Dimensioning, Cell Site Planning, and Self-Organization of 4G Radio Networks

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors
der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Informatiker,
Diplom-Mathematiker (FH)
Alexander Engels

aus Mönchengladbach

Berichter: Universitätsprofessor Dr. rer. nat. Rudolf Mathar
Universitätsprofessor Dr. rer. nat. Berthold Vöcking

Tag der mündlichen Prüfung: 27. September 2013

Diese Dissertation ist auf den Internetseiten
der Hochschulbibliothek online verfügbar.

Berichte aus der Kommunikationstechnik

Alexander Engels

Dimensioning, Cell Site Planning, and Self-Organization of 4G Radio Networks

Shaker Verlag
Aachen 2013

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: D 82 (Diss. RWTH Aachen University, 2013)

Copyright Shaker Verlag 2013
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-2315-2
ISSN 0945-0823

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Preface

This thesis was written during my time as a Research Assistant at RWTH Aachen University's Institute for Theoretical Information Technology.

First and foremost, I would like to thank my supervisor, Univ.-Prof. Dr. rer. nat. Rudolf Mathar, for giving me the opportunity to take a very unique path in pursuing my Ph.D degree. I would also like to thank Prof. Mathar for his continuous support and for being an excellent example of fair and practical leadership.

Many thanks to Univ.-Prof. Dr. rer. nat. Berthold Vöcking for taking the effort to referee this thesis.

A special thankyou goes to Michael Reyer, Gholamreza Alirezaei, Melanie Neunerdt, and Derek J. Corbett for their very helpful discussions and suggestions, and for proof-reading parts of this thesis. Furthermore, I would like to acknowledge the support of my colleague Florian Schröder who provided me with several eye-catching pictures that illustrate the principles of ray optical based path loss computation.

I would like to express my deepest gratitude to all my former and present colleagues at the Institute for Theoretical Information Technology. You helped create a comfortable and inspiring working environment, every day. Thank you for a good time, I will miss your company.

I am very proud that I had the opportunity to contribute to numerous exciting research projects that were carried out in close collaboration with industry partners. Particularly, I would like to thank all of my colleagues at QSC AG, Cologne, you made me feel like I was part of the team.

Ein besonderer Dank gilt meinen Eltern und der übrigen Familie, die mich über alle Jahre hinweg mit viel Geduld und Zuversicht unersetzlich unterstützt haben.

Finally, I am deeply grateful to all my friends and close companions for their patience and understanding throughout the whole Ph.D journey.

Aachen, October 2013 Alexander Engels

Contents

1 Introduction **1**
- 1.1 Outline . . . 2
- 1.2 Related Work, Initiatives, and Institutions . . . 3
- 1.3 Notation . . . 5

2 Mathematical Preliminaries **7**
- 2.1 Linear Programs . . . 7
- 2.2 Integer and Mixed-Integer Linear Programs . . . 8
- 2.3 Multi-Objective Optimization Problems . . . 9
  - 2.3.1 Pareto Front Exploration and the Scalarization Approach . . . 10
  - 2.3.2 Constrained Single Target Optimization . . . 12

3 Building Blocks for Radio Network Optimization **13**
- 3.1 Radio Wave Propagation Models . . . 13
  - 3.1.1 Semi-Empirical Path Loss Models . . . 15
  - 3.1.2 Ray Optical Path Loss Models . . . 17
  - 3.1.3 A Direction-Specific Land Use Based Path Loss Model . . . 19
- 3.2 Wireless Channel Models, Rate Computation, and Bandwidth Allocation 22
- 3.3 Demand Prediction Model . . . 26
- 3.4 Fundamental Problems in Radio Network Optimization . . . 27
  - 3.4.1 The Maximal Covering Location Problem . . . 28
  - 3.4.2 User Assignment in OFDMA Systems . . . 29
  - 3.4.3 Resource Allocation in OFDMA Systems . . . 30

4 Dimensioning of Multihop Networks **33**
- 4.1 System Model . . . 33
- 4.2 Optimization Model . . . 35
- 4.3 Concept Validation . . . 38
- 4.4 Summary . . . 43

5 Interference Approximation in LTE Heterogeneous Networks **45**
- 5.1 System Model . . . 46
- 5.2 Interference Approximation Model . . . 47
- 5.3 Numerical Evaluation . . . 51
- 5.4 Summary . . . 54

6 Cell Site Planning of LTE Heterogeneous Networks **55**
6.1 System Model . . . 55
6.2 Optimization Model . . . 57
6.3 Numerical Evaluation . . . 61
6.4 Summary . . . 65

7 Self-Optimization of Coverage and Capacity **67**
7.1 System Model . . . 68
7.1.1 Radio Resource Management and Scheduling . . . 71
7.1.2 Dynamic Parameter Adaption . . . 73
7.2 System Performance Measures . . . 75
7.2.1 Assessment of Network Coverage . . . 76
7.2.2 Assessment of Network Capacity . . . 77
7.3 Joint Coverage and Capacity Optimization . . . 78
7.3.1 Variants for Trade-Off Optimization . . . 82
7.4 Traffic Light Based Self-Optimization . . . 83
7.4.1 Climbing Up Principle for Monotone Performance Improvement 88
7.5 Numerical Evaluation . . . 90
7.5.1 Simulation Setup . . . 90
7.5.2 Case Study Results . . . 93
7.5.3 The Solution Space for Trade-Off Optimization . . . 100
7.6 Summary . . . 101

8 Conceptual Extensions **105**
8.1 Energy Efficiency . . . 105
8.2 Embedding User Acceptance as Decision Criterion . . . 107
8.3 Mobility Robustness . . . 109
8.4 Graph-Based Problem Partitioning and Adaptive Clustering . . . 110

9 Conclusions **113**
9.1 Summary and Contributions . . . 113
9.2 Future Research . . . 115

Acronyms **117**

List of Symbols **121**

Bibliography **125**

Index **133**

1 Introduction

Most network operators will enhance their existing radio networks by introducing fourth generation (4G) communication systems. This is a necessary step to cope with exponentially growing traffic demand as well as supporting sophisticated mobile services. Cellular 4G radio networks are based on the *Long-Term Evolution/System Architecture Evolution (LTE/SAE)* standard specification [30] and its extension *LTE Advanced* [29]. Additionally, the other prominent 4G standard is the *Worldwide Interoperability for Microwave Access (WiMAX)* standard [44, 83]. The underlying technology of all 4G standards is *Orthogonal Frequency Division Multiple Access (OFDMA)* [50, 97].

The roll-out of 4G communication systems brings many opportunities for network operators to reduce the costs and complexity of deploying and operating their networks. From the network operator's perspective, the costs of deploying and maintaining the network determine its profitability, and are therefore major criteria for roll-out decisions [33, 49].

Figure 1.1 illustrates the related tasks that have to be carried out by the network operator in a chronological order. First, the results from network dimensioning basically serve as input for business cases and strategic decisions. Second, cell site planning covers site selection and initial configuration of the network infrastructure that is actually deployed. Therefore, the optimization models for cell site planning have to be more accurate and more realistic than the ones that are utilized for network dimensioning. This requirement necessitates a greater level of detail in modeling interference-sensitive resource allocation and computation of accurate signal quality information. Finally, *self-optimization* of the deployed network allows network equipment to adapt its radio parameters autonomously, i.e., without any human intervention and without the corresponding personnel expenses.

The work presented in this dissertation provides novel concepts, optimization models, and related building blocks for the dimensioning, the planning, and the self-organized operation of 4G radio networks. Concerning the latter two tasks, it particularly focuses on *Heterogeneous Networks (HetNets)* that implement a multi-tier cell topology.

Parts of this thesis have already been published in [35, 36, 37, 38, 39, 76] and [107].

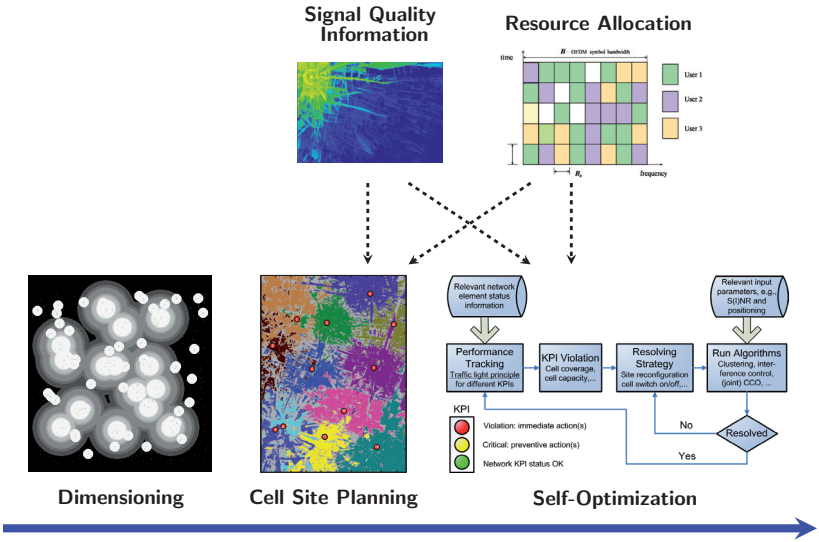


Figure 1.1: Basic tasks (lower components) and related input information (upper components) to deploy and operate fourth generation radio networks.

1.1 Outline

The subsequent section gives an overview of the related work, initiatives, and institutions. At the end of this chapter, we introduce the basic notations that are used throughout this thesis.

Chapter 2 covers the mathematical preliminaries to the different optimization problems that we consider in this work. All following chapters correspond to the components that are depicted in Figure 1.1 in a chronological order.

Chapter 3 presents a discussion of the building blocks that serve as immanent components of the subsequent optimization models. Additionally, some of the fundamental problems in radio network optimization are examined. These problems are strictly related to the optimization models that we develop in this thesis. Due to the close relation, finding optimal solutions for our optimization problems turns out to be computationally hard.

Chapter 4 considers WiMAX-based radio networks that support multihop transmission. We determine the infrastructure dimensioning of such networks according to an economically motivated approach and subject to the expected user distribution and rate demand. The presented dimensioning approach deals with inter-cell interference by ensuring a certain minimum distance between deployed transmitters.

This simple method is not appropriate for the purpose of accurate cell site planning and configuration of LTE HetNets. Hence, in Chapter 5 we develop a low-complexity interference approximation model that estimates the bandwidth requirements for macrocells and femtocells subject to inter-cell and cross-tier interference.

For the optimal cell site planning of LTE HetNets in Chapter 6, we consider macrocells and user-operated femtocells that are not necessarily active all the time. The objective of the corresponding optimization problem is to provide a minimum number of macrocells such that mobile services are area-wide guaranteed. On the other hand, it avoids dispensable cell sites for the sake of cost efficiency and low interference.

Chapter 7 extends the HetNet deployment results from Chapter 6 and proposes a model for joint coverage and capacity optimization. We present an integrated approach for the self-optimization of coverage and capacity in the time-variant system. The corresponding algorithms are designed according to a traffic light principle. They autonomously control site activity, transmission power, and antenna downtilt parameters in the operating HetNet.

Before this work is concluded in Chapter 9, Chapter 8 presents several conceptual extensions to the previous optimization problems. The conceptual extensions either provide heuristics to lower the problem complexity or they address the incorporation of additional aspects into the network optimization domain, e.g., energy efficiency and integration of user acceptance.

1.2 Related Work, Initiatives, and Institutions

Generally, the basic tasks and objectives for network planning and network operation have not changed over the evolving generations from 2G (GSM) radio networks to 3G (UMTS, HSPA) and 4G (WiMAX, LTE, LTE Advanced) communication systems. As a consequence, the basic principles, workflow descriptions, and optimization problems, e.g., discussed in [84, 101, 72, 103, 14, 45], are still relevant. The systems themselves, however, have changed substantially not only in terms of performance capability. The backbone architecture of networks has changed (E-UTRAN) as well as the supported transmission modes (coding, modulation), the multiple access technology (OFDMA), and the supported antenna techniques (MIMO, beamforming). Solid introductions and detailed technical information on these topics are provided in [30, 29, 94, 50, 100].

Due to the changes in the underlying system technology, existing optimization models have to be adapted or have to be re-developed for each network generation. The typical problems for the planning and optimization of third generation *Code Division Multiple Access (CDMA)* networks are presented in [14]. One of the main challenges for maximizing coverage and capacity of 3G networks is coping with inter-cell and intra-cell interference by choosing suitable spreading code assignments and beneficial power allocations. Since all assignment and allocation decisions at base stations are mutually interconnected, interference-sensitive optimization becomes a computationally hard combinatorial problem. The same holds for OFDMA-based networks although

the technical problem definition changes: Interference depends on the subcarrier assignment and the power allocation on the subcarriers as interference coordination takes place in the frequency domain. Related problems, optimization models, and computationally efficient heuristics are presented, e.g., in [62, 66, 27].

Modeling and handling interference becomes even more complex with the integration of multiple tiers into the network topology. A second tier refers, for instance, to relay stations in IEEE 802.16j networks (WiMAX) [83] or to pico-/femtocells in LTE advanced systems (LTE HetNets) [29]. Co-channel deployment can cause cross-tier interference if cells at different tiers use the same frequency spectrum. This cross-tier interference is complex to handle due to the cell interdependencies and the required communication overhead [69]. Radio resource management and network planning for 802.16j networks is investigated, for instance, in [80]. Concerning LTE HetNet deployment and configuration, we refer to [81] and [67].

Radio Resource Management (RRM) and cell reconfiguration are resource-intensive tasks that have to be carried out with respect to dynamic changes in the network. The required resources are particularly related to computational complexity, time consumption, communication overhead, and man power. A popular paradigm to minimize such aspects is the implementation of the network as a *Self-Organizing Network (SON)*. The underlying principle of any SON is to delegate tasks from the *Network Management System (NMS)* to the network elements [85]. This feature enables autonomous *Self-Planning*, *Self-Optimization*, and *Self-Healing* in a semi-decentralized or fully decentralized manner, e.g., as proposed in [86, 75]. The strong interest in SON topics is reflected by the recent activities under guidance of the *Next Generation Mobile Networks alliance (NGMN)* and the *3rd Generation Partnership Project (3GPP)*. While NGMN mainly provides economical and technical guidelines [77, 79], 3GPP is in charge of the standardization of related network components, e.g., see [13, 7].

The 3GPP specifications for self-organized LTE HetNets are covered by different releases. LTE Release 8 contains the fundamental specifications for *Home eNodeB (HeNB)* components, self-establishment of network equipment, and automatic neighbor relation list management [6]. LTE Release 9 covers the specifications for enhanced HeNB functionality as well as studies on self-organization, self-healing, and self-organized coverage and capacity optimization [8]. In releases 10 and 11, 3GPP extends the specifications and SON use cases successively with respect to LTE Advanced systems. These releases particularly emphasize the coordination between SON functionalities [29, 4, 5].

Several initiatives have been established to investigate and contribute to self-optimization and self-configuration in wireless communication networks [85]: As part of the *Celtic Initiative* [21], the *Celtic GANDALF* project contributed at a very early stage to automated troubleshooting and automatic control of network parameters [22]. The *End-to-End Efficiency* project, funded by the European Union within the *7th Framework Program*, covers some SON related use cases such as handover optimization and inter-cell interference coordination [1]. The SOCRATES project was established within the same EU program. This project particularly addresses SON aspects

such as integrated handover parameter optimization and load balancing, automatic generation of initial insertion parameters, and cell outage management [96]. Furthermore, several *COST (European COoperation in Science and Technology)* projects have contributed to SON-related network modeling, network planning, and network optimization [41].

1.3 Notation

We use the basic notation listed in Table 1.1 throughout the rest of this work. In the subsequent chapters, the symbols and identifiers are extended with respect to the particular context.

Symbol & domain	**Description**
$\mathcal{S}$, $\mathcal{K}$, $\mathcal{F}$	Index sets of macrocell sites, relay stations, and femtocell transmitters with representative indices $s \in \mathcal{S}$, $k \in \mathcal{K}$, $f \in \mathcal{F}$.
$\mathcal{T}$	Index set of demand nodes utilized for modeling the traffic distribution, representative index $t \in \mathcal{T}$.
$r_t \in \mathbb{R}_{\geq 0}$	Requested data rate at a demand node.
$r_t^{\text{MIN}} \in \mathbb{R}_{\geq 0}$	Minimum required data rate if the demand node is served.
B_s, B_k, $B_f \in \mathbb{R}_{\geq 0}$	Total available bandwidth at macrocells, relay stations, and femtocells.
e_{st}, e_{kt}, $e_{ft} \in \mathbb{R}_{\geq 0}$	Supported spectral efficiency (signal quality indicator) from macrocell transmitter, relay station, and femtocell transmitter to demand node t.
y_s, y_k, $y_f \in \{0, 1\}$	Binary decision variables indicating the selection of a (configured) macrocell site, relay station, or femtocell transmitter.
z_{st}, z_{kt}, $z_{ft} \in \{0, 1\}$	Binary decision variables indicating the assignment of demand node t to a certain macrocell site, relay station, or femtocell transmitter.
$z_t \in \{0, 1\}$	Auxiliary variable indicating that demand node t is assigned to a transmitter.
b_{st}, b_{kt}, $b_{ft} \in \mathbb{R}_{\geq 0}$	Amount of allocated bandwidth for transmission from macrocell transmitter, relay station, and femtocell transmitter to demand node t.
$r_t^{\text{eff}} \in \mathbb{R}_{\geq 0}$	Auxiliary variable describing the effectively served data rate at demand node t, depending on the particular signal quality and bandwidth allocation.

Table 1.1: Basic symbols and identifiers – parameters (upper part) are separated from variables (lower part) by the dashed line.

For network dimensioning, planning, and optimization we consider *macrocell* sites that are represented by the index set $\mathcal{S}$. Depending on the considered system technology, i.e., WiMAX or LTE, a site corresponds to a *Base Station (BS)* or to an *eNodeB (eNB)*. When we present our approach for the dimensioning of WiMAX multihop networks in Chapter 4, *Relay Stations (RSs)* are represented by the index set $\mathcal{K}$. Analogously, the index set $\mathcal{F}$ describes the *femtocell* related HeNBs for LTE HetNet planning and optimization in Chapter 6 and Chapter 7. For the sake of simplicity, we define the indices s, k, f as representatives that refer to all elements of the corresponding index sets $\mathcal{S}, \mathcal{K}$, and $\mathcal{F}$. If a transmitter entity s, k, or f is selected, the corresponding decision variable y_* is set to one. Please note that each selected transmitter entity applies a certain configuration. The configuration state space can include, for instance, different transmit powers or different antenna downtilt settings. Optimal network deployment and initial configuration can be considered jointly within the same optimization problem if multiple configurations of the cell sites are included in $\mathcal{S}$. This principle is applied when we consider self-organized online control of macrocell sites and femtocells in Chapter 7.

Computation time and memory are critical resources for the application of our optimization models. In order to lower the size of the problem instances, the *Demand Node (DN)* concept from [101] is adapted in Section 3.3 to abstract from single users. The demand nodes from the index set $\mathcal{T}$ model the spatial distribution of aggregated users as well as their joint rate demand for a predefined reference period. Each DN $t \in \mathcal{T}$ is associated with a certain location and a certain data rate demand r_t. Please note that t might context-specifically also represent a single user, i.e., a sufficiently small DN that contains exactly one user (receiver). Serving the requested data rate requires the assignment of the DN to a transmitter station $* \in \{s, k, f\}$ by setting the corresponding decision variables z_{*t} and z_t to one. The serving station $*$ has to allocate a sufficiently large amount of bandwidth b_{*t} to provide an effective data rate of r_t^{eff} to the DN. The required bandwidth is determined subject to the requested data rate r_t, the minimum required rate r_t^{MIN}, and the supported spectral efficiency (signal quality) e_{*t} on the link to t. This principle is formalized in Section 3.2 for the interference-free case and in Chapter 5 for an interference-limited system. Finally, the overall available bandwidth at each macrocell site, relay station, and HeNB $*$ is limited by the corresponding maximum bandwidth B_*.

Please note that symbols with a capitalized subscript or superscript refer to (constant) parameters, whereas symbols with a lowercased subscript or superscript refer to variables.

2 Mathematical Preliminaries

Even though the subsequent chapters deal with various topics and applications, most of them have two conceptual aspects in common. First, the related optimization problems are of multi-objective nature, i.e., we consider the joint optimization of multiple *Key Performance Indices (KPIs)*. The corresponding single objectives can be contradictory. And second, we formalize the optimization models as linear programs that can also contain integer constraints. The basic concepts and aspects regarding those two properties are introduced in the following.

2.1 Linear Programs

According to [18], a *general optimization problem* is defined as

$$\begin{aligned} &\min\ f_0(\boldsymbol{x}) \\ &\text{s.t. } f_i(\boldsymbol{x}) \leq 0\ , && i = 1, \ldots, m \\ &\quad\ h_j(\boldsymbol{x}) = 0\ , && j = 1, \ldots, p \end{aligned} \tag{2.1}$$

with *optimization variable* $\boldsymbol{x} \in \mathcal{X}^n$, inequality constraint functions $f_i : \mathcal{X}^n \to \mathbb{R}$, equality constraint functions $h_j : \mathcal{X}^n \to \mathbb{R}$, and *objective function* $f_0 : \mathcal{X}^n \to \mathbb{R}$. If not defined otherwise, we assume $\mathcal{X} = \mathbb{R}_{\geq 0}$. The set of points for which the objective function and all constraint functions are defined is called the domain $\mathcal{D}$ of the optimization problem. A point $\boldsymbol{x} \in \mathcal{D}$ is *feasible* if it satisfies all constraints of problem (2.1). The *optimal value* or *optimum* of problem (2.1) is defined as

$$\inf \{ f_0(\boldsymbol{x}) \mid f_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, m,\ h_j(\boldsymbol{x}) = 0,\ j = 1, \ldots, p \}$$

and a feasible point $\boldsymbol{x}^* \in \mathcal{D}$ for which the optimum is attained, is called *optimal point*.

If the objective function and all constraint functions are linear, problem (2.1) can be reformulated as

$$\begin{aligned} &\min\ \boldsymbol{c}^T \boldsymbol{x} \\ &\text{s.t. } \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i\ , && i = 1, \ldots, m \end{aligned} \tag{2.2}$$

for $\boldsymbol{x} \in \mathcal{X}^n$, vectors $\boldsymbol{c}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$, and scalars $b_1, \ldots, b_m \in \mathbb{R}$. The optimization problem (2.2) is called a *Linear Program (LP)*. If the LP considers the p equality constraints from (2.1) its feasible points are in a subspace of $\mathcal{D}$ that is reduced by the dimension $d = \dim \{ \boldsymbol{x} \in \mathcal{X}^n \mid h_j(\boldsymbol{x}) = 0,\ j = 1, \ldots, p \}$.

Since linear programs are convex by definition, they can be solved by methods that are computationally efficient, at least in practice: The *simplex method* finds optimal points in a compact set by exploring the vertices of the polyhedron that describes the solution space. The polyhedron is defined by the constraints $\boldsymbol{a}_i^T \boldsymbol{x} = b_i$, $i = 1, \ldots, m$ that follow from the inequalities of (2.2) and the incorporation of corresponding slack variables [61, 26]. There exist examples where the simplex method needs an exponential number of operations to find the optimum [59]. However, in practice the computational effort is of order $n^2 m$ assuming that $m \geq n$ [18]. If the LP considers the (linear) equality constraints from (2.1), the computational effort is of order $(n-d)^2 m$ for $m \geq (n-d)$.

Interior point methods are another prominent approach to solve LPs. The computational effort of these methods is strictly bounded by $\mathcal{O}\left(n^3 \sqrt{n}\right)$, i.e., interior point methods have a polynomial complexity [108]. Moreover, the algorithms can perform significantly better than this worst case upper bound for many practical problem instances. Interior point methods outperform the simplex algorithm on several problem classes, e.g., on large degenerate problems with many zero entries in the solution vector [43]. Most state-of-the-art LP solvers such as CPLEX [51] or Gurobi [48] support both solution approaches.

2.2 Integer and Mixed-Integer Linear Programs

The linear optimization problem (2.2) is called an *Integer Linear Program (ILP)* [61] if its optimization variables are subject to corresponding *integrality constraints*. In this case, we assume $\mathcal{X} = \mathbb{N}$. If only a proper subset of variables is restricted to the integer domain, the resulting problem is called a *Mixed-Integer linear Program (MIP)*. Both variants belong to the same problem class in terms of computational complexity. Thus, in the following we will not distinguish MIPs from ILPs.

Contrary to the problem class of LPs, solving ILPs is NP-hard [68]. The search space of binary ILPs ($\mathcal{X} = \{0, 1\}$), for instance, grows exponentially with n. Therefore, solving an ILP can become computationally intractable for large problem instances. It is often useful to determine solutions by heuristics such as *simulated annealing* or *tabu search algorithms* [14]. For ILPs of a suitable size, however, the most common techniques to compute optimal solutions are *branch-and-bound* and *branch-and-cut* algorithms [61, 68]. The state-of-the-art LP solvers mentioned above solve ILPs and MIPs by applying variants of these algorithms. Moreover, they can introduce problem-specific cutting planes to simplify the solution space. Computing optimal ILP solutions can take hours or days even for reasonably sized problem instances and very sophisticated solver implementations.

We refer to [25] for more details on integer programming and the related class of combinatorial optimization problems. This reference provides very useful information, and moreover, interesting facts about the people that have dominated this research field in the last 50 years.

2.3 Multi-Objective Optimization Problems

A *Multi-Objective Optimization (MOO)* problem is defined as a general optimization problem according to (2.1) with v objective functions $F_l : \mathcal{X}^n \to \mathbb{R}$, see [18]. Thus, the function $f_0 : \mathcal{X}^n \to \mathbb{R}^v$ is vector-valued with

$$f_0(\boldsymbol{x}) = ((F_1, \ldots, F_v)(\boldsymbol{x}))^T \ .$$

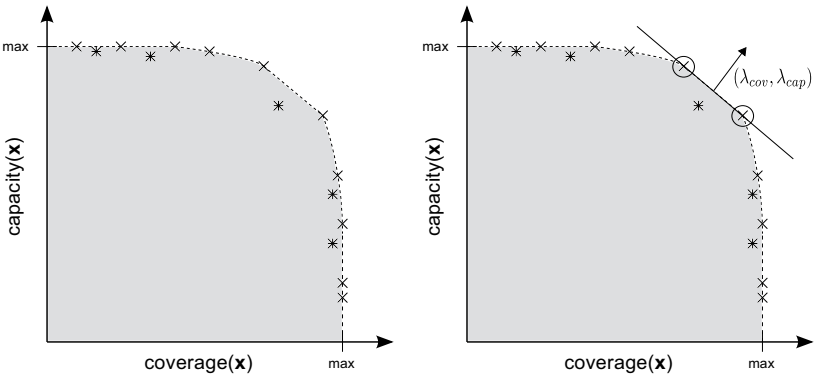
MOO problems are linear if all objectives F_l and all constraint functions are linear. In that case, the definitions and properties from Section 2.2 hold for a proper selection of $\mathcal{X}$. Compared to an optimization problem (2.1) with a scalar-valued objective function, it is not intuitive how optimal points are defined for an MMO problem.

Many relevant problems from radio network planning and optimization are MOO problems. Typically, several decision criteria have to be considered jointly, e.g., coverage, capacity, and cost are considered as separate objectives F_l. The joint optimization of multiple objectives can be a trade-off task if they are contradicting. For instance, maximization of coverage and capacity as well as maximization of coverage and minimization of cost generally are trade-off tasks [105, 55]. In such a situation, it is a problem to decide which solution should be preferred if several ones exist. In the following, we formalize this problem and introduce three approaches to deal with it.

If $\boldsymbol{x}$ is a feasible point, the l-th objective $F_l(\boldsymbol{x})$ may be interpreted as its score. If two points $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{X}^n$ are both feasible, $F_l(\boldsymbol{x}) \leq F_l(\boldsymbol{y})$ means that $\boldsymbol{x}$ is at least as good as $\boldsymbol{y}$ with respect to the l-th objective; $F_l(\boldsymbol{x}) < F_l(\boldsymbol{y})$ means that $\boldsymbol{x}$ is better than $\boldsymbol{y}$ or $\boldsymbol{x}$ beats $\boldsymbol{y}$ on the l-th objective, respectively. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are both feasible, $\boldsymbol{x}$ is *better* than $\boldsymbol{y}$, i.e., $\boldsymbol{x}$ *dominates* $\boldsymbol{y}$, if $F_l(\boldsymbol{x}) \leq F_l(\boldsymbol{y})$ for all $l = 1, \ldots, v$ and $F_u(\boldsymbol{x}) < F_u(\boldsymbol{y})$ holds for at least one u. Roughly speaking, $\boldsymbol{x}$ is better than $\boldsymbol{y}$ if $\boldsymbol{x}$ meets or beats $\boldsymbol{y}$ on all objectives and beats it on at least one objective. If there exists a non-dominated feasible point $\boldsymbol{x}^* \in \mathcal{X}^n$ that is optimal for each *scalar* problem

$$\begin{aligned} \min \ & F_l(\boldsymbol{x}) & \\ \text{s.t.} \ & f_i(\boldsymbol{x}) \leq 0 \ , & i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 0 \ , & j = 1, \ldots, p \end{aligned}$$

with $l = 1, \ldots, v$, $\boldsymbol{x}^*$ is an optimal point. Such an optimal point does not exist for many MMO problems. Hence, the following definition is the more relevant one for MMO problems. Any *Pareto optimal point* or *efficient point* $\boldsymbol{x}^{PO} \in \mathcal{X}$ satisfies the following condition: If $\boldsymbol{y} \in \mathcal{X}$ is feasible and $F_l(\boldsymbol{y}) \leq F_l(\boldsymbol{x}^{PO})$ for $l = 1, \ldots, v$ then it holds $F_l(\boldsymbol{x}^{PO}) = F_l(\boldsymbol{y})$ for all $l = 1, \ldots, v$. This means that it is impossible to improve one score of a Pareto optimal point without decreasing another one. Particularly, it holds that if a feasible point is not Pareto optimal there exists at least one other feasible point that is better. In consequence, Pareto optimal points are well suited candidates for finding beneficial solutions for MOO problems.



(a) The convex hull (dashed line) of Pareto optimal points (cross or asterisk) defines the set of supported efficient points (crosses).

(b) The scalarization approach can find the supported Pareto optimal points (crosses) of the Pareto front if $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) > \mathbf{0}$.

Figure 2.1: An exemplary two-objective optimization problem.

2.3.1 Pareto Front Exploration and the Scalarization Approach

Figure 2.1 (a) illustrates the set of Pareto optimal points for a discrete multi-objective maximization problem that considers coverage and capacity as conflicting objectives. Coverage can be interpreted as the area where users experience a minimum received signal power, whereas capacity might describe the maximum cell edge throughput.

Each objective is assumed to be a function of the optimization variable $\boldsymbol{x} \in \mathcal{D}$, which contains integer components. Furthermore, the achievable objective values are bounded by a power constraint. The image of the set of Pareto optimal points forms the *optimal trade-off surface* or so-called *Pareto front*. Its shape describes the trade-off characteristic between the objectives.

Exploring the points of the Pareto front is an intuitive approach to find a beneficial solution for joint coverage and capacity maximization – and for joint MOO in general. From the set of explored Pareto optimal points, we choose the one that is best suited in terms of a certain decision criterion. Exploring all Pareto optimal points, however, is computationally hard since there usually exist exponentially many points [57, 63]. Thus, computing the Pareto front requires the application of sophisticated methods and algorithms. In [109], the Pareto front is approximated by genetic algorithms and [24, 87] consider particle swarm optimization for this purpose.

Even if we are able to compute a set of Pareto optimal points, the following problem arises as soon as the set contains more than one element: If each Pareto optimal point beats the other ones on at least one objective, which point gives the best solution? One way to cope with this dilemma is to make use of the *scalarization approach* and its properties, which are discussed in [34].

Basically, the scalarization of an MOO problem is obtained by combining the multiple objectives of $f_0(\boldsymbol{x})$ into a single *weighted sum objective*

$$\boldsymbol{\lambda}^T \left((F_1, \ldots, F_v)(\boldsymbol{x})\right) = \sum_{l=1}^{v} \lambda_l F_l(\boldsymbol{x}) \ . \tag{2.3}$$

The factor λ_l can be interpreted as the weight attached to the l-th objective or as the importance of making F_l small. The ratio λ_i/λ_j is the relative weight or *relative importance* of the i-th objective compared to the j-th objective. Alternatively, λ_i/λ_j might be interpreted as *exchange rate* between the two objectives. Solving the MOO problem for the weighted sum objective (2.3) has the same computational complexity as solving a scalar-valued optimization problem.

For $\boldsymbol{\lambda} > \mathbf{0}$, any optimal solution of the MOO problem with objective (2.3) is Pareto optimal [18, 34]. As illustrated in Figure 2.1 (b), the weight vector $\boldsymbol{\lambda}$ gives the normal of the tangential hyperplane at the associated Pareto optimal point(s). Please note that Pareto optimal points in the interior of the convex hull (gray area) cannot be found by the scalarization approach. Therefore, they are called *unsupported efficient points*. In Figure 2.1, unsupported efficient points are depicted as an asterisk. Furthermore, it holds that:

1. A supported Pareto optimal point can have several tangential hyperplanes if the convex hull is not smooth, i.e., if it is not (multidimensionally) continuously differentiable. In this case such a point can be found by the scalarization approach for different weight vectors $\boldsymbol{\lambda}$.

2. If the scalarization approach finds multiple supported efficient points for a certain weight vector $\boldsymbol{\lambda}$, the value of the corresponding weighted sum objective is the same for all points. However, the single components of the efficient points can differ significantly, e.g., see the encircled points in Figure 2.1 (b).

Both properties are particularly relevant for integer MOO problems, where the convex hull of the discrete – and in most cases non-dense – set of Pareto optimal points usually is not smooth. If the weights λ_l are chosen with respect to a certain intention, e.g., to obtain a solution with balanced objectives, those properties can lead to solutions that do not match with this intention at all. Figure 2.1 (b) illustrates such a situation where the weights are chosen equally but the two resulting efficient points (encircled) mutually differ in one objective component. Nevertheless, the particular setting of λ_l generally influences the balance of the objectives that are obtained for multiple instances of the MOO problem.

In the ideal case, the exchange rates between the objectives are unambiguously. Then the MMO problem reduces to a scalar-valued optimization problem, which can be solved by the conventional methods that were discussed in Sections 2.1 and 2.2.

2.3.2 Constrained Single Target Optimization

An alternative but very simple approach is to solve the MMO problem in a hierarchical fashion [105]. Basically, this is done by considering a scalar-valued optimization problem for one of the objectives $F_l(\boldsymbol{x})$. After an optimal solution has been computed, the achieved objective value defines an upper bound for $F_l(\boldsymbol{x})$ when the procedure is repeated for another objective $F_{l'}(\boldsymbol{x})$, $l' \neq l$. The upper bound on $F_l(\boldsymbol{x})$ is implemented by an additional (maximum) constraint. The solutions that are obtained by this approach strictly depend on the order of the single optimization tasks, i.e, they depend on the predefined hierarchy of objectives. In consequence, the application of this method makes only sense if the considered hierarchy is reasonably defined.

3 Building Blocks for Radio Network Optimization

This chapter introduces the building blocks for the optimization of 4G radio networks at different stages of the system lifecycle. In particular, it provides the detailed definition of models for radio wave propagation prediction, channel models, and methods to describe user mobility and demand in a computationally efficient way. Furthermore, we discuss some of the fundamental problems in radio network optimization. The discussed problems are closely related to the optimization models that we develop in the subsequent chapters.

Parts of this chapter have already been published in [37, 76].

3.1 Radio Wave Propagation Models

Signal strength and *Channel State Information (CSI)* provide essential input information for any approach that considers planning and control of radio networks. They are important for all applications that deal with time-variant user positions, e.g., localization and location tracking. Basically, the information is used to describe the wireless channel characteristic that might vary over time and the spatial domain. Exact methods like channel measuring or estimation of the channel impulse response are very resource consuming and usually not practicable to obtain area-wide information. Therefore, sophisticated models have been developed in order to provide highly accurate approximations.

In the following, we neglect the signal phase information, which is justified by the long-term perspective of the subsequent applications. The amplitude of a received signal is determined by the emitted signal power and the physical effects that the signal experiences on its path from the transmitter to the receiver. A radio wave propagation model typically describes the effects on the signal path either in an empirical (stochastic) or in a semi-empirical way, where the latter one incorporates deterministic components. We refer to [88] for a comprehensive introduction to the basic principles of radio wave propagation modeling. Particularly, the document provides a good overview of some popular models like the *Hata model* and the *COST-231-Walfisch-Ikegami model*. It serves as root reference for the following definitions and descriptions.

The *path loss* information $L(t)$ is the relevant output of a radio wave propagation model. The pass loss describes the average attenuation of the emitted signal power P on its path to the *receiver point* t, i.e.,

$$L(t) = \frac{P}{P_t} \qquad \text{and} \qquad L^{\text{dB}}(t) = 10 \log_{10} \frac{P}{P_t} \; , \tag{3.1}$$

where P_t is the received signal power at t. According to [42], the effective path loss is a superposition of three essential components, namely

1. a distance-dependent basic path loss $L_0(t)$,
2. the *slow fading* (*shadowing*) effects modeled by a random variable G_{slow}, and
3. the *fast fading* effects modeled by a random variable G_{fast}.

With respect to the distance $d(t)$ between transmitter and receiver point t, the distance-dependent basic path loss is often assumed as the *Line-of-Sight (LoS)* or *free-space* path loss component, i.e.,

$$L_0(t) = \frac{(4\pi)^2 (d(t))^\gamma}{\lambda_{f_c}^2 \, G_A(\phi(t), \psi(t))}$$

for signal wavelength λ_{f_c} at *carrier (center) frequency* f_c, the *antenna gain* $G_A(\phi(t), \psi(t))$, and the free-space *path loss exponent* $\gamma = 2$. On a logarithmic scale this is

$$L_0^{\text{dB}}(t) = 20 \log_{10} \frac{4\pi}{\lambda_{f_c}} - 10 \log_{10} G_A(\phi(t), \psi(t)) + \gamma 10 \log_{10} d(t) \; . \tag{3.2}$$

Non-Line-of-Sight (NLoS) situations are modeled by a larger path loss exponent, which can increase up to a value of five for urban environments. The antenna gain $G_A(\phi(t), \psi(t))$ is given by an antenna-specific pattern according to Figure 3.1. It considers the transmitter in the center position and t in direction $(\phi(t), \psi(t))$ in spherical coordinates. The gain of beamforming-generated antenna patterns can be approximated by an equivalent representation. Please note that we do not consider an antenna gain at the receiver side for path loss computation and that (3.2) is only feasible for far-field considerations, i.e., for $d(t) \gg 0$.

Finally, the effective (overall) path loss is described as

$$L(t) = L_0(t) G_{\text{fast}} G_{\text{slow}} \quad \text{or} \quad L^{\text{dB}}(t) = L_0^{\text{dB}}(t) + 10 \log_{10} G_{\text{fast}} + 10 \log_{10} G_{\text{slow}} \; . \tag{3.3}$$

The fading components model the signal variation over time for moving receivers in time-variant channel simulations. Fast fading results from multipath propagation of the signal due to physical reflection and scattering effects. It is typically modeled by a *Rayleigh* distributed random variable G_{fast} for LoS scenarios and as *Ricean* distributed random variable for the NLoS case. Fast fading influences the signal on a very small time scale. Slow fading effects take place on a much larger time scale since they are caused by the shadowing impact of large obstacles such as buildings or hills.

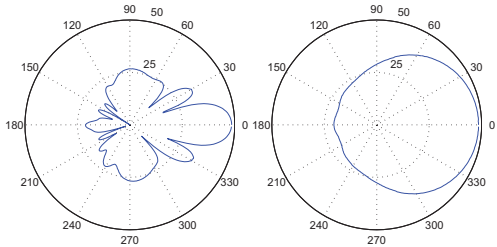


Figure 3.1: Exemplary antenna pattern from [58]: elevation and azimuth diagrams.

The shadowing-related variation of the signal magnitude in dB can be modeled suitably by a log-normal distributed random variable G_{slow}. We consider the time-invariant channel since tasks like network dimensioning and cell site planning are carried out with respect to the long-term characteristic of the channel. Fading effects are incorporated into the system model as a constant term that reflects either the average fading situation or the worst case fading situation. Therefore, in the following we concentrate on alternative models for the basic path loss $L_0^{\text{dB}}(t)$.

A widely used representation for measurement calibrated path loss models is

$$L^{\text{dB}}(t) = \Delta_0 + 20\log_{10}\frac{4\pi}{\lambda_{f_c}} + \gamma 10\log_{10} d(t) + L_{\text{effects}}^{\text{dB}}(t) \ , \tag{3.4}$$

where Δ_0 serves as offset constant in a least-squares regression fitting [40] and covers the antenna gain, potential measurement inaccuracies, and shadowing related effects. Some path loss models exploit further knowledge of certain physical conditions such as antenna height, terrain type, or building information and extend the model by a related term $L_{\text{effects}}^{\text{dB}}(t) \neq 0$. Consequently, path loss models differ in the particular term $L_{\text{effects}}^{\text{dB}}(t)$ as well as in the according model parameters that are typically estimated from different measurements.

In the following, we discuss some prominent empirical and semi-empirical path loss models. Furthermore, in Section 3.1.3 we develop a direction-specific path loss model that combines principles from empirical and ray optical path loss computation.

3.1.1 Semi-Empirical Path Loss Models

The Hata model – a variant of the well-known *Okumura* model [88] for carrier frequencies above 1.5 GHz – is intended for computing the path loss in large cells of $1 - 20$ km. The COST-231-Walfisch-Ikegami model considers cell radii in the range of 20 m to 5 km and is, therefore, a suitable choice for path loss computation in femtocells and picocells. Both models consider the antenna height and the reference height of the receiver in the term $L_{\text{effects}}^{\text{dB}}(t)$. Additionally, they apply a constant penetration term that is chosen with respect to the particular environment. For instance the Hata model

	Terrain type		
Model parameter	A hilly heavily forested	B hilly or plain lightly or moderately forested	C plain lightly forested
a	4.6	4.0	3.6
b	0.0075	0.0065	0.0050
c	12.6	17.1	20.0

Table 3.1: Terrain-specific parameters for the Erceg model.

adds a penetration constant of 3 dB for metropolitan areas. Both models serve as key components for system simulations in the SOCRATES project [95] when algorithms for self-organized LTE HetNets are assessed.

Distinguishing 13 different propagation environments, the *WINNER II* project group provides parameter sets for more than 20 semi-empirical path loss models. All those models were developed on basis of extensive measurement (channel sounding) campaigns [53]. Some of the models consider indoor propagation, which is modeled by penetration terms $L^{\mathrm{dB}}_{\mathrm{effects}}(t)$ that depend on the number of floors and the number of walls between the transmitter and the receiver. From *indoor office* over *indoor-to-outdoor* to *bad urban macrocell* and *rural macrocell* environments, almost all models distinguish LoS from NLoS situations. They apply distance-dependent parameter sets that are defined either by an absolute distance range or relatively with respect to the antenna height. If the particular environment is unknown, the LoS/NLoS-sensitive path loss models provide scenario-specific approximations for the LoS probability. The LoS probability decreases exponentially with the distance of the receiver point. The main intention of the WINNER initiative was to develop channel models for time-variant MIMO systems. Thus, the provided path loss models are combined with environment-specific fading models and both serve as key components of sophisticated MIMO channel models. We discuss some of the channel models in Section 3.2. Many recent works that deal with LTE systems utilizes the WINNER path loss models, the WINNER channel models, or variants of it.

The *Erceg path loss model* is very popular for radio wave propagation in suburban or rural areas since it distinguishes the terrain type between transmitter and receiver [40]. Neglecting the shadowing component, the Erceg model is defined by (3.4) for applying $L^{\mathrm{dB}}_{\mathrm{effects}}(t) = 0$, the path loss exponent

$$\gamma = (a - b\,h_A + c/h_A)$$

for an antenna (transmitter) height $h_A,\ 10 \leq h_A \leq 80$ [m], and parameters $a,\ b,\ c$ that are chosen from Table 3.1 according to the predominant terrain type. The Erceg model improves the prediction accuracy for the designated application scenarios compared to approaches that do not consider the terrain type. However, it is restricted to the choice of one (predominant) terrain category. Hence, all receiver points at the same distance

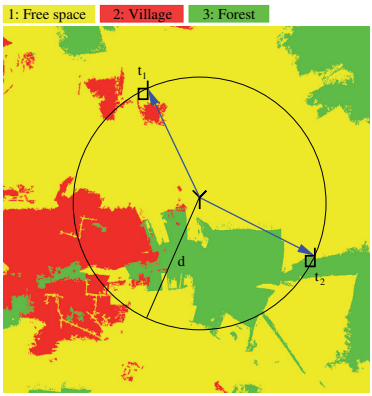


Figure 3.2: In the Erceg model, receivers t_1 and t_2 experience equal path losses due to the same distance d and the same predominant terrain type.

from the transmitter gain identical path losses if their predominant terrain type is the same. This can lead to inaccuracies and inconsistent results in some situations, see Figure 3.2. In Section 3.1.3, we propose a semi-empirical path loss model that overcomes this drawback. Since our model utilizes concepts from ray optical path loss computation, we first introduce the principles of ray optical algorithms.

3.1.2 Ray Optical Path Loss Models

Ray optical models improve the path loss prediction accuracy by incorporating environment-specific effect terms $L^{\mathrm{dB}}_{\mathrm{effects}}(t)$ in (3.4) in a deterministic manner. This approach is very resource consuming in terms of computational complexity, runtime, and expenses for providing the related input data. Therefore, *ray optical path loss computation* is usually only applied for purposes that necessarily require the supported level of accuracy. For instance, network dimensioning – as discussed in Chapter 4 – is typically carried out using non-deterministic and low-cost path loss models according to Section 3.1.1. On the other hand, cell site planning and site configuration in urban environments prerequisite very detailed information about the radio conditions in the corresponding area. Therefore, the following approach is utilized for path loss computation in the Chapters 6 and 7.

The basic principle of ray optical path loss prediction is to identify the signal propagation paths between the transmitter and the receiver point t as a set of rays. A ray is defined as a sequence of straight lines through the scene, which are connected by effect-related deflection points. The buildings in an urban environment are described as polyhedrons that are given by the buildings' surface sections (*facets*). In

(a) Physical effects. (b) Reflection. (c) Horizontal diffraction (edge). (d) Vertical diffraction (roof).

Figure 3.3: Visualization of the modeled physical effects on launched rays.

the widely used 2.5D data format, the building heights are specified but roof shapes are not considered.

Basically, two different classes of ray optical models are distinguished. In *ray tracing models*, the tracking of possible propagation paths starts at the receiver point and proceeds towards the transmitter. The set of possible propagation paths is limited by the maximum number of *deflection points*, i.e., points at facets where deflection effects occur. If the path loss is computed for multiple receiver points, it requires a high computational effort to determine all relevant deflection points for each receiver. For receiver points that are located nearby, however, the corresponding propagation paths are nearly identical. In such situations – or if the receiver locations are dynamic or not known in advance – it is more efficient to apply the following principle: Complementary to the ray tracing method, *ray launching algorithms* emit a finite set of rays from the transmitter in predetermined directions. They track the deflection points at facets and the physical effects on the paths through the scene. Each point that is reached by at least one path can be considered as receiver point. As the emitted rays disperse due to physical effects such as diffraction, important deflection points or even some relevant receiver points may not be reached. Therefore, the density of emitted rays has to be high enough to avoid that effect. Since a higher ray density increases the required computational effort, the multiplication of rays at deflection points is proposed in [89] to keep the overall number of rays low.

The ray launching method from [73, 90] is applied to generate the input information for the planning and control methods presented in Chapters 6 and 7. The basic idea of the underlying *Cube Oriented Ray Launching Algorithm (CORLA)* is to rasterize the considered urban environment into suitably and equally sized cubes. Physical effects at a cube are tracked if the cube belongs to a facet, see Figure 3.3 (a). The following *deflection effects* are distinguished: Reflection (R) at a facet surface, horizontal diffraction (H) at a facet edge, and vertical diffraction (V) due to deflection at rooftop edges. Diffraction effects cause the emission of a new bundle of rays into the diffraction cone, whereas reflection effects just change the angle of the arriving ray. Due to the diffrac-

tion effects and the multiple rays that are emitted at the transmitter, a receiver point t might be reached by a set of different paths $\mathcal{P}_t$. The effect-related path loss on each path $p \in \mathcal{P}_t$ depends on the number of associated deflection points $n_R(p), n_H(p), n_V(p)$ and the signal penetration at each deflection point i with angular change ϕ_i.

If the penetration terms are described as polynomials of degree K in ϕ_i, the effect related attenuation on the path p is given as

$$L^{\mathrm{dB}}_{\mathrm{CORLA}}(p) = \sum_{i=1}^{n_R(p)} \sum_{j=0}^{K} w_{R,j} \phi^j_{R,i} + \sum_{i=1}^{n_H(p)} \sum_{j=0}^{K} w_{H,j} \phi^j_{H,i} + \sum_{i=1}^{n_V(p)} \sum_{j=0}^{K} w_{V,j} \phi^j_{V,i} \ , \tag{3.5}$$

where $w_{R,j}, w_{H,j}, w_{V,j}$, $j = 0, \ldots, K$, are the polynomial coefficients that have to be calibrated in advance by a measurement based parameter estimation. As illustrated in Figure 3.3 (b) - (d), each considered effect has its own path loss characteristic, and particularly, the vertical diffraction is influenced by the height of the buildings. The superposition of all effects provides a highly accurate description of the radio conditions at receiver points. Please note that rays are not longer tracked from the deflection point on, where the overall attenuation on the path exceeds a certain (maximum) threshold.

According to [102], it is a reasonable approximation to consider only the strongest path that exists between transmitter and receiver, i.e., the signal path with the lowest path loss. With respect to this approximation and (3.5), the distance-dependent terms in (3.4) are replaced by

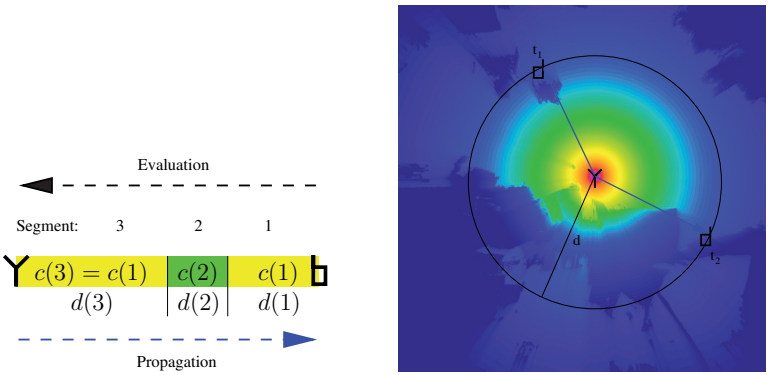
$$\gamma 10 \log_{10} d(t) + L^{\mathrm{dB}}_{\mathrm{effects}}(t) = \min_{p \in \mathcal{P}_t} \left\{ \gamma 10 \log_{10} d(p) + L^{\mathrm{dB}}_{\mathrm{CORLA}}(p) \right\} \ ,$$

where $d(p)$ denotes the length of the path to receiver point t. This approximation is not suitable for channel models that consider the signal phase information, e.g., MIMO channel models, since it neglects multipath propagation. We refer to the approach presented in [91] if multipath propagation is a required feature.

3.1.3 A Direction-Specific Land Use Based Path Loss Model

In [37], we develop a *Direction-specific Land use based Path loss model (DiLaP)* that combines principles from empirical and ray optical path loss computation. The combination counteracts the inaccuracies that can arise for the models from Section 3.1.1 due to the missing diversification on the propagation path, see Figure 3.2. Our model is intended for application to suburban/rural areas. It improves related path loss models such as the Erceg model from Section 3.1.1 by considering all land use segments – with different sizes and attenuation properties – that are passed by a straight ray from the receiver to the transmitter. The underlying principle of our model is that land use segments nearby the receiver have a strong influence on the path loss, whereas the impact of segments far away is reduced.

The set $\mathcal{C}$ contains all land use classes that we distinguish for path loss computation. In the following, we consider $\mathcal{C} = \{1 \text{ (free space)}, 2 \text{ (village)}, 3 \text{ (forest)}\}$ as illustrated



(a) Path evaluation principle.

(b) Path loss prediction according to DiLaP.

Figure 3.4: DiLaP computation for the area depicted in Figure 3.2.

in Figure 3.2. The set of land use classes is generally not limited to those three but this choice leads to good results for the investigated scenarios. In [76], we propose an approach to extract the required land use information from low-cost satellite pictures by applying dedicated classification algorithms.

Before the DiLaP computation is carried out, we first determine the $i = 1, \ldots, n(t)$ different land use segments that are intersected by the direct path from receiver t to the transmitter. Each segment i has a corresponding length (sub-distance) $d(i) > 0$ and the land use class $c(i)$. Figure 3.4 (a) sketches this information exemplarily for the receiver t_2 from Figure 3.2.

The segment information is evaluated by

$$L^{\mathrm{dB}}_{\mathrm{DiLaP}}(t) = \Delta_0 + 20\log_{10}\frac{4\pi}{\lambda_{f_c}} + \gamma_{c(1)}\,10\log_{10}\left(d(1)\right) + \sum_{i=2}^{n(t)}\left[\gamma_{c(i)}\,10\log_{10}\left(\sum_{j=1}^{i} d(j)\right) - \gamma_{c(i)}\,10\log_{10}\left(\sum_{j=1}^{i-1} d(j)\right)\right], \qquad (3.6)$$

where the parameters Δ_0, λ_{f_c} are defined according to (3.2). Each land use class $c \in \mathcal{C}$ corresponds to an individual path loss coefficient $\gamma_c \in \{\gamma_1, \gamma_2, \gamma_3\}$. The coefficients are predetermined by a parameter estimation (calibration) for the considered evaluation scenario. Figure 3.4 (a) illustrates the principle behind formula (3.6). It shows particularly the logical evaluation direction that is reversely aligned to the physical propagation direction: The path loss at the receiver is modeled as additive superposition of the segment path losses in between. For instance, the path loss contribution of segment 2 in Figure 3.4 (a) is calculated as the path loss with respect to the corresponding land use class $c(2)$ and distance $d(1) + d(2)$ to the receiver. The obtained

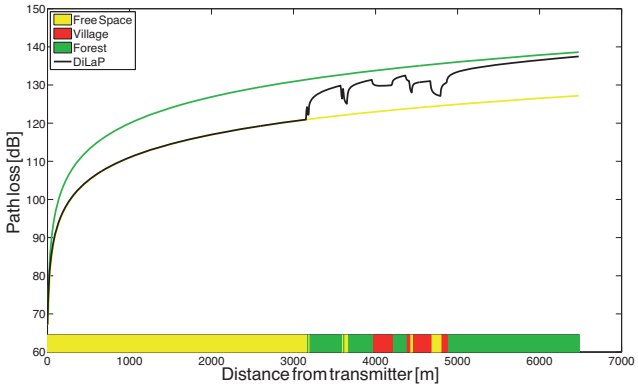


Figure 3.5: DiLaP characteristics for a straight path of increasing distance.

result is then adapted by subtracting the land use specific influence of the antecedent segment 1. The subtraction considers the land use class $c(2)$ but distance $d(1)$. Consequently, the impact of segment 2 on the overall path loss depends on its land use type and length, but particularly it depends on its distance to the receiver. Hence, the impact of segments that are located far away from the receiver is significantly smaller than the impact of segments nearby. This effect becomes clear when (3.6) is rewritten as

$$L_{\mathrm{DiLaP}}^{\mathrm{dB}}(t) = \Delta_0 + 20\log_{10}\frac{4\pi}{\lambda_{f_c}} + \gamma_{c(1)}\,10\log_{10}\left(d(1)\right) + \sum_{i=2}^{n(r)} \gamma_{c(i)}\,10\log_{10}\left(1+\frac{d(i)}{\sum_{j=1}^{i-1} d(j)}\right).$$

Furthermore, it is illustrated by Figure 3.5 where $L_{\mathrm{DiLaP}}^{\mathrm{dB}}(t)$ is computed successively for all points t on one ray that starts at the transmitter:

- The path loss according to DiLaP is upper bounded by the distance-dependent path loss for the land use class *forest* that has the largest path loss exponent. It is lower bounded by the *free-space* path loss with path loss exponent $\gamma_1 = 2$, see Table 3.2.

- Since the segment nearby the receiver point has the highest impact to the effective path loss at t, the path loss can decrease although the distance is increasing. This effect occurs in segments that cause a lower path loss than the antecedent segment on the path, e.g., in transition areas from *village* to *free-space* or from *forest* to *village*.

- The DiLaP results are asymptotic for an increasing segment length $d(i)$.

Evaluation

Our DiLaP approach is evaluated with respect to measurement data from a WiMAX measurement campaign in Regensburg, Germany. After computing the required input information by automatic land use classification of the corresponding satellite image, the DiLaP parameters are calibrated by a least-squares regression fit on a subset of measurements points. The obtained model parameters are shown in the upper left part of Table 3.2. Figure 3.4 (b) visualizes the path loss prediction results for the area from Figure 3.2 and a resolution of $6.25\,\mathrm{m}^2$ per pixel.

Approach	DiLaP	Erceg model
Model Parameter	$\Delta_0 = 3.2, \gamma_1 = 2.0$ $\gamma_2 = 2.2, \gamma_3 = 2.3$	Terrain type C $h_A = 10\,\mathrm{m}$
MSE [dB]	4.97	6.86

Table 3.2: Model parameters and evaluation results.

The DiLaP approach is compared to the Erceg model from Section 3.1.1 on a set of measurement points that are distinct from the ones that were used for DiLaP calibration. For the Erceg model, we choose the terrain type that leads to the best results in terms of the *Mean-Squared Error (MSE)* achieved on the evaluation points. Even though DiLaP does not exploit any additional information like antenna height, Table 3.2 shows that it beats the Erceg path loss model in terms of prediction accuracy. The smart direction-specific evaluation principle enables DiLaP to reflect the path loss characteristics of the underlying measuring track more precisely than the Erceg model. This property is illustrated in Figure 3.6. The improved accuracy of DiLaP comes at the cost of an increased computational complexity, mainly caused by the successive superposition of segment-wise processing. In [37], we discuss how the DiLaP runtime can be reduced by implementing the model on a parallel computer architecture.

Overall, the evaluation results demonstrate excellent path loss accuracy of our DiLaP approach and its advantages over the Erceg model. Since the path loss computation for single segments is very simple, this component might be enhanced by using more sophisticated semi-empirical models, e.g., the WINNER path loss models mentioned in Section 3.1.1.

3.2 Wireless Channel Models, Rate Computation, and Bandwidth Allocation

The level of detail that channel models have to provide depends on the particular application. If wireless networks are optimized with respect to a long-term perspective, the channel information is not required at a time scale of milliseconds, seconds, or even minutes. Hence, short-term effects like fast fading are irrelevant for tasks

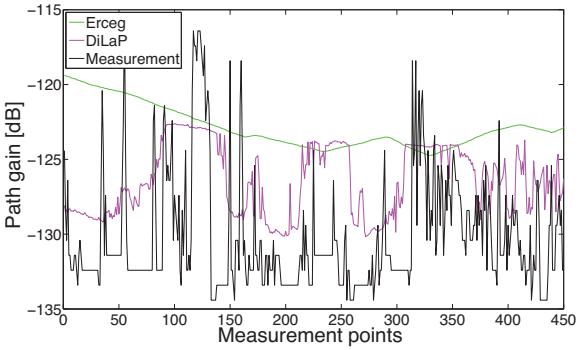


Figure 3.6: Comparison of DiLaP and the Erceg model on a measurement track.

like network dimensioning or cell site planning. They can optionally be included by a constant penalty term that serves as buffer for worst case situations. On the other hand, applications such as receiver design and dynamic system simulation require a high temporal resolution of the channel information, most likely at realtime. For this purpose, the WINNER project invented several extensions to the path loss models that were discussed in Section 3.1.1. The extensions incorporate scenario-specific (stochastic) components for fast fading and spatial components that enable MIMO channel modeling: The *Spatial Channel Model (SCM)*, its extension *SCME* (for higher carrier frequencies and transmission bandwidth), and the WINNER I and II models are state-of-the-art MIMO channel models that are widely used for LTE system simulations [74]. All WINNER channel models refer only to a small set of representative environment classes. Particularly, they are not able to consider the details of a given environment, e.g., buildings or other de facto obstacles on the signal paths. This is the motivation in [107] and [92] to combine the WINNER channel models and the ray launching method from Section 3.1.2 in order to obtain an environment-specific channel characteristic. Another recent approach that follows a similar idea is the *QuaDRiGa* channel model, which is (partly) presented in [54].

The computed channel state information and the emitted transmit power P are input for the following approach that determines the supported spectral efficiency (signal quality) e_{*t} on the link from transmitter entity $* \in \{s, k, f\}$ to user or demand node t, see Section 1.3. The corresponding results serve as input parameters for the optimization models that are proposed in the subsequent chapters.

The transmitter entities in WiMAX and LTE systems apply *adaptive modulation and coding* subject to the present link channel state and a maximum bound for the *Bit Error Rate (BER)* or *BLock Error Rate (BLER)*. Hence, the supported data rate depends on the *Signal-to-Interference and Noise Ratio (SINR)*. The LTE system specification distinguishes 16 *Channel Quality Indicators* (*CQIs*) [30]. Each CQI corresponds to a supported modulation scheme and code rate for downlink transmission, i.e., the

downlink spectral efficiency $e^{\text{CQI } i}$ can be computed in terms of bits per second per Hertz for each CQI i. The smallest non-zero spectral efficiency in present LTE systems is $e^{\text{CQI } 1} = 0.25$ [bps/Hz] for QPSK and code rate $1/8$. The largest spectral efficiency is $e^{\text{CQI } 15} = 4.8$ [bps/Hz] for 64-QAM and code rate $4/5$, both for a fixed BLER of 10^{-1}. Please note that the practically achieved spectral efficiency can slightly differ from the theoretical values due to a higher resolution in the supported code rates. Since the de facto values do not affect the presented optimization models, we keep the theoretical values for all numerical evaluations.

The *system link budget* defines what SINR is required to support a certain CQI such that the receiver can decode the data with a transport block error probability below 10% [10]. The according receiver sensitivity model typically considers thermal noise (-174 dBm/Hz) multiplied by the transmission bandwidth, the receiver noise figure (9 dB), an implementation margin (2.5 dB for QPSK, 3 dB for 16-QAM, and 4 dB for 64-QAM), and a diversity gain (-3 dB), see [94]. Additional *Quality-of-Service (QoS)* requirements can be modeled by modifying the link budget specification accordingly. Frequency-specific adaption of modulation and code rate is not supported in presently deployed system releases (8 and 9) because it does not improve the system throughput in absence of frequency-specific transmission power control [94]. The WiMAX system specification defines the system link budget in a comparable manner but distinguishes only seven transmission modes with non-zero spectral efficiency. The supported modes start with 0.25 [bps/Hz] for BPSK and code rate $1/2$ and end with 4.5 [bps/Hz] for 64-QAM and code rate $3/4$ [44].

We consider a link from transmitter $* \in \{s, k, f\}$ to user or demand node t *without interference* from other transmissions. In this case, we can select the highest CQI that is supported by the *Signal-to-Noise Ratio (SNR)* on the link and the SINR requirements from the system link budget. The parameter e_{*t} is set to the corresponding spectral efficiency of the chosen CQI. Thus, each spectral efficiency parameter is generated with respect to the applied path loss model and according to a predefined *CQI lookup table*.

In our optimization models, the link quality information is used to compute the average amount of bandwidth b_{*t} that the serving station $*$ has to allocate for transmission to demand node t. For a requested data rate r_t, the required bandwidth for downlink transmission is given by

$$b_{*t} = \frac{r_t}{e_{*t}} \tag{3.7}$$

when the location and spacing restrictions for resources in the spectrum are neglected. The sum of required bandwidth over different transmission links is always computed *overlap-free*. According to (3.7), we model the allocated bandwidth b_{*t} as a *continuous variable* although the smallest resource units that can be assigned in the considered systems have a certain minimum spacing: The smallest allocatable resource unit in WiMAX systems is a *SubCarrier (SC)*. SCs have a fixed spacing for each supported system bandwidth, e.g., 10.94 kHz for a total transmission bandwidth of 20 MHz [16]. The resource allocation in LTE systems considers a *Physical Resource Block (PRB)* as the smallest allocatable unit. Each PRB comprises 12 consecutive SCs of 15 kHz spacing, i.e., a PRB has a total spacing of 180 kHz [94]. Resource scheduling basically

allows for multiple usage of a PRB over time along different UEs. This feature is exploited in Section 7.1.1 to slice a PRB virtually into single SCs.

The total amount of resources that is required to serve the typical rate demand per link is relatively high compared to the fixed resource spacing. Therefore, clipping effects are reasonably low and we can accurately model the allocated bandwidth by a continuous variable. If the impact of clipping effects is expected to be larger, this can be modeled by an (artificially) increased rate demand, e.g., according to the robustness approach that is discussed in Section 8.3.

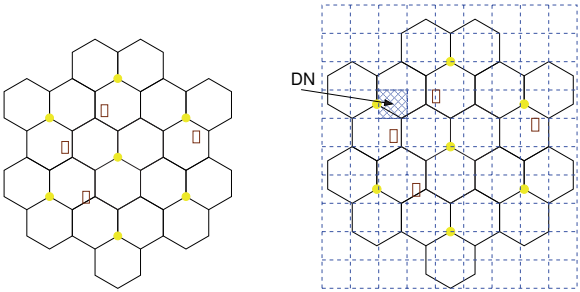
For a resource unit n with fixed bandwidth spacing b_n, the CQI lookup table can serve as *discrete rate-power function*

$$R(t,n,P_n) = \begin{cases} 0 & , \quad P_n < P_{t,n}^{\text{CQI 1}} \\ b_n e^{\text{CQI 1}} & , \quad P_{t,n}^{\text{CQI 1}} \leq P_n < P_{t,n}^{\text{CQI 2}} \\ b_n e^{\text{CQI 2}} & , \quad P_{t,n}^{\text{CQI 2}} \leq P_n < P_{t,n}^{\text{CQI 3}} \\ & \vdots \end{cases} , \tag{3.8}$$

where each power requirement is computed subject to the system link budget and the path loss dependent channel gain that t experiences on the resource unit n.

If the bandwidth allocation or the rate-power function are computed with respect to potentially interfered resources, there are basically three options to deal with it:

1. Apply (3.7) or (3.8) as above and ensure the interference-free case. This can be achieved by deploying transmitters far away from each other such that there is only marginal or no *inter-cell interference*. Please note that *Intra-cell interference* is avoided in 4G systems by the underlying OFDMA multiplexing. We apply this approach for network dimensioning in Chapter 4.
2. Compute the spectral efficiency subject to the SINR information for the related link and apply (3.7) or (3.8). SINR computation requires a full user-to-transmitter assignment and the application of dedicated bandwidth allocation algorithms. We avoid this approach for cell site planning and network control due to the high computational complexity of (optimal) bandwidth allocation, see Section 3.4.3. Instead, we utilize the approximation model that is introduced in Chapter 5.
3. Apply *Inter-Cell Interference Coordination (ICIC)*, e.g., *interference mitigation* techniques or *soft frequency reuse* [2, 110]. An according radio resource management has to organize the frequency block usage such that some blocks are aligned in the spectrum with blocks from other interfering transmitters (frequency reuse). For interference mitigation, some blocks are excluded from usage at interfering transmitters [29, 94]. Such protected frequency blocks do not suffer from interference, and hence, (3.7) or (3.8) can be applied as described above. However, also the blocked parts of the spectrum have to be considered for bandwidth allocation since they reduce the utilizable transmission bandwidth. Interference mitigation is applied as ICIC technique throughout the system simulations in Chapter 7.



(a) Considered network area with buildings (rectangles).

(b) The area is divided into equal patches.

Figure 3.7: DN generation principle.

3.3 Demand Prediction Model

We adapt the demand node (DN) concept from [101] to reduce the number of users that have to be considered in the optimization problems for the dimensioning, planning, and operation of radio networks. DNs model the spatial distribution of aggregated users as well as their joint demand attributes, e.g., their cumulative rate demand. This concept is very useful when computation time and memory are critical resources.

The abstraction from physical users is particularly reasonable for transmitter location planning and anticipative network configuration. Each DN is associated with a reference coordinate that represents all positions of the users that are covered by it. The users that are represented by the same DN should experience a similar signal quality from the network transmitters. The DN distribution as well as the corresponding demand parameters have to be chosen accurately in order to model the de facto users in the network suitably. DN information can be extracted from related data provided by network operators or it can be generated according to simulation statistics. The operator data typically includes individual forecast information that is obtained by dedicated (traffic) prediction algorithms [99]. It might also consider additional information from the marketing department that is aware of exceptional events like the launch of new services [85].

Figures 3.7 and 3.8 illustrate how DNs can be created deterministically for a certain network area that contains buildings. First, the network area shown in Figure 3.7 (a) is divided into the equal patches that are depicted in Figure 3.7 (b). Initially, each patch corresponds to a DN. Since indoor users are of special interest in many cases, buildings are modeled by own DNs that are systematically separated from the surrounding DNs. Figure 3.8 visualizes this principle for two patches that cover a building; the separation step leads to three resulting DNs.

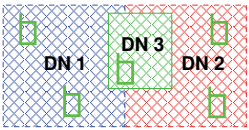


Figure 3.8: DN separation for a building.

The demand parameters that are associated with a created DN t can be generated as follows. We exemplarily choose the data rate demand r_t as target value. A snapshot of the system shall serve as representative instance to model the user behavior over the considered time period. Thus, the rate demand parameter r_t is chosen as the mean over time of the data rate distribution of aggregated users. This method is applied for network dimensioning and cell site planning in Chapter 4 and Chapter 6, respectively. We suggest to consider a higher order quantile instead of the mean if solutions have to be robust against strong demand variation.

When the DN concept is applied in a dynamic system, e.g., for optimal network operation in Chapter 7, the demand parameters can be adapted successively: We assume knowledge of the observation r_t^{past} from the previous operation cycle(s) and r_t^{future} as prediction for the next operation cycle. The *reliability indicator* $\mu \in [0, 1]$ describes the level of confidence in the prediction accuracy. The information is used to compute

$$r_t = (1 - \mu)\, r_t^{\text{past}} + \mu\, r_t^{\text{future}} \ . \tag{3.9}$$

The prediction accuracy achieved by (3.9) is tracked over the operation cycles. In case of significant differences between prediction and realization, the DN generation process is adapted accordingly.

3.4 Fundamental Problems in Radio Network Optimization

The optimization of radio networks at different stages of their life cycle is formalized by corresponding optimization problems according to Chapter 2. The underlying system structure, the technical interdependencies, and the desired optimization goals are represented by the following three basic components:

1. The *objective function* covers the KPI metrics that shall be optimized by tuning according system control parameters, i.e., the optimization variables.
2. The optimization *constraints* model interdependencies and restrictions in the system, mostly in a technical sense.
3. The *optimization input parameters* describe the spatial radio conditions for different system configurations and the distribution of user demand. The demand distribution is modeled according to the DN concept from Section 3.3.

The spatial radio conditions are computed according to the previously introduced channel models. Alternatively, in an operating system this information might be derived from system observations and receiver measurements, e.g., according to the *X-Map Estimation* approach that is proposed in [96].

The following optimization problems are closely related to the optimization models that we develop in this work.

3.4.1 The Maximal Covering Location Problem

With respect to the notation introduced in Section 1.3, we define

$$\mathcal{S} * \mathcal{T} = \{(s,t) \in \mathcal{S} \times \mathcal{T} \ : e_{st} > 0\} \tag{3.10}$$

as the set of supported supplier-DN combinations and

$$\mathcal{S}_t = \{s \in \mathcal{S} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \tag{3.11}$$

as the set of potential suppliers (serving transmitters) for demand node t. The weight for each DN is arbitrarily set to its rate demand r_t. The DN weights and the spectral efficiency information e_{st} are given as input parameters. The latter one might be computed according to Section 3.2.

The *Maximal Covering Location Problem (MCLP)* is given as

$$\max \sum_{(s,t)\in\mathcal{S}*\mathcal{T}} r_t z_{st} \tag{3.12}$$

subject to

$$\sum_{s\in\mathcal{S}_t} z_{st} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T} \tag{3.13}$$

$$z_{st} \leq y_s \ , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T} \tag{3.14}$$

$$\sum_{s\in\mathcal{S}} y_s \leq S_{max} \tag{3.15}$$

and with respect to the binary decision variables y_s, z_{st}. Supplier $s \in \mathcal{S}$ is in the solution set if $y_s = 1$. Furthermore, $z_{st} = 1$ indicates that DN t is covered by supplier s. Constraint (3.13) avoids double counting of a covered weight for the maximization of the sum weight in (3.12). The coverage of DN t by a potential supplier $s \in \mathcal{S}_t$ requires its selection in (3.14). Finally, (3.15) bounds the total number of selected suppliers to $S_{max} \in \mathbb{N}$.

Since the MCLP is a budged-constrained variant of the *Minimum Set Cover Problem (MSCP)*, see [101], it is an NP-hard optimization problem and NP-complete in its decision version [60]. Hence, there does not exist an efficient algorithm to solve the MCLP, unless P = NP. The best approximation we can hope for is a *Polynomial-*

Time Approximation Scheme (PTAS) [26], e.g., the one that is presented in [45] for a slightly modified version. Heuristic solution approaches are, for instance, the *greedy algorithm* proposed in [101], an *evolutionary algorithm* presented in [103], and the *simulated annealing* method applied in [72]. The latter two publications particularly consider variants of the MCLP that are sensitive to interference, i.e., the *Channel Assignment Problem (CAP)* for interference-limited systems is solved by an integrated approach. An optimal solution of the CAP assigns a subset of available channels to each supplier such that the number of overlapping channels is minimal for every pair of suppliers that can cover the same DNs. For frequency-division / time-division multiple access (FD/TDMA) systems like GSM networks, the channels of neighboring suppliers are separated in the frequency domain. In this case, the CAP is transformed into the fixed-spectrum *Frequency Assignment Problem (FAP)*. The FAP is known to be NP-hard since it contains the *Vertex Coloring Problem* as a subproblem [103]. An equivalent problem exists for CDMA systems where the separation is achieved by optimum spreading code assignment [14].

Recent 4G systems, such as LTE networks, apply a different approach to cope with inter-cell interference. LTE systems share the spectrum along all suppliers, i.e., all channels are used at (H)eNBs with *reuse factor* 1. The *dynamic channel assignment* takes place in the frequency domain and is carried out during network operation. This principle leads to the optimization problems that are discussed in the following.

3.4.2 User Assignment in OFDMA Systems

With respect to the notation introduced in Section 1.3, we consider a set $\mathcal{S}$ of operating suppliers (transmitters) in a 4G OFDMA network. The available bandwidth of supplier s is limited by B_s. The elements $t \in \mathcal{T}$ are interpreted as single users and each user has a rate demand r_t. Supplier s can serve the rate demand of user t if it can allocate a sufficient amount of bandwidth for t, see (3.7). If supplier s serves user t, the binary assignment indicator z_{st} is set to one. The sets $\mathcal{S} * \mathcal{T}$ and $\mathcal{S}_t$ are defined as in (3.10) and (3.11), respectively. Furthermore, we define

$$\mathcal{T}_s = \{t \in \mathcal{T} \ : (s,t) \in \mathcal{S} * \mathcal{T}\}$$

and assume that $\bigcup_{s\in\mathcal{S}} \mathcal{T}_s = \mathcal{T}$.

In the case that the required amount of bandwidth for serving user t is b_t for all potential suppliers, the *user assignment problem* can be written as the *Multiple Knapsack Problem (MKP)*

$$\max \sum_{(s,t)\in\mathcal{S}*\mathcal{T}} r_t z_{st} \tag{3.16}$$

subject to

$$\sum_{s \in \mathcal{S}_t} z_{st} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T} \tag{3.17}$$

$$\sum_{t \in \mathcal{T}_s} b_t z_{st} \leq B_s \ , \qquad \text{for all } s \in \mathcal{S} \tag{3.18}$$

and with respect to the binary decision variables z_{st}. Constraint (3.17) ensures that each user can be assigned only once and (3.18) gives the capacity constraint for each knapsack. The sum over single bandwidth requirements is overlap-free.

If

$$\sum_{t \in \mathcal{T}_s} b_t \leq B_s \tag{3.19}$$

holds for all $s \in \mathcal{S}$, the MKP has a trivial solution: Each user can be assigned to an arbitrary supplier. If (3.19) does not hold, some (or all) of the *knapsack constraints* get tight and the MKP becomes a hard combinatorial problem. The non-trivial MKP is NP-hard. Optimal solutions can be computed by branch-and-bound methods, whereas approximate solutions can be obtained by greedy algorithms or relaxation techniques [71].

The user assignment problem is an underlying component of the optimization models that we develop in the subsequent chapters. Compared to the MKP formulation from above, those user assignment problems have to cope with additional aspects. First, the required bandwidth for a user t can differ from supplier to supplier. And second, interference effects can change the required bandwidth subject to the particular assignment. The effectively required bandwidth for serving users is determined by solving the resource allocation problem that is discussed in the next section. Actually, the user assignment problem and the resource allocation problem cannot be treated separately. On the other hand, the computational complexity of every optimal integrated algorithm is too high for practical applications. For some applications, the user assignment is computed according to a suitable heuristic first, and the resource allocation problem is solved in a consecutive step.

3.4.3 Resource Allocation in OFDMA Systems

The resource allocation problem in OFDMA systems considers the assignment of subcarriers from a set $\mathcal{N}$ of available SCs to users (or DNs) t from a set $\mathcal{T}$. Furthermore, it considers power allocation P_n for each assigned subcarrier $n \in \mathcal{N}$. The total transmit power at the considered supplier is bounded by P. Any inter-symbol interference is neglected. The set of users is predetermined by a user assignment step. Each user t has a certain rate demand r_t.

For every potential assignment pair, we introduce the indicator variable $x_{tn} \in \{0, 1\}$ that is one if SC n is assigned to user t and equals zero otherwise. The *rate-power*

function $R(t,n,P_n)$ defines the achievable rate for user t on SC n subject to the transmit power P_n and the (fixed) spacing of an SC in the frequency domain. Structure and properties of $R(t,n,P_n)$ are discussed in more detail below.

The *OFDMA resource allocation* problem for *sum rate maximization* is defined as

$$\max \sum_{t\in\mathcal{T}} \sum_{n\in\mathcal{N}} R(t,n,P_n)\, x_{tn} \tag{3.20}$$

subject to

$$\sum_{n\in\mathcal{N}} R(t,n,P_n)\, x_{tn} \geq r_t \;, \qquad \text{for all } t\in\mathcal{T} \tag{3.21}$$

$$\sum_{t\in\mathcal{T}} x_{tn} \leq 1 \;, \qquad \text{for all } n\in\mathcal{N} \tag{3.22}$$

$$\sum_{n\in\mathcal{N}} P_n \leq P \;, \tag{3.23}$$

$$P_n \geq 0 \;, \qquad \text{for all } n\in\mathcal{N} \tag{3.24}$$

and with respect to the binary decision variables x_{tn} and the corresponding power allocation variables P_n. While (3.20) maximizes the overall sum rate, constraint (3.21) ensures that the (minimum) rate demand of each user is fulfilled by any feasible solution. According to (3.22), every SC can be assigned exactly to one user. Finally, the total power assigned to SCs is bounded by (3.23) and the non-negative constraint (3.24).

This formulation is consistent with the optimization model presented in [65]. It can be solved by a *dual method* approach with complexity $\mathcal{O}\big(|\mathcal{N}||\mathcal{T}|^3\big)$, see [93, 106]. For the special case of equal rate demands, the computational complexity can be reduced to $\mathcal{O}\big(|\mathcal{N}||\mathcal{T}|\big)$, see [65]. These results hold for a continuous and concave rate-power function

$$R(t,n,P_n) = \log_2\left(1+\mathrm{SNR}_{t,n}\Gamma_{t,n}\right) \tag{3.25}$$

with fitting factor $\Gamma_{t,n} > 0$ and

$$\mathrm{SNR}_{t,n} = \frac{P_n g_{tn}}{\sigma_n^2} \;. \tag{3.26}$$

The *channel gain* g_{tn} is basically determined by the path loss according to (3.3). Additionally, it can include the channel characteristics that were discussed in Section 3.2. The constant σ_n^2 denotes the noise power on subcarrier n.

The complexity bounds do not hold any longer if we consider a discrete rate-power function as introduced in Section 3.2. If the continuous domain of power variables is replaced by a discrete set of minimum power requirements and the rate-power function can only achieve discrete rates according to (3.8), the resource allocation problem becomes a binary decision problem, i.e., an ILP. In [46], it is shown that this binary decision problem can be interpreted as *Multiple-Choice Knapsack Problem (MCKP)*, which is NP-hard to solve. However, the authors propose a heuristic *relaxation and*

rounding approach that determines approximate solutions efficiently. Particularly, it provides a bound on the integrality gap that vanishes asymptotically for an increasing granularity of subcarrier spacing. The latter property is highly related to the simplified bandwidth computation in (3.7).

So far, the resource allocation problem has been considered for a single supplier (single-cell). Thus, any implementation of (3.25) as well as any discrete rate-power function depends on the SNR according to (3.26). For the multiple supplier (multi-cell) case, however, the power-rate function depends on the SINR. For a set $\mathcal{S}$ of multiple interfering suppliers, a supplier s provides the

$$\mathrm{SINR}_{s,t,n} = \frac{P_{sn} g_{stn}}{\sum\limits_{s' \in \mathcal{S},\, s' \neq s} P_{s'n} g_{s'tn} + \sigma_n^2}$$

to user t on subcarrier n. Here, the channel gains and the noise power are supplier-specific. Since the resource allocation at interfering suppliers s' affects the denominator, computation of the SINR requires a full user-to-supplier assignment. Furthermore, it requires knowledge of the subcarrier and power assignment at all other suppliers. These properties make optimal OFDMA resource allocation a computationally hard combinatorial problem if inter-cell interference has to be considered [69].

Please note that presently deployed LTE networks, i.e., Release 8 and 9 systems, do not support frequency-dependent power control [94]. In this case, the constraints (3.23) and (3.24) are dropped and the power variables P_n are replaced by constant power parameters P_s. What remains is the combinatorial problem of assigning SCs to the users subject to the interdependencies between interfering transmitters. If the constant power parameters P_s are related to optimal network configuration, the user assignment problem and the resource allocation problem are immanent subproblems. A model for optimal network configuration is presented in Chapter 7.

4 Dimensioning of Multihop Networks

The following profit-related approach for the dimensioning of multihop WiMAX networks is inspired by the economical perspective on radio network planning and operation in [84] and [85]. Multihop WiMAX networks are specified according to the IEEE 802.16j standard [83]. The proposed concept for dealing with the multi-objective nature of the optimization problem is transferable to other radio technologies and network topologies. In fact, the economically driven deployment approach is also applied for cell site planning of LTE HetNets in Chapter 6.

Parts of this chapter have already been published in [35].

4.1 System Model

Network dimensioning means the joint selection of an arbitrary number of WiMAX base stations (BSs) and relay stations (RSs) from the sets of *deployment candidates* $\mathcal{S}$ and $\mathcal{K}$, respectively. The selection process is carried out with respect to the input parameters and optimization variables introduced in Section 1.3 and the additional parameters given in Table 4.1. We consider downlink data transmission and apply the demand node model from Section 3.3 to describe the (spatial) distribution of user rate demands r_t.

Symbol & domain	**Description**
c_s, $c_k \in \mathbb{R}_{\geq 0}$	Deployment cost for BS, RS.
$\lambda_{\text{RATE}} \in \mathbb{R}_{\geq 0}$	Sum rate weighting factor for objective scalarization.
$e_{\text{MIN}} \in \mathbb{R}_{>0}$	Minimum required spectral efficiency to establish a transmission link.
$d_{ss'} \in \mathbb{R}_{\geq 0}$	Distance between BSs $s, s' \in \mathcal{S}$.
$\mathcal{G}$	Conflict graph, representing the set of conflicting BS pairs.
$z_{sk} \in \{0, 1\}$	Binary decision variable indicating the backhaul connection from RS k to donor BS s.
$b_{sk} \in \mathbb{R}_{\geq 0}$	Amount of allocated bandwidth on the supply link from donor BS s to RS k.

Table 4.1: Additional input parameters (upper part) and variables (lower part) for dimensioning two-hop WiMAX networks, complementing Table 1.1.

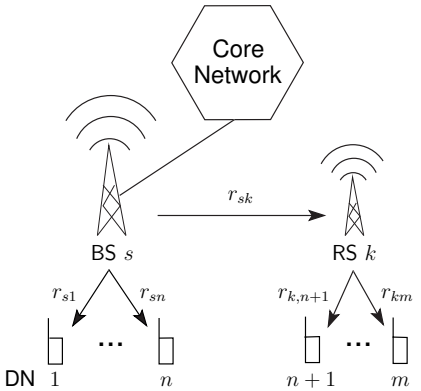


Figure 4.1: DL transmission model for non-cooperative two-hop relaying.

Each demand node (DN) $t \in \mathcal{T}$ can be served either by a deployed BS or a deployed RS. The deployment of a BS or RS is associated with cost c_s and c_k, respectively. We assume two-hop *non-cooperative relaying* in *layer three (L3)* mode, i.e., each RS receives the user data from its donor BS via radio link and forwards the data to the assigned users. Basically, L3 RSs have comparable transmission functionalities as a BS. Their capacity, however, is limited by the achievable rate on the radio link from the donor BS. Since L3 RSs do not have a wired backbone connection, they are characterized by a simpler structure and less cost than BSs. Due to their fixed and usually exposed location, we assume RSs to experience a much higher signal quality from their donor BS than an equally located user entity.

Figure 4.1 illustrates an exemplary network topology, where one RS and one BS serve m users. The overall supportable rate from RS k to its assigned users is bounded by the rate from its donor BS s, i.e.,

$$\sum_{t=n+1}^{m} r_{kt} \leq r_{sk} \ . \tag{4.1}$$

According to Section 3.2, the supportable rate from BSs to RSs and DNs is determined by the spectral efficiency on the corresponding radio links and the allocated bandwidth. Thus, BS s and RS k have to allocate enough bandwidth to serve the DN rate demands. Since the transmission spectrum in the OFDMA system is shared and limited, it holds for Figure 4.1 that

$$\sum_{t=1}^{n} b_{st} + b_{sk} \leq B_s \quad \text{and} \quad \sum_{t=n+1}^{m} b_{kt} \leq B_k \ . \tag{4.2}$$

The considered network does not suffer from any intra-cell interference if OFDMA is used as multiplex scheme. For the purpose of network dimensioning, we apply simple semi-empirical path loss models from Section 3.1.1. Hence, it is sufficient to ensure a minimum distance between deployed BSs to ease inter-cell interference on the macrocell layer. The corresponding minimum distance constraints define the *conflict graph* $\mathcal{G}$. Each element $(s, s') \in \mathcal{G}$ denotes a tuple of BSs that mutually interfere. Cross-layer interference between BSs and RSs is mitigated by applying *out-band relaying*, i.e., the transmission from RSs to DNs is separated from the BS transmission band in the frequency spectrum (type 1a relay). We do not consider any further inter-cell interdependencies such as handover aspects or resource coordination between cells.

4.2 Optimization Model

Similarly to (3.10) and (3.11), we define

$$\begin{aligned}
\mathcal{S} * \mathcal{T} &= \{(s,t) \in \mathcal{S} \times \mathcal{T} \ : e_{st} \geq e_{\text{MIN}}\} \ , \\
\mathcal{K} * \mathcal{T} &= \{(k,t) \in \mathcal{K} \times \mathcal{T} \ : e_{kt} \geq e_{\text{MIN}}\} \ , \\
\mathcal{S} * \mathcal{K} &= \{(s,k) \in \mathcal{S} \times \mathcal{K} \ : e_{sk} \geq e_{\text{MIN}}\} \ , \\
\mathcal{S}_t &= \{s \in \mathcal{S} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \ , \quad \mathcal{K}_t = \{k \in \mathcal{K} \ : (k,t) \in \mathcal{K} * \mathcal{T}\} \ , \\
\mathcal{T}_s &= \{t \in \mathcal{T} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \ , \quad \mathcal{T}_k = \{t \in \mathcal{T} \ : (k,t) \in \mathcal{K} * \mathcal{T}\}
\end{aligned}$$

to exclude decision variables and constraints that are irrelevant due to insufficient link quality. This restriction is not necessarily required since high-end optimization engines like CPLEX remove irrelevant terms in a preprocessing step within milliseconds.

The following MIP selects BSs and RSs from the candidate sets $\mathcal{S}$ and $\mathcal{K}$ for joint deployment in a WiMAX two-hop relay network. A compact representation of the optimization problem is given in Table 4.2. Optimal solutions of this MIP provide a bandwidth-limited *Profit MAXimization (ProMAX)* and serve as input for network dimensioning decisions. The return-on-invest ratio is typically a major criterion for such decisions.

The objective

$$\max \left\{ \lambda_{\text{RATE}} \sum_{t \in \mathcal{T}} r_t^{\text{eff}} - \left(\sum_{s \in \mathcal{S}} c_s y_s + \sum_{k \in \mathcal{K}} c_k y_k \right) \right\} \tag{4.3}$$

optimizes the trade-off between the achieved sum rate and the costs for deployed BSs and RSs. The optimization is carried out with respect to the optimization variables $y_s, y_k, z_{st}, z_{kt}, z_{sk}, b_{st}, b_{kt}, b_{sk}$ and the auxiliary variables z_t, r_t^{eff}.

Applying the scalarization approach from Section 2.3.1 to cope with this MOO problem, the sum rate weighting factor (exchange rate) λ_{RATE} is chosen as follows: The network operator computes the *Average Revenue Per Unit (ARPU)*, i.e., the average quotient of user revenue and cumulated data rate over a fixed time period. The user revenue typically covers basic fee, data fee, and additional services fees.

$$\max \left\{ \lambda_{\text{RATE}} \sum_{t \in \mathcal{T}} r_t^{\text{eff}} - \left(\sum_{s \in \mathcal{S}} c_s y_s + \sum_{k \in \mathcal{K}} c_k y_k \right) \right\}$$

subject to

$z_t = \sum_{s \in \mathcal{S}_t} z_{st} + \sum_{k \in \mathcal{K}_t} z_{kt} \leq 1$,	for all $t \in \mathcal{T}$
$z_{st} \leq y_s$,	for all $(s,t) \in \mathcal{S} * \mathcal{T}$
$z_{kt} \leq y_k$,	for all $(k,t) \in \mathcal{K} * \mathcal{T}$
$z_{sk} \leq y_s$,	for all $(s,k) \in \mathcal{S} * \mathcal{K}$
$b_{kt} \leq \frac{r_t}{e_{kt}} z_{kt}$,	for all $(k,t) \in \mathcal{K} * \mathcal{T}$
$\sum_{t \in \mathcal{T}_k} b_{kt} \leq B_k$,	for all $k \in \mathcal{K}$
$\sum_{s \in \mathcal{S}} z_{sk} = y_k$,	for all $k \in \mathcal{K}$
$b_{sk} \leq B_s z_{sk}$,	for all $(s,k) \in \mathcal{S} * \mathcal{K}$
$\sum_{t \in \mathcal{T}_k} e_{kt} b_{kt} \leq \sum_{s \in \mathcal{S}} e_{sk} b_{sk}$,	for all $k \in \mathcal{K}$
$b_{st} \leq \frac{r_t}{e_{st}} z_{st}$,	for all $(s,t) \in \mathcal{S} * \mathcal{T}$
$\sum_{t \in \mathcal{T}_s} b_{st} + \sum_{k \in \mathcal{K}} b_{sk} \leq B_s$,	for all $s \in \mathcal{S}$
$y_s + y_{s'} \leq 1$,	for all $(s,s') \in \mathcal{G}$
$r_t^{\text{eff}} = \sum_{s \in \mathcal{S}_t} e_{st} b_{st} + \sum_{k \in \mathcal{K}_t} e_{kt} b_{kt} \geq r_t^{\text{MIN}} z_t$,	for all $t \in \mathcal{T}$

Table 4.2: MIP formulation of the bandwidth-limited Profit MAXimization (ProMAX) problem for techno-economical network dimensioning.

In case that the ARPU varies significantly for different mobile services, the sum rate in (4.3) can be split up into linear terms

$$\sum_{m \in \mathcal{M}} \lambda_{\text{RATE}}^m \sum_{t \in \mathcal{T}_m} r_t^{\text{eff}}$$

that depend on the set of different mobile services $\mathcal{M}$, the corresponding partition $\{\mathcal{T}_m\}_{m \in \mathcal{M}}$ of $\mathcal{T}$, and the service-specific exchange rates λ_{RATE}^m. The considered ARPU [€/kbps] is normalized over the same time interval as the considered costs. Thus, objective (4.3) describes the revenue minus costs when λ_{RATE} is chosen as the ARPU. In economics, this term is called *profit*.

The maximization in (4.3) is carried out subject to the following constraints. The inequality

$$z_t = \sum_{s \in \mathcal{S}_t} z_{st} + \sum_{k \in \mathcal{K}_t} z_{kt} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T} \tag{4.4}$$

ensures that DNs are assigned to at most one BS or one RS. The variable z_t indicates whether DN t is assigned to any station or not. Since all assignments require the deployment of the related supplying station,

$$z_{st} \leq y_s,\ z_{kt} \leq y_k,\ z_{sk} \leq y_s \ , \text{ for all } (s,t) \in \mathcal{S} * \mathcal{T},\ (k,t) \in \mathcal{K} * \mathcal{T},\ (s,k) \in \mathcal{S} * \mathcal{K} \tag{4.5}$$

must hold. If DN t is assigned to RS k, i.e., $z_{kt} = 1$, the RS allocates the bandwidth

$$b_{kt} \leq \frac{r_t}{e_{kt}} z_{kt} \ , \qquad \text{for all } (k,t) \in \mathcal{K} * \mathcal{T} \tag{4.6}$$

according to (3.7). The amount is bounded by the bandwidth that is required to serve the full rate demand r_t. The overall allocated bandwidth at each RS has to stay below the available bandwidth, i.e.,

$$\sum_{t \in \mathcal{T}_k} b_{kt} \leq B_k \ , \qquad \text{for all } k \in \mathcal{K} \ . \tag{4.7}$$

As mentioned in the description of the system model, each deployed RS needs to be assigned to a donor BS by

$$\sum_{s \in \mathcal{S}} z_{sk} = y_k \ , \qquad \text{for all } k \in \mathcal{K} \ . \tag{4.8}$$

The donor BS has to allocate enough bandwidth for the corresponding supply link

$$b_{sk} \leq B_s z_{sk} \ , \qquad \text{for all } (s,k) \in \mathcal{S} * \mathcal{K} \tag{4.9}$$

such that it provides enough capacity for all the forwarded user data, i.e.,

$$\sum_{s \in \mathcal{S}} e_{sk} b_{sk} \geq \sum_{t \in \mathcal{T}_k} e_{kt} b_{kt} \ , \qquad \text{for all } k \in \mathcal{K} \ . \tag{4.10}$$

Each deployed BS can serve itself any assigned DNs by allocating the bandwidth

$$b_{st} \leq \frac{r_t}{e_{st}} z_{st} \ , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T} \tag{4.11}$$

as long as the overall allocated bandwidth – for RS links and DN links– does not exceed the total available bandwidth, i.e.,

$$\sum_{t \in \mathcal{T}_s} b_{st} + \sum_{k \in \mathcal{K}} b_{sk} \leq B_s \ , \qquad \text{for all } s \in \mathcal{S} \ . \tag{4.12}$$

The inequality

$$y_s + y_{s'} \leq 1 \ , \qquad \text{for all } (s, s') \in \mathcal{G} \ , \tag{4.13}$$

excludes BS pairs from a joint deployment if they are element of the predefined conflict graph $\mathcal{G}$. The conflicting BS pairs are predefined according to an interference criterion. If the interference-related conflict between BSs is exclusively determined by the distance between the corresponding site locations, (4.13) can be replaced by

$$(d_{ss'} + d_{\text{MIN}})\,(y_s + y_{s'}) \leq d_{\text{MIN}} + 3d_{ss'} \ , \qquad \text{for all } s, s' \in \mathcal{S} \ , \tag{4.14}$$

where d_{MIN} is a suitable minimum distance parameter. In this case, the generation of the conflict graph $\mathcal{G}$ is dispensable.

So far, a DN can be assigned to any BS or RS that provides a sufficient link quality. The supplying station allocates bandwidth according to (4.6) or (4.11) and serves the DN with an effective rate $r_t^{\text{eff}} \in [0, r_t]$. Solutions that guarantee a certain QoS level are obtained by a minimum rate requirement $r_t^{\text{MIN}} \leq r_t$ for every assigned DN, i.e.,

$$r_t^{\text{eff}} = \sum_{s \in \mathcal{S}_t} e_{st} b_{st} + \sum_{k \in \mathcal{K}_t} e_{kt} b_{kt} \geq r_t^{\text{MIN}} z_t \ , \qquad \text{for all } t \in \mathcal{T} \ . \tag{4.15}$$

Together with (4.7) and (4.12) this constraint causes a multiple knapsack problem if $r_t^{\text{MIN}} > 0$ and the load limit of (some) stations is exceeded. Compared to the MKP from Section 3.4.2, the required bandwidth for an assignment is not equal for all potential suppliers but depends on the signal quality of the particular station.

If $r_t^{\text{MIN}} = r_t$ for all $t \in \mathcal{T}$, the ProMAX problem can be rewritten in a simplified form as the *Full Rate Profit MAXimization (FR-ProMAX)* problem. The MIP for the FR-ProMAX problem is shown in Table 4.3.

4.3 Concept Validation

A representative application of the ProMAX problem(s) from above demonstrates the properties of our dimensioning approach. Exemplary test cases are generated according to the system setup described in Table 4.4. Table 4.5 shows the considered mobile services and their corresponding rate demands. For each DN, one requested mobile service is randomly chosen according to the service proportions from Table 4.5. The related rate demand is computed randomly and uniformly distributed over the service-specific rate interval. BSs, RSs, and DNs are uniformly distributed over the area. Out-band relaying is performed in a separate frequency spectrum, which provides half of the BS bandwidth. The considered RS footprint is very small compared to the BS coverage area since RSs are mounted at a much lower height and they apply less transmission power. Any deployed BS is associated with monthly *OPerational EXpenditures (OPEX)* of 2500 € covering site rental, leased line rental, air-conditioning, and maintenance [33]. RSs can be mounted much simpler than BSs, e.g., on top of traffic

$$\max \left\{ \lambda_{\text{RATE}} \sum_{t \in \mathcal{T}} r_t z_t - \left(\sum_{s \in \mathcal{S}} c_s y_s + \sum_{k \in \mathcal{K}} c_k y_k \right) \right\}$$

subject to

$z_t = \sum_{s \in \mathcal{S}_t} z_{st} + \sum_{k \in \mathcal{K}_t} z_{kt} \leq 1$,	for all $t \in \mathcal{T}$
$\sum_{t \in \mathcal{T}_k} \frac{r_t}{e_{kt}} z_{kt} \leq B_k y_k$,	for all $k \in \mathcal{K}$
$\sum_{s \in \mathcal{S}} z_{sk} = y_k$,	for all $k \in \mathcal{K}$
$b_{sk} \leq B_s z_{sk}$,	for all $(s,k) \in \mathcal{S} * \mathcal{K}$
$\sum_{t \in \mathcal{T}_k} r_t z_{kt} \leq \sum_{s \in \mathcal{S}} e_{sk} b_{sk}$,	for all $k \in \mathcal{K}$
$\sum_{t \in \mathcal{T}_s} \frac{r_t}{e_{st}} z_{st} + \sum_{k \in \mathcal{K}} b_{sk} \leq B_s y_s$,	for all $s \in \mathcal{S}$
$y_s + y_{s'} \leq 1$,	for all $(s,s') \in \mathcal{G}$

Table 4.3: MIP formulation of the Full Rate ProMAX (FR-ProMAX) problem.

lights, and they do not have a wired backbone connection. Hence, RSs account for a significantly reduced OPEX of 500 €. Since network dimensioning is carried out with respect to long-term considerations, we ignore the *CAPital EXpenditures (CAPEX)* for deploying the transmitter equipment. However, CAPEX can be additionally considered in two variants. First, by an additional budget constraint in the ProMAX problem. And second, by apportioning the CAPEX among the OPEX subject to the intended lifetime of the network.

Neglecting the DN link penetration for links from BSs to RSs, the signal strength of the supply links from BSs to RSs is 10 dB higher than the signal to an equally located DN. Furthermore, the path loss model for BS-to-RS transmission provides an extra gain compared to the DN related path loss. This is justified by the assumption that RSs can be deployed in locations that are beneficial for the BS-to-RS signal propagation. Table 4.6 specifies the SINR requirements for the discrete set of achievable spectral efficiencies. DNs can be served at seven different transmission rates that depend on the individual DN position in the BS or RS coverage area. Due to the lower transmission power and the applied path loss profile, RSs provide a small but high rate coverage area, see Figure 4.2. In this sense, RSs complement the conventional macrocell layer by bridging local coverage holes or improving the locally available data rate (capacity) [104, 17]. Inter-macrocell interference can be neglected by enforcing a minimum distance of $d_{\text{MIN}} = 500\,\text{m}$ between deployed BSs.

System parameter	Setting
Area size	1750 m × 1750 m
Number of BS / RS candidates	15 / 45, uniformly distributed
Number of DNs	650, uniformly distributed
BS / RS carrier frequency	2 GHz / 2.5 GHz
BS / RS transmission bandwidth	20 MHz / 10 MHz
BS / RS monthly cost	2500 € / 500 € (OPEX)
BS / RS / DN reference height	25 m / 5 m / 1.5 m
BS / RS Tx power	46 dBm / 24 dBm
BS / RS / DN antenna gain	14 dBi / 9 dBi / 0 dBi
BS / RS / DN noise figure	5 dB / 8 dB / 8 dB
Path loss $BS \downarrow DN$ / $BS \downarrow RS$ / $RS \downarrow DN$	WINNER II C2 NLOS / C1 NLOS / B1 [53]
Link penetration, worst case assumption	40 dB (shadowing, fast fading, etc.)
DN link penetration	10 dB (wall penetration for indoor users)
Minimum inter-site distance (BSs)	500 m

Table 4.4: Considered system parameters for network dimensioning.

Mobile service	Proportion	Rate demand [kbps]
Interactive real time	20 %	128 – 2000
Video	20 %	128 – 1000
FTP	10 %	64 – 1000
Web	20 %	64 – 512
VoIP	30 %	64

Table 4.5: Considered mobile services [78].

Each assigned DN is served by its supplying BS or RS up to the requested data rate. Assuming the users to request mobile services on 20 days per month for five hours at a charge of 0.05 €/MB, the according sum rate weighting factor (monthly ARPU) is

$$\lambda_{\mathrm{RATE}} = \frac{20 \cdot 5 \cdot 3600\,\mathrm{s}}{8 \cdot 1000} \cdot 0.05\,€/\mathrm{MB} = 2.25\,€/\mathrm{kbps}\,. \tag{4.16}$$

Please note that the denominator in (4.16) converts megabyte to kilobit, which is the reference unit for the rate specifications in Table 4.5. For $\lambda_{\mathrm{RATE}} = 2.25\,€/\mathrm{kbps}$, a VoIP DN yields a monthly revenue of 144 €. This seems reasonable when a DN models five to seven users. In consequence, a single BS or RS is already profitable if it can serve at least 18 and 4 VoIP DNs, respectively.

The ProMAX problem from Table 4.2 is applied for the planning scenario depicted in Figure 4.2 and $r_t^{\mathrm{MIN}} = 0$. We utilize the CPLEX MIP solver [51] to compute optimal solutions. The resulting network consists of 7 deployed BSs and 22 deployed RSs, which serve on average 50 DNs and 3 DNs, respectively. However, not all of the DNs are covered. According to Table 4.5, the average DN rate demand is 456 kbps, and

Rate ID	Modulation	Code rate	Spectral efficiency [bps/Hz]	Receiver SINR [dB]
0	out of range			
1	BPSK	1/2	0.5	3.0
2	QPSK	1/2	1.0	6.0
3	QPSK	3/4	1.5	8.5
4	16-QAM	1/1	2.0	11.5
5	16-QAM	3/4	3.0	15.0
6	64-QAM	2/3	4.0	19.0
7	64-QAM	3/4	4.5	21.0

Table 4.6: SINR requirements according to the WiMAX system link budget specification [52, 44], BER $= 10^{-6}$.

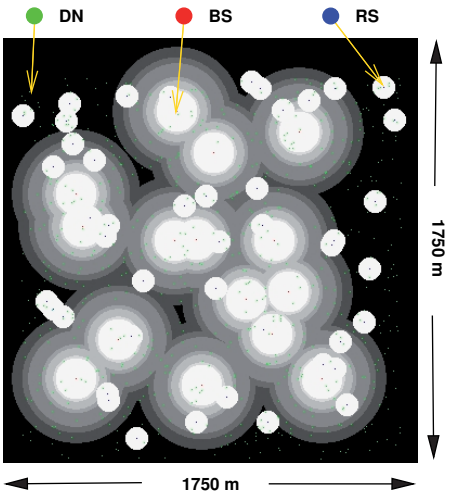


Figure 4.2: Considered BS candidates, RS candidates, and DNs. The spectral efficiency in the coverage area is visualized by a grayscale coloring.

hence, the overall revenue in (4.3) is much higher than the BS and RS expenses. It turns out that the total number of deployed BSs is limited by the minimum distance constraint (4.14): If the minimum required distance is lowered to $d_{\text{MIN}} = 250\,\text{m}$, the resulting network has 10 deployed BSs that are still profitable. The minimum distance requirement is also the reason why the number of deployed BSs does not change for a variation of the sum rate weighting factor λ_{RATE} in Figure 4.3. The number of deployed BSs stays constant whereas the number of deployed RSs increases with a rising ARPU. The maximum of 29 deployed RSs is reached for $\lambda_{\text{RATE}} = 8\,€/\text{kbps}$. In this (extreme) case, one served low rate DN is enough to achieve a positive profit at a deployed RS. Generally, we observe that profitable RSs are selected for deployment if

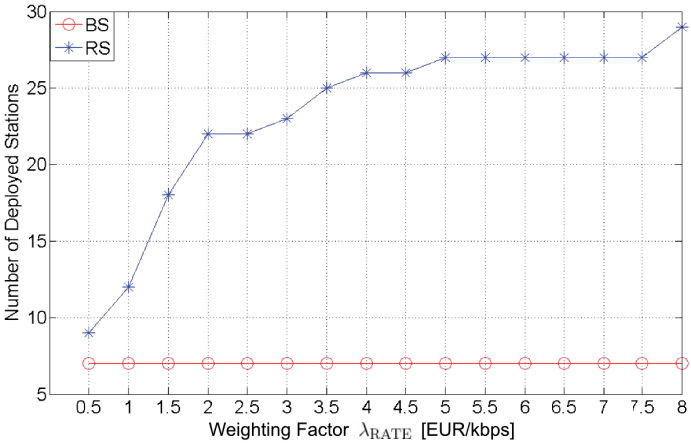


Figure 4.3: Dimensioning results for an increasing scalarization factor λ_{RATE}.

they provide coverage at the BS cell border or if they improve the achievable spectral efficiency due to an increased signal strength.

Solving the FR-ProMAX problem from Table 4.3 for planning scenarios that are generated according to the same system setup leads to almost identical dimensioning results. This outcome is not surprising since the considered transmission bandwidth is reasonably large and the knapsack problem, hence, does not affect the dimensioning problem. The knapsack problem becomes a serious issue when the rate intervals for interactive real time service and video applications in Table 4.5 are increased by a factor of five. The corresponding solutions reveal the following effects: First, the assignment of DNs gets more sensitive to their corresponding rate demand. Compared to the solutions for the ProMAX problem, more DNs with low rate demand are served. Second, the knapsack constraint (4.12) shrinks the capacity of RS backbone connections. It becomes a trade-off decision for BSs if it is more efficient to serve DNs or to provide bandwidth to RSs that have a better signal to surrounding DNs. In consequence, the number of deployed RSs decreases by roughly 50 %. And third, the knapsack problem increases the solver runtime, which typically ranges from a few minutes to one hour depending on the particular problem instance. Compared to the ProMAX problem, solving the FR-ProMAX problem for the regular rate requirements from Table 4.5 increases the average runtime by a factor of two. The runtime grows by more than one magnitude if the rate demand is increased and the knapsack constraints become tight.

4.4 Summary

While there exist many studies on the general impact of relaying to network coverage and network capacity, e.g., [104] and [17], only a few works consider the problem of *joint* BS and RS deployment. For instance, [80] investigates the placement of one RS in a single macrocell. In [64], multiple RSs are deployed but the perspective is still restricted to a single macrocell environment. An approach for joint BS and RS deployment is presented in [70]. Although the maximization of the network capacity is comparable to our approach, this work does not assume any restrictions for the available bandwidth at BSs and RSs.

Compared to the these works, our approach allows for the joint deployment of multiple BSs and multiple RSs subject to shared and limited bandwidth resources. Particularly, our optimization model applies an economically motivated approach to align the conflicting objectives of sum rate and cost in a scalar-valued overall objective. The applied simplifications to cope with interference (minimum distance constraints, out-band relaying) as well as the usage of simple semi-empirical path loss models are sufficient for the purpose of network dimensioning. Accurate network planning and network configuration, however, require a higher precision in both components. Therefore, we introduce a suitable low complexity interference approximation model in the next chapter. This approximate model is combined with ray optical path loss predictions for optimal cell site planning and network configuration in Chapter 6 and Chapter 7, respectively.

5 Interference Approximation in LTE Heterogeneous Networks

LTE systems apply full *frequency reuse* over all cells, i.e., all cells in the network share the same frequency spectrum [94]. LTE transmitters can successfully serve users or *User Equipment (UE)* even at very low SINR ($< -5\,\mathrm{dB}$). However, the achievable sum rate will be significantly degraded if too many UEs are served at low SINR. In heterogeneous networks, low power femtocells can severely suffer from interference of high power macrocell transmissions. Consequently, the SINR can become very poor for shared resources in certain parts of the HetNet. It is the task of the system RRM to assign users to supplying cells and to allocate resources in the spectrum such that the SINR is sufficient to serve the user demand. In Section 3.4.2 and Section 3.4.3 we have seen that user assignment and resource allocation are difficult problems. Furthermore, the SINR requirements for the related (discrete) rate-power-function have no linear representation, which is a prerequisite for the incorporation into our MIP-based optimization models. Therefore, we develop an approximate model that estimates the overall required bandwidth for serving users subject to inter-cell and cross-tier interference. The following aspect motivates the underlying principle of our approximate model.

The LTE standard defines several features that support the application of ICIC in the network. Interference mitigation was introduced as an ICIC technique at the end of Section 3.2. Basically, dedicated resources are protected from interference by blocking them at potentially interfering transmitters [29]. Please note that the LTE standard does not define what particular ICIC techniques should be applied in the network, i.e., the interference management is customizable. However, we assume that the application of any ICIC technique causes a certain penalty and that this penalty can be expressed in terms of additional bandwidth consumption at some of the coordinated transmitters. For interference mitigation, for instance, the penalty is given by the unused (protected) resources at the interference mitigating transmitters.

The basic idea of the developed approximation model is to consider bandwidth that is allocated for transmission to users separately from protected bandwidth and to estimate both terms subject to the SNR.

5.1 System Model

We consider two-tier LTE HetNets, i.e., a network topology with eNodeB related *macrocells (MCs)* and user-operated Home eNodeBs that are deployed indoors. We also refer to the latter ones as *femtocells (FCs)*. The wildcard symbol $a \in \{s, f\}$ for an (H)eNB refers to an arbitrary MC s or FC f. All cells transmit in the same frequency spectrum and FCs are operated according to an *open access policy*, i.e., any user is allowed to connect to any FC in the network. FCs provide coverage of 10 to 50 meters for stationary or low-mobility users in offices or at home. They are connected to the network backhaul via optical fiber or DSL [15]. The considered network topology is sketched in Figure 5.1.

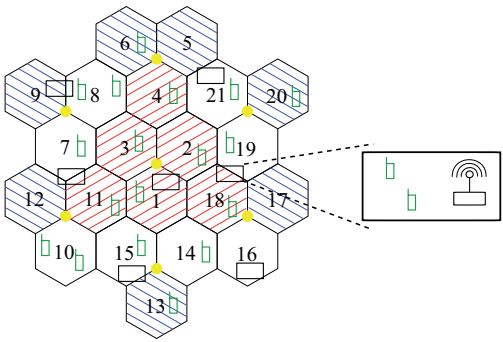


Figure 5.1: Two-tier LTE HetNet topology with hexagonal macrocell layout and indoor femtocell transmitters. We distinguish *inner macrocells* (／) from *outer macrocells* (＼).

Since LTE systems apply OFDMA, the HetNet has not to cope with intra-cell interference. However, inter-cell interference can arise and *co-channel deployment* of femtocells can cause *cross-tier interference* between eNBs and HeNBs. Cross-tier interference typically leads to an increased cardinality of problem instances for HetNet optimization when macrocells and femtocells allocate resources in the same spectrum.

The considered HetNet applies interference mitigation as ICIC technique. The symbol b_a^{itf} denotes the amount of bandwidth that (H)eNB a has to block (protect) for mitigating interference to other cells. The overall amount of required (consumed) bandwidth at (H)eNB a is denoted by b_a. The overall amount comprises the blocked bandwidth b_a^{itf} as well as the bandwidth b_a^{trm} that is allocated for transmission to users.

5.2 Interference Approximation Model

The following *Interference Approximation Model (IAM)* partitions the bandwidth that is occupied (consumed) at a transmitter entity into the amount b^{trm} for data transmission and into the amount b^{itf} that is considered as ICIC penalty term. Both terms are computed subject to the SNR on the related radio links. The total bandwidth consumption of an (H)eNB a is modeled as linear superposition

$$b_a = b_a^{\text{trm}} + b_a^{\text{itf}} \ , \tag{5.1}$$

where the component b_a^{trm} is the overlap-free sum of required bandwidth for all transmission links. The required bandwidth for each single transmission link is computed according to (3.7). The penalty term b_a^{itf} is computed as linear combination over all interfered transmission links. The allocated resources of a transmission link from (H)eNB a' to user t are interfered by (H)eNB a proportionally to the *interference impact factor* $q_{aa't} \in [0, 1]$. Thus, for eNB s it is

$$b_s^{\text{itf}} = \sum_{\substack{(s',t) \in \mathcal{S} * \mathcal{T}, \\ s' \neq s}} q_{ss't} b_{s't} \ . \tag{5.2}$$

The interference impact factor basically depends on the SNR difference on the link and the frequency reuse potential in the interfering macrocell. It can be modeled as

$$q_{ss't} = \beta_s \min \left\{ \frac{e_{st}}{e_{s't}} , 1 \right\} \ , \tag{5.3}$$

where the *frequency reuse factor* β_s particularly depends on the macrocell topology. All non-interfering neighbors of a certain cell can allocate the same resources but the cell has to block those resources only once to guarantee an interference free situation. In this sense, the reuse potential in the cell can be larger than one. This is equivalently modeled by a frequency reuse factor smaller than one. The *inner cells* in Figure 5.1 can reuse the blocked resources for several neighboring cells, whereas the *outer cells* do not have that much reuse opportunities. Hence, outer cells generally have a higher frequency reuse factor than inner cells.

The following example illustrates the principles of the IAM and shows its relation to actual resource allocation in an LTE system. In Figure 5.2, we assume an equal SNR for all transmission links and all interference links, i.e., all spectral efficiency parameters are identical. Furthermore, each eNB has to allocate two PRBs to serve the rate demand of an assigned UE. UE 2 and UE 3 are interfered by eNB 1 with an identical link quality as their supplying eNBs. Since eNB 2 and eNB 3 do not mutually interfere, the according interference impact factor of eNB 1 to both transmission links is $1/2$. Usually, it is not that simple to assess the interference situation and the reuse potential between cells. In such cases, the interference impact factors have to calibrated by a suitable parameter estimation, e.g., as carried out in Section 7.1.2.

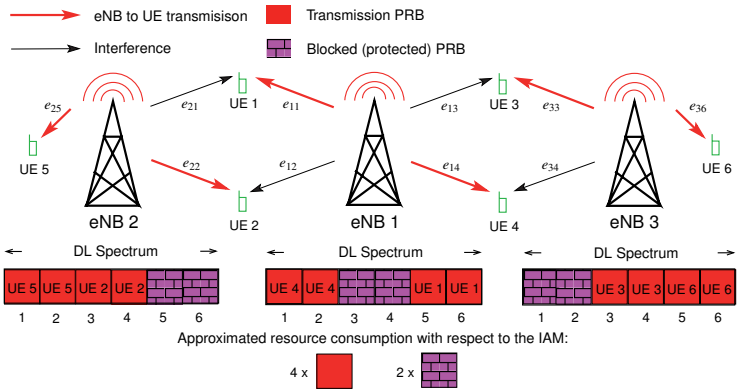


Figure 5.2: Exemplary PRB allocation and approximated resource consumption.

According to the IAM equation (5.2), eNB 1 has to block (protect) two PRBs: One for the interfered link from eNB 2 to UE 2 and one for the interfered link from eNB 3 to UE 3. Combined with the four PRBs that eNB 1 has to allocate for serving its assigned UEs 1 and 4, the overall bandwidth consumption at eNB 1 is six PRBs. This estimation is an exact match with the de facto PRB allocation that is shown in the Figure. The eNBs 2 and 3 cannot align the protected PRBs in the DL spectrum because they have only one neighbor eNB. Thus, their interference impact factors have to be chosen as one if the IAM shall hold exactly.

We assume that inter-femtocell interference can be handled by applying ICIC techniques as before and that the related penalty terms b_f^{itf} can be written analogously to (5.2). For modeling *cross-tier interference* between macrocells and femtocells, we take advantage of the fact that femtocells located nearby each other are typically interfered by the same eNB. As a consequence, the capacity of those femtocells can be increased significantly if eNBs block resources for femtocell interference mitigation [29]. Inter-femtocell interference is assumed to be low since the HeNB coverage area is generally very small. This allows particularly for a reuse of PRBs at HeNBs without a significant SINR degradation. Thus, it is beneficial for an efficient HetNet resource allocation to align the PRBs that are blocked at an eNB for femtocell interference mitigation jointly over all interfered FCs. The alignment may result in a (virtual) separation of MC spectrum and FC spectrum, which can particularly help to prevent the HetNet from overstressing its limited communication capabilities for the coordination between eNBs and HeNBs [29].

The IAM from above is extended by the discussed cross-tier interference aspects as follows. Additionally to the inter-macrocell interference component, each eNB s blocks

an amount of bandwidth that corresponds to the maximum consumed bandwidth at potentially interfered FCs $\mathcal{F}_s$. This means that (5.2) is rewritten as

$$b_s^{\text{itf}} = \underbrace{\sum_{\substack{(s',t) \in \mathcal{S} * \mathcal{T}, \\ s' \neq s}} q_{ss't} b_{s't}}_{\text{inter-cell interference}} + \underbrace{\max_{f \in \mathcal{F}_s} \{q_{sf} b_f\}}_{\text{cross-tier interference}} , \tag{5.4}$$

where b_f is computed according to (5.1) and

$$b_f^{\text{itf}} = \sum_{\substack{(f',t) \in \mathcal{F} * \mathcal{T}, \\ f' \neq f}} q_{ff't} b_{f't} \tag{5.5}$$

for all HeNBs f in the HetNet. For the cross-tier interference term in (5.4), we assume that all UEs served by a femtocell are located nearby the HeNB and that the same holds for all inter-FC interfered femtocells. Hence, the interference impact factor q_{sf} scales with the eNB signal strength to the HeNB location and relatively to the link quality that the HeNB provides to its served UEs. We suggest

$$q_{sf} = \frac{e_{sf}}{e_{\text{MAX}}} \tag{5.6}$$

if FCs predominantly serve users with maximum spectral efficiency e_{MAX} on the corresponding transmisson links. Since the protected bandwidth can be aligned along all interfered femtocells, the maximum cross-tier interference term in (5.4) eliminates interference to FCs in the eNB coverage area. Please note that due to the spatial diversity of rate demand and femtocell distribution, the amount of blocked (protected) bandwidth typically varies from eNB to eNB.

We call the model (5.1) *Bandwidth Reservation Concept (BRC)* as it approximates the expected amount of bandwidth that has to be reserved in the spectrum subject to supported UE rate demand and interference. Please note that the BRC considers bandwidth as a continuous variable and that it does not describe where the resources are allocated in the spectrum. It serves as an approximation model for the discrete combinatorial problem of PRB assignment.

The example depicted in Figure 5.3 illustrates the application of the BRC for a small HetNet with two eNBs and three femtocells:

- The bandwidth is measured in arbitrary units from the continuous domain.
- The blocked bandwidth at eNB 2, FC 2, and FC 3 is not considered.
- FC 2 has to reserve 2 bandwidth units for transmission to UE 2.
- FC 3 has to reserve 1 bandwidth unit for transmission to UE 3.

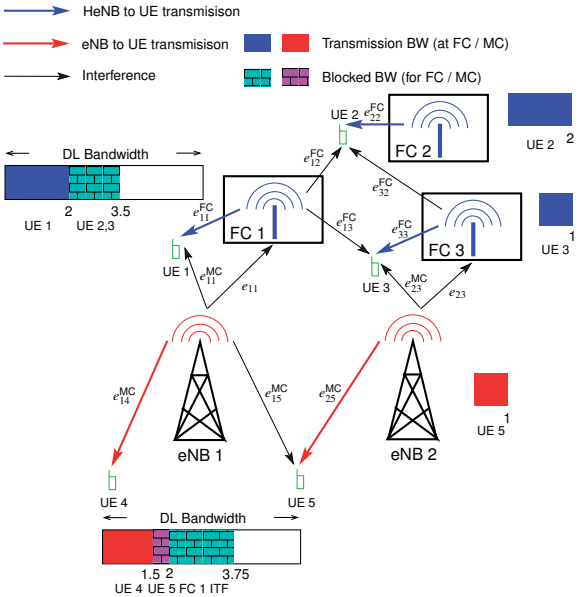


Figure 5.3: Bandwidth reservation at eNB 1 and HeNB 1 (FC 1) in a small HetNet.

- FC 1 allocates 2 bandwidth units for transmission to UE 1. It interferes the transmission link from FC 2 to UE 2 and the link from FC 3 to UE 3, both with an interference impact factor of 0.5. According to (5.5), it additionally reserves (blocks) $1 + 0.5 = 1.5$ bandwidth units.

- eNB 2 has to reserve 1 bandwidth unit for transmission to UE 5.

- eNB 1 has to reserve 1.5 units for transmission to UE 4. It interferes the link from eNB 2 to UE 5 with and interference impact factor of $q_{125} = 0.5$. According to (5.2), it additionally reserves (blocks) 0.5 bandwidth units.

The overall reserved bandwidth at eNB 1 follows from the BRC equation (5.1) and the cross-tier IAM equation (5.4). The cross-tier related amount of blocked bandwidth is determined by the maximum affected femtocell, i.e., FC 1, and its bandwidth consumption of 3.5 units. The cross-tier interference factor q_{11} is assumed as $e_{11}/e_{\mathrm{MAX}} = 0.5$. Therefore, eNB 1 reserves 1.75 additional bandwidth units to cope with cross-tier interference. Finally, the BRC model leads to an estimated bandwidth consumption of 3.75 units for eNB 1.

5.3 Numerical Evaluation

The BRC and its immanent IAM are evaluated exemplarily for the following simulation setup. We consider a hexagonal grid topology as depicted in Figure 5.1. Each of the related 3-sectorized eNBs transmits at a center frequency of 2 GHz and with 43 dBm Tx power, 14 dBi antenna gain, 5 dB noise figure, and 8 dB shadowing standard deviation. The received signal power in the area is computed according to the semi-empirical path loss model from [3] for 500 m inter-site distance. We apply the RRM procedures that are explained later on in Section 7.1.1. The considered spectrum comprises 50 PRBs, i.e., the effective transmission bandwidth is 9 MHz. Interference mitigation is applied as ICIC technique. Please note that femtocells are not considered in the evaluation. However, the basic insights obtained for the macrocell scenario are transferable.

500 UEs are randomly and uniformly distributed over the network area. As some users can experience a high signal power from eNBs that are not located nearby, the *relevant interferers* are identified for each UE. This are the eNBs that are received with a signal power of more than 1 % of the maximum received signal power. This approach reduces the set of potential interferers to the neighboring cells and some cells nearby. The UE service requests are varied in order to evaluate the BRC for different cell load situations: Table 5.1 shows the considered mobile services, which are basically a variant of the service profile from Table 4.5. Interactive real time, video, and FTP services are aggregated into a new service class *data* and the service proportions are modified such that the average load of the cells is either low or high.

Mobile service	Rate demand [kbps]	Proportion low	Proportion high
Data	$512 - 2000$	10 %	30 %
Web	$128 - 512$	20 %	40 %
VoIP	64	70 %	30 %

Table 5.1: Considered mobile services and their proportions in order to obtain low and high cell loads.

As discussed in the previous section, a higher number of neighboring cells typically provides more opportunities for frequency reuse. Hence, in Figure 5.1 the *inner cells*, i.e., cells $1, 2, 3, 4, 11, 18$, are distinguished from *outer cells*, i.e., cells $5, 6, 9, 12, 13, 17, 20$. First, the BRC is evaluated for a constant interference impact factor $q_{ss't} = \beta_s = 0.5$ for all cells and for all interfered transmission links. This setting results from the fact that we consider a high interference scenario where the signal power to interfered links is large, i.e., $e_{st}/e_{s't} \approx 1$. Furthermore, all cells have at least two neighbors that can benefit from a full frequency reuse, compare (5.3). Table 5.2 shows the according evaluation results for the inner cells and for two different cell loads. The approximated bandwidth consumption is compared to the consumption that is obtained by the RRM algorithms for a (static) system snapshot. All depicted evaluation measures are computed as mean over the cell-wise results. The mean gap denotes the difference between approximation and actual consumption, relatively to the actual consumption.

BW consumption	Simulation	BRC approximation	Mean gap	Deviation
Low cell load				
Transmission [kHz]	1027.5	962.6	−7.5 %	3.4 %
Blocked [kHz]	4300.0	4693.6	+8.4 %	18 %
Total [kHz]	5327.5	5656.2	+6 %	15 %
High cell load				
Transmission [kHz]	3245.0	1575.5	−52.3 %	12.6 %
Blocked [kHz]	5187.5	6744.0	+30.4 %	19.4 %
Total [kHz]	8432.5	8319.5	−1.7 %	10 %

Table 5.2: Bandwidth (BW) consumption for inner cells according to the BRC approximation versus results from system simulation.

The results do not show a significant variation over different simulation instances. This is mainly an effect of the high number of uniformly distributed UEs. According to Table 5.2, we observe the following effects:

i) For a low cell load, all UEs in the system simulation are served at full demand. For a high cell load, this holds only for 83 %, i.e., the system is partially overloaded. For the sake of a fair comparison, the identical set of served UEs is considered for the BRC evaluation.

ii) For a low cell load, the BRC achieves almost full accordance in all bandwidth approximations. This leads to a mean gap of +6 %. Due to the constant interference impact factor $q_{ss't} = 0.5$, the approximated amount of blocked (protected) bandwidth slightly exceeds the amount that is actually required in the system simulation. This effect results from the fact that the resource allocation in the simulation benefits from the degree of freedom to allocate resources in a sparsely occupied spectrum (low cell load).

iii) For a high cell load, the mean gap for the overall consumption results shrinks to −1.7 %. This results from an underestimated amount of transmission bandwidth and an overestimated amount of blocked bandwidth, which balance out when they are added up in (5.1). This is a general property of the BRC when the spectrum is highly occupied. The reason is the following: The required transmission bandwidth and the blocked (protected) bandwidth are approximated subject to the SNR, whereas the RRM in the simulation considers the (lower) SINR.

Table 5.3 shows the obtained evaluation measures for the outer cells of the network, still for the constant interference impact factor $q_{ss't} = 0.5$. The results affirm the assumption that outer cells have a reduced frequency reuse potential, and hence, the considered interference impact factor is chosen too low. The corresponding effect is that the BRC results suffer from a general underestimation of consumed bandwidth for all considered cell load situations. If the interference impact factor is increased to $q_{ss't} = 0.8$, the BRC accuracy improves significantly, see Table 5.4. Particularly, these results are comparable to the evaluation for the inner cells.

BW consumption	Simulation	BRC approximation	Mean gap	Deviation
Low cell load				
Transmission [kHz]	1107.9	1020.3	−8 %	1.4 %
Blocked [kHz]	3932.1	2100.6	−46.5 %	3.2 %
Total [kHz]	5040.0	3120.9	−38.5 %	4.9 %
High cell load				
Transmission [kHz]	2637.9	1682.7	−36.2 %	16.2 %
Blocked [kHz]	3865.7	2926.5	−21.1 %	16.3 %
Total [kHz]	6503.6	4609.2	−27.6 %	13.2 %

Table 5.3: Bandwidth (BW) consumption for outer cells according to the BRC approximation versus results from system simulation.

BW consumption	Simulation	BRC approximation	Mean gap	Deviation
Low cell load				
Transmission [kHz]	1107.9	1020.3	−8 %	1.4 %
Blocked [kHz]	3932.1	3360.9	−14.4 %	5.1 %
Total [kHz]	5040.0	4381.2	−13 %	4 %
High cell load				
Transmission [kHz]	2637.9	1682.7	−36.2 %	16.2 %
Blocked [kHz]	3865.7	4682.4	+26.3 %	26 %
Total [kHz]	6503.6	6365.1	−2.1 %	15 %

Table 5.4: Bandwidth (BW) consumption for outer cells according to the BRC approximation for $q_{ss't} = 0.8$ versus results from system simulation.

Overall, the evaluation results show that the BRC and its immanent IAM are suitable approximation models to describe the bandwidth consumption in cellular LTE systems subject to interference. The BRC results can be influenced by an adaption of the approximation parameters, i.e., the interference impact factors. This property allows for a calibration of the approximation model with respect to the RRM components of the considered system. In Section 7.1.2, we present an approach for adaptive BRC (IAM) parameter calibration with respect to dynamic system measurements. An according parameter calibration over several simulation instances of the static network scenario from above leads to

$$\beta_s^{\mathrm{INNER}} \in [0.4, 0.5] \,, \quad \beta_s^{\mathrm{OUTER}} \in [0.8, 0.9] \,, \tag{5.7}$$

where we still assume ${}^{e_{st}}\!/_{e_{s't}} \approx 1$, i.e., $q_{ss't} \approx \beta_s$, compare (5.6).

5.4 Summary

User assignment and resource allocation in multicell networks are typically carried out subject to the SINR in the system. In LTE heterogeneous networks, these tasks have additionally to consider cross-tier interference between macrocells and femtocells that transmit in the same spectrum. SINR-based decision problems have no linear representation, and hence, cannot be integrated into the MIP-based network optimization models that we develop. Therefore, we have introduced an approximate model that estimates the overall required bandwidth at eNBs and HeNBs with respect to the SNR in the system. The underlying principle of the corresponding bandwidth reservation concept (BRC) is to compute the required transmission bandwidth separately from the bandwidth that is blocked to mitigate interference to other cells. Both terms are computed subject to the SNR on the related radio links.

The interference approximation model (IAM) considers inter-cell interference as well as cross-tier interference. The cross-tier interference is modeled by the maximum consumed bandwidth at interfered femtocells, see (5.4). The related maximum operator can be transformed into a linear representation for any MIP formulated optimization problem. The linear representation requires one additional optimization variable and one additional constraint in the MIP.

In the following chapters, the BRC and the IAM serve as integral components of the MIP-based models that we develop for the optimization of LTE HetNets.

6 Cell Site Planning of LTE Heterogeneous Networks

In contrast to the network dimensioning approach presented in Chapter 4, cell site planning aims at the precise selection of base station locations and at the initial downtilt configuration of the related antennas. For the optimal cell site planning of LTE heterogeneous networks, we consider macrocell sites and and user-deployed femtocells that are not necessarily active all the time. The objective of the planning task is to provide a minimum number of macrocell sites such that mobile services are area-wide available. On the other hand, the deployment of dispensable cell sites has to be avoided for the sake of cost efficiency and low interference. The availability of mobile services is related to the network coverage, whereas the support of high rate services depends on the achievable network capacity. Hence, both KPIs are distinguished in the objective function of the corresponding planning problem in Section 6.2.

The economically motivated approach for multiple objective alignment from Section 4.2 is applied to deal with the trade-off between the conflicting objectives of the MOO problem. Instead of avoiding interference by minimum distance constraints, the interference approximation model from the previous chapter is utilized to incorporate inter-cell and cross-tier interference efficiently into the optimization model.

At the end of this chapter, we apply the developed optimization approach to a representative LTE HetNet planning scenario. The obtained results define the reference HetNet that is considered in Chapter 7 for the self-optimization of coverage and capacity.

6.1 System Model

We consider a multi-tier LTE HetNet according to the system model from Section 5.1. The system is supposed to apply interference mitigation – or any other comparable method – as ICIC technique.

The demand node (DN) model from Section 3.3 is applied to abstract from single users. Rate computation and bandwidth allocation for downlink transmission from eNBs and HeNBs to DNs are carried out according to Section 3.2. The planning model selects eNBs from the candidate set $\mathcal{S}$ for deployment. The selection depends on the input parameters and the optimization variables from Section 1.3 as well as on the additional parameters and variables that are given in Table 6.1.

Symbol & domain	Description
$c_s \in \mathbb{R}_{\geq 0}$	eNB deployment cost.
$\rho_f \in [0,1]$	Average HeNB activity rate.
$p_t \in \mathbb{R}_{>0}$	DN priority level, related to either the DN service type or the corresponding customer type.
$\lambda_{\text{RATE}} \in \mathbb{R}_{\geq 0}$	Sum rate weighting factor for objective scalarization.
$\lambda_{\text{BASIC}} \in \mathbb{R}_{\geq 0}$	Basic coverage weighting factor for objective scalarization.
e_{MIN}, $e_{\text{MAX}} \in \mathbb{R}_{>0}$	Minimum required spectral efficiency for transmission (LTE: CQI 1) and maximally achievable spectral efficiency (CQI 15).
$\mathcal{G}$	Conflict graph, containing all pairs of eNB deployment candidates that correspond to the same cell site but apply different antenna configurations.
a	Wildcard symbol for either an eNB s or an HeNB f.
$q_{aa't} \in [0,1]$	Interference impact factor from (H)eNB a to the transmission link from (H)eNB a' to t.
$b_a^{\text{itf}} \in \mathbb{R}_{\geq 0}$	Auxiliary variable describing the amount of protected bandwidth at (H)eNB a.
$b_a \in \mathbb{R}_{\geq 0}$	Auxiliary variable describing the amount of overall consumed bandwidth (transmission plus protected) at (H)eNB a.

Table 6.1: Additional input parameters (upper part) and variables (lower part) for cell site planning in LTE HetNets, complementing Table 1.1.

Network *coverage* and network *capacity* are distinguished in the objective function in order to obtain deployment solutions that guarantee area-wide mobile services and that support high data rates in demand hotspots. The corresponding objectives are weighted with the factors λ_{BASIC} and λ_{RATE} to implement the scalarization approach from Section 2.3.1. This is done in analogy to (4.3). We introduce *priority level* parameters p_t to allow for prioritization of DN coverage. While the data rate is related to the requested service, the priority level can correspond to either the service class or the customer type. For instance, voice services (VoIP) might have a higher priority than data services or business clients may be favored over private customers.

Every macrocell site can apply different antenna configurations, i.e., the initial *antenna downtilt* has to be chosen for every deployed eNB. Each configuration of an eNB is considered as a separate deployment candidate in $\mathcal{S}$. The joint deployment of overlapping candidates is avoided by adding all pairs of different configurations of the same eNB to the conflict graph $\mathcal{G}$. The deployment of a macrocell site s is associated with cost c_s, whereas femtocell equipment is purchased by users and, in that sense, free of charge for the network operator. This also means that the operator cannot influence where HeNBs are deployed and when they are activated by the users. However, we

assume that the network operator knows at least the distribution of potential femtocell locations as well as the average activity rates ρ_f .

The interference approximation model from Chapter 5 is utilized to estimate the amount of protected (blocked) bandwidth b_a^{itf} and the amount of overall consumed bandwidth b_a for each (H)eNB a.

6.2 Optimization Model

We introduce the MIP from Table 6.2 to formalize the following optimization problem: Given the set $\mathcal{F}$ of potentially active femtocells and given the rate demand distribution, modeled by the set $\mathcal{T}$ of DNs. Which eNBs from the set $\mathcal{S}$ of candidates shall be selected for deployment in order to cover and serve the DNs optimally?

The sets

$$\begin{aligned}
\mathcal{S} * \mathcal{T} &= \{(s,t) \in \mathcal{S} \times \mathcal{T} \ : e_{st} \geq e_{\text{MIN}}\} \ , \\
\mathcal{S}_t &= \{s \in \mathcal{S} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \ , \quad \mathcal{T}_s = \{t \in \mathcal{T} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \ , \\
\mathcal{F} * \mathcal{T} &= \{(f,t) \in \mathcal{F} \times \mathcal{T} \ : e_{ft} \geq e_{\text{MIN}}\} \ , \\
\mathcal{F}_t &= \{f \in \mathcal{F} \ : (f,t) \in \mathcal{F} * \mathcal{T}\} \ , \quad \mathcal{T}_f = \{t \in \mathcal{T} \ : (f,t) \in \mathcal{F} * \mathcal{T}\} \ , \\
\mathcal{F}_s &= \{f \in \mathcal{F} \ : e_{sf} \geq e_{\text{MIN}}\}
\end{aligned}$$

are defined to exclude decision variables and constraints that are irrelevant due to insufficient link quality. We consider the optimization problem

$$\max \left\{ \lambda_{\text{BASIC}} \underbrace{\sum_{t \in \mathcal{T}} p_t z_t}_{\text{coverage}} + \lambda_{\text{RATE}} \underbrace{\sum_{t \in \mathcal{T}} r_t^{\text{eff}}}_{\text{capacity}} - \underbrace{\sum_{s \in \mathcal{S}} c_s y_s}_{\text{cost}} \right\} \tag{6.1}$$

with respect to the optimization variables $y_s, z_{st}, z_{ft}, b_{st}, b_{ft}$ and the auxiliary variables $z_t, b_f, b_s^{\text{itf}}, r_t^{\text{eff}}$ in order to answer the initial question from above. Compared to the objective function of the dimensioning approach (4.3), network coverage and network capacity are distinguished by separate terms. The sum of covered priorities is considered as HetNet coverage and the effectively served sum rate models the HetNet capacity. DNs with an infinitesimal rate demand r_t may serve as pure *coverage points*. DNs with a high rate demand are relevant for the coverage term in (6.1) as well as for the capacity component. The conflicting objectives of the MMO problem are aligned into one scalar-valued objective according to the economically motivated approach from Section 4.2. The corresponding weighting factors (exchange rates) λ_{BASIC} and λ_{RATE} are chosen as ARPU over the considered time period. Thus, optimal solutions for (6.1) define a cell site deployment – and initial antenna configuration – for *HetNet Profit MAXimization (HetNet-ProMAX)* in a techno-economical sense.

$$\max \left\{ \lambda_{\text{BASIC}} \sum_{t \in \mathcal{T}} p_t z_t + \lambda_{\text{RATE}} \sum_{t \in \mathcal{T}} r_t^{\text{eff}} - \sum_{s \in \mathcal{S}} c_s y_s \right\}$$

subject to

$$z_t = \sum_{s \in \mathcal{S}_t} z_{st} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T}$$

$$z_{st} \leq y_s \ , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T}$$

$$y_s + y_{s'} \leq 1 \ , \qquad \text{for all } (s,s') \in \mathcal{G}$$

$$\sum_{f \in \mathcal{F}_t} z_{ft} \leq z_t \ , \qquad \text{for all } t \in \mathcal{T}$$

$$b_{ft} \leq \tfrac{r_t}{e_{ft}} z_{ft} \ , \qquad \text{for all } (f,t) \in \mathcal{F} * \mathcal{T}$$

$$b_f = \sum_{t \in \mathcal{T}_f} b_{ft} + \sum_{\substack{(f',t) \in \mathcal{F} * \mathcal{T}, \\ f' \neq f}} q_{ff't} b_{f't} \leq \rho_f B_f \ , \qquad \text{for all } f \in \mathcal{F}$$

$$b_{st} \leq \frac{r_t z_t - \sum_{f \in \mathcal{F}_t} e_{ft} b_{ft}}{e_{st}} \ , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T}$$

$$b_{st} \leq B_s z_{st} \ , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T}$$

$$r_t^{\text{eff}} = \sum_{s \in \mathcal{S}_t} e_{st} b_{st} + \sum_{f \in \mathcal{F}_t} e_{ft} b_{ft} \geq r_t^{\text{MIN}} z_t \ , \qquad \text{for all } t \in \mathcal{T}$$

$$b_s^{\text{itf}} = \sum_{\substack{(s',t) \in \mathcal{S} * \mathcal{T}, \\ s' \neq s}} q_{ss't} b_{s't} + \max_{f \in \mathcal{F}_s} \left\{ \tfrac{e_{sf}}{e_{\text{MAX}}} b_f \right\} \ , \qquad \text{for all } s \in \mathcal{S}$$

$$\sum_{t \in \mathcal{T}_s} b_{st} + b_s^{\text{itf}} \leq B_s + (1 - y_s) \cdot \infty \ , \qquad \text{for all } s \in \mathcal{S}$$

Table 6.2: MIP formulation of the profit maximization problem for techno-economical LTE HetHet deployment (HetNet-ProMAX).

The maximization in (6.1) is carried out subject to the following constraints. The inequality

$$z_t = \sum_{s \in \mathcal{S}_t} z_{st} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T} \tag{6.2}$$

ensures that DNs are assigned to at most one eNB. The variable z_t indicates whether DN t is assigned (covered) or not. Moreover, this constraint defines coverage as *covered by an eNB*, i.e., the basic HetNet coverage does not depend on any femtocell activity. According to the inequalities

$$z_{st} \leq y_s \ , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T} \tag{6.3}$$

$$y_s + y_{s'} \leq 1 \ , \qquad \text{for all } (s,s') \in \mathcal{G} \tag{6.4}$$

only deployed eNBs are available for DN assignment and the joint deployment of conflicting eNBs is not allowed. Here, the conflict graph $\mathcal{G}$ contains all pairs of eNB deployment candidates that correspond to the same cell site but apply different antenna configurations. Each different (initial) antenna configuration of a cell site is a separate element in the candidate set $\mathcal{S}$. The main effect of switching the antenna configuration of a cell site is a change of the spatial signal power distribution. Thus, the related spectral efficiency parameters can differ significantly over the available antenna configurations.

We assume that the location of each potentially active femtocell $f \in \mathcal{F}$ is known. Furthermore, we assume estimations for their activity rate $\rho_f \in [0, 1]$. The following principles are applied to deal with the uncertainty of FC activity. First, femtocells are excluded from contributing to the basic HetNet coverage. And further:

1. HeNBs can (partially) take over rate demand of DNs. These DNs have to be assigned to eNBs. This models the opportunity that users can be served by femtocells in the operating system.

2. The available bandwidth at HeNBs is scaled down proportionally to their activity rate.

These principles are modeled by

$$\sum_{f \in \mathcal{F}_t} z_{ft} \leq z_t \ , \qquad \text{for all } t \in \mathcal{T} \tag{6.5}$$

$$b_{ft} \leq \frac{r_t}{e_{ft}} z_{ft} \ , \qquad \text{for all } (f,t) \in \mathcal{F} * \mathcal{T} \tag{6.6}$$

$$b_f = \sum_{t \in \mathcal{T}_f} b_{ft} \ + \sum_{\substack{(f',t) \in \mathcal{F} * \mathcal{T}, \\ f' \neq f}} q_{ff't} b_{f't} \leq \rho_f B_f \ , \qquad \text{for all } f \in \mathcal{F} \tag{6.7}$$

where (6.5) ensures that only DNs which are covered by an eNB can be served by at most one HeNB. Constraint (6.6) follows from (3.7) and the inequality (6.7) bounds the approximated bandwidth consumption b_f at HeNBs to the (scaled) maximum bandwidth. The bandwidth consumption is approximated by the BRC and the IAM according to (5.1) and (5.5).

Optionally, an additional bandwidth buffer might be considered in order to cope with unexpected inter-FC and cross-tier interference. A reasonable buffer can be obtained by a replacement of the interference term in (6.7) with

$$\sum_{\substack{(f',t) \in \mathcal{F} * \mathcal{T}, \\ f' \neq f}} q_{ff't} \frac{r_t}{e_{ft}} z_{f't} \ .$$

By this replacement, all interfered FC transmission links are considered with fully served rate demand even though the actual transmission rate can be lower, see (6.6).

The DN rate that is not (partially) served by any femtocell can be served by the assigned eNB, i.e.,

$$b_{st} \leq \frac{r_t z_t - \sum\limits_{f \in \mathcal{F}_t} e_{ft} b_{ft}}{e_{st}} , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T} \tag{6.8}$$

$$b_{st} \leq B_s z_{st} , \qquad \text{for all } (s,t) \in \mathcal{S} * \mathcal{T} . \tag{6.9}$$

As before in the dimensioning model, compare (4.15), the effectively served rate of each covered DN

$$r_t^{\text{eff}} = \sum_{s \in \mathcal{S}_t} e_{st} b_{st} + \sum_{f \in \mathcal{F}_t} e_{ft} b_{ft} \geq r_t^{\text{MIN}} z_t , \qquad \text{for all } t \in \mathcal{T} \tag{6.10}$$

has to exceed a QoS related minimum rate threshold r_t^{MIN}. The threshold can be very small for coverage related DNs or sufficiently high for capacity related DNs. Depending on the minimum rate requirements and the spectrum occupation at (H)eNBs, (6.10) can cause a multiple knapsack problem. In contrast to the MKP from Section 3.4.2, the required bandwidth for an assignment is not necessarily equal for all potential suppliers. Moreover, the available bandwidth for data transmission, i.e., the knapsack capacity, can change with the assignment of users. This effect is related to the applied interference approximation model and the underlying concept of bandwidth protection.

According to the HetNet IAM (5.4) and suggestion (5.6), the interference related bandwidth consumption at eNBs is given by

$$b_s^{\text{itf}} = \sum_{\substack{(s',t) \in \mathcal{S} * \mathcal{T}, \\ s' \neq s}} q_{ss't} b_{s't} + \max_{f \in \mathcal{F}_s} \left\{ \frac{e_{sf}}{e_{\text{MAX}}} b_f \right\} , \qquad \text{for all } s \in \mathcal{S} . \tag{6.11}$$

This term is considered in

$$\sum_{t \in \mathcal{T}_s} b_{st} + b_s^{\text{itf}} \leq B_s + (1 - y_s) \cdot \infty , \qquad \text{for all } s \in \mathcal{S} \tag{6.12}$$

in order to meet the bandwidth limitation restrictions. Non-deployed cell sites and configuration candidates can gain large positive values for b_s^{itf} in (6.11). The infinity term on the right hand side of (6.12) avoids feasibility problems that are irrelevant for these candidates. In practice, the infinity term is replaced by a sufficiently large number, e.g., the sum of available bandwidth over all deployment candidates.

Each optimal solution of the HetNet-ProMAX problem provides an approximate bandwidth consumption for all (H)eNBs in the solution set. The approximation is computed according to the BRC and the IAM. This information can be used in the operating HetNet for an initial configuration of the applied ICIC algorithms.

6.3 Numerical Evaluation

Table 6.3 shows the parameter setup for planning eNB cell sites and their initial antenna downtilt configuration in an urban LTE HetNet scenario. Figure 6.1 (b) depicts the 60 potential eNB site locations as (red) dots. The building data is taken from [31]. Since the cell site planning is carried out with respect to effects that take place on a long-term scale, we assume average values for volatile system parameters such as transmit power and shadowing. An omnidirectional antenna pattern and two supported downtilt configurations are considered for the antenna configuration of each eNB. Thus, the set $\mathcal{S}$ contains 120 different deployment candidates and the conflict graph $\mathcal{G}$ contains 60 eNB pairs. Figure 6.1 (a) illustrates the effect of switching the antenna downtilt for an exemplary eNB: The signal strength at short distance increases with a higher downtilt but at the cost of a shrinking cell footprint. Please note that we use the highly accurate ray optical path loss model from Section 3.1.2 to compute the spatial signal power distribution for each eNB deployment candidate. The signal power distribution for (indoor) HeNBs is obtained by applying a free space path loss model as introduced at the beginning of Section 3.1. The signal power information is precomputed prior to the optimization task. The spectral efficiency parameters are derived from the discrete rate-power function according to (3.8) and the LTE SINR requirements defined in Table 6.4.

The eNB operational expenses (OPEX) of 1800 € per month cover site rental, leased line rental, air-conditioning, and maintenance [33]. The investment costs (CAPEX) for eNB deployment are neglected analogously to the network dimensioning approach, see Section 4.3 for details. Contrarily to eNB deployment, user-deployed femtocells do not cause any expenses for the network operator. Nevertheless, they can contribute to the HetNet performance when the corresponding HeNBs are active. We consider an HeNB activity rate of $\rho_f = 50\,\%$ and a transmit power of 3 dBm. Femtocells operate in the same transmission spectrum as eNBs. We uniformly distribute 400 HeNBs at random over the round about 2000 buildings in the planning area. HeNBs are located only indoors.

The constant DN rate demand of 320 kbps results from an aggregation of service-related average rates. Each DN t has the priority $p_t = 1$ and a minimum rate requirement $r_t^{\text{MIN}} = r_t$. The 750 DNs are uniformly distributed over the city area at random. We apply cell site planning for 50 % and 80 % indoor DNs. The scalarization parameters $\lambda_{\text{BASIC}} = 50$ € and $\lambda_{\text{RATE}} = 0.5\,€/_{kbps}$ transfer objective (6.1) into the techno-economical domain of monthly profit for the network operator. The basic fee λ_{BASIC} covers the monthly basic charges, the fees for basic services (sms, voice), and all extra fees (flatrates, special options, etc.) for the DN related users. The rate fee λ_{RATE} denotes the charge per transmitted data unit.

We assume the interference impact factor $q_{ss'} = 1/3$ for all eNBs and $q_{sf} = 1$ for all interfered HeNB transmissions, compare (5.3) and (5.6). The low impact factor for inter-macrocell interference is motivated by the fact that all eNB deployment candidates have a large resource reuse potential: First, the cell density is typically very high in urban areas. Therefore, there are many neighboring cells that allow for a reuse

System parameter	Setting
Planning area	Urban, 2.5 km × 3.5 km, 5 m resolution
Number of buildings	Approximately 2000
Wall penetration loss	10 dB
Shadowing standard deviation	8 dB
Carrier frequency	2 GHz
Transmission bandwidth	10 MHz
Number of eNB (macrocell site) location candidates	60
eNB antenna downtilt profile	{0° (low), 5° (high)}
eNB Tx power	43 dBm
Propagation model eNB	Ray optical, omnidirectional
eNB antenna gain and noise figure	14 dBi and 5 dB
Number of HeNBs (femtocells)	400
HeNB distribution	Random, uniformly distributed over buildings
Minimum distance HeNB to eNB	35 m
HeNB activity rate	50 %
HeNB Tx power	3 dBm
Propagation model HeNB	Free space + wall penetration, omnidirectional
Noise figure HeNB	8 dB
Number of DNs	750
Indoor DNs	50 % and 80 %
DN rate demand	320 kbps
DN antenna gain and noise figure	0 dBi and 9 dB

Table 6.3: Considered system parameters for HetNet deployment.

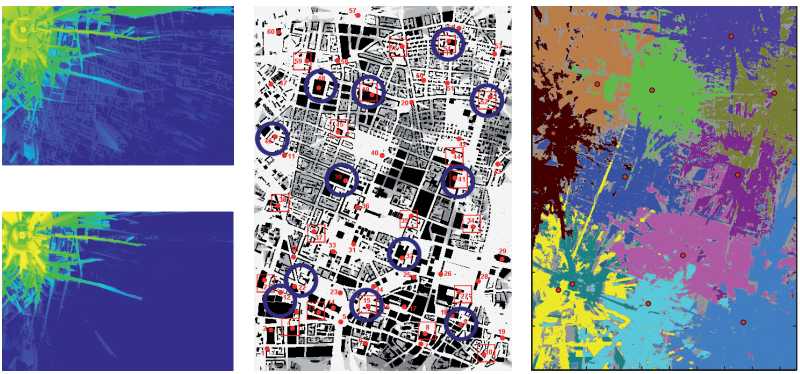
of (blocked) bandwidth. And second, the signal propagation in urban environments leads to further reuse candidates that are not located nearby, see Figure 6.1 (c) for an illustration. This enables the reuse option also for the outer cells in the network.

The HetNet-ProMAX problem is solved by utilizing the Gurobi Optimizer [48]. The obtained results are summarized in Table 6.5 and Figure 6.1 (b) illustrates the deployment solution for 80 % indoor DNs. The 12 optimally deployed eNBs of the HetNet solution are marked by circles, whereas the 19 eNBs from the alternative single-tier solution are framed by squares. Moreover, Figure 6.1 (c) shows the best server plot for the corresponding HetNet macrocells. Light gray areas indicate that the SNR values from different eNBs do not differ more than 2 dB, i.e., such areas indicate potential *handover zones*.

The eNBs in the single-tier solution and the deployed eNBs in the HetNet cover almost the same number of DNs and serve a similar amount of requested data rate. Please note that indoor DNs are served by eNBs if no femtocells are available. The number of deployed eNBs in the HetNet is reduced by 37 % due to the availability of femtocells. In fact, half of the served data rate in the HetNet is covered by femtocells. Figure 6.2 illustrates how HeNBs improve the network capacity: They either provide higher signal strength in low power macrocell regions (blue) or they serve DNs in medium power

CQI	Modulation	Code rate	Spectral efficiency [bps/Hz]	Receiver SINR [dB]
0		out of range		
1	QPSK	1/8	0.25	−5.1
2	QPSK	1/5	0.40	−2.9
3	QPSK	1/4	0.50	−1.7
4	QPSK	1/3	0.66	−1.0
5	QPSK	1/2	1.00	2.0
6	QPSK	2/3	1.33	4.3
7	QPSK	3/4	1.50	5.5
8	QPSK	4/5	1.60	6.2
9	16-QAM	1/2	2.00	7.9
10	16-QAM	2/3	2.66	11.3
11	16-QAM	3/4	3.00	12.2
12	16-QAM	4/5	3.20	12.8
13	64-QAM	2/3	4.00	15.3
14	64-QAM	3/4	4.50	17.5
15	64-QAM	4/5	4.80	18.6

Table 6.4: LTE SINR requirements according to [94] for 10 MHz transmission bandwidth and BLER $= 10^{-1}$.



(a) eNB applying 0° (top) and 5° (below) antenna downtilt.

(b) Candidate sites (dots) and deployed eNBs in the HetNet (circles) and the single-tier solution (squares).

(c) Best server plot for the eNB related macrocells in the HetNet deployment solution.

Figure 6.1: Considered planning area and deployment solution for 80 % indoor DNs .

macrocell regions (green) to lower the eNB load. Following from the IAM estimations, eNBs have to block on average 13 % of the spectrum in order to mitigate cross-tier interference. Combined with the bandwidth that is blocked due to inter-macrocell interference, eNBs utilize on average round about half of the spectrum for transmission and the other half for ICIC. Most of the deployed single-tier eNBs (85 %) apply

	Single-tier	**HetNet**
80% indoor DNs		
Number of deployed eNBs	19	12
Rate served by eNBs [%]	100	50
Average transmission bandwidth at eNBs according to BRC [kHz]	5155	5630
Average protected bandwidth at eNBs (MC + FC) according to IAM [kHz]	3975 + 0	3070 + 1300
Runtime [h]	6.5	259
50% indoor DNs		
Number of deployed eNBs	17	11
Rate served by eNBs [%]	100	52
Average transmission bandwidth at eNBs according to BRC [kHz]	4930	5530
Average protected bandwidth at eNBs (MC + FC) according to IAM [kHz]	4400 + 0	3170 + 1300
Runtime [h]	12	181

Table 6.5: Result statistics for two different indoor DN ratios and two different network topologies.

the higher antenna downtilt, whereas most eNBs in the HetNet solution select a low antenna downtilt. The dense single-tier cell deployment might be interpreted as an alternative solution to the HetNet topology.

By solving the HetNet-ProMAX problem for 50 % indoor DNs, we investigate its sensitivity to the indoor DN ratio. The corresponding results are shown in the lower part of Table 6.5. It turns out that the relative results for the single-tier solution and the HetNet solution are roughly the same as for 80 % indoor DNs. However, the total number of required eNBs decreases slightly as the signal power from eNBs to outdoor users is generally higher.

Since the spectrum of the deployed HetNet eNBs is highly occupied, inactive femtocells can barely be compensated. As a consequence, the HeNB activity rate is a very important parameter that should not be overestimated.

The runtime to solve the HetNet problem is more than one magnitude higher as the runtime for computing an optimal single-tier solution. This holds for both evaluation variants. Mainly two reasons cause this effect: First, the lower number of possible transmitter-to-DN assignments reduces the size of the single-tier problem instance. And second, the trade-off between revenue and cost is more difficult to assess for an eNB deployment candidate in the HetNet than in the single-tier network. Particularly, there can exist alternative solutions where DNs that are served by a single eNB can also be served by a combination of another eNB and HeNB.

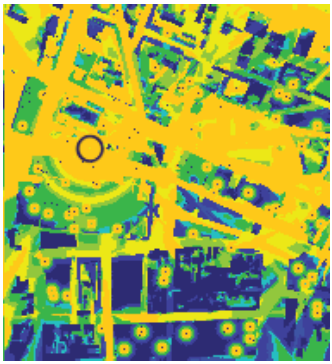


Figure 6.2: Signal strength distribution for a deployed eNB (large circle) and multiple HeNBs (small circular areas).

6.4 Summary

Most of the related work in the context of LTE HetNet deployment addresses performance improvement and flexibility gain when a single-tier network is switched to a multi-tier topology [81]. Some works consider the planning of femtocells and pico nodes. However, in most cases the planning is carried out either subject to an existing macrocell environment [56] or it ignores the macrocell tier at all [67].

Our approach deals with the problem of planning cell sites and their initial configuration jointly over both network tiers. Particularly, it takes the corresponding resource allocation into account. Macrocells are selected for deployment subject to user-deployed femtocells that are not necessarily active all the time. The interference approximation model from Chapter 5 is integrated into the optimization problem in order to incorporate inter-cell and co-channel interference. Although not mentioned explicitly, *pico nodes* can be modeled in our planning approach as *low capability* eNBs. These nodes might support lower power ranges or have fixed antenna configurations.

Numerical evaluation results demonstrate the feasibility of our optimization approach (HetNet-ProMAX) for the purpose of cell site planning of heterogeneous LTE networks. The HetNet planning solution for 80 % indoor DNs serves as reference network in the following chapter, where we present an approach for the autonomous self-control of cell parameters subject to coverage and capacity optimization.

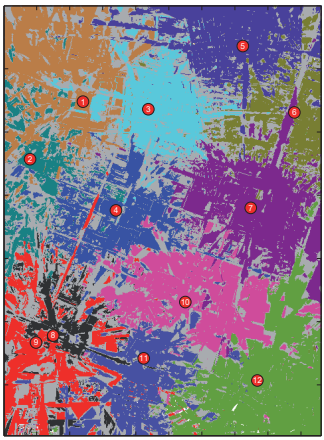
7 Self-Optimization of Coverage and Capacity

In this chapter, we develop an approach for *closed-loop, autonomous self-optimization of coverage and capacity* [85] in operating heterogeneous LTE networks. *Autonomous self-optimization* means that there is no need for any human intervention once the optimization parameters have been initially specified. According to [8], this feature is one of the essential 3GPP requirements for *self-organizing networks*. It is a prerequisite to benefit maximally from the opportunities that self-organization provides to minimize OPEX and operational effort.

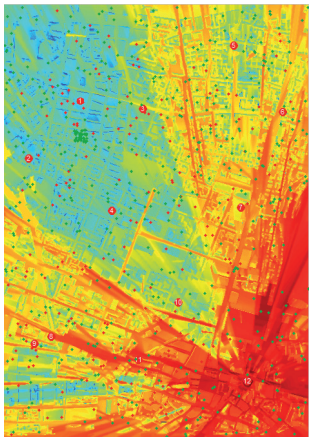
Self-organized *Coverage and Capacity Optimization (CCO)* considers the online configuration of existing transmitters in an operating network. It is the consecutive step to network dimensioning and cell site planning in the system lifecycle. All CCO related procedures have to be aligned with the existing network operation components. The optimization task is carried out subject to the dynamically changing network situation. Overall, this leads to a network operation paradigm where network elements autonomously configure themselves at certain time instances, and the configuration takes the expected network situation of the next operation cycle into account. An *operation* cycle denotes the time interval before the next self-configuration is triggered.

The cell site planning results from Section 6.3 (80 % indoor DNs) serve as reference network in this chapter, see Figure 7.1. We introduce the corresponding HetNet model, its operation components, and the automatic adaption of dynamic system parameters in Section 7.1. At a first glance, Figure 7.1 (a) suggests a smooth macrocell topology. Figure 7.1 (b), however, illustrates that we have to keep in mind that the interference-related interdependencies between cells can be very strong and spatially diverse. The system performance measures defined in Section 7.2 serve as assessment criteria for the results of joint coverage and capacity maximization in Section 7.3. The corresponding optimization model is embedded into a traffic-light based network operation scheme in Section 7.4. The *integrated CCO approach* autonomously controls transmission power, antenna tilt, and activity status (switch on/off) of the network elements. The evaluation in Section 7.5 investigates achievable performance improvements and illustrates the underlying principles of our CCO approach.

This chapter extends the work presented in [39].



(a) 12 initial macrocells, light gray areas indicate potential handover zones (best server difference ≤ 2 dB).



(b) Signal strength distribution of eNB 12, FCs (red dots), UEs (green dots), and hotspot (bunch of spots).

Figure 7.1: Considered urban reference HetNet, resulting from the HetNet cell site planning presented in Section 6.3.

7.1 System Model

We basically consider the HetNet system model from Section 5.1. In the following, this specification is extended by the components that are missing to model the operating system and to implement self-organized CCO.

Self-organized CCO requires coordination and information exchange between transmitters in certain situations. The following components enable the related communication between network elements: HeNBs are connected to the backhaul – in most cases via DSL – which generally allows for centralized optimization of femtocells. However, the coordination between HeNBs and eNBs can only take place at coarse time scales, e.g., tens of frames, but not at the granularity of every frame [98]. This is not a crucial drawback for our purposes because the time interval between any two consecutive CCO executions is sufficiently long. Communication between eNBs is supported by the X2 interface [12] (direct interconnection) and the S1 interface [11] via backbone (indirect interconnection). HeNB to (H)eNB information exchange and coordination can utilize the Type 1 interface via the HeNB Management System (HeMS) [9]. The communication between (H)eNBs and UEs follows several signaling schemes that are defined for dedicated control channels: The *Physical Downlink Control CHannel (PDCCH)* is used for UE-specific information and the *Physical Uplink Control CHannel (PUCCH)* carries, for instance, the CQI information. Overall, we assume fully meshed communication capabilities between all network elements.

The developed optimization model utilizes the demand node (DN) model from Section 3.3 to abstract from single users. Rate computation and bandwidth allocation for downlink transmission from eNBs and HeNBs to DNs are carried out according to Section 3.2. The optimization depends on the input parameters and optimization variables from Section 1.3 and on the ones that are given in Table 7.1. The CCO objective is the optimal configuration of all eNBs and HeNBs in the HetNet subject to the joint maximization of network coverage and capacity.

The considered HetNet applies interference mitigation as ICIC technique. The interference approximation model from Chapter 5 is utilized to estimate the corresponding amount of protected (blocked) bandwidth b_a^{itf} and the amount of overall consumed bandwidth b_a for each (H)eNB a.

Symbol & domain	Description
$\mathcal{S}^{\text{MC}}$, $\mathcal{F}^{\text{FC}}$	Index set of available MC transmitters and FC transmitters.
$\mathcal{S}$, $\mathcal{F}$	Index set of different configurations at available MC transmitters (eNBs) and FC transmitters (HeNBs).
a	Wildcard symbol for either a configured eNB s or a configured HeNB f, i.e., a transmitter with a particular configuration.
$p_t \in \mathbb{R}_{>0}$	DN priority level, related to either the DN service type or the corresponding customer type.
$p_a^{\text{MAX}} \in \mathbb{R}_{>0}$	Maximum feasible sum of covered priorities at the (H)eNB a.
$\delta_{\text{ITF}} \in [0,1]$	Minimum threshold for coverage definition over pairwise spectral efficiency ratios.
$\lambda_{\text{COV}} \in \mathbb{R}_{\geq 0}$	Coverage weighting factor for objective scalarization.
$\lambda_{\text{CAP}} \in \mathbb{R}_{\geq 0}$	Sum rate (capacity-related) weighting factor for objective scalarization.
e_{MIN}, $e_{\text{MAX}} \in \mathbb{R}_{>0}$	Minimum required spectral efficiency for transmission (LTE: CQI 1) and maximally achievable spectral efficiency (CQI 15).
$q_{aa't} \in [0,1]$	Interference impact factor from (H)eNB a to the transmission link from (H)eNB a' to t.
y_s, $y_f \in \{0,1\}$	Binary decision variables indicating the selection of a certain configuration at eNBs and HeNBs.
$\tilde{z}_{st}$, $\tilde{z}_{ft} \in \{0,1\}$	Binary decision variables indicating the coverage of DN t by a certain (configured) eNB or HeNB.
$\tilde{z}_t \in \{0,1\}$	Auxiliary variable indicating that DN t is covered.
$b_a^{\text{itf}} \in \mathbb{R}_{\geq 0}$	Auxiliary variable describing the amount of protected bandwidth at (H)eNB a.
$b_a \in \mathbb{R}_{\geq 0}$	Auxiliary variable describing the amount of overall consumed bandwidth (transmission plus protected) at (H)eNB a.

Table 7.1: Additional or reinterpreted input parameters (upper part) and variables (lower part) for self-organized CCO in LTE HetNets.

The meaning of some variables that were introduced in previous chapters changes when network configuration is considered for the operational stage: The set $\mathcal{S}$ and the set $\mathcal{F}$ are now associated with the indices of all *configurations* that can be chosen at eNBs and HeNBs. The elements s and f are created in the configuration index sets $\mathcal{S}$ and $\mathcal{F}$ for each configuration that is selectable for an MC transmitter (eNB) $i \in \mathcal{S}^{\mathrm{MC}}$ and FC transmitter (HeNB) $j \in \mathcal{F}^{\mathrm{FC}}$, respectively. The associated configurations are selected if y_s and y_f are equal to one. We denote the set of configuration indices that belong to the same eNB $i \in \mathcal{S}^{\mathrm{MC}}$ or HeNB $j \in \mathcal{F}^{\mathrm{FC}}$ as $\mathcal{Y}_i$ and $\mathcal{Y}_j$, respectively. Thus, it is $\mathcal{S} = \bigcup_{i \in \mathcal{S}^{\mathrm{MC}}} \mathcal{Y}_i$ and $\mathcal{F} = \bigcup_{j \in \mathcal{F}^{\mathrm{FC}}} \mathcal{Y}_j$. The eNB configuration state space considers transmission power and antenna downtilt as control parameters, whereas HeNBs can only configure their *transmission power*. We ensure the selection of one unique configuration for each HetNet transmitter by applying constraints

$$\sum_{s \in \mathcal{Y}_i} y_s = 1\,, \quad i \in \mathcal{S}^{\mathrm{MC}} \text{ (exactly one configuration per eNB)} \tag{7.1}$$

$$\sum_{f \in \mathcal{Y}_j} y_f = 1\,, \quad j \in \mathcal{F}^{\mathrm{FC}} \text{ (exactly one configuration per HeNB)}\,. \tag{7.2}$$

Since the association from a selected configuration to the related transmitter is unique, we use the notation *transmitter* or (H)eNB $s \in \mathcal{S}$, $f \in \mathcal{F}$, even though s and f just refer to certain configurations. Furthermore, we use the wildcard symbol $a \in \mathcal{S} \cup \mathcal{F}$ for short notation whenever expressions are related to transmitters $s \in \mathcal{S}$ as well as to transmitters $f \in \mathcal{F}$.

Compared to the previous chapter, we redefine the meaning of system *coverage*. The related variables are separated from those ones that refer to data transmission, i.e., to system capacity. The decision variables $\tilde{z}_{st},\ \tilde{z}_{ft}$ indicate which DN is covered by what (H)eNB. Coverage is defined by a minimum *Signal-to-Interference Ratio (SIR)* requirement subject to the threshold value δ_{ITF}. The variables $\tilde{z}_{st},\ \tilde{z}_{ft}$ are distinguished from the decision variables $z_{st},\ z_{ft}$ that describe the (H)eNB-to-DN assignment for data transmission. The same holds for the binary coverage indicator variables $\tilde{z}_t$ and their counterparts z_t that now strictly serve as capacity-related variables.

The maximum number of DN priorities that can be covered by a single transmitter is bounded since the resources at eNBs and HeNBs are limited. This shall prevent from degraded CCO solutions, where only a few transmitters are in charge of the whole provided network coverage. Such solutions can lead to infeasible situations if all UEs that are modeled by DNs become active at the same time. A reasonable maximum feasible sum of covered priorities for an (H)eNB a can be obtained as follows. We assume an average priority $\bar{p}_a$ that might be derived from (simulation) statistics. Furthermore, we consider a coverage-related average rate r^{COV} and the bandwidth B_a that is available at a on average. For an average spectral efficiency e_a on the transmission links from a to arbitrary users, we suggest

$$p_a^{\mathrm{MAX}} = \frac{B_a\, e_a}{r^{\mathrm{COV}}} \bar{p}_a \tag{7.3}$$

as an estimation for the maximum feasible sum of covered priorities.

The weighting factors λ_{COV} and λ_{CAP} are introduced for the coverage term and the capacity term in the CCO objective function. Several variants of the scalarization approach from Section 2.3.1 are applied in order to achieve Pareto optimal solution or to implement constrained single target maximization.

Although our optimization models utilize the DN model to abstract from single UEs, it is facilitating in some of the subsequent sections to interpret DNs as UEs. UEs can be located indoors or outdoors and have a time-varying rate demand and mobility. Indoor UEs move according to a random walk model and do not leave the building area, whereas outdoor UEs are either pedestrians or vehicles. Pedestrians move according to a random walk model and vehicles move along the roads following a random waypoint model. The mobility models are defined more precisely in Section 7.5.1.

7.1.1 Radio Resource Management and Scheduling

We apply the following model to disregard the scheduling of PRBs over time from our investigations. Consequently, the RRM can consider the tasks of admission control, user assignment, PRB allocation, and interference mitigation independently over consecutive time instances.

Typically, the RRM assigns available PRBs to UEs such that the UE rate demands are fulfilled. However, UEs with a low rate demand may not require a full PRB for the whole assignment period but need one PRB at some time instances. In practical LTE systems, the multiple usage of identical PRBs over time is implemented by an according scheduling algorithm. We model this scheduling in the time domain by an equivalent scheduling in the frequency domain. More precisely, we allow a separate allocation of SCs over the assignment period, and hence, consider a finer granularity in the frequency domain as usually supported. This approach is motivated exemplarily in Figure 7.2, where we assume that serving UE 1 requires the allocation of 45 kHz transmission bandwidth. This is realized by an assignment of every fourth PRB, where T describes the smallest time interval before a PRB reallocation is possible. We transform this scheduling over time into an assignment in the frequency domain such that UE 1 gets the first, fifth, and ninth subcarrier instead of the first, fifth, and ninth PRB. This assignment is equivalently realizable in the time domain for choosing the assignment period as $T' = 12T$. The period T' is assumed to be sufficiently small such that the channel conditions do not change (significantly) over T'.

The RRM tasks are decoupled for consecutive time instances as described above. At each time instance, the following procedures are carried out successively:

1. (Initial cell assignment) First, all users in the system are initially assigned to the supplier cell with the best pilot SINR. The SINR considers full interference in the whole spectrum.
2. (User reassignment & drop users) In general, the user distribution is non-uniform, and therefore, most of the HetNet cells typically experience a different cell load. This fact can lead to overloaded cells, i.e., some users cannot be served sufficiently

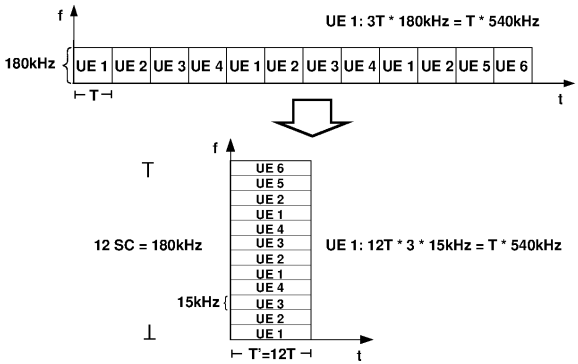


Figure 7.2: Modeling the scheduling of PRBs over time by subcarrier allocation.

by their associated supplier cell. In such a situation, we try to reassign the users that cannot be served to neighboring cells which are not overloaded. Users that require only a small amount of additional resources are preferred for the reassignment step. The user reassignment is repeated until there are no overloaded cells anymore or if there are no further candidates for a reassignment. If there are no reassignment candidates left but there still exists an overloaded cell, some users in that cell have to be dropped. The first iteration of this block does not consider interference, whereas all subsequent calls take the current interference situation into account.

3. (Resource protection) The considered system applies interference mitigation for ICIC. Therefore, (some) neighboring cells can be asked to avoid the usage of particular resources, i.e., to protect (block) certain SCs in the spectrum. The potential benefit of a resource protection is assessed before sending the request to the neighboring cells. This keeps the related communication overhead low.

4. (Resource allocation) The information from the preceding steps serves as input for the resource allocation step. The conditionally optimal resource allocation is performed for all cells in parallel. The parallel execution is possible since the resource assignment – for allocation and protection – is fixed at this stage. If it turns out that there does not exists any feasible allocation that supports the rate requirements of all considered users, more users have to be reassigned or dropped in the second iteration step. All subsequent steps are repeated.

Please note that some systems might (re-)assign users to a supplier cell with respect to a *load balancing* strategy [96]. This strategy can be included into the second iteration step.

7.1.2 Dynamic Parameter Adaption

In contrast to the relatively dense UE distribution that is typically considered for network dimensioning and cell site planning, the UEs in the operating system can emerge locally bounded and very sparsely over the network area. Furthermore, the UE distribution and the rate demand distribution can change significantly over time. For instance, just a few UEs at night can turn into thousands of UEs during lunch time. The spatial and temporal diversity of the UE distribution influences the input parameters for our optimization approach. In particular, the DN information and the frequency reuse factor, which predominantly determines the interference impact factors for MCs in the IAM, see (5.3). Therefore, those input parameters have to be adapted dynamically and subject to the present – or near future – network situation.

Demand Node Parameters

According to the approach for DN generation from Section 3.3, we create a DN grid that partitions the whole network area into a set of (non-equal) patches. The DN grid is computed for the scenario from Figure 7.1 prior to the optimization step. It remains constant throughout the whole system simulation.

Resulting from an initial patch length of 100 m (quadratic), we obtain a total number of $|\mathcal{T}| \approx 2400$ DNs. These DNs subdivide into outdoor DNs and indoor DNs by roughly one third and two-thirds, respectively. For each DN $t \in \mathcal{T}$, we compute the expected rate demand r_t for the next operation cycle according to (3.9). Analogously, the expected amount of accumulated UE priorities p_t is computed. Please note that we generate the DN information for the numerical evaluations in Section 7.5 with respect to identically repeated UE simulations. Consequently, we can assume a perfect traffic forecast with reliability indicator $\mu = 1$ in (3.9).

Interference Approximation Parameters

We apply the following approach to calibrate the IAM parameters for all MCs in the HetNet subject to the DN information that we have obtained according to the previous section. Please recall that we do not assume frequency reuse potential for FCs ,and hence, we do not have to calibrate any FC-related IAM parameters.

First, the implemented RRM algorithms compute the (average) bandwidth consumption components $b^{\mathrm{trm}}_{\mathrm{RRM}}$ and $b^{\mathrm{itf}}_{\mathrm{RRM}}$ for each MC subject to the considered DNs. This are the reference consumption values that the system RRM achieves for the present transmitter configurations. Please note that these results depend on the particular RRM of the considered system. Thus, the presented approach allows for a system-specific calibration of the IAM parameters.

Second, the (average) bandwidth consumption components $b^{\mathrm{trm}}_{\mathrm{BRC}}$ and $b^{\mathrm{itf}}_{\mathrm{BRC}}$ are computed according to the BRC equation (5.1). This step considers (3.7), (5.2), (5.3) with $\beta_s = 1$, and the cell assignment from the RRM step.

The frequency reuse factor in (5.3) is estimated as

$$b_{\mathrm{BRC}}^{\mathrm{trm}} + \hat{\beta}_s \, b_{\mathrm{BRC}}^{\mathrm{itf}} \stackrel{!}{=} b_{\mathrm{RRM}}^{\mathrm{trm}} + b_{\mathrm{RRM}}^{\mathrm{itf}} \quad \Longleftrightarrow \quad \hat{\beta}_s = \frac{b_{\mathrm{RRM}}^{\mathrm{trm}} + b_{\mathrm{RRM}}^{\mathrm{itf}} - b_{\mathrm{BRC}}^{\mathrm{trm}}}{b_{\mathrm{BRC}}^{\mathrm{itf}}} \tag{7.4}$$

with respect to the values $b_{\mathrm{RRM}}^{\mathrm{trm}}, b_{\mathrm{RRM}}^{\mathrm{itf}}, b_{\mathrm{BRC}}^{\mathrm{trm}}, b_{\mathrm{BRC}}^{\mathrm{itf}} \, (> 0)$ for MC s. This estimation particularly takes into account that an underestimation of blocked (protected) bandwidth can compensate an overestimated transmission bandwidth and vice versa. This property of the BRC was discussed in Section 5.3.

The consumption values might be provided by corresponding monitoring entities that have logged the values in the previous operation cycle(s). This mechanism avoids the computation of consumption values from scratch. The logged information is sufficient if we assume a comparable system behavior for the next operation cycle. Furthermore, the estimated reuse factor for the present configuration $s \in \mathcal{Y}_i$ of MC $i \in \mathcal{S}^{\mathrm{MC}}$ is assumed to hold for all configurations from $\mathcal{Y}_i$. Nevertheless, different configurations still yield different interference impact factors in the IAM since the ratio of spectral efficiencies in (5.3) changes with the reconfiguration of transmitters.

The parameter calibration according to (7.4) is exemplarily carried out in a dynamic network simulation for the HetNet from Figure 7.1 – the simulation setup is discussed in detail in Section 7.5.1. We obtain estimation results that are consistent with the results for the hexagonal network topology in Section 5.3: As illustrated by Figure 7.3, the outer MCs in the network, i.e., the cells $1, 5, 9$ and 12, yield the high frequency reuse factors. According to the definition in (5.3), this corresponds to a low frequency reuse potential. On the other hand, the inner cells 4 and 10 yield the lowest frequency reuse factors for both considered configurations. Both effects are caused by the individual number of reuse neighbors and the particular UE distribution during the simulation period.

Overall, the average frequency reuse factors vary from 0.65 to 1.12. Please note that frequency reuse factors larger than one indicate strong compensation effects between $b_{\mathrm{BRC}}^{\mathrm{trm}}$ and $b_{\mathrm{BRC}}^{\mathrm{itf}}$ due to the gap to SINR related bandwidth allocation. This can happen if MCs are (almost) isolated from interfering neighbors. This is the case for the outer MC 12, where the eNB is shadowed towards the middle parts of the network area by high buildings, compare Figure 7.1 (b). The estimated frequency reuse factor for MC 12 decreases significantly when the eNB is powered up in the second configuration and the interference situation gets stronger.

Generally, the bandwidth approximations according to the BRC and the IAM get more accurate the stronger the interference situation between neighboring cells is. This is not a serious drawback since we are mostly interested in high cell load situations that typically suffer from strong interference. On the other hand, neither interference nor reconfiguration is a crucial matter if cells are only slightly loaded. Throughout the HetNet simulations in Section 7.5, we observe that the total bandwidth consumption is approximated very accurately with deviations of less than 2% on average.

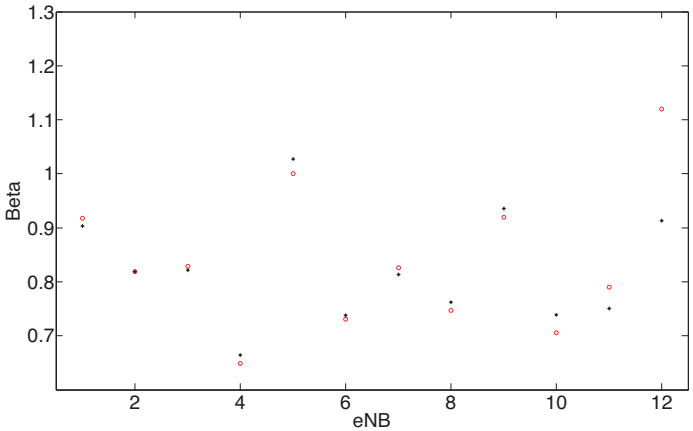


Figure 7.3: Average frequency reuse factors for two different configurations (○ and ∗) of the MCs in a HetNet simulation over 60 minutes.

7.2 System Performance Measures

We introduce coverage and capacity related performance measures that assess the status of these KPIs in the operating (simulated) HetNet. These measures allow to track the effects that our CCO approach yields. It is important to note that we distinguish KPI *performance measures* from KPI related *optimization metrics*. The latter ones serve as maximization targets in the objective function of the CCO model in Section 7.3. Naturally, the performance measure of a KPI and the corresponding optimization metric are highly correlated.

Measure (Metric)	**Related KPI**	**Reference period**
System operation (simulation)		
Cumulated priorities of covered UEs	Coverage	Observation period
Cumulated priorities of served UEs	Capacity	Observation period
Sum rate of served UEs	Capacity	Observation period
Coverage indicator	Coverage & Capacity	One time step
Capacity indicator	Capacity	One time step
System optimization		
Cumulated priorities of covered DNs	Coverage	Operation cycle
Sum rate of served DNs	Capacity	Operation cycle

Table 7.2: Considered system performance measures and optimization metrics.

The considered KPI performance measures and optimization metrics are summarized in Table 7.2. The measures refer either to the assessment of the related KPI status at one time step or they describe the KPI performance for the whole observation period. An observation period typically comprises one operation cycle. While the operation cycle that is considered for CCO takes place in the (near) future, an observed period refers to the time instances of the last operation cycle. Consequently, the effects of a CCO related (re-)configuration of the HetNet can be observed earliest at time instances of the next operation cycle. The effects can be fully assessed at the end of the next observation period. The optimization metrics refer to a whole operation cycle since it is impractical to carry out the CCO procedure for every time instance.

All cells in the HetNet monitor the relevant network status information, e.g.,

- the number of served calls and the number of unattended calls,
- the data rate demand and the priority of each UE, and
- the allocated (utilized) and the blocked (protected) bandwidth.

7.2.1 Assessment of Network Coverage

The *cumulated priorities of covered DNs*

$$\sum_{t \in \mathcal{T}} p_t \tilde{z}_t \tag{7.5}$$

serve as optimization metric for network coverage. A DN is covered if the signal level at its reference coordinate fulfills a minimum SINR requirement.

The *cumulated priorities of covered UEs* over the whole operation cycle serve as corresponding performance measure. The measure assesses the coverage performance over an observation period in the operating (simulated) system. Averaging the measure over the length (duration) of the observation period allows for a direct comparison of the obtained value with the optimization metric. Please recall that the DNs from the optimization metric model average values for the considered operation cycle.

The following *coverage indicator* describes the coverage status at a certain time instance. It can be used to investigate the time-variant coverage performance during simulation. Since coverage is not directly measurable in the operating system, we define

$$\frac{\text{no. of served calls}}{\text{no. of served calls + no. of unattended calls}} \tag{7.6}$$

as coverage indicator. In contrast to the coverage definition from before, an UE is defined as covered if its call, i.e., its service related data rate demand, can be served by any (H)eNB. Thus, the coverage indicator takes the actual user distribution into account as well as the specific UE rate demands and the present resource allocation at (H)eNBs. The *unattended calls* comprise all UEs that experience an insufficient SINR (non-covered UEs) and the UEs that cannot be served due to a lack of available transmission bandwidth (non-served UEs). Due to the latter property, the indicator (7.6) is

also related to the system capacity, particularly in situations where the assessed cell is overloaded. This ambiguity does not apply if (7.6) is evaluated over artificial coverage measurement UEs, i.e., UEs that request a minimal data rate and that have static locations in the network area.

7.2.2 Assessment of Network Capacity

The *sum rate of served DNs*

$$\sum_{t \in \mathcal{T}} r_t^{\text{eff}} \,, \tag{7.7}$$

serves as optimization metric for the system capacity. A DN is served if the link quality and the available resources at the associated (H)eNB are sufficient to provide at least a certain minimum data rate to the DN.

The *sum rate of served UEs* over the whole operation cycle serves as corresponding performance measure. The measure assesses the capacity performance over an observation period in the operating (simulated) system. Analogously to the coverage performance measure from above, the capacity measure can be compared directly to the optimization metric after averaging it over the observation period. The *cumulated priorities of served UEs* over the observation period might serve as an additional capacity performance measure. This measure decreases if covered UEs cannot be served due to a lack of bandwidth for serving all users jointly. Particularly, the second measure can help to resolve the ambiguity problems in the coverage performance measure from Section 7.2.1: If the capacity measure is smaller than the cumulated priorities of covered UEs it clearly indicates capacity problems.

Since the HetNet applies interference mitigation as ICIC technique, we distinguish the bandwidth that is allocatable (free) for serving users from bandwidth that is blocked (protected) to mitigate interference to other cells. Thus, the *free bandwidth* is given by the difference between the *total available transmission bandwidth* and the *blocked bandwidth*. We define the ratio of the bandwidth that is utilized for serving users and the free bandwidth as *utilization ratio* at an (H)eNB. The term

$$1 - \frac{\text{bandwidth utilized for serving users}}{\text{free bandwidth}} = 1 - \text{utilization ratio}$$

might serve as *capacity indicator* for a certain time instance. The indicator can become quite small even though only a few UEs in the cell are served. This effect arises if the amount of bandwidth that is blocked to mitigate interference gets very large and the utilization ratio consequently increases. Since the capacity indicator from above can overemphasize the influence of surrounding cells in terms of blocked bandwidth, we redefine it as

$$1 - \omega \text{ utilization ratio} - (1 - \omega)\frac{\text{free bandwidth}}{\text{total bandwidth}} \,. \tag{7.8}$$

The last term adds a partial amount of blocked bandwidth to the cell capacity subject to the chosen parameter $\omega \in [0, 1]$. This partial amount of blocked bandwidth can be

interpreted as *potential capacity*. It bears analogy to the compensation effects in the BRC in Section 5.3 and Section 7.1.2.

In contrast to the capacity measures, the capacity indicator describes the *remaining capacity* in terms of bandwidth that is left for serving (additional) users. The system RRM typically aims at serving all requested user rates. Hence, we assume that the maximal utilization of available bandwidth is equivalent to the maximization of the served sum rate when cells are highly loaded. Consequently, an optimization of the capacity metric from above leads to effects that are directly measurable by the capacity indicator, at least in capacity shortage situations. This can be seen as a *duality* relation between the sum rate and the capacity indicator. The duality property is particularly relevant for choosing the correct minimum thresholds in the evaluation framework in Section 7.5, where the different performance measures and indicators are compared.

The spectral efficiency might be additionally taken into account to assess the system capacity. However, we do not consider this measure since our RRM implementation does not maximize the spectral efficiency. The simulation results would not show any corresponding improvement that is (potentially) achieved by our CCO approach.

7.3 Joint Coverage and Capacity Optimization

The MIP in Table 7.3 formalizes the optimization problem of *Joint Coverage and Capacity MAXimization (JoCoCaMAX)* with respect to the parameters and variables from Table 7.1. The optimization parameters are adapted according to Section 7.1.2. The sets

$$\begin{aligned}
\mathcal{S} * \mathcal{T} &= \{(s,t) \in \mathcal{S} \times \mathcal{T} \ : e_{st} \geq e_{\text{MIN}}\} \ , \\
\mathcal{S}_t &= \{s \in \mathcal{S} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \ , \quad \mathcal{T}_s = \{t \in \mathcal{T} \ : (s,t) \in \mathcal{S} * \mathcal{T}\} \ , \\
\mathcal{F} * \mathcal{T} &= \{(f,t) \in \mathcal{F} \times \mathcal{T} \ : e_{ft} \geq e_{\text{MIN}}\} \ , \\
\mathcal{F}_t &= \{f \in \mathcal{F} \ : (f,t) \in \mathcal{F} * \mathcal{T}\} \ , \quad \mathcal{T}_f = \{t \in \mathcal{T} \ : (f,t) \in \mathcal{F} * \mathcal{T}\} \ , \\
\mathcal{F}_s &= \{f \in \mathcal{F} \ : e_{sf} \geq e_{\text{MIN}}\} \ , \\
\mathcal{T}_i &= \bigcup_{s \in C_i} \mathcal{T}_s
\end{aligned}$$

are defined to exclude decision variables and constraints that are irrelevant due to insufficient link quality (CQI 0). The set $\mathcal{T}_i$ covers all DNs that can be assigned to MC $i \in \mathcal{S}^{\text{MC}}$ for at least one configuration of the related eNB.

We consider the optimization problem

$$\max \left\{ \lambda_{\text{COV}} \sum_{t \in \mathcal{T}} p_t \tilde{z}_t + \lambda_{\text{CAP}} \sum_{t \in \mathcal{T}} r_t^{\text{eff}} \right\} \tag{7.9}$$

with respect to the optimization variables y_a, $\tilde{z}_{at}$, z_{at}, b_{at} and the auxiliary variables b_a^{itf}, b_a, $\tilde{z}_t$, z_t, r_t^{eff}.

$$\max \left\{ \lambda_{\mathrm{COV}} \sum_{t \in \mathcal{T}} p_t \tilde{z}_t + \lambda_{\mathrm{CAP}} \sum_{t \in \mathcal{T}} r_t^{\mathrm{eff}} \right\}$$

subject to

$$\sum_{s \in \mathcal{Y}_i} y_s = 1 \ , \qquad \text{for all } i \in \mathcal{S}^{\mathrm{MC}}$$

$$\sum_{f \in \mathcal{Y}_j} y_f = 1 \ , \qquad \text{for all } j \in \mathcal{F}^{\mathrm{FC}}$$

$$\tilde{z}_{at} \leq y_a \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T}$$

$$\tilde{z}_t = \sum_{s \in \mathcal{S}_t} \tilde{z}_{st} + \sum_{f \in \mathcal{F}_t} \tilde{z}_{ft} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T}$$

$$\frac{e_{at}}{e_{a't}} \geq (\tilde{z}_{at} + y_{a'} - 1)\, \delta_{\mathrm{ITF}} \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T},\ a' \in \mathcal{S}_t \cup \mathcal{F}_t$$

$$\sum_{t \in \mathcal{T}_a} p_t \tilde{z}_{at} \leq p_a^{\mathrm{MAX}} \ , \qquad \text{for all } a \in \mathcal{S} \cup \mathcal{F}$$

$$z_{at} \leq y_a \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T}$$

$$z_t = \sum_{s \in \mathcal{S}_t} z_{st} + \sum_{f \in \mathcal{F}_t} z_{ft} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T}$$

$$r_t^{\mathrm{eff}} = \sum_{s \in \mathcal{S}_t} e_{st} b_{st} + \sum_{f \in \mathcal{F}_t} e_{ft} b_{ft} \geq r_t^{\mathrm{MIN}} z_t \ , \qquad \text{for all } t \in \mathcal{T}$$

$$b_a = \sum_{t \in \mathcal{T}_a} b_{at} + b_a^{\mathrm{itf}} \leq B + (1 - y_a) \cdot \infty \ , \qquad \text{for all } a \in \mathcal{S} \cup \mathcal{F}$$

$$b_f^{\mathrm{itf}} = \sum_{\substack{(f',t) \in \mathcal{F} * \mathcal{T}, \\ f' \neq f}} q_{ff't} b_{f't} \ , \qquad \text{for all } f \in \mathcal{F}$$

$$b_s^{\mathrm{itf}} = \sum_{\substack{(s',t) \in \mathcal{S} * \mathcal{T}, \\ s' \neq s}} q_{ss't} b_{s't} + \max_{\substack{f \in \mathcal{F}_s, \\ y_f = 1}} \left\{ \frac{e_{sf}}{e_{\mathrm{MAX}}} b_f \right\} \ , \qquad \text{for all } s \in \mathcal{S}$$

$$\sum_{t \in \mathcal{T}_i} p_t \tilde{z}_t \geq p_i^{\mathrm{MIN}} \ , \quad \sum_{t \in \mathcal{T}_i} r_t^{\mathrm{eff}} \geq r_i^{\mathrm{MIN}} \ , \qquad \text{for all } i \in \mathcal{S}^{\mathrm{MC}}$$

Table 7.3: MIP formulation of the joint coverage and capacity maximization problem for LTE HetNets (JoCoCaMAX).

According to the discussion in Section 7.2, we assume the maximization of covered priorities to maximize the system coverage and assume the maximization of the effective sum rate to maximize the system capacity. Although the terms in objective (7.9) look similar to the ones that are considered for cell site planning in (6.1), the coverage indicator $\tilde{z}_t$ and the effectively served data rate r_t^{eff} differ substantially from before: Coverage of a DN is now defined by a minimum SINR requirement and each DN is uniquely assigned to either one eNB or one HeNB for data transmission. In the following, both properties are defined by corresponding constraints. The weigh-

ting factors λ_{COV} and λ_{CAP} determine the trade-off relation between the conflicting objectives of this MOO problem.

Generally, a DN – and its related priority value – can be covered by a uniquely assigned (H)eNB a. This is modeled by the constraints

$$\tilde{z}_{at} \leq y_a \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T} \tag{7.10}$$

$$\tilde{z}_t = \sum_{s \in \mathcal{S}_t} \tilde{z}_{st} + \sum_{f \in \mathcal{F}_t} \tilde{z}_{ft} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T} \ . \tag{7.11}$$

The unique configuration of (H)eNB a is ensured by the constraints (7.1) and (7.2). The DN coverage requirement is modeled as

$$\frac{e_{at}}{e_{a't}} \geq (\tilde{z}_{at} + y_{a'} - 1)\, \delta_{\mathrm{ITF}} \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T},\ a' \in \mathcal{S}_t \cup \mathcal{F}_t \tag{7.12}$$

with respect to all potentially interfering transmitters a'. The constraint becomes a tautology for all decision variables $y_{a'}$ that equal zero. For $\delta_{\mathrm{ITF}} = 1$, only best link coverage is allowed. Please note that the coverage requirement (7.12) can also be written analogously to (6.4), i.e., a conflict graph constraint prohibits the joint selection of certain DN-to-(H)eNB associations and (H)eNB configurations. However, we prefer the formulation from above since it reflects the technical relationship. Furthermore, it avoids the precomputation of a (large) conflict graph. As noise is considered for the computation of the spectral efficiency parameters, constraint (7.12) actually models a minimum SINR condition. For sufficiently large signal powers, however, the interpretation as a minimum SIR requirement is adequate.

We want to avoid degraded CCO solutions that can cause a coverage overkill at some (H)eNBs, see Section 7.1. Therefore, the sum of maximally coverable priorities is bounded by

$$\sum_{t \in \mathcal{T}_a} p_t \tilde{z}_{at} \leq p_a^{\mathrm{MAX}} \ , \qquad \text{for all } a \in \mathcal{S} \cup \mathcal{F} \tag{7.13}$$

where p_a^{MAX} is predefined according to (7.3) and can differ for various transmitters.

Analogously to (7.10) and (7.11),

$$z_{at} \leq y_a \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T} \tag{7.14}$$

$$z_t = \sum_{s \in \mathcal{S}_t} z_{st} + \sum_{f \in \mathcal{F}_t} z_{ft} \leq 1 \ , \qquad \text{for all } t \in \mathcal{T} \tag{7.15}$$

are applied for the capacity (transmission) related decision variables z_{st}, z_{ft} and z_t. Consequently, the DN rate demand can be served by either one configured eNB or one configured HeNB. The data transmission to a DN is directly connected to the

allocation of bandwidth b_{at}, which is computed according to (3.7). The maximum served rate is bounded by the actual requested rate r_t, i.e.,

$$b_{at} \leq \frac{r_t}{e_{at}} z_{at} \ , \qquad \text{for all } (a,t) \in (\mathcal{S} \cup \mathcal{F}) * \mathcal{T} \ . \tag{7.16}$$

Even though this means that the total DN rate demand has not necessarily to be fulfilled, the effectively served rate

$$r_t^{\text{eff}} = \sum_{s \in \mathcal{S}_t} e_{st} b_{st} + \sum_{f \in \mathcal{F}_t} e_{ft} b_{ft} \geq r_t^{\text{MIN}} z_t \ , \qquad \text{for all } t \in \mathcal{T} \tag{7.17}$$

has to exceed at least the minimum rate requirement r_t^{MIN}. Otherwise, the DN cannot be assigned to any supplying station. This problem can arise if the available (remaining) bandwidth resources at all potential suppliers are not sufficient, i.e., the maximum bandwidth constraints

$$b_a = \sum_{t \in \mathcal{T}_a} b_{at} + b_a^{\text{itf}} \leq B + (1 - y_a) \cdot \infty, \qquad \text{for all } a \in \mathcal{S} \cup \mathcal{F} \tag{7.18}$$

would be violated by serving another DN. We assume $B = B^{\text{MC}} = B^{\text{FC}}$. The infinity term on the right hand side of (7.18) avoids irrelevant feasibility problems for non-selected configurations, compare (6.12). Analogously to Section 6.2, the constraints (7.17) and (7.18) can cause a multiple knapsack problem if $r_t^{\text{MIN}} > 0$.

The IAM from Chapter 5 is applied as follows. The interference related bandwidth consumption b_f^{itf} at HeNBs is computed as

$$b_f^{\text{itf}} = \sum_{\substack{(f',t) \in \mathcal{F} * \mathcal{T}, \\ f' \neq f}} q_{ff't} b_{f't} \qquad \text{for all } f \in \mathcal{F} \tag{7.19}$$

whereas the eNBs block (protect) bandwidth according to

$$b_s^{\text{itf}} = \sum_{\substack{(s',t) \in \mathcal{S} * \mathcal{T}, \\ s' \neq s}} q_{ss't} b_{s't} \ + \ \max_{\substack{f \in \mathcal{F}_s, \\ y_f = 1}} \left\{ \frac{e_{sf}}{e_{\text{MAX}}} b_f \right\} \qquad \text{for all } s \in \mathcal{S} \ . \tag{7.20}$$

The cross-tier interference term follows suggestion (5.6), i.e., we assume that all UEs served by a femtocell are located nearby the HeNB and that the same holds for all inter-FC interfered femtocells.

Finally, the constraints

$$\sum_{t \in \mathcal{T}_i} p_t \tilde{z}_t \geq p_i^{\text{MIN}} \ , \quad \sum_{t \in \mathcal{T}_i} r_t^{\text{eff}} \geq r_i^{\text{MIN}} \ , \qquad \text{for all } i \in \mathcal{S}^{\text{MC}} \tag{7.21}$$

enforce a minimum supported level of coverage and capacity for each MC in the network. The requirements are independent of the selected configuration at the related eNB. During the iterative optimization procedure in Section 7.4, these constraints restrict the degradation of coverage and capacity if configurations are changed. The preservation of the performance from previous optimization steps particularly allows for a monotone improvement of solutions in an iteratively conducted CCO processing.

7.3.1 Variants for Trade-Off Optimization

According to the discussion in Section 2.3.1, every optimal JoCoCaMAX solution of (7.9) is Pareto optimal for $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) > \mathbf{0}$. However, the particular weighting factors determine which Pareto optimal points are found. Generally, the weights can be chosen arbitrarily by the network operator. In Section 7.4, we show how a reasonable setting can be derived from the reference thresholds that the network operator typically considers for performance tracking in its network monitoring center.

We consider the JoCoCaMAX problem from (7.9) for $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) > \mathbf{0}$. Previously achieved KPI values are preserved according to (7.21) whenever the present coverage status and the present capacity status are both at a sufficient level. This leads to a monotone performance improvement. The following approach deals with insufficient KPIs whenever one or both KPIs are degraded intolerably: We choose weighting factors $\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}} \in \{0, 1\}$ in order to enable constrained single target maximization of a primary (worst case) KPI, see Section 2.3.2. This allows the system to counteract insufficient performance in a hierarchical manner. The restriction of the potential degradation of the secondary KPI is realized by defining according parameters r_i^{MIN} and p_i^{MIN} for each MC i in (7.21). This leads to the following two optimization variants.

Restricted Coverage Maximization

The *Restricted Coverage MAXimization (RCovMAX)* problem for an MC $i \in \mathcal{S}^{\mathrm{MC}}$ is defined as the JoCoCaMAX problem with $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) = (1, 0)$. According optimization parameters $p_j^{\mathrm{MIN}}, r_j^{\mathrm{MIN}}$, $j \in \mathcal{S}^{\mathrm{MC}}$, ensure a minimum achieved sum of covered priorities and a minimum achieved sum rate in MC i. Furthermore, they preserve the present KPI status in all MCs $j \neq i$ for the sake of a global improvement.

The objective of the RCovMAX problem is the improvement of a single MC, i.e., the worst case MC i. Therefore, the considered DNs in (7.9) are restricted to $\mathcal{T}_i$ while the optimization variables are still the same as for the JoCoCaMAX problem. This means that RCovMAX is solved with respect to the whole configuration state space and all optimization parameters. For the particular case of applying RCovMAX for MC i, objective (7.9) can be rewritten as

$$\max \sum_{t \in \mathcal{T}_i} p_t \tilde{z}_t \ .$$

The subsequently presented integrated CCO approach allows a degradation of the capacity performance in MC i by choosing r_i^{MIN} smaller as the presently achieved sum rate. This gains more degrees of freedom to improve the coverage in MC i.

Restricted Capacity Maximization

The *Restricted Capacity MAXimization (RCapMAX)* problem for an MC $i \in \mathcal{S}^{\mathrm{MC}}$ is defined as the JoCoCaMAX problem with $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) = (0, 1)$. According optimization parameters $p_j^{\mathrm{MIN}}, r_j^{\mathrm{MIN}}$, $j \in \mathcal{S}^{\mathrm{MC}}$, ensure a minimum achieved sum of covered priorities and a minimum achieved sum rate in MC i. Furthermore, they preserve the present KPI status in all MCs $j \neq i$ for the sake of a global improvement.

Analogously to above, the DNs in (7.9) are restricted to $\mathcal{T}_i$. The optimization variables for capacity maximization at MC i are the same as for the JoCoCaMAX problem. This means that the RCapMAX is solved with respect to the whole configuration state space and all optimization parameters. For the particular case of applying RCapMAX for MC i, objective (7.9) can be rewritten as

$$\max \sum_{t \in \mathcal{T}_i} r_t^{\mathrm{eff}} \ .$$

The subsequently presented integrated CCO approach allows a degradation of the coverage performance in MC i by choosing p_i^{MIN} smaller as the presently achieved sum of covered priorities. This gains more degrees of freedom to improve the capacity in MC i.

7.4 Traffic Light Based Self-Optimization

We propose the following *integrated CCO approach* for autonomous self-optimization of the operating HetNet. It is based on the KPI performance measures that were introduced in Section 7.2 and utilizes the CCO model from the previous section. The system continuously tracks the performance of all cells. The observation of insufficient system performance might automatically trigger optimization procedures, i.e., an *aperiodic optimization*. Such a mechanism requires complex algorithms for degradation detection and network status classification. Particularly, it has to be defined how long a detection phase has to be in order to provide reliable detection results. Hence, we apply time triggered optimization, i.e., *periodic optimization*, and consider the implementation of an aperiodic optimization trigger for future research.

Our integrated CCO approach is applicable for a single MC, for a cluster of MCs, or for the whole network. For sufficiently long trigger intervals it makes sense to apply CCO at least in a semi-centralized way, i.e., cluster-wise established control instances (master units) collect all necessary information and perform the CCO approach centrally in their cooperation cluster. Clustering can provide a beneficial trade-off between

computational complexity and optimization quality. The chance to find the global optimum, however, increases with the internal cardinality of the considered coordination clusters. We define a *cooperation cluster* as a group of MCs such that the strongest interferers to the contained eNBs are elements of the group. Since the interference situation depends on the particular user distribution, the clustering step is performed at the beginning of the CCO approach. In the following, we assume to have a suitable clustering algorithm on hand.

The following consecutive subroutines are executed everytime the integrated CCO approach is triggered:

1. DN parameters and IAM parameters are adapted according to Section 7.1.2 subject to the present – or the near future – network situation.

2. Cooperation clusters are identified with respect to the computed DN and IAM parameters. The DN priority values and the DN demands are assigned to the cooperation clusters proportionally to the probability that they are served by an MC or FC from the cluster. The probability for being served by a certain (H)eNB can be derived from the UE assignment statistics over the last operation cycle(s). The total priority and the total rate demand of a DN do not depend on any particular clustering, though a DN can be considered for CCO by multiple cooperation clusters.

3. The master unit of each cluster applies the traffic light related CCO loop that is depicted in Figure 7.4. The CCO loop is explained below.

4. If the CCO results can improve the HetNet performance, the corresponding optimal configurations are applied.

The *Traffic Light (TL)* related *CCO loop* from Figure 7.4 works as follows: First, the KPI traffic lights are computed for all CCO candidates by solving the JoCoCaMAX problem from Section 7.3 for the presently applied (H)eNB configurations. Since all transmitter configurations are fixed for this evaluation step, the solution space is significantly reduced and the traffic lights are obtained very quickly. Please note that these are the expected future KPI traffic lights if the DN information includes the prediction component as introduced in (3.9). The sum of covered DN priorities and the sum rate that a CCO candidate MC achieves in the JoCoCaMAX solution determine its coverage and capacity measure, respectively. Each KPI related traffic light indicates a green, yellow, or red KPI status subject to the corresponding measure and the traffic light thresholds. The traffic light threshold parameters from Table 7.4 are customizable and have to be predefined by the network operator. The coverage and capacity metrics that are obtained in the evaluation step provide feasible values for the minimum required number of covered DN priorities p_i^{MIN} and the minimum required amount of served rate r_i^{MIN} in (7.21). Thus, we know the (minimum) achievable KPI performance of each candidate MC i for all further optimization steps.

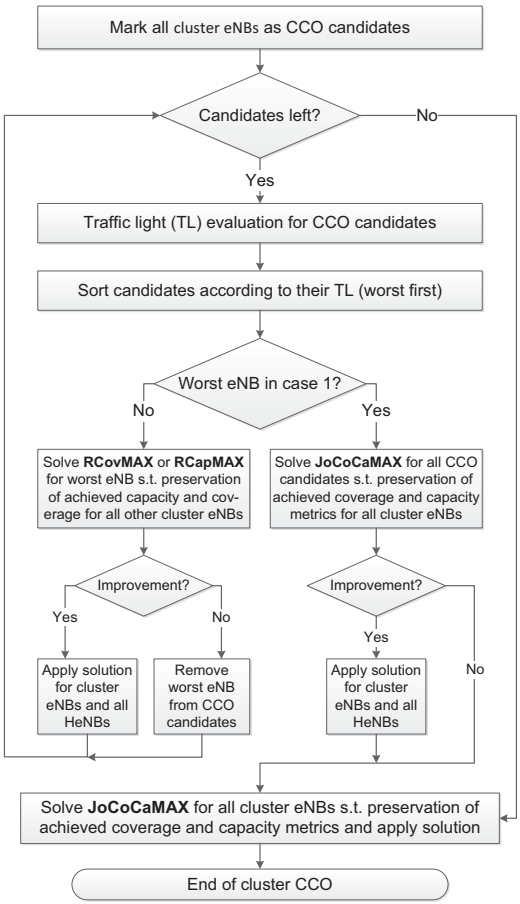


Figure 7.4: Closed-loop autonomous CCO in a cooperation cluster, the related traffic light cases are depicted in Figure 7.5.

Depending on the particular constellation of its KPI traffic lights, each CCO candidate MC belongs to one of the *performance cases (categories)* that are listed in Figure 7.5. All CCO candidate MCs are sorted according to their performance cases. The worst case MC is chosen for the optimization step. If there are multiple MCs in the worst case category, we choose the one with the worst coverage performance. There are basically three different CCO variants that can be applied for the worst case MC subject to the optimization strategy that is defined for each performance case in Figure 7.5:

Symbol & domain	Description
$\mathrm{TH}_{\mathrm{G}}^{\mathrm{COV}},\ \mathrm{TH}_{\mathrm{Y}}^{\mathrm{COV}} \in [0,1]$	Threshold factors referring to sufficient (green), critical (yellow), and insufficient (red) coverage performance.
$\mathrm{TH}_{\mathrm{G}}^{\mathrm{CAP}},\ \mathrm{TH}_{\mathrm{Y}}^{\mathrm{CAP}} \in [0,1]$	Threshold factors referring to sufficient (green), critical (yellow), and insufficient (red) capacity performance.
$\theta^{\mathrm{COV}},\ \theta^{\mathrm{CAP}} \in [0,1]$	Threshold factors referring to lower bounds for coverage performance and capacity performance if constrained single target optimization is applied.

Table 7.4: Traffic light related threshold parameters that determine the KPI status in the CCO loop and define the minimum KPI performance requirements..

1. If the worst case MC belongs to performance category 1, all CCO candidate MCs have a sufficient (expected) coverage performance and a sufficient (expected) capacity performance. The JoCoCaMAX problem from Section 7.3 is solved with respect to all CCO candidate MCs and subject to the preservation of all green traffic lights. The performance of all MCs that are not a CCO candidate is preserved as well.

2. If the worst case MC belongs to performance category $3, 4, 7, 8$ or 9, the RCovMAX problem from Section 7.3.1 is solved for this MC. The corresponding minimum required sum rate for the MC is guaranteed by every feasible solution. For all other MCs in the cluster, the minimum constraints (7.21) are chosen such that the related traffic lights can not get worse.

3. If the worst case MC belongs to performance category $2, 5$ or 6, the RCapMAX problem from Section 7.3.1 is solved for this MC. The corresponding minimum required sum of covered DN priorities for the MC is guaranteed by every feasible solution. For all other MCs in the cluster, the minimum constraints (7.21) are chosen such that the related traffic lights can not get worse.

If the optimization step yields any improvement, the corresponding configurations are applied for all eNBs and HeNBs in the cooperation cluster. In case that we do not achieve an improvement, all configurations stay as before. If it is not possible to improve a worst case MC that has at least one non-green traffic light, this MC is removed from the CCO candidate list. The CCO loop is repeated until there is no MC left in the candidate list. Please note that the MCs which are removed from the CCO candidate list are still considered for the configuration state space of the optimization models. They are just not considered anymore for the selection of the worst case MC.

The CCO loop is repeated as long as there are CCO candidate MCs left and at least one candidate MC is not in the performance category 1. If these conditions do not hold anymore, the JoCoCaMAX problem is finally solved one time for all cluster MCs subject to the preservation or improvement of all achieved KPI traffic lights. The CCO loop terminates after this step.

Case	Coverage / Capacity	KPI Status: Green (TH_G)	KPI Status: Yellow (TH_Y)	KPI Status: Red (θ)	Applied Optimization Model and KPI Bounds
1	Coverage	●			JoCoCaMAX
	Capacity	●			$p_i^{MIN}(TH_G^{COV})$ $r_i^{MIN}(TH_G^{CAP})$
2	Coverage	●			RCapMAX
	Capacity		●		$p_i^{MIN}(TH_G^{COV})$
3	Coverage		●		RCovMAX
	Capacity	●			$r_i^{MIN}(TH_Y^{CAP})$
4	Coverage		●		RCovMAX
	Capacity		●		$r_i^{MIN}(TH_Y^{CAP})$
5	Coverage	●			RCapMAX
	Capacity			●	$p_i^{MIN}(TH_Y^{COV})$
6	Coverage		●		RCapMAX
	Capacity			●	$p_i^{MIN}(TH_Y^{COV})$
7	Coverage			●	RCovMAX
	Capacity	●			$r_i^{MIN}(\theta^{CAP})$
8	Coverage			●	RCovMAX
	Capacity		●		$r_i^{MIN}(\theta^{CAP})$
9	Coverage			●	RCovMAX
	Capacity			●	$r_i^{MIN}(\theta^{CAP})$

Figure 7.5: Traffic light related performance cases (best first) and the corresponding optimization strategy for an MC i.

The ordering of the performance cases and the corresponding optimization strategies in Figure 7.5 clearly indicate the preference to consider coverage as primary KPI whenever the performance of this KPI is not indicated by a green traffic light. Particularly, for status 7 – 9 (insufficient coverage) we allow the capacity in a solution to degrade to a red traffic light in favor of the maximal degree of freedom to maximize coverage. The CCO loop yields a monotone improvement of the cluster MCs, starting with the worst performing MC first. A mathematical interpretation of this *climbing up principle* is presented in the next section.

We consider the parameter settings of the optimization strategies in Figure 7.5 and take the predefined threshold parameters from Table 7.4 into account. For MC i, we define

$$p_i^{\mathrm{MIN}}(\mathrm{TF}^{\mathrm{COV}}) = \mathrm{TF}^{\mathrm{COV}} \cdot \sum_{t \in \mathcal{T}_i} p_t$$

for threshold factor $\mathrm{TF}^{\mathrm{COV}} \in \{\mathrm{TH}_{\mathrm{G}}^{\mathrm{COV}}, \mathrm{TH}_{\mathrm{Y}}^{\mathrm{COV}}, \theta^{\mathrm{COV}}\}$ and

$$r_i^{\mathrm{MIN}}(\mathrm{TF}^{\mathrm{CAP}}) = \mathrm{TF}^{\mathrm{CAP}} \cdot \sum_{t \in \mathcal{T}_i} r_t$$

for threshold factor $\mathrm{TF}^{\mathrm{CAP}} \in \{\mathrm{TH}_{\mathrm{G}}^{\mathrm{CAP}}, \mathrm{TH}_{\mathrm{Y}}^{\mathrm{CAP}}, \theta^{\mathrm{CAP}}\}$.

Furthermore, we suggest to choose the scalarization parameters λ_{COV} and λ_{CAP} in the JoCoCaMAX objective (7.9) proportionally to the relation between the coverage and the capacity threshold factors $\mathrm{TH}_{\mathrm{G}}^{\mathrm{COV}}$ and $\mathrm{TH}_{\mathrm{G}}^{\mathrm{CAP}}$, e.g., as

$$\lambda_{\mathrm{COV}} = \frac{1}{\sum_{t \in \mathcal{T}} p_t} \ , \tag{7.22}$$

$$\lambda_{\mathrm{CAP}} = \frac{{}^{\mathrm{TH}_{\mathrm{G}}^{\mathrm{CAP}}}/_{\mathrm{TH}_{\mathrm{G}}^{\mathrm{COV}}}}{\sum_{t \in \mathcal{T}} r_t} \ . \tag{7.23}$$

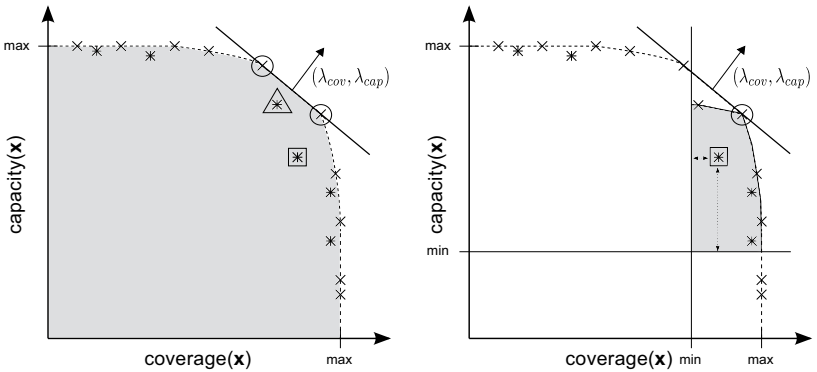
The final setting of the threshold parameters is up to the network operator. The network operator typically considers comparable parameters for system evaluation and black list generation in its network monitoring center.

7.4.1 Climbing Up Principle for Monotone Performance Improvement

Generally, the CCO loop from Figure 7.4 tries to maximally improve the worst performing MC towards performance case 1 from Figure 7.5. It drops the MC from the CCO candidate list if an improvement is not possible anymore. The way upwards can take place stepwise if the improved MC is not directly shifted to performance case 1: If another CCO candidate MC is in a lower performance category (higher performance case number) after the improvement step, the worst case MC for the next CCO iteration switches. Any change of the HetNet configuration in this next iteration step can influence the CCO solution space of the priorly improved MC. Particularly, a change of the solution space might support a further improvement of the prior MC when it is considered the next time for CCO.

The minimum constraints (7.21) are the key components of this improvement strategy. They are automatically parametrized with respect to the results from the traffic light evaluation step, i.e., according to the solution of the JoCoCaMAX problem for fixed (H)eNB configurations. Please note that the present configurations can result from a previous CCO iteration. Figure 7.6 (a) illustrates the set of feasible points (gray area) and the Pareto front (dashed line) of the MOO problem at the beginning of the CCO loop, compare Section 2.3.1. The initial solution point (boxed asterisk) is obtained as solution of the JoCoCaMAX problem for fixed configurations at all (H)eNBs. Since the fixed configurations are usually not optimal for the present HetNet situation, the initial solution point is most likely not located on the Pareto front of the unrestricted JoCoCaMax problem. However, we use the obtained performance metrics to define minimum bounds for the coverage performance and the capacity performance. The corresponding minimum values have to be achieved by any feasible solution in the next optimization step.

Our CCO approach implements a *local search heuristic*, which is inspired by [57]. It works according to the following *climbing up principle*:



(a) Initial solution point (boxed asterisk) obtained by solving the JoCoCaMAX problem for $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) > \mathbf{0}$ and fixed configurations.

(b) Minimum performance constraints reduce the maximum degradation of single objectives and the set of feasible points.

Figure 7.6: Constrained CCO approach for monotone performance improvement.

1. The JoCoCaMAX problem is solved for the present (H)eNB configurations. The obtained solution gives the feasible point $\boldsymbol{x}$, see the boxed asterisk in Figure 7.6.
2. We attempt to find a better feasible point $\boldsymbol{x}'$ in the neighborhood of $\boldsymbol{x}$ by a constrained optimization approach. In Figure 7.6 (b), the corresponding constraints determine the maximum allowed gap in the coverage and the capacity domain subject to the considered threshold factors. Please recall that we allow a degradation of KPIs for certain constellations. Thus, the gap can be larger than zero. The constraints particularly reduce the set of feasible points, i.e., the solution space for the subsequent optimization step. In Figure 7.6 (b), the convex hull of the constrained feasible set excludes the upper encircled point from Figure 7.6 (a) – that was formerly feasible and Pareto optimal – because it violates the minimum coverage constraint. On the other hand, the formerly unsupported point that is marked with a triangle in Figure 7.6 (a), now becomes a supported point on the Pareto front of the JoCoCaMax problem.
3. The predetermined optimization strategy defines the applied CCO variant, i.e., JoCoCaMax, RCovMax, or RCapMax. The corresponding solution gives the feasible point $\boldsymbol{x}'$. Note that we can always find a feasible point since the initial point $\boldsymbol{x}$ is still an element of the feasible set. For the example in Figure 7.6 (b), we solve the JoCoCaMAX problem for scalarization parameters $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) > \mathbf{0}$. These parameters define the depicted tangential line and lead to the encircled Pareto optimal point, compare Section 2.3.1.
4. If the point $\boldsymbol{x}'$ is better than $\boldsymbol{x}$ in terms of the applied optimization strategy, we apply the corresponding configurations, replace $\boldsymbol{x}$ by $\boldsymbol{x}'$, and continue with the CCO loop.

7.5 Numerical Evaluation

In the following, we investigate performance and behavior of our integrated CCO approach for autonomous self-organized optimization of coverage and capacity. For this purpose, we consider several representative case studies. Our integrated CCO approach includes the dynamic adaption of DN parameters and IAM parameters according to Section 7.1.2. The parameter adaption is carried out with perfect knowledge of the considered optimization cycle, i.e., we assume best case parameter estimation with reliability $\mu = 1$, compare (3.9). The aim of this section is to provide a proof of concept for our approach and to illustrate achievable performance gains. With respect to this purpose, we consider a reduced set of possible transmitter configurations in order to keep the computational effort low. Furthermore, the CCO loop is applied for one cluster, i.e., for the whole HetNet. The investigation of further improvements due to a larger configuration state space is open for future work. The same holds for the analysis of potential performance loss due to clustering.

The HetNet system simulation is implemented in MATLAB, whereas we utilize *CPLEX* [51] as MIP solver to compute (optimal) solutions for the CCO variants.

7.5.1 Simulation Setup

We consider the urban HetNet from Figure 7.1. The HetNet has 12 MCs (eNBs). The applied system model includes the RRM implementation as introduced in Section 7.1.1. Users are simulated with time-variant rate demand and mobility as follows: UEs enter the network area with a randomly chosen lifetime such that the expectation of active UEs stays constant during simulation. Similarly to the setup in Section 4.3, each active UE requests a certain mobile service that is specified by its service type, a uniformly distributed data rate demand, and its proportion on the overall traffic. Table 7.5 shows the available mobile services and their corresponding priority level. This service specification is a modified version of Table 4.5. It is chosen such that the HetNet operates at the limit of its capabilities most of the time and is temporarily overloaded. Since we consider priority one for all services, the total sum of coverable priorities equals the number of active UEs in the network.

Mobile service	Priority	Proportion	Rate demand [kbps]
Data	1	10 %	128 – 2000
Web	1	50 %	128 – 512
VoIP	1	40 %	64 – 128

Table 7.5: Considered mobile services for UEs.

The associated UE *mobility model* determines how a UE moves across the simulation area over time. Partially adopted from [95], we consider the following UE mobility models: Indoor UEs move according to a *random walk model* and do not leave the building area. Outdoor UEs are either pedestrians that move according to the random

walk model or vehicles that follow a *random waypoint model*. The waypoints are modeled by an appropriate set of road points that describe the irregular course of the roads. At each crossing point, the vehicle UE chooses randomly the direction to go on. The way back is excluded from that decision. All related mobility model parameters are listed in Table 7.6.

Parameter	**Setting**
Duration of one simulation step	1 s
Pause time	Deterministic at 0 s
Random walk direction	Uniform $[0, 2\pi)$
Velocity pedestrian/indoor	Uniform $[0, 1.5]$ m/s
Velocity vehicle	Uniform $[10, 20]$ m/s
Velocity traffic hotspot	3 m/s

Table 7.6: User mobility parameters.

The following *moving traffic hotspot model* generates locally bounded overload situations: We define a certain number of hotspot UEs and a hotspot radius. The hotspot UEs are located circularly around a *hotspot center UE* and stay within the given radius. The hotspot center UE is a vehicle UE that moves along the roads at a given velocity. All hotspot UEs become active when a predefined hotspot activity period starts and change their state to inactivity if the hotspot period is over. The considered traffic hotspot moves along the roads in the middle part of the network area. In Figure 7.1 (b), it is depicted as bunch of small circles and surrounded by eNBs $1-4$.

Table 7.7 shows the considered HetNet system parameters. The setup is basically the same as for the evaluation of the cell site planning approach in Section 6.3. However, some of the settings slightly differ from before due to the fact that we now consider the operating HetNet. The (H)eNB transmission power can be configured as follows: Each of the 12 eNBs can adjust its transmission power according to the specified power profile, i.e., it can apply a low, medium, or high transmission power. Moreover, it has the option to switch off. Each eNB can apply 0° (low) or 5° (high) antenna downtilt for an omnidirectional antenna pattern, compare Figure 6.1 (a). All 200 HeNBs provide just one antenna pattern (omnidirectional) and support the power profile that is specified in the according part of Table 7.7. Thus, eNBs and HeNBs can select from a set of seven and three different configurations, respectively.

The ray optical path loss model from Section 3.1.2 is utilized to compute the spatial signal power distribution for each eNB configuration. The free space path loss model according to Section 3.1 provides this information for HeNB configurations. We precompute the spectral efficiency parameters for all transmitter configurations subject to the discrete rate-power function (3.8) and the LTE SINR requirements that are defined in Table 6.4.

System parameter	Setting
Simulation area	Urban, 2.5 km × 3.5 km, 5 m resolution
Number of buildings	Approximately 2000
Wall penetration loss	10 dB
Shadowing standard deviation	8 dB
Carrier frequency	2 GHz
Effective transmission bandwidth	9 MHz (50 PRBs in LTE 10 MHz mode)
Number of eNBs (macrocells)	12
eNB Tx power profile	$\{-\infty$ (off), 40 (low), 43 (medium), 46 (high)$\}$ dBm
eNB antenna downtilt profile	$\{0°$ (low), $5°$ (high)$\}$
Propagation model eNB	Ray optical, omnidirectional
eNB antenna gain and noise figure	14 dBi and 5 dB
Number of HeNBs (femtocells)	200, uniformly distributed over buildings
Minimum distance HeNB to eNB	35 m
HeNB Tx power profile	$\{-\infty$ (off), 3 (normal), 10 (high)$\}$ dBm
Propagation model HeNB	Free space + wall penetration, omnidirectional
Noise figure HeNB	8 dB
Average number of (active) UEs	215
Number of hotspot UEs	60
Outdoor UEs	80%
Vehicular-to-pedestrian ratio	20%
UE initial distribution	Random, uniformly distributed
UE antenna gain and noise figure	0 dBi and 9 dB

Table 7.7: Considered simulation parameters for the operating HetNet.

Parameter	Setting
DN prediction reliability μ	1
Minimum threshold for coverage δ_{ITF}	0.6
Coverage bound for MCs p_s^{MAX} and FCs p_f^{MAX}	125 and 15
Utilization weighting factor ω	0.9
Coverage threshold factors $\mathrm{TH}_{\mathrm{G}}^{\mathrm{COV}}$, $\mathrm{TH}_{\mathrm{Y}}^{\mathrm{COV}}$, θ^{COV}	0.98, 0.95, 0.90
Capacity threshold factors $\mathrm{TH}_{\mathrm{G}}^{\mathrm{CAP}}$, $\mathrm{TH}_{\mathrm{Y}}^{\mathrm{CAP}}$, θ^{CAP}	0.90, 0.80, 0.70

Table 7.8: Considered CCO parameters.

The integrated CCO approach from Section 7.4 is applied with respect to the optimization parameters that are shown in Table 7.8. According to (7.22) and (7.23), the traffic light factors in Table 7.8 determine the objective weights for joint CCO as

$$\lambda_{\mathrm{COV}} = \frac{1}{\sum\limits_{t\in\mathcal{T}} p_t}, \quad \lambda_{\mathrm{CAP}} = \frac{0.92}{\sum\limits_{t\in\mathcal{T}} r_t}.$$

Please note that the optimization parameters are defined only once in advance to the start of the HetNet simulation (operation). Any further modification of parameters

is possible before each operation cycle. However, this requires human intervention, which should only take place in emergency situations.

Our integrated CCO approach determines the optimum HetNet configuration for each operation cycle. When the optimal configurations have been applied at all (H)eNBs, we go back to the beginning of the considered operation cycle and repeat the UE simulation. This evaluation scheme allows for a fair comparison between non-optimized and optimized configuration results.

7.5.2 Case Study Results

The following case studies present the results of a comparison between the HetNet performance without CCO and the performance when our CCO approach is applied. For comparison of coverage and capacity performance, we mainly consider the coverage indicator (7.6) and the capacity indicator (7.8) for $\omega = 0.9$. The indicators assess the system performance at each time instance of the simulation. The corresponding traffic light status of both KPIs is computed subject to the traffic light threshold values that follow from the threshold factors given in Table 7.8 and the maximum achievable KPI.

The threshold values for capacity have to consider the duality relation between the capacity indicator (7.8) and the capacity metric (sum rate), see Section 7.2.2. Hence, the traffic light thresholds for capacity follow from Table 7.8 as 0.2 and 0.1 for the yellow and the red traffic light status. Any capacity indicator above 0.2 corresponds to a green traffic light. The traffic light status can be depicted or we can accumulate the number of different traffic lights over a certain period.

Furthermore, we consider the coverage measure and the capacity measures from Table 7.2. Consequently, we assess the number of covered UEs, the number of served UEs, and the sum rate for each operation cycle (observation period). Please recall that in our simulation setup the number of UEs equals the number of coverable priorities.

For any MC (eNB) specific discussion, we refer to the enumeration in Figure 7.1.

Obliteration of Events

We start with a medium- to long-term CCO trigger: We consider an operation cycle of one hour and a traffic hotspot activity of five minutes, taking place from 10 to 15 minutes. As explained above, the whole operation cycle is monitored before the CCO loop is triggered for dynamically adapted input parameters. Afterwards, the UE simulation is repeated identically but with optimized configurations at the (H)eNBs. Figure 7.7 shows the coverage indicator and the capacity indicator for each second of the operation cycle and averaged over all MCs in the non-optimized HetNet. The HetNet operates predominantly at performance status one, i.e., with green coverage and capacity traffic lights. This holds for all MCs if we consider the average over the whole operation cycle. The comparison with the performance of the optimized system shows almost no improvement in the CCO results. Contrariwise, the KPI deficiencies at the eNBs 3 and 4 – that are located nearby the traffic hotspot – slightly increase.

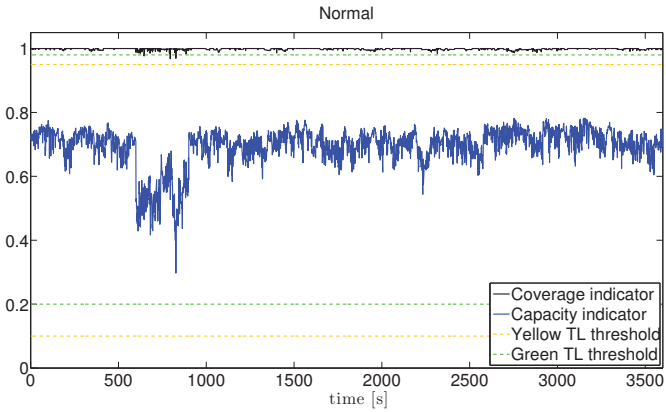


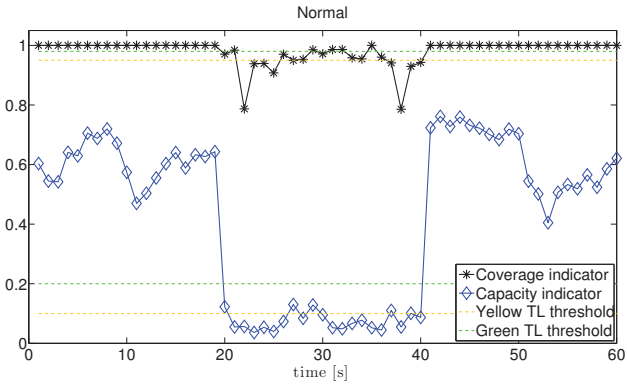
Figure 7.7: Performance indicators for each second of the one hour operation cycle, averaged over all MCs.

This effect is reasonable in the following sense: The generated DN information is averaged over one hour. This leads to an obliteration of the five minutes of hotspot activity, and hence, the system is mainly optimized for the period when the hotspot UEs are inactive. In consequence, there is a performance degradation at the eNBs 3 and 4 during the hotspot activity period. More precisely, the number of time instances where these eNBs operate with a red coverage traffic light and a red capacity traffic light increases from one to seven seconds and from zero to one seconds, respectively.

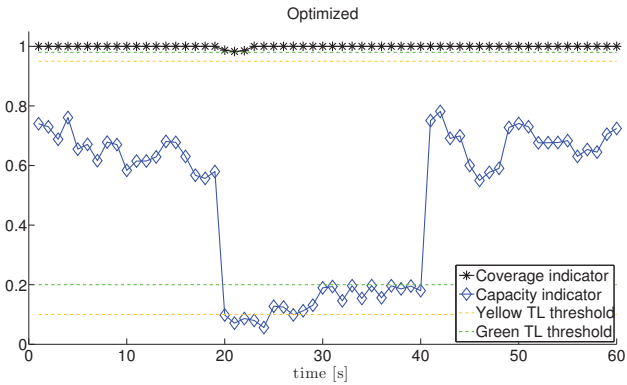
We conclude that averaging the DN information over periods with a significant variation in the UE characteristics is crucial. Hence, we suggest to consider such periods separately for optimization. This can be achieved by applying a trigger rate that attains a large ratio between the duration of the exceptional event (traffic hotspot) and the length of the operation cycle that is considered for CCO. We define this ratio as *event-to-trigger ratio*. The event-to-trigger ratio in the considered example is $1/12$. Apparently, this is too small to achieve a performance improvement.

Temporary Traffic Hotspot

The following setup considers a much higher event-to-trigger ratio: The overall observation period is reduced to one minute and the hotspot UEs are active from 20 to 40 seconds. The CCO loop is triggered every five seconds, i.e., all periods of significantly different UE behavior are fully separated. Moreover, we attain an event-to-trigger ratio of four. This can be interpreted as an oversampling of the hotspot period. In this sense, the considered case study investigates the potential of our CCO approach under ideal assumptions. Although realistic trigger periods and the hotspot duration



(a) HetNet applies default configurations.



(b) HetNet applies configurations according to the CCO results.

Figure 7.8: Second-wise performance indicators at eNB 4, which suffers most from the traffic hotspot.

are typically much longer than in the considered setup, we expect the results to be scale-invariant under the same event-to-trigger ratio.

Figure 7.8 depicts the coverage and capacity indicators at eNB 4, which is most affected by the traffic hotspot. The HetNet applies the default configuration according to the HetNet planning results in Figure 7.8 (a) and the optimum CCO configuration in Figure 7.8 (b). During the hotspot period, the coverage and capacity indicators decrease significantly when the HetNet operates in its default configuration. For the optimum CCO configuration, the coverage indicator is improved to an uncritical performance

status whereas the capacity indicator, though improving, remains in the critical zone when the hotspot is active. Table 7.9 shows the exact number of time instances where the KPIs at eNB 4 are in a critical (non-green) status: The coverage problems are fully resolved by applying the optimized configurations, whereas the capacity deficiencies are improved from a red traffic light to a yellow light for ten time instances.

Traffic light status	Default configuration	CCO configuration
Red coverage indicator	8	0
Yellow coverage indicator	8	0
Red capacity indicator	16	6
Yellow capacity indicator	5	15

Table 7.9: Number of time instances where the KPIs at eNB 4 are not in a green traffic light status.

The significant improvement during the hotspot period is mainly achieved by reducing the interference at the eNB that serves the hotspot UEs. When the hotspot UEs become active, the eNBs $1, 2, 3$ and 7 increase their downtilt (tilt down) and decrease their transmission power. Consequently, they reduce their interference to the service area of eNB 4. On the other hand, eNB 4 increases its downtilt and applies the maximum available transmission power in order to focus on the hotspot area. When the hotspot activity is over, the eNBs $1, 2, 3$ and 7 decrease their antenna downtilt to $0°$ (tilt up) and power up to enlarge the coverage footprint.

Figure 7.8 illustrates highly volatile values for the capacity indicator, which performs sometimes clearly worse than in the non-optimized case. Both effects are caused by our RRM implementation: First, the volatility is explained by the fact that the RRM algorithms are executed at each time instance independently of the assignments and allocations that were computed in previous simulation steps. And second, our RRM does not necessarily maximize the capacity indicator. It terminates when a feasible assignment and allocation have been reached. Thus, our RRM does not take full advantage of the CCO benefits as long as it has not to cope with coverage and capacity problems. In other words, the KPI values of the non-optimized system can outperform the optimized values but *only* when the non-optimized system does not suffer from any KPI deficiencies.

Coverage and Capacity Trade-Off

So far, the unattended calls in the coverage indicator were mostly caused by a lack of available resources (bandwidth), i.e., the coverage indicator mainly accounted for capacity problems. Therefore, the following case study investigates our CCO approach for coverage problems in terms of missing signal power: eNB 2 and the eNBs 7 to 10 are permanently switched off in order to cause artificial coverage problems, particularly in the coverage area of the eNBs 4, 11 and 12, see Figure 7.9 (a). Otherwise, we keep the simulation setup from the previous case study, including the hotspot period from 20

to 40 seconds. First, the HetNet performance is evaluated for the initial (default) cell configurations from the planning stage. Second, the CCO loop is triggered *one time* with respect to the UE traffic that is averaged over the whole observation period of one minute. And third, the 60 seconds are separated into three isochronous operation and optimization cycles such that the period of hotspot activity is exactly covered by the middle cycle. This corresponds to an event-to-trigger ratio of one. Since coverage problems and capacity shortage appear simultaneously during the hotspot period, trade-off handling becomes a serious challenge in the CCO loop.

Performance indicators / Configuration	**Number of users which are not covered**	**covered**	**served**	**Average served rate [Mbps]**
1st cycle (coverage compensation)				
Initial configuration	3	4152	3815	53.382
Optimized once	0	4155	3957	56.255
Optimized per cycle	0	4155	3944	55.932
2nd cycle (hotspot activity)				
Initial configuration	5	5255	4106	53.288
Optimized once	0	5260	4265	55.502
Optimized per cycle	3	5257	4763	67.568
3rd cycle (regular operation)				
Initial configuration	2	3958	3631	52.392
Optimized once	2	3958	3828	56.569
Optimized per cycle	2	3958	3840	57.252

Table 7.10: Covered users, served users, and served rate in the HetNet over the whole observation period of one minute.

Table 7.10 shows that the coverage performance and the capacity performance significantly improve when the initial HetNet configuration is switched to the optimum CCO configuration. One-time CCO and cyclic CCO reveal similar behavior for the first and third cycles. In the second cycle, however, we observe a significant difference: The cyclic CCO reduces the coverage performance only marginally while it considerably improves the capacity performance, i.e., the number of served users and the served rate. The one-time optimization has to take into account the coverage problems from the first and third cycles. This reduces the degrees of freedom for improving the capacity performance in the second cycle. Please note that the simulation results can slightly differ from the computed CCO solutions, and hence, minor discrepancies are possible.

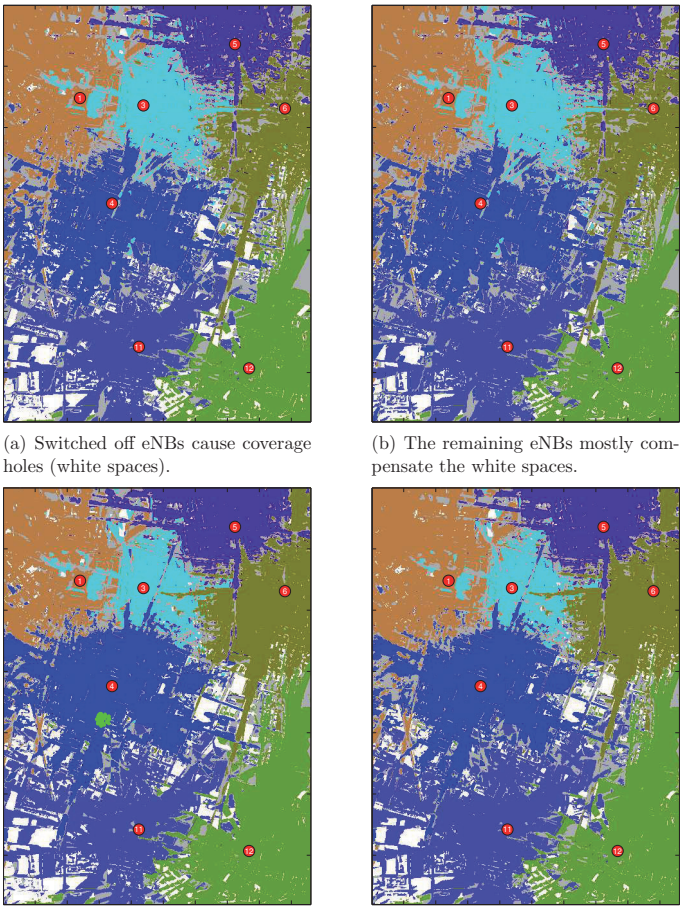
Table 7.11 and the (re-)configuration related MC footprints that are depicted in Figure 7.9 illustrate how the autonomous CCO approach basically works: The initial cell configurations lead to *coverage holes* in the HetNet area with a received signal power below $-121\,\mathrm{dBm}$. In Figure 7.9 (a), coverage holes particularly occur at the cell border of the MCs $4, 11$ and 12. Thus, the CCO loop for the first cycle considers these coverage holes as well as severe capacity problems in the other cells. The corresponding CCO solution keeps the initial antenna tilts but powers up all eNBs and many of the HeNBs. This solution improves the coverage situation significantly, compare the white

eNB	**1**	**3**	**4**	**5**	**6**	**11**	**12**
Initial config. (tilt/power)	high/med.	high/med.	low/med.	low/med.	low/med.	low/med.	high/med.
1st cycle (coverage compensation)							
Main CCO problems	Most eNBs suffer from capacity problems and eNBs 11 and 12 have additional coverage problems.						
Resulting config. (tilt/power)	high/high	high/high	low/high	low/high	low/high	low/high	high/high
Main effects	eNBs keep initial tilt but power up, 40% FCs operate at high power. eNBs 1, 4 and 11 reduce coverage holes.						
2nd cycle (hotspot activity)							
Main CCO problems	All eNBs have capacity problems but the coverage performance is tolerable. eNB 4 is most affected by the traffic hotspot.						
Resulting config. (tilt/power)	high/low	high/low	high/high	high/high	high/high	low/med.	high/high
Main effects	eNBs 4, 5, and 6 tilt down to increase the near distance link quality while eNBs 1, 3, and 11 reduce their interference.						
3rd cycle (regular operation)							
Main CCO problems	Coverage problems dominate at eNBs 1, 4, and 11 but also the degraded network capacity is still a problem.						
Resulting config. (tilt/power)	high/med.	high/low	high/high	high/high	high/high	low/high	high/high
Main effects	eNBs 1 and 11 power up to compensate coverage holes. The number of high power FCs slightly increases.						

Table 7.11: Actions taken by the autonomously running CCO to cope with joint coverage and capacity problems, see Figure 7.9 for a visualization.

spaces in the middle part of Figure 7.9 (b) to subfigure (a). Furthermore, it reduces the capacity problems. In the second cycle, the active hotspot causes strong capacity problems, particularly at eNB 4. The coverage performance is still tolerable. As expectable, eNB 4 applies the high antenna downtilt and operates at full power in order to serve most of the hotspot users. Moreover, all stations – except for eNB 11 – use a high downtilt to reduce their interference to the neighboring cells and, particularly, to the coverage area of eNB 4. Overall, the CCO results for the second cycle lead to a cell footprint which shows more coverage problems as in the first cycle, but in favor of an improved HetNet capacity. In the last cycle, the coverage problems become more important again whereas capacity is still an issue. The coverage problems cannot be fully resolved without reducing the capacity performance significantly. In consequence, the CCO leads – similarly to the first cycle – to power changes at the eNBs and to a slightly increased capacity performance.

This case study points out the coverage and capacity trade-off strategy that our CCO approach applies: Coverage performance is the preferred optimization target but only up to the point when the capacity performance gets intolerable. In this case, the coverage is allowed to decrease slightly in favor of more degrees of freedom for improving the capacity. The particular preferences, however, can be adapted – or even be inverted – by choosing according optimization parameters. This requires a modification of the ordering and the optimization strategies in Figure 7.5. The CCO related reconfiguration provides the operating network the opportunity to focus locally on the traffic



(a) Switched off eNBs cause coverage holes (white spaces).

(b) The remaining eNBs mostly compensate the white spaces.

(c) Reconfiguration with respect to the traffic hotspot (bunch of spots).

(d) Handling of joint coverage and capacity problems.

Figure 7.9: MC footprints for the consecutive handling of coverage problems (switched off eNBs), capacity problems (traffic hotspot), and joint KPI degradation.

hotspot when it is necessary. This is clearly indicated by the resource consumption at eNB 4, which can raise the bandwidth for serving users from round about 66 % in the operation cycles one and three to more than 90 % in cycle two. Moreover, the one-time CCO results demonstrate that there is significant potential for improvements by applying our approach even though the degradation period is not perfectly separated.

7.5.3 The Solution Space for Trade-Off Optimization

The following investigation validates the assumptions that we made on the structure of the CCO solution space for the climbing up principle in Section 7.4.1. Furthermore, it justifies our suggestion for an appropriate selection of the JoCoCaMAX parameters λ_{COV} and λ_{CAP}.

We make use of the property that each optimal solution of the JoCoCaMAX problem (7.9) is a Pareto optimal point on the Pareto front for $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}}) > \mathbf{0}$, see Section 2.3.1. This property helps us to compute a *Pareto front approximation* for the following setup: We consider the JoCoCaMAX problem for MC 4 in the second cycle of the previous case study, i.e., during the traffic hotspot period. Corresponding solutions are computed for an increasing sequence of the scalarization parameters λ_{COV} and λ_{CAP} on the interval $[0.2, 1]$ with step size 0.2. The scalarization parameters λ_{COV} and λ_{CAP} are normalized by

$$\frac{1}{\sum\limits_{t \in \mathcal{T}_4} p_t} \quad \text{and} \quad \frac{1}{\sum\limits_{t \in \mathcal{T}_4} r_t} \,,$$

respectively. The parameters p_i^{MIN} and r_i^{MIN} are equal to zero for all considered cells. Figure 7.10 depicts the coverage and capacity results that are attained by the JoCoCaMAX solutions for eNB 4. The results are normalized to the maximum number of DN priorities and the maximum sum rate that can be achieved. Please note that the plot shows only the JoCoCaMAX results for eNB 4, though, the configuration state space of the related JoCoCaMAX problem includes the configurations of all relevant neighbor MCs and FCs. Redundant points are dropped. In order to have the labeling neatly arranged, we additionally dropped some points that are very close to the depicted ones. Some of the points are not directly located on the approximated Pareto front (dashed line) even though they have to be supported points by definition, see Section 2.3.1. This effect is caused by a timeout bound for the MIP solver. The gap to the optimum capacity value can reach up to 0.5 % when the solver terminates. In consequence, the depicted Pareto front represents an artificial best fit curve and not the actual convex hull of the solution set.

The points are labeled with their corresponding (normalized) vector $(\lambda_{\mathrm{COV}}, \lambda_{\mathrm{CAP}})$. The results confirm the assumption from Section 2.3.1 that the dedicated balance or imbalance of the weighting factors influences the balance or imbalance of the achieved coverage and capacity performance in the related solution: The higher valued KPI of a solution point with an imbalanced weight vector corresponds to the higher KPI weight. The two points on the lower right of Figure 7.10, however, illustrate that this is only a relative property and not an absolute one. This finding is in accordance with the statement from Section 2.3.1 that the particular selection of weights influences just the *tendency* of the ratio between the effectively achieved KPIs.

Compared to Figure 7.6, the obtained structure of the solution space is close to our expectations. Moreover, the (encircled) solution point – that corresponds to our weight suggestions (7.22) and (7.23) – meets our intention to find a JoCoCaMAX solution

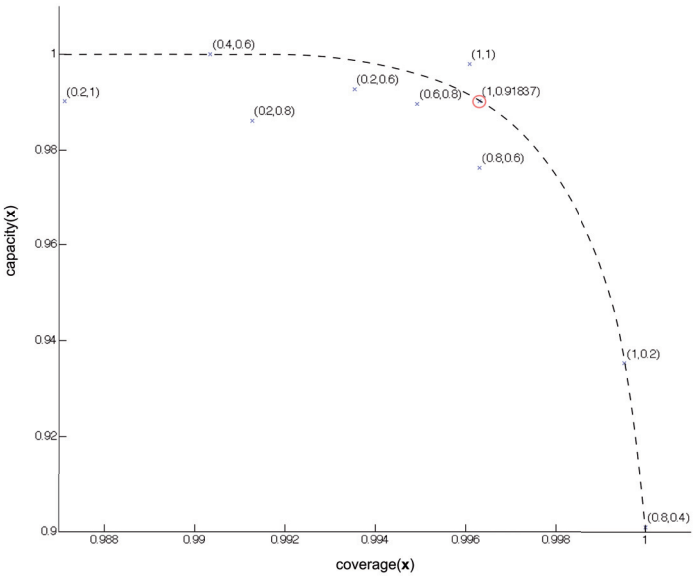


Figure 7.10: Pareto front approximation and selected optimal point (encircled).

that is almost balanced but has a slight preference for a higher coverage performance. The according solution point can be located somewhere else for a different problem instance. However, we expect it to be located in this region of the solution space as long as there are any Pareto optimal points in this region at all.

7.6 Summary

Some works in recent literature provide algorithms for self-organized CCO that have the same intention as our approach. However, most of them follow a contrary implementation principle. For instance, [75] and [86] apply reinforcement learning methods to create an expert system (knowledge database) that defines rule-based decisions for the self-organized configuration of cell parameters. Such a *model-free approach* does not require much a-priori information since it does not consider an explicit system model. Instead, it is solely based on the system feedback to its decisions and actions. In contrast to a model-based approach, the model-free approach can make it very hard to interpret the decisions that are taken. It can even be impossible to influence the

processing of a model-free approach intentionally. Further related aspects are studied in [96], where cell outage management and self-organized eNB downtilt optimization are considered. However, this work do not deal with joint multi-tier configuration but keeps the femtocell environment static.

Most works in the SON domain consider the hexagonal grid network topology according to [3]. We observed effects that were specifically caused by that artificial topology, e.g., uniform frequency reuse potential in MCs due to cell symmetry. Therefore, we have switched to a more realistic simulation framework that considers the propagation characteristics of an urban environment. This allows for the evaluation of our CCO approach under realistic assumptions.

Our model-based CCO approach describes the underlying system dependencies in an explicit system model. It adapts its optimization parameters dynamically subject to the system feedback. Particularly, our approach considers both tiers, i.e., macrocells and femtocells, jointly for the optimization of *user-centric* performance metrics. The user-centric definition of system coverage that we consider might look unusual compared to the common spatial definition. The main advantage of our definition is the additional degree of freedom that we gain for improving the capacity if necessary. We expect this perspective to become more and more important in future when the cells are getting smaller, the cell deployment gets more dense, the average cell load increases, and the systems can react very quickly to dynamic changes in the network.

In Section 7.4, we have introduced a traffic light based scheme for self-organized CCO. The corresponding approach finds Pareto optimal solutions for the joint optimization of coverage and capacity. It confines to a hierarchical constrained single target maximization in the case that a single performance metric predominates. The CCO loop as well as the overall integrated CCO approach meet the definition of a closed-loop optimization framework that is given in [85]. Our CCO model integrates the interference approximation approach that we have introduced in Chapter 5. Thus, cell configurations are optimized subject to inter-cell and cross-tier interference. The cell configuration can particularly influence the transmitter activity status (switch on/off). This provides the opportunity to correct problems that are caused by *peppered* FCs. The transmission power of disruptive FCs can be tuned down to a minimum if necessary. For application in a real system, the periodic CCO trigger can be combined with an aperiodic trigger mechanism that is able to detect system degradations automatically.

The evaluation results demonstrate that our integrated CCO approach can cope with coverage hole situations as well as with a (temporal) traffic hotspot. Both scenarios are defined by 3GPP as reference scenarios for the assessment of CCO solutions in self-organized networks [8]. Moreover, our approach fulfills the following three requirements for self-organized CCO that 3GPP defines in [8]:

1. The need for human intervention is minimized. Our CCO approach does not require any human intervention once the optimization parameters have been initially specified. In fact, the optimization parameters are calibrated automatically subject to the dynamic changes in the network.

2. The static optimization components are fully customizable by the network operator. The static components of our integrated CCO approach are the performance measurement functions, the corresponding traffic light thresholds, and the available configuration state space.
3. The CCO approach is configurable individually for different areas of the network as the optimization parameters are customizable for each cell.

Overall, the presented CCO approach improves the network performance when the HetNet has to cope with traffic hotspots, coverage shortage, and objective trade-off situations. We expect further improvements if the configuration state space is increased, e.g., by more power diversity or sectorized antennas. Related investigations are open for future work.

8 Conceptual Extensions

This chapter presents some conceptual extensions to the optimization models that we have introduced in the chapters before. On the one hand, the extensions incorporate further aspects into the optimization models. The approach in Section 8.1 considers energy efficiency as additional decision criterion for optimal network configuration. Section 8.2 provides a concept to embed user acceptance into the network planning task, whereas Section 8.3 considers the robustness of the related optimization problem. On the other hand, the extensions deal with methods to reduce the computational complexity of the optimization algorithms. Related methods for graph-based problem partitioning and adaptive clustering are presented in Section 8.4.

We briefly discuss the aim of each conceptual extension and explain how it can be applied to the related optimization models. Please note that we skip any comprehensive formalism even though it might be required to apply the extension. Instead, we sketch the extension principles and refer to the related publications.

8.1 Energy Efficiency

Nowadays, energy efficiency plays an important role for any kind of technical development. Concerning the optimization models that we have presented in the previous chapters, the CCO problem from Chapter 7 is the only one that allows for an optimization of power parameters. The corresponding objective (7.9) can be extended to

$$\max \left\{ \lambda_{\mathrm{COV}} \sum_{t \in \mathcal{T}} p_t \tilde{z}_t + \lambda_{\mathrm{CAP}} \sum_{t \in \mathcal{T}} r_t^{\mathrm{eff}} - \lambda_{\mathrm{POW}} \left(\sum_{s \in \mathcal{S}} P_s y_s + \sum_{f \in \mathcal{F}} P_f y_f \right) \right\} , \tag{8.1}$$

where P_s and P_f denote the transmission power that is associated with the according eNB configuration s and the HeNB configuration f, respectively. The weighting factor λ_{POW} depends on the importance of energy efficiency compared to the importance of maximal coverage and capacity. Furthermore, the factor might differ for the power consumption of MC transmitters and FC transmitters. In this case, the power term is separated into differently weighted terms for eNB sum power and HeNB sum power. We suggest to incorporate energy efficiency as a minor criterion, i.e., to choose the weighting factor λ_{POW} relatively to the factors λ_{COV} and λ_{CAP} such that each (average) DN that can be covered or served gets a higher priority in the overall optimization objective than any power reduction.

The integration of energy efficiency into the optimization objective may increase the computational effort to solve the corresponding MIP. Therefore, it can be a beneficial alternative to apply the *power minimization* step consecutively to the CCO loop: After the CCO loop has terminated, we bound the minimum constraints (7.21) by the accumulated priorities and the sum rate that have been achieved for the optimum CCO solution. Afterwards, the JoCoCaMAX problem is solved one more time, but in this optimization step the objective (7.9) is replaced with

$$\min \left\{ \sum_{s \in \mathcal{S}} P_s y_s + \sum_{f \in \mathcal{F}} P_f y_f \right\} . \tag{8.2}$$

The importance of the sum power for MCs can optionally be distinguished from the importance of the FC power consumption by introducing according weighting factors. Please note that any obtained solution to this problem yields the same coverage and capacity performance as before but, perhaps, with a network configuration that consumes less power as the former solution. This consecutive optimization step is very fast since the tight minimum constraints reduce the feasible solution space maximally. Thus, this reduction can also mean that there do not exist alternative feasible solutions.

Operational effort can be considered as a further aspect of energy efficiency or as an additional optimization metric. It is directly related to the configuration changes in CCO solutions. The number and the type of (electrical) antenna downtilt modifications can cause a higher or lower effort. Optimal solutions that take the minimization of configuration changes into account could be favored over solutions that maximize the coverage and capacity performance solely.

For each (H)eNB with configuration a, the parameter $c_{a'} \in \mathbb{R}_{\geq 0}$ denotes the cost for a change from the presently applied configuration to configuration a'. The cost may be interpreted as a penalty term. For instance, it can consider the number of antenna tilt modifications or the difference in total transmission power. The cost parameters are chosen relatively to the presently applied configuration a. The parameters are recomputed each time the (physical) configurations in the HetNet have changed. Naturally, it holds $c_a = 0$. The minimization of operational effort can be incorporated into the JoCoCaMAX problem – or into any of its variants – analogously to (8.1). The corresponding weighting factor λ_{OPE} has to be chosen suitably. Alternatively, the operational effort can be minimized in a secondary optimization step according to (8.2). We recommend the second variant to incorporate it into the CCO procedure, i.e., to solve the optimization problem

$$\min \left\{ \sum_{s \in \mathcal{S}} c_s y_s + \sum_{f \in \mathcal{F}} c_f y_f \right\}$$

for tight minimum constraints after the CCO loop has finished. Due to the identical problem formulations it is possible to define cost parameters that include the related power consumption. The corresponding optimization problem minimizes power consumption and operational effort jointly.

8.2 Embedding User Acceptance as Decision Criterion

The acceptance of people becomes more and more important – and sometimes even crucial – for the planning and the optimization of wireless networks. Even though wireless networks are deployed, expanded, and (re-)configured subject to electromagnetic compatibility, there are controversial discussions on the impact of wireless communication technology to health and environment. Nowadays, web 2.0 applications (facebook, twitter, etc.) provide detractors a powerful platform to broadcast their negative propaganda. As a consequence, the acceptance of mobile communication technology can decrease. Furthermore, worried people will most likely not become customers for the related devices and services.

In [36], we propose an approach for the integration of *user acceptance* as immanent component of the optimization models for the deployment and configuration of wireless networks. In the following, we describe the integration concept mainly for the deployment stage, i.e., with respect to Chapter 6. However, the network operator can also benefit from our approach in the network configuration phase if people appreciate this feature. This is questionable since people cannot see the effects of a reconfiguration, whereas it is very simple for them to distinguish if a transmit antenna is deployed or not.

Symbol & domain	**Description**
$\mathcal{A} = \{1, \ldots, \mathrm{N}_A\}$	Index set of acceptance preferences (APs).
$(i_{j1}, \ldots, i_{j\mathrm{N}_S}) \in \{0,1\}^{\mathrm{N}_S}$	Indication vector of relevant transmitters for AP j.
$n_j \in \mathbb{N}_0$	Maximum number of tolerated transmitters in AP j.
$w_j \in \mathbb{R}_{\geq 0}$	Penalty (weight) if AP j is violated.
$u_j \in \{0,1\}$	Binary decision variable indicating that AP j is violated.

Table 8.1: Additional input parameters (upper part) and variables (lower part) to incorporate user acceptance.

We consider the (additional) input parameters and optimization variables from Table 8.1 and model the acceptance of certain network deployments as follows: First, a socio-economic analysis of the considered planning area provides the input for an appropriate acceptance model, see [36]. The acceptance model identifies (computes) the *Acceptance Preferences (APs)* of user groups that are potentially affected by the network deployment. Each AP of a user group describes the subset of cell site deployment candidates that is relevant for the acceptance in the group. We define $\mathrm{N}_S = |\mathcal{S}|$. All relevant candidates $s \in \mathcal{S}$ have an entry $i_{js} = 1$ in the indication vector of the corresponding AP j. The indicator entries of irrelevant candidates are equal to zero. If an unpreferred cell site can have different configurations, all the corresponding entries from the set of configured transmitters $\mathcal{S}$ are equal to one.

Typically, transmit antennas that are deployed in the neighborhood are more relevant to a user group than the ones that are deployed far away. Nevertheless, there can be

user groups that are not related to any local part of the network, e.g., the citizens of a small city that do not want the network to exceed a certain maximum size. The size of a user group, the acceptance weights that are derived from the acceptance model, and the rank of the group members determine the importance of an AP j. The importance is reflected by the penalty w_j that is caused if the AP is violated. It is not necessary to consider a constant penalty for each unpreferred cell site. Instead, the overall penalty w_j for violating AP j is triggered if a minimum number n_j of unpreferred candidates is selected for deployment. Thus, the specific setting of n_j defines the particular intention of AP j:

1. *The deployment of certain transmitters is not appreciated*: The AP penalty is triggered whenever one or more unpreferred cell sites $s \in \mathcal{S}$ are selected for deployment. The corresponding entries i_{js} are equal to one and $n_j = 0$. Separate APs are created if the penalty shall increase with the number of violations.

2. *The deployment of at most n certain transmitters is tolerated*: The indication vector for the tolerated transmitters is created as before and $n_j = n$.

3. *The total size of the network shall not exceed n cell sites*: All entries of the indication vector are equal to one and $n_j = n$.

Acceptance can be integrated into the HetNet-ProMAX problem from Section 6.2 by replacing objective (6.1) with

$$\max \left\{ \lambda_{\text{BASIC}} \underbrace{\sum_{t \in \mathcal{T}} p_t z_t}_{\text{coverage}} + \lambda_{\text{RATE}} \underbrace{\sum_{t \in \mathcal{T}} r_t^{\text{eff}}}_{\text{capacity}} - \underbrace{\sum_{s \in \mathcal{S}} c_s y_s}_{\text{cost}} - \lambda_{\text{ACP}} \underbrace{\sum_{j \in \mathcal{A}} w_j u_j}_{\text{acceptance}} \right\} . \tag{8.3}$$

The binary variable u_j indicates the violation of AP j. The indicators serve as additional optimization variables. The constraints (6.2) to (6.12) are extended by

$$\sum_{s \in \mathcal{S}} i_{js} y_s - n_j \leq \mathrm{N}_S u_j \ , \qquad \text{for all } s \in \mathcal{S} \ , \tag{8.4}$$

which ensures $u_j = 1$ for each violated AP j. The extended problem has N_A extra optimization variables and N_A additional constraints compared to the original HetNet-ProMAX problem.

The embedding of acceptance into the network deployment decision depends on the weighting factor $\lambda_{\text{ACP}} \geq 0$ for the acceptance (penalty) term in (8.3). The influence of the acceptance term can reach from a harmless integration (low weighting factor) to an equal or dominant criterion (high weighting factor). We refer to [36] for further investigations and discussions. Please note that it can be an alternative to consider the acceptance penalty term analogously to (8.2) or to bound the penalty term by a corresponding maximum constraint.

8.3 Mobility Robustness

The solution quality of all the optimization models that have been presented in this thesis particularly depends on the accuracy of the DN input parameters. The achieved system performance will be poor if the predicted rate demand does not match the effective demand of users in the post-optimization phase. In the operating system, discrepancies in the DN parameter estimation according to (3.9) can be corrected dynamically before the next CCO loop starts. The network dimensioning approach in Chapter 4 and the cell site planning from Chapter 6, however, do not have any chance to correct inaccurate input parameters dynamically.

In [23], the authors show how an approach from the field of *robust optimization* can help to cope with uncertainty in the DN parameters. Basically, this work considers a planning (dimensioning) problem that is similar to the FR-ProMAX problem from Section 4.2. The problem does not consider relay stations. Instead of a fixed rate demand r_t for DN t, the demand is modeled as a bounded random variable over the interval $[\bar{r}_t - \hat{r}_t, \bar{r}_t + \hat{r}_t]$. The parameter $\bar{r}_t$ denotes the nominal rate demand and $\hat{r}_t$ its maximum deviation. The mobility robustness parameter $\Gamma \in [0, \ldots, |\mathcal{T}|]$ describes the number of DNs that deviate from their nominal rate demand. The capacity constraint of the corresponding multiple knapsack problem from Table (4.3) is rewritten as

$$\sum_{t \in \mathcal{T}_s} \frac{\bar{r}_t}{e_{st}} z_{st} + \max_{\substack{\mathcal{T}' \subseteq \mathcal{T}_s, \\ |\mathcal{T}'| \leq \Gamma}} \sum_{t \in \mathcal{T}'} \frac{\hat{r}_t}{e_{st}} z_{st} \leq B_s \ , \qquad \text{for all } s \in \mathcal{S} \ .$$

This constraint includes the worst case constellation into the optimization problem:

i) The subset $\mathcal{T}'$ contains the (maximally) Γ DNs that degrade the capacity constraint most if they simultaneously request a rate at the upper bound of their rate interval.

ii) All DNs of subset $\mathcal{T}'$ simultaneously request a rate at the upper bound of their rate interval.

We refer to [23] for the details how this non-linear constraint can be transformed into an equivalent ILP formulation. The robust optimization problem can be solved with the same methods as the original problem. Compared to the original problem, the robust optimization problem has $|\mathcal{S}| + |\mathcal{S} * \mathcal{T}|$ extra variables and $|\mathcal{S} * \mathcal{T}|$ additional constraints.

Generally, it is preferable to consider too much rate demand (demand buffer) instead of a parameter setup that leads to a permanently overloaded network. On the other hand, a large demand buffer can cause the deployment of dispensable stations. More precisely, a cell site planning problem – or a network dimensioning problem – that considers too much demand buffer will have a solution with many cells that are underloaded in the operating system. Nevertheless, these cells are deployed and operated at full CAPEX and OPEX. Besides the increased problem complexity, this can be interpreted as *the price of robustness*.

8.4 Graph-Based Problem Partitioning and Adaptive Clustering

The integrated CCO approach presented in Section 7.4 is intended to run autonomously in a semi-decentralized manner, i.e., preferably in small- to middle-sized cooperation clusters. We define a cluster as a reasonably sized group of network elements such that the strongest interferers to the contained elements are covered by the group. Thus, (inter-cell) interference is the major decision criterion for network clustering. The network entities are clustered in order to reduce the size of the problem instances, and hence, to reduce the computational effort for solving the particular optimization problems. Moreover, clustering can help to reduce the communication overhead that is caused by self-organization in the system. All the relevant information has to be shared only within the cluster and not along the whole network.

Clustering is also useful to reduce the computational effort that is required to solve the network dimensioning problem from Chapter 4 and the cell site planning problem according to Chapter 6. In the pre-deployment stage, however, the cell site locations are not fixed yet. Thus, the computation of potential clusters – defined by a *strong intra-cluster interference* criterion – becomes a combinatorial problem. The corresponding clustering algorithms cannot be used to partition the planning problem into smaller sub-problems prior to the optimization step. Therefore, the methods for clustering in the operating network differ from the methods that can be applied to find suitable clusters for the planning problem. The clusters of a planning problem can be considered as independent problem instances.

In [38], we present a graph-based approach to partition a network planning problem automatically into an appropriate number of n sub-problems. The n corresponding problem clusters have a strong intra-cluster relationship (interference) and a weak inter-cluster relationship. The basic idea is to model the planning instance as bipartite graph $G = (S, T, E)$ as depicted in Figure 8.1: The set S contains all cell site deployment candidates (BSs or eNBs) from $\mathcal{S}$. The set T covers the considered receiver points (DNs or UEs) from $\mathcal{T}$. The weights of the undirected edges E are equal to the spectral efficiency e_{st} on the corresponding links. If $\mathcal{S}$ contains different configurations for identical cell site locations, each location is considered only once in S. The corresponding weight of each undirected edge to an element of T equals the maximum spectral efficiency over all possible configurations of the cell site.

We connect an *artificial source node* to all nodes from S and consider each node from T successively as *sink node*. The partitioning of the bipartite graph is obtained by the overall *minimum cut* that we find for this setup. The cost of a cut is defined by measures that suitably reflect the intra-cluster interference and inter-cluster interference of the corresponding (two) partitions. All subsequent partitions are clustered in a hierarchical manner until a stop criterion is fulfilled. Basically, the clustering in a partition branch stops when the cut cost of a subsequent partitioning exceeds the relative cost of its parent cluster. We refer to [38] for all technical details.

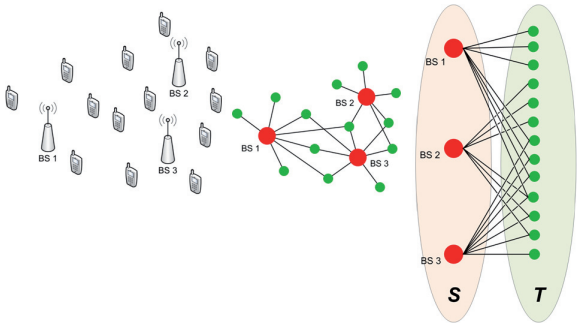
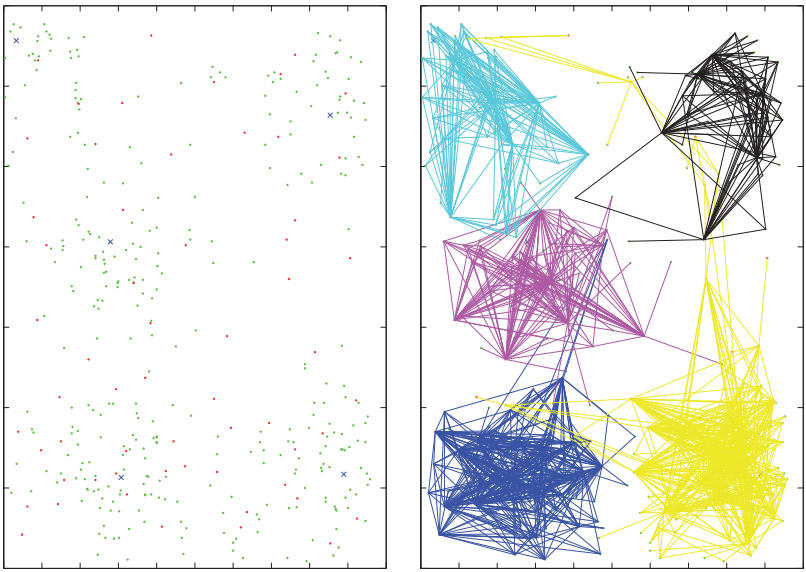


Figure 8.1: Graph-based representation of a wireless network.

Related works from the literature propose alternative approaches for problem partitioning, e.g., variants of the k-means clustering algorithm or partitioning around medoids. In contrast to these approaches, our method does not require any a priori knowledge about the intended number of clusters and does not depend on randomly chosen initial cluster centers. In fact, our approach arranges clusters automatically around their native centers. This is illustrated in Figure 8.2, where the receiver points are distributed over the urban planning area from Figure 6.1 (b) according to a 2-*dimensional Gaussian sampling* [32] on a set of quasi-uniformly distributed points. The distribution means of the five randomly sampled Gaussian distributions are depicted by crosses. Numerical evaluations show that our graph-based partitioning approach achieves significant savings (80 % to 90 %) in the computation time that is required to solve the cell site planning problems. The savings including the time for problem partitioning. The corresponding solution quality is very close to the optimum one (95 % to 99 %). However, we can not find any reasonable partitioning with more than one cluster for planning problems with many densely distributed DNs and deployment candidates.

We can use a much simpler approach for the purpose of dynamic clustering in the operating network: We estimate the interference impact of each transmitter to all transmission links that are assigned subject to a best server criterion. The estimation utilizes the interference approximation model from Chapter 5 and the related IAM equation (5.2). Consequently, each transmitter knows the bandwidth that it has to block (protect) in order to mitigate interference to each particular cell in the network. We define a suitable threshold parameter $b^{\mathrm{CC}} \geq 0$. Each cell with an amount of blocked bandwidth above b^{CC} is joined to the cluster of the interfered transmitter. This step is carried out once for every cell in the network. Existing clusters are merged successively if the threshold condition holds for any two transmitters of both clusters. The threshold value is chosen such that we obtain reasonable cluster sizes. A reasonably sized cluster provides a beneficial trade-off between optimization runtime and solution quality.



(a) Clustered DN distribution.

(b) Computed clustering.

Figure 8.2: Min-cut based partitioning for five Gaussian distributed DN clusters (with cluster centers ×) in the planning area from Figure 6.1 (b).

9 Conclusions

9.1 Summary and Contributions

This thesis contributes to multiple stages along the typical lifecycle of broadband wireless networks by introducing novel concepts, models, and algorithms for

- the estimation of transmission quality and bandwidth requirements subject to path loss predictions,
- the dimensioning of wireless networks,
- interference modeling in multi-tier networks,
- HetNet cell site planning,
- the self-organized optimization of operating HetNets, and
- the conceptual extension of several components from above.

First, the key components for wireless network modeling have been introduced. Path loss information serves as basic input for these components. We have shown how the required bandwidth for data transmission can be computed via the discrete rate-power function that is given by the system link budget. Since the spectrum in OFDMA multi-cell networks is shared and limited, all presented optimization models are subject to a variant of the multiple knapsack problem. This becomes an issue when users expect a minimum quality-of-service and the corresponding rate demand exceeds the load limit of single cells.

We have considered multihop WiMAX networks and determined their infrastructure dimensioning with respect to the expected user distribution and rate demand. The related optimization problem has been formalized as a mixed-integer linear program that covers all relevant technical system aspects. Particularly, we have shown how an economical perspective can help to find a closed-form representation for conflicting objectives like the trade-off alignment between network coverage, network capacity, and deployment cost. Even though the dimensioning model has been developed for multihop WiMAX networks, the approach is transferable to further OFDMA based wireless networks.

Compared to the dimensioning approach, accurate cell site planning and network configuration require a higher precision, particularly in terms of the applied inter-cell interference model. For this purpose, a low-complexity interference approximation has been developed that estimates the overall required bandwidth at eNBs and HeNBs subject to inter-cell and cross-tier interference. This approximate model serves as immanent component of the optimization models that we have introduced for the planning and the self-organized operation of LTE HetNets. The interference approximation can be calibrated with respect to the dynamic changes in the network.

For the optimal cell site planning of LTE HetNets, we have considered macrocells and user-operated femtocells that are not necessarily active all the time. The objective of the corresponding optimization problem is to provide a minimum number of macrocells such that mobile services are area-wide guaranteed. On the other hand, it avoids dispensable cell sites for the sake of cost efficiency and low interference.

We have developed an integrated approach for the self-optimization of coverage and capacity in time-variant LTE HetNets. The corresponding algorithms were designed according to a traffic light principle and they autonomously control site activity, transmission power, and antenna downtilt parameters. The presented approach finds Pareto optimal solutions for the joint optimization of coverage and capacity. It confines to a hierarchical constrained single target maximization in the case that a single performance metric predominates. Simulative evaluations have demonstrated that our approach improves the overall performance when the HetNet has to cope with traffic hotspots, coverage shortage, and objective trade-off situations.

Finally, we have presented several conceptual extensions to the previous optimization problems. The extensions incorporate further decision criteria into the optimization problems, e.g., power minimization and integration of user acceptance. Furthermore, they provide heuristics to cope with the computational complexity of the presented planning problems (problem partitioning) and the online optimization of operating HetNets (adaptive clustering).

The work presented in this dissertation has a strong focus on modeling aspects. Therefore, the development of dedicated optimization models and low-complexity approximations had a higher priority than the implementation of efficient algorithms to solve the optimization problems. In fact, we have utilized state-of-the-art tools such as CPLEX or Gurobi to solve the integer and mixed-integer linear programs without any further modifications. The development of concepts and algorithms for acceleration of the presented optimization procedures is open for future work. Promising approaches are, for instance, relaxation and rounding methods and heuristics like simulated annealing or evolutionary algorithms.

Although not mentioned explicitly, *pico nodes* can be incorporated into our planning and operation approaches as *low capability* eNBs. Here, low capability can mean lower power profiles or fixed antenna configurations. In that sense, this thesis covers multiple aspects that are relevant for future wireless networks *towards and beyond IMT advanced* [82]. These networks are typically characterized by dense picocell deployments and advanced interference coordination.

9.2 Future Research

This dissertation provides several aspects for future research beyond the development of methods for an optimization speed up. The first aspect addresses the quantitative evaluation of the uplink-downlink duality [20, 47] that we implicitly assume. Even though we can expect channel reciprocity in the LTE TDD mode [29, 94], we have not investigated the de facto correlation between the solution quality for the downlink scenario and the uplink scenario yet. A suitable investigation requires particularly the integration of time-variant and frequency selective channel models into the simulation framework.

A further open aspect is the analysis of the impact of an extended configuration state space to the joint coverage and capacity optimization. This analysis might be combined with the assessment of potential performance loss due to (adaptive) clustering in the network.

Finally, the automatic detection of performance degradation effects in the operating network is a missing component. Such a mechanism can improve the applied trigger mechanism for the autonomous self-optimization. It would particularly allow to combine the periodic optimization trigger with a suitable aperiodic trigger. Models and techniques from the field of *Complex Adaptive Systems (CAS)* are promising candidates for automatic degradation detection. Their typical purpose is the detection of structure and emergence in complex and self-organizing systems [28, 19].

Acronyms

2^m-QAM	Quadrature Amplitude Modulation, quantized with respect to 2^m different symbols
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
AP	Acceptance Preference
ARPU	Average Revenue Per Unit
BER	Bit Error Rate
BLER	BLock Error Rate
BPSK	Binary Phase-Shift Keying
BRC	Bandwidth Reservation Concept
BS	Base Station
CAP	Channel Assignment Problem
CAPEX	CAPital EXpenditures
CAS	Complex Adaptive System
CCO	Coverage and Capacity Optimization
CDMA	Code Division Multiple Access
CORLA	Cube Oriented Ray Launching Algorithm
CQI	Channel Quality Indicator
CSI	Channel State Information
DiLaP	Direction-specific Land use based Path loss model
DN	Demand Node
DSL	Digital Subscriber Line
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
eNB	eNodeB
FC	FemtoCell
FDMA	Frequency-Division Multiple Access
FR-ProMAX	Full Rate Profit MAXimization
FTP	File Transfer Protocol
GSM	Global System for Mobile communications
HeMS	HeNB Management System
HeNB	Home eNodeB

HetNet	Heterogeneous Network
HSPA	High-Speed Packet Access
IAM	Interference Approximation Model
ICIC	Inter-Cell Interference Coordination
ILP	Integer Linear Program
JoCoCaMAX	Joint Coverage and Capacity MAXimization
KPI	Key Performance Index
LMSE	Least-Mean-Squared Error
LoS	Line-of-Sight
LP	Linear Program
LTE/SAE	Long-Term Evolution/System Architecture Evolution
MC	MacroCell
MCKP	Multiple-Choice Knapsack Problem
MCLP	Maximal Covering Location Problem
MIMO	Multiple-Input Multiple-Output
MIP	Mixed-Integer linear Program
MKP	Multiple Knapsack Problem
MOO	Multi-Objective Optimization
MSE	Mean-Squared Error
NGMN	Next Generation Mobile Networks alliance
NLoS	Non-Line-of-Sight
NMS	Network Management System
NP	Non-deterministic Polynomial
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	OPerational EXpenditures
PDCCH	Physical Downlink Control CHannel
PRB	Physical Resource Block
ProMAX	Profit MAXimization
PTAS	Polynomial-Time Approximation Scheme
PUCCH	Physical Uplink Control CHannel
QoS	Quality-of-Service
QPSK	Quadrature Phase-Shift Keying
RCapMAX	Restricted Capacity MAXimization
RCovMAX	Restricted Coverage MAXimization
RRM	Radio Resource Management
RS	Relay Station
SC	SubCarrier
SCM	Spatial Channel Model
SINR	Signal-to-Interference and Noise Ratio

SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
SON	Self-Organizing Network
TDMA	Time-Division Multiple Access
TL	Traffic Light
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
VoIP	Voice over Internet Protocol
WiMAX	Worldwide Interoperability for Microwave Access
WINNER	Wireless world INitiative NEw Radio

List of Symbols

$\subseteq, \subset$	Subset and proper subset, respectively.
$\emptyset$	Empty set.
$\mathbb{N}, \mathbb{N}_{>0}$	Set of natural numbers (including zero) and positive natural numbers, respectively.
$\mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{R}_{>0}$	Set of real numbers, non-negative real numbers, and positive real numbers, respectively.
$\lvert\cdot\rvert$	Cardinality of a set.
$\boldsymbol{x}, \boldsymbol{X}$	Bold face lower case letters and upper case letters denote column vectors and matrices, respectively.
$\boldsymbol{x}^T, \boldsymbol{X}^T$	Transpose of a vector $\boldsymbol{x}$ and a matrix $\boldsymbol{X}$, respectively.
$\mathbf{0}$	Zero vector.
$\boldsymbol{x} < \boldsymbol{y}$	Vector $\boldsymbol{x}$ is component-wise smaller than $\boldsymbol{y}$.
$\dim\{\cdot\}$	Dimension of a space.
$\mathcal{O}(\cdot)$	Landau symbol, defining an upper bound for the computational complexity.
a	Wildcard symbol for either a (configured) eNB s or a (configured) HeNB f, p. 48.
b_{*t}	Amount of bandwidth allocated at transmitter $* \in \{s, k, f\}$ for transmission to demand node t, p. 8.
b_a^{itf}	Amount of protected bandwidth at (H)eNB a to cope with interference to other cells, p. 48.
b_a^{trm}	Amount of allocated bandwidth at (H)eNB a for data transmission, p. 48.
b_a	Amount of overall consumed bandwidth at (H)eNB a, p. 48.
B_*	Total available bandwidth at transmitter entity $* \in \{s, k, f\}$, p. 8.
c_*	Cost for the selection of a transmitter entity $* \in \{s, k, f\}$ and the corresponding configuration, p. 36.
$c(\cdot)$	Corresponding land use class of a land use segment, p. 22.
$d_{ss'}$	Distance between the transmitter entities s and s', p. 40.
$d(\cdot)$	Distance to a reference point or length of a path loss segment, p. 16.

e_{*t}	Supported spectral efficiency on the link from transmitter entity $* \in \{s, k, f\}$ to user or demand node t, p. 25.
f	Representative femtocell from the index set $\mathcal{F}$, p. 8.
f_c	Carrier (center) frequency, p. 16.
g_{*tn}	Channel gain from transmitter entity $* \in \{s, k, f\}$ to t on subcarrier n, p. 33.
$G_A(\phi, \psi)$	Antenna gain in direction (ϕ, ψ), p. 16.
G_{fast}	Fast fading component of a path loss model, p. 16.
G_{slow}	Slow fading (shadowing) component of a path loss model, p. 16.
k	Representative relay station from the index set $\mathcal{K}$, p. 8.
$L(t)$	Path loss on the signal path from a transmitter to receiver point t, p. 16.
$L^{\text{dB}}(t)$	Path loss on the signal path from a transmitter to receiver point t on a logarithmic scale, p. 16.
$L_0(t)$	Distance-dependent basic path loss component of $L(t)$, p. 16.
P	Transmit power, p. 16.
P_n	Transmit power on subcarrier n, p. 32.
p_t	Priority level of demand node t, related to either the service type or the corresponding customer type, p. 58.
p_a^{MAX}	Maximum feasible sum of covered priorities at (H)eNB a in CCO solutions, p. 72.
p_i^{MIN}	Minimum required sum of covered DNs (priorities) in the coverage area of MC i, p. 84.
$q_{aa't}$	Interference impact factor from (H)eNB a to the transmission link from (H)eNB a' to t, p. 51.
$R(\cdot)$	Rate-power function, p. 33.
r_t	Requested data rate at demand node t, p. 8.
r_t^{MIN}	Minimum required data rate if demand node t is served, p. 8.
r_t^{eff}	Effectively served data rate at demand node t, depending on the particular signal quality and bandwidth allocation, p. 8.
r_i^{MIN}	Minimum required sum rate that is achieved in the coverage area of MC i, p. 84.
s	Representative cell site from the index set $\mathcal{S}$, p. 8.
t	Representative demand node from the index set $\mathcal{T}$, might context-specifically also represent a single user (receiver), p. 8.
$\text{TH}_{\{\text{G,Y}\}}^{\text{KPI}}$	KPI related threshold factor, referring to sufficient (G), critical (Y), and insufficient performance, p. 88.
u_j	Binary decision variable indicating the violation of acceptance preference j, p. 110.
w_j	Penalty that is associated with the violation of acceptance preference j, p. 110.

x_{tn}	Binary decision variable indicating the assignment of subcarrier n to user t, p. 32.
y_*	Binary decision variable indicating the selection of a transmitter entity $* \in \{s, k, f\}$, p. 8.
z_{*t}	Binary decision variable indicating the assignment of demand node t to a transmitter entity $* \in \{s, k, f\}$, p. 8.
$\tilde{z}_{at}$	Binary decision variable indicating the coverage of demand node t by (H)eNB a, p. 72.
z_t	Auxiliary variable indicating that demand node t is assigned to a transmitter entity, p. 8.
$\tilde{z}_t$	Binary decision variable indicating the coverage of demand node t, p. 72.
$\mathcal{A}$	Index set of acceptance preferences, p. 109.
$\mathcal{C}$	Set of land use classes that are distinguished for path loss computation, p. 21.
$\mathcal{D}$	Domain of an optimization problem, p. 9.
$\mathcal{F}$	Index set of (configured) femtocell HeNBs, p. 8.
$\mathcal{F}^{\mathrm{FC}}$	Index set of available femtocell transmitters in a HetNet, p. 72.
$\mathcal{G}$	Conflict graph, p. 37.
$\mathcal{K}$	Index set of (configured) relay stations, p. 8.
$\mathcal{N}$	Set of subcarriers, p. 32.
$\mathcal{S}$	Index set of (configured) macrocell sites, p. 8.
$\mathcal{S}^{\mathrm{MC}}$	Index set of available macrocell transmitters in a HetNet, p. 72.
$\mathcal{T}$	Index set of demand nodes, p. 8.
$\mathcal{Y}_a$	Index set of all different configurations for an (H)eNB a, p. 72.
β_s	Frequency reuse factor in the related macrocell of site s, p. 49.
Δ_0	Offset constant in path loss models, p. 17.
δ_{ITF}	Minimum threshold for coverage definition, p. 72.
γ	Path loss exponent, p. 16.
Γ	Mobility robustness parameter, p. 111.
λ_{BASIC}	Basic coverage weighting factor for objective scalarization, p. 59.
λ_{RATE}	Sum rate weighting factor for objective scalarization, p. 37.
λ_{COV}	Coverage weighting factor in the CCO approach, p. 73.
λ_{CAP}	Capacity weighting factor in the CCO approach, p. 73.
λ_{POW}	Sum power weighting factor if power consumption is considered in the CCO objective, p. 107.
λ_{ACP}	Weighting factor for the acceptance penalty term if acceptance is considered as objective component, p. 110.

μ	Reliability indicator for demand predictions, p. 29.
ω	Weighting factor for the utilization ratio in the capacity indicator, p. 79.
ρ_f	Average activity rate of HeNB (femtocell) f, p. 59.
σ^2	Thermal noise power, p. 33.
θ^{KPI}	KPI related threshold factor, referring to the lower performance bound if constrained single target optimization is applied, p. 88.

Bibliography

[1] End-to-End Efficiency (E3) project. [Online]. Available: https://ict-e3.eu

[2] 3GPP R1-050507, "Soft frequency reuse scheme for UTRAN LTE," May 2005.

[3] 3GPP R4-092042, "Simulation assumptions and parameters for FDD HeNB RF requirements," Alcatel-Lucent, picoChip Designs, Vodafone, May 2009. [Online]. Available: www.3gpp.org

[4] 3GPP Rel. 10, "Overview of 3GPP release 10," Sep. 2012, version 0.1.6. [Online]. Available: www.3gpp.org

[5] 3GPP Rel. 11, "Overview of 3GPP release 11," Sep. 2012, version 0.1.2. [Online]. Available: www.3gpp.org

[6] 3GPP Rel. 8, "Overview of 3GPP release 8," Sep. 2012, version 0.2.8. [Online]. Available: www.3gpp.org

[7] 3GPP TR 36.902, "Self-configuring and self-optimizing network (SON) use cases and solutions," Dec. 2010, version 9.3.0. [Online]. Available: www.3gpp.org

[8] 3GPP TS 32.521, "SON policy network resource model (NRM) integration reference point (IRP) requirements," May 2011, version 10.1.0. [Online]. Available: www.3gpp.org

[9] 3GPP TS 32.591, "LTE; Telecommunication management; Home enhanced Node B (HeNB) Operations, Administration, Maintenance and Provisioning (OAM&P); Concepts and requirements for Type 1 interface HeNB to HeNB Management System (HeMS)," Oct. 2012, version 11.0.0. [Online]. Available: www.3gpp.org

[10] 3GPP TS 36.213, "LTE evolved universal terrestrial radio access (E-UTRA) physical layer procedures," Oct. 2010, version 9.3.0. [Online]. Available: www.3gpp.org

[11] 3GPP TS 36.410, "E-UTRAN S1 general aspects and principles," Jun. 2010, version 9.1.0. [Online]. Available: www.3gpp.org

[12] 3GPP TS 36.420, "E-UTRAN X2 general aspects and principles," Dec. 2009, version 9.0.0. [Online]. Available: www.3gpp.org

[13] 3GPP Work Items on SON, "Work items on self-organising networks," Oct. 2010, version 0.0.6. [Online]. Available: www.3gpp.org

[14] E. Amaldi, A. Capone, and F. Malucelli, "Radio planning and coverage optimization of 3G cellular networks," *Wireless Networks*, vol. 14, pp. 435–447, Aug. 2008.

[15] Y. Bai, J. Zhou, and L. Chen, "Hybrid spectrum usage for overlaying LTE macrocell and femtocell," in *IEEE Global Communications Conference*, Honolulu, USA, Dec. 2009.

[16] C. Ball, T. Hindelang, I. Kambourov, and S. Eder, "Spectral efficiency assessment and radio performance comparison between LTE and WiMAX," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2008*, Cannes, France, Sep. 2008.

[17] T. Beniero, S. Redana, J. Hamalainen, and B. Raaf, "Effect of relaying on coverage in 3GPP LTE-Advanced," in *IEEE Vehicular Technology Conference Spring 2009*, Barcelona, Spain, Apr. 2009.

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

[19] N. Brodu, "Reconstruction of epsilon-machines in predictive frameworks and decisional states," May 2011. [Online]. Available: arXiv:0902.0600

[20] D. Catrein, L. Imhof, and R. Mathar, "Power control, capacity, and duality of up- and downlink in cellular CDMA systems," *IEEE Transactions on Communications*, vol. 52, no. 10, pp. 1777–1785, Sep. 2004.

[21] Celtic-Plus Initiative. [Online]. Available: http://www.celtic-initiative.org/

[22] ——. GANDALF project. [Online]. Available: http://www.celtic-initiative.org/Projects/Celtic-projects/Call2/GANDALF/gandalf-default.asp

[23] G. Claßen, A. M. C. A. Koster, and A. Schmeink, "A robust optimisation model and cutting planes for the planning of energy-efficient wireless networks," *Computers and Operations Research*, vol. 40, no. 1, pp. 80–90, Jan. 2013.

[24] C. Coello Coello and M. Lechuga, "MOPSO: a proposal for multiple objective particle swarm optimization," in *Proceedings of the 2002 Congress on Evolutionary Computation*, vol. 2, 2002, pp. 1051–1056.

[25] W. Cook, "Fifty-plus years of combinatorial integer programming," 2009. [Online]. Available: http://www2.isye.gatech.edu/∼wcook/papers/ip50.pdf

[26] W. Cook, W. Cunningham, W. Pulleyblank, and A. Schrijver, *Combinatorial Optimization*. John Wiley & Sons Ltd, 1998.

[27] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *IEEE International Conference on Communications*, Ottawa, Canada, Jun. 2012, pp. 2485–2489.

[28] J. P. Crutchfield, "The calculi of emergence: computation, dynamics and induction," *Physica D: Nonlinear Phenomena*, vol. 75, pp. 11–54, Aug. 1994.

[29] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Academic Press, 2011.

[30] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed. Academic Press, 2008.

[31] E. Damosso, *COST Action 231: Digital mobile radio towards future generation systems, Final Report*. Office for Official Publications of the European Communities, 1999.

[32] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: John Wiley & Sons, 2001.

[33] ECOSYS Deliverable 19, "Final techno-economic results on mobile services and technologies beyond 3G," Sep. 2006. [Online]. Available: http://ecosys.optcomm.di.uoa.gr/deliverableslist.html

[34] M. Ehrgott, "A discussion of scalarization techniques for multiple objective integer programming," *Annals of Operations Research*, vol. 147, pp. 343–360, 2006.

[35] A. Engels, M. Reyer, and R. Mathar, "Profit-oriented combination of multiple objectives for planning and configuration of 4G multi-hop relay networks," in *IEEE International Symposium on Wireless Communication Systems 2010*, Sep. 2010, pp. 330–334.

[36] A. Engels, M. Neunerdt, R. Mathar, and H. M. Abdullah, "Acceptance as a success factor for planning wireless network infrastructure," in *International Symposium on Wireless Communication Systems 2011*, Aachen, Germany, Nov. 2011, pp. 889–893.

[37] A. Engels, M. Reyer, and R. Mathar, "A direction-specific land use based path loss model for suburban/rural areas," in *IEEE AP-S/URSI International Symposium 2010*, Toronto, Ontario, Canada, Jul. 2010.

[38] A. Engels, M. Reyer, A. Steiger, and R. Mathar, "Min-cut based partitioning for urban LTE cell site plannin," in *The 10th IEEE Consumer Communications & Networking Conference CCNC 2013*, Las Vegas, Nevada, USA, Jan. 2013.

[39] A. Engels, M. Reyer, X. Xu, R. Mathar, J. Zhang, and H. Zhuang, "Autonomous self-optimization of coverage and capacity in LTE cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1989–2004, Jun. 2013.

[40] V. Erceg, L. Greenstein, S. Tjandra, S. Parkoff, A. Gupta, B. Kulic, A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1205–1211, Jul. 1999.

[41] European Cooperation in Science and Technology (COST). Actions in Information and Communication Technologies. [Online]. Available: http://www.cost.eu/domains_actions/ict/Actions

[42] R. C. French, "The effect of fading and shadowing on channel reuse in mobile radio," *IEEE Transactions on Vehicular Technology*, vol. VT-28, no. 3, pp. 171–181, 1979.

[43] R. M. Freund and S. Mizuno, "Interior Point Methods: Current Status and Future Directions," *Optima*, vol. 51, pp. 1–9, 1998.

[44] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, "Broadband wireless access with WiMAX/802.16: current performance benchmarks and future potential," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 129–136, Feb. 2005.

[45] C. Glasser, S. Reith, and H. Vollmer, "The complexity of base station positioning in cellular networks," *Discrete Applied Mathematics*, vol. 148, no. 1, pp. 1–12, 2005.

[46] S. Görtzen and A. Schmeink, "Optimality of dual methods for discrete multiuser multicarrier resource allocation problems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3810–3817, Oct. 2012.

[47] O. Grondalen, P. Gronsund, T. Breivik, and P. Engelstad, "Fixed WiMAX field trial measurements and analyses," in *16th IST Mobile and Wireless Communications Summit*, Budapest, Hungary, Jul. 2007.

[48] Gurobi Optimization, Inc. Gurobi optimizer 5.0. [Online]. Available: http://www.gurobi.com/products/gurobi-optimizer

[49] A. Hoikkanen, "Economics of 3G Long-Term Evolution: the business case for the mobile operator," in *International Conference on Wireless and Optical Communications Networks*, Jul. 2007.

[50] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: A survey," *IEEE Transactions on Vehicular Communications*, vol. 58, pp. 1673–1693, May 2009.

[51] IBM ILOG. CPLEX optimization 12.2. [Online]. Available: http://www.ilog.com/products/cplex

[52] IEEE Computer Society and IEEE Microwave Theory and Techniques Society, "IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems," *IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005*, pp. 349–352, 2006.

[53] IST-4-027756 WINNER II, *D1.1.2 V1.2 WINNER II Channel Models*, Feb. 2008.

[54] S. Jaeckel, K. Borner, L. Thiele, and V. Jungnickel, "A geometric polarization rotation model for the 3-D spatial channel model," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 12, pp. 5966–5977, Dec. 2012.

[55] S. Jamaa, Z. Altman, J. Picard, and B. Fourestie, "Multi-objective strategies for automatic cell planning of UMTS networks," in *IEEE Vehicular Technology Conference Spring 2004*, vol. 4, May 2004, pp. 2420–2424.

[56] S. Kaneko, T. Matsunaka, and Y. Kishi, "A cell-planning model for HetNet with CRE and TDM-ICIC in LTE-Advanced," in *IEEE Vehicular Technology Conference Spring 2012*, May 2012.

[57] G. K. Kao and S. H. Jacobson, "Finding preferred subsets of Pareto optimal solutions," *Computational Optimization and Applications*, vol. 40, pp. 73–95, 2008.

[58] KATHREIN-Werke KG. KATHREIN scala devision: Professional antennas and filters. [Online]. Available: http://www.kathrein-scala.com

[59] V. Klee and G. J. Minty, "How good is the simplex algorithm?" in *Inequalities*, O. Shisha, Ed. New York: Academic Press, 1972, vol. III, pp. 159–175.

[60] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, 3rd ed. Germany: Springer, 2006.

[61] A. Koster and X. Muñoz, *Graphs and Algorithms in Communication Networks*. Springer Berlin Heidelberg, 2010.

[62] G. Lee, D. Park, and H. Seo, *Wireless Communications Resource Management*. John Wiley & Sons Ltd, 2009.

[63] J. Legriel, C. Le Guernic, S. Cotton, and O. Maler, "Approximating the Pareto front of multi-criteria optimization problems," in *TACAS 16th International Conference*, 2010.

[64] B. Lin, P.-H. Ho, L.-L. Xie, X. S. Shen, and J. Tapolcai, "Optimal relay station placement in broadband wireless access networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 2, pp. 259–269, 2010.

[65] C. Liu, A. Schmeink, and R. Mathar, "Equal rate resource allocation for multiuser OFDM," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, pp. 1–20, Sep. 2012.

[66] C. Liu, P. Wang, A. Schmeink, and R. Mathar, "Joint BS selection and subcarrier assignment for multicell heterogeneous OFDM unicasting," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, Canada, Sep. 2011, pp. 1361–1365.

[67] J. Liu, T. Kou, Q. Chen, and H. Sherali, "Femtocell base station deployment in commercial buildings: A global optimization approach," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 652–663, Apr. 2012.

[68] A. Lodi, "Mixed integer programming computation," in *50 Years of Integer Programming 1958-2008*, M. J¸nger, T. M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, and L. A. Wolsey, Eds. Springer Berlin Heidelberg, 2010, pp. 619–645.

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

[70] H.-C. Lu and W. Liao, "Joint base station and relay station placement for IEEE 802.16j networks," in *IEEE Global Communications Conference*, Dec. 2009.

[71] S. Martello and P. Toth, *Knapsack Problems - Algorithms and Computer Implementations*. Chichester, West Sussex, England: John Wiley & Sons Ltd, 1990.

[72] R. Mathar and T. Niessen, "Optimum positioning of base stations for cellular radio networks," *Wireless Networks*, vol. 6, no. 4, pp. 421–428, 2000.

[73] R. Mathar, M. Reyer, and M. Schmeink, "A cube oriented ray launching algorithm for 3D urban field strength prediction," in *IEEE International Conference on Communications*, Glasgow, Scotland, UK, Jun. 2007.

[74] M. Narandzic, C. Schneider, R. Thoma, T. Jamsa, P. Kyosti, and X. Zhao, "Comparison of SCM, SCME, and WINNER channel models," in *IEEE Vehicular Technology Conference Spring 2007*, Apr. 2007, pp. 413–417.

[75] M. Naseer ul Islam and A. Mitschele-Thiel, "Reinforcement learning strategies for self-organized coverage and capacity optimization," in *IEEE Wireless Communications and Networking Conference 2012*, Apr. 2012, pp. 2819–2823.

[76] M. Neunerdt, A. Engels, and R. Mathar, "Land use classification as a key component for path loss prediction in rural areas," in *IEEE International Symposium on Wireless Communication Systems 2010*, York, United Kingdom, Sep. 2010, pp. 666–670.

[77] NGMN Alliance, "Next generation mobile networks beyond HSPA and EVDO," Dec. 2006, version 3.0. [Online]. Available: www.ngmn.org

[78] ——, "Next generation mobile networks radio access performance evaluation methodology," Jan. 2008. [Online]. Available: www.ngmn.org

[79] ——, "NGMN recommendation on SON and O&M requirements," Dec. 2008, version 1.23. [Online]. Available: www.ngmn.org

[80] D. Niyato, E. Hossain, D. I. Kim, and Z. Han, "Relay-centric radio resource management and network planning in IEEE 802.16j mobile multihop relay networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6115–6125, Dec. 2009.

[81] S. Parkvall, E. Dahlman, G. Jöngren, S. Landström, and L. Lindbom, "Heterogeneous network deployments in LTE," *Ericsson Review*, vol. 90, pp. 34–38, 2011.

[82] S. Parkvall, A. Furuskar, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 84–91, 2011.

[83] S. Peters and R. Heath, "The future of WiMAX: Multihop relaying with IEEE 802.16j," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 104–111, Jan. 2009.

[84] S. Raghavan and G. Anandaligam, *Telecommunications Planning: Innovations in Pricing, Network Design and Management*, 1st ed. Springer Science, 2006.

[85] J. Ramiro and K. Hamied, *Self-Organizing Networks (SON): Self-Planning, Self-Optimization and Self-Healing for GSM, UMTS and LTE*, 1st ed. Wiley, 2012.

[86] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in *International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2010, pp. 1865–1870.

[87] R. Santana, M. Pontes, and C. Bastos-Filho, "A multiple objective particle swarm optimization approach using crowding distance and roulette wheel," in *International Conference on Intelligent Systems Design and Applications*, Dec. 2009, pp. 237–242.

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

[89] M. Schmeink and R. Mathar, "Preprocessed indirect 3D-ray launching for urban microcell field strength prediction," in *AP 2000 Millennium Conference on Antennas and Propagation*, Davos, Suisse, Apr. 2000.

[90] F. Schröder, M. Reyer, and R. Mathar, "Fast radio wave propagation prediction by heterogeneous parallel architectures with incoherent memory," in *Proceedings of Wave Propagation and Scattering in Communication, Microwave Systems and Navigation*, Chemnitz, Germany, Dec. 2010, pp. 89–93.

[91] ——, "Efficient implementation and evaluation of parallel radio wave propagation," in *5th European Conference on Antennas and Propagation*, Apr. 2011, pp. 2466–2470.

[92] ——, "Field strength prediction for environment aware MIMO channel models," in *6th European Conference on Antennas and Propagation*, Prague, Czech, Mar. 2012, pp. 1–4.

[93] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *IEEE International Symposium on Information Theory (ISIT)*, Jul. 2006.

[94] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution*. John Wiley & Sons Ltd, 2009.

[95] SOCRATES Deliverable D2.3, "Assessment criteria for self-organising networks," EU STREP SOCRATES (INFSO-ICT-216284), Jun. 2008. [Online]. Available: www.fp7-socrates.eu

[96] SOCRATES Deliverable D5.9, "Final report on self-organisation and its implications in wireless access networks," EU STREP SOCRATES (INFSO-ICT-216284), Jan. 2011. [Online]. Available: www.fp7-socrates.eu

[97] S. Srikanth, P. Murugesa Pandian, and X. Fernando, "Orthogonal frequency division multiple access in WiMAX and LTE: A comparison," *IEEE Communications Magazine*, vol. 50, no. 9, pp. 153–161, Sep. 2012.

[98] K. Sundaresan and S. Rangarajan, "Efficient resource management in OFDMA femto cells," in *Proc. of the 10th ACM international symposium on Mobile ad hoc networking and computing*, ser. MobiHoc '09. New York, NY, USA: ACM, 2009, pp. 33–42. [Online]. Available: http://doi.acm.org/10.1145/1530748.1530754

[99] D. Tikunov and T. Nishimura, "Traffic prediction for mobile network using holt-winter's exponential smoothing," in *International Conference on Software, Telecommunications and Computer Networks*, Sep. 2007.

[100] G. Tsoulos, *MIMO System Technology for Wireless Communications*. CRC Press, 2006.

[101] K. Tutschku, "Demand-based radio network planning of cellular mobile communication systems," in *IEEE INFOCOM '98*, vol. 3, Mar-Apr 1998, pp. 1054–1061.

[102] R. Wahl, G. Wölfle, P. Wertz, P. Wildbolz, and F. Landstorfer, "Dominant path prediction model for urban scenarios," in *Proceedings of the IST Mobile and Wireless Communications Summit*, Dresden, Germany, Jun. 2005.

[103] N. Weicker, G. Szabo, K. Weicker, and P. Widmayer, "Evolutionary multiobjective optimization for base station transmitter placement with frequency assignment," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 189–203, Apr. 2003.

[104] E. Weiss, S. Max, O. Klein, G. Hiertz, and B. Walke, "Relay-based vs. conventional wireless networks: Capacity and spectrum efficiency," in *IEEE Conference on Personal, Indoor and Mobile Communications 2007*, Sep. 2007.

[105] R. Whitaker, L. Raisanen, and S. Hurley, "A model for conflict resolution between coverage and cost in cellular wireless networks," in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, Jan. 2004.

[106] G. Wunder and T. Michel, "Optimal resource allocation for parallel gaussian broadcast channels: Minimum rate constraints and sum power minimization," *IEEE Transactions on Information Theory*, vol. 53, pp. 4817–4822, Dec. 2007.

[107] X. Xu, M. Reyer, F. Schröder, A. Engels, and R. Mathar, "A semi-stochastic radio propagation model for wireless MIMO channels," in *International Symposium on Wireless Communication Systems 2011*, Aachen, Germany, Nov. 2011, pp. 619–623.

[108] Y. Ye, *Interior Point Algorithms: Theory and Analysis*, ser. Wiley Series in Discrete Mathematics and Optimization. Wiley, 1997.

[109] J. Zhang, Y. Shang, R. Gao, and Y. Dong, "An improved multi-objective genetic algorithm based on Pareto front and fixed point theory," in *International Workshop on Intelligent Systems and Applications*, May 2009, pp. 1–5.

[110] J. Zhang, H. Tian, P. Tian, Y. Huang, and L. Gao, "Dynamic frequency reservation scheme for interference coordination in LTE-Advanced heterogeneous networks," in *IEEE Vehicular Technology Conference Spring 2012*, May 2012.

Index

3GPP, 6, 69

adaptive modulation and coding, 25
antenna
  downtilt, 58, 65, 93
  gain, 16
average revenue per unit, 37

bandwidth reservation concept, 51, 61, 80
bandwidth utilization ratio, 79
base station, 8
Bit Error Rate, 25
BLock Error Rate, 25

capacity, 58
  duality, 80, 95
  indicator, 79, 95
  measure, 79, 95
  metric, 81
  remaining, 80, 95
capital expenditures, 41, 63
CDMA, 5
channel
  gain, 33
  quality indicator, 25
  state information, 15
channel assignment problem, 31
climbing up principle, 89, 90, 102
co-channel deployment, 48
conflict graph, 37, 58
cooperation cluster, 86, 112
CORLA, 20
coverage, 58, 72
  hole, 99
  indicator, 78, 95
  measure, 78, 95
  metric, 78, 79, 81
Coverage and Capacity Optimization, 69

  integrated approach, 85, 92
  loop, 86, 90
  model-free approach, 103

deflection
  effects, 20
  points, 20
demand node, 8
deployment candidate, 35
dynamic clustering, 113

efficient point, 11
  unsupported, 13
eNodeB, 8
Erceg path loss model, 18
event-to-trigger ratio, 96

facets, 19
fading
  fast, 16, 24
  slow, 16
feasible point, 9, 11, 90
femtocell, 8
  open access policy, 48
frequency reuse, 47
  factor, 31, 49, 75
Full Rate Profit MAXimization, 40

handover
  zone, 64
heterogeneous networks, 3, 47, 57, 69
home eNodeB, 6

integrality constraints, 10
inter-cell interference coordination, 27
interference
  approximation model, 49, 61, 75, 113
  cross-tier, 48, 50
  impact factor, 49, 75

inter-cell, 27, 31, 48
intra-cell, 27, 48
intra-cluster, 112
mitigation, 27, 47
interior point methods, 10

Joint Coverage and Capacity MAXimization, 80

key performance index, 9
traffic light, 86
knapsack
constraint, 32, 44
multiple-choice problem, 33
problem, 31, 40, 62, 83

linear program, 9
integer, 10
mixed-integer, 10
local search heuristic, 90
LTE Advanced, 3
LTE / SAE, 3

macrocell, 8
inner, 48, 49, 53, 76
outer, 48, 49, 53, 76
maximal covering location problem, 30
minimum cut, 112
mobility robustness, 111
MSE estimation, 24, 55, 75

network management system, 6
NGMN, 6

objective function, 9, 29
OFDMA, 3
resource allocation, 33, 49
operation cycle, 69, 95
operational effort, 108
operational expenditures, 40, 63, 69
optimal
point, 9
value, 9
optimization metric, 77
optimization problem
general, 9
multi-objective, 11, 37, 59, 82
optimization variable, 9, 29
optimum, 9

Pareto
front, 12, 90
front approximation, 102
optimal point, 11, 84, 102
path loss, 16
exponent, 16
performance case, 87
performance measure, 77
physical resource block, 26, 73
potential capacity, 80
power minimization, 108
priority level, 58
problem partitioning, 112
profit maximization, 37
HetNet, 59

quality-of-service, 26

radio resource management, 6
rate-power function, 33
continuous, 33
discrete, 27
ray optical path loss model, 19
ray launching, 20
ray tracing, 20
receiver point, 16
relative importance, 13
relay station, 8
relaying
non-cooperative, 36
out-band, 37
Restricted Capacity MAXimization, 85
Restricted Coverage MAXimization, 84

scalarization approach, 12, 37, 58, 73
self-healing, 6
self-optimization, 6
autonomous, 69
coverage and capacity, 69, 85
self-organizing network, 6, 69
self-planning, 6
shadowing, 16
simplex method, 10

SINR, 25, 34
SNR, 26, 34
soft frequency reuse, 27
spatial channel model, 25
subcarrier, 26, 32
sum rate maximization
    LTE HetNet, 79
    OFDMA, 33
system link budget, 26

traffic hotspot, 93

user acceptance, 109
user assignment problem, 31
user mobility model, 92

weighted sum objective, 13
WiMAX, 3

Curriculum Vitae

Alexander Engels	Born on the 12th of October 1978 in Mönchengladbach, Germany

Education

1985 - 1989	Elementary school in Mönchengladbach, Germany
1989 - 1998	Academic high school in Mönchengladbach, Germany
3rd of June 1998	General qualification for university entrance (Abitur)
07/1998 - 07/1999	Military service (community service) in Viersen, Germany
09/1999 - 09/2002	Apprenticeship as Scientific Programer at RWTH Aachen University, Aachen, Germany
27th of September 2002	Final examination at the Chamber of Industry and Commerce (IHK), Cologne, Germany
15th of July 2004	Diploma degree in mathematics at the Aachen University of Applied Sciences, Aachen, Germany
15th of August 2008	Diploma degree in computer science at RWTH Aachen University, Aachen, Germany
08/2008 - 09/2013	Pursuing the Ph.D degree in electrical engineering

Professional Experience

08/2002 - 09/2004	Nonscientific Staff Member (Programer) at the Chair of Electrical Engineering and Computer Systems of RWTH Aachen University, Aachen, Germany
09/2004 - 08/2008	Nonscientific Staff Member (Programer) at the Institute for Theoretical Information Technology of RWTH Aachen University, Aachen, Germany
Since 08/2008	Scientific Staff Member (Research Assistant) at the Institute for Theoretical Information Technology of RWTH Aachen University, Aachen, Germany