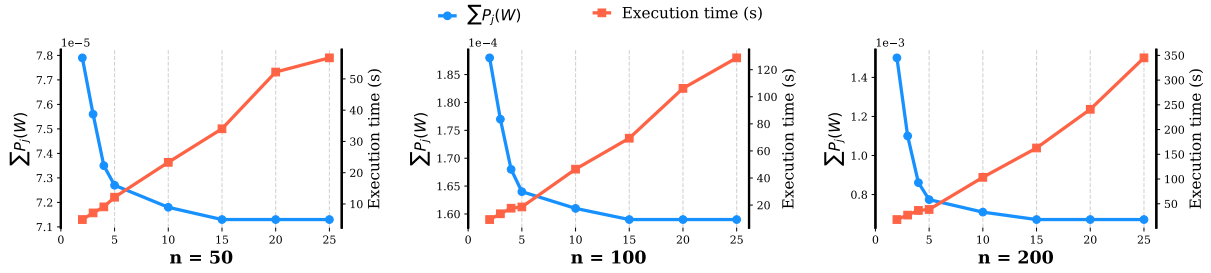


Figure 4.11: Optimal cost $\sum_{j=1}^{N} P_j$ and execution time considering different values of m and number of users n.

Geometric Program, leveraging the Big-M method to handle the problem's combinatorial nature. Additionally, a piecewise power function approximation of the Shannon-Hartley Theorem is employed, followed by variable transformation to tackle the problem's non-convex characteristics effectively. The optimization problem involves determining the optimal users' working bandwidth, base station (BS) transmission powers, and users' associations. Remarkably, unlike many existing methods, this approach does not rely on successive or iterative algorithms and does not necessitate prior knowledge of a feasible initial condition for the problem. Moreover, it is shown that the problem becomes convex when simple users' association rules are employed.

The effectiveness of the proposed optimization problem is demonstrated through extensive testing in a highly realistic scenario with a considerable number of users. Channel gains are calculated using a high-performance propagation solver, ensuring a practical evaluation.

Overall, this work presents a promising optimization approach for enhancing the energy efficiency of heterogeneous networks while accommodating the increasing demands of connected devices, paving the way for greener and more efficient cellular networks in the future. We plan to share our code, so the MIGP can be easily compared to other approaches in the literature.

Chapter 5

Log-normal stochastic channel-gains: a Robust MIGP

The channel gains between UEs and BSs have a fundamental importance when optimizing mobile networks efficiency. In general, these parameters can be evaluated by means of site specific ray tracing solvers that simulate the radio waves propagation. These strategies (or any other relying on physics-based models) have the limitation of depending on how accurate the environment is described and reconstructed into the propagation solvers/models. However, even if the buildings geometry and electromagnetic properties of the construction materials are known, parameters such as motion of objects (people, cars, etc) make these quantities stochastic in practice [38]. In real life situations the electromagnetic waves propagated by the antennas will suffer not only from shadowing, but also combined effects of reflection and diffraction. Consequently, the channel gain between a UE/BS pair is inherently a random quantity. Additionally, the log-normal distribution has been empirically proved to be an accurate model for the channel gains in indoor and outdoor environments [39].

In this Chapter, we extend the work presented in the previous one to the case where the channel gains are random variables. We propose a joint optimization algorithm that minimizes the BSs transmission powers in OFDMA heterogeneous networks, while respecting chance-constrained individual users quality of service (QoS) requirements.

5.1 Problem extension

In Chapter 4, we propose the (4.15) (MIGP) which is an upper-bound solution to the non-convex integer optimization problem (3.15)(non-convex-OFDMA). The goal of the latter one is to minimize transmission powers of base stations while respecting individual users' throughput constraints.

5.1.1 The channel-gains as random quantities

In real-world environments, the channel gains are unknown random quantities, due to the combined effects of shadowing, multipath, diffraction, etc. In particular, as shown by [39], these quantities respect a log-normal distribution, i.e., the dB gains defined as

$$g_{ij}^{(dB)} \doteq 10 \log_{10} g_{ij}, \quad i \in [n], j \in [N] \tag{5.1}$$

obey a Gaussian distribution $g_{ij}^{(dB)} \sim \mathcal{N}(\tilde{\mu}_{ij}, \tilde{\sigma}_{ij}^2)$. We write this as $g_{ij} \sim \mathcal{LN}(\mu_{ij}, \sigma_{ij}^2)$. Furthermore, we can define the normalized gains as

$$\rho_{ij} \doteq \frac{10 \log_{10} g_{ij} - \tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}} \sim \mathcal{N}(0,1), \quad i \in [n], j \in [N]. \tag{5.2}$$

From (5.2) and with $c = \frac{\log 10}{10}$, we have

$$g_{ij} = e^{c(\tilde{\mu}_{ij} + \rho_{ij}\tilde{\sigma}_{ij})}, \quad i \in [n], j \in [N]. \tag{5.3}$$

5.1.2 Stochastic MIGP

Since the channel gains are random variables and the users' throughput levels are highly dependent on them, constraint (4.15f) becomes stochastic, and must be respected with a probability $\mathbb{P}_i$ given by $1 - \alpha_i$, leading to the following chance-constrained optimization problem

$$\min_{\bar{z}_{ij}, u_{ij}, q_j} \quad \sum_{j=1}^{N} e^{q_j} \tag{CC-MIGP}$$

s.t.:

$$e^{u_{ij}} \leq 1, \quad i \in [n], j \in [N] \tag{5.4a}$$

$$\bar{z}_{ij} \in \{0,1\}, \quad i \in [n], \quad j \in [N], \tag{5.4b}$$

$$e^{q_j} \leq \tilde{P}_j, \quad j \in [N], \tag{5.4c}$$

$$\sum_{j=1}^{N} \bar{z}_{ij} = N - 1, \quad i \in [n], \tag{5.4d}$$

$$\sum_{i=1}^{n} e^{u_{ij}} \leq 1, \quad j \in [N], \tag{5.4e}$$

$$\mathbb{P}_i \left(\hat{f}_\ell(q_j, u_{ij}) \leq \frac{\log\left(\frac{B_j a_\ell}{t_i}\right)}{b_\ell} + M\bar{z}_{ij} \right) \geq 1 - \alpha_i$$
$$\ell \in [m], i \in [n], j \in [N]. \tag{5.4f}$$

Chance constrained optimization problems might be difficult to handle. Then, inspired by the work of [40] we present a robust formulation based solution.

5.2 Robust MIGP

Proposition 5 (Box uncertainty) *Let ρ_{ij}, $i \in [n]$, $j \in [N]$ be Gaussian random variables with zero mean and unitary standard deviation, as in* (5.2)*. For each user, compute the quantities $[\underline{\rho}_{ij},\overline{\rho}_{ij}]$ such that*

$$\begin{cases} \mathbb{P}(\underline{\rho}_{ij} \leq \rho_{ij} \leq \overline{\rho}_{ij}) = \dfrac{erf\left[\frac{\overline{\rho}_{ij}}{\sqrt{2}}\right] - erf\left[\frac{\underline{\rho}_{ij}}{\sqrt{2}}\right]}{2} = \varphi_j, \;\; j \in [N], \\ \prod_{j=1}^{N} \varphi_j = 1 - \alpha_i. \end{cases} \tag{5.5}$$

Then a robust optimization problem considering the uncertainty box can be derived such that its optimal solution is feasible to the CC-MIGP.

Note that there are many solutions to Proposition 5, meaning that different boxes of uncertainties can be constructed and lead to the same probability level φ_j. Physically, one could use real-world data to set better interval values: if a specific BS j provides to a user i channel gains with low standard deviation, the respective interval $(\underline{\rho}_{ij} \leq \rho_{ij} \leq \overline{\rho}_{ij})$ can be smaller when compared to the channel gains provided by the other BSs. If no information is available, one could simply set intervals with the same length for the channel gains provided by all BSs, as later presented in the *application use cases*.

Theorem 2 *For uncertainty boxes given by Proposition 1, the optimal solution $u_{ij}^{\star}, \bar{z}_{ij}^{\star}, q_j^{\star}$ to*

$$\min_{\bar{z}_{ij},u_{ij},q_j} \quad \sum_{j=1}^{N} e^{q_j} \tag{Robust-MIGP}$$

s.t.:

$$e^{u_{ij}} \leq 1, \quad i \in [n], j \in [N] \tag{5.6a}$$

$$\bar{z}_{ij} \in \{0,1\}, \quad i \in [n], \quad j \in [N], \tag{5.6b}$$

$$e^{q_j} \leq \tilde{P}_j, \quad j \in [N], \tag{5.6c}$$

$$\sum_{j=1}^{N} \bar{z}_{ij} = N - 1, \quad i \in [n], \tag{5.6d}$$

$$\sum_{i=1}^{n} e^{u_{ij}} \leq 1, \quad j \in [N], \tag{5.6e}$$

$$\log(\eta^2 e^{\beta_{ij\ell}} + \sum_{k \neq j} e^{\zeta_{ik\ell}}) \leq \frac{\log\left(\frac{B_j a_\ell}{t_i}\right)}{b_\ell} + M\bar{z}_{ij},$$
$$\forall \rho_{ij} \in [\underline{\rho}_{ij},\overline{\rho}_{ij}] \;\; and \;\; \forall \rho_{ik} \in [\underline{\rho}_{ik},\overline{\rho}_{ik}]$$
$$\ell \in [m], i \in [n], j \in [N], \tag{5.6f}$$

with

$$\beta_{ij\ell} = -c(\tilde{\mu}_{ij} + \underline{\rho}_{ij}\tilde{\sigma}_{ij}) - q_j - \frac{u_{ij}}{b_\ell},$$
$$\zeta_{ik\ell} = c(\tilde{\mu}_{ik} + \overline{\rho}_{ik}\tilde{\sigma}_{ik} - \tilde{\mu}_{ij} - \underline{\rho}_{ij}\tilde{\sigma}_{ij}) + q_k - q_j - \frac{u_{ij}}{b_\ell},$$

is always feasible to the chance constrained optimization problem in (5.4)*, i.e.*

$$\textit{Robust MIGP}(u^\star_{ij}, x^\star_{ij}, P^\star_j) \implies \textit{Stochastic MIGP}.$$

Proof: Constraint (5.6f) is obtained by replacing the channel gains as given in (5.3) into (4.11). Additionally, the terms $\beta_{ij\ell}$ and $\zeta_{ik\ell}$ consider the worst-case scenario inside the uncertainty box, i.e. the problem is solved for $\underline{\rho}_{ij}$ (lowest channel gain) and $\overline{\rho}_{ik}$ (highest interference) for each user. Therefore, the optimal solution $u^\star_{ij}, x^\star_{ij}, P^\star_j$ is feasible for any other possible combination of $\rho_{ij} \in [\underline{\rho}_{ij},\overline{\rho}_{ij}]$ and $\rho_{ik} \in [\underline{\rho}_{ik},\overline{\rho}_{ik}]$ $\forall k \neq j$ inside the uncertainty box, concluding the proof.

Note that we can easily retrieve $\hat{f}_\ell(q_j,u_{ij})$ when the channel gains are given constants by making $\tilde{\sigma}_{ij} = \tilde{\sigma}_{ik} = 0$.

5.3 Application use cases

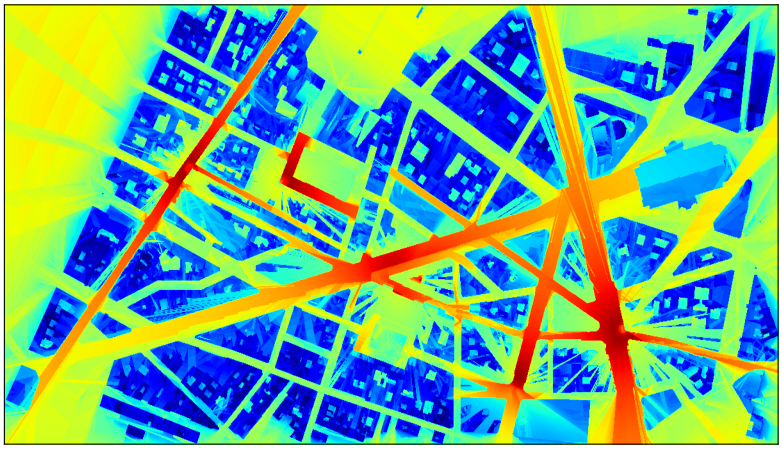
5.3.1 Channel gains expected values

We estimate the channel gains from a neighbourhood in a large European city by reconstructing it (considering building height, city layout, BSs location, etc) into a 3D ray tracing propagation software. Figure 5.1.(a) shows the channel gains expected values $\tilde{\mu}_{ij}$ and Figures 5.1.(b) and 5.1.(c) expose how the randomness in these quantities can lead to a different quality of signal and coverage. Therefore, a scenario optimized considering only the expected values might not respect all constraints in real-life applications. Specifically to the optimization problem at hand, not treating the channel gains as random quantities might lead to a solution where many UEs have less throughput than their respective required threshold, estimated according to [41].

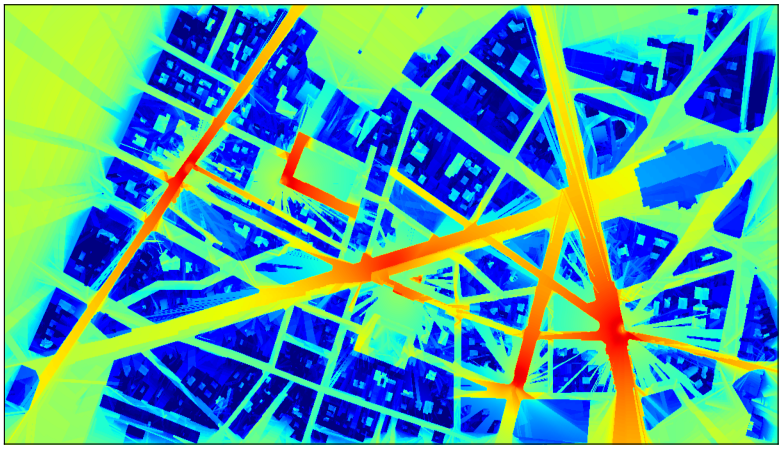
5.3.2 Optimization Scenario

All results obtained in the following subsections were obtained considering a neighbourhood from a large European city. The scenario consists of $N = 5$ base stations, $n = 130$ users, the channel gains expected values are the ones presented in Figure 5.1 (a), and $M = 10^6$. Additionally, we use $m = 5$ functions in the piecewise power function approximation, whose parameters are

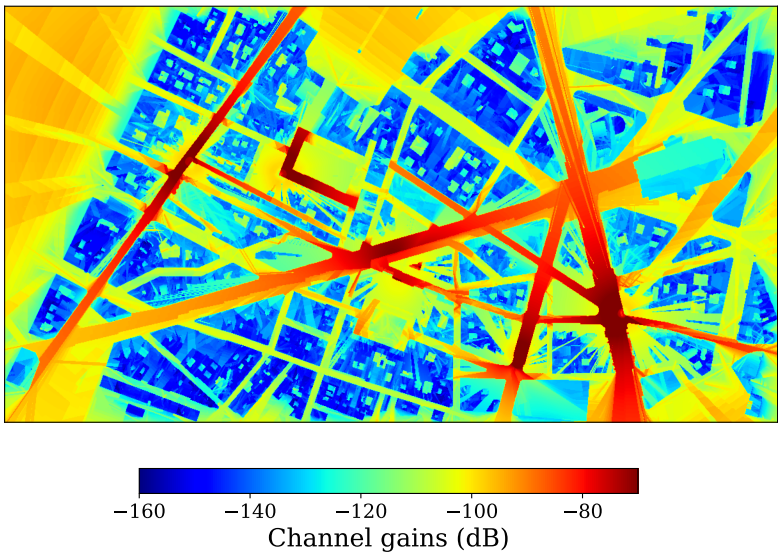
$$a_\ell = [1.4080, 0.7720, 1.3436, 2.0641, 2.8584] \quad \text{and}$$
$$b_\ell = [1, 0.79940.39280.25380.1840].$$



(a)



(b)



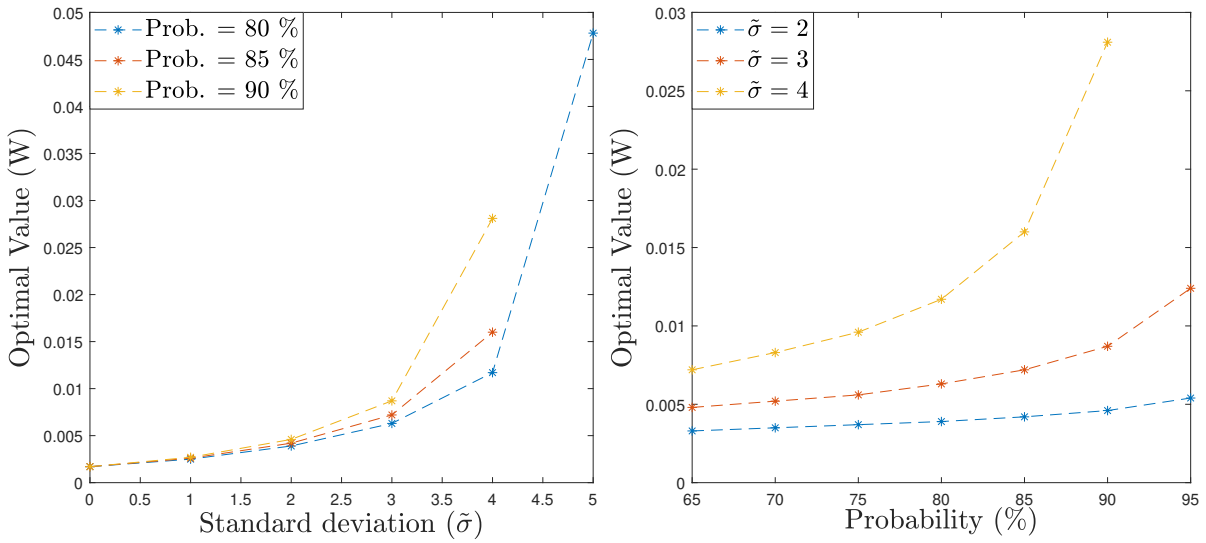
(c)

Figure 5.1: Channel gains (dB) for $\tilde{\sigma} = 3$ and (a) $\tilde{\mu}_{ij}$, (b) $\tilde{\mu}_{ij} - 2\tilde{\sigma}$, and (c) $\tilde{\mu}_{ij} + 2\tilde{\sigma}$.

5.3.3 On the probability of the chance constrained optimization problem: $\sigma$and ρ

The channel gains standard deviation give information on how disperse these values are with respect to the expected ones obtained with procedure in Section 5.3.1. Therefore, increasing $\tilde{\sigma}$ means that the robust approach needs to be solved for worse cases given a fixed probability in the constraint. The impact of this parameter in the optimal value is shown in Figure 5.2.(a) (here we used $\tilde{\sigma}_{ij} = \tilde{\sigma}$ $i \in [n], j \in [N]$). Note that $\tilde{\sigma} = 5$ and probability levels of 85% and 90% lead to infeasible problems.

We can fix the standard deviation of the channel gains and vary the probability level in the constraints by modifying the uncertainty intervals given by $\rho_{ij} \in [\underline{\rho}_{ij}, \overline{\rho}_{ij}]$ and $\rho_{ik} \in [\underline{\rho}_{ik}, \overline{\rho}_{ik}]$. Similarly to the previous case, when we increase these intervals (and the box of uncertainties), the robust optimization problem needs to respect these constraints for worse scenarios, degrading the optimal value as in Figure 5.2.(b). The problem becomes infeasible considering a probability level of 95% and $\tilde{\sigma} = 4$.



(a) (b)

Figure 5.2: The optimal value as a function of (a) standard deviation and (b) constraint probability.

In another experiment, optimization problem (5.6) is solved by considering the uncertainty box intervals

$$[\underline{\rho}_{ij}, \overline{\rho}_{ij}] = [-2.04, \infty), \quad [\underline{\rho}_{ik}, \overline{\rho}_{ik}] = (-\infty, 2.04],$$

and standard deviation $\tilde{\sigma} = 3$ for all variables ρ_{ij} and ρ_{ik}. Therefore $\mathbb{P}(\rho_{ij} > -2.04) =$

0.9793, $\mathbb{P}(\rho_{ik} < 2.04) = 0.9793$, and

$$\prod_{j=1}^{N} \varphi_j = 0.9793^5 = 0.9007.$$

For each UE, 10^5 combination samples with different values for each standard Gaussian variable ρ_{ij} and ρ_{ik} were generated. We then calculated the throughput of the respective UE considering the transmission powers, resource allocation, and association given by optimization problem (5.6) for each combination of ρ_{ij} and ρ_{ik}. Notice in Figure 5.3 that even if around 10% of the samples do not belong to the uncertainty box, there is no constraint violation for many of those scenarios. For instance consider UE 130: 10.088% of the samples combination do not belong to the proposed interval, however in just 1.408% the respective throughput constraint was violated.

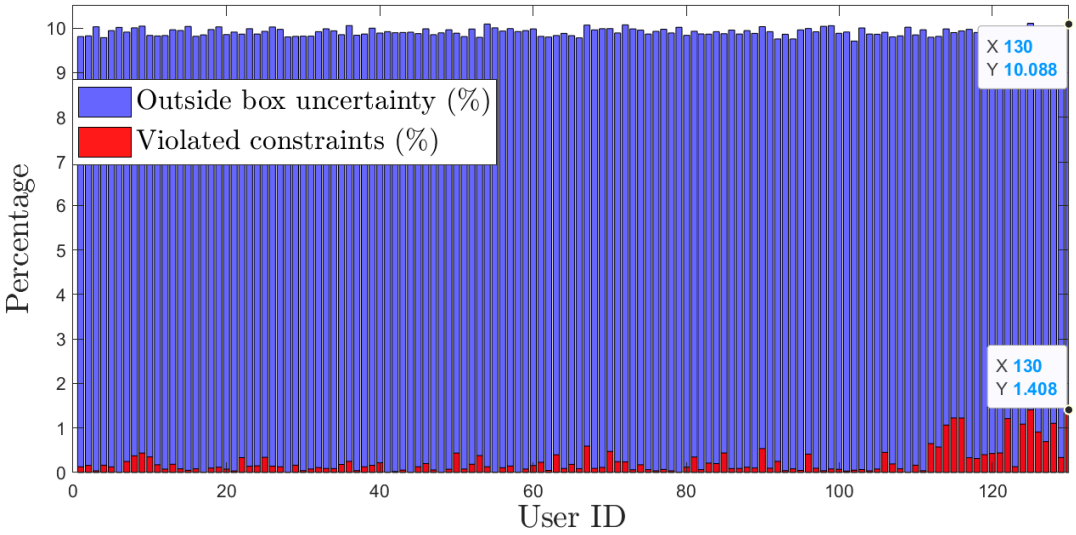


Figure 5.3: Percentage of channel gains combination outside box of uncertainties $i \in [n]$ and violated constraints.

5.3.4 Comparison

Comparisons between the (4.15) (OP-MIGP) and (3.15) (non-convex-OFDMA) were performed in Chapter 4, where the latter one was solved with 1) successive linearization and gradient descent (GD), 2) difference of convex (DC) functions programming, and 3) with the commercial solver MIDACO (Ant Colony Optimization). In all cases, the GP based optimization problem provides remarkably greater optimal value with substantially smaller solving time. Therefore here we perform comparisons only between the deterministic and stochastic approaches, namely (OP-MIGP) and (Robust-MIGP) respectively. We solved both problems and tested the optimal solutions in 10^5 different channel gains combinations (with log-normal distribution and a given σ) for each user. The percentage

of times that the throughput constraints were violated is presented in Table 5.1. Note that, despite providing a smaller optimal value $\sum_j P_j = 0.0017$ W, the deterministic approach fails to respect users demands in many more situations, since they were designed just considering the channel gains expected values, i.e. $\sigma = 0$.

Table 5.1: Percentage of violated constraints.

σ	(Robust-MIGP)	$\sum_j P_j$ (W)	(OP-MIGP)
2	0.29 %	0.0046	29.40 %
3	0.24 %	0.0087	34.93 %
4	0.12 %	0.0281	38.40 %

5.3.5 Robust MIGP and different channel gains distribution

Even if the Robust MIGP assumes log-normal distribution for the channel gains, it is able to respect the chance-constrained optimization problem in different scenarios. Notice in Table 5.2 that the proposed approach can easily guarantee that at least 90% of the users' throughput demands are respected when the channel gains combinations come from a log-normal distribution. This is expected since the problem is formulated under that assumption. If the optimization problem is solved considering log-normal distribution, but the channel gains respect a uniform one, then the 90% of the users' throughput demands are respected for the cases where the channel gains belong to the following intervals: $[\tilde{\mu}-\tilde{\sigma}, \tilde{\mu}+\tilde{\sigma}]$, $[\tilde{\mu}-2\tilde{\sigma}, \tilde{\mu}+2\tilde{\sigma}]$, and $[\tilde{\mu}-3\tilde{\sigma}, \tilde{\mu}+3\tilde{\sigma}]$. Considering a Student's t distribution with degree of freedom $dof = 2$, the 10% violation goal was also respected. As $Dof \to \infty$, the Student's t distribution tends to the Gaussian one (log-normal channel gains, which are Gaussian considering the respective values in dB).

Table 5.2: Percentage of unsatisfied users' throughput requirements.

$\tilde{\sigma}$	Log-Normal	Uniform $[\tilde{\mu}-3\tilde{\sigma}, \tilde{\mu}+3\tilde{\sigma}]$	Student's t Distribution $Dof = 2$	Optimal Value
2	0.29%	3.86 %	6.52 %	0.0046
3	0.24%	4.20 %	7.82 %	0.0087
4	0.12%	3.74 %	8.53 %	0.0281

To achieve the 90% goal even for the case of uniform distribution and $[\tilde{\mu}-4\tilde{\sigma}, \tilde{\mu}+4\tilde{\sigma}]$, one could increase the uncertainty box, according to

$$[\underline{\rho}_{ij}, \overline{\rho}_{ij}] = [-2.25, \infty), \quad [\underline{\rho}_{ik}, \overline{\rho}_{ik}] = (-\infty, 2.25],$$

with the cost of increasing the optimal value, as presented in Table 5.3.

Table 5.3: Percentage of unsatisfied users' throughput requirements under uniform distribution.

	Uncertainty box: 90%		**Uncertainty box: 94%**	
$\tilde{\sigma}$	$[\tilde{\mu} - 4\tilde{\sigma}, \tilde{\mu} + 4\tilde{\sigma}]$	$\sum_j P_j$ (W)	$[\tilde{\mu} - 4\tilde{\sigma}, \tilde{\mu} + 4\tilde{\sigma}]$	$\sum_j P_j(\tilde{\sigma} = 0)$
2	12.91 %	0.0046	9.88 %	0.0052
3	14.34 %	0.0087	10.79 %	0.0113
4	14.21 %	0.0281	9.92 %	0.1403

5.3.6 Robustness with respect to users' traffic demands

To examine the robustness of our approach with respect to the user traffic demands, we solved the problem considering the original throughput levels given by r, $\rho_{ij} = -2.04$, $\rho_{ik} = 2.04$. Subsequently we increased the traffic demands and verify the amount of times that those new constraints were not satisfied (with the optimal solution obtained for the original throughput levels). The results are given below.

Table 5.4: Percentage of unsatisfied users' throughput requirements.

$\tilde{\sigma}$	r	1.1 r	1.2 r	1.4 r	1.5 r	2.3 r
2	0.29%	0.70 %	1.73 %	7.18 %	11.33 %	40.00 %
3	0.24%	0.43 %	0.77 %	2.44 %	3.96 %	23.09 %
4	0.12%	0.19 %	0.29 %	0.72 %	1.13 %	10.20 %

It is clear that for larger values of standard deviation $\tilde{\sigma}$, for instance $\tilde{\sigma} = 4$, the uncertainty box increases leading to an optimal solution that is able to respect the probabilistic constraint even when the users throughput are up to about 130% higher than the ones used in the optimization problem. From one side, this shows that our approach is able to deal with fast and large variations in the throughput demands; however one may notice that the approaches gets more conservative when the channel gains uncertainty increases.

5.4 Final considerations

The minimization of base stations transmission powers in OFDMA networks, subject to individual UEs throughput constraints is studied in this chapter. The channel gains between UE and base stations are considered to be random variables with log-normal distribution. By using change of variables with a suitable power function approximation, the highly non-convex chance constrained optimization problem is formulated as a robust mixed-integer Geometric Program. The proposed approach was tested in a realistic scenario: a neighbourhood from a large European city was reconstructed into a ray tracing

propagation solver (considering building height, city layout, BSs location, etc), so the expected values for the channel gains could be estimated. The results show that considering the channel gains as deterministic might lead to considerable amount of constraints violation.

Chapter 6

MPC Control of Cellular Network Transmission Powers and Resource Allocation over Failure Environments

Different from the previous Chapters, now we aim to control the transmission powers (rather then only minimize them) by considering that the users are moving through the network, their trajectory can be predicted [42,43], and one of the BS is abruptly switched off. As we reconstructed a neighbourhood of a large European city into a ray tracing simulator, it is possible to estimate the users channel gains according to their foreseen position and apply the Model Predictive Control strategy (MPC). MPC fits our problem requirements due to two main reasons: the change in the transmission power levels between consecutive steps might have a physical limit and, since we can predict the users channel gains, MPC allows us to change P_j $j \in [N]$ by considering these quantities. Additionally, at each iteration, any trajectory prediction or channel gain errors are attenuated since the real-world values can be measured.

6.1 Problem extension

Since we are dealing with time-varying network scenarios, optimization problem (4.15) decision variables and channel gains are, from now on, referred as $q_j(t)$, $u_{ij}(t)$, $z_{ij}(t)$, and $g_{ij}(t)$, respectively.

6.1.1 Initial condition and time-varying reference

The network initial condition is given by the power and resource allocations decision variables $q_j(t)$, $u_{ij}(t)$ and $z_{ij}(t)$, respectively. They are obtained by solving the optimization problem (4.15) for $t = 0$ and considering that the cellular network has N active BSs. Subsequently, we aim to solve a problem where a BS is switched off, as in a network

failure, and just $N-1$ ones are able to provide the UEs connections. In such a scenario, the remaining operational BSs might not have enough resources to allocate all the UEs and respect their throughput constraints. Therefore, we introduce slack variables λ_i to the optimization (4.15), leading to

$$\begin{aligned}
\min_{\substack{\bar{z}_{ij}(t),u_{ij}(t),\\ q_j(t),\lambda_i(t)}} \quad & \omega_1 \sum_{j=1}^{N-1} e^{q_j(t)} + \omega_2 \sum_{i=1}^{n} \lambda_i(t) + \omega_3 ||\bar{\mathcal{Y}}(t)||_\infty \\
\text{s.t.:} & \\
1) \quad & \bar{z}_{ij}(t) \in \{0,1\}, \quad i \in [n], \quad j \in [N], \\
2) \quad & \sum_{j=1}^{N} \bar{z}_{ij}(t) = N-1, \quad i \in [n], \\
3) \quad & \sum_{i=1}^{n} e^{u_{ij}(t)} \leq 1, \quad j \in [N], \\
4) \quad & \hat{f}_\ell(q(t),u_{ij}(t)) \leq \frac{\log\left(\frac{B_j a_\ell}{t_i(t)}\right)}{b_\ell} + \lambda_i(t) + M\bar{z}_{ij}(t), \\
& \qquad \ell \in [m], i \in [n], j \in [N], \\
5) \quad & \lambda_i(t) \geq 0, \quad i \in [n].
\end{aligned} \tag{6.1}$$

$\bar{\mathcal{Y}}(t)$ is a vector containing the number of UEs connected to each one of the BSs at time t. ω_1, ω_2, and ω_3 strongly impact the optimal solution and might lead to different UEs association and resource allocation, as discussed below.

- ω_1: this parameter concerns the cost given to the expenditure of the transmission powers. Therefore, high values of ω_1 leads to an approach that is less power consuming, however might violate constraints when $\omega_1 \gg \omega_2$.

- ω_2: weight parameter associated to the constraints violation. Increasing such value leads to approaches that prioritize respecting the constraints rather than minimizing the transmission powers. With an appropriate value, $\lambda_i(t) \to 0$ when the respective user's throughput constraint can be respected.

- ω_3: weight parameter responsible for the cellular network balancing. Appropriate values for ω_3 lead to solutions that do not associate many users to a single BS.

The optimal solution obtained with (6.1) can be strongly different from the one when all N BSs work properly, both in terms of transmission power levels and association between UEs and BSs. In real life cellular networks, there can be a physical limit between

two consecutive transmission power levels, that is

$$\beta P_j(t-1) \leq P_j(t) \leq \alpha P(t-1), \quad j \in [N], \tag{6.2}$$

and since $P_j(t) = e^{q_j}(t)$

$$\log \beta \leq q(t) - q(t-1) \leq \log \alpha. \tag{6.3}$$

Furthermore, a physical limit on how many new users can be associated to a BS at a time t given by

$$\sum_{i}^{n} z_{ij}(t) - \sum_{i}^{n} z_{ij}(t-1) \leq \xi, \quad j \in [N-1] \tag{6.4}$$

can also be observed. Consequently, the solution given by (6.1) at a time t might not be applicable to real-world scenarios if the aforementioned physical limits are not considered.

6.1.2 The MPC

As further addressed, a methodology to predict the UEs next positions inside a prediction horizon T is used. By doing so, the channel gains $g_{ij}(t+1),\ldots, g_{ij}(t+T)$, $i \in [n]$, $j \in [N]$ in the corresponding foreseen positions can be estimated. Then, we propose to solve optimization problem (6.1) considering the channel gains at time $(t+T)$, and the optimal transmission power levels $q_j^\star$ become the output reference for an MPC that considers not only the constraints in (6.1), but also the physical limitations previously presented at all time steps inside the T interval. Algorithm 3 describes this process.

Note that ω_5 and ω_6 are equivalent to ω_2 and ω_3, respectively. ω_4 is the weight parameter associated to transmission powers reference tracking. Increasing such value will lead to a solution that prioritizes the power level control rather than respecting the throughput constraints or the network balancing.

6.1.3 Trajectory prediction

At an instant t, the blue box in Algorithm 3 is solved by considering that the users' trajectories are predicted for the next T time steps. Therefore, the GP-MPC is solved with the estimated channel gains for the respective foreseen locations. For any UE, its next position is given by a small displacement in the direction of the line passing through the two previous ones if the predicted position has at least one BS providing a channel gain greater than -110 dB, that is

$$C_x(t+1) = C_x(t) + [C_x(t) - C_x(t-1)],$$
$$C_y(t+1) = C_y(t) + [C_y(t) - C_y(t-1)],$$

with C_x and C_y being the respective geographic coordinates (longitude and latitude). If such a condition is not satisfied (usually indicating indoor environments), the UE foreseen

Algorithm 3: GP-MPC

Solve OP (6.1) **at** $t = 0$ **with real UEs positions:**

$q_j(t) = q_j(0),\ \ j \in [N]$
$u_{ij}(t) = u_{ij}(0),\ i \in [n],\ j \in [N]$
$z_{ij}(t) = z_{ij}(0),\ i \in [n],\ j \in [N]$

1 BS switched off

while *True* **do**

Predict UE next T **positions:**

$g_{ij}(t+1), \ldots, g_{ij}(t+T),\ i \in [n],\ j \in [N-1]$

Solve OP (6.1) **with predicted positions at** $t+T$**:**

$q_j^\star(t) = q_j(t+T),\ \ j \in [N-1]$

GP-MPC:

$$\min_{\bar{z}_{ij}(\delta), u_{ij}(\delta), q_j(\delta), \lambda_i(\delta)} \sum_{\delta=t+1}^{t+T} \omega_4 \sum_{j=1}^{N-1} (q_j(\delta) - q_j^\star(t)) + \omega_5 \sum_{i=1}^{n} \lambda_i(\delta) + \omega_6 ||\bar{\mathcal{Y}}(\delta)||_\infty$$

subject to:

for $\delta = (t+1) : (t+T)$ **do**

Constraints 1, 2, 3, 4, and 5 in (6.1)

$\log \beta \leq q_j(\delta) - q_j(\delta - 1) \leq \log \alpha,\ \ j \in [N-1]$

$\sum_i^n z_{ij}(\delta) - \sum_i^n z_{ij}(\delta - 1) \leq \xi,\ \ j \in [N-1]$

end

Update network:

$q_j(t) = q_j(t+1), \quad j \in [N]$
$u_{ij}(t) = u_{ij}(t+1),\ i \in [n],\ j \in [N]$
$z_{ij}(t) = z_{ij}(t+1),\ i \in [n],\ j \in [N]$

$t = t + 1$

end

location is considered to be the same of the previous time for all remaining steps in the prediction horizon T. These assumptions allow the MPC controller to smoothly adjust the BSs transmission powers even when the users change their walking direction, since such a dynamic is captured by the aforementioned model.

6.2 Example

The efficacy of the proposed approach is shown in the same real-world dataset from the previous Chapters. This information, combined with the network parameters in Table 6.1, allows us to estimate the channel gains for each user inside the studied environment with respect to all BSs before and after one BS is switched off, as presented in Figure 6.1.

Table 6.1: Network parameters.

Network	Value
Technology	5G NR
Transmission technique	OFDMA
RB per BS N_{RB}	500
Number of users n	100
Number of base stations N	5
Noise power η^2	-174 dbm/Hz
Carrier Wave frequency	3.5 GHz
BS Bandwidth B_j	100 MHz
Throughput $[\min t_i, \max t_i]$	[3.80 $\times 10^3$, 2.50 $\times 10^6$] bits/s

The scenario consists of 100 UEs placed in different areas of the neighbourhood. Their trajectories have 22 samples, i.e, for each UE 22 consecutive locations were generated with the restriction that all users must remain outside the dark blue areas in Figure 6.1 (indoor buildings), since these places have severe signal attenuation. The BSs location, the UEs initial position, and the trajectories of some UEs are shown in Figure 6.2. The users' throughput requirement $\hat{t}_i$ generation is based on [44].

6.2.1 The network output reference optimization

As previously shown in Algorithm 3, the network initial condition ($t = 0$) is calculated with optimization problem (6.1). Note that when $t = 0$, all BSs work properly and the optimal solution is obtained with the current UEs locations and their respective channel gains. The reference transmission powers are also obtained with optimization problem (6.1), however considering that a BS is switched off and the users' channel gains are given at the predicted position ($t + T$). The simulation hyper-parameters are given in Table 6.2.

The reference transmission powers q_j, the number of users connected to each BS, and the network total throughput violation, defined as

$$-\sum_{i=1}^{n} \min(t_i - \hat{t}_i, 0), \tag{6.5}$$

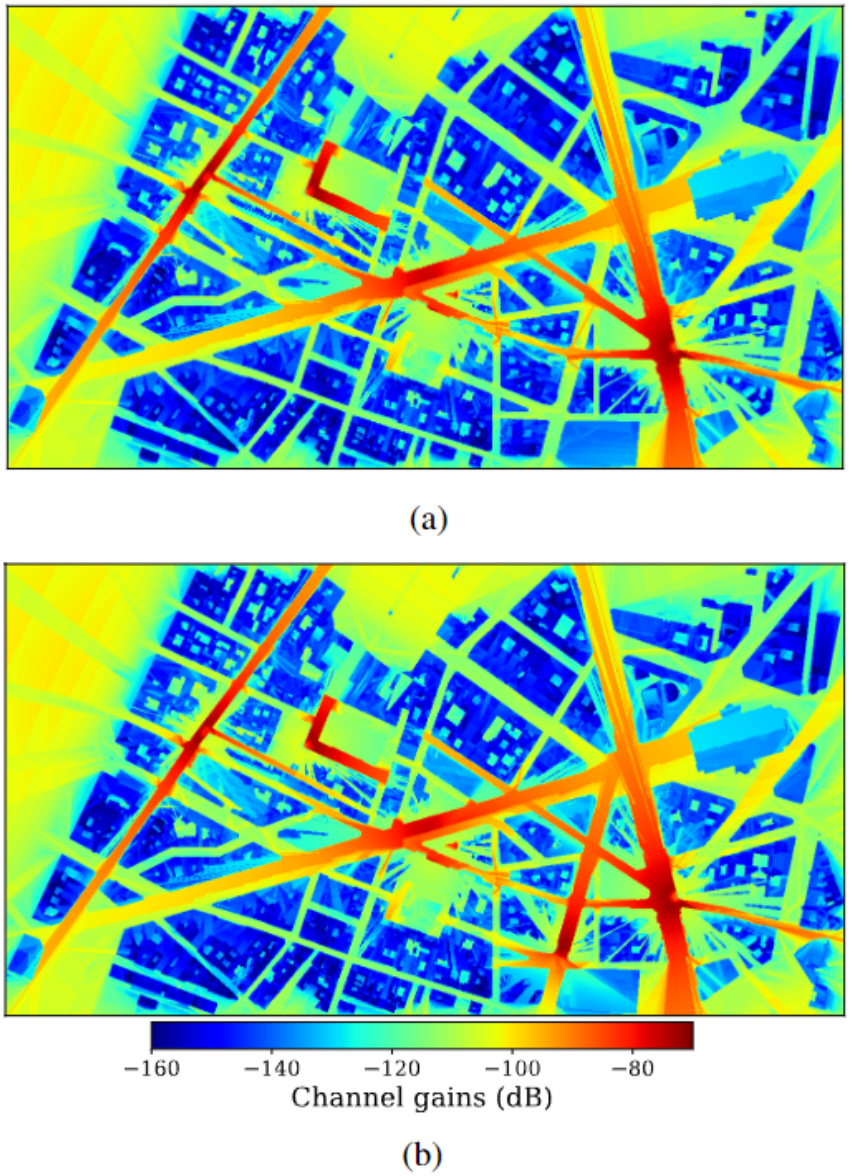


Figure 6.1: Channel gains (dB scale) for the European neighbourhood. (a) $N-1$ BSs and (b) N BSs.

Table 6.2: Initial condition and reference calculation hyper-parameters.

Hyper-parameter	Value
T	4
ω_1	1
ω_2	1
ω_3	10^{-5}
a_d	[0.7330 1.3150 1.9061 3.1232 4.3638]
b_d	[0.7821 0.4201 0.2589 0.1697 0.1161]
M	10^6

are shown in Figure 6.3.

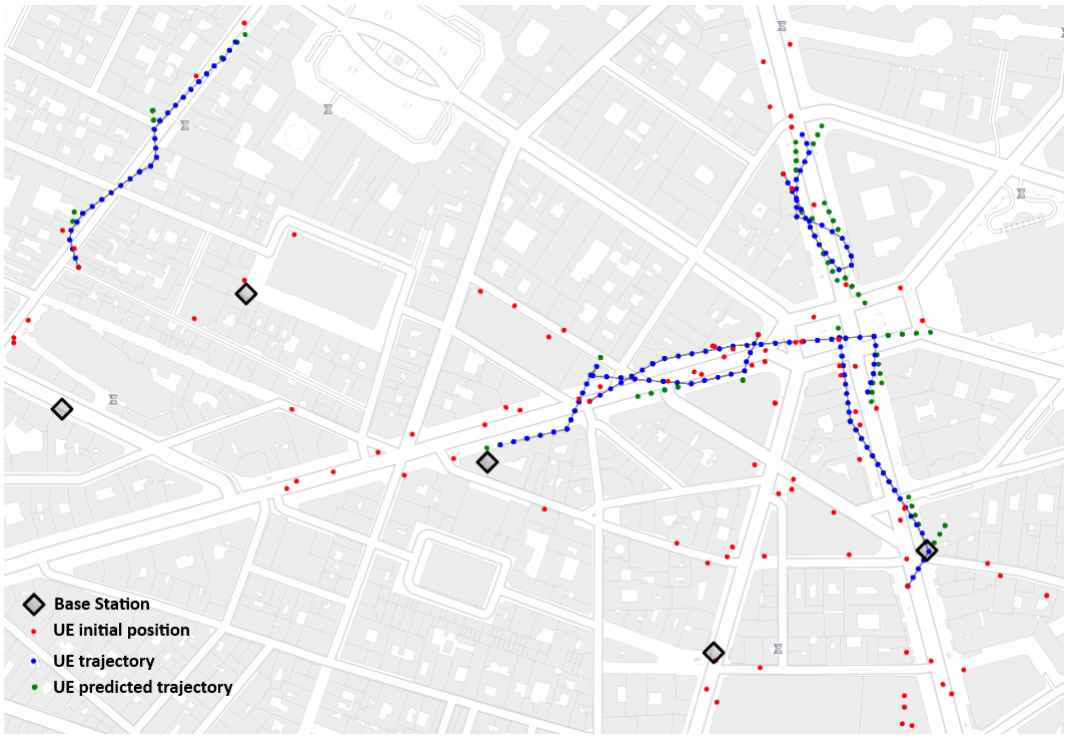


Figure 6.2: UEs location/trajectory and BSs deployed in an European neighbourhood.

6.2.2 The network control with MPC

Given the reference BSs transmission powers at a time $(t+T)$ and the hyper-parameters in 6.3, we use the GP-MPC control strategy to smoothly take the respective $q_j(t+T)$ to the reference level.

Table 6.3: GP-MPC hyper-parameters.

Hyper-parameter	Value
ω_4	1
ω_5	1
ω_6	10^{-5}
β	0.2
α	5
ξ	25

As presented in Algorithm 3 and as an MPC standard, the network manipulated variables are updated just with the ones at time $(t+1)$, i.e., $q_j(t+1)$, $u_{ij}(t+1)$, and $z_{ij}(t+1)$. The subsequent ones are discarded due to possible trajectory prediction or channel gains estimation errors. The reference and real transmission powers applied to the network are shown in Figure 6.4.

Note that in all time-steps of the simulation, there is an error associated to the reference powers and the ones applied to the network. The main reason is based on the fact that the

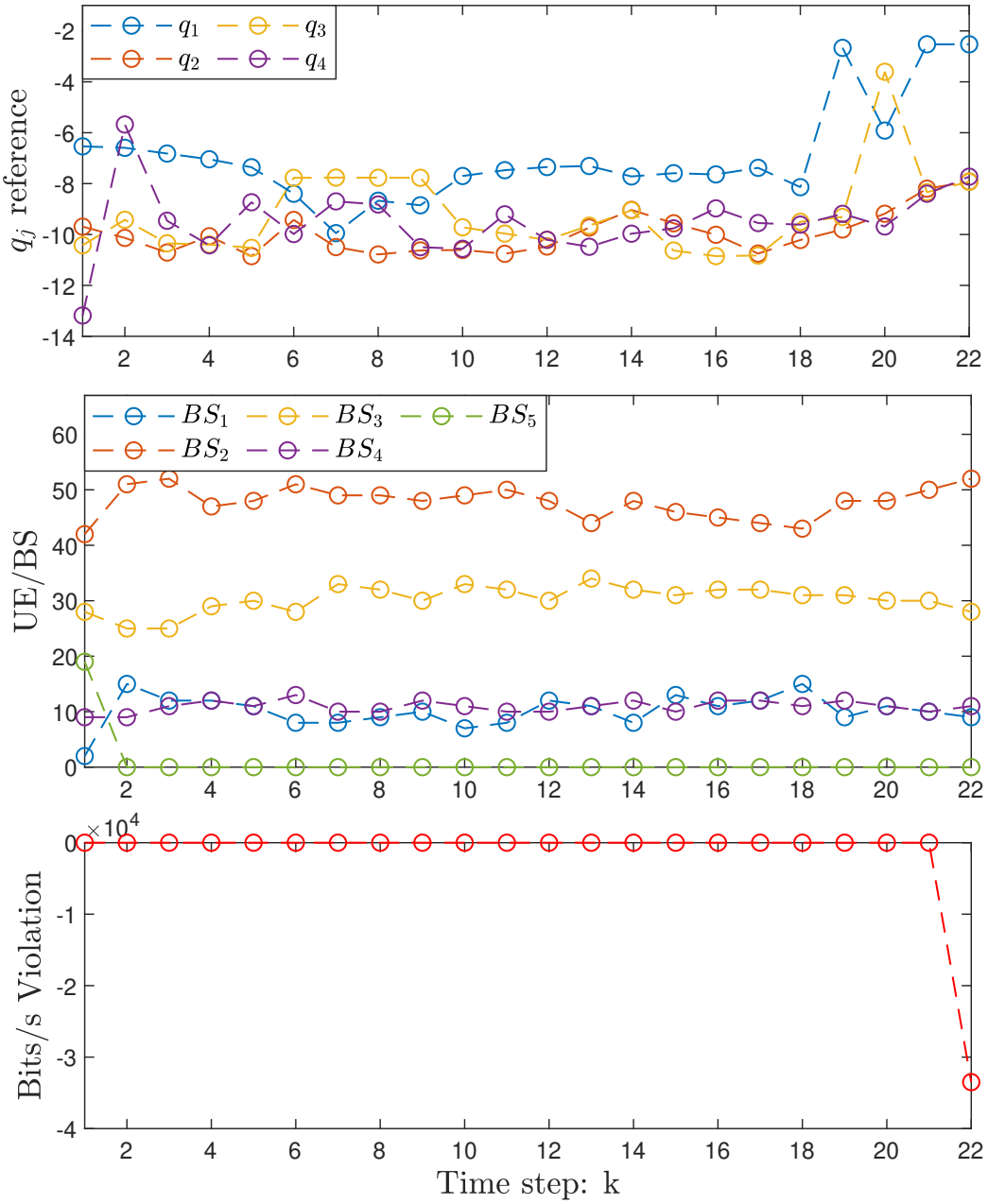


Figure 6.3: BSs reference transmission powers, consequently UE/BSs association, and total constraint violation.

cost function to be minimized by the GP-MPC depends not only on the power reference tracking, but also on the throughput constraints violation and on the network balance. Therefore, one could decrease the power tracking error by increasing the value of ω_4 relatively to ω_5 and ω_6, which could degrade the network balance or constraints violation. The network total throughput violation and the number of users associated to each BS can be seen in Figure 6.5.

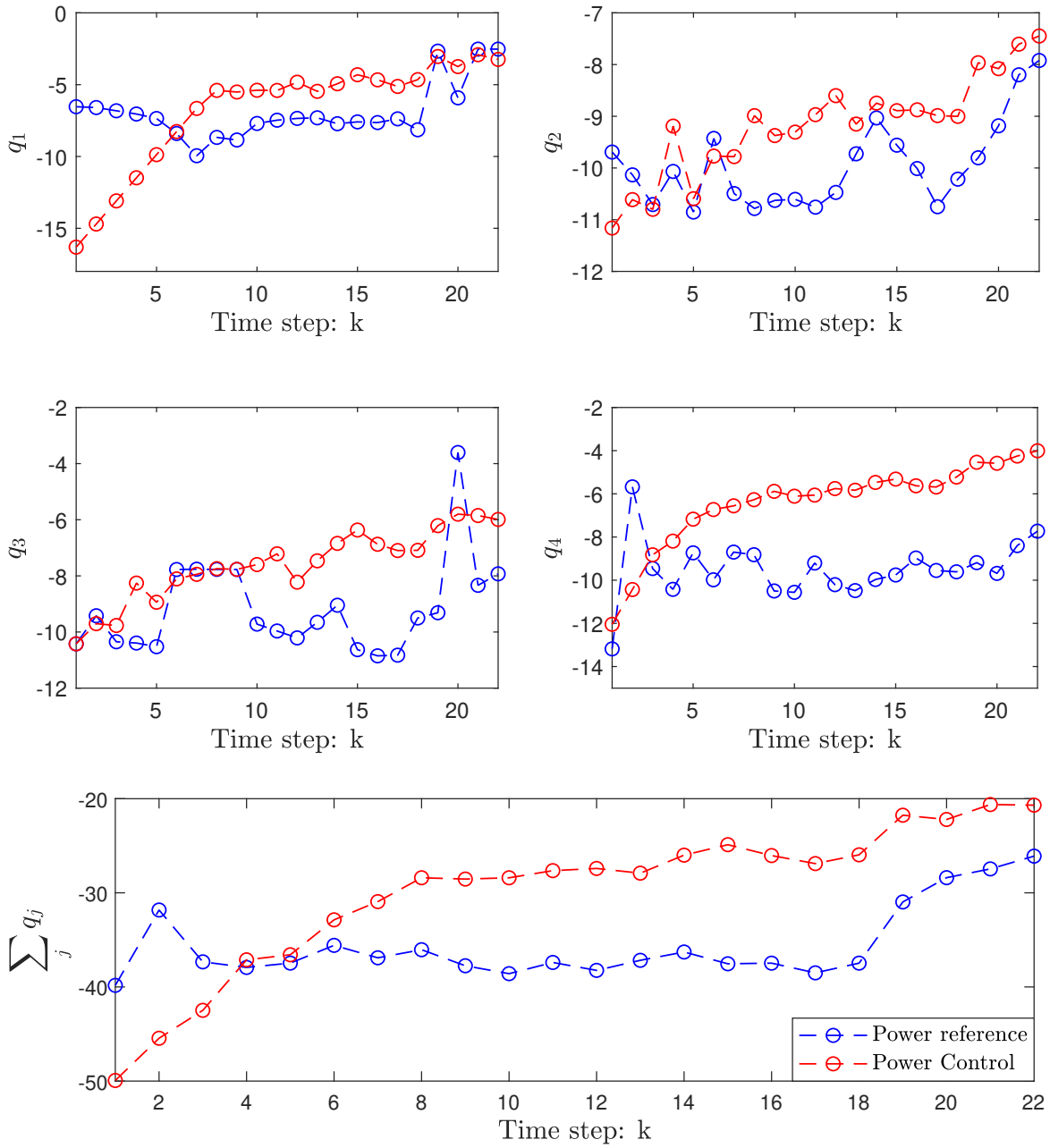


Figure 6.4: Transmission powers reference and respective ones applied to the BSs in the network.

6.3 Final considerations

This Chapter introduces two optimization problems applied to OFDMA heterogeneous cellular networks. The first one is used to minimize the base stations transmission powers while allocating the network resources and respecting individual UE throughput

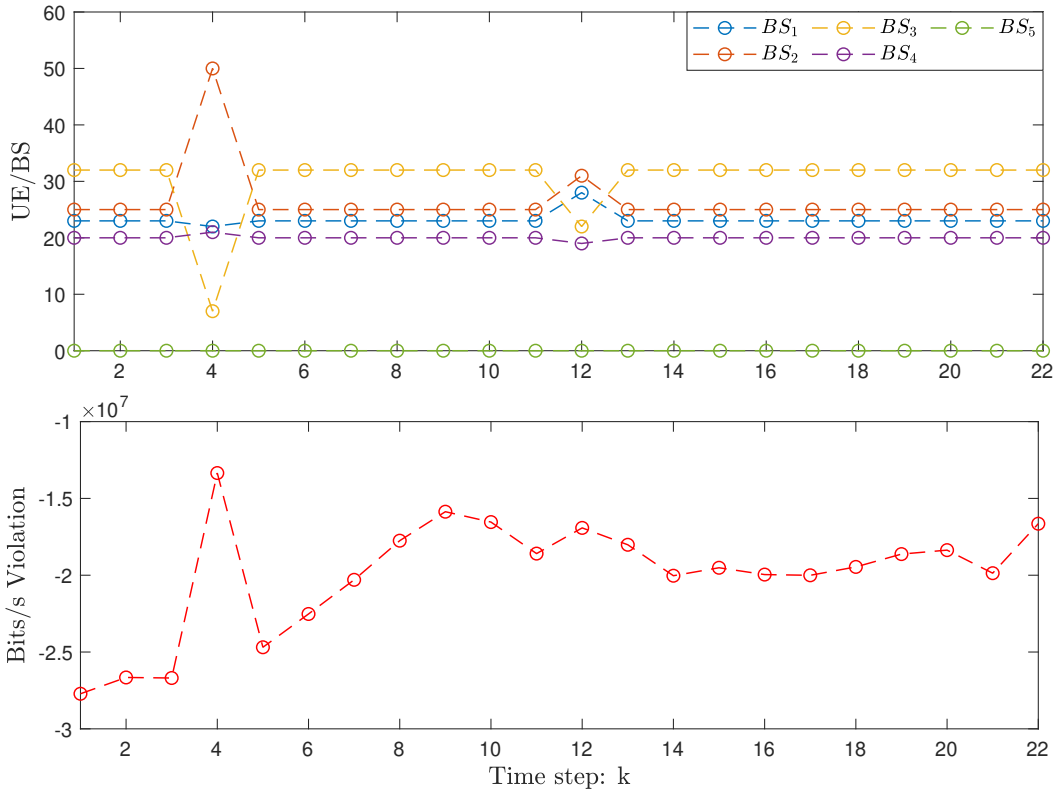


Figure 6.5: Number of users associated to each BS and network total throughput violation.

constraints. The latter one makes use of Model Predictive Control and UE trajectory prediction to control the network transmission powers and resource allocation when one of the BSs is abruptly switched off, as in failure environments. Both optimization problems are mixed-integer Geometric Programs (mixed-integer convex) and use a piecewise concave approximation of the results of the Shannon-Hartley theorem. We reconstructed a neighbourhood of a large European city in a high performance propagation solver and used real-world BSs locations to test the efficacy of our approach in a realistic scenario.

Chapter 7

Conclusion

In this work, a novel approach to minimize the transmission power of base stations (downlink) in OFDMA heterogeneous networks is proposed. The new technique is based on piecewise power function approximations followed by change of variables, leading to a class of convex optimization problems known as *Geometric Programming*. The global optimal solution for GPs can always be found (for feasible problems) and they can provide fast solutions when solved with interior-points method.

The optimization problem jointly solves the problems of user association, bandwidth allocation, and transmission power minimization. Among the constraints, one may cite the individual users' throughput requirements (bits/s), i.e the proposed solution is able to provide the proper amount of capacity for each user individually. The efficacy of the technique was tested in highly realist scenarios of a large European city, where the channel gains between each base station and user equipment are calculated with a high-performance propagation solver, which was able to emulate the electromagnetic waves propagated by each base station from a real-world neighbourhood. The results show not only substantially smaller optimal values, but also considerably smaller runtime when compared to successive linearization approaches, DC programming, and MIDACO (commercial software for global optimization).

Some works in the literature empirically show that, due to blockage caused by objects in motion (cars, people, etc), the channel-gains are log-normal stochastic variables both outdoor and indoor environments. Therefore, the work previously mentioned was extended to deal with the resulting chance constrained optimization problem. In summary, a robust optimization problem which is able to respect the users' throughput constraints inside a probability level is then proposed.

Lastly, we apply MPC to control the base stations transmission powers when there is a network failure (one base station abruptly is switched off). In such a case, it is not possible to respect all users constraints simultaneously, however the controller, based on users' trajectory prediction, is able to minimize a weighted objective function depending

on base stations transmission power and constraints violation level simultaneously.

7.1 Future Works

Despite the fact the conditions developed present promising results, some topics could be further investigated, such as:

1. Extend the work to allow the users connection to more than one base station/antenna. Since the SINR is also critical in such a context, the approach proposed in this thesis- piecewise power function approximation followed by change of variables and logarithmic transformation (leading to a GP)- will also be useful in that extension. This technology (MIMO- Multiple Input Multiple Output) is extremely related to the development of 5G and beyond wireless communication systems. In massive MIMO, every base station is equipped with a large number of antennas aiming at optimized spectral and energy efficiency. Massive MIMO also significantly improves the network throughput and coverage, playing a crucial role in meeting the ever growing data demand from mobile devices [45]. Therefore, the extension of the work presented in this thesis to the massive MIMO technology is of great relevance.

2. Extend the heterogeneous network scenario to the case where the users are moving through the neighborhood, and their trajectory include a more realistic model for trajectory prediction. In this context, the spectral and energy efficiency can be improved by predicting not only areas with more traffic demand, but also the time of such peaks. Spatiotemporal information can be extracted from mobile devices data, such as: human mobility patterns, urban hot-spot detection, work or home location [46]. By combining these insights with time-series prediction models (as NN based), one could create extremely realistic urban environments and optimize the network power consumption with the proposed MIGP technique. Additionally, the human transportation modes and mobility could be simulated with SUMO (Simulation of Urban MObility). In such dynamic scenarios, the challenge of users' handovers will be a critical consideration to address.

3. Since the channel gains between UE/BS are, in practice, random quantities, the problem to be solved becomes a chance-constrained optimization problem. We proposed to circumvent that issue by relaxing it to a robust version, where a box of uncertainties is designed. Many different sets can be constructed, including even non-convex ones, which could lead to solutions with smaller objective values when compared to the box version. Therefore, the different sets where the channel gains

could lie should be further investigated. Specifically, data from channel gains measurements in a given location could be used to statistically draw these sets, and subsequently approximate them with convex representations.

4. Another possible work extension is to include Reinforcement Learning (RL) when optimizing the scenario where the users are moving through the neighborhood. By using the feedback policy to make decisions of resource allocation (base station assignment, bandwidth allocation), the optimization model could identify points of signal shadowing and properly allocate resources to users at those specific locations. Since the channel gains are constantly changing, the Reinforcement Learning strategy could even allow us to calculate different sets of uncertainty for them according to the environment seasonality (the shadowing during rush hours could lead to higher standard deviation to the gains due to blockage effects). Additionally, RL can also be applied to the failure environment scenario due to its capability of learning from the new environment.

Appendix A

From non-convex to a GP representation

At a given interval, the following inequality is valid for an user i

$$B_j x_{ij} \log_2\left(1+\frac{g_{ij}P_j}{\sum_{k\neq j} g_{ik}P_k+\eta^2}\right) \geq B_j x_{ij} a_\ell \left(\frac{g_{ij}P_j}{\sum_{k\neq j} g_{ik}P_k+\eta^2}\right)^{b_\ell} \geq t_i,$$

$$a_\ell (x_{ij}^{\frac{1}{b_\ell}})^{b_\ell} \left(\frac{g_{ij}P_j}{\sum_{k\neq j} g_{ik}P_k+\eta^2}\right)^{b_\ell} = a_\ell \left(\frac{g_{ij}P_j x_{ij}^{\frac{1}{b_\ell}}}{\sum_{k\neq j} g_{ik}P_k+\eta^2}\right)^{b_\ell} \geq \frac{t_i}{B_j},$$

$$a_\ell \left[\left(\sum_{k\neq j} g_{ik}P_k\right) g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}} + \eta^2 g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}})^{-1}\right]^{b_\ell} \geq \frac{t_i}{B_j},$$

$$\log\left\{a_\ell \left[\left(\sum_{k\neq j} g_{ik}P_k\right) g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}} + \eta^2 g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}}\right]^{-b_\ell}\right\} \geq \log\left(\frac{t_i}{B_j}\right),$$

$$\log(a_\ell) + \log\left\{\left[\left(\sum_{k\neq j} g_{ik}P_k\right) g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}} + \eta^2 g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}}\right]^{-b_\ell}\right\} \geq \log\left(\frac{t_i}{B_j}\right),$$

$$-b_\ell \log\left[\left(\sum_{k\neq j} g_{ik}P_k\right) g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}} + \eta^2 g_{ij}^{-1}P_j^{-1}x_{ij}^{-\frac{1}{b_\ell}}\right] \geq \log\left(\frac{t_i}{B_j}\right) - \log(a_\ell).$$

With the variables transformation $P_j = e^{q_j}$, $P_k = e^{q_k}$, and $x_{ij} = e^{u_{ij}}$:

$$-b_\ell \log\left(\sum_{k\neq j} \frac{g_{ik}}{g_{ij}} e^{q_k} e^{-q_j} e^{-\frac{u_{ij}}{b_\ell}} + \frac{\eta^2}{g_{ij}} e^{-q_j} e^{-\frac{u_{ij}}{b_\ell}}\right) \geq \log\left(\frac{t_i}{B_j}\right) - \log(a_\ell),$$

$$-b_\ell \log\left(\sum_{k\neq j} \frac{g_{ik}}{g_{ij}} e^{q_k - q_j - \frac{u_{ij}}{b_\ell}} + \frac{\eta^2}{g_{ij}} e^{-q_j - \frac{u_{ij}}{b_\ell}}\right) \geq \log\left(\frac{t_i}{B_j}\right) - \log(a_\ell).$$

Finally, we have a **log-sum-exp** expression, which is convex:

$$\log\left(\sum_{k\neq j} \frac{g_{ik}}{g_{ij}} e^{q_k - q_j - \frac{u_{ij}}{b_\ell}} + \frac{\eta^2}{g_{ij}} e^{-q_j - \frac{u_{ij}}{b_\ell}}\right) \leq \frac{\log(a_\ell) - \log\left(\frac{t_i}{B_j}\right)}{b_\ell}. \tag{A.1}$$

Note that for the linear case, one just needs to certify that $b_\ell = 1$.

References

[1] C. J. Bernardos and M. A. Uusitalo, "European vision for the 6G network ecosystem," june 2021.

[2] R. Borralho, A. Mohamed, A. U. Quddus, P. Vieira, and R. Tafazolli, "A survey on coverage enhancement in cellular networks: Challenges and solutions for future deployments," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1302–1341, 2021.

[3] D. López-Pérez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, "A survey on 5G radio access network energy efficiency: Massive mimo, lean carrier design, sleep modes, and machine learning," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 653–697, 2022.

[4] A. De Domenico, E. Calvanese Strinati, and A. Capone, "Enabling green cellular networks: A survey and outlook," *Computer Communications*, vol. 37, pp. 5–24, 2014.

[5] Ericsson, "Ericsson mobility report: Q4 2020 update." Available at: `https://ict.moscow/en/research/ericsson-mobility-report-q4-2020-update/`, 2020. Acessed: July 26, 2023.

[6] A. Turner, "Over 90% of operators concerned about rising energy costs for 5G and edge." Available at: https://www.mobileeurope.co.uk/more-than-90-of-operators-concerned-about-rising-energy-costs-for-5g-and-edge/, 2019. Acessed: July 26, 2023.

[7] Nokia, "How artificial intelligence reduces the carbon footprint of telco networks." Available at: `https://www.nokia.com/networks/bss-oss/ava/energy-efficiency/`, Acessed: July 26, 2023.

[8] 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, 2013.

[9] F. Wang, W. Chen, H. Tang, and Q. Wu, "Joint optimization of user association, subchannel allocation, and power allocation in multi-cell multi-association OFDMA heterogeneous networks," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2672–2684, 2017.

[10] N. Sapountzis, T. Spyropoulos, N. Nikaein, and U. Salim, "Optimal downlink and uplink user association in backhaul-limited HetNets," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, 2016.

[11] N.-T. Le, L.-N. Tran, Q.-D. Vu, and D. Jayalath, "Energy-efficient resource allocation for OFDMA heterogeneous networks," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7043–7057, 2019.

[12] J.-S. Liu, C.-H. R. Lin, and Y.-C. Hu, "Joint resource allocation, user association, and power control for 5G LTE-based heterogeneous networks," *IEEE Access*, vol. 8, pp. 122654–122672, 2020.

[13] W. Bao and B. Liang, "Structured spectrum allocation and user association in heterogeneous cellular networks," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 1069–1077, 2014.

[14] W. Bao and B. Liang., "Radio resource allocation in heterogeneous wireless networks: A spatial-temporal perspective," in *IEEE INFOCOM 2015 - IEEE Conference on Computer Communications*, pp. 334–342, 2015.

[15] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 668–695, 2021.

[16] G. Sun, X. Wang, R. Jiang, and Y. Xu, "Beamforming and resource allocation in multi-cell OFDMA systems based on deep transfer reinforcement learning," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pp. 1–6, 2022.

[17] C. Pan, R. Liu, and G. Yu, "Joint User Association and Resource Allocation for mmWave Communication: A Neural Network Approach," *Journal of Communications and Information Networks*, vol. 6, no. 2, pp. 125–133, 2021.

[18] M. Johansson, L. Xiao, and S. Boyd, "Optimal Routing and SINR Target Selection for Power-Controlled CDMA Wireless Networks." WiOpt'03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, Mar. 2003. Poster.

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

[20] G. O. Ferreira, C. Ravazzi, F. Dabbene, G. C. Calafiore, and M. Fiore, "Forecasting network traffic: A survey and tutorial with open-source comparative evaluation," *IEEE Access*, vol. 11, pp. 6018–6044, 2023.

[21] G. O. Ferreira, A. Felipe Zanella, S. Bakirtzis, C. Ravazzi, F. Dabbene, G. C. Calafiore, I. Wassell, J. Zhang, and M. Fiore, "A Joint Optimization Approach for Power-Efficient Heterogeneous OFDMA Radio Access Networks," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 11, pp. 3232–3245, 2024.

[22] G. O. Ferreira, C. Ravazzi, F. Dabbene, and G. C. Calafiore, "Power minimization and resource allocation in HetNets with uncertain channel gains," *IEEE Communications Letters*, pp. 1–1, 2024.

[23] A. Goldsmith, *Overview of Wireless Communications*, p. 1–26. Cambridge University Press, 2005.

[24] A. Goldsmith, *Multiuser Systems*, p. 452–504. Cambridge University Press, 2005.

[25] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

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

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

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

[30] L. Grüne and J. Pannek, *Nonlinear Model Predictive Control: Theory and Algorithms*. Springer Cham, 2 ed., 01 2017.

[31] L. Wang, *Model Predictive Control System Design and Implementation Using MATLAB®*. Springer-Verlag London Limited, 2009.

[32] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows (2nd Ed.)*. USA: John Wiley & Sons, Inc., 1990.

[33] M. Asghari, A. M. Fathollahi-Fard, S. M. J. Mirzapour Al-e hashem, and M. A. Dulebenets, "Transformation and linearization techniques in optimization: A state-of-the-art survey," *Mathematics*, vol. 10, no. 2, 2022.

[34] D. He, B. Ai, K. Guan, L. Wang, Z. Zhong, and T. Kürner, "The design and applications of high-performance ray-tracing simulation platform for 5G and beyond wireless communications: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 10–27, 1st Quart., 2018.

[35] T. K. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *IEEE Antennas Propag. Magazine*, vol. 45, no. 3, pp. 51–82, Jun. 2003.

[36] M. Schlueter, S. O. Erb, M. Gerdts, S. Kemble, and J.-J. Rückmann, "MIDACO on MINLP space applications," *Advances in Space Research*, vol. 51, no. 7, pp. 1116–1131, 2013.

[37] P. Parida and S. S. Das, "Power allocation in OFDM based NOMA systems: A DC programming approach," in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 1026–1031, 2014.

[38] S. Bakirtzis, I. Wassell, M. Fiore, and J. Zhang, "Stochastic evaluation of indoor wireless network performance with data-driven propagation models," in *Proc. IEEE Global Commun. Conf.*, pp. 3587–3592, 12 2022.

[39] A. Goldsmith, *Path Loss and Shadowing*, p. 27–63. Cambridge University Press, 2005.

[40] K.-L. Hsiung, K. Seung-Jean, and S. Boyd, "Power control in lognormal fading wireless channels with uptime probability specifications via robust geometric programming," in *American Control Conference*, pp. 3955–3959, 2005.

[41] A. F. Zanella, A. Bazco-Nogueras, C. Ziemlicki, and M. Fiore, "Characterizing and modeling session-level mobile traffic demands from large-scale measurements," in *Proceedings of the 2023 ACM on Internet Measurement Conference*, IMC '23, (New York, NY, USA), p. 696–709, Association for Computing Machinery, 2023.

[42] J. T. Matamalas, M. De Domenico, and A. Arenas, "Assessing reliable human mobility patterns from higher order memory in mobile communications," *Journal of The Royal Society Interface*, vol. 13, p. 20160203, Aug. 2016.

[43] H. Zhang and L. Dai, "Mobility prediction: A survey on state-of-the-art schemes and future applications," *IEEE Access*, vol. 7, pp. 802–822, 2019.

[44] A. F. Zanella, A. Bazco-Nogueras, C. Ziemlicki, and M. Fiore, "Characterizing and modeling session-level mobile traffic demands from large-scale measurements," in *Proceedings of the 2023 ACM on Internet Measurement Conference*, pp. 696–709, 2023.

[45] A. Mchangama, J. Ayadi, V. P. G. Jiménez, and A. Consoli, "Mmwave massive mimo small cells for 5g and beyond mobile networks: An overview," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2020.

[46] M. Okmi, L. Y. Por, T. F. Ang, and C. S. Ku, "Mobile phone data: A survey of techniques, features, and applications," *Sensors*, vol. 23, no. 2, 2023.