

OPTIMIZATION AND PARALLELIZATION  
METHODS FOR THE DESIGN  
OF NEXT-GENERATION  
RADIO NETWORKS

Lucas Benedičić

Doctoral Dissertation  
Jožef Stefan International Postgraduate School  
Ljubljana, Slovenia, September 2013

*Assoc.*

Supervisor: ~~Assoc.~~ Prof. Dr. Peter Korošec, Jožef Stefan Institute, Ljubljana, Slovenia  
Co-Supervisor: Assist. Prof. Dr. Tomaž Javornik, Jožef Stefan Institute, Ljubljana, Slovenia

**Evaluation Board:**

Assist. Prof. Dr. Jurij Šilc, Chairman, Jožef Stefan Institute, Ljubljana, Slovenia

Assist. Prof. Dr. Aleš Šviglej, Member, Jožef Stefan Institute, Ljubljana, Slovenia

Prof. Dr. Marian Vajteršic, Member, Department of Computer Science, University of Salzburg,  
Austria

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA  
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Lucas Benedičič

OPTIMIZATION AND PARALLELIZATION  
METHODS FOR THE DESIGN  
OF NEXT-GENERATION  
RADIO NETWORKS

Doctoral Dissertation

OPTIMIZACIJSKE IN VZPOREDNE METODE ZA  
NAČRTOVANJE RADIJSKIH OMREŽIJ  
NASLEDNJE GENERACIJE

Doktorska disertacija

*Assoc.*

*Supervisor:* Assist. Prof. Dr. Peter Korošec

*Co-Supervisor:* Assist. Prof. Dr. Tomaž Javornik

Ljubljana, Slovenia, September 2013



# Abstract

The complexity of the design of radio networks has grown with the adoption of modern standards. Therefore, the role of the computer for the faster delivery of accurate results has become increasingly important. In this thesis, novel methods for the planning and automatic optimization of radio networks are developed and discussed.

The state-of-the-art metaheuristic algorithms, which compare a large number of different network configurations, rely on model-based simulations for the evaluation of the solution quality and the exploration of the search space. However, current radio-network solutions, based on snapshot simulations, have major weaknesses with respect to the simulation time and flexibility provided. In particular, the size of networks that can be analyzed in a feasible time is typically very limited.

The new unified framework developed in this thesis significantly outperforms the currently available solutions for snapshot-based, radio-network simulations. It brings together novel and state-of-the-art parallelization methods, in order to allow for a detailed analysis of very large networks within an acceptable amount of time for everyday planning. This is achieved by the parallel features of the framework, which are exploitable on a single multi-core CPU, as well as on a network of standard PCs with GPU devices. Clearly, the significant speed-up achieved at the simulation stage allows for an increased level of detail of the simulations, which improves the accuracy of the results.

Increasing the performance of the simulations involved during the objective-function evaluation is only the first step towards a practical running-time reduction for radio-network optimization. In addition to this, also the optimization algorithms have to be improved in terms of speed, but not at the expense of the quality of results. In this sense, a novel agent-based algorithm is presented and tailored to a classic optimization problem in radio networks. The algorithm, which is based on several techniques of cellular automata and population-based metaheuristics, shows considerable gains with respect to the size of problem instances it may handle, as well as regarding its speed performance and solution quality.

Another way to take advantage of the achieved speed-up at the simulation stage is to tackle problems of greater complexity. This thesis identifies a new optimization problem in 3G radio networks that deals with soft-handover balancing of downlink and uplink areas. Using a black-box approach, different metaheuristic algorithms are employed for solving the problem, the solutions of which show a substantial improvement of downlink and uplink balance.

From the practical point of view, the automation of time-consuming, radio-planning tasks is yet another way to exploit the benefits of faster and more precise simulations. One such task is the configuration of the control parameters of empirical radio-propagation models. Another one is identified with the environmental adaptation of the signal losses due to clutter. Both tasks are automated, using the parallel framework as a central component of the system.

The utility and performance of the developed methods is assessed against an enterprise, commercial tool for radio-network planning. The objective of such comparison is to deter-

mine the performance and usage properties of the parallel framework in everyday planning, using real-world scenarios.

## Parallel Planning in Everyday Planning

Planning is a key component of many mobile applications. In particular, mobile navigation systems often require users to plan their routes in advance. This is because route planning is a time-consuming process that requires significant computation power.

One way to improve the performance of route planning is to use parallel processing. Parallel processing involves dividing a task into smaller sub-tasks and then executing them simultaneously on multiple processors or cores. This can significantly reduce the time required to complete a task.

In this paper, we propose a parallel framework for route planning in everyday planning. Our framework uses a divide-and-conquer approach to divide the route planning task into smaller sub-tasks, which are then executed in parallel on multiple cores.

We evaluate the performance of our parallel framework using real-world scenarios. Our results show that our framework is able to significantly reduce the time required to plan routes, especially for complex routes involving many turns and stops.

Overall, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Future work will focus on further optimizing our parallel framework for route planning. We also plan to evaluate our framework on more complex scenarios, such as those involving multiple drivers or passengers.

In conclusion, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Overall, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Future work will focus on further optimizing our parallel framework for route planning. We also plan to evaluate our framework on more complex scenarios, such as those involving multiple drivers or passengers.

In conclusion, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Overall, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Future work will focus on further optimizing our parallel framework for route planning. We also plan to evaluate our framework on more complex scenarios, such as those involving multiple drivers or passengers.

In conclusion, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Overall, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Future work will focus on further optimizing our parallel framework for route planning. We also plan to evaluate our framework on more complex scenarios, such as those involving multiple drivers or passengers.

In conclusion, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

Overall, our work provides a promising approach for improving the performance of route planning in everyday planning. By using parallel processing, we are able to significantly reduce the time required to plan routes, making it easier for users to navigate through complex environments.

# Contents

ABSTRACT  
List of -  
- - -

Abbreviations	xxi
<b>1 Introduction</b>	1
1.1 Problem statement . . . . .	2
1.2 Hypotheses and approaches . . . . .	3
1.3 Scientific contributions . . . . .	4
1.4 Organization . . . . .	5
1.5 Publications . . . . .	5
<b>2 Overview of optimization methods</b>	7
2.1 Gradient-based methods . . . . .	8
2.2 Linear and non-linear programming . . . . .	9
2.3 Metaheuristics . . . . .	10
2.3.1 Differential evolution . . . . .	12
2.3.2 Differential ant-stigmergy algorithm . . . . .	13
2.3.3 Simulated annealing . . . . .	14
2.4 Black-box optimization . . . . .	14
<b>3 Principles of mobile radio networks</b>	17
3.1 Handover . . . . .	17
3.1.1 Hard handover . . . . .	18
3.1.2 Soft handover . . . . .	18
3.2 Pilot signal and power . . . . .	19
<b>4 Overview of radio-network optimization</b>	21
4.1 Optimizing base-station locations . . . . .	22
4.1.1 Related work . . . . .	23
4.2 Optimizing antenna parameters . . . . .	23
4.2.1 Related work . . . . .	24
4.3 Optimizing coverage . . . . .	25
4.3.1 Related work . . . . .	25
4.4 Discussion . . . . .	26
4.5 Summary . . . . .	26
<b>5 Principles of GPU programming</b>	27
5.1 CUDA . . . . .	27
5.2 OpenCL . . . . .	29
5.3 Summary . . . . .	29
<b>6 A parallel framework for radio-network planning and optimization</b>	31
6.1 Motivation . . . . .	32

6.2	Related work . . . . .	33
6.3	Radio-coverage prediction for mobile networks . . . . .	34
6.3.1	Background . . . . .	34
6.3.2	Radio-propagation model . . . . .	35
6.4	Design and implementation . . . . .	36
6.4.1	Geographic Resources Analysis Support System . . . . .	36
6.4.2	Multi-paradigm parallel programming . . . . .	36
6.4.3	Design of the serial version . . . . .	36
6.4.3.1	Input parameters . . . . .	37
6.4.3.2	Isotropic path-loss calculation . . . . .	37
6.4.3.3	Antenna diagram influence . . . . .	37
6.4.3.4	Transmitter path-loss prediction . . . . .	38
6.4.3.5	Coverage prediction . . . . .	39
6.4.4	Computational complexity . . . . .	39
6.4.5	Design of the parallel version . . . . .	40
6.4.5.1	Master process . . . . .	42
6.4.5.2	Worker processes on CPU . . . . .	43
6.4.5.3	Worker processes on GPU . . . . .	44
6.4.5.4	Master-worker communication . . . . .	45
6.5	Simulations . . . . .	46
6.5.1	Test networks . . . . .	47
6.5.2	Weak scalability . . . . .	47
6.5.2.1	Results . . . . .	47
6.5.3	Strong scalability . . . . .	49
6.5.3.1	Results . . . . .	49
6.5.3.2	Speedup . . . . .	50
6.5.3.3	Efficiency . . . . .	51
6.5.4	Performance analysis . . . . .	52
6.6	Summary . . . . .	53
<b>7</b>	<b>The service-coverage problem</b>	<b>55</b>
7.1	Motivation . . . . .	56
7.2	Related work . . . . .	56
7.3	Radio-network model . . . . .	57
7.3.1	Basic elements . . . . .	57
7.3.2	Coverage . . . . .	57
7.4	Problem definition . . . . .	58
7.5	Optimization approaches . . . . .	58
7.5.1	Attenuation-based approach . . . . .	58
7.5.2	Parallel-agent approach . . . . .	59
7.5.2.1	Objective-function evaluation . . . . .	59
7.5.2.2	Autonomous agents . . . . .	60
7.6	Simulations . . . . .	62
7.6.1	Test networks . . . . .	62
7.6.2	Parameter settings of the parallel-agent approach . . . . .	63
7.6.3	Experimental environment . . . . .	64
7.6.4	Results . . . . .	64
7.6.4.1	Performance analysis . . . . .	65
7.7	Summary . . . . .	68

# Abbreviations

*and Symbols*

## ~~Abbreviations~~

2G	= Second Generation of mobile networks.
3G	= Third Generation of mobile networks.
4G	= Fourth Generation of mobile networks.
ACO	= Ant-colony optimization.
BSs	= Base station.
CDMA	= Code division multiple access.
CUDA	= Compute unified device architecture.
DASA	= Differential ant-stigmergy algorithm.
DB	= Database system.
DE	= Differential evolution.
GPU	= Graphics processing unit.
GRASS	= Geographic Resources Analysis Support System.
GSM	= Global System for Mobile communications.
HPC	= High-perfomance computing.
HPCC	= High-performance computing.
HSDPA	= High Speed Downlink Packet Access.
HSPA	= High Speed Packet Access.
HSUPA	= High Speed Uplink Packet Access.
I/O	= Input/output.
KPI	= Key performance indicator.
LTE	= Long Term Evolution.
MW	= Master-worker parallel paradigm.
MWD	= Master-worker-database parallel paradigm.
NACC	= Nagasaki advanced computing center.

PRATO	= Parallel radio-prediction tool.
QoS	= Quality of service.
RSG	= Regular square grid.
SA	= Simulated annealing.
SHO	= Soft handover.
SIMT	= Single instruction multiple thread.
SM	= Streaming multiprocessor.
SP	= Streaming processor.
SPMD	= Single-program multiple-data.
TDMA	= Time division multiple access.
UE	= User equipment.
UMTS	= Universal Mobile Telecommunications System.
WCDMA	= Wide-code division multiple access.

## Symbols

$\gamma^{\text{cov}}$	= Signal-to-interference ratio (SIR) coverage threshold.
$\gamma^{\text{sho}}$	= SHO window.
$A_{\text{covered}}$	= Area under service coverage of the mobile network.
$A_{\text{total}}$	= Complete geographical area under optimization.
$as^{\max}$	= Maximum number of neighbouring cells in the active set.
$C$	= Set of antenna installations (cells) in a mobile network.
$C_m$	= Subset of cells, $C_m \subset C$ , that cover a mobile $m \in M$ .
$c_m^*$	= Best-serving cell of mobile $m \in M$ .
$cov(c, m)$	= Binary function to assert the coverage of a mobile $m \in M$ by from a cell $c \in C$ .
$cov(x, y)$	= Returns 1 if the coordinate $(x, y)$ is under mobile-network coverage.
$f_{\text{cov}}$	= Objective function for the service-coverage optimization problem.
$f_{\text{sho}}$	= Objective function of the SHO-balancing problem.
$L_{cm}^{\downarrow}$	= Downlink attenuation factor between cell $c \in C$ and mobile $m \in M$ .
$L_{mc}^{\uparrow}$	= Uplink attenuation factor between mobile $m \in M$ and cell $c \in C$ .
$M$	= Set of mobile devices or users of a mobile network.
$P_c$	= Set of candidate CPICH power settings for cell $c \in C$ .

## 1 Introduction

Many researchers believe the computer has become the third method to do research, behind theory and experimentation, for both science and engineering. Although there is no complete agreement on the position intended for scientific computing with respect to the other two methods, it is undeniable that computational methods are an essential tool in most disciplines, particularly in those related to decision making.

Nowadays, decision making is present practically everywhere. As scientists, engineers and managers have to make decisions in more complex and competitive circumstances every day, decision making involves dealing with rational and optimal approaches. According to Talbi [159], decision making consists of the following steps:

- formulating the problem,
- modeling the problem,
- optimizing the problem, and
- implementing a solution for the problem.

Formulating a decision problem means making an initial statement about it. Although this first formulation may be imprecise, the objectives of the problem are outlined, together with the internal and external factors that have some degree of influence over it. During the modeling of the problem, an abstract mathematical model is built for it. Sometimes this model is inspired by similar models in the literature, making it possible to tackle the problem with well-studied methods. After a model of the problem is available, the optimization step, i.e., generating “good” solutions for the problem, may begin. It is important to note that the resulting solutions are given for the abstract model, and not for the original problem itself. Therefore, the performance of the obtained solution is indicative when the model is an accurate one [159]. In the last step, the obtained solution is practically tested by the decision maker and implemented if it is an “acceptable” or “good” one. In case of “bad” or “unacceptable” solutions, the decision-making process is repeated, possibly improving the model and/or the optimization algorithm. The process, as described here, is depicted in Figure 1.1.

Scientific computing, by means of computer-science methodology, enables the study of problems that are too complex to be treated analytically, or those that are very expensive

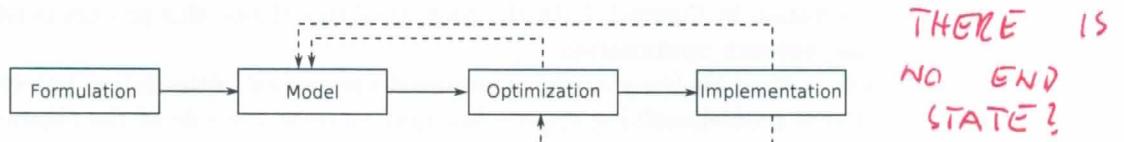


Figure 1.1: A diagram of the classic decision-making process, adapted from [159]. Multiple iterations of this process improve the optimization algorithm and/or model until an acceptable solution is found.

or dangerous to be studied by direct experimentation. Real-world problems are typically very complex systems to be directly assessed by analytical models, and require a numerical simulation for their study. Computer simulations provide a resource to mimic the behavior of complex systems, by numerically evaluating a model and gathering its data to estimate their true characteristics [96].

A model is a simplified representation of a studied problem, and one of its purposes is to predict the effects of the variations within the system. A good model is a balance between realism and simplicity. The system simulation, on the other hand, is the operation of the model. Its configuration can be changed, thus allowing multiple experimental executions, something that might not be possible with the real system it represents [103]. However, it is important to understand that the models used in scientific simulations and engineering never offer a perfect representation of the system they resemble, but only a subset of its composition and dynamics. For this reason, experimentation and expert observation will always be essential as reference points for understanding the studied phenomena. Consequently, problems categorized as of large size and of considerable complexity represent a challenge, because of the different involved disciplines for their study and the degree of difficulty of their modeling. Cellular radio networks in general, and those of the lastest generations in particular, fall under this categorization.

## 1.1 Problem statement

Radio networks represent one of the most fast-growing technology markets since the introduction of the Global System for Mobile communications (GSM) [1]. As an implementation of the second generation of mobile networks (2G), GSM appeared more than twenty years ago. Its successor, the Universal Mobile Telecommunications System (UMTS) [2] marks an evolution from 2G, representing a milestone for the third generation of mobile radio networks (3G). In recent years, the first commercial networks implementing the Long Term Evolution (LTE), also known as fourth generation (4G) LTE, have also appeared [61]. The always increasing demand for more bandwidth has been one of the main forces behind the standardization and later implementation of systems delivering higher-speed data services in order to improve the user's experience.

This evolution, first from 2G to 3G and later from 3G to 4G, has introduced not only the technology needed to increase data-transfer capacity and voice quality, but also a greater complexity in terms of radio-network planning, deployment, and configuration. This fact has attracted the attention of the research community into areas such as the design and optimization of radio-networks.

During the design phase of a radio network, a traditional or manual approach comprises the software tool, i.e., the model in Figure 1.1, that executes the analysis, and the human that makes the configuration changes, i.e., the optimization in Figure 1.1. Therefore, a radio engineer manually adjusts the network parameters and the software tool analyzes the given configuration. If the obtained results are not acceptable, the analysis process has to be repeated several times, until the goal is achieved and the changes are implemented, i.e., the implementation in Figure 1.1. In the context of this thesis, this process is referred to as manual radio-network optimization.

Advances in the past few years have improved the manual optimization process by introducing different problem-solving approaches that increase the role of the computer during the optimization of radio networks, consequently enlarging the scope of problems and instance sizes that may be subject to optimization. Still, there are some important aspects that restrict the automation level of these methods, not only in real-world environments, but also when doing research in the area of radio networks:

- A selected optimization method is typically a compromise between solution quality and computational-time complexity. The proposed state-of-the-art approach for the evaluation of radio networks is the Monte-Carlo snapshot analysis [165, 172]. However, real-world environments, where radio-network design is carried out, require the evaluation of networks with thousands of base stations in a reasonable amount of time. Moreover, for applications involving radio-network optimization, usually millions of evaluations are required to find a good solution, in which case also snapshot simulations are too time-consuming for practical use. Therefore, for such applications and environments, methods with improved time efficiency are required.
- A considerable number of publications in the field of radio-network optimization, a few references of which are given for illustration purposes [8, 29, 32, 66, 143, 146], base their simulations on platforms for which it is not possible to reproduce the experiments, either because they have used proprietary software or because the data is not available. This fact reduces the possibilities for comparing different approaches among each other, and significantly contrasts with other research areas, such as evolutionary computing or numerical optimization, where several sets of open and well-known benchmarks are available for the community to use. Consequently, an open and unified framework should allow researchers to compare different methods and results in a simplified and objective manner.
- Available commercial tools for radio-network evaluation present several drawbacks. In particular, regarding to the size of networks that can be considered and their computational-time complexity. Yet, even if precise and fast methods were commercially available, they would lack the level of flexibility required by the scientific community. Consequently, an essential attribute of the framework is to be open source, so that anyone can extend it to meet some specific requirements. In the long term, this process should also extend the set of built-in functionality.
- Particularly for path-loss predictions, estimated with empirical or deterministic mathematical models, the inaccuracy of the input data directly deteriorates the precision of the calculated results. Moreover, since the physical properties that influence the propagation of radio signals are not constant and every environment introduces its own deviations, the calculated prediction may be considered as not more than a rough approximation. Therefore, there is a need for a technique that improves the accuracy of the state-of-the-art mathematical models for radio propagations, despite the various sources of error noted before.

This thesis introduces methods and tools to mitigate the above-mentioned drawbacks from a radio-planning perspective.

## 1.2 Hypotheses and approaches

The work presented in this thesis is based on the following hypotheses and their related approaches:

- Applying parallelization techniques should reduce the computational-time requirement of the radio-propagation prediction, thus making it possible to process larger, real-world radio networks.
  - Unlike several examples in the literature of radio-network simulators, which do not include support for parallel methods in order to reduce the computational-time complexity, the unified framework includes this feature in its basic structure.

In this sense, parallel techniques are used even when targeting a single computing host.

- Using hardware, specialized for parallel execution (e.g., GPU), should improve the execution time of parallel algorithms using threads or message-passing mechanisms.
  - From its arrival a few years ago, general GPU programming is continuously gaining the attention of the research community and it is extending to several areas of science. In particular, the combination of parallel-programming techniques with GPU hardware should reduce the computational-time requirement of the radio-propagation prediction even further.
- Using the parallel framework for the objective-function evaluation of different optimization problems should improve the quality of solutions and enable tackling larger problem instances.
  - On the one hand, a larger number of evaluations in the same amount of time translates in a better search of the search space by the optimization algorithm used. On the other hand, more complex or exact models may be evaluated in the same amount of time, which implies more accurate solutions.
- Using the parallel framework to evaluate the objective function of an automatic-optimization system should provide the means for solving new, previously inaccessible optimization problems.
  - The improved performance at the evaluation level opens new possibilities for formally defining and tackling optimization problems that were previously out-of-reach of state-of-the-art methods, either because of their complexity or size.

### 1.3 Scientific contributions

The contributions of this thesis to the fields of telecommunications and computer sciences include the following:

- Design and development of an unified framework for radio-network planning and optimization, incorporating the following two elements, as well as experimentation on real-world radio networks. The experimentation includes a comparison with a commercial tool that is currently being used for real radio-network planning. Moreover, the parallel implementation of the framework exploits the computing resources of computer clusters and GPUs.
- Proposal of a new approach for parallel programming that combines a classic master-worker method with an external database. The proposed approach improves the scalability of the classic master-worker paradigm by preventing the master process to become the bottleneck of a parallel system.
- Quality improvement of renowned radio-propagation models by applying metaheuristic-optimization techniques that enhance the results of current state-of-the-art methods.
- A new algorithm, based on autonomous agents, to tackle the service-coverage problem in radio networks. The algorithm deals with problem instances that are out-of-reach of other state-of-the-art techniques, reaching good quality of solutions if compared to

*WHICH?*

*WAN, DA JE TO V DE. POKAZANOV?*  
*TO JE ZELO NEČNA TRDITEV?*

*ONREŠE:  
 ČE DA,  
 POTEN JE  
 TO OK!*

classic network-planning techniques. This approach is applicable for large real-world radio networks.

- Identification of a new optimization problem in 3G radio networks that deals with soft-handover alignment of downlink and uplink areas. By solving this problem, network malfunctioning is avoided in areas where there is soft-handover capability in the uplink, but none in the downlink. So far this problem has never been solved by means of automatic optimization.

MI MICO OOMRO, PA MI TO E E VREJ  
✓ TEKSTU RAZLOZIL: ZEO NA KAIKO.

## 1.4 Organization

So far, a context has been provided for the following content of this thesis. An introduction of the three research areas addressed by this work, i.e., optimization, radio networks, and parallel methods, is given in Chapters 2, 3 and 5, respectively. They provide an overview of the theoretical elements needed for a better understanding of the rest of the thesis. Additionally, Chapter 4 provides a short literature survey about the optimization of radio networks.

The unified framework for radio-network planning and optimization, incorporating elements from the above-mentioned fields of study, is presented in Chapter 6.

Chapters 7 and 8 demonstrate how the framework is applied to tackle two optimization problems. In Chapter 7, the problem of minimizing the total amount of pilot power subject to a full coverage constraint is addressed with a novel approach. Next, Chapter 8 formally introduces a new optimization problem that deals with the balancing of downlink and uplink soft-handover areas in UMTS networks.

The applicability and performance of the framework for radio-network planning is validated in Chapters 9 and 10. Chapter 9 further extends the framework with automated-tuning capabilities that enable its adaptation to different environments, thus improving the accuracy of the radio-propagation predictions. In Chapter 10, the framework is tested in real network-planning conditions and compared to a commercial, enterprise-level tool in terms of solution quality and speed performance.

Finally, Chapter 11 gives a short summary of the thesis, outlines its main contributions and discusses further work.

## 1.5 Publications

The work presented in this thesis is supported by a number of previous publications ...???

A comprehensive list of related publications is given in Appendix ???.

(Continued)

the first time that the author has been able to do so. The author would like to thank the editor for his kind and patient guidance and for his valuable suggestions which have greatly improved the manuscript.

This book is an attempt to introduce the reader to the field of quantum mechanics. It is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is divided into two parts. The first part is concerned with the basic concepts and principles of quantum mechanics.

The second part is concerned with the applications of quantum mechanics to various fields of science and technology.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

The book is intended to be a comprehensive introduction to the subject, covering all the basic concepts and principles of quantum mechanics.

## 2 Overview of optimization methods

Optimization may be informally defined as the procedure of finding better solutions to a given problem that usually models some physical phenomenon. In our every day life, we are constantly solving small optimization problems, like choosing the shortest route to a friend's house, or organizing the appointments in our agenda. In general, these problems are small enough for us to find a good solution without extra help, but as they become larger and more complex, the aid of computers for their resolution is unavoidable.

Complex multidimensional optimization problems are popular in engineering, economics, physics and other scientific fields. When solving an optimization problem, the objective is to find a "good" solution in a "reasonable" computational time. In this respect, the field of mathematical optimization has received a lot of attention by the scientific community during the last decades. However, both "good" and "reasonable" are problem, application and context-specific concepts, in which the biggest challenge of selecting an appropriate optimization approach usually lays.

Mathematical optimization involves the process of finding solutions from a group of possible decisions, which may be defined as:

$$\min f(\vec{x}) \quad \vec{x} \in \Omega \subseteq \mathbb{R}^n, \quad (2.1)$$

where  $\vec{x} = (x_1, \dots, x_n)$  is a vector representing the decision variables,  $f(\vec{x})$  is the objective function measuring the quality of the decisions and  $\Omega$  is the set of feasible solutions of the problem, also known as search space. Note that the objective function  $f$  makes it possible to define a total order relation between any pair of solutions in  $\Omega$ .

The search space  $\Omega$  may also be expressed as a solution to a system of equalities or inequalities, e.g.:

$$\begin{aligned} g(x_1, \dots, x_n) &\leq 0 \\ h(x_1, \dots, x_n) &= 0. \end{aligned} \quad (2.2)$$

Optimization problems involving the maximization of the objective function also fall into this category, since:

$$\max f(\vec{x}) = -\min(-f(\vec{x})). \quad (2.3)$$

A point  $\vec{x}^*$  is considered to be an unrestricted local minimum of a function  $f$  if it holds a better value than all its neighbours, i.e., there exists  $\epsilon > 0$  so that:

$$f(\vec{x}^*) \leq f(\vec{x}) \quad \forall \vec{x} \in \mathbb{R}^n \quad |\vec{x} - \vec{x}^*| < \epsilon. \quad (2.4)$$

Similarly, a point  $\vec{x}^*$  is considered to be an unrestricted global minimum of a function  $f$  if it holds a better value than all others, i.e.:

$$f(\vec{x}^*) \leq f(\vec{x}) \quad \forall \vec{x} \in \mathbb{R}^n. \quad (2.5)$$

ni cista  
kes, ni  
nuwo na  
so vse  
re sitve  
feasible.  
vra raig  
"possible"

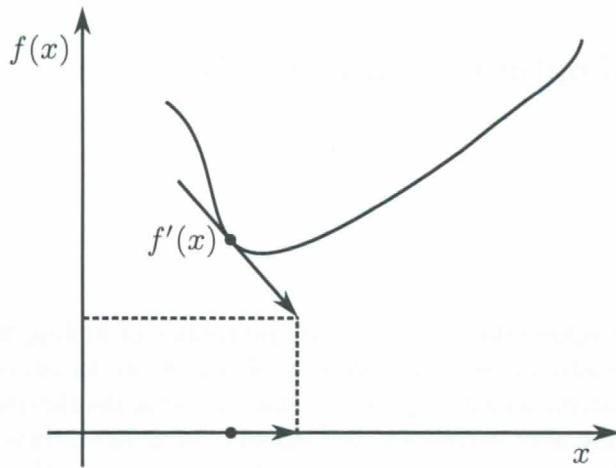


Figure 2.1: Gradient descent with a negative slope, i.e.  $x$  is increasing.

The concepts of local and global minimum are considered strict if the inequalities of Equations 2.4 and 2.5 are strict. Likewise, the definition of local and global maximum is given by the existing relation between a minimization and a maximization problem, as specified in Equation 2.3, i.e., a point  $\vec{x}^*$  is a local or global maximum of a function  $f$  if and only if  $\vec{x}^*$  is a local or global minimum of function  $-f$ , respectively.

## 2.1 Gradient-based methods

Gradient-based methods are among the oldest and most studied optimization approaches. They are based on the derivative of the optimized function, using the first and even the second derivative of a function  $f$ . The name gradient follows from the derivative of multidimensional functions,  $\nabla f(\vec{x})$ , which is simply a vector where each element is the slope of  $\vec{x}$  in that dimension, i.e.,  $\left< \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right>$  [101].

The principle behind gradient-based methods is rather simple. Starting from an arbitrary value for  $x$ , a subtraction (or addition) of a small positive value is iteratively applied to it, e.g., for gradient descent:

$$x \leftarrow x - \alpha f'(x), \quad \text{OPISI TA } f \text{ KAJ JE (IRVI OPNUO...)} \quad (2.6)$$

where  $\alpha$  is a small positive value. Consequently, a positive slope will make  $x$  decrease, whereas a negative slope will make it increase. Figure 2.1 shows an example of this behaviour. Therefore,  $x$  will gradually move down the function until it finds its minimum, where  $f'(x)$  is zero, causing it to stop.

However, gradient methods have certain drawbacks that make them unsuitable for tackling a wide range of optimization problems. Take, for example, the time they take to converge. As gradient descent approaches a function minimum, it will skip this point and land on the other side. In the next step, something similar will happen, but this time from the other side of the minimum point, thus slowly approaching to the target in a “zig-zag” way. This behavior is directly related to the slope of the function at the given point, i.e., a steepest slope translates into a larger jump, and may be alleviated by adjusting the value of  $\alpha$ . However, some functions (or regions of functions) may require smaller values, while for

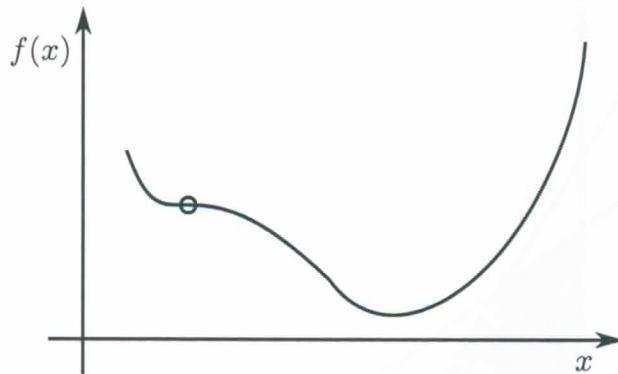


Figure 2.2: A saddle point or point of inflection, where the derivative is zero.

others a bigger value would be more appropriate. Newton's method improves this by taking the second derivative of the function into account, i.e.:  $\frac{f''(x)}{f'(x)}$

$$x \leftarrow x - \alpha \frac{f'(x)}{f''(x)}, \quad (2.7)$$

thus adjusting the value of  $\alpha$  as it converges towards a point with zero slope [101].

Another issue is how other points are handled. Beside maxima and minima points, some functions also contain saddle points (known as inflection points in one-dimensional functions). Clearly, the first derivative of a saddle point is zero, meaning gradient descent will stop looking for the minimum, even though it hasn't found it (see Figure 2.2). Newton's method, on the other hand, does not help either, even trying to divide by zero in this case. These observations clearly show how gradient methods get caught in local optima. Local optima of a function is defined as the optima (or minima in this case) of a local region. Similarly, global optima are defined as the optima of the whole domain of a function. It follows that gradient methods, as gradient descent or Newton's method, are local optimization algorithms [101].

But maybe the biggest concern with gradient-based methods is they assume the function under optimization is derivable. This assumption holds only when optimizing a well-formed mathematical function. Unfortunately, this is generally not true, since in most cases the gradient is not computable because the function is not known. The only available approach in such situations is creating inputs to the function in order to assess their quality. Meta-heuristics (see Section 2.3) are good candidates for this class of problems, for solving both moderate and large instances.

## 2.2 Linear and non-linear programming

It was in the early 40s of the twentieth century, through the work of teams formed by mathematicians, economists and physicists, that the basis were established for the resolution of problems with a set of techniques known as linear and non-linear programming. Their initial goal was to solve different kinds of logistic problems during the second world war.

In a linear programming optimization problem, both the objective function  $f$  and a given set of constraints are linear functions. The constraints impose restrictions over  $\vec{x}$ , i.e., they must meet certain requirements as, for example, fullfil a limited availability of resources. A problem may be formulated as follows:

$$\min f(\vec{x}) = c \cdot \vec{x} \quad (2.8)$$

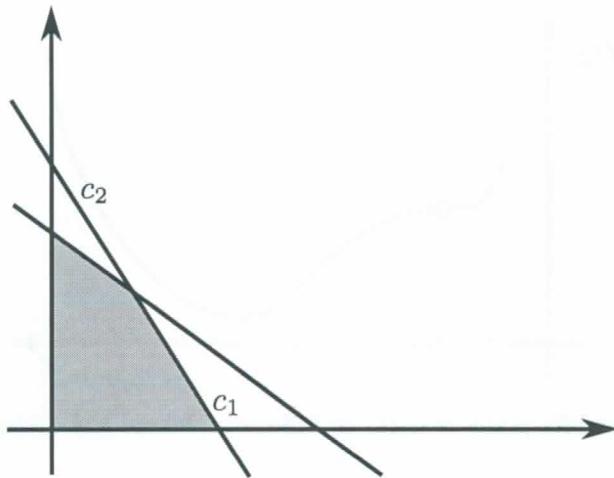


Figure 2.3: Graphical representation of a linear-programming example with two constraints,  $c_1$  and  $c_2$ . The greyed area is the polytope representing the region of feasible solutions.

subject to

$$\begin{aligned} A \cdot \vec{x} &\leq b \\ \vec{x} &\geq \vec{0} \end{aligned} \quad (2.9)$$

PÖENETI  
ENKRAT  
INAT WA  
EGORNDI  
GNACTI  
DRUGIC  
WA SPOHN

In the example above, the inequalities defined in 2.9 are the constraints to the linear program defined in 2.8.

For solving continuous, linear-optimization problems, efficient exact algorithms exist, such as the simplex method [43] or the interior-points method [88]. Indeed, linear programming is one of the most satisfactory models for solving optimization problems, since the feasible region of the problem is a convex set and the objective function is a convex function. It follows that the global optimum is a node of the polytope representing the feasible region [159]. See Figure 2.3 for a linear-programming example with several constraints.

Non-linear programming models, on the other hand, consider problems where the objective function  $f$  and/or the constraints are non-linear [15]. However, non-linear continuous problems are more difficult to solve. Despite several existing techniques to linearize such models, they often not only introduce extra variables and constraints, but also some degree of approximation [63]. Moreover, some problem properties such as high dimensionality, parameter interaction, and multi-modality make these approaches ineffective.

Generally speaking, when dealing with real-world problems, the availability of analytical optimization models, such as those required by gradient methods or (non-)linear programming, is not guaranteed. Indeed, for some applications, only simulations or physical models are the available means for objective-function evaluation [53]. Once again, metaheuristics appear as good candidates to solve different instance sizes of this class of problems.

## 2.3 Metaheuristics

Metaheuristics, a term proposed by Glover in [64], represent a group of approximation algorithms designed to combine basic, heuristic principles with advanced high-level guidance methods, targeted at improving the efficiency of a search process. These techniques are

meant to find good solutions to a given problem, for which the mathematical function is not available or its search space is big enough for an exhaustive search to be unfeasible [91].

From the theoretical point of view, metaheuristics represent a subset of stochastic optimization, since they use some degree of randomness to find optimal (or as good as possible) solutions to hard problems. They are the most general of these kinds of algorithms, and are applied to a wide range of problems [101]. Unlike the exact optimization methods introduced in the previous sections, metaheuristics do not ~~guarantee~~<sup>guarantee</sup> the optimality of the obtained solutions [159]. Moreover, they do not define how close the obtained solutions are from the optimal ones, as approximation algorithms do.

The characterization given by Blum and Roli [25] provides a clear overview of the fundamental properties associated with metaheuristics:

- metaheuristics are strategies that “guide” the search process;
- their goal is to efficiently explore the search space in order to find optimal or near-optimal solutions;
- they build upon techniques which range from simple local search procedures to complex learning processes;
- they are approximate and usually non-deterministic;
- they may incorporate mechanisms to avoid getting trapped in confined areas of the search space;
- their basic concepts permit an abstract-level description, which is not problem-specific;
- they may make use of domain-specific knowledge in the form of heuristics that are controlled by the upper level strategy;
- advanced metaheuristics use search experience (implemented as some form of memory) to guide the search process.

The strategies used by metaheuristics should provide a dynamic balance between the exploitation of the accumulated search experience (commonly called intensification) and the exploration of the search space (commonly called diversification) [25]. This balance provides the necessary means to quickly identify promising regions, and early discarding those which have already been explored or don't provide solutions of better quality. Promising regions within the search space, which are identified by the obtained “good” solutions, are thoroughly explored during the intensification phase, hoping to find better solutions. On the other hand, during the diversification phase, not-yet-visited regions are explored, making sure the search space as a whole is evenly explored, thus avoid confining the search to a reduced number of regions. In this context, the ultimate search algorithm in terms of diversification is random search. Random search generates a random solution in the search space at each iteration, without using memory [159]. In terms of intensification, iterative local search<sup>is a</sup> is a representative algorithm. The steepest local search algorithm selects, at each iteration, the best neighboring solution that improves the current one [159].

Metaheuristics are applicable where state-of-the-art exact algorithms cannot tackle the given instances within the required time, either because of their size or structure. The meaning of “required time” within this context directly depends on the target optimization problem itself. A feasible or acceptable time may vary from some seconds to several months, again, depending on the target optimization problem, e.g., real-time decisions against structural-design problems.

metaheuristics  
metaheuristic  
metaheuristics  
metaheuristics  
metaheuristics  
metaheuristics

Based on the characterization given by Talbi [160], a summary of the essential properties of optimization problems that justify the use of metaheuristics follows:

- Very large problem instances. Even though exact polynomial-time algorithms might be known for solving the target problem, they are too expensive due to the size of the instances.
- Problems with hard real-time constraints, where a “good solution” has to be found online. Metaheuristics appear as an alternative to exact algorithms in order to reduce the search time.
- A difficult problem of moderate size, which input instances have an intricate structure.
- Optimization problems with time-consuming objective function(s) and/or constraints. Indeed, various real-world optimization problems are characterized by the huge computational cost of the objective functions. Several radio-network design problems fall into this category.
- Problems that cannot be solved with exhaustive search due to the non-analytical models on which they are based. These problems are defined by a black-box evaluation of the objective function (see Section 2.4).

The influence of these conditions may increase in the presence of non-deterministic optimization models, e.g., problems with complex Monte Carlo simulations [39].

Undoubtedly, metaheuristics are rapidly gaining popularity as optimization problems are increasing in both size and complexity. Indeed, as the computing power of commodity hardware increases, the possibility of building models of greater complexity is available for developing more accurate models of real-world problems in engineering and science.

Related literature groups metaheuristic algorithms due to their behaviour, for example:

- differential evolution, from the family of evolutionary algorithms;
- differential ant-stigmergy algorithm, from the family of swarm-intelligence algorithms; and
- simulated annealing, from the group of classic metaheuristic algorithms.

Each of these algorithms minimize an objective-function value by adopting essentially distinct approaches. In the following sections, a short description about their functioning is given.

### 2.3.1 Differential evolution

Differential Evolution (DE) [153] is a simple and powerful evolutionary algorithm proposed for numerical optimization. The version of DE used for this thesis is known as DE/rand/1/bin, or "classic DE" [125]. At the beginning, the classic DE initializes a population of  $N$ ,  $D$ -dimensional vectors with parameter values that are distributed with random uniformity between pre-specified lower and upper initial parameter bounds,  $x_{i,\text{low}}$  and  $x_{i,\text{high}}$ , respectively [130], i.e.:

$$x_{i,j,g} = x_{i,\text{low}} + \text{uniform}[0, 1] \cdot (x_{i,\text{high}} - x_{i,\text{low}}), \quad i = (1, 2, \dots, D), j = (1, 2, \dots, N), g = 0. \quad (2.10)$$

ARNO IZHEUMA'

In Equation 2.10,  $g$  denotes the generation index, while  $i$  and  $j$  indicate the parameter and population indices, respectively. Consequently,  $x_{i,j,g}$  is the  $i^{\text{th}}$  parameter of the  $j^{\text{th}}$  population vector in generation  $g$ . The (pseudo)-random number generator uniform[0, 1] is represented by a function for selecting a uniformly-distributed random number from the interval [0, 1].

DE generates solutions by applying the mutation and cross-over operations. A vector from the current population is mutated by adding the scaled difference of two other vectors from the same population, combined with the mutation-scale factor. A proportion of the parameters of the mutated vector are crossed with those belonging to the  $j^{\text{th}}$  population vector, resulting in a new solution vector. The cross-over constant controls the proportion of parameters that the mutant vector contributes to the new vector.

The generated solution vector is retained if it yields a lower objective-function value than the  $j^{\text{th}}$  population vector, otherwise the  $j^{\text{th}}$  vector is kept for at least one more generation.

The four parameters to control the search process of DE are: the population size ( $N_p$ ), the number of generations for the algorithm to run ( $G_{\max}$ ), the crossover constant ( $CR$ ), and the mutation scaling factor ( $F$ ).

An extensive description of DE and its variants may be found in [126]. A wide range of optimization problems have been solved by applying DE [44], also in the area of radio-network optimization [19, 42, 107].

### 2.3.2 Differential ant-stigmergy algorithm

The Differential Ant-Stigmergy Algorithm (DASA) [93] is based on the metaheuristic Ant-Colony Optimization (ACO) [46]. ACO was pioneered by Dorigo in 1992 and is based on the foraging behaviour of social ants, which use pheromone as a chemical messenger. When looking for a food source, a swarm of ants interacts in their local environment. An ant is able to follow the route marked with pheromone laid by other ants. When an ant finds a food source, it will mark the trail to and from it with pheromone.

The pheromone concentration,  $p$ , evaporates at a constant rate,  $\gamma$  over time,  $t$ , i.e.:

$$p(t) = p_0 e^{-\gamma t}, \quad (2.11)$$

where  $p_0$  is the initial pheromone concentration at  $t = 0$ . The evaporation is important, as it provides the means for algorithm convergence and self-organization.

The pheromone concentration varies over time and the ants follow the route with higher concentration. In turn, the pheromone concentration increases with the number of ants. That is, as more and more ants follow the same route, it becomes the favored path, producing a positive feedback mechanism. Consequently, as the system evolves in time, it converges to a self-organized state, which is the essence of any ant-based algorithm.

As it was mentioned before, DASA is an ant-based algorithm that provides a framework to successfully cope with high-dimensional numerical optimization problems. It creates a fine-grained discrete form of the search space, representing it as a graph. This graph is then used as the walking paths for the ants, which iteratively improve the temporary best solution.

At the end of every iteration, and after all the ants have created solutions, they are evaluated to establish if any of them is better than the best solution found so far.

There are six parameters that control the way DASA explores the search space: the number of ants ( $m$ ), the discrete base ( $b$ ), the pheromone dispersion factor ( $q$ ), the global scale-increasing factor ( $s_+$ ), the global scale-decreasing factor ( $s_-$ ), and the maximum parameter precision ( $\epsilon$ ).

$\epsilon$  (epsilon)

To se tipico  
S (no)

NOGO ČE  
BI MILA  
DORIGO  
ZA RUVAK  
CISE IT  
EVANDI  
IGUNOKOD  
OPERAMITL  
KAKJEW  
PREH  
FONT  
(LAISE  
LOCI IN  
HERE)

**Algorithm 2.1** Pseudo-code of the simulated-annealing metaheuristic. Adapted from [23].

```

Require:  $x \leftarrow$  initial solution
Require:  $T_1 \leftarrow$  temperature parameter
for  $k=1,2,\dots$  do
     $y \leftarrow$  random solution from  $\Omega$ 
    if  $f(y) \leq f(x)$  then
         $x \leftarrow y$ 
    else if  $e^{\frac{f(x)-f(y)}{T_k}} \leq \text{uniform}[0, 1]$  then
         $x \leftarrow y$ 
    end if
     $T_{k+1} \leftarrow$  update temperature  $T_k$ 
end for

```

A more in-depth explanation about these parameters and the DASA algorithm itself can be found in [93].

### 2.3.3 Simulated annealing

This metaheuristic was introduced in the area of combinatorial optimization by Kirkpatrick et al. [89]. The algorithm resembles the physical annealing process, where particles of a solid material arrange themselves into a thermal equilibrium. Simulated Annealing (SA) is based on the principle of local-search heuristics, and uses a pre-defined neighborhood structure on the search space  $\Omega$ . The temperature parameter controls the search behavior. Its name follows the analogy to the physical annealing process. In each iteration, a solution  $y$ , which is a neighbour of the current solution  $x$ , is computed. If  $y$  has a better objective-function value than  $x$ , it becomes the new solution, i.e., the current solution  $x$  is replaced by  $y$ . If, on the other hand,  $y$  has a worse objective-function value than  $x$ , it is only accepted with a certain probability depending on two factors:

- the difference of the objective-function values of  $x$  and  $y$ , and
- the temperature parameter.

Intensification in SA is provided by the local-search nature of the algorithm, while diversification is produced with non-zero temperatures, by “uphill” moves. The pseudo-code of the SA metaheuristic can be represented as in Algorithm 2.1. In this pseudo-code,  $T_1, T_2, \dots$  represent an usually decreasing sequence of values for the temperature parameter, and  $\text{uniform}[0, 1]$  is a function for selecting a uniformly-distributed random number from the given interval, whereas the meaning of the rest of the symbols is as defined above.

SA has two parameters to control the search process: the initial temperature ( $t_{initial}$ ) and the total number of iterations or evaluations ( $it$ ). SA also allows to define the way the temperature is lowered during the annealing process.

SA has proved to be a solid optimization algorithm, capable of giving high-quality solutions to a wide scope of optimization problems [155]. Additionally, several works employ the SA metaheuristic for solving optimization problems in the area of radio network [19, 47, 54, 105, 141, 182].

*NARSKA ... "A MORE DETAILED ... CAN BE FOUND IN {}"*

### 2.4 Black-box optimization

The complexity of a problem is equivalent to the complexity of the best known algorithm solving that problem [160]. If there exists a polynomial-time algorithm to solve a problem,

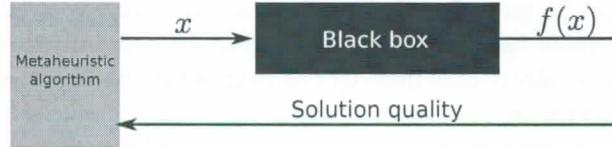


Figure 2.4: A metaheuristic algorithm process using a black box for the objective-function evaluation,  $f(x)$ , of a solution,  $x$ .

the problem is said to be easy or tractable. Similarly, if a problem is difficult or intractable, there is no known polynomial-time algorithm to solve it.

Many optimization problems cannot be formulated with a clear analytical mathematical notation. In such cases, the objective function may become a black box [87]. This is one of the main advantages when using metaheuristics, i.e., there is no need of a complete knowledge of the targeted model. Indeed, in a black-box optimization, no analytical formulation of the objective exists [160], as Figure 2.4 shows.

More specifically, a function  $f(\vec{x})$ ,  $\vec{x} \in \mathbb{R}^n$ , is a black-box function if and only if [160]:

- the domain  $\vec{x}$  is known,
- it is possible to get the value of  $f$  for each  $\vec{x}$  based on simulation, and
- there is no other information available for function  $f$ .

Typically, the experiments associated with these kind of problems are very expensive in terms of time and cost, since a simulation must be forced to evaluate the solution. Generally speaking, the most time-consuming part of a metaheuristic optimization process is the evaluation of the objective function [159]. This is especially true when dealing with real-world problems of areas such as structural design [14], molecular docking [161] and, the field on which this thesis focuses, radio-network design [18, 19, 20, 21, 22]. A possible substitution for lengthy evaluations is to reduce their complexity by approximating the objective function, thus replacing it with an approximation during the optimization process. This approach is known as meta-modeling [159]. However, when dealing with approximations, some degree of solution quality is inevitably sacrificed. As it will be shown in the following chapters, there is a very fine balance between the number of evaluations and the quality of the achieved solutions. Consequently, reducing the time spent in objective-function evaluation should favorably influence the solution quality achieved by a preferred metaheuristic algorithm. A major portion of this thesis is dedicated to improve this specific aspect in the area of radio-network optimization, starting with a high-performance, unified framework for radio-network planning, which is presented in Chapter 6.

In practice, however, the black-box evaluation of the objective function presents a problem. In the context of research about radio-network optimization there is an inherent difficulty of providing the black-box used to evaluate a given approach. A quick review of the state-of-the-art in radio-network optimization indicates that this fact has become increasingly regular in several published works [8, 29, 32, 66, 143, 146]. Clearly, this fact creates a barrier to one of the most important phases of scientific methodology: experimental reproducibility [57].

There are several reasons behind this situation. For example, it is a known fact that proprietary software, providing good computational models for radio-network simulation, is a very expensive tool for science. Even neglecting the economical aspect, but considering the great variety of software packages and license combinations, it is practically impossible

for a research laboratory to have the whole palette of commercially-available solutions at its disposal. Moreover, genuine users of these applications are generally not allowed to mention the formats, protocols, or algorithms used by the proprietary software, since their disclosure is often explicitly forbidden by restrictive licenses.

Some non-commercial and open-source packages for radio-network simulation [106, 85, 81, 123, 124, 179] present two main drawbacks: poor documentation and/or low scalability. The scalability issues shown by some projects prevent these packages to be used in larger, real-world environments, where big problem instances are the rule. Despite this, the merit and acknowledgement go to the authors of these frameworks, not only for providing them to the scientific community, but fundamentally because for providing an environment in which different kinds of simulations are completely reproducible. Additionally, the lack of documentation represents a big hurdle for extending the base code, which becomes a difficult task without the help of the original authors of the package. In such cases, author's knowledge is required for effectively expanding the functionality of an open-source tool. This is especially true when dealing with complex simulation frameworks, as the ones used for radio networks.

*VESTUJ NEN*

### 3 Principles of mobile radio networks

A cellular mobile radio network is a collection of individual cells that are served by several transmitters, called Base Stations (BSs). Each BS gives radio coverage to a small geographical area. The integration of the coverage of various BSs provides radio coverage over a much larger geographical area, thus defining a cellular radio network. The two basic functions of a radio-network system are:

- locate and track both active and idle mobile devices, called User Equipments (UEs); and
- attempt to connect each UE to the best available BS.

The first task involves a location-update procedure, which allows an UE to inform the network about its movement from one location area to the next. This process is called mobility management. The second task requires the constant evaluation of the radio-link quality with the serving BS, and the radio-link qualities of alternate BSs. This process is called radio-resource management, and is performed by the network using knowledge about the link-quality evaluations of the reference channels, e.g., the pilot channels.

The radio communications between an UE and a grid of BSs use low power. However, the movement of the UE causes highly irregular radio-link conditions, thus consistent monitoring and precise control are required to keep the radio-link quality at an acceptable level. At the core of the evaluation of radio-link quality is a statistical-measurement process based on a previous knowledge of the expected characteristics of the pilot channel. On the one hand, the link quality, and the size and distribution of the cells of a modern radio cellular system are limited by the speed of the link-quality measurement and network control. On the other hand, the spectral efficiency of a radio network is bounded by the cell sizes, the ability of radio links to withstand interference, and the ability of the system to react to variations in traffic [154].

Cellular radio systems partition the available spectrum among the BSs, and a given frequency is reused at the closest possible distance that the radio link will allow. Consequently, smaller cells have a shorter distance between reused frequencies, and this results in an increased spectral efficiency and traffic-carrying capacity. The radio links of a high-capacity radio network interfere with each other due to frequency reuse. For this reason, it is always desirable to use the lowest possible transmit power while maintaining each radio link above a given Quality-of-Service (QoS) threshold. Therefore, radio links should not significantly exceed their target QoS, since doing so will cause unnecessary interference to other radio links [154]. This particular situation is further discussed in Chapter 7, where an optimization approach, that minimizes the total transmit power used in a radio network, is presented.

#### 3.1 Handover

In radio networks, handover (or handoff) is one of the main features that allows the mobility of UEs [78]. The concept behind the handover operation is simple: when a UE moves from

the coverage area of a cell to the coverage area of a neighboring cell, the system creates a new connection with the latter cell and disconnects the user from the former one, so that an acceptable link quality can be maintained. Otherwise, the increase in transmit power, that is required to compensate for path loss, results in excessive interference. The handover procedure consists of two processes [154]:

- the link-quality evaluation before handover initiation, and
- the allocation of radio and network resources.

Generally speaking, radio networks with smaller cell sizes require faster and more reliable handover algorithms. Indeed, it has been shown that the number of cell-boundary crossings is inversely proportional to the cell size [94]. Since there is certain probability of dropping a connection whenever a handover is attempted, it is clear that the role of handover configuration becomes more important as the cell sizes decrease. Therefore, if the radio network does not detect poor signal quality fast enough, or makes too many handovers, the capacity is diminished due to increased interference and/or excessive control traffic [154].

### 3.1.1 Hard handover

During a hard handover, an UE can connect to only one BS at a time. A unique decision initiates and executes a handover without making a number of simultaneous connections among candidate BSs. Based on link measurements, the target BS is selected prior to executing the handover, and the active connection is instantly transferred to it. Moreover, the connection even experiences a brief interruption during the actual transfer, because the UE can only connect to one BS at a time. In contrast to soft handover (see next section), hard handovers do not take advantage of the diversity gain, where the signals from two or more BSs arrive at comparable strengths. Hard handover is a simple and inexpensive way to support UE mobility. It is used in Time-Division, Multiple-Access (TDMA) cellular systems such as GSM [154].

### 3.1.2 Soft handover

~~Hard Handover~~ Soft Handover (SHO) enhances handover functionality by allowing an UE to potentially operate on multiple radio links at a time. During SHO, the target BS is selected as the best candidate from among the available BSs. The UE performs the necessary link-quality measurements by monitoring the signals from the surrounding BSs. Simultaneously keeping multiple connections means that SHO enhances the system performance through diversity reception.

Despite the advantages it provides, SHO is complex and expensive to implement. Additionally, interference actually increases with SHO, since several BSs can connect to the same UE. This increase in forward interference can become a problem. If the handover region is large, such that there are many UEs in SHO mode, the increased interference due to several BSs connected to the same UE can become a problem [154].

The SHO procedure is important in systems using the channel-access method known as Code Division Multiple Access (CDMA), and especially in Wide CDMA (WCDMA), which is employed by the UMTS. CDMA systems are interference-limited meaning their capacities are closely related to the level of interference they can tolerate. Specifically, a CDMA system cell is affected by the interference within its own cell, and also interference from its neighboring cells. In order to mitigate the level of interference, and thus increase the capacity and quality, CDMA systems use power control. The main idea behind power control is to prevent the UEs and the BSs from transmitting more power than is strictly

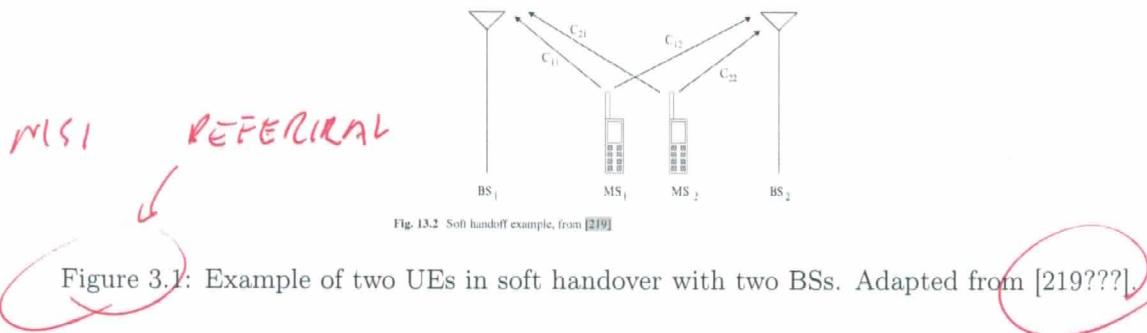


Figure 3.1: Example of two UEs in soft handover with two BSs. Adapted from [219??].

necessary to meet the target QoS level. For the power control to work properly, the system must ensure that each UE is connected to the BS having the least attenuation at all times. If this is not the case, a positive-feedback problem appears, and it can potentially destabilize the entire system [175]. The SHO procedure helps to prevent such situations by ensuring that each UE is served by the best BS most of the time, i.e., by allowing connections to multiple BSs.

The SHO condition depends on the relative, received-signal quality from different cells and the SHO window, which triggers the addition of a cell to the active set of the UE. Depending on radio propagation characteristics, the radio transmission can gain more than 3 dB out of a SHO situation [78]. From this point of view, SHO is a method to reduce interference and improve radio quality, particularly at the cell border where the radio coverage is of inferior quality. In UMTS Release 99 [163], SHO is specified to work from the BS towards the UE (downlink), and from the UE towards the BS (uplink).

With the introduction of the High Speed Packet Access (HSPA) as an improvement of the performance existing in WCDMA protocols, the role SHO plays in mobile network configuration and functioning slightly changed. The key difference is that the High Speed Downlink Packet Access (HSDPA) does not support SHO, whereas the High Speed Uplink Packet Access (HSUPA) does. This particular distinction is further discussed in Chapter 8, because it has some important implications in the balanced distribution of SHO areas, and thus in the quality and capacity of HSPA services [80].

### 3.2 Pilot signal and power

<Not yet finished ...???

*NPNSKP*



Antenna