



Forschungsberichte
aus
dem Institut
für Höchstfrequenztechnik
und Elektronik
der
Universität Karlsruhe (TH)

Herausgeber:
Prof. Dr.-Ing. W. Wiesbeck

Xuemin Huang

Automatic Cell Planning for Mobile Network Design: Optimization Models and Algorithms

Band 30

Copyright: Institut für Höchstfrequenztechnik und Elektronik
Universität Karlsruhe (TH)

alle Rechte vorbehalten

Druck: Druckerei Gunter Dünnbier, 02779 Großschönau,
Tel. 035845-36757

ISSN: 0942-2935

Foreword

Will future mobile communication systems need network planning or not? There are presently numerous activities going on for the development of products for ad hoc networks. Ad hoc networks consist of transceivers, mostly mobile, which connect themselves with their neighbors and thereby create a communications network without any planning process. A deeper look into this subject shows, however, that ad hoc networks are a solution for small limited areas. They can not handle high-density traffics and they can not work over long distances. The conclusion is that the planning of mobile communication networks, especially base station placement will be one of the primary tasks also in the future.

The Ph.D. thesis of Dr. Huang delivers significant new ideas, concepts and results for mobile network design. The design process of base station placement and dimensioning is modeled and optimized by a multi-objective genetic algorithm. Genetic algorithms allow to include easily a number of incommensurable criteria, for example cost and coverage. The planning process is, as it follows biological evolution, easily understandable. I wish that the results of this thesis are well accepted by the community and the ideas spread out throughout the world.

Prof. Dr.-Ing. Werner Wiesbeck
- Institutsleiter -

Forschungsberichte aus dem Institut für Höchstfrequenztechnik und Elektronik der Universität Karlsruhe (TH)

Herausgeber: Prof. Dr.-Ing. Werner Wiesbeck

- Band 1 Daniel Kähny
Modellierung und meßtechnische Verifikation polarimetrischer, mono- und bistatischer Radarsignaturen und deren Klassifizierung
- Band 2 Eberhardt Heidrich
Theoretische und experimentelle Charakterisierung der polarimetrischen Strahlungs- und Streueigenschaften von Antennen
- Band 3 Thomas Kürner
Charakterisierung digitaler Funkssysteme mit einem breitbandigen Wellenausbreitungsmodell
- Band 4 Jürgen Kehrbeck
Mikrowellen-Doppler-Sensor zur Geschwindigkeits- und Wegmessung- System-Modellierung und Verifikation
- Band 5 Christian Bornkessel
Analyse und Optimierung der elektrodynamischen Eigenschaften von EMV-Absorberkammern durch numerische Feldberechnung
- Band 6 Rainer Speck
Hochempfindliche Impedanzmessungen an Supraleiter / Festelektrolyt-Kontakten
- Band 7 Edward Pillai
Derivation of Equivalent Circuits for Multilayer PCB and Chip Package Discontinuities Using Full Wave Models
- Band 8 Dieter J. Cichon
Strahlenoptische Modellierung der Wellenausbreitung in urbanen Mikro- und Pikofunkzellen
- Band 9 Gerd Gottwald
Numerische Analyse konformer Streifenleitungsantennen in mehrlagigen Zylindern mittels der Spektralbereichsmethode
- Band 10 Norbert Geng
Modellierung der Ausbreitung elektromagnetischer Wellen in Funksystemen durch Lösung der parabolischen Approximation der Helmholtz-Gleichung

**Forschungsberichte aus dem
Institut für Höchstfrequenztechnik und Elektronik
der Universität Karlsruhe (TH)**

- Band 11 Torsten C. Becker
Verfahren und Kriterien zur Planung von Gleichwellennetzen für den Digitalen Hörrundfunk DAB (Digital Audio Broadcasting)
- Band 12 Friedhelm Rostan
Dual polarisierte Microstrip-Patch-Arrays für zukünftige satellitengestützte SAR-Systeme
- Band 13 Marcus Demmler
Vektorkorrigiertes Großsignal-Meßsystem zur nichtlinearen Charakterisierung von Mikrowellentransistoren
- Band 14 Andreas Froese
Elektrochemisches Phasengrenzverhalten von Supraleitern
- Band 15 Jürgen v. Hagen
Wide Band Electromagnetic Aperture Coupling to a Cavity: An Integral Representation Based Model
- Band 16 Ralf Pötzschke
Nanostrukturierung von Festkörperflächen durch elektrochemische Metallphasenbildung
- Band 17 Jean Parlebas
Numerische Berechnung mehrlagiger dualer planarer Antennen mit koplanarer Speisung
- Band 18 Frank Demmerle
Bikonische Antenne mit mehrmodiger Anregung für den räumlichen Mehrfachzugriff (SDMA)
- Band 19 Eckard Steiger
Modellierung der Ausbreitung in extrakorporalen Therapien eingesetzter Ultraschallimpulse hoher Intensität
- Band 20 Frederik Küchen
Auf Wellenausbreitungsmodellen basierende Planung terrestrischer COFDM-Gleichwellennetze für den mobilen Empfang
- Band 21 Klaus Schmitt
Dreidimensionale, interferometrische Radarverfahren im Nahbereich und ihre meßtechnische Verifikation

**Forschungsberichte aus dem
Institut für Höchstfrequenztechnik und Elektronik
der Universität Karlsruhe (TH)**

- Band 22 Frederik Küchen, Torsten C. Becker, Werner Wiesbeck
**Grundlagen und Anwendungen von Planungswerkzeugen für
den digitalen terrestrischen Rundfunk**
- Band 23 Thomas Zwick
**Die Modellierung von richtungsaufgelösten Mehrwege-
gebäudefunkkanälen durch markierte Poisson-Prozesse**
- Band 24 Dirk Didascalou
**Ray-Optical Wave Propagation Modelling in Arbitrarily Shaped
Tunnels**
- Band 25 Hans Rudolf
Increase of Information by Polarimetric Radar Systems
- Band 26 Martin Döttling
**Strahlenoptisches Wellenausbreitungsmodell und System-
studien für den Satellitenmobilfunk**
- Band 27 Jens Haala
**Analyse von Mikrowellenheizprozessen mittels selbstkon-
sistenter finiter Integrationsverfahren**
- Band 28 Eberhard Gschwendtner
**Breitbandige Multifunktionsantennen für den konformen
Einbau in Kraftfahrzeuge**
- Band 29 Dietmar Löffler
**Breitbandige, zylinderkonforme Streifenleitungsantennen für
den Einsatz in Kommunikation und Sensorik**
- Band 30 Xuemin Huang
**Automatic Cell Planning for Mobile Network Design:
Optimization Models and Algorithms**

To my wife, Danya

Automatic Cell Planning for Mobile Network Design: Optimization Models and Algorithms

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

der Fakultät für
Elektrotechnik und Informationstechnik
der Universität Fridericiana Karlsruhe

genehmigte

DISSERTATION

von

M. S. Xuemin Huang
aus Jiangxi, China

Tag der mündlichen Prüfung:

10. Mai 2001

Hauptreferent:

Prof. Dr.-Ing. Werner Wiesbeck

Korreferent:

Prof. Dr. rer. nat. Rudolf Mathar

Acknowledgements

The research work presented in this dissertation was done during my staying at the Institut für Höchstfrequenztechnik und Elektronik (IHE), Department of Electrical Engineering and Information Technology, University of Karlsruhe, Germany and sponsored by LS telcom AG, Lichtenau, Germany.

First I am deeply grateful to Professor Dr.-Ing. Werner Wiesbeck, the director of IHE, for giving me the opportunity to achieve a Ph.D. degree (Dr.-Ing.) and for his fruitful advice and guidance throughout this work. My appreciation is also extended to Professor Dr. rer. nat. Rudolf Mathar of Aachen University of Technology, Germany for undertaking the co-examiner of my thesis.

I am also very grateful to Dr.-Ing. Georg Schöne and Dr.-Ing. Manfred Lebherz, the founders and managing directors of LS telcom AG, for the great support in enabling me to carry out a Ph.D. work in Germany. Their encouragement and stimulation helped drive the project to a successful conclusion.

My thanks go to all members (past and present) of IHE and LS telcom AG for providing me a pleasant and stimulating environment to work. Among the numerous colleagues, Dr. rer. nat. Ulrich Behr and Dipl.-Inform. Kai Rohrbacher deserve a special mention for many invaluable suggestions and helpful discussions. They are always available to answer my questions and listen to my ideas.

I would also like to thank my colleagues Dr.-Ing. Jürgen Kehrbeck, Dr.-Ing. Jürgen von Hagen and Dipl.-Ing. Ralph Schertlen for reviewing this thesis. Their constructive criticisms and corrections are sincerely appreciated. I am particularly indebted to Stephen Hurley of Cardiff University, UK for the careful English proofreading.

Again, I wish to thank the people who make my stay in Germany productive and pleasurable.

Finally, I thank my parents for their love and constant support from my first day of life. My most heartfelt gratitude goes to my wife, Danya, for her dedication and patience. Without her support I could never have come this far and completed this work.

Contents

List of figures	v
List of tables	vii
1 Introduction	1
1.1 Quality, capacity, and economic issues of mobile network design.....	1
1.2 Research scope and objectives	2
1.3 Thesis outline	4
2 Cell planning fundamentals	7
2.1 What is cell planning?	7
2.1.1 Elements of cellular system design.....	7
2.1.2 Objectives of cell planning	14
2.1.3 Design process	15
2.2 Radio wave propagation aspects of cell planning	16
2.2.1 Macrocell propagation models.....	17
2.2.2 Microcell propagation models	19
2.2.3 Indoor propagation models	21
2.3 Coverage prediction	22
2.3.1 Coverage probability.....	22
2.3.2 Assignment probability.....	25
2.4 Traffic capacity.....	28
2.4.1 Traffic estimation.....	28
2.4.2 The number of channels required	29
2.4.3 Demand vector	30
2.5 Interference analysis.....	32
2.5.1 Mutual interference.....	32
2.5.2 Compatibility matrix.....	34
2.6 Channel assignment.....	35
2.6.1 Channel assignment strategies	36
2.6.2 A heuristic algorithm for channel assignment	37
2.7 Summary	40
3 Cell planning as an optimization problem	41

3.1	Problem description.....	41
3.2	Overview of cell planning problems	43
3.2.1	Cell selection.....	43
3.2.2	Cell dimensioning.....	44
3.2.3	Automatic base station placement and dimensioning	45
3.2.4	Growth planning	45
3.3	Local conditions	46
3.3.1	Traffic load.....	47
3.3.2	Interference probability.....	48
3.4	Global conditions	49
3.4.1	Coverage rate	49
3.4.2	Spectral cost.....	52
3.4.3	Financial cost	54
3.5	Multi-objective optimization.....	56
3.5.1	Basic concepts and definitions.....	56
3.5.2	Methods for solving multi-objective problems.....	60
3.6	Summary	63
4	Genetic algorithms	65
4.1	A basic genetic algorithm.....	65
4.2	Design aspects of genetic algorithms	68
4.2.1	Representation.....	68
4.2.2	Evaluation and fitness assignment	69
4.2.3	Selection methods	71
4.2.4	Genetic operators	72
4.2.5	Replacement scheme.....	74
4.3	Real-valued genetic algorithms	74
4.3.1	Genetic operators for real-valued chromosomes	75
4.4	Multi-objective genetic algorithms	78
4.4.1	Multi-objective search techniques	78
4.4.2	Elitist non-dominated sorting genetic algorithm.....	82
4.4.3	Integrated search and decision making	86
4.5	Summary	87
5	Disk graph models and the application of genetic algorithms	89
5.1	An introduction to disk graph models	89
5.1.1	Disk graph model and cellular system	89
5.1.2	Formulation.....	91
5.2	Unit disk graph model for cell selection	93
5.2.1	Test problems.....	93
5.2.2	Simulation results	94
5.3	Cell dimensioning model	102

5.3.1	Test problems.....	102
5.3.2	Binary GA vs. real-valued GA.....	102
5.4	Base station placement model.....	104
5.4.1	Hierarchical approach.....	104
5.4.2	Simulation results.....	105
5.5	Summary.....	108
6	Cell planning experiment results and analyses	109
6.1	Simulation environment.....	109
6.1.1	Geographical data.....	109
6.1.2	The wave propagation model.....	110
6.1.3	The database.....	112
6.2	Simulation results and analyses.....	113
6.2.1	Cell selection.....	113
6.2.2	Cell dimensioning.....	117
6.2.3	Automatic base station placement and dimensioning.....	118
6.2.4	Growth planning.....	124
6.3	Summary.....	128
7	Conclusions	129
7.1	Summary of contributions.....	129
7.2	Future work.....	130
A	Abbreviations and acronyms	133
	Bibliography	135

List of figures

2.1	Frequency reuse pattern.....	9
2.2	Erlang B curves	12
2.3	Coverage-oriented cell splitting and capacity-oriented cell splitting	13
2.4	Cellular system design workflow	15
2.5	An example of the coverage probability calculation of a single cell	24
2.6	An example of the assignment probability calculation	27
2.7	Mutual interference between two cells.....	32
3.1	Illustrative example of Pareto optimality in a 2D objective space.....	57
3.2	Possible relations between solutions in multi-objective space.....	58
3.3	Graphical interpretation of the weighting method for solving multi-objective problems.....	61
3.4	The constraint method for solving multi-objective problems	62
4.1	Main procedures of a genetic algorithm.....	66
4.2	Relations between individual space, decision space, and objective space.....	69
4.3	One-point crossover for bit-string chromosomes	73
4.4	Crowding distance of an individual in the non-dominated front.....	84
5.1	An intersection model of equal-sized circles and the corresponding unit disk graph	90
5.2	Hexagonal distribution of equal-sized cells	93
5.3	Simulation results of test problem 1 using elitist NSGA	95
5.4	Illustration of non-dominated solutions obtained in the simulation of test problem 1	96
5.5	Multi-objective GA vs. single objective GA.....	97
5.6	Binary tournament selection vs. roulette wheel selection	98

5.7	Simulation results of test problem 2 with different mutation rates	99
5.8	Comparisons between one-point crossover, two-point crossover, and fusion crossover	99
5.9	Simulation results of non-elitism and the controlled elitism with different reduction rates.....	100
5.10	Simulation results of test problem 3 considering all the three objectives	101
5.11	Structure of the hierarchical approach.....	105
5.12	Illustration of the solutions obtained by the hierarchical approach.....	106
5.13	Simulation results using the hierarchical approach, compared with that of a non-hierarchical simulation.....	107
6.1	Geographical data used in cell planning experiments	110
6.2	Tables of the database used in cell planning experiments.....	112
6.3	Pareto optimal solutions obtained in the cell selection experiment	115
6.4	Cell selection experiment	116
6.5	Workflow of the hierarchical approach for solving ABSPAD.....	119
6.6	Breadth-first search method used in the hierarchical approach in order to avoid local optima.....	121
6.7	Objective scores of the solutions obtained by the hierarchical approach.....	122
6.8	Coverage areas of the cells obtained by ABSPAD	124
6.9	Nominal cell splitting	125
6.10	Experimental results of growth planning in an existing network.....	127

List of tables

2.1	Different cell characteristics.....	8
2.2	The numbers of CCHs and FCHs corresponding to the number of TCHs.....	31
2.3	The values of the channel separation.....	33
3.1	Classification of cell planning problems.....	46
5.1	GA parameters used in the cell selection simulation of test problem 1	94
5.2	Simulation results of test problem 4 using binary GA and real-valued GA.....	103
5.3	Simulation results of test problem 5 using binary GA and real-valued GA.....	103
5.4	GA parameters used in BS placement simulations.....	106
6.1	Morpho correction factors used in Okumura-Hata model.....	111
6.2	Technical parameters used in cell planning experiments.....	113
6.3	Configurations of potential cells in the cell selection experiment.....	114
6.4	Performance of the selected cells.....	115
6.5	Performance of the selected cells after cell dimensioning.....	118
6.6	Performance of the cells in the final solution of the ABSPAD experiment.....	123

Chapter 1

Introduction

The future is mobile. On the one hand, more and more people will spend much of their time on the road, whether working or during vacation. People want access to the telecommunication services in any place at any time. On the other hand, more and more communication devices are portable, such as pagers, mobile phones, personal digital assistants (PDAs), laptops, and palmtops. The development of wireless communications takes these facts into account. New technologies and services are continuing to come forth. The number of subscribers to mobile services is tremendously growing worldwide. Therefore, the proper design of mobile networks, in order to meet the increases in demand, is of utmost importance.

This thesis is devoted to the network planning aspects of mobile communication systems. The following section highlights the background and motivation for this research. The research scope and objectives will be introduced in Section 1.2. This chapter concludes with an outline of the thesis.

1.1 QUALITY, CAPACITY, AND ECONOMIC ISSUES OF MOBILE NETWORK DESIGN

The increasing demand for mobile communications leads mobile service providers to look for ways to improve the quality of service and to support increasing numbers of users in their systems. Since the amount of frequency spectrum available for mobile communications is very limited, efficient use of the frequency resource is needed. Currently, cellular systems such as the global system for mobile communications (GSM) are widely deployed because the cellular concept uses frequencies in a more efficient way compared to the conventional mobile radio systems [MacDonald (1979)]. It allows frequency reuse such that the same frequency channel (FCH) can be simultaneously used in different cells.

Theoretically a hexagonal geometry is assumed for cellular systems [Lee(1989)], but this assumption is generally not fulfilled in real mobile networks. Since terrain and other factors may result in weak radio signals in certain areas, the position of base stations (BSs) must be changed in order to provide satisfactory coverage. Thus the design of cellular systems needs a way to optimize the placement of BSs.

Furthermore, the rapid development of cellular mobile networks increases the need for economic use of the available frequency resource. Hence, worldwide research activities have been done on the channel assignment problem (CAP), also called the frequency assignment problem (FAP) in some literature [Koster(1999)]. Once the cell sites and dimensions are determined, there is a lower bound i.e. the minimum frequency spectrum required for assigning each cell a sufficient number of channels [Gamst(1986)]. Such a lower bound is defined as spectral cost in this thesis. In order to minimize the spectral cost, cell sites and dimensions must be optimized. This is one of the tasks of cell planning.

Distinguished from the spectral cost, the cost for building a mobile network is called financial cost. In an extremely competitive market, the financial cost is a key element in determining the economic feasibility of any communication system. However, economic aspects of mobile network design are not addressed effectively in most of the existing network planning tools. These tools are basically “prediction tools” [Cheung(1994)] because they merely predict the radio signal coverage and some system performance measures, such as bit error rate (BER), outage probability, and blocking probability, according to several established propagation models. The handling of economic issues is usually left to the expertise of network planners. A good design method should be able to minimize the financial cost, and to consider tradeoffs among factors such as system performance criteria, service demand, and system upgrade. This is another task of cell planning.

Currently, cellular system design is challenged by the need for a better quality of service and the need for serving an increased number of subscribers. Cell planning is becoming a key issue in the current scenario, with exceedingly high growth rates in many countries which force operators to reconfigure their networks virtually on a monthly basis. Therefore, the search for intelligent techniques, which may considerably alleviate planning efforts (and associated costs), becomes extremely important for operators in a competitive market.

1.2 RESEARCH SCOPE AND OBJECTIVES

Cell planning is a very complex task, as many aspects must be taken into account, including the topography, morphology, traffic distribution, existing infrastructure and so on. Things become more complicated because a handful of constraints are involved, such as the system capacity, service quality, frequency bandwidth, and coordination requirements. Nowadays, it is the network planner’s task to manually place BSs and to specify their parameters based on personal experience and intuition. These manual processes have to go through a number of iterations before achieving satisfactory performance and do not necessarily guarantee an optimum solution. It could work well when the demand for mobile services was low. However, the

explosive growth in the service demand has led to a need for an increase in cell density. This in turn has resulted in greater network complexity, making it extremely difficult to design a high-quality network manually.

Furthermore, the forthcoming third-generation (3G) systems such as the universal mobile telecommunications system (UMTS) aim to offer a multitude of telecommunication services, ranging from speech communication and video telephony to high data rate file transfer up to 2 Mb/s. The traffic density generated will vary substantially as different services require different transmission rates that can vary up to two orders of magnitude [Holma(2000)]. As a result, the traffic distribution will be dynamic. This also makes the conventional design methods insufficient for planning mobile networks in the future. Thus more advanced and intelligent network planning tools are required. A promising planning tool should be able to aid the human planner by automating the lower levels of the design processes [Cheung(1994)]. These include automatic placement and dimensioning of BSs.

This thesis presents a novel automatic cell planning approach that aims to design a high-quality mobile network which guarantees the system performance i.e. a network which meets the requirements of coverage, capacity, and interference level, while trying to minimize system costs, including the spectral cost and financial cost. Since BSs can be located at any place in the service area and the cell dimensions are very changeable, there are lots of variables to be determined, including the number of cells and parameters of each cell. Suppose that the service area is digitized into $n \times m$ pixels and each pixel is a potential cell site. The number of possible cell sets, i.e., different combinations of potential cells, is 2^{nm} when only considering the BS site. In principle, this is a nondeterministic polynomial (NP) combinatorial optimization problem [Garey(1979)]. Since the parameters of each cell, such as the transmitted power, antenna type, height, and orientation, also need to be optimized, both combinatorial and parametric optimization problems are involved in cell planning.

Although many studies have been reported in the area of optimization problems in terms of algorithm and complexity, relatively few studies have been done regarding the application of optimization algorithms for mobile network design. Basically, most available channel assignment algorithms can be used to minimize the spectral cost, such as those presented in [Box(1978), Kunz(1991), Mathar(1993), Ngo(1998), and Beckmann(1999)], but they do not consider the optimization of cell sites and dimensions. On the other hand, algorithms for minimizing the financial cost has not been extensively investigated. [Calégari(1997), Chamaret(1997), and Molina(1999)] studied genetic algorithms (GAs) for choosing a minimum number of cells from the given potential cells, while [Tutschku(1997)] used neural network techniques in an integrated cellular network planning tool. In [Hao(1997)], an algorithm based on simulated annealing (SA) was applied to solve an economic optimization model for cellular system design. SA algorithm was also used in [Hurley(2000)]. However, none of them take spectrum efficiency into account.

One important characteristic of the cell planning problem is that the design criteria are interrelated: increasing the coverage probably causes higher financial cost, improving the system capacity probably requires more channels. Thus there are several objectives needed to be considered simultaneously. Usually the spectral cost and financial cost are to be minimized, while maximum coverage or capacity are desired. Such a problem is classified as multi-objective optimization problem (MOP) because it involves simultaneous optimization of several objectives.

The common way of dealing with this class of problems is to use a weighted sum of the normalized objective functions, or to optimize one of the objectives while treating the others as constraints, such that the MOP is converted into and subsequently solved as a single objective optimization problem (SOP). In MOP, however, because the objectives may be conflicting, there does not exist a unique optimal solution that is global optimum with respect to all the objectives. Instead, there is a set of optimal solutions which represent tradeoffs between the conflicting objectives. These solutions are optimal in the wider sense that no other solutions in the search space are superior to them when all objectives are considered. Ideally we would like to obtain a set of tradeoff solutions, from which a good compromise solution can be chosen subject to the preference of the decision maker.

The potential of GAs to become a powerful method for multi-objective optimization has long been recognized [Schaffer(1985)]. GAs seem particularly suitable to solve MOPs because they deal simultaneously with a set of possible solutions, namely, a population. This inherent characteristic allows us to find a set of tradeoff solutions in a single optimization run. Therefore, multi-objective genetic algorithms (MOGAs) are studied in this work for the purpose of solving the cell planning problem. This is the first step in the direction of automatic cell planning and forms the basis for further research in this area.

1.3 THESIS OUTLINE

This thesis is arranged in seven chapters. Chapter 2 provides the fundamentals and basic techniques of cell planning. The conceptual basis of cellular system design and design objectives are first discussed. This is followed by a review of radio wave propagation in a mobile communications environment. A typical design process of cell planning is also outlined. The basic steps, including the coverage prediction, capacity prediction, interference analysis, and channel assignment, are discussed in detail from Section 2.3 to 2.6.

In Chapter 3 the cell planning problem is analyzed from the viewpoint of system performance optimization. First, a general description of the problem is given. Four kinds of problems can be distinguished: cell selection, cell dimensioning, automatic base station placement and dimensioning (ABSPAD), and growth planning. For the

first time, system performance criteria are addressed by the local conditions and global conditions. The local conditions, including the traffic load and interference probability, are defined on each single cell, while the global conditions are defined on the whole system (a set of cells), including the coverage rate, spectral cost, and financial cost. The global conditions are used as optimization objectives. Cell planning problems are typical MOPs as several objectives need to be considered simultaneously. The principle of multi-objective optimization is introduced in Section 3.5 which also reviews classical methods for solving MOPs.

The insufficiency of classical methods gives rise to the research on MOGAs. A general introduction to GAs is given in Section 4.1. This is followed by a more detailed analysis of the design aspects of GAs. Subsequently, real-valued GAs are presented for continuous parameter optimization. Section 4.4 is devoted to MOGAs. First, the key elements of GAs for multi-objective search are discussed, then, a MOGA is presented which makes use of non-dominated sorting, crowded comparison, and controlled elitism techniques [Deb(2000a), Deb(2000b)]. Hence it is called an elitist non-dominated sorting genetic algorithm (NSGA). Moreover, a new approach is proposed for solving MOPs by integrating the search and decision making processes. This approach, together with the NSGA, will be applied to the cell planning problems.

In Chapter 5 disk graph models (DGMs) are introduced for the purpose of investigating algorithms for cell planning problems. In Section 5.2 test problems for cell selection model are presented and the elitist NSGA is applied. Various GA parameters are investigated in order to improve the algorithm performance. In Section 5.3 real-valued GAs are applied to cell dimensioning problems, in comparison with binary GAs. Afterwards, the BS placement problem is considered in Section 5.4. The computational complexity makes this problem intractable. To overcome this problem, a novel method termed hierarchical approach is presented in this work, aiming to effectively reduce the computational time.

The methodology and algorithms presented in this work have been tested in real design scenarios. Numerical experiments have been made on all the four cell planning problems: cell selection, cell dimensioning, ABSPAD, and growth planning. First, a description of the simulation environment is given in Section 6.1. Afterwards, in Section 6.2, simulation results are presented and analyzed for each of the experiments.

Finally, Chapter 7 summarizes main contributions of this thesis and offers future perspectives.

Chapter 2

Cell planning fundamentals

This chapter provides background material and techniques for cell planning in cellular radio networks. We start with the conceptual basis of cellular system design. The design objectives of cell planning are then specified, and a typical design process is outlined. This is followed by a discussion about radio wave propagation in a mobile communications environment in Section 2.2. Some widely used propagation models are reviewed. Afterwards, we will discuss the basic design steps in the following sections. Section 2.3 presents a probabilistic model to predict the cell coverage. Section 2.4 gives a traffic estimation model and an algorithm to calculate the number of traffic channels (TCHs). In Section 2.5 the interference probability between a pair of cells is calculated. Based on the interference probability the compatibility matrix can be obtained which will be used for the channel assignment. The channel assignment problem (CAP) itself is considered in Section 2.6.

2.1 WHAT IS CELL PLANNING?

2.1.1 Elements of cellular system design

A cell is the area served by a base station transmitter, which is the basic geographical unit of a cellular system. In practice, engineers draw hexagonal-shaped cells on a layout to simplify the planning and design of a cellular system because it approaches a circular shape that is the ideal coverage area of a base station [Lee(1989)]. In the real world, however, each cell varies depending on the radio environment. Due to constraints imposed by the natural and man-made terrain, the real shape of cells is very irregular. Based on the size of cells, they can be classified as one of the following three types:

Macrocells provides overall coverage, especially to fast-moving mobiles like those in cars. The base station antenna is mounted above the surrounding buildings, providing a cell radius from around 1 to 30 km.

Microcells are used in areas with high subscriber density such as urban and suburban areas. Base station antennas are placed at an elevation of street lamps, so the shape of microcells is defined by the street layout. The cell length is up to 2 km.

Picocells are designed for very high mobile user density or high data rate applications, typically in indoor environments. Base station antennas are below the rooftop or at the elevation of bookshelves so the coverage is dictated by the shape and characteristics of rooms and the service quality is affected by the presence of furniture and people. The radius of picocells is between 10 and 200 m.

Table 2.1 summarizes the three types of cells [Geng(1998), Prasad(1998)].

Table 2.1 Different cell characteristics

Cell type	Cell radius*	Transmitted power	Antenna height	Use
Macrocell	1 ~ 30 km	1 ~ 10 W	> 30 m, top of tall buildings	Large area coverage, support high-speed mobiles
Microcell	0.2 ~ 1 km	0.01 ~ 1 W	< 10 m, street lamp elevation	High subscriber density areas
Picocell	< 200 m	< 100 mW	Ceiling/top of bookshelves	Mainly for indoor

* Cell radius is used to indicate the cell size, although the cell shape is not circular.

Traditional mobile systems were constructed in a fashion similar to radio or television broadcasting: One very powerful transmitter located at the highest spot in an area broadcasts in a radius of up to 50 kilometers. The cellular concept structures the mobile telephone network in a spectral efficient way. Instead of using one powerful transmitter, many low-power transmitters are placed throughout the service area. For example, by dividing a metropolitan region into one hundred different areas (cells) with low-power transmitters using 12 channels each, the system capacity theoretically can be increased from 12 channels using one powerful transmitter to 1,200 channels using one hundred low-power transmitters. This means that the number of channels per area is greatly increased. The cellular concept includes the following basic elements:

- Frequencies (channels) used in one cell can be reused in other cells some distance away.
- Conversations can be handed off between cells to maintain seamless service as mobile users travel from cell to cell.
- Variable power levels allow cells to be dimensioned according to the subscriber density and service demand within a particular region.
- As the demand for mobile services grows the system should be able to accommodate that growth by subdividing a congested cell into smaller cells.

In the following, these basic elements will be explained from the point of view of cellular system design.

A Frequency reuse

Frequency reuse is the core concept of the cellular radio system, in which users in different geographical locations (different cells) may simultaneously use the same frequency channel (FCH). The frequency reuse concept can drastically increase the spectral efficiency, but if the system is not properly designed, serious interference may occur. Interference due to the common use of the same channel is called co-channel interference (CCI) and is our major concern in cell planning.

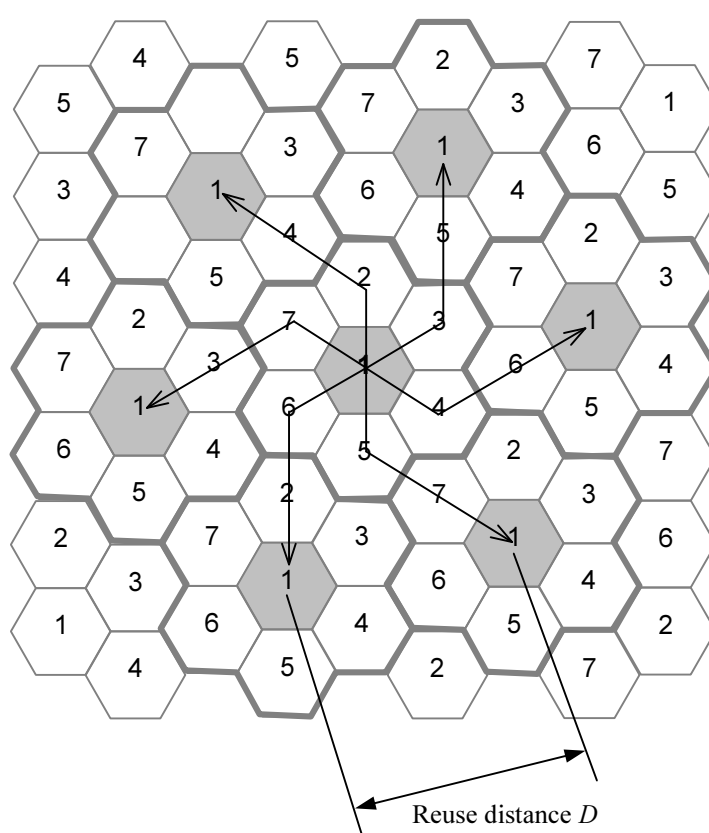


Figure 2.1 Frequency reuse pattern. In this example, the cluster size is 7 i.e. each cell uses 1/7 of the available FCHs. Cells with the same number have the same set of channels. Each cell has six co-channel interfering cells in the first tier.

The set of frequencies available in the system is assigned to a group of cells constituting a cluster as shown in Figure 2.1. Cells are assumed to have a regular hexagonal shape and the number of cells per cluster determines the frequency reuse pattern. Only certain reuse patterns can tessellate because of the hexagonal geometry. The cluster size K (the number of cells per cluster) is given by

$$K = i^2 + ij + j^2 \quad (2.1)$$

where i and j are integers. Thus, the clusters can accommodate only certain numbers of cells such as 3, 4, 7, 9, 12, 13, 16, 19, 21, and so on.

In a cellular system, when the size of each cell is approximately the same, CCI is independent of the transmitted power and becomes a function of the cell radius R and the distance between the nearest co-channel cells D . The ratio D/R , called the co-channel reuse ratio, is related to the cluster size. In a hexagonal geometry it is found that

$$D/R = \sqrt{3K}. \quad (2.2)$$

Obviously, a small cluster size provides a higher spectral efficiency that is defined as the number of subscribers per FCH supported by the system. The fewer cells per cluster, the larger the number of channels per cell and the more the reuse times of channels. And thus more subscribers can be accommodated. In practice, however, the cluster size is selected according to the requirement of the carrier-to-interference ratio (C/I). Let n be the number of co-channel interfering cells, the C/I ratio at a mobile unit can be expressed as

$$\frac{C}{I} = \frac{C}{\sum_{i=1}^n I_i} \quad (2.3)$$

where C is the carrier signal power and I_i is the interference power caused by the i th co-channel cell. It is demonstrated in [Rappaport(1996)] that the C/I ratio is directly proportional to the cluster size K :

$$\frac{C}{I} \cong \frac{(D/R)^\gamma}{n} = \frac{(\sqrt{3K})^\gamma}{n} \quad (2.4)$$

where γ is the path loss exponent, typically ranging from 2.7 to 4 for ground-to-ground mobile communications. In a hexagonal cellular system there are six co-channel interfering cells in the first tier (see Figure 2.1). The interference caused by co-channel cells in the second and higher tiers can be neglected, thus $n = 6$. Suppose the path loss exponent $\gamma = 4$. If a C/I ratio of 21 dB is required for satisfactory channel performance, i.e.

$$\frac{C}{I} = \frac{(\sqrt{3K})^\gamma}{n} = \frac{(3K)^2}{6} \geq 21 \text{ dB}, \quad (2.5)$$

we get $K \geq 12$. Therefore the 12-cell reuse pattern can be employed for maximum spectral efficiency. This example indicates that the CCI determines the transmission

quality, which in turn dictates the frequency reuse plan and the overall capacity of cellular systems.

B Traffic engineering

A decisive factor in determining how many subscribers can share a mobile network is the amount of traffic caused by each subscriber. The traffic per subscriber is defined by the call rate and average duration of a call. Because the call rate is measured in calls/second and call duration is in seconds/call, the traffic is a dimensionless quantity that, however, is given in units of Erlang. Since the traffic fluctuates with the time of day and the day of the week or year, a cellular system is usually dimensioned for the busiest hour, the so-called busy hour. Providing the average conversation time is T seconds and the number of calls per subscriber at the busy hour is λ , the traffic offered by an average subscriber is

$$a = \frac{\lambda \cdot T}{3600} \text{ Erlang.} \quad (2.6)$$

This can be interpreted as the fraction of time that each user occupies a channel.

The maximal traffic load that can be carried by a cell with n channels depends on how the system handles the calls when the n channels are fully occupied. There are three major blocking formulas [Minoli(1993)]:

Poisson: Blocked calls will wait no longer than their average holding time. If a channel becomes idle before the holding time expires, the call will seize it and use it for the remaining part or duration of its holding time.

Erlang B: Blocked calls are not willing to wait, and abandon the call attempt immediately. The user is not making another call attempt.

Erlang C: Blocked calls wait an indefinite time to obtain a channel.

A common practice in the cellular industry is to assume a loss system (all blocked calls are cleared), when evaluating the amount of traffic that can be serviced while keeping the blocking probability at or below a certain level e.g. 2%. Therefore the Erlang B formula is often used for this purpose, which is expressed as

$$P_{block} = \frac{\frac{A^n}{n!}}{\sum_{i=0}^n \frac{A^i}{i!}}. \quad (2.7)$$

It describes the relationship between three variables: the blocking probability P_{block} , the traffic load A (in Erlang), and the number of channels n . Obviously, the blocking

probability increases with the traffic load and decreases with the number of channels, as shown graphically in Figure 2.2. Because of its importance, the formula has been widely tabulated. Readers are referred to [Lee(1989)] for a complete Erlang B table. By knowing any two of the three variables (A , n , and P_{block}), one can derive the third.

If the number of channels and the maximum allowed blocking probability are known, the traffic capacity of a cell can be determined, say A_i Erlang. Providing there are N cells in a cellular system, the system capacity is then

$$A = \sum_{i=1}^N A_i \text{ Erlang.} \quad (2.8)$$

And thus, the number of subscribers that can be supported during the busy hour is

$$M = A / a. \quad (2.9)$$

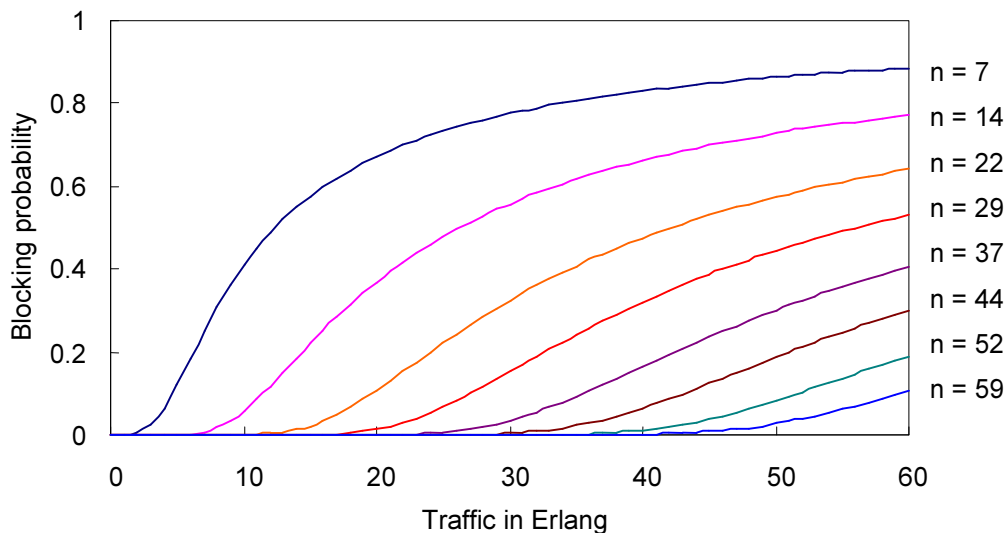


Figure 2.2 Erlang B curves. Each curve is marked with the number of channels.

C Handoff

Handoff, also called handover, is a process to transfer an ongoing call from one cell to an adjacent cell as a mobile user approaches the cell boundary. Handoffs must be performed successfully, be imperceptible to the users, and as infrequently as possible. In order to meet these requirements, the following design factors must be considered in cell planning:

- The cellular system must be designed in such a way that there are adequate overlapping area between cells.

- In high traffic density areas where microcells are normally used, high speed users may suffer many handoffs. This situation can be improved by employing the so-called umbrella cells [Rappaport(1996)], and thus a hierarchical cellular system is constructed.
- A handoff is initiated when the received signal strength reaches the threshold for handoff P_{ho} which is higher than the minimum usable signal strength P_{min} for acceptable voice quality. The handoff margin, given by $\Delta_{ho} = P_{ho} - P_{min}$, must be carefully chosen. If Δ_{ho} is large, many unnecessary handoffs may occur simply because the handoffs are taken too early. If Δ_{ho} is small, there may be insufficient time to complete a handoff before a call is lost due to weak signal conditions. It is evident that the handoff margin gives influence on the coverage and capacity of cells. This will be discussed in Subsection 2.3.2.

D Cell splitting and sectoring

There are two kinds of cell splitting: coverage-oriented splitting and capacity-oriented splitting. Figure 2.3(a) illustrates the coverage-oriented cell splitting in which four new cells are placed in the vicinity of the original cell. The radius of the new cells is the same as that of the original cell, and the coverage increases by a factor of

$$\frac{4R^2 + 2\pi R^2}{\pi R^2} \cong 3.27. \quad (2.10)$$

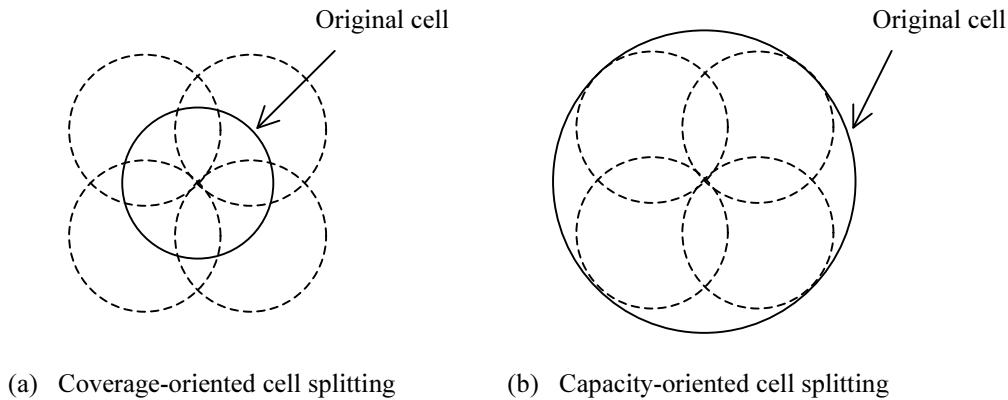


Figure 2.3 Coverage-oriented cell splitting and capacity-oriented cell splitting. In (a) the radius of new cells is the same as that of the original cell. The new cells cover a larger area. In (b) the radius of new cells is half of the original cell. The traffic capacity increases.

This cell splitting technique can be used if the coverage area needs to be increased. The second is capacity-oriented cell splitting, see Figure 2.3(b). A cell is split into smaller cells, each having the same number of channels as the original large cell.

Suppose the splitting is executed in such a way that the radius of the new cells is half of the original cell. To cover the complete area of the original cell at least four new cells are needed. Thus, the traffic capacity increases by four times. For the new cells to be smaller in size, the transmitted power P_t of these cells must be reduced by 12 dB [Lee(1993)] i.e.

$$P'_t = P_t - 12 \text{ dB.} \quad (2.11)$$

Another way to increase capacity is to keep the cell radius unchanged and decrease the cluster size as well as the co-channel reuse ratio D/R . From the above discussion it is clear that the CCI determines the cluster size, and thus the frequency reuse plan of cellular systems. The CCI may be reduced by replacing a single omnidirectional antenna at the base station by several directional antennas. This technique is called sectoring. A cell is normally partitioned into three 120° sectors or six 60° sectors. In [Lee(1998)], it is demonstrated that the C/I improvement of sectoring for three sectors per cell is 3.16 dB.

As described before, a cluster size of 12 should be used for the required C/I ratio of 21 dB. If there are, e.g., 396 channels, each cell may have 33 channels. Assume that the maximum allowed blocking probability is 1%. From the Erlang B table, it can be found that each cell may accommodate 22.9 Erlang traffic for an omni-directional antenna system. For a three-sector system, the cluster size of 7 can be used as $C/I = (3 \times 7)^2 / 6 + 3.16 \text{ dB} = 21.82 \text{ dB} > 21 \text{ dB}$. There are about 57 channels per cell and 19 channels per sector. For the same blocking probability, it can be found that each sector can handle 11.2 Erlang traffic, giving a cell capacity of 33.6 Erlang. The capacity increases by a factor of $33.6/22.9 \cong 1.47$.

2.1.2 Objectives of cell planning

One important goal when planning a cellular system is to achieve a high spectral efficiency. In other words, we want that a large number of subscribers per square kilometer would be able to use the system while maintaining an acceptable grade of service (GOS¹) and minimizing the system cost. This includes the following aspects:

Coverage. The radio signal coverage must be guaranteed. The received signal strength must always be higher than a threshold according to the receiver sensitivity, and holes in the coverage area should be avoided.

Capacity. In each cell, a sufficient number of channels must be available in order to meet its traffic demand for new calls and handoffs. Usually, handoffs should be given a priority. An acceptable GOS must be maintained for the whole system.

¹ GOS is a measure of congestion which is specified as the blocking probability (for Erlang B), or the probability of a call being delayed beyond a certain time (for Erlang C).

Transmission quality. The C/I ratio of radio channels must satisfy the requirement of transmission quality. The compatibility matrix used for channel assignment specifies the restrictions for channel separation that guarantees the required signal quality.

Cost. Since the cost and complexity of a mobile network is closely related to base stations (BSs), the number of BSs must be minimized so as to reduce the cost and the time of system deployment. Furthermore, it is essential to use the spectrum efficiently.

System growth capability. Cell planning is an endless task because the demand for mobile services keeps growing. New cells are needed to accommodate that growth. The new deployed cells must be integrated with the existing infrastructure. This usually means that the cell design should fit some coordination requirements. It is also required for the evolution to new generation systems.

These design objectives are all reflected in cell planning which is one of the basic steps to take when a cellular system is to be implemented.

2.1.3 Design process

Figure 2.4 shows the typical workflow of cellular system design. The traffic demand i.e. how many subscribers will join the mobile system and how much traffic they will generate, provides the basis for all the cellular network engineering. The geographical distribution of traffic demand can be estimated by using demographic data such as the population distribution, land usage, income level, and other factors.

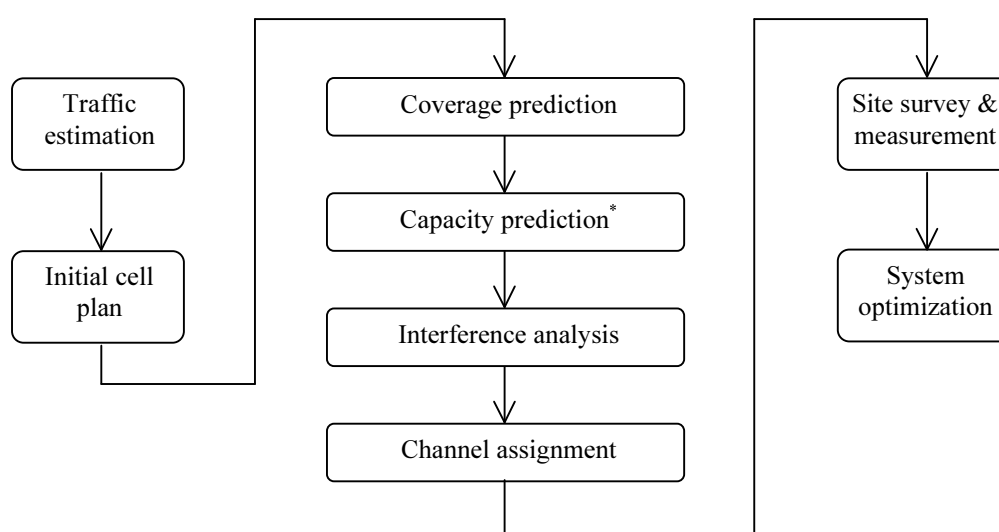


Figure 2.4 Cellular system design workflow.

* Capacity prediction may be skipped in some cases.

The initial cell plan gives a preliminary structure of the cellular system based on available sites of the network operator, called friendly sites. Besides, the traffic distribution is taken into account for specifying the BS configuration. For example, a BS at a rural site may have 3 sectors and 2 transceivers (TRXs), whereas an urban BS may have 3 sectors and 6 TRXs. The design principles presented in Subsection 2.1.1 can be used in this step.

Next, the field strength is calculated for each cell. Since radio propagation is affected by the terrain profile as well as the land usage type, e.g., the building structure and vegetation type, the topographical data and morphological data of the service area should be used for field strength prediction. The following steps including the coverage prediction, capacity prediction, interference analysis, and channel assignment are all based on the field strength prediction. These steps will be discussed in detail from Section 2.3 to Section 2.6.

In order to get accurate radio coverage prediction and information about the signal quality, site surveys and field measurements should be performed. Each BS site must be visited in order to determine whether to accept it, or reject it and find a new one.

System optimization means analyzing the extensive measured data from an operating network in order to tune the system parameters so that the service quality of the whole system can be improved. For instance, add FCHs to cells if they are often congested, and reduce the number of channels in cells with lower traffic than expected.

2.2 RADIO WAVE PROPAGATION ASPECTS OF CELL PLANNING

A profound knowledge of the radio wave propagation or the radio channel characteristics is a prerequisite to achieving an optimum system design. In this section, an overview is presented on several propagation models that are widely used in practice. More detailed discussions of wave propagation models can be found in [Steele(1992), Rappaport(1996), Geng(1998), Saunders(1999), C atedra(1999)].

Modeling the radio channel is one of the most difficult parts of mobile system design. Hence, worldwide research activities have been done to develop propagation models which aim at providing an accurate estimation of the mean received power or path loss (PL) for a specified frequency band based on more or less detailed geographical information about the environment. According to their nature, the propagation models can be classified as empirical, semi-empirical (or semi-deterministic), and deterministic models [Fleury(1996), COST 231(1999)].

Empirical models are described by equations or curves derived from statistical analysis of a large number of measured data. These models are simple and do

not require detailed information about the environment. They are also easy and fast to apply because the estimation is usually obtained from closed expressions. However, they can not provide a very accurate estimation of the path loss.

Deterministic models are calculation methods which physically simulate the propagation of radio waves. Therefore the effect of the environment on the propagation parameters can be taken into account more accurately than in empirical models. Another advantage is that deterministic models provide a prediction for the wideband characteristics of radio channels, which could be used as system performance criteria instead of (or in addition to) the signal strength. Most of the deterministic models are based on ray-optical modeling approaches [Lebherz(1992), Kürner(1993), Cichon(1994), Didascalou(2000)]. The serious drawback of deterministic methods is the computational complexity.

Semi-deterministic models result from an empirical modification of deterministic models in order to improve the agreement with measurements. These methods require more detailed information about the environment than the empirical methods but not as much as the deterministic models. They are relatively easy and fast to apply because the results are obtained from closed expressions in a similar way to the empirical models.

For the practical application of propagation models there is an important tradeoff between the accuracy of the prediction and the speed with which the prediction can be made. In general, there is a relationship between these three types of models and the types of cells for which they are suitable. Due to their inherent empirical nature, the empirical and semi-deterministic models are suitable for macrocells. The semi-deterministic models are also suitable for homogeneous microcells where the parameters considered by the method characterize the whole scenario well. The deterministic models are suitable for microcells and picocells independently of their shapes, but they are not adequate for macrocells because the CPU time required in such environments makes these methods impractical.

2.2.1 Macrocell propagation models

In macrocells, it is important to predict the average received signal strength at a given location. Usually radio transmission takes place over irregular terrain. The terrain profile of a particular area, as well as the presence of trees, buildings, and other obstacles, must be taken into account for estimating the path loss. The data needed to describe such environments would be very large and the computational effort involved would be excessive. As a result, empirical models are widely used for propagation prediction in macrocells because these models have the advantage of implicitly taking account of all propagation factors, both known and unknown,

through actual field measurements.

Among numerous propagation models, the Okumura-Hata model is considered to be one of the most significant. It is a fully empirical model developed on the basis of an extensive series of measurements in urban environments. Hata presented a standard formula in [Hata(1980)], which has been extended to cover the frequency range above 1500 MHz by COST 231 study group of the European Cooperation in the Field of Scientific and Technical Research [COST 231(1991)]. The general path loss formula of the model is

$$L_{Hata} = a_1 + a_2 \log(f) - 13.82 \log(h_{BS}) + [44.9 - 6.55 \log(h_{BS})] \log(d) - L_{mh} \quad (2.12)$$

where

- L_{Hata} : Path loss in dB;
- f : Frequency in MHz ($150 \text{ MHz} \leq f \leq 2000 \text{ MHz}$);
- h_{BS} : Effective BS antenna height² in meters ($30 \text{ m} \leq h_{BS} \leq 200 \text{ m}$);
- d : Distance between BS and the mobile in kilometers ($1 \text{ km} \leq d \leq 20 \text{ km}$);
- a_1 = 69.55 for $150 \text{ MHz} \leq f \leq 1500 \text{ MHz}$
= 46.3 for $1500 \text{ MHz} < f \leq 2000 \text{ MHz}$;
- a_2 = 26.16 for $150 \text{ MHz} \leq f \leq 1500 \text{ MHz}$
= 33.9 for $1500 \text{ MHz} < f \leq 2000 \text{ MHz}$.

L_{mh} is the mobile antenna height correction given by

$$L_{mh} = 3.2 \log^2(11.75 \cdot h_{mobile}) - 4.97 \quad (2.13)$$

where h_{mobile} is the mobile antenna height in meters ($1 \text{ m} \leq h_{mobile} \leq 10 \text{ m}$).

Equation (2.12) is formulated for flat terrain in urban environments. Corrections have to be applied for hilly terrain and other land usage types (i.e., morpho classes) such as suburban, forest, and open area. Thus the path loss is calculated as

$$L_{macro} = L_{Hata} + L_{diff} - L_{morpho} \quad (2.14)$$

L_{diff} is the diffraction attenuation in dB, which is caused by obstacles in the propagation path. Various algorithms are implemented to calculate it. Among them, Epstein-Peterson model [Epstein(1953)] and Deygout model [Deygout(1966)] are commonly used. L_{morpho} is the morpho correction in dB. In practice, a map containing detailed morphological information should be used. The correction factor corresponding to each morpho class is obtained from measurement data. A table of

² The effective BS antenna height is measured with reference to the mean height above sea level along the terrain profile in range between 3 km and 15 km from BS.

morpho correction factors is given in Chapter 6.

2.2.2 Microcell propagation models

Experimental observations reveal that in microcells the path loss as a function of the distance between the transmitter and receiver is well described by a power law with the exponent close to two up to a distance called “breakpoint”. Beyond this breakpoint a power decay with a considerably larger exponent ($\approx 4\sim 5$) is observed [Green (1990)]. In this case the path loss is modeled as

$$L_{micro} = \begin{cases} L_1 + 10\gamma_1 \log(d), & d \leq d_b \\ L_1 + 10\gamma_1 \log(d_b) + 10\gamma_2 \log(d/d_b), & d > d_b. \end{cases} \quad (2.15)$$

This is known as the dual-slope model. L_1 is the reference path loss at $d = 1$ m, d_b is the breakpoint distance given by

$$d_b = \frac{4 \cdot h_{BS} \cdot h_{mobile}}{\lambda}, \quad (2.16)$$

where h_{BS} and h_{mobile} are BS and mobile antenna heights, respectively. λ is the wavelength. Typical values of the exponents are found by measurements to be around $\gamma_1 = 2$ and $\gamma_2 = 4$, with breakpoint distances of 200~500 m. In order to plan the sites of microcells effectively, it is important to ensure that co-channel cells have coverage areas which do not overlap within the breakpoint distance.

A Two-dimensional models

For microcells the BS antennas can be above, below, or at the same level as the surrounding buildings. Usually, two situations are distinguished according to the relative locations of the transmitter and receiver: line-of-sight (LOS) in case of no obstacles in the propagation path, otherwise non-line-of-sight (NLOS).

The COST 231-Walfisch-Ikegami model [COST 231(1999)] provides good results in situations where wave propagation above rooftops is dominant. This semi-deterministic model takes account of statistical parameters describing the character of urban environments, including the average building height h_{roof} , the average building separation b , the street width w , and the street orientation φ with respect to the propagation path. The path loss is calculated as

$$L_{WI} = \begin{cases} L_{sc}, & \text{for LOS} \\ L_0 + L_{rsd} + L_{msd}, & \text{for NLOS.} \end{cases} \quad (2.17)$$

L_{sc} is the propagation loss due to the street canyon phenomenon, see [Döttling(1997)]. It is calculated as

$$L_{sc} = 42.6 + 20 \log(f) + 26 \log(d) \quad (2.18)$$

where f is the frequency in MHz, d is the distance between the transmitter and receiver in kilometers. L_0 is the free space loss defined as

$$L_0 = 32.44 + 20 \log(f) + 20 \log(d). \quad (2.19)$$

L_{rsd} is the loss due to the diffraction and scattering from the rooftops of adjacent buildings to the street, and L_{msd} accounts for the multiple screen diffraction along rooftops. L_{rsd} is given by

$$L_{rsd} = -16.9 + 10 \log(f) - 10 \log(w) + 20 \log(\Delta h_{mobile}) + L_{ori} \quad (2.20)$$

where w is the width (in meters) of the street where the mobile user is. Δh_{mobile} is the difference between the average height of the buildings and the mobile antenna height in meters i.e. $\Delta h_{mobile} = h_{roof} - h_{mobile}$, and L_{ori} is a function of the street orientation φ .

$$L_{ori} = \begin{cases} -10 + 0.354 \cdot \varphi, & 0^\circ \leq \varphi < 35^\circ \\ 2.5 + 0.075 \cdot (\varphi - 35), & 35^\circ \leq \varphi < 55^\circ \\ 4 + 0.114 \cdot (\varphi - 55), & 55^\circ \leq \varphi < 90^\circ. \end{cases} \quad (2.21)$$

Finally, the multiple screen diffraction loss L_{msd} is given by

$$L_{msd} = L_{bsh} + k_a + k_d \log(d) + k_f \log(f) - 9 \log(b) \quad (2.22)$$

where b is the average building separation. L_{bsh} is the BS antenna height correction with $\Delta h_{BS} = h_{BS} - h_{roof}$.

$$L_{bsh} = \begin{cases} -18 \log(1 + \Delta h_{BS}), & h_{BS} > h_{roof} \\ 0, & h_{BS} \leq h_{roof}. \end{cases} \quad (2.23)$$

The other empirical correction factors are given as follows:

$$k_a = \begin{cases} 54, & h_{BS} > h_{roof} \\ 54 - 0.8 \cdot \Delta h_{BS}, & h_{BS} \leq h_{roof}, \quad d \geq 0.5 \text{ km} \\ 54 - 0.8 \cdot \Delta h_{BS} \cdot \frac{d}{0.5}, & h_{BS} \leq h_{roof}, \quad d < 0.5 \text{ km}. \end{cases} \quad (2.24)$$

$$k_d = \begin{cases} 18, & h_{BS} > h_{roof} \\ 18 - 15 \frac{\Delta h_{BS}}{h_{roof}}, & h_{BS} \leq h_{roof}. \end{cases} \quad (2.25)$$

$$k_f = -4 + a \left(\frac{f}{925} - 1 \right). \quad (2.26)$$

a is 1.5 if the mobile is in metropolitan centers, otherwise a is 0.7.

The COST 231-Walfisch-Ikegami model is valid for the following ranges: 800 MHz $\leq f \leq$ 2000 MHz, 4 m $\leq h_{BS} \leq$ 50 m, 1 m $\leq h_{mobile} \leq$ 3 m, 0.02 km $\leq d \leq$ 5 km. This model exclusively considers the propagation in the vertical plane.

Two-dimensional (2D) models considering the propagation in the horizontal plane have been proposed for the case where the antenna height is below the rooftops and the building height is assumed to be infinite. The propagation paths are determined by means of the ray-tracing method. The path loss is computed either with the uniform geometrical theory of diffraction (UTD) [Wart(1993)], by placing virtual sources at the building corners [Wagen(1993)], or by using the free space propagation model with a hypothetical distance recursively determined as a function of the number of street intersections along the paths [Berg(1995)].

B Three-dimensional models

Three-dimensional (3D) models are proposed for an arbitrary antenna height, e.g., the 3D-URBAN-MICRO model developed at IHE, the University of Karlsruhe [Kürner(1993), Cichon(1994)]. This model takes account of the propagation paths in both the vertical and the transverse plane. Multiple knife-edge diffraction effects are evaluated by successively computing single knife-edge diffraction based on the UTD. Additional 3D propagation paths via scattering at ground and building walls are separately regarded. A modified Kirchhoff Method with a scalar approximation is applied for surface scattering calculations.

2.2.3 Indoor propagation models

It has been observed that propagation within buildings is strongly influenced by specific features such as the layout of the building, construction materials, and the building type [COST 231(1999)]. An empirical method that considers the attenuation through the individual walls and floors is given by

$$L_{indoor} = L_0 + L_C + \sum_{i=1}^W n_{wi} L_{wi} + (n_f)^e L_f \quad (2.27)$$

where

- L_0 : Free space loss between the transmitter and receiver;
- L_C : Constant loss;
- W : Number of wall types;

- n_{wi} : Number of penetrated walls of type i ;
 L_{wi} : Penetration loss through a wall of type i ;
 n_f : Number of penetrated floors;
 L_f : Loss between adjacent floors;
 $e = (n_f + 2)/(n_f + 1) - k$, where k is a constant.

L_C and k must be determined empirically. Some recommended values at 1800 MHz are [Cátedra(1999)]: $L_{w1} = 3.4$ dB for light walls, $L_{w2} = 6.9$ dB for heavy walls, $L_f = 18.3$ dB and $k = 0.46$.

Indoor radio propagation is dominated by the same mechanisms as outdoor. Ray-tracing and the geometrical theory of diffraction (GTD) have been applied for indoor propagation prediction. These deterministic methods can be used for site-specific predictions, provided that sufficient details of the building geometry and materials are available. However, there are limitations in using such methods for practical picocell planning due to the difficulty of obtaining and using sufficiently accurate physical data. In addition, the influence of furniture and the movement of people can have a significant (and time-varying) effect on signal coverage. Therefore, the combination of deterministic modeling with statistical measurements is expected to be more practical for the prediction of indoor radio propagation [Zwick(2000)].

2.3 COVERAGE PREDICTION

A proper description of cell shapes is important for many design processes, e.g. coverage and capacity prediction, interference analysis, and channel assignment. Nominal cell descriptions assume hexagonal cell shapes. The hexagonal cells are based on the assumption of isotropic propagation conditions and a regular distribution of base stations. These assumptions are generally not fulfilled in real radio networks. This section presents a probabilistic model to describe the realistic cell shapes based on path loss prediction.

2.3.1 Coverage probability

Providing the service area is digitized as $m \times n$ pixels and the pixel size is $\Delta x \times \Delta y$. The values of Δx and Δy depend on the accuracy of available maps. Each pixel can be viewed as a mobile location, which is denoted by (x, y) , $x \in [1, m]$ and $y \in [1, n]$. The local mean received power \bar{P}_r (in dB) i.e. field strength at a specific location can be calculated as

$$\bar{P}_r = P_t + G_t + G_r - L_p \quad (2.28)$$

where P_t is the transmitted power. G_t and G_r are the transmitter and receiver antenna

gains, respectively. L_p is the path loss that can be predicted using a suitable propagation model. A location is considered to be covered by a cell if the received power level is higher than a given threshold. Due to the random shadowing effects, however, the actual received power P_r is a Gaussian distributed random variable around the mean value \bar{P}_r with standard deviation σ . Typically in an urban environment $\sigma = 8$ dB. It could be any value between 4 dB and 12 dB, depending on the actual surroundings. The distribution of P_r is therefore described as the Gaussian (normal) probability density function (pdf)

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \bar{P}_r)^2}{2\sigma^2}\right). \quad (2.29)$$

The probability that the field strength received from a transmitter is higher than a particular value θ is the integral of the pdf over the range (θ, ∞) , expressed as

$$\begin{aligned} P\{P_r > \theta\} &= \int_{\theta}^{\infty} g(x) dx \\ &= \int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \bar{P}_r)^2}{2\sigma^2}\right) dx \\ &= Q\left(\frac{\theta - \bar{P}_r}{\sigma}\right) \end{aligned} \quad (2.30)$$

where $Q(z)$ is the so-called Q-function defined as

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} \exp\left(-\frac{x^2}{2}\right) dx. \quad (2.31)$$

Obviously,

$$Q(z) = 1 - Q(-z). \quad (2.32)$$

The coverage probability i.e. the probability of a location (x, y) being covered by a cell c_i , is then given by

$$P_{cov}(x, y, c_i) = Q\left(\frac{\theta - F_i(x, y)}{\sigma}\right) \quad (2.33)$$

where θ represents the desired received power threshold and $F_i(x, y)$ is the predicted field strength i.e. \bar{P}_r .

Figure 2.5 shows an example of the coverage probability calculation of a single

cell. The field strength shown in Figure 2.5(a) is predicted by using the COST 231-Okumura-Hata model. The dark area in Figure 2.5(b), in which the field strength is below the given threshold, is taken as the uncovered area according to the simple threshold method. The coverage prediction based on the probabilistic model is shown in Figure 2.5(c). The probability of each pixel is calculated by applying (2.33) with $\theta = -92$ dBm and $\sigma = 6$ dB. Because of the radio signal characteristics, the boundary of cells is not crisp. Therefore, the probabilistic model gives a more realistic description of the cell shape.

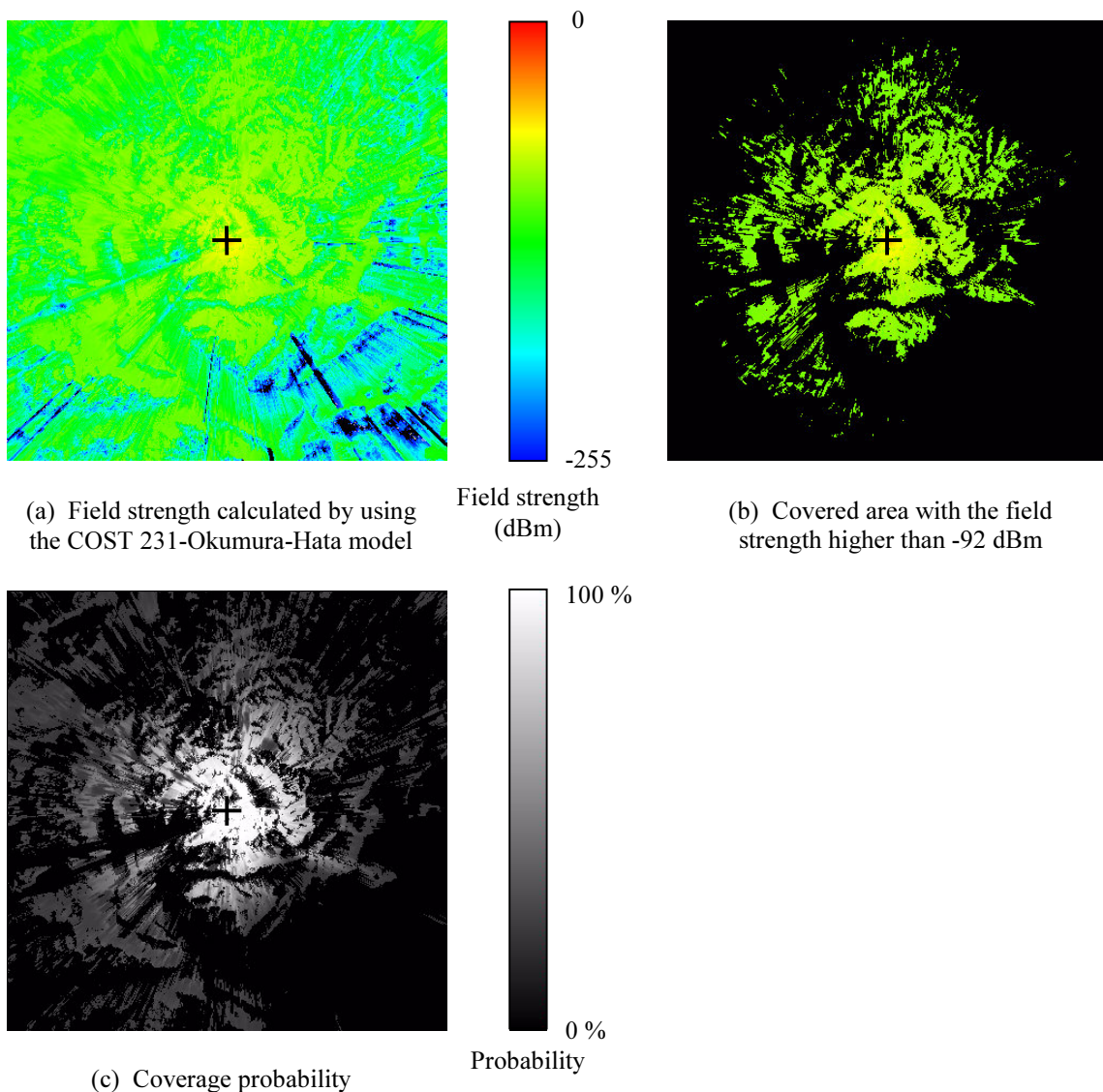


Figure 2.5 An example of the coverage probability calculation of a single cell. (a) shows the field strength in the whole area (30×30 km²). In (b) a simple threshold method is used. The dark area in which the field strength is below the given threshold is taken as uncovered area. In (c) the coverage probability is calculated with $\theta = -92$ dBm and $\sigma = 6$ dB. A pixel with higher field strength has larger coverage probability.

2.3.2 Assignment probability

Traditionally the coverage areas are assigned to cells using the best server model. This model is based on the field strength prediction. It assigns a mobile location to a cell that provides the strongest signal. The drawback of this model is that a dedicated assignment of each location to a single cell is assumed i.e. overlapping of cells is neglected. A more advanced method is to assign a location to a cell with a certain probability, considering the handoff between cells.

The assignment probability is defined as the probability of a mobile unit at (x, y) being served by a cell c_i . In principle, it is calculated according to the signal strength received from c_i relative to the strongest signal strength at (x, y) i.e.

$$F_{max}(x, y) = \max_i \{F_i(x, y)\}. \quad (2.34)$$

Handoff strategies influence the assignment of a mobile to cells [Kürner(1998)]. Cells from which the signal strength plus the handoff margin Δ_{ho} is still less than the strongest signal strength i.e.

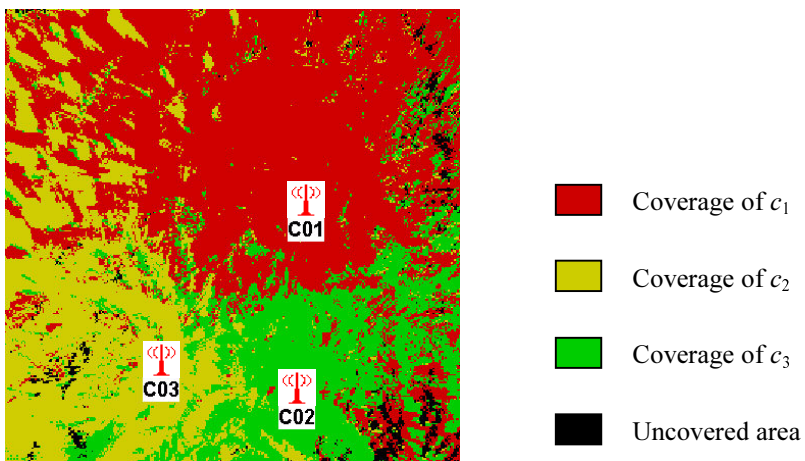
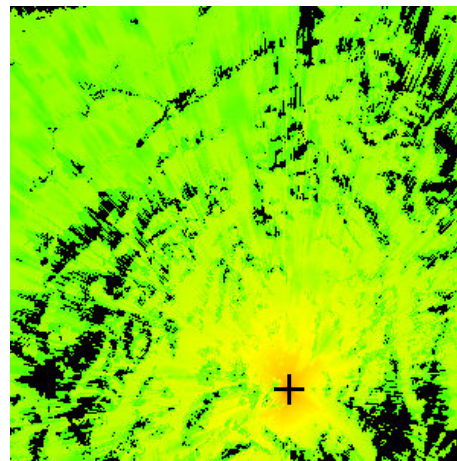
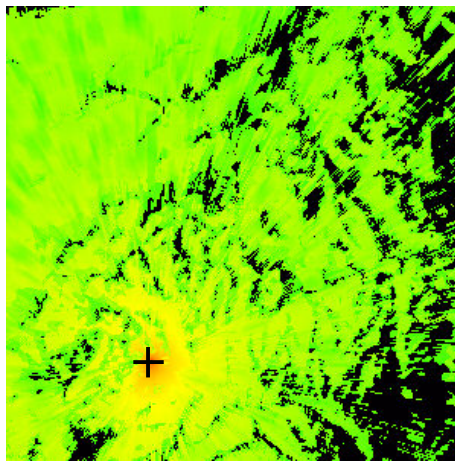
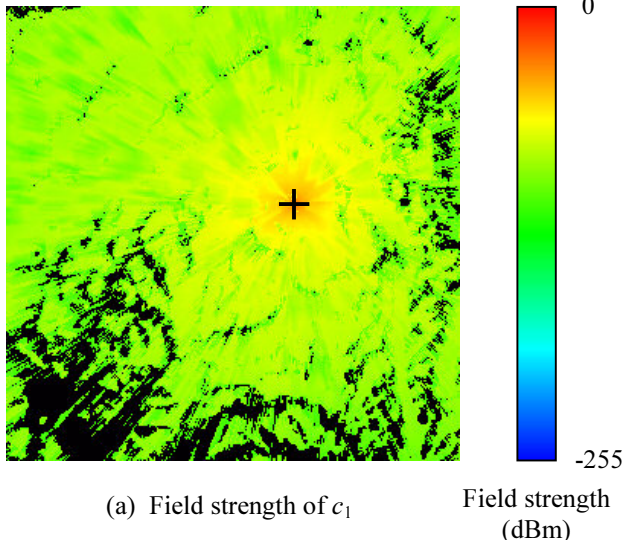
$$F_i(x, y) < F_{max}(x, y) - \Delta_{ho}, \quad (2.35)$$

will be discarded. The assignment probability of the discarded cells is 0. This means that these cells can not serve a mobile unit at the location (x, y) . The assignment probability of each of the remaining cells is calculated as

$$P_{ass}(x, y, c_i) = \frac{F_i(x, y) - F_{max}(x, y) + \Delta_{ho}}{\sum_{j=1}^{N'} (F_j(x, y) - F_{max}(x, y) + \Delta_{ho})} \quad (2.36)$$

where N' is the number of remaining cells.

Figure 2.6 shows a cellular system containing three cells. The field strength of these cells are shown in Figure 2.6(a), (b), and (c), respectively. The best server coverage is shown in Figure 2.6(d), where a pixel is exclusively assigned to only one cell. Consequently the coverage area breaks into fragments. As seen in the figure, there are numerous small holes in the coverage area. This means that frequent handoffs are inevitable when the mobile user moves. Fortunately, this is not the case in reality because the handoff margin is considered. Therefore, a mobile unit in the coverage area could be served by one of several cells if they overlap each other. The assignment probability is calculated pixelwise by applying (2.35) and (2.36) with $\Delta_{ho} = 5$ dB. The results are shown in Figure 2.6(e), (f), and (g) for each cell correspondingly. The assignment probability also indicates that the boundary of cells is not crisp.



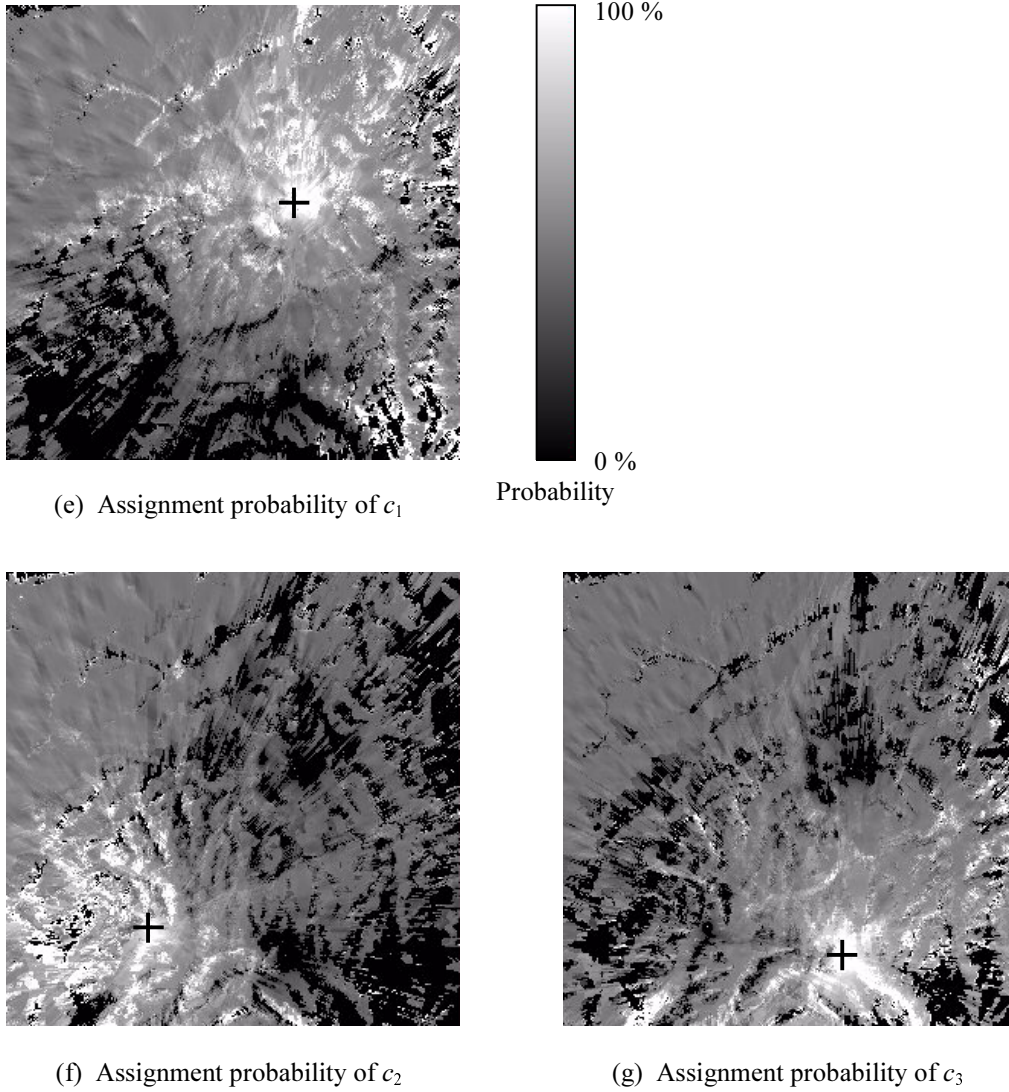


Figure 2.6 An example of the assignment probability calculation. (a), (b), and (c) show the field strength of cells c_1 , c_2 , and c_3 , respectively. The result of the coverage prediction based on the best server model is shown in (d). The assignment probability is calculated with $\Delta_{ho} = 5$ dB. Results are shown in (e) - (g).

In the preceding section, we use the coverage probability to define the area covered by a single cell. Besides, the assignment probability presented in this section is useful for a set of cells. This probabilistic model gives a realistic description of the cell shapes. Next we will use this model to predict the cell coverage. The coverage of a cell c_i is calculated as the sum of the pixels covered by it i.e.

$$A_{area}(c_i) = \sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, c_i) \cdot P_{ass}(x, y, c_i) \cdot \Delta x \cdot \Delta y . \quad (2.37)$$

There are $m \times n$ pixels and the pixel size is $\Delta x \times \Delta y$. Each pixel is weighted by the coverage probability $P_{cov}(x, y, c_i)$ and assignment probability $P_{ass}(x, y, c_i)$. In case there is only one cell, the assignment probability is set to be 1 for every pixel.

2.4 TRAFFIC CAPACITY

2.4.1 Traffic estimation

As discussed in Subsection 2.1.3, traffic estimation is a basic step of cellular system design. The primary task of cell planning is to locate and configure transmission facilities i.e. BSs in the service area. These facilities should be located close to the expected traffic in order to increase system efficiency. A traffic model that describes the traffic distribution as observed from the nonmoving network elements e.g. BSs and switches, is required to estimate the spatial extended traffic demand. In [Tutschku(1998)], a simple linear model is presented for this purpose.

The traffic density $\tau^{(t)}(x, y)$ (in Erlang/km²) represents the amount of traffic caused in a unit area at time t . It is usually defined on a 2D grid with a pixel size of $\Delta x \times \Delta y$. The values of Δx and Δy are the same as those used for coverage prediction.

$$\tau^{(t)}(x, y) = \lambda^{(t)}(x, y) \cdot T \quad (2.38)$$

where $\lambda^{(t)}(x, y)$ is the number of call attempts in a unit area element at location (x, y) during time interval $(t, t + \Delta t)$. T is the mean call duration. In practice, it is almost impossible to directly calculate the call attempts $\lambda^{(t)}(x, y)$. The traffic density therefore has to be derived from measurements of a real network. Generally, the mobile system must be dimensioned to offer a minimum GOS (i.e. blocking probability) during the busy hour. Consequently, we assume that all figures and equations refer to the busy hour traffic and thus the time-varying traffic model is reduced to a stationary model describing the peak traffic.

The spatial traffic distribution heavily depends on the morphological and demographic factors like population, vehicular traffic, income per capita, and so on. Providing morphological information of the service area is available, the whole area is divided into clusters such that each cluster is composed of the adjacent pixels with the same morpho class (e.g. urban, suburban, rural). To estimate the traffic demand, a linear prediction model is applied for every cluster j .

$$E_j = \sum_{\text{all factors } i} \mu_{i,k} + \tau_k \quad (2.39)$$

where E_j represents the Erlang value of cluster j . k is the cluster type i.e. morpho

class. Assume that the area of cluster j is S_j , then

$$\tau_k = c_k \cdot S_j \quad (2.40)$$

is the primary traffic load based on the cluster type and size, and

$$\mu_{i,k} = b_{i,k} \cdot x_i \quad (2.41)$$

is the expected traffic resulting from factor i under the circumstances of morpho class k . x_i is the quantity of the factor i , e.g. the population of a city or vehicular traffic of a highway. The parameters $b_{i,k}$ and c_k can be determined by training this linear model using measured data from an operating network. Finally, the traffic density in cluster j is derived as

$$\tau(x, y) = \frac{E_j}{S_j}. \quad (2.42)$$

By applying the linear model i.e. equations from (2.39) to (2.42), in the service area, we can obtain a 2D traffic density matrix that describes the spatial traffic distribution. The proposed model appears to be simple. However, due to its structure this model can be adapted to the proper parameters. This capability enables its application for mobile network design.

Obviously, the total traffic inside the service area is expected to be

$$A_T = \sum_{x=1}^m \sum_{y=1}^n \tau(x, y) \cdot \Delta x \cdot \Delta y. \quad (2.43)$$

And the traffic coverage of a cell c_i is calculated as the sum of the estimated traffic offered to this cell i.e.

$$A_{trf}(c_i) = \sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, c_i) \cdot P_{ass}(x, y, c_i) \cdot \tau(x, y) \cdot \Delta x \cdot \Delta y \quad (2.44)$$

where $\tau(x, y)$ is the traffic density at location (x, y) , weighted by the coverage probability $P_{cov}(x, y, c_i)$ and assignment probability $P_{ass}(x, y, c_i)$ which are defined in (2.33) and (2.36), respectively. In case of a uniform traffic distribution, the traffic density $\tau(x, y) \equiv 1$, thus the traffic coverage is equivalent to the area coverage, i.e. $A_{trf}(c_i) = A_{area}(c_i)$, see (2.37).

2.4.2 The number of channels required

The amount of traffic offered to a cell may not be equal to the traffic carried by the

cell. Some of the traffic may be blocked (or potentially delayed). The Erlang B formula, see (2.7), describes the blocking probability P_{block} as a function of the traffic load A (in Erlang) and the number of traffic channels n . A traffic channel (TCH) is a logical channel over which the information is exchanged between mobile users during a connection.

For every cell, the blocking probability must be lower than an acceptable level in order to guarantee the desired GOS of the cellular system. If the traffic offered to a cell is known, the blocking probability is then a function of the number of channels n . From (2.7) we get

$$P_{block}(n) = \frac{A \cdot P_{block}(n-1)}{A \cdot P_{block}(n-1) + n}. \quad (2.45)$$

It is clear that $P_{block}(0) = 1$. Therefore, the blocking probability can be calculated by iteratively applying (2.45).

In cellular system design, it is more useful to determine the number of TCHs required to achieve the desired GOS. Based on (2.45), an iterative algorithm can be applied to calculate the number of channels required by a cell with certain amount of traffic. The algorithm is described as follows:

Algorithm 2.1 (Number of traffic channels)

Input: A – The traffic in Erlang
 P_b – Maximum allowed blocking probability

Output: The number of channels n

Step 1 (Initialization):

 Set $n = 0$ and $p = 1.0$

Step 2 (Iteration):

 While $p > P_b$ do

 1) $n = n + 1$;

 2) Calculate the blocking probability $p = A \cdot p / (A \cdot p + n)$.

This algorithm is quite efficient because it is not necessary to calculate the factorial n . Just imagine how large $100!$ would be, i.e. completely impractical.

2.4.3 Demand vector

The number of TCHs is not equal to the number of FCHs. Usually a frequency band is

divided into a number of channels by frequency division multiple access (FDMA), and a FCH may be shared by multiple users with a certain access scheme such as time division multiple access (TDMA) or code division multiple access (CDMA). The GSM system for example uses TDMA. Each FCH is divided into eight timeslots to enable eight simultaneous communications between users i.e. each FCH can provide a maximum of eight TCHs. Besides, control channels (CCHs) are needed to transmit the system control signal. Table 2.2 gives the required number of CCHs corresponding to a given number of TCHs. The number of FCHs is simply determined by the fact that each channel can hold eight timeslots, regardless whether they carry traffic or control signal i.e.

$$n_{FCH} = \left\lceil \frac{n_{TCH} + n_{CCH}}{8} \right\rceil, \quad (2.46)$$

where n_{FCH} , n_{TCH} , and n_{CCH} are the number of FCHs, TCHs, and CCHs, respectively. The operation $\lceil x \rceil$ makes a real value x to be rounded up to the closest larger integer value.

Table 2.2 The numbers of CCHs and FCHs corresponding to the number of TCHs

Traffic channels*	Number of channels required	
	Control channels	Frequency channels
1 ~ 7	1	1
8 ~ 14	2	2
15 ~ 22	2	3
23 ~ 29	3	4
30 ~ 37	3	5
38 ~ 44	4	6
45 ~ 52	4	7
53 ~ 59	5	8

* If more than 59 TCHs are required, the number of CCHs is set to 5 and the number of FCHs is calculated as 8 timeslots per frequency.

In the following, a vector \vec{W} is used to represent the number of FCHs required by a set of cells:

$$\vec{W} = (w_1, w_2, \dots, w_N) \quad (2.47)$$

where N is the number of cells. \vec{W} is called the demand vector that will be used in the

channel assignment procedure.

2.5 INTERFERENCE ANALYSIS

The next step is to assign a sufficient number of FCHs to each cell. Before assigning channels an interference analysis must be done because the interference resulting from frequency reuse is the major concern in cell planning. A large number of references can be found in the literature for detailed discussions about the interference in cellular systems [Lee(1993), Prasad(1993), Perez(1998)]. This section is concerned with the mutual interference calculation in order to determine the necessary channel separation between each pair of cells.

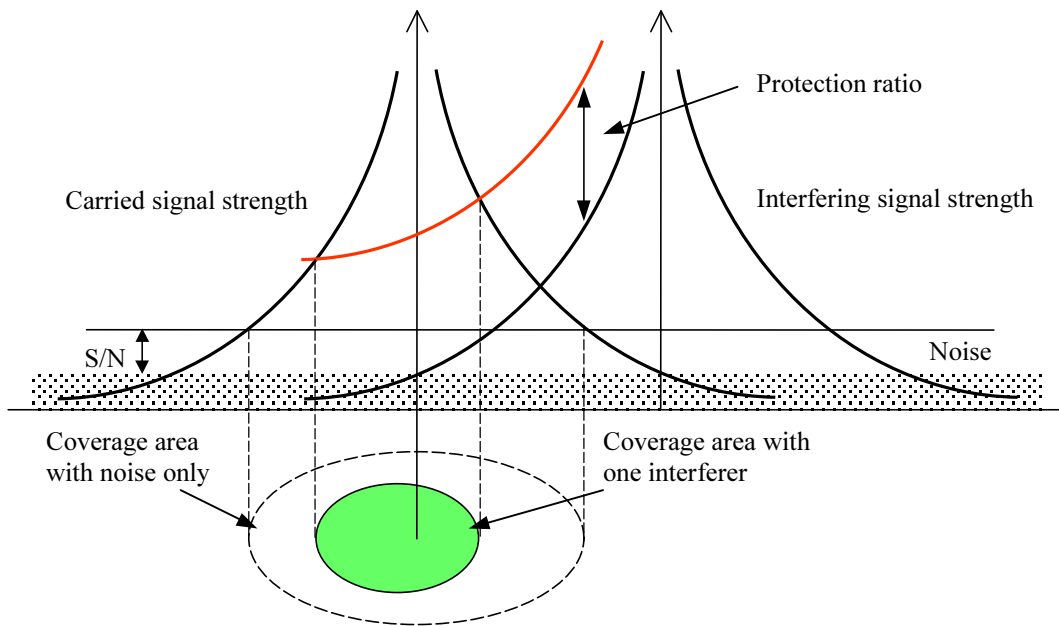


Figure 2.7 Mutual interference between two cells (adapted from [Quellmalz(1993)]).

2.5.1 Mutual interference

Mutual interference is calculated for a pair of cells: the serving cell and interfering cell. The interference between two cells is illustrated in Figure 2.7. For each pixel (x, y) in the calculation area, the actual C/I ratio is calculated based on the field strength prediction. Assume that the received power at (x, y) is $F_i(x, y)$ for the serving cell c_i and $F_j(x, y)$ for the interfering cell c_j , the C/I ratio of these two cells is given by

$$C_i / I_j(x, y) = F_i(x, y) - F_j(x, y). \quad (2.48)$$

The CCI probability at this pixel is then calculated as

$$P_{co}(x, y, c_i, c_j) = 1 - \operatorname{erf}\left(\frac{C_i / I_j(x, y) - \theta_{C/I}}{\sqrt{2}\sigma}\right) \quad (2.49)$$

where $\theta_{C/I}$ is the threshold for C/I ratio in dB and σ is the standard deviation of the received signal power. $\operatorname{erf}(z)$ is the error function defined as

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-x^2) dx. \quad (2.50)$$

On adjacent channels, the interfering power level is attenuated by an additional loss due to the filter characteristics of the receiver. Assume that the filter has a slope of ρ dB/oct, and that the bandwidth of each channel is f_B . If the frequency difference between two adjacent channels is Δ_f , the k -th adjacent channel attenuation is given by [Lee(1993)]

$$L(k) = \frac{\rho}{0.3} \log\left(\frac{k \cdot \Delta_f}{f_B}\right), \quad k = 1, 2, 3, \dots \quad (2.51)$$

k is just the channel separation. Table 2.3 gives some values of the channel separation for GSM900 system, $k = 0$ is considered for co-channel usage. Providing $\rho = 6$ dB/oct, $f_B = 25$ kHz, and $\Delta_f = 200$ kHz, then the first adjacent channel attenuation is

$$L(1) = \frac{6}{0.3} \times \log\left(\frac{200}{25}\right) \cong 18 \text{ dB}. \quad (2.52)$$

Table 2.3 The values of the channel separation

Channel separation	Frequency difference* (kHz)	Protection ratio (dB)
0	0	9
1	200	-9
2	400	-41
3	600	-49

* The values of frequency difference are specified for GSM900, see [ETSI 05.05(2000)].

Since the protection ratio is very small if the channel separation is greater than 3, the adjacent channel interference (ACI) can be neglected for $k > 3$. Therefore, the interference probabilities at the pixel (x, y) for channel separations from 1 to 3 are calculated as

$$P_k(x, y, c_i, c_j) = 1 - \operatorname{erf}\left(\frac{C_i / I_j(x, y) + L(k) - \theta_{C/I}}{\sqrt{2}\sigma}\right), \quad k \in [1, 3]. \quad (2.53)$$

After the interference probabilities are calculated for each pixel in the calculation area, the average CCI probability and ACI probabilities can be expressed as follows:

$$\bar{P}_{co}(c_i, c_j) = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{co}(x, y, c_i, c_j) \cdot P_{cov}(x, y, c_i) \cdot P_{ass}(x, y, c_i) \cdot \Delta x \cdot \Delta y}{A_{area}(c_i)}, \quad (2.54)$$

$$\bar{P}_k(c_i, c_j) = \frac{\sum_{x=1}^m \sum_{y=1}^n P_k(x, y, c_i, c_j) \cdot P_{cov}(x, y, c_i) \cdot P_{ass}(x, y, c_i) \cdot \Delta x \cdot \Delta y}{A_{area}(c_i)}, \quad k \in [1, 3]. \quad (2.55)$$

The coverage probability and assignment probability of c_i at the pixel (x, y) are $P_{cov}(x, y, c_i)$ and $P_{ass}(x, y, c_i)$, respectively. $A_{area}(c_i)$ is the area coverage of c_i , see (2.37). Note that, because of the adjacent channel attenuation, the following relationship exists:

$$\bar{P}_{co}(c_i, c_j) > \bar{P}_1(c_i, c_j) > \bar{P}_2(c_i, c_j) > \bar{P}_3(c_i, c_j). \quad (2.56)$$

2.5.2 Compatibility matrix

The compatibility matrix, also known as channel separation matrix, is a symmetric matrix represented as $S_{N \times N}$, which has as many rows and columns as the number of cells N . Each row, as well as each column, corresponds to a cell. The element s_{ij} of the matrix indicates the channel separation between cells c_i and c_j . The channel separation specifies the minimum frequency difference between two cells. A channel separation of zero allows co-channel usage, while a channel separation of one forbids co-channel but allows the first adjacent channel, and so on. The values of s_{ij} may be 0, 1, 2, or 3 according to the mutual interference between the corresponding cells.

First, the compatibility matrix is initialized, considering some restrictions such as the minimum channel separation inside the same cell MIN_{cc} and the minimum channel separation between co-site cells MIN_{cs} .

$$s_{ij}^{(0)} = \begin{cases} MIN_{cc}, & \text{if } c_i = c_j \\ MIN_{cs}, & \text{if } c_i \text{ and } c_j \text{ are located on the same site} \\ 0, & \text{otherwise.} \end{cases} \quad (2.57)$$

Typically, MIN_{cc} is greater than or equal to five, MIN_{cs} is greater than or equal to three.

In the next step, the interference probabilities for co-channel and adjacent channels are used to calculate the channel separation.

$$s_{ij}^{(1)} = \begin{cases} 0, & \text{if } \bar{P}_{co}(c_i, c_j) < \theta_{int} \\ 1, & \text{if } \bar{P}_1(c_i, c_j) < \theta_{int} < \bar{P}_{co}(c_i, c_j) \\ k+1, & \text{if } \bar{P}_{k+1}(c_i, c_j) < \theta_{int} < \bar{P}_k(c_i, c_j), \text{ for } k \in [1, 2]. \end{cases} \quad (2.58)$$

θ_{int} is the threshold for interference probability. For $s_{ij}^{(1)}$ the cell c_i is the server and c_j is the interferer, and for $s_{ji}^{(1)}$ it is just the reverse case. Hence it usually holds that $s_{ij}^{(1)} \neq s_{ji}^{(1)}$. The maximum of $s_{ij}^{(1)}$ and $s_{ji}^{(1)}$ will be taken as the final channel separation. Finally, the initial value $s_{ij}^{(0)}$ needs to be taken into account. Thus

$$s_{ij} = \max(s_{ij}^{(0)}, s_{ij}^{(1)}, s_{ji}^{(1)}), \quad \text{for } i, j \in [1, N]. \quad (2.59)$$

A typical compatibility matrix may look as follows.

$i \backslash j$	1	2	3	4	...
1	5	3	1	1	
2	3	5	0	0	
3	1	0	5	2	
4	1	0	2	5	
...					

2.6 CHANNEL ASSIGNMENT

The channel assignment problem (CAP), also known as the frequency assignment problem (FAP), has two basic aspects. On the one hand, the demand vector \vec{W} must be satisfied. This means that a sufficient number of FCHs must be assigned to each cell in order to satisfy the traffic demand. On the other hand, the frequencies used simultaneously in different cells may incur interference resulting in loss of quality of the signal. This requires that the assignment of the channels to the cells must comply with the compatibility constraints, as specified by the compatibility matrix S . Both aspects should be taken into account for allocating channels. Numerous algorithms for channel assignment in cellular systems are available in the literature. Readers are

referred to [Katzela(1996)] and [Koster(1999)] for a good survey. This section starts with a general description of the CAP, and then presents a heuristic algorithm to solve the CAP.

2.6.1 Channel assignment strategies

The frequency band $[f_{min}, f_{max}]$ available to some provider of mobile communications is usually partitioned in a set of channels that are all with the same bandwidth f_B and equally spaced by Δ_f . For this reason the channels are usually numbered from 1 to a given maximum M_{max} ,

$$M_{max} = \frac{f_{max} - f_{min}}{\Delta_f}. \quad (2.60)$$

The available channels are denoted by the domain $Z = \{1, \dots, M\}$. In mobile communication systems the traffic is usually bidirectional. Therefore, one needs two channels for the data transmission between the BS and the mobile. One is for the uplink direction (from the mobile to the BS), another is for the downlink direction (from the BS to the mobile). Instead of one band $[f_{min}, f_{max}]$, two bands $[f^1_{min}, f^1_{max}]$ and $[f^2_{min}, f^2_{max}]$ are used in real networks: one with the channels $\{1, \dots, M\}$, and one with the channels $\{s + 1, \dots, s + M\}$, where $s \gg M$. Thus, the downlink uses a channel which is shifted s channels up. The choice of s prevents any interference of downlink channels with uplink channels. Each assignment for the downlink channels can directly be transformed to an assignment for the uplink channels with similar performance. As a consequence, only downlink (or uplink) needs to be considered for the channel assignment.

Channel assignment strategies can be classified as either fixed or dynamic. The choice of channel assignment strategy impacts the system performance, particularly as to how calls are managed when a mobile user is handed over from one cell to another [Rappaport(1996)]. In a fixed channel assignment (FCA), each cell is allocated a predetermined set of channels. Any call attempt within the cell can only be served by the unused channels in that particular cell. If all the channels in that cell are occupied, the call is blocked, even though there are channels available in adjacent cells. In order to overcome the deficiencies of FCA schemes the dynamic channel assignment (DCA) strategy is introduced. In DCA, channels are not allocated to different cells permanently. Instead, each time a call request is made, the serving BS requests a channel from the mobile switching center (MSC). The switch then allocates a channel to the requested cell following an algorithm that takes into account the likelihood of future blocking within the cell, the usage frequency of the candidate channel, compatibility constraints, and other cost functions. At the cost of higher complexity, the DCA strategy provides flexibility and adaptability to the time-varying traffic distribution. It also increases the trunking capacity of the system since all the

available channels are accessible to all of the cells.

In general, there is a tradeoff between the spectrum utilization efficiency and implementation complexity of the channel assignment algorithms. It is proved in [Johri(1994)] that DCA schemes perform better than FCA schemes under light traffic. However, FCA schemes become superior at heavy traffic load, especially in the case of uniform traffic distribution. To obtain a better overall performance, hybrid channel assignment (HCA) algorithms are designed by combining the FCA and DCA schemes. In one HCA strategy, called the borrowing strategy, a cell is allowed to borrow channels from a neighboring cell if all of its own channels are already occupied. The MSC supervises such borrowing procedures and ensures that the borrowing of a channel does not disrupt or interfere with any of the calls in progress in the donor cell.

It is widely recognized that FCA gives a bound on the performance of the DCA schemes. Besides, in case the DCA scheme allows for complete rearrangement of the channel allocation a FCA problem has to be solved every time the situation changes. For these reasons, we concentrate on the FCA problem in the following.

2.6.2 A heuristic algorithm for channel assignment

A lot of research can be found in the literature for solving the CAP. Most of the investigations are based on graph theoretic or heuristic approaches [Box(1978), Gamst(1982), Sivarajan(1989), Smith(1998)]. Besides, algorithms employing neural networks [Kunz(1991), Funabiki(1992), Chan(1994), Kim(1997), Smith(1997)], simulated annealing [Mathar(1993), Duque-Anton(1993)], and genetic algorithms [Lai(1996), Ngo(1998), Beckmann(1999)] have also been proposed.

Providing a system of N cells is represented by a set $C = \{c_1, c_2, \dots, c_N\}$. Each cell c_i requires w_i channels ($w_i \geq 1$). This forms the demand vector $\vec{W} = (w_1, w_2, \dots, w_N)$, see Subsection 2.4.3. The assignment of the channels to the cells subjects to the constraints specified by the compatibility matrix S , in which the element s_{ij} indicates the minimum separation between channels assigned to cell c_i and those assigned to cell c_j .

Definition 2.1 (Channel assignment problem): Let a triple $P = (C, \vec{W}, S)$ represent a CAP, let $\{1, \dots, M\}$ be a set of channels and H_i the set of channels assigned to cell c_i . A solution of P is to find the minimum value of M such that there exists an assignment pattern $H = \{H_1, H_2, \dots, H_N\}$, which satisfies the following conditions:

$$|H_i| = w_i, \quad \text{for } i = 1, \dots, N;$$

$$\forall h \in H_i \text{ and } h' \in H_j, |h - h'| \geq s_{ij}, \quad \text{for } i, j = 1, \dots, N$$

where $|H_i|$ denotes the number of channels in H_i . h and h' denote an arbitrary channel in H_i and H_j , respectively.

This problem is equivalent to a generalized graph coloring problem [Sivarajan(1989)]. A CAP can be described by a graph $G(V, E)$, where V is the vertex set and E is the edge set. Each cell c_i is represented by a vertex v_i with weight w_i . If $s_{ij} > 0$, the vertices v_i and v_j are joined together by an edge e_{ij} with label s_{ij} . The resulting graph is called an interference graph. The CAP is equivalent to assigning positive integers $\{1, 2, \dots, M\}$ to the vertices such that each vertex v_i has w_i integers assigned. The difference between the integers assigned to two adjacent vertices must not be less than the edge label. The objective is to minimize the maximum integer used. In the special case, where only CCI is considered, s_{ij} is either zero or one. This problem is then reduced to the classical graph coloring problem. Since the graph coloring problem is known to be nondeterministic polynomial (NP) complete [Garey(1979)], the CAP is also NP-complete i.e. the computational complexity of searching for the optimum solution grows exponentially with the problem size.

Based on the graph coloring model, a heuristic algorithm for solving the CAP is proposed in [Sung(1997)]. Before introducing the algorithm, we have to define some terms. First of all, we define the neighborhood of v , $N(v)$ as the set of v 's adjacent vertices. A set of vertices in a graph, which are all completely interconnected, is called a clique. The weight of a clique is defined as the sum of weights of all the vertices in it. A vertex typically belongs to more than one clique. We denote $C_{max}(v)$ as the clique, which contains v and has maximum weight. The maximum clique weight of v , denoted as $W(v)$, is defined as the clique weight of $C_{max}(v)$. When it is necessary to make the corresponding graph G explicit, we write it as $W(v|G)$.

Basically the heuristic algorithm uses sequential strategy described in [Gamst (1982)], where a color z_i is picked up and assigned to the vertices one by one until no further assignment of that color is possible. Then, the next color z_{i+1} is used, and the procedure is started over again. The question is how to determine which vertices should be colored by z_i . Here the vertex with the greatest weight is taken as the first vertex. To choose the subsequent vertices, the principle of the maximum overlap of denial areas³ as defined in the third method in [Gamst(1982)] is used. This principle states that a channel should be assigned to the cell whose denial area has maximum overlap with the already existing denial area of that channel. The overlap is defined as the number of cells within the intersection of the two denial areas. When there is a tie, it is broken by choosing the vertex with the largest maximum clique weight with respect to the topology graph induced by the intersection of the denial areas. The rationale of this rule is that the larger the maximum clique weight is, the more

³ The denial area for a cell c is the set of neighboring cells which can not share the same frequency with c due to co-channel interference. The denial area of a channel is the set of vertices which can not be assigned with this channel [Sung(1997)].

difficult it is for that vertex to be colored. This algorithm is called the sequential packing (SP) algorithm, which is described in Algorithm 2.2. Note that K denotes the set of vertices being colored by the current color z , and F denotes the vertices which are forbidden to be colored by z :

Algorithm 2.2 (Sequential packing algorithm for channel assignment)

Input: $G(V, E)$ – The graph representing the CAP

Output: Channel assignment pattern $H = \{H_1, H_2, \dots, H_N\}$

Step 1:

Set $z = 1$;

Step 2:

Set $K = \emptyset$ and $F = \emptyset$;

Let v be a vertex in $\{v_i: v_i \in V \text{ and } w_i = \max_{v_j \in V} w_j\}$;

Step 3:

While $F \neq V$ do

1) $K = K \cup \{v\}$;

$F = F \cup \{v\} \cup N(v)$;

$m = \max_{v_i \in V-F} |N(v_i) \cap F|$;

2) If $m = 0$ then

let v be a vertex in $\{v_i: v_i \in V-F \text{ and } w_i = \max_{v_j \in V-F} w_j\}$,

else

a) $T = \{v_i: v_i \in V-F \text{ and } |N(v_i) \cap F| = m\}$;

b) If $|T| = 1$ then let v be the only vertex in T , else

i) For each $v_i \in T$ do

Construct a subgraph $G(v_i)$ induced by the vertex set
 $(N(v_i) \cap F) \cup \{v_i\}$;

ii) Let v be a vertex in $\{v_i: v_i \in T \text{ and } W(v_i|G(v_i)) = \max_{v_j \in T} W(v_j|G(v_j))\}$;

Step 4:

For each $v_i \in K$ do

1) $H_i = H_i \cup \{z\}$;

$w_i = w_i - 1$;

2) If $w_i = 0$ then

a) $V = V - \{v_i\}$;

b) Delete all the edges connecting to v_i ;

Step 5:

$z = z + 1$;

If $V = \emptyset$ then return $\{H_1, H_2, \dots, H_N\}$,
else go to Step 2.

This algorithm ensures that the cells to which a channel is assigned are packed as close to each other as possible. It turns out that this algorithm always yields an optimal solution for cellular systems with a certain special structure. The proof is given in [Sung(1997)]. The computational complexity of the algorithm is $O(MN^2)$, where M is the number of channels used and N is the number of cells.

2.7 SUMMARY

This chapter discussed the fundamentals and basic techniques of cell planning. First, the elements of cellular system design were considered, which include the frequency reuse concept, traffic engineering, handoff mechanisms, and cell splitting techniques. The cellular system design objectives were also specified.

Afterwards, radio wave propagation was taken into account. Propagation models can be classified in three categories: empirical, semi-empirical (or semi-deterministic), and deterministic models. An empirical model (COST 231-Okumura-Hata model) and a semi-deterministic (COST 231-Walfisch-Ikegami model) were reviewed, which are widely used in practice to predict the propagation loss for macrocells and microcells, respectively.

A typical design process of cell planning was outlined. The basic steps, including the coverage prediction, capacity prediction, interference analysis, and channel assignment, were discussed in detail. First, the probabilistic model based on the coverage probability and assignment probability gives a more realistic description of the cell coverage. Second, a linear model is used to estimate the traffic distribution. Next, the number of TCHs required by a single cell is calculated using an algorithm based on the Erlang B formula. The demand vector is then obtained. Afterwards, the CCI probability and the ACI probabilities are calculated for each pair of cells. Based on the interference probabilities, the compatibility matrix is generated. Both the demand vector and compatibility matrix are used as input data in the channel assignment procedure. Channel assignment strategies can be classified as fixed (FCA), dynamic (DCA), or hybrid. The channel assignment problem is equivalent to a graph coloring problem that is known to be NP-complete. A heuristic algorithm for channel assignment was introduced.

Chapter 3

Cell planning as an optimization problem

In this chapter, we will discuss the cell planning from the viewpoint of system performance optimization. Section 3.1 gives a general description of the cell planning problem, followed by an overview of cell planning problems. Four kinds of problems can be distinguished in cellular system design practice. They will each be discussed in Section 3.2. The requirements of system performance, such as the coverage, capacity, and signal quality, are addressed in the optimization objectives. The performance requirements are classified as local conditions and global conditions because some of them are defined on a single cell and the others are defined on the whole system (a set of cells). The definitions of these conditions will be given in Section 3.3 and 3.4, respectively. Cell planning is a typical multi-objective optimization problem (MOP) as several objectives need to be considered simultaneously. In Section 3.5, the principles of multi-objective optimization will be introduced, and the methods for solving multi-objective problems will be discussed.

3.1 PROBLEM DESCRIPTION

Cell planning is a very complex task, as many aspects must be taken into account, including the topography, morphology, traffic distribution, existing infrastructure, and so on. Things become more complicated because several constraints are involved, such as the system capacity, service quality, frequency bandwidth, and coordination requirements. Nowadays, most radio network planning tools can only aid the network planners in radio frequency calculations such as the coverage prediction, interference analysis, and channel assignment. The network itself, however, has to be designed manually by the planners based on personal experience and intuition. The manual design process has to go through a number of iterations before achieving a satisfactory performance, and does not necessarily guarantee an optimum solution.

The problem is how to place BSs at the optimum sites and how to specify the optimum values for the parameters of each BS such that the optimal system performance is achieved and the system cost is minimized. In order to solve this problem, we first give a general description of the problem as follows:

Let $c_i = (x_i, y_i, p_i, h_i, a_i)$ denote a cell, where (x_i, y_i) is the BS site, p_i is the

transmitted power, h_i is the antenna height, and a_i is the antenna orientation. Several cells may share the same BS site (called co-site cells). Let $C = \{c_1, c_2, \dots, c_N\}$ be a cell set, N is the number of cells. These cells influence each other with respect to the coverage and interference. C will be called an admissible cell set if the following conditions hold:

1. The traffic load of each cell is less than some maximum value T_{max} corresponding to the available channels so as to guarantee the desired GOS of the cellular system.
2. The interference probability of each cell is lower than a given threshold θ_{int} , corresponding to the required signal quality.
3. The coverage rate i.e. the proportion of the area (or traffic) covered by these cells, is higher than a given threshold θ_{cov} , e.g., 95%.
4. The spectral cost, which is defined as the number of frequency channels required for assigning all the cells a sufficient number of channels, does not exceed the maximum available frequencies M_{max} .
5. The financial cost for building and maintaining these BSs should be limited to the budgetary infrastructure investment E_{budget} .

Conditions 1 and 2 are called local conditions as they are defined on each single cell, while the other three conditions (3-5) are called global conditions because they are defined on a set cells. The local conditions and global conditions will be discussed in more details in Section 3.3 and 3.4, respectively.

It is worth mentioning that these conditions are system-oriented. This means that their definitions are based on the system performance requirements. Different systems e.g. GSM or UMTS, could have their own definitions of the local conditions and global conditions. Other conditions can also be included as necessary. In this work the definitions are based on GSM system specification, but the methods presented here are also applicable to other systems.

Conditions 1-5 are required to guarantee the system performance, such as the coverage, capacity, and signal quality. The task of cell planning is to design an optimal cellular system i.e. to determine the number of cells, cell sizes, BS locations, and the parameters of each BS, which can guarantee the required performance goals while minimizing the total system costs, including the spectral cost and financial cost.

Essentially the cell planning is an optimization problem, which can be viewed as finding an optimal cell set from a large number of potential cells. If there are N potential cells, the number of possible combinations of these cells is 2^N , thus the computational complexity of searching for an optimal cell set grows exponentially with the problem size. Such a combinatorial optimization problem is NP-complete [Garey(1979)]. Since many factors and complicated constraints are involved, none of the available algorithms can be directly used for cell planning problems to obtain an

optimal solution in reasonable time. The key objective is therefore to develop an algorithm, which can obtain approximate optimal results with reasonable computational effort, taking into account all the above mentioned design factors.

Various optimization methods have been considered for this purpose. In principle, most available channel assignment algorithms can be used to minimize the spectral cost. The optimization of the financial cost, however, has seldom been studied. Some studies, e.g., [Calégari(1997)], [Chamaret(1997)], and [Molina(1999)], only considered the minimization of the number of cells. In [Hao(1997)], an economic optimization model was presented. It considered the economic efficiency of cellular system design, but did not address the efficiency of spectrum. The optimization strategy presented in this work has the advantage of taking into account all the requirements of system performance. It has the ability to minimize both the spectral cost and financial cost and to make tradeoffs among the system performance criteria.

3.2 OVERVIEW OF CELL PLANNING PROBLEMS

As discussed in the preceding section, there are numerous variables to be determined for an optimal cellular system design, including the number of cells N , cell site (x_i, y_i) , and parameters of each cell (p_i, h_i, a_i) . In practice, some of the variables could be given, or restricted in a very limited range. This gives rise to a variety of cell planning problems, although their tasks are the same. Four kinds of problems can be distinguished according to the given conditions. They are cell selection, cell dimensioning, automatic base station placement and dimensioning, and growth planning. We will discuss each of them next.

3.2.1 Cell selection

One of the key issues a mobile communication operator must face when deploying a cellular network is the selection of good locations for BSs. In principle, BSs can be located at any place in the service area. For some operators, however, the available sites for BSs are very limited due to the economic or technical reasons. The available sites where BSs may be installed are taken as input variables, and the goal is to find a minimum subset of the given sites that allows a good service in the planning area. This is the so-called cell selection problem i.e. to find an optimal cell set from a large number of potential cells.

By changing the selection of cells affects not only the number of cells but also the whole system performance and costs. Hence the cell selection is not a simple optimization problem. Assume that there are N potential cells according to the available BS sites, and furthermore the configurations of these cells are also given. As analyzed in the preceding section, the complexity of cell selection problem is

exponential (NP-hard). Several approaches have been investigated for this problem. One of them, as presented in [Calégari(1997)], models the cell selection as the minimum dominating set problem from graph theory and a genetic algorithm is employed to solve the combinatorial optimization problem. Another approach uses so-called demand nodes to represent control points in the service area [Tutschku (1997) and Tutschku(1998)], and applies neural network techniques to find the optimal cell set. In [Huang(2000a)], an approach based on genetic algorithms (GAs) was proposed for solving the cell selection problem, taking advantage of an integrated objective function of the coverage rate, spectral cost, and financial cost.

3.2.2 Cell dimensioning

In realistic mobile radio environments, the size of cells depends on the BS parameters such as the transmitted power p_i , antenna height h_i , and orientation a_i . The task of cell dimensioning is to specify these parameters so as to meet the requirements of system performance, which are addressed by the local conditions and global conditions. This is a parametric optimization problem. Since the number of cells and cell sites are given, the financial cost is determined, it does not need to be optimized further. The objective of cell dimensioning is then to minimize the spectral cost while keeping the other requirements satisfactory.

As demonstrated in Section 2.6, the spectral cost i.e. the number of frequencies required for channel assignment, depends mainly on the number of channels required by each cell and the mutual interference between cells. Hence, the local conditions of each cell must be fulfilled, which include the traffic load and interference as described previously.

The complexity of the cell dimensioning problem is also NP-complete. It is clear that the cells influence each other with respect to the local conditions. Any slight change on one cell could result in substantial differences on the conditions of other cells. The result of such a change is unpredictable, and can not be described by a simple function. Therefore, stochastic optimization methods e.g. genetic algorithms (GAs) and simulated annealing (SA), are good candidates for solving this problem. Although no report can be found in the literature for the application of stochastic optimization algorithms to the cell dimensioning problem, these algorithms have already been widely used to solve channel assignment problems e.g. [Duque-Anton(1993)], [Mathar(1993)], and [Beckmann(1999)]. However, the cell dimensioning is much more difficult than the channel assignment as the configurations of cells are unknown. The cell dimensioning problem was first considered in [Huang(2000c)], where a fuzzy expert system was developed to use the expert experience to adjust the BS parameters.

3.2.3 Automatic base station placement and dimensioning

For cell selection and cell dimensioning, the sites or dimensions of the potential cells must be given by the planner. This does not fulfil the task of automatic cell planning. A real automatic planning approach should have the capability to search potential cells (including BS locations and configurations) by itself. This kind of problems are called automatic base station placement and dimensioning (ABSPAD).

Obviously, ABSPAD is more difficult than the cell selection and cell dimensioning problems as more variables need to be determined. Suppose that the service area is digitized into $m \times n$ pixels, each pixel is a potential cell site as the BS can be located at any place in the service area. The number of possible cell sets, i.e., the combinations of potential cells, is 2^m when only considering the cell site (x_i, y_i) . Since the other parameters such as the transmitted power p_i , antenna height h_i , and orientation a_i are also unknown, the exponent is certainly much greater than $n \cdot m$. Therefore ABSPAD is also NP-complete.

Automatic cell planning has not been extensively studied yet. In [Hurley(2000)], an approach based on SA is used for cell site selection and configuration. Four specific neighborhood structures are used to generate new solutions based on given probabilities. A SA algorithm controls the choice and acceptance of new solutions. In [Huang(2000c)], a hierarchical approach is proposed for automatic BS placement. The idea is to search optimal cell sites on a series of grids with different resolutions from rough to fine. On each level only promising areas are searched, such that the size of search space can be controlled and will not expand exponentially. The hierarchical approach will be discussed in more detail in Chapter 5. Also, in Chapter 6, this approach will be applied to real planning problems.

3.2.4 Growth planning

It is clear that a smaller cell size increases the traffic capacity. However, a smaller cell size also means more BSs and a higher cost of the infrastructure. Obviously, we do not want to work with an unnecessarily small cell size. What is needed is in fact a method that matches cell sizes to the capacity requirements. So the system is usually started using large cells. The cell size is as large as possible. When the system capacity needs to grow, the cell size is decreased and the number of cells is therefore increased in order to meet the new requirements [Bäck(1991)].

The method commonly used in practice is cell splitting i.e. reducing the existing cell sizes and adding new cells. The problem is then to find optimal BS sites for new cells and optimize dimensions of both original cells and additional cells so that the growing demand can be met effectively, while offering minimum disturbance to the existing network structure. This is the so-called growth planning problem.

The growth planning problem becomes more and more important and frequently faced by the network planners due to the tremendous growth in demand for mobile services. This problem was first studied in [Vasudevan(1999)]. In an existing network, if the subscriber growth makes some of the existing cells unable to carry the increased traffic, the coverage of these cells is reduced by lowering the transmitted power level, so that the traffic in the smaller area can be supported. At the same time, new cells are needed to cover the remaining uncovered traffic. In [Huang(2000d)], the above-mentioned hierarchical approach was applied to find the optimal new cell sites, taking into account the existing cells. The automatic growth planning algorithm will be presented in Chapter 5.

Table 3.1 summarizes the above-mentioned cell planning problems. The difficulty of the problem increases from the cell selection to growth planning.

Table 3.1 Classification of cell planning problems

Problem	Number of Cells N	Cell sites (x_i, y_i)	Cell dimensions (p_i, h_i, a_i)	Existing cells*
Cell selection	Unknown	Known	Known	No
Cell dimensioning	Known	Known	Unknown	No
ABSPAD	Unknown	Unknown	Unknown	No
Growth planning	Unknown	Unknown	Unknown	Yes

* The dimensions of existing cells also need to be considered. This makes the growth planning problem more difficult.

3.3 LOCAL CONDITIONS

As discussed in Section 3.1, two local conditions need to be considered for each cell: traffic load and interference condition. Obviously, these two conditions are very important for a high-quality cellular system design. Besides, these conditions have great influence on the channel assignment, and therefore, are also important for a spectrum efficient system design.

In the following, the two local conditions will be defined. We use a triple $P = (C, \vec{W}, S)$ to represent a channel assignment problem, where $C = \{c_1, c_2, \dots, c_N\}$ is the cell set, N is the number of cells, $\vec{W} = (w_1, w_2, \dots, w_N)$ is the demand vector, S is the

compatibility matrix. The spectral cost i.e. the minimum number of channels for solving P , is denote by $F_0(P)$.

3.3.1 Traffic load

According to traffic theory, the blocking probability increases with the traffic load offered to a cell and decreases with the number of available channels. For each cell, the blocking probability must be lower than an acceptable level in order to guarantee the desired GOS. If the available channels are given, the traffic load within the cell should be less than some maximum value T_{max} corresponding to the allowed blocking probability.

Theorem 3.1: Let $\{1, \dots, M\}$ be the available frequency channels, MIN_{cc} be the minimum separation between the channels used by the same cell. w_i is the number of channels required by the cell c_i . For an admissible cell set, it is necessary that

$$w_i \leq \left\lfloor \frac{M-1}{MIN_{cc}} \right\rfloor + 1, \quad \text{for } i = 1, \dots, N. \quad (3.1)$$

The operation $\lfloor x \rfloor$ makes a real value x to be rounded up to the closest smaller integer value.

Proof: According to the first-level lower bound given in [Gamst(1986)], we have

$$F_0(P) \geq (w_i - 1) \cdot s_{ii} + 1, \quad \text{for } i = 1, \dots, N,$$

where $s_{ii} = MIN_{cc}$ (typically greater than or equal to five). Since $F_0(P)$ must be less than M in order to solve the channel assignment problem with the available channels, obviously,

$$M \geq (w_i - 1) \cdot MIN_{cc} + 1, \quad \text{for } i = 1, \dots, N.$$

Therefore (3.1) is obtained. ■

Assume that the maximum allowed blocking probability P_{block} is known e.g. 2%. Furthermore, let

$$n_{max} = \left\lfloor \frac{M-1}{MIN_{cc}} \right\rfloor + 1,$$

which is the maximum number of channels that can be used by a cell. By knowing P_{block} and n_{max} , we can derive the traffic capacity of a cell according to the Erlang B formula. The result gives the maximum traffic load T_{max} .

Suppose that the service area is digitized as $m \times n$ pixels and the pixel size is

$\Delta x \times \Delta y$. The traffic density at the location (x, y) is represented by $\tau(x, y)$. There are $\tau(x, y) \cdot \Delta x \cdot \Delta y$ Erlang of traffic at (x, y) . The traffic load of a cell c_i is just its traffic coverage $A_{trf}(c_i)$ i.e. the sum of the traffic covered by c_i . The traffic coverage has already been defined in Section 2.4. Here we give the equation once again:

$$A_{trf}(c_i) = \sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, c_i) \cdot P_{ass}(x, y, c_i) \cdot \tau(x, y) \cdot \Delta x \cdot \Delta y \quad (3.2)$$

where $P_{cov}(x, y, c_i)$ and $P_{ass}(x, y, c_i)$ are the coverage probability and assignment probability, respectively, see Section 2.3.

Now the condition of traffic load can be expressed as

$$A_{trf}(c_i) \leq T_{max}, \quad \text{for } i = 1, \dots, N. \quad (3.3)$$

3.3.2 Interference probability

The mutual interference between a pair of cells has been considered in Section 2.5. The mutual interference probabilities are used to calculate the channel separation and element s_{ij} of the compatibility matrix indicates the channel separation between cells c_i and c_j . This compatibility constraint must be fulfilled for a feasible channel assignment.

Let $P = (C, \bar{W}, S)$ be a channel assignment problem. Suppose that there exists an assignment that can satisfy the constraints specified by the compatibility matrix S , such that the separation between channels assigned to cell c_i and that assigned to cell c_j is greater than or equal to s_{ij} . Also, let $F_i(x, y)$ be the received power at the location (x, y) from the cell c_i . Then, the C/I ratio of c_i at (x, y) can be calculated as follows, taking into account all the possible interference with the other cells.

$$C_i / I(x, y) = F_i(x, y) - \sum_{j=1, \neq i}^N (F_j(x, y) - L(s_{ij})), \quad (3.4)$$

where $L(s_{ij})$ is the adjacent channel attenuation given by (2.51). In case of co-channels i.e. $s_{ij} = 0$, there is no attenuation, thus $L(0) = 0$.

The interference probability of c_i at the pixel (x, y) is then calculated as

$$P_{int}(x, y, c_i) = 1 - \text{erf}\left(\frac{C_i / I(x, y) - \theta_{C/I}}{\sqrt{2}\sigma}\right) \quad (3.5)$$

where $\theta_{C/I}$ is the required C/I ratio in dB and σ is the standard deviation of the received signal power. $\text{erf}(z)$ is the error function defined in (2.50). Furthermore, the

average interference probability of c_i can be obtained by

$$P_{int}(c_i) = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{int}(x, y, c_i) \cdot P_{cov}(x, y, c_i) \cdot P_{ass}(x, y, c_i) \cdot \Delta x \cdot \Delta y}{A_{area}(c_i)}. \quad (3.6)$$

Again, $P_{cov}(x, y, c_i)$ and $P_{ass}(x, y, c_i)$ are the coverage probability and assignment probability, respectively. $A_{area}(c_i)$ is the area coverage of c_i , as given by (2.37).

Now, the condition of interference probability can be expressed as

$$P_{int}(c_i) \leq \theta_{int}, \quad \text{for } i = 1, \dots, N. \quad (3.7)$$

θ_{int} is an important system performance parameter, which is specified corresponding to the required radio transmission quality.

3.4 GLOBAL CONDITIONS

The design objectives can be addressed by three conditions: coverage rate, spectral cost, and financial cost. These conditions are called global conditions because they are defined on a set of cells $C = \{c_1, c_2, \dots, c_N\}$. An admissible cell set has to comply with three threshold values (θ_{cov} , M_{max} , E_{budget}) corresponding to the global conditions. In this section, we will discuss how to calculate the coverage rate, spectral cost, and financial cost of a given cell set.

3.4.1 Coverage rate

First, the coverage rate is considered, which is defined as the proportion of the area (or traffic) covered by the cells $\{c_1, c_2, \dots, c_N\}$. Although two coverage rates can be distinguished i.e. the area coverage rate and traffic coverage rate, they are highly correlated (even equivalent in case of uniform traffic distribution). Therefore, either the area coverage rate or the traffic coverage rate can be used as the measurement of coverage condition, but not necessary both of them.

A Coverage probability of a cell set

The coverage prediction for the cellular system design has already been discussed in Section 2.3. Because of radio signal characteristics, a mobile unit at a location (x, y) can receive signals from a cell c_i with a coverage probability $P_{cov}(x, y, c_i)$, which specifies the probability of a location being covered by the cell.

In cellular mobile systems, there are overlapping areas between cells. This means that a location could be covered by several cells simultaneously. In order to predict

the coverage of a cell set $C = \{c_1, c_2, \dots, c_N\}$, we introduce a random variable r_i , $i \in [1, N]$, to describe the event that c_i covers (x, y) . Such a random variable is defined as

$$r_i : F_i(x, y) > \theta,$$

where $F_i(x, y)$ is the field strength of the signal received from c_i and θ represents the desired received power threshold. Obviously, the probability of r_i being true, denoted as $P\{r_i\}$, is just the coverage probability of c_i , i.e.

$$P\{r_i\} = P_{cov}(x, y, c_i) \quad (3.8)$$

and

$$P\{\bar{r}_i\} = 1 - P_{cov}(x, y, c_i) \quad (3.9)$$

where \bar{r}_i denotes the complement of r_i . It means that c_i fails to cover (x, y) . If r_i and r_j are unrelated, the following relationship exists:

$$P\{r_i \cup r_j\} = 1 - P\{\bar{r}_i \cap \bar{r}_j\} = 1 - P\{\bar{r}_i\} \cdot P\{\bar{r}_j\}. \quad (3.10)$$

A mobile unit is considered to be served if the field strength of at least one of the cells is above the threshold θ , so the event of a location (x, y) being covered by a cell set C is the union of r_i . Assume that all the random events r_i are independent i.e. r_i and r_j are unrelated, for $i, j \in [1, N]$ and $i \neq j$. The coverage probability of C is then calculated as

$$\begin{aligned} P_{cov}(x, y, C) &= P\left\{\bigcup_{i=1}^N r_i\right\} = 1 - P\left\{\bigcap_{i=1}^N \bar{r}_i\right\} \\ &= 1 - \prod_{i=1}^N (1 - P_{cov}(x, y, c_i)). \end{aligned} \quad (3.11)$$

The above equation means that a location (x, y) is uncovered if all of the cells fail to cover it. Otherwise, it is considered to be covered by the cell set C .

B Area coverage rate

Based on the coverage probability given by (3.11), we can now calculate the area coverage A_{area} as

$$A_{area}(C) = \sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C) \cdot \Delta x \cdot \Delta y. \quad (3.12)$$

There are $m \times n$ pixels and the pixel size is $\Delta x \times \Delta y$. Each pixel is weighted by the

coverage probability $P_{cov}(x, y, C)$. The area coverage rate of C , $R_{area}(C)$, is then calculated as the proportion of the covered area to the total area size, i.e.

$$R_{area}(C) = \frac{A_{area}(C)}{A_S} = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C) \cdot \Delta x \cdot \Delta y}{\sum_{x=1}^m \sum_{y=1}^n \Delta x \cdot \Delta y} \quad (3.13)$$

$$= \frac{\sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C)}{m \cdot n}$$

where A_S represents the service area size.

C Traffic coverage rate

Since the traffic is usually distributed unevenly in the service area, the traffic coverage rate is not equal to the area coverage rate. Assume that the traffic distribution is known and $\tau(x, y)$ is the traffic density at the location (x, y) , then the traffic coverage of the cell set C , $A_{trf}(C)$, can be calculated in a similar way to the area coverage, but weighted by the traffic density:

$$A_{trf}(C) = \sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C) \cdot \tau(x, y) \cdot \Delta x \cdot \Delta y, \quad (3.14)$$

where $P_{cov}(x, y, C)$ is the coverage probability of C , as given by (3.11). Next, we can get the traffic coverage rate R_{trf} as

$$R_{trf}(C) = \frac{A_{trf}(C)}{A_T} = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C) \cdot \tau(x, y) \cdot \Delta x \cdot \Delta y}{\sum_{x=1}^m \sum_{y=1}^n \tau(x, y) \cdot \Delta x \cdot \Delta y} \quad (3.15)$$

where A_T is the total amount of the estimated traffic in the service area. In case that a uniform traffic distribution is assumed, the traffic density $\tau(x, y) \equiv 1$. From (3.15) we have

$$R_{trf}(C) = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C) \cdot \Delta x \cdot \Delta y}{\sum_{x=1}^m \sum_{y=1}^n \Delta x \cdot \Delta y} = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C)}{m \cdot n}. \quad (3.16)$$

We can see that the traffic coverage rate is just equal to the area coverage rate i.e. $R_{trf}(C) = R_{area}(C)$.

Hence, the area coverage rate can be viewed as a special case of the traffic coverage rate. Besides, a satisfactory coverage of the traffic is one of the necessary requirements for designing a cellular system in an efficient way. It is easy to understand that a small region in a metropolis is more important than a vast desert, even though the former is smaller in the area size. Therefore the traffic coverage rate, rather than the area coverage rate, is often used in cell planning.

3.4.2 Spectral cost

The spectral cost is defined as the minimum number of frequency channels required for assigning each cell a sufficient number of channels. This is the lower bound for the channel assignment problem (CAP). In Section 2.6, the CAP is formulated as a generalized graph coloring problem, which is known to be NP-complete [Garey(1979)]. An exact search for the optimal solution is impractical for a large-scale system due to its exponentially growing computation time. Instead of solving the CAP itself, the lower bound for this kind of problem can be calculated analytically. The computations of such a lower bound can be found in [Gamst(1986)] and [Sung(1997)].

First, we recall the definition of the generalized graph coloring problem. Consider the graph $G(V, E)$ obtained by representing each cell c_i by a vertex v_i . The number of channels w_i required by c_i is taken as the weight of v_i . If the vertices v_i and v_j are not allowed to use the same channels i.e. the channel separation $s_{ij} > 0$, they are connected by an edge e_{ij} with label s_{ij} . An M -coloring of G is to assign positive integers $\{1, 2, \dots, M\}$ to the vertices such that each vertex v_i has w_i integers assigned. The difference between the integers assigned to two adjacent vertices must not be less than the edge label. By determining the chromatic number M , we also determine the minimum possible number of channels required.

Definition 3.1 (Complete subset): Let $P = (C, \vec{W}, S)$ be a CAP and k be a positive integer. A subset Q of C is called k -complete if

$$s_{ij} \geq k, \quad \text{for all } c_i, c_j \in Q.$$

Note that a 1-complete subset is equivalent to a clique. The concept of a k -complete subset is just a generalization of a clique.

Theorem 3.2: Let $P = (C, \vec{W}, S)$ be a CAP and Q be a 1-complete subset of C . For each cell $c_i \in Q$, there exist an integer $u \in [1, s_{ii}]$ and a subset R of Q such that $c_i \notin R$ and $s_{ij} \geq u$, for all $c_j \in R$. Furthermore, let

$$w_R = \sum_{c_j \in R} w_j. \quad (3.17)$$

If $s_{ii} - 2u + 1 \leq 0$

$$F_0(P) \geq (w_i - 1) \cdot s_{ii} + 1 + w_R \quad (3.18a)$$

else

$$F_0(P) \geq (w_i - 1) \cdot s_{ii} + 1 + \max(w_R - (w_i - 1)(s_{ii} - 2u + 1), 0). \quad (3.18b)$$

Proof: Define $P' = (C, \bar{W}', S')$ with \bar{W}' having only two nonzero components

$$w'_i = w_i, \quad \text{for } c_i \in Q,$$

and

$$w'_j = w_R, \quad \text{for } c_j \in R.$$

The entries of the compatibility matrix S' are

$$s'_{ii} = s_{ii}, \quad s'_{ij} = s'_{ji} = u, \quad \text{and } s'_{jj} = 1.$$

According to lemmas 4 and 5 in [Gamst(1986)], $F_0(P) \geq F_0(P')$. Now we want to find $F_0(P')$. It is obvious that

$$F_0(P') \geq (w'_i - 1) \cdot s_{ii} + 1$$

since the channels used in c_i must be separated with minimum distance s_{ii} .

For any two channels in c_i , spaced with distance s_{ii} , the number of channels that can be used by cells in R between the gaps is

$$\rho = (s_{ii} - 1) - 2(u - 1) = s_{ii} - 2u + 1.$$

If $\rho \leq 0$, no gap exists and w'_j more channels are needed. Hence (3.18a) is obtained. On the contrary, if $\rho > 0$, there are $\rho(w'_i - 1)$ usable channels left inside all the gaps. If $w'_j \leq \rho(w'_i - 1)$, no additional channel is needed. Otherwise, we need $w'_j - \rho(w'_i - 1)$ more channels. Hence (3.18b) is obtained. ■

For a cell set C , once the cell sites and dimensions are fixed, the triple $P = (C, \bar{W}, S)$ is determined. At the same time, the lower bound of $F_0(P)$ is also determined. The conclusion of Theorem 3.2 can be used to calculate the spectral cost of C , denoted as $M_{req}(C)$. Assume that C has K 1-complete subsets, Q_1, Q_2, \dots, Q_K . Let

$$W_{i,u}(Q_k) = \begin{cases} (w_i - 1) \cdot s_{ii} + 1 + w_R, & s_{ii} - 2u + 1 \leq 0 \\ (w_i - 1) \cdot s_{ii} + 1 + \max(w_R - (w_i - 1)(s_{ii} - 2u + 1), 0), & s_{ii} - 2u + 1 > 0 \end{cases} \quad (3.19)$$

and

$$F_k(P) = \max_{c_i \in Q_k} \max_{u=1, s_{ii}} W_{i,u}(Q_k). \quad (3.20)$$

The spectral cost of C is then calculated as

$$M_{req}(C) = \max_{k=1, K} F_k(P). \quad (3.21)$$

If only the co-channel interference is considered i.e. $s_{ij} \in \{0, 1\}$, the 1-complete subset Q_k is reduced to a clique. There are K cliques in the corresponding graph G . It is clear that $s_{ii} = u = 1$ and therefore $s_{ii} - 2u + 1 \equiv 0$. From (3.17) and (3.19) we have

$$W_{i,u}(Q_k) = w_i + w_R = \sum_{c_i \in Q_k} w_i. \quad (3.22)$$

From the above equation, we see that $W_{i,u}(Q_k)$ has the same value for each cell $c_i \in Q_k$. This value is defined as the weight of a clique, denoted as $W(Q_k)$. In this case, $F_k(P)$ is just equal to the clique weight $W(Q_k)$. As a result, the spectral cost

$$M_{req}(C) = \max_{k=1, K} W(Q_k). \quad (3.23)$$

This means that the maximum clique weight provides a lower bound for a pure co-channel CAP since the channels assigned to the cells in the same clique must be different.

3.4.3 Financial cost

In a mobile communications network, BSs constitute its backbone. The infrastructure cost is therefore mainly determined by the BSs deployed in the system. In contrast to the spectral cost, the cost for building and maintaining the BSs is called financial cost. Usually the cost of a BS comprises two parts: the cost of the site and the cost of the equipment.

In [Hao(1997)], an economic optimization model was presented for designing a low-cost cellular mobile communication system. The cost of BS equipment was considered in this model, but it did not include the cost of BS site. In our optimization model, we define the financial cost for building and maintaining a mobile network as follows:

$$E_{fin}(C) = E_0 + \sum_{c_i \in C} (E_{site}(c_i) + E_{eq}(c_i) + E_m(c_i)) \quad (3.24)$$

where

- $E_{fin}(C)$: Financial cost of the cell set C ;
- E_0 : Fixed cost for installing and maintaining the network;
- $E_{site}(c_i)$: Cost of the BS site of a cell c_i ;
- $E_{eq}(c_i)$: Cost of the equipment of c_i ;
- $E_m(c_i)$: Cost for maintaining the cell c_i .

The system fixed cost E_0 includes the cost for switching centers, BS links, cables, and so on. In some cases, it can be dropped from the financial cost as it does not affect the optimization of the system design.

The site cost $E_{site}(c_i)$ is caused by renting or purchasing a place for building the BS. It is difficult to calculate because such a cost strongly depends on the environment where the BS is located. In a simple model, the site cost is just defined as a function of the morpho class e.g. the urban, suburban, or rural areas. Geographic information systems (GIS) are useful for evaluating the site cost. The more detailed information we have, the more accurate cost we can obtain. In practical system design, site surveys are usually made in order to get an exact cost of sites. Moreover, the site cost also includes the cost of buildings and towers if required.

The equipment installed in the BS mainly consists of the antennas and transceivers, so the equipment cost $E_{eq}(c_i)$ is determined by the antennas and transceivers used. Since the price of antennas and transceivers can be found in the products catalog, it is not difficult to calculate the equipment cost.

After a mobile network is installed, the cost for maintaining the network should be considered. It is clear that the cost $E_m(c_i)$ for running the cell c_i is proportional to its transmitted power. Thus, a linear relationship between the maintenance cost and the transmitted power is assumed in our optimization model.

$$E_m(c_i) = E'_m + \rho_t \cdot P_t, \quad \text{for } P_{min} \leq P_t \leq P_{max} \quad (3.25)$$

where P_t is the transmitted power, P_{min} and P_{max} are the minimum and maximum power, respectively. For GSM systems, $P_{min} = 0.01$ W and $P_{max} = 10$ W [ETSI 05.05(2000)]. E'_m is the basic cost for $P_t = P_{min}$. ρ_t is the cost coefficient with respect to the transmitted power. E'_m and ρ_t must be chosen empirically. The recommended values are: $E'_m = 100$, $\rho_t = 10$.

From (3.24) we can see that the financial cost depends mainly on the number of cells in C . Generally, the financial cost of C becomes higher if it consists of more cells. This is why the number of required cells is often used as the cost function in

some simple optimization models. Obviously (3.24) gives a more accurate evaluation of the financial cost.

3.5 MULTI-OBJECTIVE OPTIMIZATION

Having measurements of system performance is only the first step toward solving the cell planning problem. In order to find an optimal solution, an efficient optimization strategy is needed. One important characteristic of the cell planning problem is that the global conditions are interrelated. Increasing the coverage rate probably causes higher financial cost. Improving the system capacity probably requires more frequency channels. So these conditions need to be considered simultaneously. Usually the spectral cost and financial cost are to be minimized, while maximum coverage rate is desired. Such a problem is classified as multi-objective optimization problem (MOP) as it involves simultaneous optimization of several objectives, which is distinguished from the single objective optimization problem (SOP).

Almost every real-world engineering design or decision making problem can be classified as MOP. Unfortunately, most optimization algorithms confine to SOP, where only one “best” solution could be found based on the single objective. In MOP, because the objectives may be conflicting, there does not exist a unique optimal solution that is a global minimum (or maximum) with respect to all the objectives. Instead, there is a set of optimal solutions which represent tradeoffs between the conflicting objectives. These solutions are optimal in the wider sense that no other solutions in the search space are superior to them when all objectives are considered. They are called Pareto optimal solutions, named after the economist and sociologist Vilfredo Pareto (1848-1923). The concept of Pareto optimality will be given later in this section.

This section is an introduction of multi-objective optimization, which consists of two parts. In the first part, the principles of multi-objective optimization are outlined and basic concepts are defined. The second part gives a review of classical methods for solving MOPs and in particular their potential disadvantages. More discussions about multi-objective optimization can be found in [Fonseca(1995a), Deb(1999), and Zitzler(1999a)].

3.5.1 Basic concepts and definitions

Multi-objective optimization problems are characterized by several incommensurable and often conflicting measures of performance i.e. objectives. Formally, this can be defined as follows:

Definition 3.2 (Multi-objective optimization problem): A general MOP includes a set of m objective functions that map a decision vector \bar{x} to an objective vector \bar{y} .

The optimization goal is to

$$\text{minimize/maximize } \bar{y} = f(\bar{x}) = (f_1(\bar{x}), f_2(\bar{x}), \dots, f_m(\bar{x}))$$

where $\bar{x} = (x_1, x_2, \dots, x_n) \in X$,

$$\bar{y} = (y_1, y_2, \dots, y_m) \in Y.$$

X is called the decision space, and Y is the objective space.

Without loss of generality, a maximization problem is assumed here. For minimization or mixed maximization/minimization problems the definitions presented in this section are similar. In single objective optimization the optimal solution is usually clearly defined, as the solutions can be totally ordered according to the objective function f . For any two solutions $\bar{x}_1, \bar{x}_2 \in X$, either $f(\bar{x}_1) \geq f(\bar{x}_2)$ or $f(\bar{x}_2) \geq f(\bar{x}_1)$. The solution that gives the maximum value of f is the optimal solution. However, when several objectives are involved, the situation changes. The solutions are, in general, not totally ordered, but partially ordered. Let us consider the cell planning problem. For the purpose of illustration, we assume that only two objectives are to be maximized, namely, the coverage rate (f_1) and cheapness (f_2) - the inverse of financial cost. The 2D objective space is shown in Figure 3.1. Obviously, the solution represented by point B is better than the solution represented by point C because B provides higher coverage at lower cost. It would even be preferable if it would only improve one objective, as is the case for C and D : despite equal cost, C achieves better coverage than D . In order to express this situation mathematically, the relations $=$, \geq , and $>$ are extended to objective vectors by analogy to the single objective case.

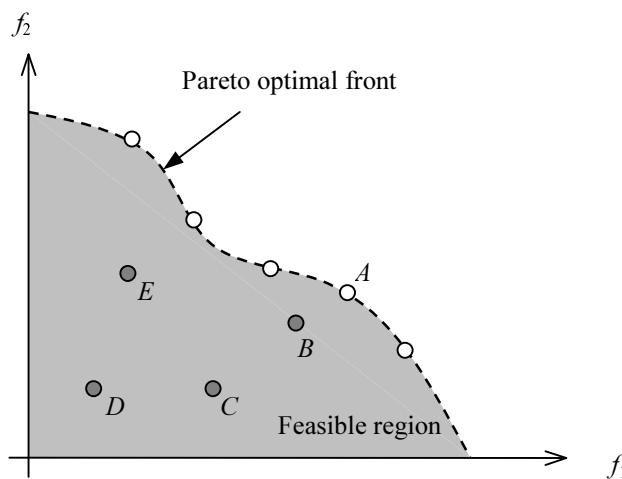


Figure 3.1 Illustrative example of Pareto optimality in a 2D objective space.

Definition 3.3: For any two objective vectors \bar{y}_1 and \bar{y}_2 ,

$$\bar{y}_1 = \bar{y}_2 \quad \text{iff} \quad y_{1,i} = y_{2,i}, \quad \text{for all } i \in [1, m];$$

$$\bar{y}_1 \geq \bar{y}_2 \quad \text{iff} \quad y_{1,i} \geq y_{2,i}, \quad \text{for all } i \in [1, m];$$

$$\bar{y}_1 > \bar{y}_2 \quad \text{iff} \quad \bar{y}_1 \geq \bar{y}_2, \text{ and } \exists i \in [1, m], y_{1,i} > y_{2,i}.$$

The relations \leq and $<$ can be similarly defined.

Using this definition, it holds that $B > C$, $C > D$, and, as a consequence, $B > D$. However, when comparing B and E , neither can be said to be superior, since $B \not\geq E$ and $E \not\geq B$. Although the solution associated with E is cheaper, it provides lower coverage than the solution represented by B . Therefore, two decision vectors \bar{x}_1 , \bar{x}_2 can have three possibilities with MOPs regarding the relation “ \geq ” (in contrast to two with SOPs): $f(\bar{x}_1) \geq f(\bar{x}_2)$, $f(\bar{x}_2) \geq f(\bar{x}_1)$, or $f(\bar{x}_1) \not\geq f(\bar{x}_2) \wedge f(\bar{x}_2) \not\geq f(\bar{x}_1)$. In order to classify the different situations, the following symbols and terms are used.

Definition 3.4 (Pareto dominance): For any two decision vectors \bar{x}_1 and \bar{x}_2 ,

$$\bar{x}_1 \succ \bar{x}_2 \quad (\bar{x}_1 \text{ dominates } \bar{x}_2) \quad \text{iff} \quad f(\bar{x}_1) > f(\bar{x}_2)$$

$$\bar{x}_1 \succcurlyeq \bar{x}_2 \quad (\bar{x}_1 \text{ partially dominates } \bar{x}_2) \quad \text{iff} \quad f(\bar{x}_1) \geq f(\bar{x}_2)$$

$$\bar{x}_1 \sim \bar{x}_2 \quad (\bar{x}_1 \text{ is indifferent to } \bar{x}_2) \quad \text{iff} \quad f(\bar{x}_1) \not\geq f(\bar{x}_2) \wedge f(\bar{x}_2) \not\geq f(\bar{x}_1)$$

The definitions for a minimization problem (\prec , \preceq , \sim) are analogous.

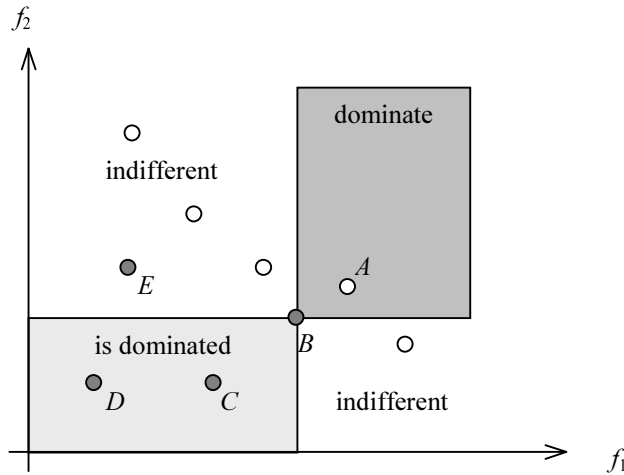


Figure 3.2 Possible relations between solutions in multi-objective space.

In Figure 3.2, the light gray rectangle encapsulates the region in objective space that is dominated by the decision vector represented by B . The dark gray rectangle

contains the objective vectors whose corresponding decision vectors dominate the solution associated with B . All solutions for which the resulting objective vector is in neither rectangle are indifferent to the solution represented by B .

Based on the concept of Pareto dominance, the optimality criterion for MOPs can be introduced. Referring to Figure 3.1, A is unique among B , C , D , and E as its corresponding decision vector \bar{a} is not dominated by any other decision vector. That means, \bar{a} is optimal in the sense that it can not be improved in any objective without causing a degradation in at least one other objective. Such solutions are known as Pareto optimal. They are also called non-dominated or non-inferior solutions.

Definition 3.5 (Pareto optimality): A decision vector \bar{a} is said to be non-dominated regarding a set $A \subseteq X$ iff

$$\nexists \bar{x} \in A : \bar{x} \succ \bar{a}$$

Furthermore, \bar{a} is said to be Pareto optimal iff \bar{a} is non-dominated regarding X .

In Figure 3.1 the white points represent Pareto optimal solutions and are tradeoffs between the conflicting objectives. Improvement in any one of the objectives may impair the others, and vice versa. None of these solutions can be recognized as better than the others unless preference information is included (e.g. a ranking of the objectives). The entirety of all Pareto optimal solutions is called the Pareto optimal set. The corresponding objective vectors form the Pareto optimal front or surface.

Definition 3.6 (Non-dominated sets and fronts): Let $A \subseteq X$. The non-dominated set regarding A , represented by A_p , is defined as

$$A_p = \{ \bar{a} \in A \mid \bar{a} \text{ is non-dominated regarding } A \}.$$

The corresponding set of objective vectors is defined as the non-dominated front regarding A , given by

$$F(A_p) = \{ f(\bar{a}) \mid \bar{a} \in A_p \}.$$

Furthermore, the non-dominated set regarding X , represented by X_p , is called the Pareto optimal set, and $Y_p = F(X_p)$ is the Pareto optimal front.

The Pareto optimal set comprises the global optimal solutions. However, as with SOPs there may also exist local optima which constitute a non-dominated set within a certain neighborhood. Moreover, MOPs may have uncountable Pareto optimal solutions. It is impossible, and also unnecessary, to find all of them. What is interesting is an optimization method that can find a Pareto optimal set representing alternative tradeoffs between the objectives, such that a suitable compromise solution can be chosen.

3.5.2 Methods for solving multi-objective problems

From the above discussions, we observe that there are primarily three goals that a multi-objective optimization algorithm must achieve:

- The resulting non-dominated front should approximate the global Pareto optimal front, although it may not be exactly a global optimum.
- The obtained solutions should be distributed uniformly so as to represent the entire Pareto optimal set.
- The spread of the resulting non-dominated front should be maximized, such that for each objective as many as possible values could be covered by the obtained solutions.

Traditional approaches to MOPs usually transform the initial multi-objective problem into a mathematically manageable format, such that well-studied algorithms for SOPs can be used. Among the traditional approaches, the weighting method and constraint method are most popular. They will be briefly reviewed in the following.

A Weighting method

The weighting method converts the original MOP into an SOP by forming a linear combination of the objectives:

$$\text{maximize } y = f(\bar{x}) = w_1 \cdot f_1(\bar{x}) + w_2 \cdot f_2(\bar{x}) + \dots + w_m \cdot f_m(\bar{x}),$$

where $w_i, i \in [1, m]$, is the weight for the i -th objective f_i . Usually, $0 < w_i < 1$, and

$$\sum_{i=1}^m w_i = 1.$$

The weight w_i indicates the relative importance of the objective f_i and it must be specified for each of the m objectives beforehand. Solving the above SOP for a certain number of different weight combinations yields a set of solutions which approximates the Pareto optimal set.

[Deb(1999)] and [Zitzler(1999a)] mentioned the disadvantages of the weighting method. The main disadvantage is that it can not find Pareto optimal solutions in MOPs with a non-convex Pareto optimal front. This is illustrated in Figure 3.3 with a 2D objective space. For fixed weights w_1, w_2 , a solution \bar{x} is sought to maximize $y = w_1 \cdot f_1(\bar{x}) + w_2 \cdot f_2(\bar{x})$. This equation can be reformulated as

$$f_2(\bar{x}) = -\frac{w_1}{w_2} f_1(\bar{x}) + \frac{y}{w_2},$$

which defines a line with slope $-w_1/w_2$ and intercept y/w_2 in objective space (solid line in Figure 3.3). Graphically, the optimization process corresponds to moving this line upwards until it is tangent with the feasible region. The tangent points (here A and D) are found to be optimal solutions. One can now imagine finding other solutions such as the point E by changing the slope of the line, namely choosing different values of w_1 and w_2 . However, the points on the non-convex surface e.g. B and C , will never be found.

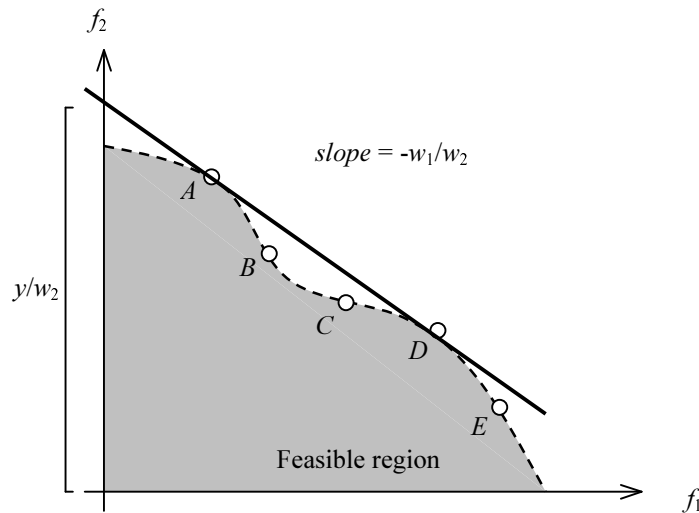


Figure 3.3 Graphical interpretation of the weighting method for solving multi-objective problems.

B Constraint method

Another technique of handling MOP is to optimize one objective (possibly the most important one) and to treat the remaining $m - 1$ objectives as constraints. Thus the MOP is converted into and subsequently solved as an SOP:

$$\text{maximize } y = f(\bar{x}) = f_j(\bar{x})$$

subject to

$$f_i(\bar{x}) \geq \theta_i, \quad i = 1, \dots, m, \text{ and } i \neq j.$$

The lower bounds, θ_i , should be chosen by the optimizer for each of the $m - 1$ objectives. The Pareto optimal set can be found by solving the above constraint SOP with different values of θ_i .

Figure 3.4 shows that the constraint method is able to find optimal solutions associated with non-convex parts of the Pareto optimal front. Setting $j = 1$ and $\theta_2 = l$

(solid line) makes the solution represented by D infeasible regarding the given constraint, while the solution related to C maximizes the objective f_1 . The non-convex Pareto optimal solution is obtained. However, there also exists a difficulty with this method. As depicted in Figure 3.4, if the lower bounds are not chosen appropriately (e.g. $\theta_2 = l'$), no feasible solution can be obtained for the corresponding SOP. In order to avoid this situation, a suitable range of values for the θ_i has to be known a priori.

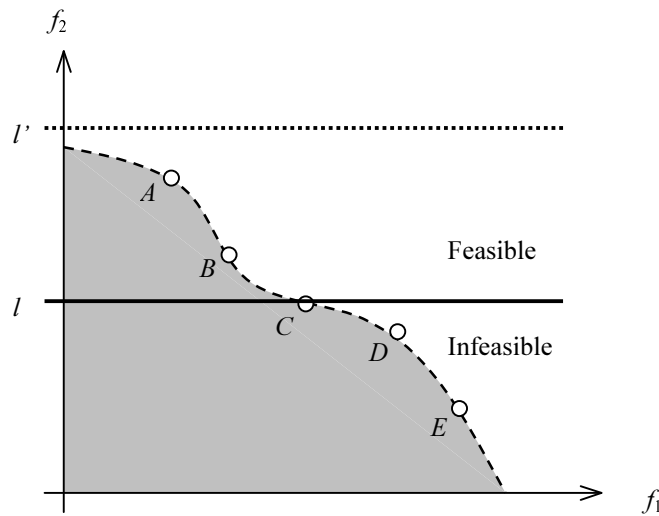


Figure 3.4 The constraint method for solving multi-objective problems.

C Need for better algorithms

Although there exist a few other methods such as goal programming [Steuer(1986)], min-max approach [Koski(1984)], and others, all these methods have some difficulties with MOPs. Based on the above discussions on the weighting method and constraint method, it can be shown that

- Some kind of knowledge about the application problem is required, which may not be available.
- Some techniques, e.g., the weighting method, are sensitive to the shape of the Pareto optimal front.

In spite of these difficulties, classical methods are still widely used. What makes them popular may be attributed to the fact that well-studied algorithms for SOPs can be used, such as random search algorithms, simulated annealing, tabu search, and Monte Carlo methods. However, in order to find multiple Pareto optimal solutions, a single objective algorithm has to be executed several times. As stated in [Deb(1999)],

“The most profound difficulty with all these (classical) methods is that all need to be applied many times, hopefully finding one different Pareto optimal solution each time.”

The above difficulties can be handled by using an evolutionary search algorithm. Evolutionary algorithm (EA) is a generic term for several stochastic search methods which computationally simulate the natural evolutionary process. EA embodies the techniques of genetic algorithms (GAs), evolution strategies (ESs), and evolutionary programming (EP). These techniques are loosely based on natural evolution and the Darwinian concept of “survival of the fittest”.

The potential of GAs to become a powerful method for multi-objective optimization has long been recognized [Schaffer(1985)]. GAs seem particularly suitable to solve MOPs because they deal simultaneously with a set of possible solutions, namely, population. This inherent characteristic allows them to find a set of Pareto optimal solutions in a single optimization run, instead of having to perform a series of separate runs as in the case of the classical methods. In addition, GAs are less susceptible to the shape or continuity of the Pareto optimal front. Therefore, GA-based approaches have recently become established as an alternative to classical methods.

The next chapter is devoted to GAs, where the principles of GAs, as well as the multi-objective GAs, will be discussed in detail.

3.6 SUMMARY

In this chapter, we have analyzed the cell planning problem from the viewpoint of system performance optimization. First, a general description of the problem has been given. Four kinds of problems i.e. cell selection, cell dimensioning, automatic base station placement and dimensioning, and growth planning, arise from the cellular system design practice. Although these problems are different in the given conditions, their tasks are the same i.e. to design an optimal cellular system which can guarantee the required performance goals while minimizing the system costs.

The system performance criteria are addressed by the local conditions and global conditions. The local conditions, including the traffic load and interference probability, are defined on each single cell, while the global conditions are defined on the whole system (a set of cells), including the coverage rate, spectral cost, and financial cost. The global conditions are used as optimization objectives.

Cell planning is a typical multi-objective optimization problem as several objectives need to be considered simultaneously. The principles of multi-objective optimization have been outlined in Section 3.5. This section has also reviewed the classical methods for solving multi-objective problems. The insufficiency of classical

methods gives rise to the research on genetic algorithms for multi-objective optimization.

Chapter 4

Genetic algorithms

This chapter seeks to draw attention to genetic algorithms (GAs) for solving cell planning problems. It begins with a brief introduction to GAs, after which a basic structure of GAs is described. In order to tailor a GA for a particular problem, the following issues need to be considered:

- Representing solutions.
- Defining the fitness of individuals (solutions).
- Selecting individuals for producing new solutions.
- Designing genetic operators.
- Reinserting new solutions into the population.

These design aspects will be discussed in Section 4.2. Usually, GAs do not directly deal with decision variables. Each solution is encoded as a chromosome-like bit-string. However, some researchers argue that such binary GAs have some drawbacks when the decision variables are naturally continuous [Michalewicz(1996), Haupt (1998)]. The direct manipulation of real-valued chromosomes will be considered in Section 4.3.

As mentioned in Chapter 3, GAs were recognized to be well qualified to tackle multi-objective optimization problems. Currently, there is rapidly growing interest in the area of multi-objective genetic algorithms (MOGAs). Section 4.4 will be concentrated on MOGAs. First, the key elements of GAs for multi-objective search will be discussed. Then, an algorithm based on non-dominated sorting and elitism techniques will be introduced. Finally, two tasks (search and decision making) of multi-objective optimization will also be discussed.

4.1 A BASIC GENETIC ALGORITHM

GAs are stochastic search algorithms inspired by the genetic mechanisms of natural species evolution [Holland(1975)]. During the course of evolution, natural population evolves according to the principle of natural selection. Individuals that are adapted to the environment have a better chance of reproducing, while those that are less fit are eliminated. This means that the “genes” from the highly fit individuals will spread to an increasing number of individuals in successive generations. The combination of

good characteristics from highly fit ancestors may produce the offspring with better adaptability. Consequently, species evolve to become more and more adaptable to the environment.

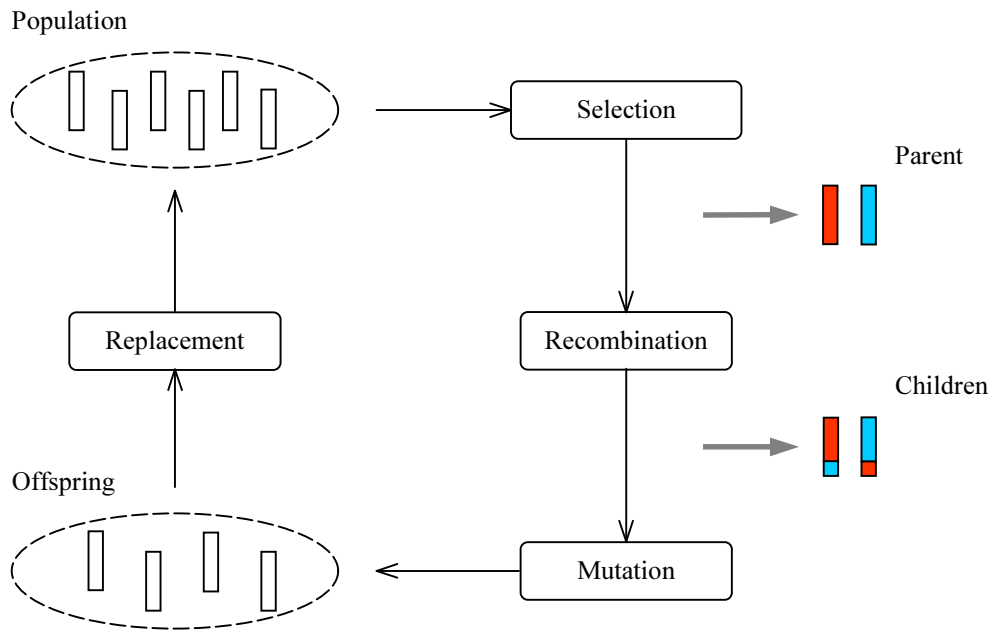


Figure 4.1 Main procedures of a genetic algorithm.

In GAs, four main procedures can be identified (see Figure 4.1). Usually, individuals are encoded as chromosome-like bit-strings. GAs simulate the evolutionary process by generating an initial population of individuals and iteratively applying genetic operators i.e. selection, recombination, and mutation, in each reproductive cycle. First, selection is executed to select a couple of individuals based on their fitness values. This may be perceived as an allocation of reproductive opportunities: the higher the fitness value of an individual, the more likely it is to be selected. On each pair of selected individuals (parents), recombination is applied with probability p_r . In this procedure, the parents exchange their genetic information to generate two new individuals (children). The selection and recombination operators compose the cooperation step of the GA. Afterwards, a mutation operator is applied with a probability of p_m on each child by randomly altering some genes in the child's chromosome. This procedure is the self-adaptation step of the GA, and prevents premature convergence of the population. The new generated children are put into an intermediate population, called offspring. The reproduction process is repeated until the offspring is full i.e. a predefined number of children have been produced. The offspring can either replace all, or a part of the individuals in the initial population. The so-called incremental replacement is commonly used, in which the less fit individuals are replaced. The execution terminates when the population is convergent or a given number of generations have been run through. More detailed discussion of

the various aspects of GAs can be found in [Goldberg(1989), Fogel(1995), and Michalewicz(1996)].

In the following, the basic structure of GAs is summarized. This basic GA is a necessary foundation for building more advanced GAs, and consists of six procedures as described in Algorithm 4.1, where the population and offspring are represented by P and P' , respectively.

Algorithm 4.1 (Basic genetic algorithm)

Input: T – The given number of generations
 K – Population size
 K' – Offspring size
 p_r – Recombination probability
 p_m – Mutation probability

Output: The best solution in P

Step 1 (Initialization):

Set $t = 0$ and $P = \emptyset$;
 For $i = 1, \dots, K$ do
 1) Create an individual \bar{a}_i randomly or heuristically;
 2) $P = P \cup \{\bar{a}_i\}$.

Step 2 (Evaluation):

For each individual $\bar{a}_i \in P$ do
 1) Calculate the objective function $f(\bar{a}_i)$;
 2) Assign a fitness value z_i to \bar{a}_i .

Step 3 (Reproduction):

Set $P' = \emptyset$;
 For $i = 1, \dots, K'$ do
 1) Select two individuals $\bar{a}, \bar{b} \in P$ according to a given scheme and based on the fitness values;
 2) Recombine \bar{a} and \bar{b} with the probability p_r , resulting a child \bar{c}_i ;
 3) Mutate \bar{c}_i with the probability p_m ;
 4) $P' = P' \cup \{\bar{c}_i\}$.

Step 4 (Replacement):

Replace the less fit individuals in population P with the children in offspring P' .

Step 5 (Termination):

$t = t + 1$;

If $t \geq T$ or another stopping criterion is satisfied then terminate,
else go to Step 2.

It must be emphasized that this algorithm does not reflect a GA in its most general form as, for example, the population size need not be restricted and recombination can also involve more than two parents. Moreover, a large number of selection, crossover, and mutation operators have been proposed for different representations, applications, and so forth. These aspects will be discussed in the next section.

4.2 DESIGN ASPECTS OF GENETIC ALGORITHMS

The mechanism of GAs is neither governed by the use of differential equations nor does it behave like a continuous function. It possesses the unique ability to search and optimize a solution of a complex system, where other mathematical oriented techniques may have failed to compile the necessary design specifications. This property makes GAs be able to solve a number of engineering problems.

In order to design a GA for solving an application problem, the following questions are inevitable:

- How to represent solutions?
- How to evaluate the solutions?
- How to select individuals for mating?
- How to produce offspring?
- How to insert the offspring into the population?

These aspects will be discussed next.

4.2.1 Representation

GAs do not directly deal with the decision variables of the optimization problem. They work with chromosome-like codes which represent the solutions. Thus, the first issue in a GA application is how to represent solutions of the problem under study i.e. how to use a data structure that is appropriate for the problem. The method of representation has a major impact on the performance of the GA. Different representation schemes might cause different performance in terms of accuracy and computation time.

The number of possible representation schemes is endless. In the GALib [Wall(1999)] several kinds of genomes are implemented, such as bit-string, vector, matrix, and tree. Among them, two representations are commonly used in GAs. The

preferred one is the bit-string representation. The reason for this method being popular is that the binary alphabet offers the maximum number of schemata per bit compared to other coding techniques [Michalewicz(1996)]. The second representation method is to use a vector of integers or real numbers, with each integer or real number representing a single parameter of the decision vector.

[Zitzler(1999a)] distinguished the decision vectors and their representation by specifying two state spaces: an individual space \mathbf{P} and a decision space \mathbf{X} , as shown in Figure 4.2. Given a decision vector $\bar{x}_i \in \mathbf{X}$, the mapping function g encodes it based on an appropriate representation scheme to get the corresponding individual $\bar{a}_i \in \mathbf{P}$. In brief, $\bar{a}_i = g(\bar{x}_i)$. The individual \bar{a}_i is then manipulated by the GA.

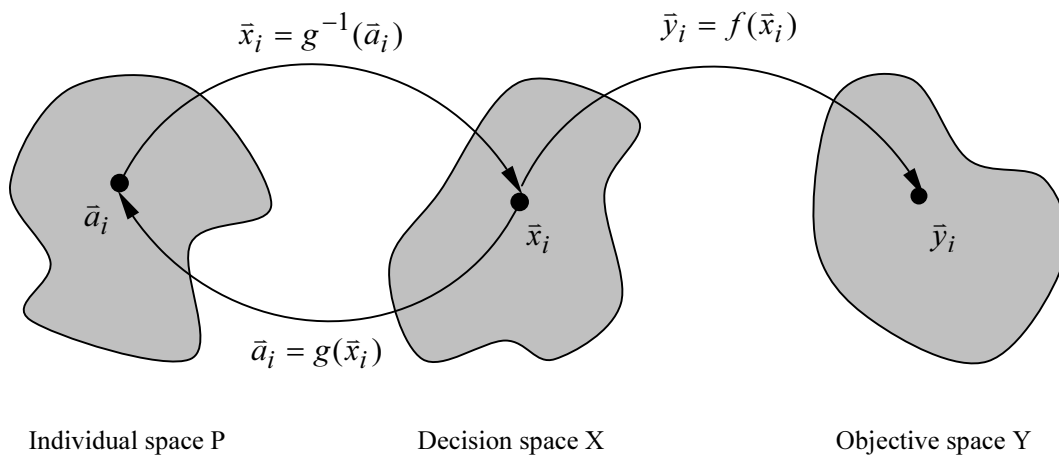


Figure 4.2 Relations between individual space, decision space, and objective space.

4.2.2 Evaluation and fitness assignment

GAs are more attractive than gradient search methods because they do not require complicated differential equations or a smooth search space. The GA needs only a single measure of how good a solution is compared to the other solutions. The objective function (also called cost function for a minimization problem) provides such a measure: given a solution, how good is it relative to others?

Since the objective function is defined on the decision vector, each individual must be decoded before its objective function can be calculated. This situation is illustrated in Figure 4.2. First, the inverse mapping function g^{-1} derives the decision vector \bar{x}_i from the individual \bar{a}_i i.e. $\bar{x}_i = g^{-1}(\bar{a}_i)$. Then, the objective function f is applied to \bar{x}_i such that $\bar{y}_i = f(\bar{x}_i)$. As a result, the objective vector of the corresponding solution is obtained.

The quality of an individual is evaluated by a scalar value, the so-called fitness. The fitness can be viewed as the number of offspring an individual is expected to produce through selection. It is fundamentally different from the objective vector (even in single objective optimization), which is intrinsic to the problem and reflects the GA's optimality condition. On the other hand, the fitness assignment relates to the selection process, being a customizable part of the optimization strategy. There are essentially two types of fitness assignment approaches: scaling and ranking [Goldberg(1989), Fonseca(1995a)].

A Scaling

Scaling is the more traditional approach, only applicable to a single objective. The fitness is computed as a monotonic, possibly non-linear, function of the objective score. Caution must be taken to ensure non-negative fitness for all individuals, while giving the best individual in the population a dominated advantage over the others.

Three scaling methods are suggested in [Goldberg(1989)]: linear scaling, power law scaling, and sigma scaling. As the name implies, linear scaling simply maps an objective score to a scaled fitness using a linear function. In power law scaling, the scaled fitness is taken as some specified power of the objective score. Sigma scaling is named from the population standard deviation (σ), and uses population variance information to preprocess the objective scores prior to scaling.

B Ranking

With scaling, an individual much stronger than all the others may be assigned a very large relative fitness and, through selection, rapidly dominate the population. However, the advantage of the best individual over the rest of the population will become trifling if most individuals perform more or less equally well, and the search will degenerate into an aimless walk. To circumvent this scaling problem, a ranking method could be used [Fonseca(1995a)].

Ranking is performed by sorting the population based on the objective scores, and subsequently assigning fitness to each individual according to its position (or rank) in the population. Note that here the fitness is related to the whole population, whereas with scaling and some other techniques, an individual's fitness value is calculated independently of other individuals. Hence, the best individual is always assigned the same relative fitness. The would-be "super" individuals can never reproduce excessively. On the other hand, when all individuals perform almost equally well, the best individual is still unequivocally preferred to the rest.

Rank-based fitness assignment is characterized by the rank-to-fitness mapping, which is usually chosen to be linear or exponential. Assume that the population size is K , and the best individual is ranked 0 and the worst is $K - 1$. Linear mapping can be

written as follows:

$$z_i = Z \cdot \left(1 - \frac{r_i}{K-1}\right), \quad (4.1)$$

where z_i is the fitness and r_i is the rank. Z is the relative fitness desired for the best individual and $Z > 0$ such that $z_i \in [0, Z]$. Exponential mapping derives the fitness value from the individual's rank using a function of the form:

$$z_i = Z \cdot \rho^{r_i}, \quad 0 < \rho < 1. \quad (4.2)$$

Again, Z is the desired fitness of the best individual in the population.

The ranking method can also be applied to multi-objective optimization. Fitness assignment regarding multiple objectives will be discussed in Section 4.4. The relative fitness z_i , not the objective vector \bar{y}_i , is used in the selection process.

4.2.3 Selection methods

To generate good offspring, a good parent selection method is necessary. Selection is a process used for choosing individuals to participate in reproduction. The chance of selecting an individual for mating should be directly proportional to its fitness so that more children are produced by highly fit individuals. The selection procedure determines which individuals actually influence the production of the next generation and, consequently, is an essential part of the search strategy. It has a significant influence on driving the search towards a promising area and finding good solutions in a short time. On the other hand, the diversity of the population must be maintained to avoid premature convergence and to reach the global optimum.

In GAs, selection is usually performed stochastically. The best known method is roulette wheel selection [Goldberg(1989)]. This method consists of a sequence of independent selection trials where the probability of each individual being selected is proportional to its fitness. The so-called reproductive probability is defined as

$$p_i = \frac{z_i}{\sum_{j=1}^K z_j} \quad (4.3)$$

where z_i is the fitness of individual \bar{a}_i , and K is the population size.

There is another well-known selection method, called tournament selection [Hancock(1994)]. In this selection k individuals are randomly picked from the population at a time (k -tournament), and the individuals are compared to each other.

The one with the best fitness is declared the winner of that tournament and selected. Tournament selection is particularly convenient when it is easier to compare individuals than to compute their absolute performance, as is often the case in genetic programming.

However, tournament selections are criticized for their strong tendency to select individuals with better fitness. The worse ones have no chance to be selected. This may result in premature convergence. Thus, we use a probabilistic binary tournament selection. Each time two individuals \bar{a}_1 and \bar{a}_2 are randomly chosen to join a tournament. The probability for \bar{a}_1 winning the tournament is $p = z_1/(z_1 + z_2)$, and the probability for \bar{a}_2 is $1 - p$. In this method, both individuals have a chance to be selected, depending on their fitness.

4.2.4 Genetic operators

Repeated selection from the same population would produce nothing better than multiple copies of the best individual originally in it. For improvement to be able to occur, some novelty must be introduced into the population between selection steps. Genetic operators modify the selected individuals by manipulating their chromosomes. The genetic operators can be classified into two main categories - recombination and mutation.

The implementation of recombination and mutation necessarily depends on the underlying genotypic representation. Further, the performance of the same operator generally depends on the problem under consideration. A great deal of research has been put into identifying the best genetic operators for various classes of problems. The development of the edge recombination operator for the travelling salesman problem (TSP) is a good example [Davis(1991)].

A Recombination

This operation causes individuals to exchange genetic information with one another, in pairs or larger groups. The main idea is to generate new individuals by integrating good characteristics of highly fit individuals. An effectively designed recombination can significantly accelerate the search process, especially in the case where decision variables are loosely coupled.

A typical recombination operator for binary and other string-like chromosomes is one-point crossover, whereby two individuals exchange a portion (right or left) of their chromosomes to produce offspring, as illustrated in Figure 4.3. The crossing point is selected at random. Other recombination operators commonly used with string-like chromosomes are: two-point crossover where two crossing points are selected instead of one, and uniform crossover where each gene is exchanged

independently, based on a randomly generated crossover mask.

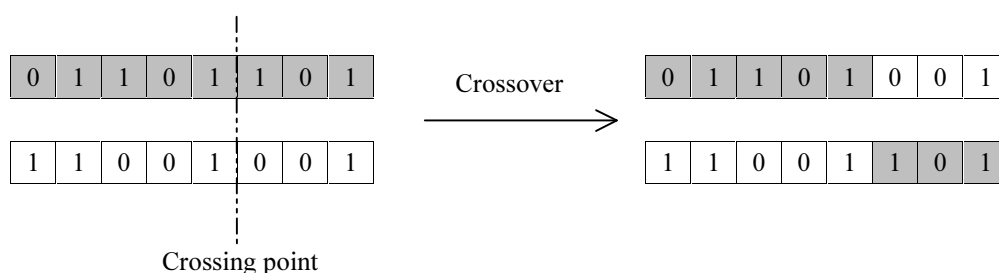


Figure 4.3 One-point crossover for bit-string chromosomes.

A so-called fusion crossover is suggested in [Hifi(1997)], which takes account of both the structure and relative fitness of the parents. Assume that the parents are represented by two vectors \bar{a}_1 and \bar{a}_2 , their fitness values are z_1 and z_2 , respectively. The child is represented by \bar{c} . For each gene c_i of the child, the fusion crossover does

1. If $a_{1,i} = a_{2,i}$ then set $c_i = a_{1,i}$;
2. If $a_{1,i} \neq a_{2,i}$ then set $c_i = a_{1,i}$ with a probability $p = z_1/(z_1 + z_2)$, or $c_i = a_{2,i}$ with probability $1 - p$.

The advantage of this crossover is that it is able to create more different children when the parent individuals are very close in structure than the frequently used one-point or two-point crossover operators, so as to maintain the necessary diversity of the population.

B Mutation

The mutation operator is applied for each new generated child after crossover. It works by randomly altering some genes in the child's chromosome. Usually, only a small part of the chromosome is changed by mutation, causing offspring to inherit most of the features of the parents. Mutation is recognized to be an important, active search operator. It forces the algorithm to search new areas. Eventually, it helps GAs avoid premature convergence and find the global optimum.

Mutation means different things for different chromosome types. For instance, a typical mutation for binary chromosomes (bit-strings) flips each bit independently with a given probability (i.e., mutation probability), while a mutation for real number chromosomes would alter the values of some genes. Mutation operators for real-valued genetic algorithms will be discussed later in Section 4.3.

Selecting the various GA parameters, such as the recombination rate and mutation rate, is still more of an art than a science. A typical value between 0.6 and

1.0 is used as the probability of recombination (p_r), while mutation rates (p_m) are normally below 0.1. For string-like chromosomes, p_m is usually set to about the inverse of the chromosome length. This has been shown theoretically to be optimal (under restrictions) [Bäck(1993)], and observed experimentally to perform well on a wide range of problems.

4.2.5 Replacement scheme

After a predefined number of children have been produced through the genetic operators, a replacement operation is required in order to modify the old population with the new generation. Replacement can be implemented in a variety of ways. In the simplest GA, for example, the parental population is completely replaced with the offspring. This is known as generational replacement.

There is also interest in GAs where children have the chance to compete with some of the parent individuals. In the case known as incremental replacement, a small number of offspring (smaller than the population size) is generated. These offspring are then evaluated, and possibly inserted into the population by replacing

- their own parents,
- the worst individuals in the parental population,
- the oldest individuals in the parental population,
- random individuals in the parental population.

Actual replacement may occur

- unconditionally,
- only if the offspring are better than the individuals they are to replace,
- probabilistically, depending on how good the offspring are compared to the individuals they are to replace.

Besides, when replacing an individual, care must be taken to prevent a duplicate individual from entering the population. A duplicate individual is identical to some member of the population. It is undesirable because the population with duplicate individuals could come to consist of all identical individuals, thus severely limiting the GA's ability to search new areas to find the global optimum.

4.3 REAL-VALUED GENETIC ALGORITHMS

Originally, the concept of GAs was based on binary chromosomes. Regardless of the problem, the candidate solutions would be mapped into bit-strings and then appropriate genetic operators are applied. Such binary GAs have some drawbacks when applied to parameter optimization problems with variables over continuous domains e.g. the cell dimensioning problem. In such a problem, each parameter

requires many bits to fully represent it. If the number of parameters is large, the chromosome length grows quite quickly, making the GA perform poorly. Thus, it is more logical to represent the continuous parameter with real numbers.

When the parameters are naturally discrete, the binary representation fits nicely. However, when the parameters are continuous, a vector of real numbers should be used instead of bit-strings. For continuous parameter optimization problems, real-valued GAs have the following advantages compared to binary GAs, as observed in [Haupt(1998)]:

- The real-valued GA has a more accurate representation of the continuous parameter, since the binary GA has its precision limited by the chromosome length.
- The real-valued GA evaluates the solutions more accurately as the objective function is calculated with real parameters.
- The real-valued GA requires less storage than the binary GA because a single floating-point variable is used instead of a large number of bits.
- The real-valued GA simplifies the problem of coding and decoding of individuals and decision vectors.

4.3.1 Genetic operators for real-valued chromosomes

Real-valued GAs have the same workflow as binary GAs, as described in Algorithm 4.1. What makes them different is the representation method where parameters are no longer represented by bits of 0s and 1s, but instead by real numbers over whatever range is deemed appropriate. This difference makes the genetic operators (recombination and mutation) designed for bit-strings unable to be used directly for real number vectors. In this section, we will present some genetic operators that are particularly designed for real-valued chromosomes. More discussions about real-valued genetic operators can be found in [Michalewicz(1996), Haupt(1998), and Cvetković(2000)].

A Recombination

Assume that the parents are represented by two vectors \bar{a}_1 and \bar{a}_2 . There are several recombination operators for producing the child, which is represented by a vector \bar{c} .

Discrete recombination: In this case, the child's gene $c_i \in \{a_{1,i}, a_{2,i}\}$. This corresponds to the uniform crossover in the binary case. The probability of choosing $a_{1,i}$ or $a_{2,i}$ could also be specified, as in the fusion crossover described in Subsection 4.2.4.

The problem with this recombination is that no new information can be introduced. The parameter values of the initial population are propagated to the

next generation, only in different combinations.

Linear recombination: The linear recombination remedies the above problem by blending the values of the parent's elements. The child is generated by applying a linear function in the form

$$c_i = \beta \cdot a_{1,i} + (1 - \beta) \cdot a_{2,i}, \quad 0 \leq \beta \leq 1, \quad (4.4)$$

where β is a random number. It could be fixed for all the elements or vary with different elements. Choosing which elements to blend is the next issue. Sometimes, this linear combination is done for all elements to the left, or right, of some crossover point.

The linear recombination is also called intermediate recombination as the values of new parameters are chosen between the values bracketed by the parents. It does not allow introduction of values beyond the extremes already represented in the population. To do this requires an extrapolating method, making $-d \leq \beta \leq 1 + d$. However, the parameter d must be carefully chosen to avoid any value outside the range $[L_i, U_i]$, where L_i, U_i represent the lower bound and upper bound for the element c_i , respectively.

Simulated binary crossover (SBX): SBX produces two children, represented by \bar{c}_1 and \bar{c}_2 , from the parents \bar{a}_1 and \bar{a}_2 , where the distance between the children is proportional to that of the parents. For each element, a probability distribution function is specified for the proportion $r = |(c_{1,i} - c_{2,i}) / (a_{1,i} - a_{2,i})|$:

$$P(r) = \begin{cases} 0.5 \cdot (1 + \eta) \cdot r^\eta, & \text{if } r \leq 1 \\ 0.5 \cdot (1 + \eta) / r^{2+\eta}, & \text{if } r > 1 \end{cases} \quad (4.5)$$

where η is a non-negative real number. The procedure of computing $c_{1,i}$ and $c_{2,i}$ is described as follows [Deb(1995)].

1. Generate a random number $u \in (0, 1)$;
2. The parameter β is determined in such a way that the area under the probability curve from 0 to β just equals to the chosen random number u , i.e.,

$$u = \int_0^\beta P(r) dr. \quad (4.6)$$

So we have

$$\beta = \begin{cases} \frac{1}{(2u)^{1+\eta}}, & \text{if } u \leq 0.5 \\ \left(\frac{1}{2(1-u)}\right)^{1+\eta}, & \text{if } u > 0.5. \end{cases} \quad (4.7)$$

3. Finally,

$$c_{1,i} = 0.5 \cdot \left((1 + \beta) \cdot a_{1,i} + (1 - \beta) \cdot a_{2,i} \right), \quad (4.8)$$

$$c_{2,i} = 0.5 \cdot \left((1 - \beta) \cdot a_{1,i} + (1 + \beta) \cdot a_{2,i} \right). \quad (4.9)$$

SBX has two important properties:

- The extent of the offspring is proportional to that of the parent individuals.
- Values near the parent individuals are more likely to be chosen than those far away from them.

B Mutation

The common mutation operator for the binary case where only one or two bits are flipped can not be used for real-valued chromosomes because the concept of complementary value is not defined. For a real number vector \bar{a} , the following mutation operators can be used:

Uniform mutation: For each element a_i that is going to be mutated, choose a random value (uniform probability distribution) from the range $[L_i, U_i]$ and assign this value to the element. Here, L_i and U_i are the lower bound and upper bound of a_i , respectively.

Gauss mutation: This mutation is similar to the previous one. The only difference is that the random value is chosen according to Gaussian distribution - smaller jumps are much more probable than large jumps. This is a standard evolutionary strategic mutation.

Exponential mutation: This mutation comes from the idea that the role of mutation at the beginning is to make large jumps whereas later on, as the search progresses, it should be used more for fine-tuning so small jumps are more desirable.

[Michalewicz(1996)] presented such a mutation. If an element a_i is selected to be mutated, its value is changed by

$$a'_i = \begin{cases} a_i - \delta(t, a_i - L_i), & \text{if a random binary digit is 0} \\ a_i + \delta(t, U_i - a_i), & \text{if a random binary digit is 1} \end{cases} \quad (4.10)$$

where t is the generation number. Again, L_i and U_i are the lower bound and upper bound of a_i , respectively. $\delta(t, x)$ is an exponential function, which returns a value in the range $[0, x]$ such that the probability of $\delta(t, x)$ being close to 0 increases as t increases. This property causes the mutation operator to search the space uniformly at the beginning (when t is small), and very locally at later stages. An example of the exponential function can be defined as follows:

$$\delta(t, x) = x \cdot u \cdot \left(1 - \frac{t}{T}\right)^\eta, \quad 0 \leq u \leq 1, \quad (4.11)$$

where u is a random number, T is the maximum number of generations, and η is an exponent that determines the probability distribution.

4.4 MULTI-OBJECTIVE GENETIC ALGORITHMS

The application of GAs to multi-objective optimization seems a logical next step. Early in their development, it was recognized that GAs could be possibly well-suited to multi-objective optimization because of their ability to capture multiple Pareto optimal solutions in a single run and to exploit similarities of solutions by recombination. Some researchers suggest that multi-objective optimization and decision making might be an application area where GAs do better than other optimization algorithms [Zitzler(1999a)].

The first practical GA-based algorithm for multi-objective optimization was presented by [Schaffer(1985)]. Thereafter, a few different implementations of MOGAs were proposed such as [Fonseca(1993)], [Horn(1994)], [Srinivas(1994)], and [Zitzler(1999b)]. These approaches were successfully applied to various MOPs and to real-world engineering design problems. With the development of many new algorithms and numerous applications, some researchers have attempted to summarize the studies in the area of multi-objective optimization from different perspectives [Fonseca(1995b), Horn(1997), Deb(1999), Coello(1999), and Zitzler(1999a)].

4.4.1 Multi-objective search techniques

What makes MOGAs attractive can be attributed to several desirable characteristics. The most important is the population-based search mechanism. Multiple individuals can search for multiple solutions in parallel. Additionally, the ability to handle complex problems, involving features such as discontinuities, multi-modality, disjoint feasible spaces and noisy function evaluations, reinforces the potential effectiveness of GAs in multi-objective search and optimization.

However, according to the goals of multi-objective optimization, as mentioned in

Subsection 3.5.2, two major issues must be considered when applying a GA for multi-objective optimization [Zitzler(1999a)]:

- How to accomplish fitness assignment and selection, respectively, in order to guide the search towards the global Pareto optimal set.
- How to maintain population diversity in order to prevent premature convergence and achieve a well distributed and well spread non-dominated set.

In order to handle the above problems, many different multi-objective search techniques have been proposed. This results in different implementations of MOGAs. In the following we will discuss the most popular methods for fitness assignment (also selection) and techniques for maintaining the population diversity.

A Pareto-based fitness assignment and selection

As mentioned in Subsection 4.2.2, fitness is understood as reproduction probability i.e. the number of offspring an individual is expected to produce through selection. Also in Subsection 4.2.2, fitness assignment for single objective optimization is considered and two types of fitness assignment approaches, scaling and ranking, are presented. However, since the population can not be totally ordered in the case of multi-objective optimization, fitness assignment and selection must allow for several alternative tradeoffs of the objectives. This is the most difficult task for designing an MOGA.

The idea of calculating an individual's fitness on the basis of Pareto dominance was first suggested by David Goldberg in his book [Goldberg(1989)]. He presented an iterative Pareto ranking procedure. First, the currently non-dominated individuals are assigned rank 1 and removed from the population. Next, a new set of non-dominated individuals is identified and assigned rank 2, and also removed. This process continues until all members of the population are ranked. Finally, the fitness of an individual is assigned according to its rank.

This idea has been taken up by many researchers, resulting in several Pareto-based approaches for fitness assignment and selection. Srinivas and Deb implemented Goldberg's idea most directly [Srinivas(1994)]. In their non-dominated sorting genetic algorithm (NSGA), the above Pareto ranking procedure is used to sort the population according to the individual's rank. The individuals with the same rank constitute a tradeoff front. For each front, fitness sharing is performed separately in order to maintain population diversity. Fitness sharing techniques will be discussed later in this section.

Fonseca and Fleming implemented Goldberg's suggestion in a slightly different way [Fonseca(1993)], whereby the population is ranked according the "degree of domination". An individual's rank depends on how many individuals dominate it. All non-dominated individuals are therefore assigned rank 1, while dominated ones are

penalized - the more members of the current population that dominate a particular individual, the lower is its rank. This ranking scheme is more fine-grained than Goldberg's because it can distinguish more ranks.

Horn, Nafpliotis, and Goldberg implemented a selection method, rather than fitness assignment, using a Pareto domination tournament [Horn(1994)]. Two randomly selected competitors are compared with a sample of the current population. If one of the competitors is dominated by a member of the sample, and the other competitor is not dominated, then the non-dominated individual wins the tournament. If both or neither are dominated, then fitness sharing is used to determine the winner. The Pareto domination tournament can be seen as a locally calculated, stochastic approximation to Fonseca and Fleming's globally calculated degree-of-domination ranking.

B Fitness sharing

Maintaining a diverse population is crucial for the efficiency of an MOGA. Fitness sharing is the most frequently used technique, aims at promoting and maintaining multiple sub-populations (called niches in the literature) along the Pareto optimal front. It is based on the idea that individuals in a particular niche have to share the available resources. The more individuals are located in the neighborhood of a certain individual, the more its fitness value is degraded.

The neighborhood is defined in terms of the distance d_{ij} between individuals and specified by the so-called niche radius σ_{share} . Mathematically, the shared fitness z_i of an individual $\bar{a}_i \in P$ is equal to its old fitness z'_i divided by its niche count:

$$z_i = \frac{z'_i}{\sum_{\bar{a}_j \in P} Sh(d_{ij})} \quad (4.12)$$

where $Sh(d_{ij})$ represents a sharing function. An individual's niche count is calculated as the sum of sharing values between itself and the other individuals in the population. A commonly used sharing function is

$$Sh(d_{ij}) = \begin{cases} 1 - \left(\frac{d_{ij}}{\sigma_{share}} \right)^\alpha, & \text{if } d_{ij} \leq \sigma_{share} \\ 0, & \text{otherwise} \end{cases} \quad (4.13)$$

with an exponent $\alpha > 1$. In Srinivas and Deb's NSGA, for instance, $\alpha = 2$ is used.

Obviously, the niche radius σ_{share} is an important parameter for fitness sharing. [Deb(1989)] suggested that

$$\sigma_{share} \approx \frac{1}{2 \cdot \sqrt[n]{q}}, \quad (4.14)$$

where q is the desired number of distinct Pareto optimal solutions and n is the number of decision variables i.e. the length of the decision vector. It has been shown that the use of above equation with $q \approx 10$ works well in many test problems [Deb(1999)].

Furthermore, depending on how the distance d_{ij} is measured, three types of fitness sharing can be distinguished:

- Fitness sharing in individual space ($d_{ij} = \|\bar{a}_i - \bar{a}_j\|$),
- Fitness sharing in decision space ($d_{ij} = \|\bar{x}_i - \bar{x}_j\|$),
- Fitness sharing in objective space ($d_{ij} = \|\bar{y}_i - \bar{y}_j\|$).

where $\|\cdot\|$ denotes an appropriate distance metric.

C Pareto strength and elitism

Zitzler and Thiele have recently suggested a multi-objective optimization algorithm [Zitzler(1999b)], called strength Pareto evolutionary algorithm (SPEA). The fitness assignment of SPEA is also based on the concept of Pareto dominance, but, rather than ranking, it assigns fitness according to the so-called Pareto strength of non-dominated individuals. The Pareto strength of an individual is proportional to the number of individuals it dominates.

Let \bar{P}_d denote the non-dominated set regarding the population P , and P_d denote the dominated set. For each individual $\bar{a}_i \in \bar{P}_d$, its strength s_i is calculated as

$$s_i = \frac{|\{\bar{a}_j \mid \bar{a}_j \in P_d \wedge \bar{a}_i \succ \bar{a}_j\}|}{|P_d| + 1}. \quad (4.15)$$

The fitness of \bar{a}_i is equal to its strength i.e. $z_i = s_i$. Obviously, $0 \leq z_i < 1$ for all $\bar{a}_i \in \bar{P}_d$. Here, the minimization of fitness is desired - smaller fitness value, higher reproduction probability. The fitness of an individual $\bar{a}_j \in P_d$ is calculated by summing the strengths of all the individuals $\bar{a}_i \in \bar{P}_d$, by which \bar{a}_j is dominated. One is added to the sum such that $z_j > 1$, guaranteeing that members of \bar{P}_d have better fitness than members of P_d .

$$z_j = 1 + \sum_{\bar{a}_i \in \bar{P}_d \wedge \bar{a}_i \succ \bar{a}_j} s_i. \quad (4.16)$$

This fitness assignment ensures that the search is directed towards the non-dominated solutions and simultaneously diversity among dominated and non-dominated solutions are maintained.

Zitzler and Thiele also suggested maintaining an external population at every generation, storing all non-dominated solutions discovered so far. This external population is called elite. The strategy of preserving the currently best solutions is called elitism. It has been shown that elitism helps in achieving better convergence for MOGAs.

Besides maintaining an external elite population, there are other implementations of elitism. The method originally suggested by [De Jong(1975)] is achieved by simply copying a certain percent of the best individuals to the next generation. Another often used method is to combine the parent population and the offspring together. The best 50% of the combined population is preserved. In this way, previously generated good solutions are given a chance to carry over to subsequent generations. This method is implemented in the following MOGA.

4.4.2 Elitist non-dominated sorting genetic algorithm

The original non-dominated sorting genetic algorithm (NSGA) was proposed by Srinivas and Deb in [Srinivas(1994)]. This algorithm comprises two main procedures. A Pareto ranking procedure is used to emphasize the non-dominated individuals and a fitness sharing procedure is used to maintain diversity in the population. Later on, motivated by improving the original NSGA, [Deb(2000a)] proposed a new version of NSGA, called NSGA-II, which differs from the old version in three aspects. First, the new non-dominated sorting procedure is much faster. Second, NSGA-II uses a crowded comparison operator instead of fitness sharing such that it does not require the sharing parameter σ_{share} . Finally, an elitism policy is introduced to prevent the loss of good solutions. This algorithm is presented next.

A Non-dominated sorting approach

It is demonstrated that the computational complexity of sorting a population of size K according to the level of non-domination is $O(mK^3)$, where m is the number of objectives [Deb(2000a)]. In case of large population sizes, it is very time-consuming. Algorithm 4.2 gives a fast non-dominated sorting approach which requires at most $O(mK^2)$ computations.

Two entities are used to measure the dominance status for each solution \bar{a}_i : (i) n_i , the number of solutions which dominate \bar{a}_i , and (ii) D_i , a set of solutions which are dominated by \bar{a}_i . The calculation of these two entities requires $O(mK^2)$ comparisons. All solutions with $n_i = 0$ are put in a list F_1 , called the current front. For each solution \bar{a}_i in the current front, each member \bar{a}_j of D_i i.e. the dominated solution of \bar{a}_i , is modified by reducing its n_j count by one. If $n_j = 0$ for any member \bar{a}_j , it will be put in a separate list H . When all members of the current front have been checked, F_1 is declared as the first non-dominated front, and the process continues using the newly

found H as the current front. Each such iteration requires $O(K)$ computations. The above procedure is repeated until all fronts are identified. Since at most there can be K fronts, the worst case complexity of this loop is $O(K^2)$. The overall complexity of the algorithm is therefore $O(mK^2) + O(K^2)$ or $O(mK^2)$.

Algorithm 4.2 (Non-dominated sorting)

Input: P – Population

Output: The non-dominated fronts $F_r, r = 1, 2, \dots$

Step 1:

Set $F_1 = \emptyset$;

For each $\bar{a}_i \in P$ do

1) Set $n_i = 0$ and $D_i = \emptyset$;

2) For each $\bar{a}_j \in P$ do

If \bar{a}_i dominates \bar{a}_j then $D_i = D_i \cup \{\bar{a}_j\}$,

else if \bar{a}_j dominates \bar{a}_i then $n_i = n_i + 1$;

3) If $n_i = 0$ then $F_1 = F_1 \cup \{\bar{a}_i\}$.

Step 2:

Set $r = 1$;

While $F_r \neq \emptyset$ do

1) Set $H = \emptyset$;

2) For each $\bar{a}_i \in F_r$ do

For each $\bar{a}_j \in D_i$ do

a) $n_j = n_j - 1$;

b) If $n_j = 0$ then $H = H \cup \{\bar{a}_j\}$.

3) Let $r = r + 1$ and $F_r = H$.

B Crowded comparison operator

Traditional approaches of maintaining diversity in a population are based on the concept of fitness sharing, which requires a pre-specified parameter σ_{share} . The performance of some MOGAs is very sensitive to this parameter. In order to avoid this difficulty, [Deb(2000a)] proposed a crowded comparison method, where the local density of a particular solution is measured by the distance between its neighboring solutions along each of the objectives. This distance is defined as crowding distance. As shown in Figure 4.4, the crowding distance of the i -th solution (represented by point i) in its front (marked with a dashed curve) is the side-length of the cuboid enclosing it.

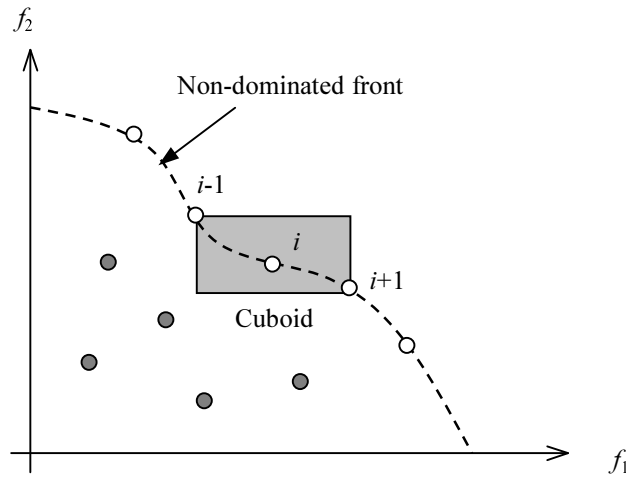


Figure 4.4 Crowding distance of an individual in the non-dominated front.

The crowded comparison operator uses two criteria to compare different solutions: the non-domination rank (r_i) and crowding distance (d_i). Solution \bar{a}_i is considered better than \bar{a}_j , if $r_i < r_j$, or in case $r_i = r_j$, $d_i > d_j$. This method emphasizes the solution having lower rank i.e. the solution close to the Pareto optimal front. For two solutions in the same front, the one with larger crowded distance is preferred. By this way, solutions from less dense regions in the search space are given more chance in the selection process. This method makes sure that the search is guided towards a uniformly outspread Pareto optimal front.

C Elitism

It is well established through a number of studies that elitism can improve the performance of GAs. However, the implementation of elitism in MOGAs is substantially more complex. In the context of multi-objective optimization, there is an elite set instead of one best individual in single objective optimization. All solutions that belong to the current non-dominated front are best solutions in the population because none of them can be said to be better than the other if no preference information is given. Thus, all these solutions are elite solutions. The elite set size can be considerable compared to the population. On many occasions, the population may be mostly composed of current non-dominated solutions. This could cause premature convergence.

In [Deb(2000b)], a controlled elitism mechanism was introduced for NSGA-II. In each front, a predefined number of individuals are maintained. Here, a geometric distribution of the front size is used:

$$n_r = \rho \cdot n_{r-1} \quad (4.17)$$

where n_r is the maximum number of allowed individuals in the r -th front and ρ is the reduction rate, $0 < \rho < 1$. Assume that the elitist population size is K and the combined population is ranked according to Pareto dominance, resulting R non-dominated fronts. The controlled size of the r -th front is

$$n_r = K \cdot \frac{1 - \rho}{1 - \rho^R} \cdot \rho^{r-1}. \quad (4.18)$$

Since $\rho < 1$, each front is restricted to have an exponentially reducing number of solutions. The first front, which comprises the current non-dominated solutions, is largest.

This controlled elitism approach allows solutions from many different fronts to co-exist in the population, making the recombination operator able to create diverse solutions. This is an important feature for preventing premature convergence.

The elitist non-dominated sorting genetic algorithm is described in Algorithm 4.3. This algorithm is based on the structure of Algorithm 4.1. Some steps, e.g. initialization and reproduction, are identical. Moreover, the above discussed multi-objective search techniques, such as non-dominated sorting, crowded comparison, and controlled elitism, are implemented.

Algorithm 4.3 (Elitist non-dominated sorting genetic algorithm)

Input: T – The given number of generations
 K – Population size
 p_r – Recombination probability
 p_m – Mutation probability
 ρ – Reduction rate of the controlled elitism

Output: Non-dominated solutions in P

Step 1 (Initialization):

 Set $t = 0$ and $P' = \emptyset$;
 Generate an initial population P according to Step 1 in Algorithm 4.1;
 Calculate the objective functions for each individual.

Step 2 (Non-dominated sorting):

$P = P \cup P'$;
 Do non-dominated sorting according to Algorithm 4.2, resulting the non-dominated fronts (F_1, F_2, \dots, F_R) .

Step 3 (Controlled elitism):

Set $r = 1$ and $P = \emptyset$;

While $|P| < K$ do

- 1) Calculate n_r according to (4.18);
- 2) Sort F_r in descending order using the crowded comparison;
- 3) Put the first n_r members of F_r in P , i.e., $P = P \cup F_r [1 : n_r]$;
- 4) $r = r + 1$.

Step 4 (Fitness assignment):

Assign fitness to each individual according to its position in P .

Step 5 (Reproduction):

Generate an offspring P' from P according to Step 3 in Algorithm 4.1 (with offspring size of K);

Calculate the objective functions for each individual in P' .

Step 6 (Termination):

$t = t + 1$;

If $t \geq T$ or another stopping criterion is satisfied then terminate,
else go to Step 2.

4.4.3 Integrated search and decision making

Having an algorithm to find the Pareto optimal front is not the final step for solving a multi-objective problem. A decision must be made among multiple tradeoffs. The choice of a suitable solution depends not only on the problem itself, but also the subjective preference of the human decision maker. Thus, the final solution to the problem is the result of a search process and a decision making process.

Generally, there are three different ways to handle these two processes involved in multi-objective optimization [Fonseca(1995a), Zitzler(1999a)]:

Decision making before search: The multiple objectives are aggregated into a single objective which implicitly includes preference information given by the decision maker.

Search before decision making: The search process is first performed without any preference information to find a set of candidate solutions. The final choice is then made by the decision maker.

Integrated search and decision making: Search and decision making work in an interactive manner. A preliminary search is performed to give the decision maker some idea of the range of possible tradeoffs. The decision maker then gives some preference information to guide the search, which, in turn, presents further solutions to the decision maker. The interactive process between the

decision maker and optimization algorithm continues until a satisfactory solution is found.

The classical multi-objective optimization methods belong to decision making before search. As discussed in Subsection 3.5.2, these methods may have to run the optimization algorithm several times if the preferences of the decision maker change, or the found solution is unsatisfactory. They have also the disadvantage of requiring a priori knowledge about the application problem and being sensitive to the shape of the Pareto optimal front.

MOGAs can be classified as search before decision making. They have the ability to find multiple Pareto optimal solutions in a single optimization run. Since no preference is assumed, the whole tradeoff surface must be carefully sampled. In case an intractably large and complex search space is involved, it would be very expensive, or even impossible, to produce a necessary number of tradeoff solutions.

The integration of search and decision making overcomes the weakness of the other two approaches and combines the strength of both. Interactively refining preferences has the potential advantage of reducing computational effort by concentrating the search in the interesting area from which promising solutions are more likely to emerge, while simultaneously providing the decision maker with updated tradeoff solutions, thereby refined decision can be made.

A relatively small amount of research has been done on integrated search and decision making. Fonseca and Fleming proposed a goal attainment method for their MOGA [Fonseca(1993)]. First, the MOGA is performed for a few generations. Based on the non-dominated solutions found by MOGA, the decision maker chooses a target point to focus subsequent search. The MOGA is then run for a few more generations using a modified multi-objective ranking scheme that takes account of the individual's distance to the target.

In this work, we use an approach based on the principle of goal attainment. Rather than a target point, a target region is defined by specifying the lower bound and upper bound for each of the objective functions. The target region is usually a small portion of the whole search space, and can be adaptively altered according to the actual tradeoff information. Solutions inside the target region are assigned better fitness. Thus, the search is guided towards the desired part of the Pareto optimal front, such that the decision maker can zoom in and get a more accurate description of the global Pareto optimal front.

4.5 SUMMARY

This chapter began with a general introduction to GAs. This was followed by a more detailed analysis of GAs, aiming at designing a particular algorithm for solving an

application problem. GAs are able to tackle complex problems with features such as mixed variables and ill-defined objective function, which are difficult to handle by other optimization techniques.

It is argued that real-valued GAs are superior to the conventional binary GAs for continuous parameter optimization problems. In real-valued GAs, parameters are represented by real numbers instead of bit-strings. This makes the bit-string based genetic operators (recombination and mutation) not applicable to real-valued GAs. Several genetic operators exclusively designed for real-valued chromosomes have been introduced.

GAs are particularly well-suited to multi-objective optimization because of their ability to find multiple Pareto optimal solutions in a single run. A multi-objective GA has been presented, making use of the non-dominated sorting, crowded comparison, and controlled elitism techniques. Moreover, an approach has been proposed for solving multi-objective problems by integrating the search and decision making processes. This approach, together with the multi-objective GA, will be applied to the cell planning problems in the following chapters.

Chapter 5

Disk graph models and the application of genetic algorithms

The purpose of this chapter is to provide a model for studying cell planning problems and to investigate genetic algorithms, especially, the non-dominated sorting genetic algorithm (NSGA). The so-called disk graph model (DGM) is used for this purpose. First, a short introduction to DGM is given, then the local and global conditions are formulated for DGMs. Afterwards, the DGM is used to study the cell selection problem in Section 5.2, where NSGA is also investigated. Section 5.3 gives a model for cell dimensioning. Simulations are made using NSGA with two different representation schemes: bit-strings and integer vectors. Finally, the BS placement problem is considered in Section 5.4. The computational complexity makes this problem intractable, therefore a hierarchical approach is proposed to solve it.

5.1 AN INTRODUCTION TO DISK GRAPH MODELS

5.1.1 Disk graph model and cellular system

Unit disk graph theory has been studied in [Clark(1990) and Marathe(1995)]. A unit disk graph is defined as an intersection graph of equal-sized circles in a plane. The intersection graph of n circles is an n -vertex graph with each vertex corresponding to a circle. Two vertices are connected by an edge if the corresponding circles intersect (tangent circles are assumed to intersect), as shown in Figure 5.1. Such an intersection graph is called a unit disk graph. The set of circles is called an intersection model of the disk graph. There are several other models, such as the containment model and the proximity model, but they can be easily transformed to an intersection model [Clark(1990)]. Therefore, we consider only the intersection model of disk graphs what follows.

Unit disk graphs are a narrow class of graphs. They have some special characteristics due to the equivalent circle size. A disk graph is a generalization of unit disk graph, in which the circles are not necessary equal-sized. Potential applications of disk graphs to cellular system design was first recognized by

[Hale(1980)], where disk graphs were used to represent frequency assignment problems (FAPs). In this work, we use DGMs to study cell planning problems.

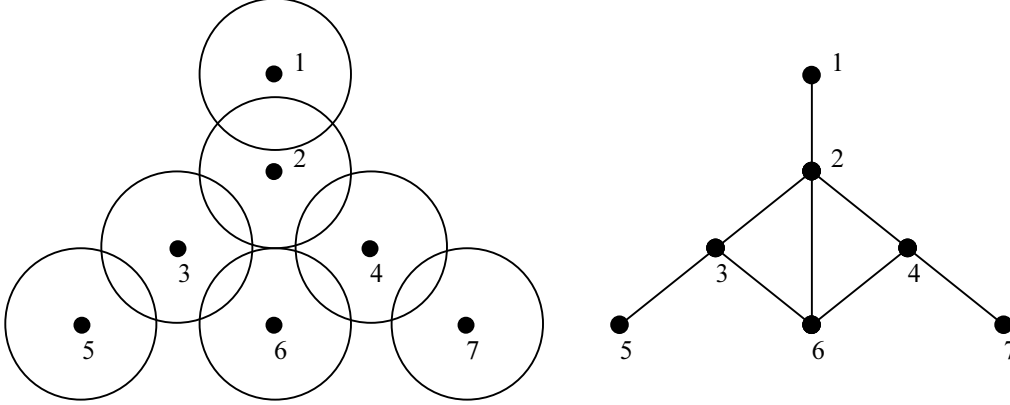


Figure 5.1 An intersection model of equal-sized circles and the corresponding unit disk graph. Tangent circles intersect each other, e.g., circle 2 and 6.

As described in Section 2.3, the coverage of a cell is the geographical area in which the received power P_r is higher than the sensitivity θ_r of mobile receivers i.e. $P_r > \theta_r$. In free space, the received power at a distance d from the transmitter is given by

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2}, \quad (5.1)$$

where P_t is the transmitted power. G_t and G_r are the transmitter and receiver antenna gains, respectively, and λ is the wavelength. The shape of a cell can now be viewed as a circle, the radius of which is

$$d = \frac{\lambda}{4\pi} \sqrt{\frac{P_t G_t G_r}{\theta_r}}. \quad (5.2)$$

Thus, the cellular system can be described using a DGM. Furthermore, if all the transmitters have equal transmitted power and the antenna gains are the same, the cells become equal-sized as their radii are identical. The cellular system is then simplified as a unit disk graph model (UDGM).

Actually, the DGM represents an ideal cellular system, where flat terrain and uniform traffic distribution are assumed. Thus, no computation of the field strength and traffic coverage is needed. The computational time can be significantly reduced. Here DGMs are used to study cell planning problems, including cell selection, cell dimensioning, and BS placement. Growth planning is not considered because no traffic data is used. Although the design scenario is somewhat simplified, the

complexity of the problems is still the same. Therefore, the algorithms designed for DGMs are also applicable to real planning problems.

5.1.2 Formulation

The cell planning problem has already been considered in Chapter 3. In order to formulate these problems using DGMs, some definitions should be reconsidered.

A Coverage

A DGM is a set of cells, $C = \{c_1, c_2, \dots, c_N\}$, where a cell is represented by a circle $c_i = (x_i, y_i, r_i)$ with the center (x_i, y_i) and the radius r_i . Suppose that the area under consideration is digitized as $m \times n$ pixels and the pixel size is 1. In DGMs, the coverage probability $P_{cov}(x, y, c_i)$ i.e. the probability of a point (x, y) being covered by a cell c_i , degenerates to a two-valued function

$$P_{cov}(x, y, c_i) = \begin{cases} 1, & \text{if } (x - x_i)^2 + (y - y_i)^2 \leq r_i^2 \\ 0, & \text{otherwise.} \end{cases} \quad (5.3)$$

A point (x, y) is said to be covered by a cell c_i if its distance to the center of the circle is not larger than the radius. Furthermore, we assume that the assignment probability is always 1 i.e. $P_{ass}(x, y, c_i) \equiv 1$. Here, the probabilistic model is not used. Thus the coverage of c_i is just the circle area size

$$A_{area}(c_i) = \pi \cdot r_i^2. \quad (5.4)$$

B Local conditions

Now we redefine the local conditions for DGMs. Two local conditions of a cellular system were given in Section 3.3: traffic load and interference probability. The traffic load within a cell must be less than some maximum value corresponding to the allowed blocking probability. Since uniform traffic distribution is assumed for the DGM, the traffic load is directly proportional to the cell size. Therefore, this condition can be addressed by specifying a lower bound and an upper bound for cell radius, i.e.,

$$r_{min} \leq r_i \leq r_{max}, \quad i = 1, \dots, N.$$

As we know, mutual interference is caused by the overlapping of cells. In DGMs, two cells are said to be interfering with each other if the corresponding circles intersect. Thus the interference probability of a cell c_i at a point (x, y) is given by

$$P_{int}(x, y, c_i) = \begin{cases} 1, & \text{if } P_{cov}(x, y, c_i) = 1 \text{ and } P_{cov}(x, y, c_j) = 1, j \in [1, N], j \neq i \\ 0, & \text{otherwise.} \end{cases} \quad (5.5)$$

Furthermore, the average interference probability of c_i is defined as the proportion of the overlapping area to the coverage area, i.e.

$$P_{int}(c_i) = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{int}(x, y, c_i)}{A_{area}(c_i)}. \quad (5.6)$$

It is clear that the interference probability must be lower than 1. If $P_{int}(c_i) = 1$, c_i is completely covered by other cells. Such a cell can be removed without any loss in the system coverage.

C Global conditions

Corresponding to the global conditions of a cellular system, three conditions are also defined on the DGM: coverage rate, spectral cost, and financial cost. Here we consider only area coverage as uniform traffic distribution is assumed. The area coverage rate is the same as that given in (3.13), i.e.

$$R_{area}(C) = \frac{\sum_{x=1}^m \sum_{y=1}^n P_{cov}(x, y, C)}{m \cdot n}, \quad (5.7)$$

where $P_{cov}(x, y, C)$ is the coverage probability of the cell set C , as already defined in (3.11),

$$P_{cov}(x, y, C) = 1 - \prod_{i=1}^N (1 - P_{cov}(x, y, c_i)). \quad (5.8)$$

However, due to $P_{cov}(x, y, c_i) \in \{0, 1\}$, $P_{cov}(x, y, C)$ is also two-valued. This is to say that a point (x, y) is considered to be covered, if and only if there is a cell $c_i \in C$ covering it.

In Subsection 3.4.2, we discussed the spectral cost in detail. In DGMs, the spectral cost of a cell set C is calculated by finding the maximum clique in the corresponding disk graph. It is given by

$$M_{req}(C) = \max_{k=1, K} |Q_k|, \quad (5.9)$$

where K is the number of cliques found, and $|Q_k|$ represents the size of the clique Q_k . All the available maximum clique algorithms are applicable to disk graphs. For unit disk graphs, even a polynomial time algorithm can be used to find cliques [Clark(1990)].

Finally, the financial cost of a DGM is simply calculated as the number of cells used i.e. $E_{fin}(C) = |C|$, because all the cells are assumed to be the same cost.

5.2 UNIT DISK GRAPH MODEL FOR CELL SELECTION

As discussed in Subsection 3.2.1, the task of cell selection is to find an optimal cell set from a large number of potential cells. Since the cell sizes are fixed, without loss of generality, they can be assumed to be equal-sized. Thus UDGMs can be used. Formally, the cell selection problem is described as follows:

Definition 5.1 (Cell selection): Given a set of N potential cells $C = \{c_1, c_2, \dots, c_N\}$, find a subset $C' \subset C$, such that the coverage rate $R_{area}(C')$ is maximized, the spectral cost $M_{req}(C')$ is minimized, and the financial cost $|C'|$ (i.e. the number of cells in C') is also minimized.

5.2.1 Test problems

In order to study algorithms for solving the cell selection problem, the following three different test problems were investigated, where the area size is 100×100 and all the cells are equal-sized with the radius being 10.

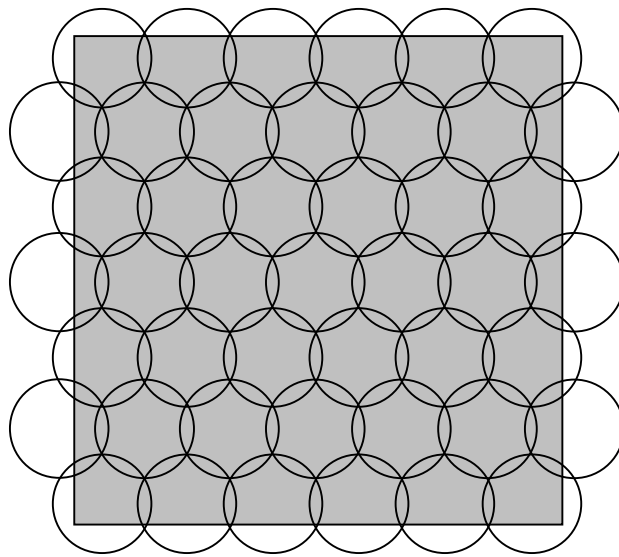


Figure 5.2 Hexagonal distribution of equal-sized cells. This is generally acknowledged to be the optimum.

Test problem 1: There are 100 potential cells. Among them, 45 cells are distributed in a hexagonal structure to cover the whole area, as shown in Figure 5.2. The other 55 cells are randomly placed in the area.

Test problem 2: There are 200 potential cells randomly generated.

Test problem 3: Cells are placed on a grid with a resolution of 4, such that the first cell is at the position (4, 4) and the last one is at (96, 96). There are 576 cells in total.

5.2.2 Simulation results

Numerical simulations of test problem 1 were made using the elitist NSGA, see Algorithm 4.3. A solution is represented by a bit-string $\bar{a} = (a_1, a_2, \dots, a_N)$ in the following way:

$$a_i = \begin{cases} 1, & \text{if } c_i \text{ is selected} \\ 0, & \text{otherwise.} \end{cases} \quad (5.10)$$

For the purpose of illustration, we first consider only two objectives: the coverage rate (f_1) and the financial cost (f_2) i.e. the number of cells selected. GA parameters used in this simulation are listed in Table 5.1:

Table 5.1 GA parameters used in the cell selection simulation of test problem 1

Parameter	Value
Population size K	100
Recombination rate p_r	0.9
Mutation rate p_m	0.01
Elitist reduction rate ρ	0.5

Pareto optimal solutions after running a number of generations are shown in Figure 5.3. The initial population is randomly generated. At the beginning there are only a few non-dominated solutions, which constitute a non-dominated front (the short curve in Figure 5.3). One of the initial non-dominated solutions is shown in Figure 5.4(a), where 62 cells are used and the coverage rate is 0.9087. After 10 generations, many more non-dominated solutions are found and the non-dominated front becomes fully outspread. In subsequent generations, better solutions are obtained. For instance, compare the solutions in generation 10 and those in generation 50. The latter uses less cells for the same coverage rate (as shown in Figure 5.4(b) and

(c)), or from another perspective, gets higher coverage rate using the same number of cells. Observe the history from the beginning to generation 100, we can clearly see that the search is directed towards the global Pareto optimal set. As shown in Figure 5.4(d), the solution obtained after 100 generations is very close to the hexagonal geometry that is generally acknowledged to be the optimal cell distribution, see Figure 5.2. Finally, it requires only 37 cells to achieve a coverage rate of 0.909.

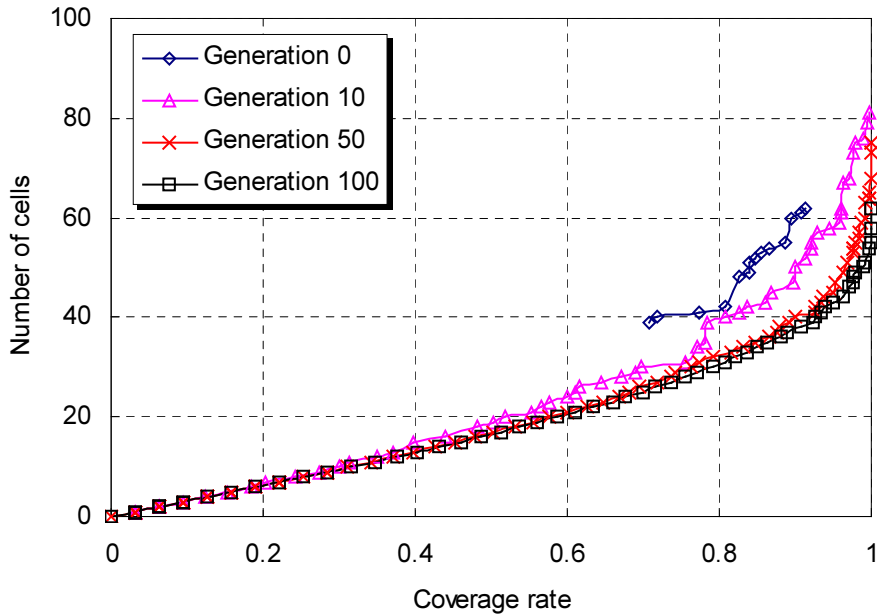
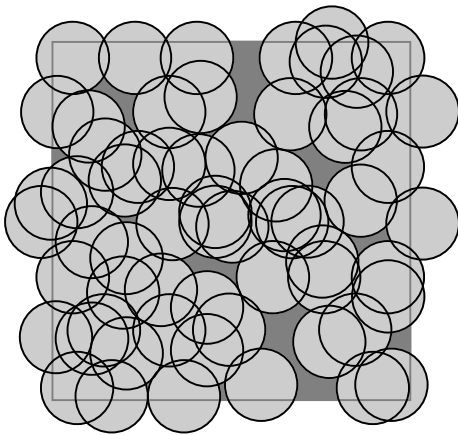
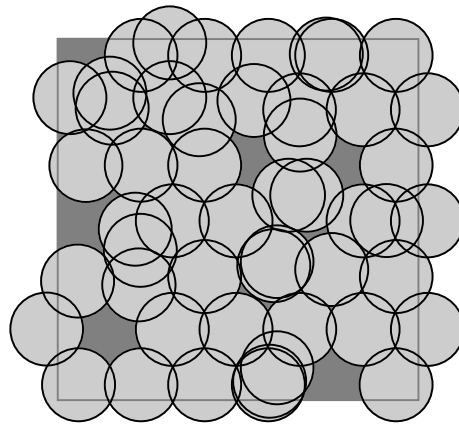


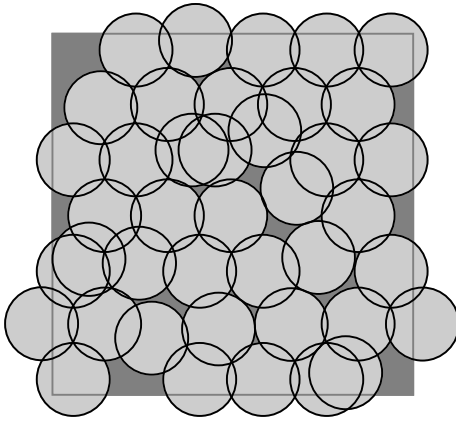
Figure 5.3 Simulation results of test problem 1 using elitist NSGA.



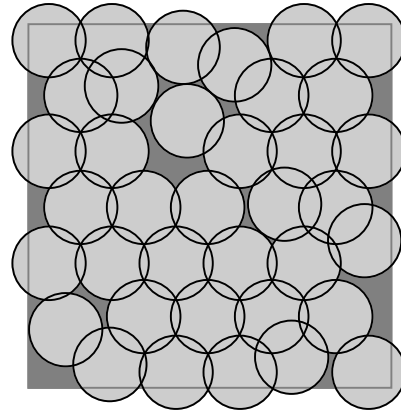
(a) One of the initial non-dominated solutions. There are 62 cells and the coverage rate is 0.9087.



(b) A non-dominated solution obtained after 10 generations. There are 50 cells and the coverage rate is 0.9013.



(c) A non-dominated solution obtained after 50 generations. There are 41 cells and the coverage rate is 0.9126.



(d) A non-dominated solution obtained after 100 generations. There are 37 cells and the coverage rate is 0.909.

Figure 5.4 Illustration of non-dominated solutions obtained in the simulation of test problem 1.

Also notice how the non-dominated solutions get uniformly distributed over the Pareto optimal front. Based on the obtained tradeoff information, the decision maker can find a compromise solution between the conflicting objectives.

A MOGA vs. single objective GA

Furthermore, the NSGA is investigated in comparison with a single objective GA. Here, Algorithm 4.1 is used. Simulations of test problem 3 were made for this purpose. Since the problem size is large, we set the population size as 200. The other parameters are the same as those used in the above simulations. Results after running 100 generations are shown in Figure 5.5. For clarity only solutions with a coverage rate higher than 0.9 are shown.

Results of the single objective GA are obtained using the constraint method, as described in Subsection 3.5.2. Each time a coverage rate constraint is specified (from 0.9 to 1). The optimization objective is then to minimize the number of cells selected. We use the same GA parameters for both algorithms. For each different constraint, the single GA runs 100 generations and the best solution with minimum number of cells is shown in the figure.

The constraint method is very time-consuming because the optimization algorithm has to run many times in order to get different solutions. In contrast, NSGA can find various tradeoff solutions in a single run. Moreover, the solutions obtained by NSGA are even better than those of the constraint method.

In the following, we use the DGM to study the NSGA from various design aspects. The purpose of this study is to improve algorithm effectiveness for future application to cell planning problems.

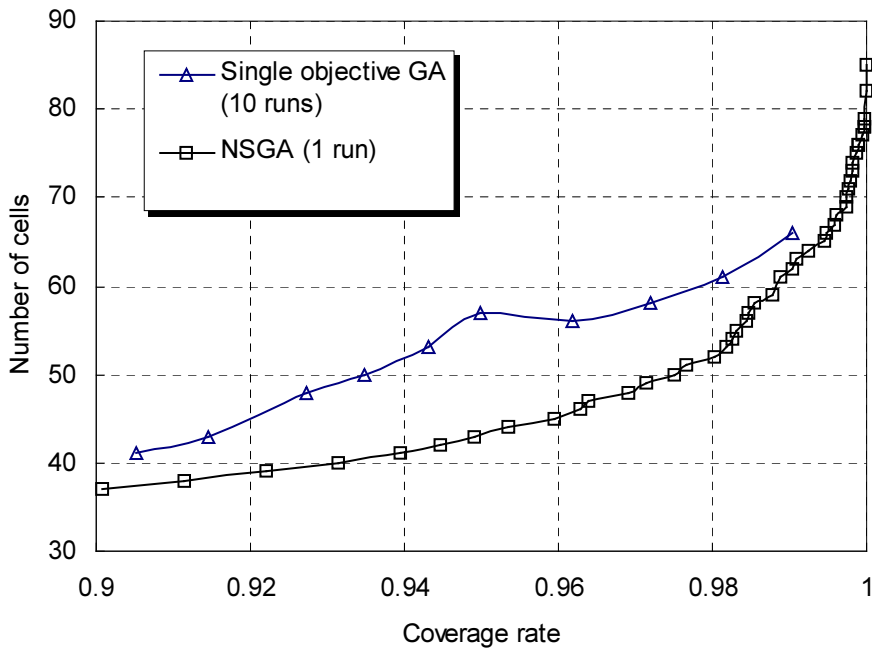


Figure 5.5 Multi-objective GA vs. single objective GA.

B Selection method

Two different selection methods are investigated: a roulette wheel selection and a binary tournament selection, as described in Subsection 4.2.3. Simulations have been performed for all three test problems with both selection methods. The population sizes are 100 for problem 1, and 200 for problem 2 and 3, respectively. The number of generations is 100 for all the simulations. The other parameters are the same as before. Simulation results are shown in Figure 5.6. We can see that the binary tournament selection has better performance than the roulette wheel selection for all the test problems. The reason is that tournament selection does not rely on scale information. So it is well-suited to ranking based fitness assignment such as the non-dominated sorting.

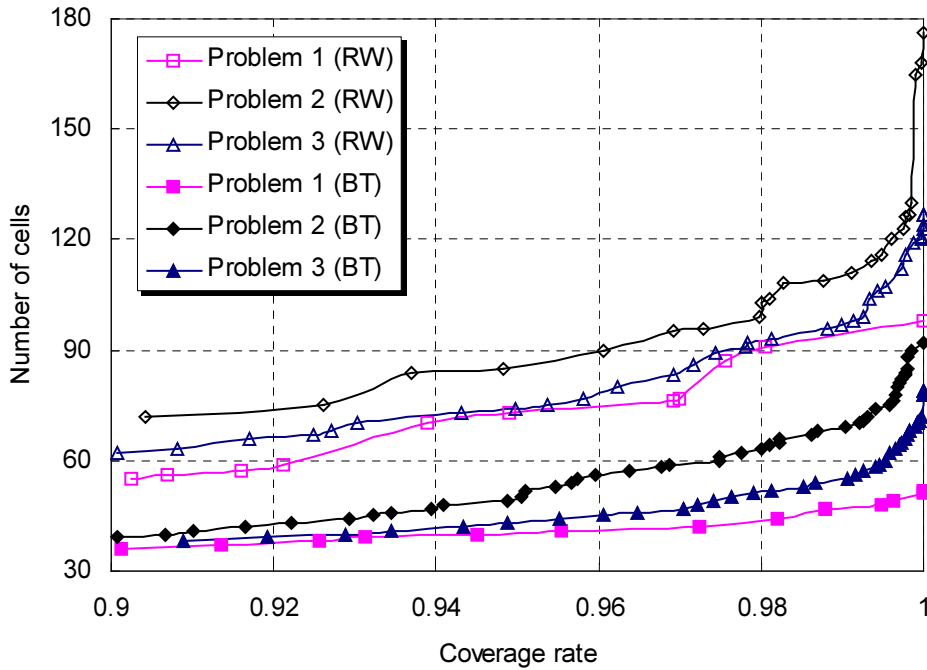


Figure 5.6 Binary tournament (BT) selection vs. roulette wheel (RW) selection.

C Mutation and recombination

Now we consider the influence of genetic operators, including mutation and recombination. In order to find the optimal mutation rate p_m and recombination rate p_r for NSGA, numerous simulations were made with p_m ranging from 0 to 0.1 and p_r from 0.6 to 1. The mutation rate is recognized to have a great influence on the convergence characteristics of GAs. This has also been observed in our simulations, which used test problem 2 with the population size of 200. Simulation results after running 100 generations are shown in Figure 5.7. Both $p_m = 0.01$ and 0.005 achieve very good convergence. The performance becomes worse if the mutation rate is high ($p_m = 0.1$) or no mutation is used at all ($p_m = 0$). This observation corroborates Bäck's statement that the optimal value of mutation rate is about the inverse of the chromosome length [Bäck(1993)]. For test problem 2, the chromosome length is 200 as there are 200 potential cells. Thus $p_m = 0.005$ is the optimum.

Concerning the recombination rate, we find that it does not have a significant influence, but $p_r = 0.9$ or 1 performs slightly better. Furthermore, three recombination operators presented in Subsection 4.2.4 are investigated. They are one-point crossover, two-point crossover, and fusion crossover. Simulations of test problem 2 were made using the above recombination methods. The population size is still 200. In order to investigate the effect of different crossovers, we set mutation rate $p_m = 0$

and recombination rate $p_r = 1$. For each of the three crossovers, 100 generations was used. Figure 5.8 shows the simulation results. It is clearly shown that the fusion crossover works much better than the other two crossovers, and the two-point crossover is slightly better than one-point crossover.

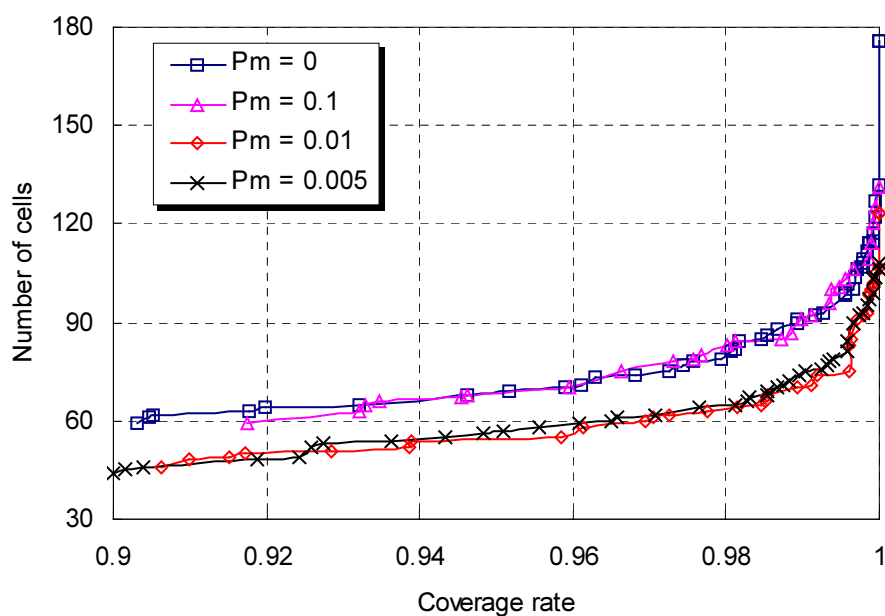


Figure 5.7 Simulation results of test problem 2 with different mutation rates.

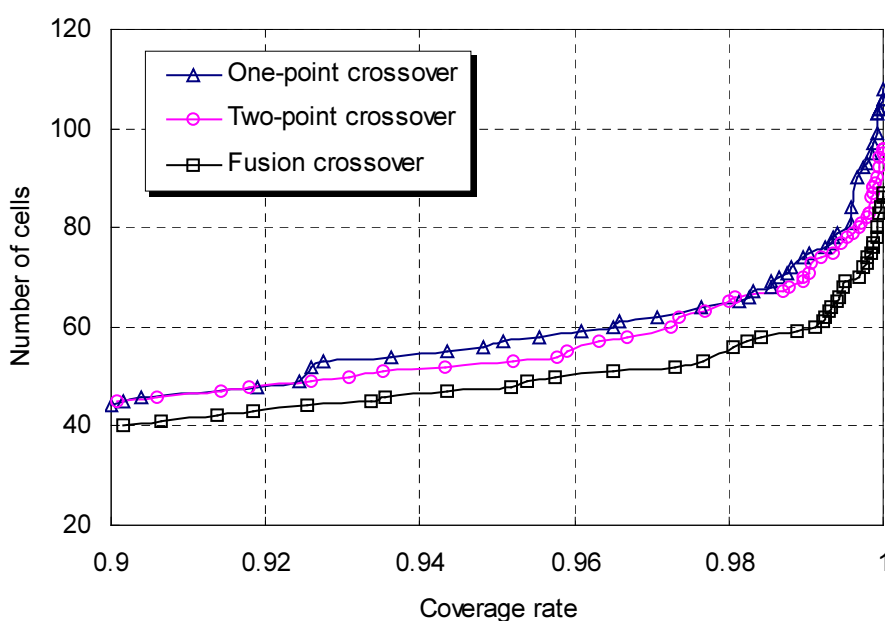


Figure 5.8 Comparisons between one-point crossover, two-point crossover, and fusion crossover.

D Influence of elitism

As observed in [Deb(2000b)], elitism helps in achieving better convergence in MOGAs. In the following, we investigate the influence of elitism using test problem 2. As usual, the population size is 200 and the number of generations is 100. Fusion crossover is used with $p_r = 0.9$ and $p_m = 0.005$. Simulation results are shown in Figure 5.9. The advantage of elitism is obvious - it keeps all previously found good solutions. Therefore, it gives many more non-dominated solutions than the non-elitist method.

In order to investigate the effect of the controlled elitism with different reduction rates, 20 simulations were made using $\rho \in [0, 0.95]$ with an interval of 0.05. We find that none of them is obviously better than the others. The results of four simulations ($\rho = 0, 0.25, 0.5,$ and 0.75) are shown in Figure 5.9. Although $\rho = 0.5$ looks somewhat better, it is not guaranteed that it excels everywhere. Investigation of different elitist strategies is an interesting subject of future research in the area of MOGAs.

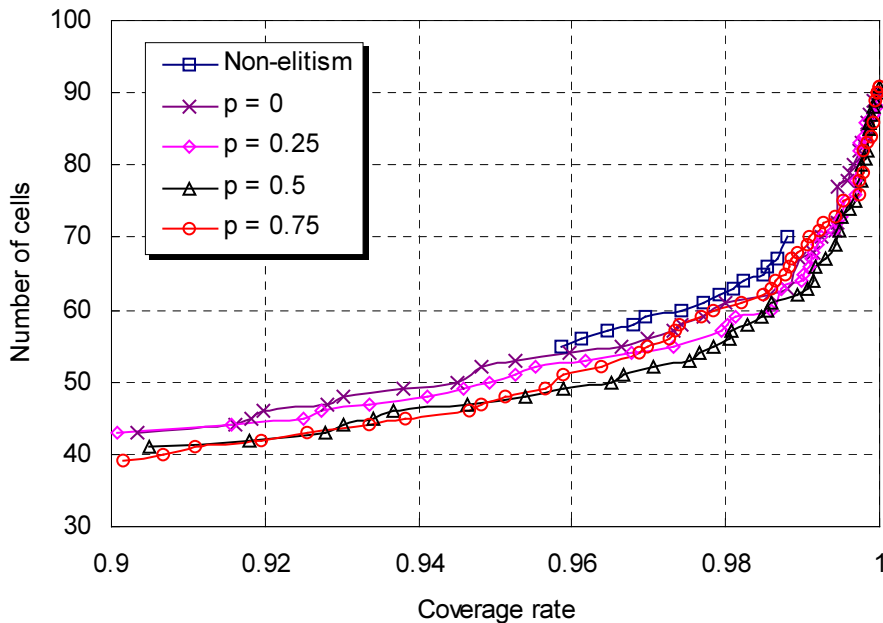


Figure 5.9 Simulation results of non-elitism and the controlled elitism with different reduction rates.

E Simulations with three objective functions

Based on the above studies, we find that the following GA parameters are suitable for the application of NSGA to cell planning problems:

- Binary tournament selection,
- Fusion crossover with $p_r = 0.9$,
- Mutation rate $p_m = 1/N$, where N is the chromosome length,
- Reduction rate of the controlled elitism $\rho = 0.5$.

Finally, simulations were made using NSGA with the above given parameters for a three objective optimization: coverage rate (f_1), spectral cost (f_2), and financial cost (f_3). Simulations using all three test problems have been made. The results of test problem 3 are shown in Figure 5.10, where each point corresponds to a Pareto optimal solution. These solutions are tradeoffs of the three objectives. For a given number of channels, the higher the coverage rate, the more cells are needed. On the other hand, if the number of cells is fixed, more channels are required for a higher coverage rate. We can also see the relation between the spectral cost and financial cost. Generally, more channels are required if more cells are in use.

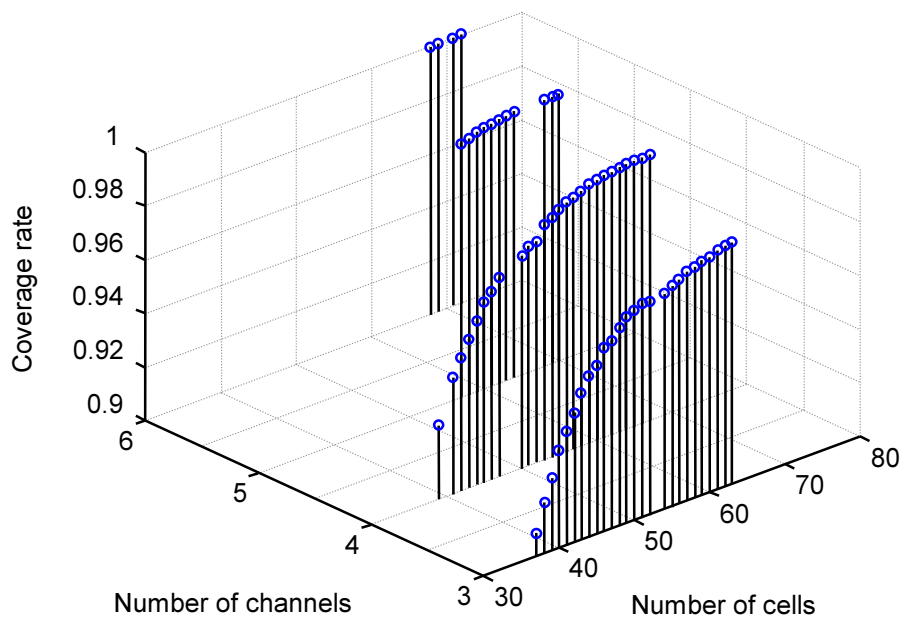


Figure 5.10 Simulation results of test problem 3 considering all the three objectives: coverage rate, spectral cost, and financial cost. The points indicate the Pareto optimal solutions.

5.3 CELL DIMENSIONING MODEL

In contrast to the cell selection problem, in cell dimensioning the number of cells and their positions are given. The task is then to adjust the cell sizes in order to minimize the spectral cost, while achieving maximum coverage rate. This problem can be formally defined as follows:

Definition 5.2 (Cell dimensioning): Given a set of N cells, $C = \{c_1, c_2, \dots, c_N\}$, and the lower bound and upper bound for cell radius, $r_{min} \leq r_i \leq r_{max}$, $i = 1, \dots, N$, change the cell radius such that the coverage rate $R_{area}(C)$ is maximized and the spectral cost $M_{req}(C)$ is minimized.

5.3.1 Test problems

The following two different sized problems were investigated for cell dimensioning:

Test problem 4: The area size is 100×100 . The lower bound and upper bound for cell radius are given such that $r_i \in [0, 15]^4$. There are 45 cells regularly distributed in a hexagonal grid (see Figure 5.2), but their radii are randomly initialized.

Test problem 5: Here the area size is 1000×1000 , and $r_i \in [0, 255]$. 100 cells are randomly placed in the area. Their radii are also randomly initialized.

Both binary GA and real-valued GA, as discussed in Section 4.3, were applied to these test problems.

5.3.2 Binary GA vs. real-valued GA

We first use a bit-string chromosome to represent the solution. The number of bits required to encode a solution is calculated as

$$B = \lceil \log_2(r_{max} - r_{min} + 1) \rceil, \quad (5.11)$$

where r_{min} , r_{max} are the lower bound and upper bound for cell radius r_i , respectively. $\lceil x \rceil$ is the ceiling function that makes a real value x to be rounded up to the closest larger integer value. Thus, $B = 4$ bits are required for test problem 4 as there are 16 possible values of cell radius, and $B = 8$ for test problem 5. Accordingly, the string length is $B \cdot N$, where N is the number of cells. A bit-string $\bar{a} = (a_1, a_2, \dots, a_{BN})$ is mapped to the cell radius in the following way:

⁴ If the radius of a cell is 0, this means that this cell is not in use so it can be discarded.

$$r_i = \sum_{b=1}^B 2^{b-1} \cdot a_{(i-1) \cdot B + b}, \quad i = 1, \dots, N. \quad (5.12)$$

In case floating-point parameters are to be optimized, the coding procedure is more complicated and much longer bit-strings are needed.

As an alternative, an integer vector is used to represent the solution i.e. $\bar{a} = (a_1, a_2, \dots, a_N)$ with $a_i = r_i$, for $i = 1, \dots, N$. These two representation methods were investigated using NSGA with the same GA parameters given in Table 5.1. As to genetic operators, we use fusion crossover and flip mutation for the binary GA, and simulated binary crossover and Gauss mutation for the real-valued GA.

We first made simulations of test problem 4. Results after running 100 generations are shown in Table 5.2. We see that both the binary method and real-valued method find that at least 4 different channels are required for 100% coverage rate. However, in other cases with 1, 2, or 3 channels, the binary GA achieves better performance.

Table 5.2 Simulation results of test problem 4 using binary GA and real-valued GA

Binary GA		Real-valued GA	
Channels	Coverage rate	Channels	Coverage rate
4	1	4	1
3	0.9979	3	0.9944
2	0.8080	2	0.5985
1	0.4613	1	0.2821

Table 5.3 Simulation results of test problem 5 using binary GA and real-valued GA

Binary GA		Real-valued GA	
Channels	Coverage rate	Channels	Coverage rate
6	0.999751	6	1
5	0.991796	5	0.999751
4	0.947964	4	0.989415
3	0.840031	3	0.857232
2	0.663907	2	0.690286
1	0.229016	1	0.036745

Furthermore, we also made simulations of test problem 5 using NSGA with the same GA parameters. The main difference between test problem 4 and 5 is that the latter uses a longer bit-string (800 bits, instead of 180 bits for test problem 4). Pareto optimal solutions after 100 generations are listed in Table 5.3. The solutions obtained by the real-valued GA are better in most cases, except the one with 1 channel. These results show that real-valued GAs are superior to binary GAs for continuous parameter optimization such as the cell dimensioning problem, especially when a long bit-string is required for encoding the solution.

5.4 BASE STATION PLACEMENT MODEL

BS placement is an automatic process, where no potential cells need to be given beforehand. The algorithm explores all the possible places in the given area in order to find optimal cell sites. In the first stage, for the purpose of studying algorithms, all the cells are assumed to be equal-sized such that the problem is not intractable. Generally the BS placement problem can be described as follows:

Definition 5.3 (Base station placement): Place cells, represented by $C = \{c_1, c_2, \dots, c_N\}$, in a given area such that the coverage rate $R_{area}(C)$ is maximized and the spectral cost $M_{req}(C)$ is minimized, using the minimum number of cells $|C|$.

It is clear that BS placement problem is more difficult than cell selection because it has to search for potential cells itself, whereas in cell selection, the potential cells are given. Since BSs can be located at any place in the area, the computational complexity of BS placement problem increases exponentially with the area size. For instance, there are 10,000 potential sites in an area of 100×100 pixels, whereas in case of 1000×1000 pixels, the number rapidly rises to 1,000,000. No algorithm is able to directly tackle such a large problem. Unfortunately, this is a very common case in mobile network design practice.

5.4.1 Hierarchical approach

In order to reduce the computational complexity, a hierarchical approach is proposed for the first time [Huang(2000c)]. It starts from a coarse resolution grid, of which each pixel is taken as a potential cell site. An optimization algorithm is then performed to find an optimal cell set from the potential cells. The obtained cells are further used to search potential cells on the next grid with a higher resolution. Here, the coverage-oriented cell splitting technique is applied (see Subsection 2.1.1). This search and optimization process continues until the finest grid is reached, so that accurate BS positions can be found.

The hierarchical approach can be illustrated by a pyramid structure shown in Figure 5.11. On the first level all pixels are taken as potential cell sites, some of them

are selected after the optimization e.g. the shadowed pixels on Level 1. On the next level, each pixel is divided into a number of subpixels (usually the number is four, as shown in Figure 5.11). Only subpixels corresponding to the selected one are taken as potential sites on the finer grid. Thus we have only 16 potential cells on the second level. By this way, the hierarchical approach can effectively reduce the computational complexity because it continuously eliminates unfit cells. Cell sites are searched for only in promising areas such that the search space will not expand exponentially.

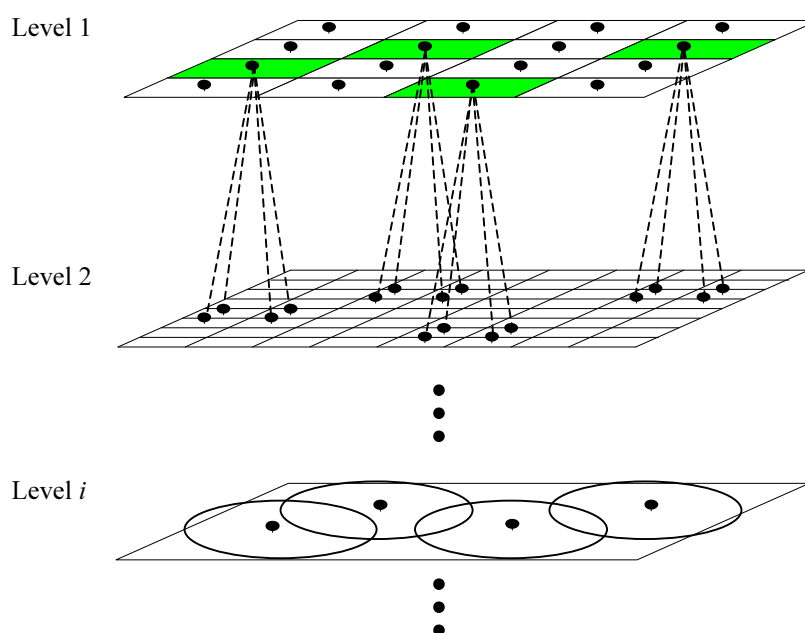


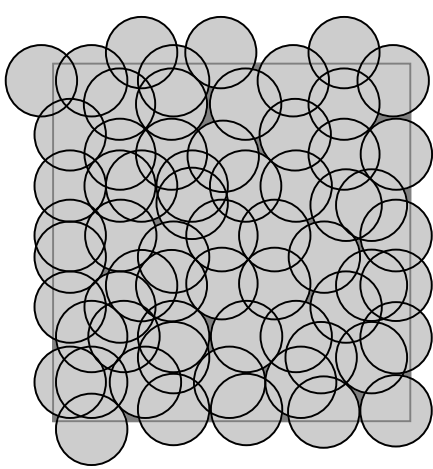
Figure 5.11 Structure of the hierarchical approach.

5.4.2 Simulation results

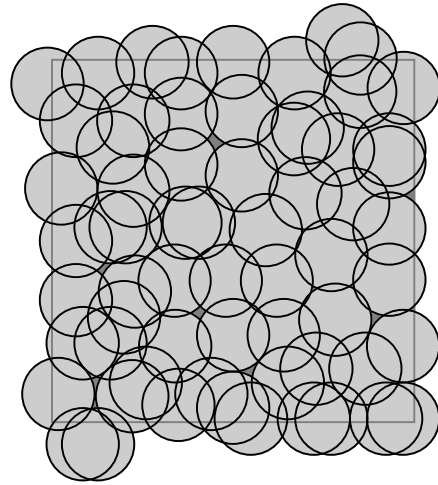
In order to investigate the hierarchical approach, simulations were made for automatically placing cells in a given area. The area size is 100×100 . The cell radius is assumed to be fixed, say 10. We start the hierarchical approach with a resolution of 8. Initially, 256 cells are evenly distributed and the whole area is covered. NSGA is then applied to find Pareto optimal solutions. The following GA parameters are used:

Table 5.4 GA parameters used in BS placement simulations

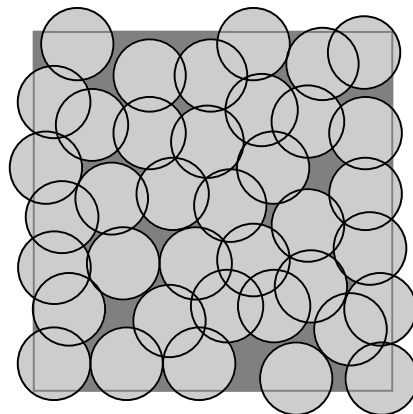
Parameter	Value
Number of generations T	100
Population size K	200
Recombination rate p_r	0.9
Mutation rate p_m	1/256
Elitist reduction rate ρ	0.5



(a) On the first level (with a resolution of 8) a solution of 64 cells is chosen, which has a coverage rate of 0.9906 and spectral cost of 6.



(b) Also, on the second level (with a resolution of 4) a solution of 64 cells is chosen, which has a coverage rate of 0.9984 and spectral cost of 6.



(c) One of the non-dominated solutions obtained on the last level (with a resolution of 2) has 38 cells. The coverage rate is 0.9042 and the spectral cost is 4.

Figure 5.12 Illustration of the solutions obtained by the hierarchical approach

One of the tradeoff solutions should be chosen by the decision maker (here, man-machine interaction is required). In our simulation, we chose solution C_1 that has 64 cells and a coverage rate $R_{area}(C_1) = 0.9906$ and spectral cost $M_{req}(C_1) = 6$, as shown in Figure 5.12(a). On the next level, the resolution is 4, and 256 potential cells are generated according to the cells in C_1 . The optimization algorithm is executed again and a solution C_2 is then chosen, see Figure 5.12(b). It has also 64 cells, but somewhat better performance of $R_{area}(C_2) = 0.9984$ and $M_{req}(C_2) = 6$. The above process is repeated on the next level (now the resolution is 2). The obtained Pareto optimal front is shown in Figure 5.13.

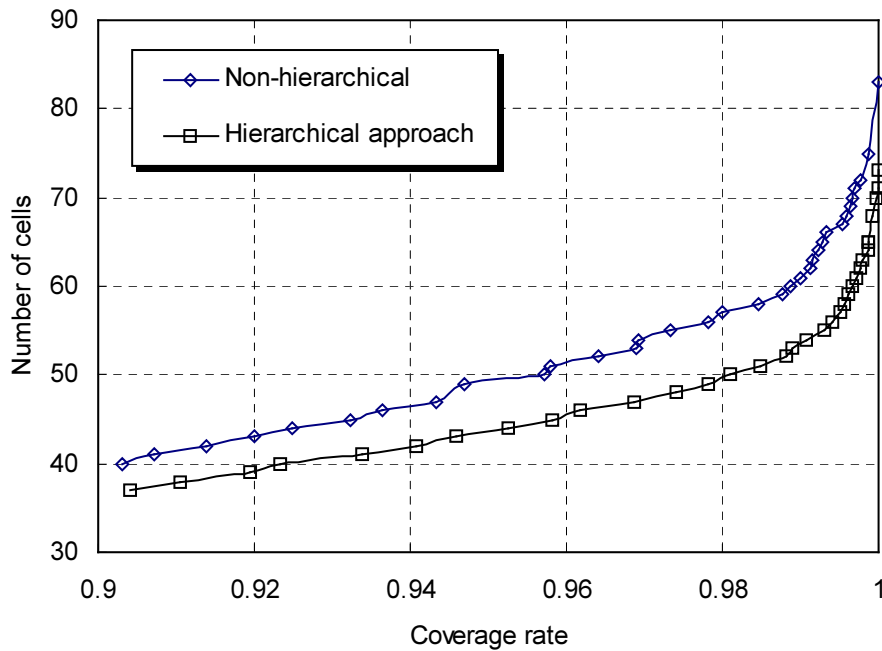


Figure 5.13 Simulation results using the hierarchical approach, compared with that of a non-hierarchical simulation.

For the purpose of comparison, a non-hierarchical simulation was also made, where cells are placed on a grid with the resolution of 2, one cell per pixel. Thus there are 2500 potential cells. NSGA is applied and the same GA parameters are used. Simulation results are also shown in Figure 5.13, where only values of two objectives are displayed in order to make the comparison clear. Since there is no significant difference between their spectral costs, we just consider the coverage rate and the number of cells in this comparison.

This figure makes it clear that the hierarchical approach achieves better performance. The reason is that the search space is too large to be managed if the optimization algorithm is used directly. The larger the search space is, the more

difficult it is for the algorithm to reach the global optimum. Moreover, the computational time is unendurable if the hierarchical approach is not used. It takes more than 15 hours for the above-mentioned non-hierarchical simulation. Using the hierarchical approach, however, each level requires about 1.5 hours, namely 4.5 hours in total. Suppose that we go one step further i.e. place BS on a grid with the resolution of 1. It takes just another 1.5 hours for the hierarchical approach. However, since there are 10,000 potential cells if the search is performed directly on the finest grid, the computational time would soar to more than 10 days!

5.5 SUMMARY

In this chapter DGMs were presented to study the cell planning problem. We first discussed the relation between DGMs and the corresponding cellular systems. The local conditions and global conditions are reformulated for DGMs as the scenario is simplified.

We then presented a cell selection model and applied the elitist NSGA to solve the given test problems. It has been shown by the simulation results that NSGA is well-suited to cell planning problems, and able to achieve very good performance with good GA parameters. The following parameters are found to be suitable:

- Binary tournament selection,
- Fusion crossover with a crossover rate around 0.9,
- Mutation rate around the inverse of the chromosome length,
- Reduction rate between 0.25 and 0.75 for the controlled elitism.

The population size and the number of generations depend on the problem size.

We also studied the cell dimensioning problem using DGMs. Both binary and real-valued GAs are applied to test problems, and simulation results are compared. It is demonstrated that, for continuous parameter optimization, real-valued GAs are more convenient and may achieve better performance.

Moreover, a hierarchical approach is proposed to solve the automatic BS placement problem that is intractable due to the computational complexity. The hierarchical approach reduces the computational effort by guiding the search towards promising areas. Although such a search is not ergodic, simulation results show that the hierarchical approach is practical and achieves satisfactory performance.

Chapter 6

Cell planning experiment results and analyses

The methodology and algorithms presented in previous chapters have been tested on real design scenarios. Numerical experiments have been made on all the four cell planning problems: cell selection, cell dimensioning, automatic BS placement and dimensioning (ABSPAD), and growth planning. First, a description of the simulation environment is given in Section 6.1. Subsequently, in Section 6.2, simulation results are presented and analyzed for each of the experiments.

6.1 SIMULATION ENVIRONMENT

Numerical experiments were made for GSM network planning. This means that system performance evaluations are based on GSM system specification. However, the optimization strategies used in these experiments are also applicable to other systems. The only requirement is to redefine the objective functions.

What makes the following simulations different to those presented in the previous chapter is that realistic design scenarios are considered. Geographical data, that are not used in DGMs, play an important role in real planning. Hence, a wave propagation model is required for predicting the cell coverage. This section gives a description of the simulation environment, including geographical data, wave propagation model, and a database for organizing the input parameters and optimization results.

6.1.1 Geographical data

In mobile network planning, three types of geographical data are used. Topographical data depicts the terrain of a given area, and morphological data defines the land usage type such as city, forest, water, and so on. Besides, traffic data describes the traffic distribution in the service area. The data used in our simulations are shown in Figure 6.1, which are provided by LS telcom AG, Lichtenau, Germany. This is an area in Switzerland. The area size is $60 \text{ km} \times 60 \text{ km}$ and the resolution is $100 \text{ m} \times 100 \text{ m}$ per pixel. Thus, there are 600×600 pixels in total. Figure 6.1(a) displays the

topographical data, and Figure 6.1(b) is the morphological data. 12 morpho classes can be distinguished. The traffic data are shown in Figure 6.1(c), where a non-uniform traffic distribution is illustrated. Totally there are 103.1516 Erlang traffic.

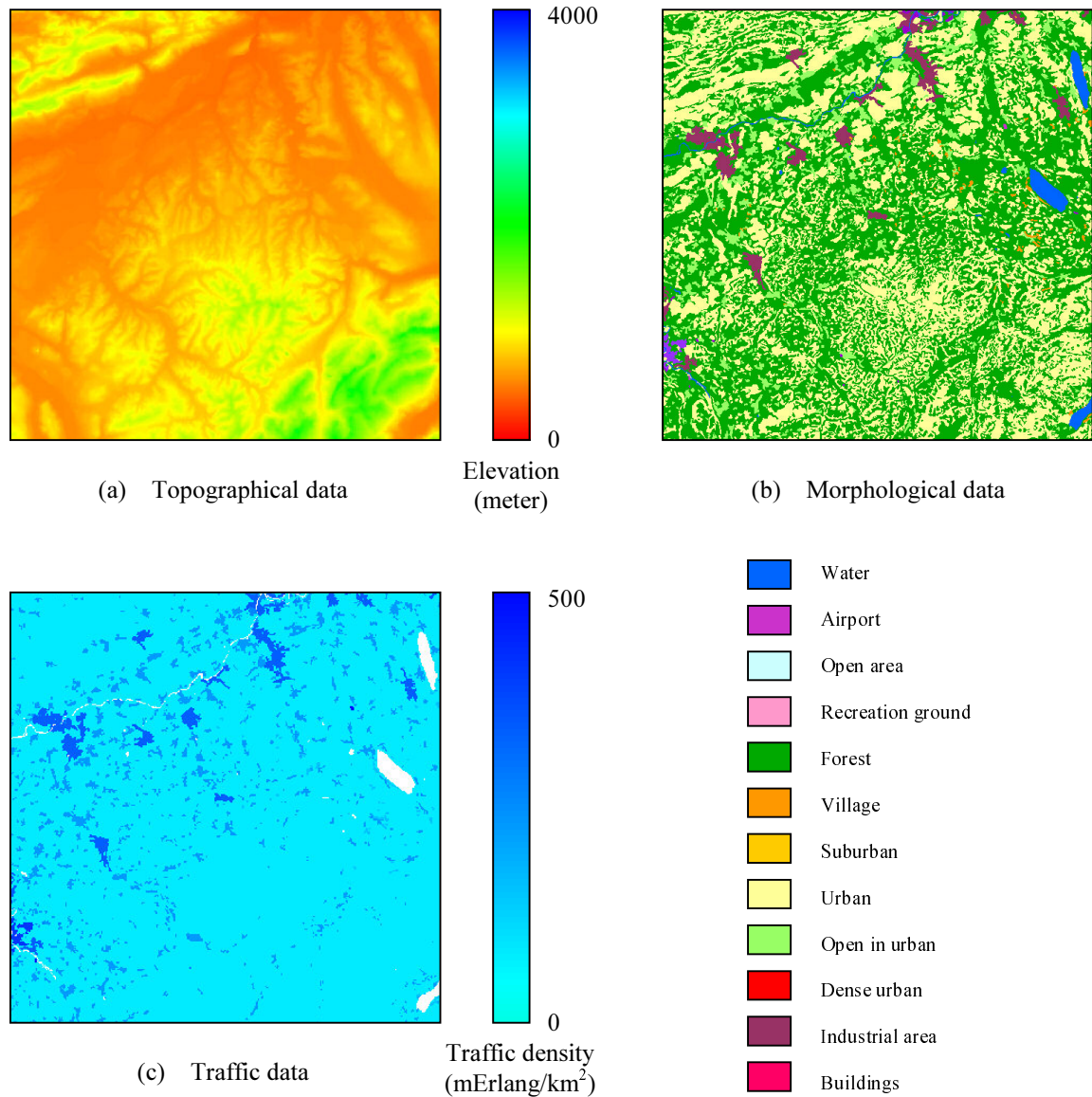


Figure 6.1 Geographical data used in cell planning experiments. The size of the area (part of Switzerland) is $60 \text{ km} \times 60 \text{ km}$ and the resolution is $100 \text{ m} \times 100 \text{ m}$ per pixel. There are 12 morpho classes. Non-uniform traffic distribution is considered.

6.1.2 The wave propagation model

In realistic mobile radio environment, the shape of a cell is extremely irregular because of the uneven terrain and the presence of trees, buildings, and other obstacles.

In order to predict the cell coverage, field strength must be first calculated using a suitable propagation model. Since macrocells are considered in the following simulations, the COST 231 Okumura-Hata model can be employed, which has been discussed in Subsection 2.2.1. Due to its simplicity and well-development, this model is widely used in mobile network planning practice. The other advantages include that this model is computationally efficient - it takes about 10 seconds for such an area with 600×600 pixels using a common PC (Pentium-III CPU at 450 MHz and 256 MB RAM), and that it does not require very detailed geographical information. For predicting coverage of microcells or picocells, more sophisticated models should be used, but more accurate data are required to describe the propagation environment.

An advanced Okumura-Hata model is developed in LS telcom AG, which is used in the following simulations. Considering that the terrain in Switzerland is quite hilly, a multiple knife-edge diffraction model is implemented to calculate the diffraction attenuation. The Epstein-Peterson model [Epstein(1953)] is used, which is combined with the Okumura-Hata model. Here, only the topographical height is considered. The morpho height is not used at the moment, i.e. it is set as 0 for all the morpho classes.

In this propagation model the morpho correction is taken into account. The correction factor corresponding to each morpho class is listed in Table 6.1.

Table 6.1 Morpho correction factors used in Okumura-Hata model

Morpho class	Correction (dB)
Water	28
Airport	28
Open area	23
Recreation ground	23
Forest	3
Village	10
Suburban	10
Urban	5
Open in urban	5
Dense urban	3
Industrial area	0
Buildings	0

6.1.3 The database

A database has been built to manage the input data and optimization results. It consists of 10 tables. Their relationships are shown in Figure 6.2.

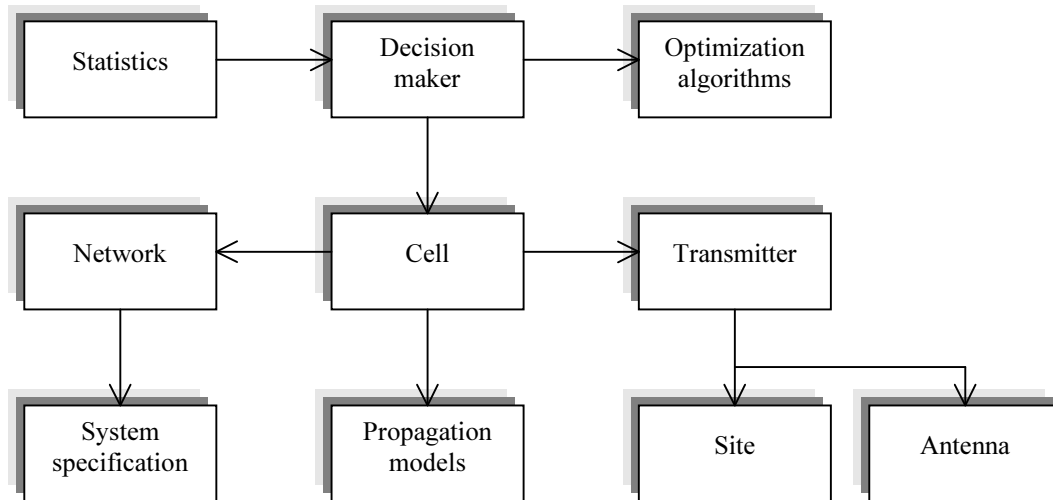


Figure 6.2 Tables of the database used in cell planning experiments.

The statistics table takes care of statistical information of the optimization process and keeps the optimization results. It gives the system information about the service area size, traffic amount etc. as well as system performance, such as the coverage rate, spectral cost, and financial cost. Also, it records how many BSs are used and how long the optimization process has been run. In case the hierarchical approach is applied, the grid resolution of each level is also recorded in addition to the other statistical information.

The statistics table has a link to the decision maker table, by which the preference of the human decision maker is accessible. The decision maker chooses which algorithm to use and specifies constraints for each of the local conditions and global conditions. There are several optimization algorithm tables as different optimization algorithms could have their own parameters. For instance, the population size, recombination rate, mutation rate etc. must be given for GAs. These parameters are specified in the corresponding optimization algorithm table.

The cell table gives the description of each cell, such as the area coverage, traffic load, interference probability, and frequency channels used by the cell. It also specifies which propagation model is used to predict the cell coverage. The cell shape is described by three raster data: field strength data, coverage probability data, and assignment probability data. These data are stored in separate files, but they are accessible to the cell table. There might be a number of different propagation model

tables. Each of them gives parameters for a specific model.

Each cell has a transmitter placed at a certain site. In the transmitter table, parameters of each transmitter are given, including the transmitted power, frequency, cable attenuation, loss in the transmitting path as well as in the receiving path, and so on. The site table gives geographical information about the cell site, such as the coordinate and the building height if it exists. Available antennas are listed in the antenna table, which gives the antenna gain, polarization, azimuth, downtilt, and antenna diagram. Besides, the costs of transmitters, sites, and antennas are retrievable by respective tables.

Cells belonging to one system are linked to the same network record. The geographical data, that are stored in separate files, are retrievable by the corresponding network record. It also gives the description of the service area, such as the area size and resolution. In case a non-rectangular area is involved, a vector is required to describe the border. Also, the network table has a link to the system specification table, where a number of technical parameters are specified. The parameters used in our simulations are listed in Table 6.2. Their values are given according to GSM system specification.

Table 6.2 Technical parameters used in cell planning experiments

Parameter	Value
Received power threshold	-92 dBm
Received power deviation	6 dB
Handoff margin	5 dB
Blocking probability	0.02
C/I threshold	10 dB
Interference probability threshold	0.1

6.2 SIMULATION RESULTS AND ANALYSES

6.2.1 Cell selection

We start with an experiment on cell selection. There are 36 potential cells i.e. $C = \{c_1, c_2, \dots, c_{36}\}$, distributed in the service area. BS sites of these cells are shown in Figure 6.4(a). Since the configurations of cells are unchangeable in cell selection, without

loss of generality, it can be assumed that all the potential cells have the same parameters. Their values are given in Table 6.3.

Table 6.3 Configurations of potential cells in the cell selection experiment

Parameter	Value
Transmitted power*	50 dBm
Frequency	900 MHz
Antenna type	Omni
Antenna gain	15 dBi
Polarization	Vertical

* The transmitted power is represented by effective isotropic radiated power (EIRP).

The task is to find a subset $C' \subset C$, such that the following three objectives are satisfied:

$$\begin{aligned} &\text{maximize} && f_1(C') = R_{trf}(C'), \\ &\text{minimize} && f_2(C') = M_{req}(C'), \\ &\text{minimize} && f_3(C') = E_{fin}(C'). \end{aligned}$$

Here, the traffic coverage rate $R_{trf}(C')$ given by (3.15) is used due to the non-uniform traffic distribution. Since co-channel interference is considered in this experiment, the spectral cost $M_{req}(C')$ given by (3.23) is used. Furthermore, we assume that all the potential cells have equal site cost and equipment cost. Thus the financial cost can be simplified as $E_{fin}(C') = |C'|$ i.e. the number of cells in C' .

NSGA is applied to solve the multi-objective optimization problem. Each solution is encoded as a bit-string with each bit corresponding to a cell. The GA parameters given in Table 5.1 are used, but now the mutation rate $p_m = 1/36$. Decision making is done by a human decision maker during the multi-objective search. First, the NSGA is performed for a few generations (10 generations in this experiment). Based on the non-dominated solutions found, the decision maker specifies a target region to adjust subsequent search. Here, $0.9 \leq f_1(C') \leq 1$ and $1 \leq f_2(C') \leq 30$ are set for the coverage rate and spectral cost, respectively. The financial cost is taken as an open condition. Afterwards, the NSGA runs for another 10 generations and the target region is then modified according to the available tradeoff information as $0.95 \leq f_1(C') \leq 1$ and $1 \leq f_2(C') \leq 20$. The NSGA is continuously executed for 50 more generations. A number of non-dominated (Pareto optimal) solutions are obtained, as shown in Figure 6.3.

Finally, one of the Pareto optimal solutions is chosen. Objective scores of the

final solution are as follows: $R_{trf}(C') = 0.9706$, $M_{req}(C') = 15$, and $|C'| = 7$. It is a compromise of the three objectives. Although a higher coverage rate can be achieved, it might require more BSs and frequency channels. On the other hand, if less BSs are used, the coverage rate would drop substantially. Performance of the selected cells is listed in Table 6.4. The corresponding coverage areas of these cells are shown in Figure 6.4(b).

Table 6.4 Performance of the selected cells

Cell no.	Traffic load (Erlang)	Number of traffic channels	Interference probability
4	18.6238	27	0.04
6	10.7476	18	0.06
9	20.1194	28	0.04
10	5.0611	10	0.04
12	20.6474	29	0.07
20	13.9504	21	0.04
26	10.9648	18	0.05

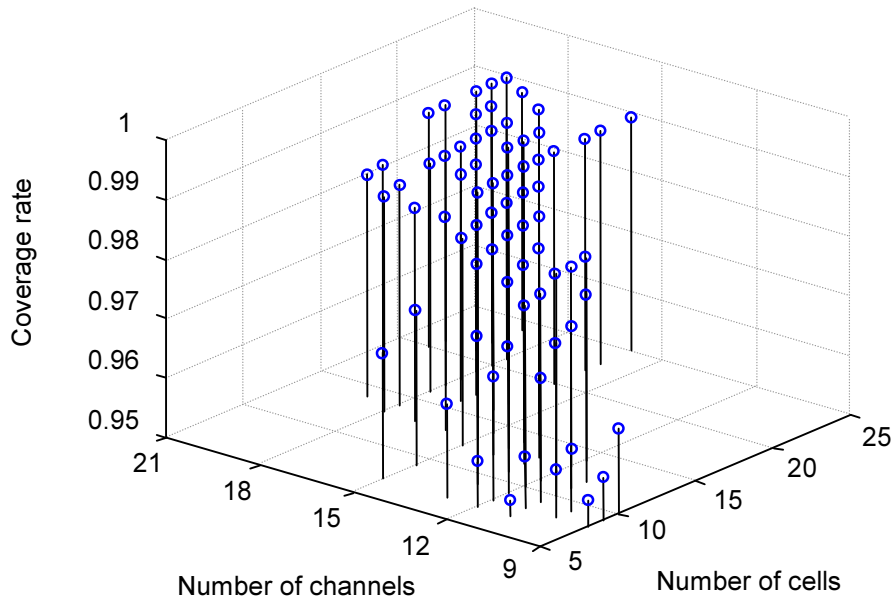


Figure 6.3 Pareto optimal solutions obtained in the cell selection experiment.

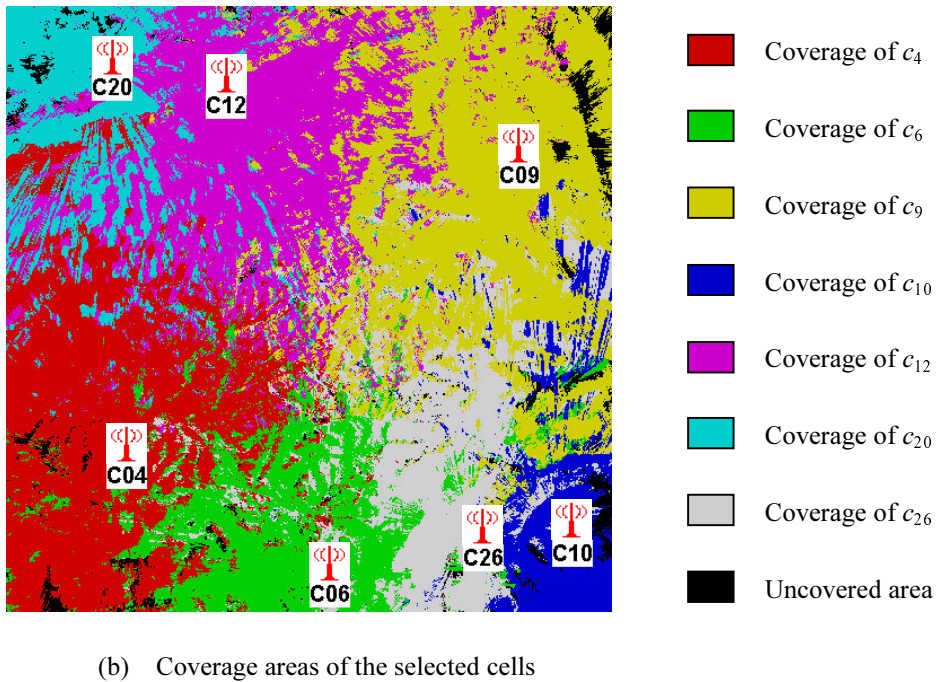
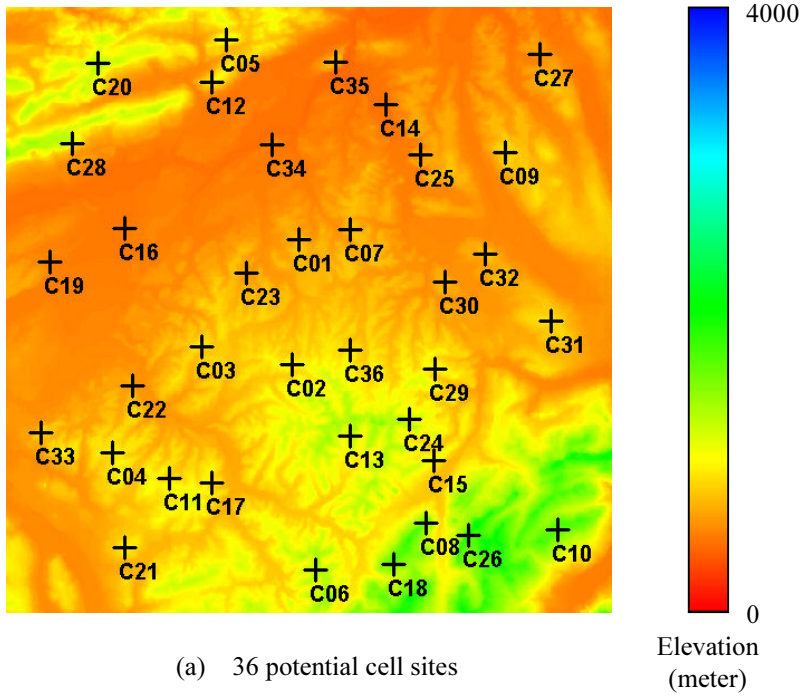


Figure 6.4 Cell selection experiment. In (a) there are 36 potential cell sites displayed on the topographical map, from which 7 cells are selected. Coverage areas of the selected cells are shown in (b). Dark areas are uncovered.

We can see from Figure 6.4 that potential cells situated at high positions are more likely to be selected in the final solution. The reason is that these cells provide better coverage than the others. It fulfils our optimization objectives: maximum coverage rate with minimum cost. Due to the large coverage, however, the overlapping areas become large too. This could cause higher interference probability in some areas e.g. the center area in Figure 6.4(b). The interference problem can be solved by a good channel assignment, but it could require more channels, and thus higher spectral cost. Next, we will discuss how to optimize cell dimensions in order to reduce the spectral cost.

6.2.2 Cell dimensioning

One reason for cell dimensioning is due to the traffic load condition. Assume that each cell can use at most 4 frequency channels, corresponding to 29 traffic channels (see Table 2.2). Since the blocking probability is also given, the maximum traffic load can be derived according to the Erlang B formula. In the experiments, $P_{block} = 0.02$, thus $T_{max} = 21.04$ Erlang. As we see in Table 6.4, the traffic loads fluctuate from 5.0611 to 20.6474. Although the traffic load condition is satisfied for all the cells, some of them (e.g. c_{12}) require much more traffic channels than the others. This could heavily influence the spectral efficiency. Cell dimensioning can be applied to solve this problem.

In cellular mobile systems, the size of a cell depends on several parameters such as the transmitted power, antenna height and orientation. In this work, an experiment was made on cell dimensioning by adjusting the transmitted power. Although altering the antenna height and orientation (for directional antenna) can also change the cell size, these aspects are not considered at this moment.

The seven cells listed in Table 6.4 are tested in this experiment i.e. $C' = \{c_4, c_6, c_9, c_{10}, c_{12}, c_{20}, c_{26}\}$. Initially, the transmitted power (represented by EIRP) is 50 dBm for all the cells. The task is to change the transmitted power of each cell, and therefore the cell size, such that the coverage rate is maximized and the spectral cost is minimized. Here, two objective functions are considered simultaneously:

$$\begin{aligned} \text{maximize} \quad & f_1(C') = R_{trf}(C'), \\ \text{minimize} \quad & f_2(C') = M_{req}(C'). \end{aligned}$$

The traffic coverage rate $R_{trf}(C')$ and spectral cost $M_{req}(C')$ are given by (3.15) and (3.23), respectively.

The NSGA is applied again, but now an integer vector, instead of a bit-string, is used to represent a solution. Each element of the vector corresponds to the transmitted power of a cell. The range of the transmitted power is given as $0 \leq p_i \leq 80$. Simulated binary crossover and Gauss mutation (see Subsection 4.3.1) are employed by the real-

valued GA. The GA parameters are the same as those used in the above cell selection experiment. The target region is set as $0.97 \leq f_1(C') \leq 1$ for the coverage rate and $1 \leq f_2(C') \leq 15$ for the spectral cost in order to emphasize better solutions. One of the Pareto optimal solutions is chosen, which has the following objective scores: $R_{rr}(C') = 0.9732$ and $M_{req}(C') = 12$.

Comparing with the results of the cell selection experiment, it is clearly shown that better scores are achieved after cell dimensioning. The transmitted power and other performance of the selected cells are listed in Table 6.5. The traffic loads slightly fluctuate from 12.5719 to 16.8373, and none of the cells is saturated. This obviously improves the spectral efficiency. Besides, the interference condition is also improved. The interference probability decreases for some cells, such as c_6 , c_{10} , c_{12} , and c_{20} .

Table 6.5 Performance of the selected cells after cell dimensioning

Cell no.	Transmitted power (dBm)	Traffic load (Erlang)	Number of traffic channels	Interference probability
4	37	14.7798	22	0.04
6	51	13.4831	21	0.05
9	33	14.1258	22	0.04
10	58	12.5719	20	0.03
12	32	15.7442	23	0.04
20	42	16.8373	25	0.03
26	54	12.8417	20	0.05

6.2.3 Automatic base station placement and dimensioning

Automatic base station placement and dimensioning (ABSPAD) is a very hard problem due to the exponentially expanding search space. In [Huang(2000c)], a hierarchical approach was proposed to solve this problem. It has been demonstrated by the BS placement simulation in Section 5.4 that the hierarchical approach can effectively reduce the computational time. However, things become more complicated in realistic design scenarios. Thus the hierarchical approach is somewhat different with that one used in DGMs. In the following, we will discuss the approach for solving ABSPAD and present the experimental results obtained.

A Hierarchical approach workflow

The hierarchical approach workflow is shown in Figure 6.5. It starts from a course resolution and finds potential cells on an initial coarse grid. For each of the potential cells, field strength is calculated using the propagation model. This is followed by the multi-objective optimization process. Here, the integrated search and decision making method is applied to guide the search towards the global Pareto optimal front and to find a satisfactory tradeoff solution. Afterwards, cell dimensioning is performed to optimize the parameters of the selected cells. Next, the grid resolution is increased and the previous selected cells are used to search potential cells on a finer grid. As the resolution increases from rough to fine, results become more and more accurate.

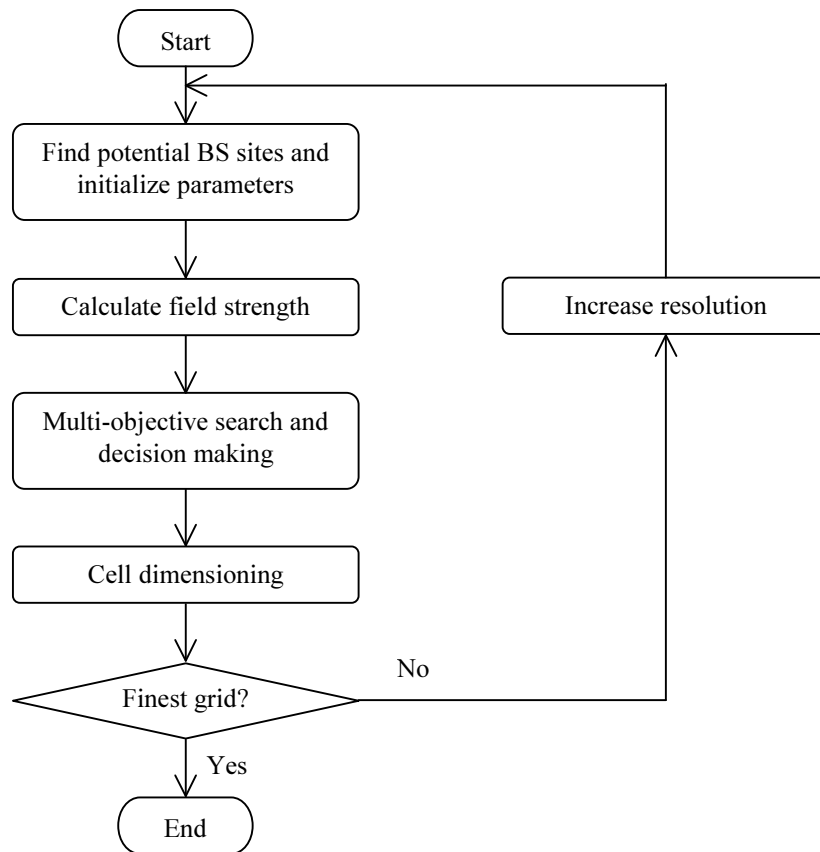


Figure 6.5 Workflow of the hierarchical approach for solving ABSPAD.

B Geographical cost

In ABSPAD, BSs are possibly located at any place in the service area. However, there are some places unable to be used e.g. locations in water, or those without permission e.g. the White House. In order to avoid such unsuitable situations, we introduce a geographical cost function $f_g(x_i, y_i)$ that evaluates how good a place is for building a

BS. The geographical cost is determined based on the information about topography, morphology, and existing infrastructure. The forbidden locations are set with very high cost such that it is impossible to place a BS in water or on the roof of the White House. On the other hand, network operators may have their own preference for some available sites. Such preferred locations are set with low cost. The geographical cost function can be defined by the network operator in the form of raster data.

The geographical cost is integrated into the financial cost, such that it can take effect in the automatic process. The financial cost function given by (3.24) is then modified as follows:

$$E_{fin}(C) = E_0 + \sum_{c_i \in C} \left(\frac{f_g(x_i, y_i)}{\theta_g} \cdot E_{site}(c_i) + E_{eq}(c_i) + E_m(c_i) \right) \quad (6.1)$$

where E_0 , E_{site} , E_{eq} , and E_m are just the same as defined in (3.24). (x_i, y_i) represents the site of cell c_i . θ_g is a threshold. A location is regarded as suitable for a BS if its geographical cost is lower than θ_g .

C Breadth-first search

The BS placement simulation in DGMs has demonstrated that the hierarchical approach can effectively reduce the computational complexity because it continuously eliminates unfit cells and searches potential cell sites only in promising areas. Since in DGMs cells are placed in a plane, the hierarchical approach works very well for that simulation. If real topography is used, however, this approach could cause the automatic process to be trapped in a local optimum as the available area for cell sites shrinks step by step. Especially, in case a mountainous area (e.g., Switzerland) is considered, more and more BSs are needed in order to achieve satisfactory coverage rate. What is even worse is that these BSs are situated in clusters. As a consequence, there are very large overlapping areas between them and the resulting interference makes this solution impracticable.

To overcome this problem, a backward search method was proposed in [Huang(2000b)]. This method continuously predicts the performance of the next level and measures the possibility of being trapped in a local optimum. If it seems to happen, the algorithm goes backward to explore new promising areas for more cells so as to prevent the automatic process from stopping at the local optima. Such a method can be classified as a depth-first search method because the search resumes from the previous level.

As an alternative, a breadth-first search method can also be used. If the search is blocked (i.e. it may be trapped in a local optimum), more potential cells will be placed on the next level. As shown in Figure 6.6, the pyramid structure is used to illustrate

the breadth-first search. Now each cell of the parent level will have $(2n)^2$ child cells, where n is a natural number. Thus, for each parent cell, the number of child cells could be 4, 16, 36, 64, and so forth. By adding more potential cells, new areas in the search space will be explored so as to make the search jump out of the local optima. In the following experiment, a hierarchical approach with the breadth-first search is used, where the parameter n should be specified by the decision maker based on the tradeoff information obtained.

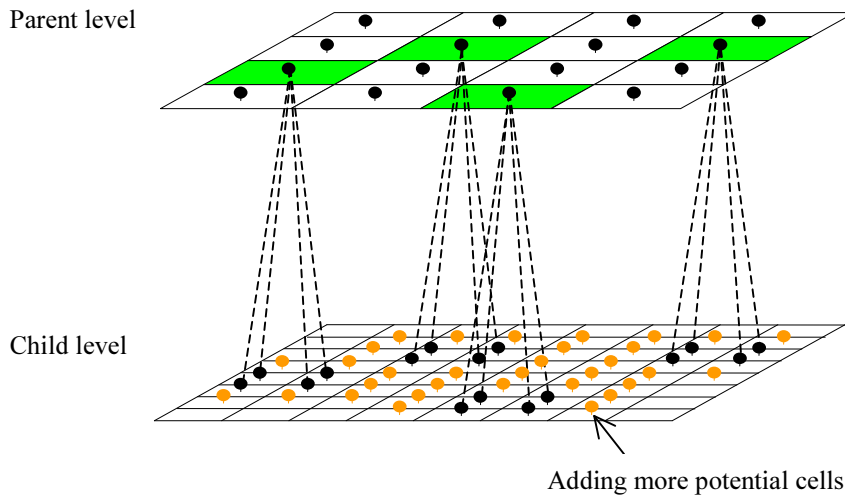


Figure 6.6 Breadth-first search method used in the hierarchical approach in order to avoid local optima.

D Simulation results

A numerical experiment was made for ABSPAD. The geographical data shown in Figure 6.1 is used. The grid resolution ranges from 6,400 m to 100 m, thus there are seven levels in total. On each level, the multi-objective search and decision making is applied and one of the tradeoff solutions is chosen by the decision maker. The following three objectives are considered simultaneously:

$$\begin{aligned} &\text{maximize} && f_1(C) = R_{nr}(C), \\ &\text{minimize} && f_2(C) = M_{req}(C), \\ &\text{minimize} && f_3(C) = E_{fin}(C). \end{aligned}$$

Objective scores of the chosen solutions according to the grid resolutions are shown in Figure 6.7, where the number of BSs is used as the financial cost simply because the BSs are assumed to be of equal cost.

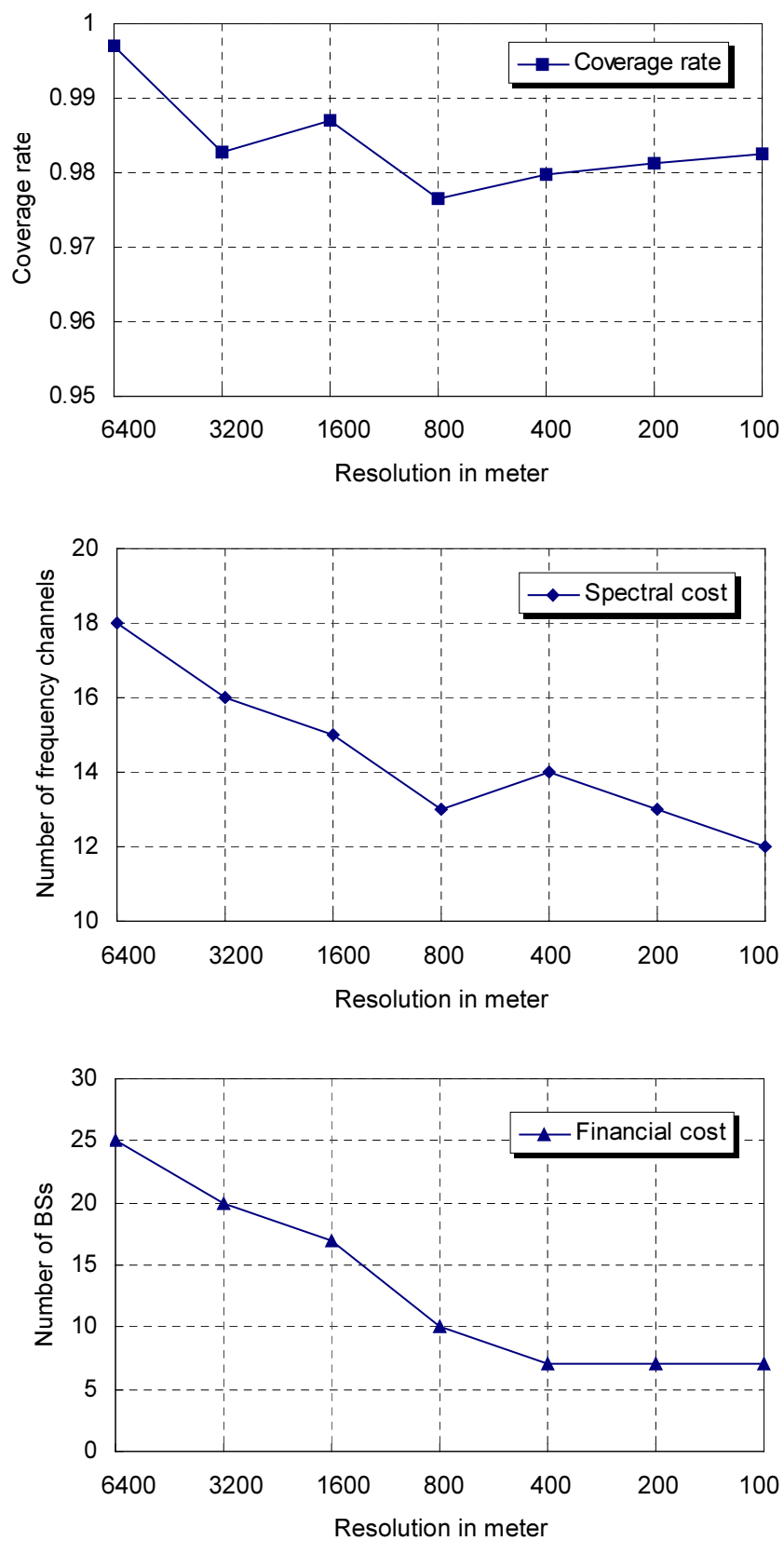


Figure 6.7 Objective scores of the solutions obtained by the hierarchical approach. There are seven levels with the grid resolutions from 6,400 m to 100 m.

On the first level, potential cells are placed on a grid with the resolution of 6,400 m, and hence, there are 100 potential cells in total. For the first 10 generations, the target region is set as $0.90 \leq f_1(C) \leq 1$ and $1 \leq f_2(C) \leq 30$. Later on it is modified as $0.95 \leq f_1(C) \leq 1$ and $1 \leq f_2(C) \leq 20$. After 100 generations, a solution with 25 cells is chosen. The coverage rate and spectral cost are 0.9969 and 18, respectively.

On the next level, the grid resolution increases to 3,200 m. Each selected cell is split into 4 subcells. The multi-objective search and decision making is applied again. The same process is repeated in the following levels. On the level of 800 m, the breadth-first search is performed with the parameter $n = 2$, such that each cell has 16 subcells. After this level the target region is set as $0.975 \leq f_1(C) \leq 1$ and $1 \leq f_2(C) \leq 15$, in order to emphasize the desired solutions. On the finest grid with the resolution of 100 m, a solution with 7 cells is chosen from the Pareto optimal solutions obtained. The objective scores it achieves are as follows: $R_{rf}(C) = 0.9823$, $M_{req}(C) = 12$.

Performance of the selected cells in the final solution is listed in Table 6.6. The corresponding coverage areas of these cells are shown in Figure 6.8. Since the optimization objectives are the same as those of the cell selection, ABSPAD also places BSs at high positions because they provide better coverage. This is why the solution obtained by ABSPAD looks similar to that one in the cell selection experiment (see Figure 6.4), but ABSPAD achieves better performance because it searches for the optimal cell set in the whole service area instead of only a number of potential cells.

Besides, it is worth mentioning that, due to the geographical cost, no BS is placed in water in the final solution. In this experiment, we set the threshold $\theta_g = 1$, and the geographical cost of water is 100. Therefore, if a solution involves any BSs located in water, it will result in a very high cost. Thus it can not be included in the non-dominated set.

Table 6.6 Performance of the cells in the final solution of the ABSPAD experiment

Cell no.	Transmitted power (dBm)	Traffic load (Erlang)	Number of traffic channels	Interference probability
1	39	13.6984	21	0.04
2	34	14.8219	22	0.04
3	54	13.8551	21	0.03
4	31	14.8530	22	0.05
5	52	15.3801	23	0.05
6	32	14.7746	22	0.04
7	49	13.9465	21	0.03

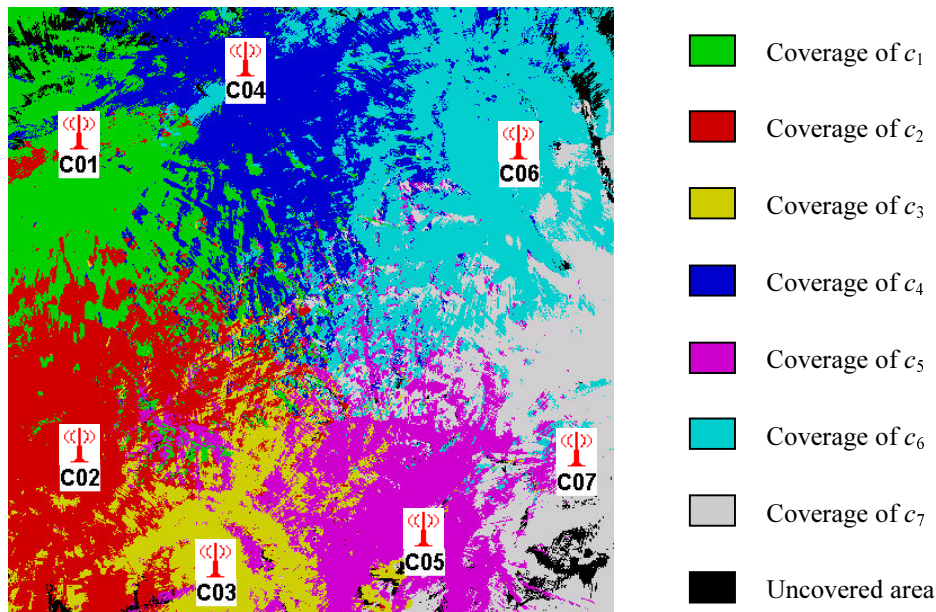


Figure 6.8 Coverage areas of the cells obtained by ABSPAD.

6.2.4 Growth planning

A Heuristic algorithm

Until now we consider only one cell per site and the cell size is as large as possible. What happens if the traffic demand increases? What can we do if the existing cells can not afford the scaled-up traffic? This is the so-called growth planning problem, which has already been described in Subsection 3.2.4. In the following, an algorithm will be presented for solving this problem, which is a heuristic algorithm based on the principle of capacity-oriented cell splitting (see Subsection 2.1.1).

Nominal cell splitting is illustrated in Figure 6.9. Initially largest possible cell sizes are used, one cell per site. In the next step, a cell is divided into a number of sectors, three and six being the most common arrangements. Here the 3-sector case is considered, but the method can also be used for the 6-sector case. Each of the sectors is served by a different set of channels and illuminated by a directional antenna. The sector can therefore be considered as a new cell. As a consequence, there are 3 cells per site using the original BS sites. BSs are located at the corner of cells, as shown in Figure 6.9(b). Now the number of BS sites is the same, but the number of cells is three times larger than before. The following step is to do further cell splitting i.e. reducing the size of existing cells and adding new cells. As we see in Figure 6.9(c), the former sites are still used in the new cell plan, but additional sites are now required for serving new cells.

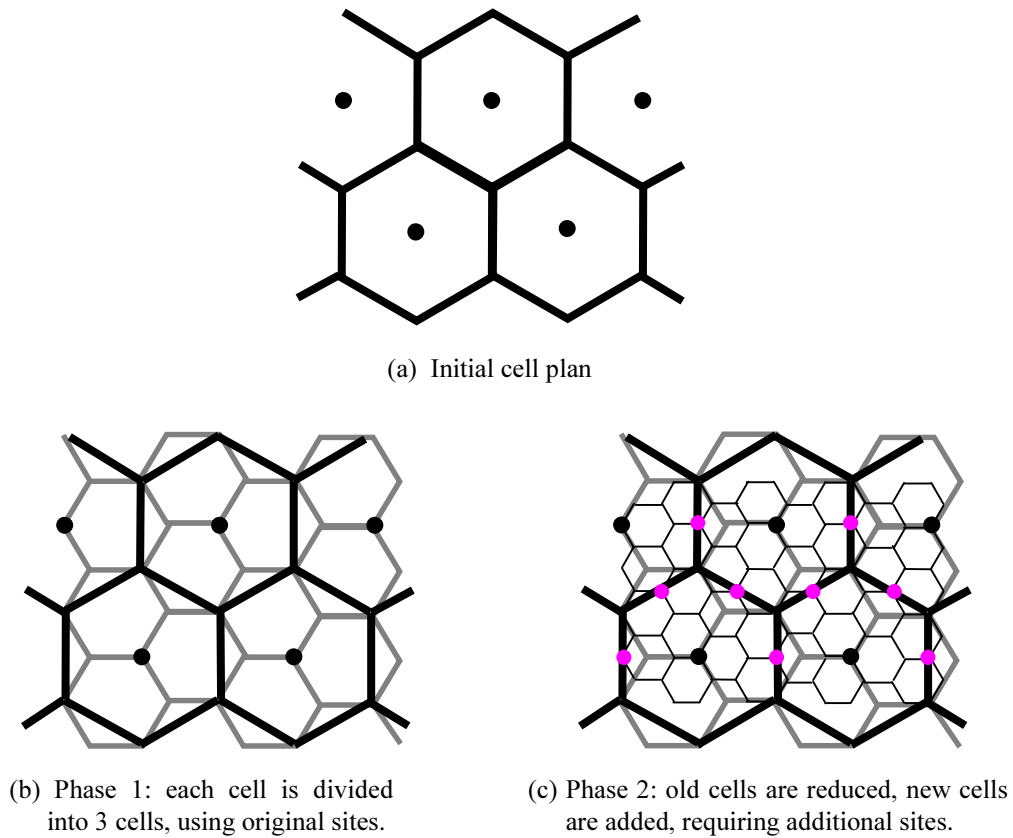


Figure 6.9 Nominal cell splitting (adapted from [Bäck(1991)]).

Based on the cell splitting technique, an algorithm for automatic growth planning is developed. The main steps of the algorithm are outlined as follows:

Algorithm 6.1 (Growth planning)

Input: C – Original cell set $C = \{c_1, c_2, \dots, c_N\}$
 T_{max} – Maximum traffic load
 θ_{int} – Interference probability threshold

Output: New cell set $C' = \{c'_1, c'_2, \dots, c'_N\}$, $N' \geq N$

Step 1:

Calculate the local condition functions of each cell according to (3.2) and (3.6);

Pick out all unfit cells c_i if $A_{trf}(c_i) > T_{max}$ or $P_{int}(c_i) > \theta_{int}$.

Step 2:

Choose the cell c_i with the greatest traffic load, i.e., $A_{trf}(c_i)$ is greater than the traffic load of other cells.

Step 3:

If c_i has no co-site cells, divide it into three sectors (cell splitting phase 1);

Otherwise, reduce the cell size such that $A_{Tf}(c_i) \leq T_{max}$, and apply the hierarchical approach to find optimal sites for new cells (cell splitting phase 2).

Step 4:

Assign frequencies for new cells.

Step 5:

Go to Step 1 and repeat the above steps until no unfit cells exist.

Since the cells influence each other with respect to the coverage and interference, values of the local condition functions must be recalculated after any cell has been changed. In Step 3, cell splitting is executed. Only the vicinity of c_i is taken into account when searching for new cell sites so as to save CPU time. Besides, during the cell splitting, neighboring cells of c_i must be respected because they exert an influence on the new cells.

B Simulation results

This algorithm is used in the following experiment. The service area is the same as that in the above ABSPAD experiment, but having much higher traffic demand. Now there are 411.0955 Erlang traffic, about three times more than before (103.1516 Erlang). The original cell set consists of the seven cells found in the ABSPAD experiment. Due to the growth in traffic demand, the traffic loads of these cells increase greatly. Each cell now covers more than 50 Erlang traffic. This results in very high blocking probability ($> 45\%$). Obviously, the existing cells are not able to carry the new scaled-up traffic. Growth planning is inevitable.

Algorithm 6.1 is applied with $T_{max} = 21.04$ Erlang and $\theta_{int} = 0.1$. In the first step, each cell is divided into three sectors such that there are 21 cells, but some of them still have very high traffic load. Therefore, the cell size must be reduced and new cells are required to cover the gaps. Three new BS sites are found by using the hierarchical approach. Their positions are shown in Figure 6.10(a), marked as C08 – C10.

In the final solution, 10 BS sites are used and each site serves three cells with directional antennas. Hence there are 30 cells. Their coverage areas are shown in Figure 6.10(b). All the cells meet the requirement of traffic load and interference probability. The objective scores of the final solution are as follows: the coverage rate is 0.9834, and the spectral cost is 27. Since there are three cells for each BS site, the co-site channel separation has to be taken into account. As a result, more frequency channels are required. Finally 10 BSs are used, but seven of them are already in use.

Only three new BSs need to be built.

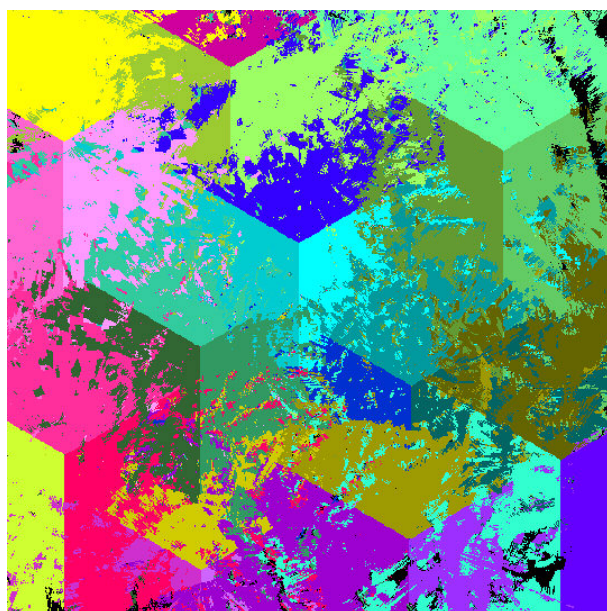
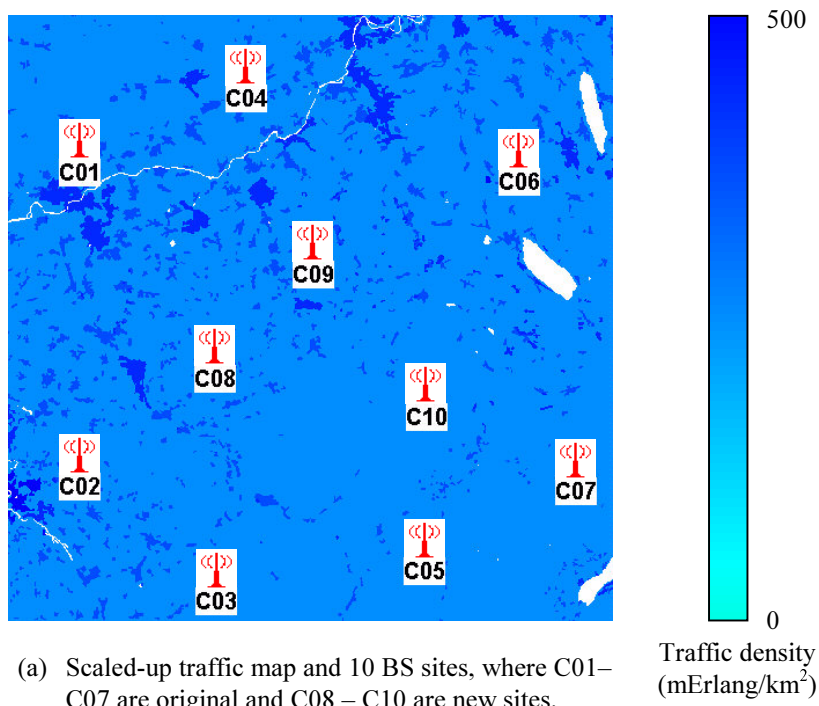


Figure 6.10 Experimental results of growth planning in an existing network. The seven original cells are obtained from the ABSPAD experiment.

6.3 SUMMARY

In this chapter, cell planning experiment results were presented and analyzed. Realistic geographical data are used in these experiments, which include the topographical data, morphological data, and traffic data. Field strength calculations were done by employing the COST 231 Okumura-Hata model. A database was developed to manage the input data and optimization results.

Cell selection was first considered, where an optimal cell set was selected from a number of potential cells. Cell dimensioning was applied to optimize the cell configurations such that the system performance could be improved. Another experiment was made for ABSPAD, which is able to find the optimal BSs sites and simultaneously determine the optimal parameters of each cell. Finally, the growth planning problem has been considered. A heuristic algorithm based on the technique of cell splitting was presented. It is demonstrated by the experimental results that the methodology and algorithms presented in this work can effectively solve the cell planning problems arising from mobile network design practice.

Chapter 7

Conclusions

The goal of this thesis is the development of an efficient optimization strategy to solve the cell planning problems arising from mobile network design practice. These problems are NP-hard and involve numerous constraints. In this work, several original contributions have been produced to achieve the research goal. Major contributions are summarized in Section 7.1. Based on the results of this research, Section 7.2 identifies promising research areas within which further work may be beneficial.

7.1 SUMMARY OF CONTRIBUTIONS

The main contributions of this thesis are:

Definition and classification of cell planning problems. For the first time, the cell planning problem is defined and analyzed from the viewpoint of system performance optimization. Four kinds of problems are distinguished: cell selection, cell dimensioning, automatic base station placement and dimensioning (ABSPAD), and growth planning. Experiments were made for all the four problems.

Local conditions and global conditions. Cellular system performance criteria are addressed by local conditions and global conditions. Local conditions, including the traffic load and interference probability, are defined on each single cell, while global conditions are defined on the whole system (a set of cells), including the coverage rate, spectral cost, and financial cost. The global conditions are used as optimization objectives and the local conditions are taken as hard constraints.

Integrated multi-objective search and decision making. Cell planning is a typical multi-objective optimization problem as several objectives need to be considered simultaneously. After analyzing the insufficiency of classical methods, we propose an approach for solving multi-objective problems by integrating the search and decision making processes.

Disk graph models (DGMs). For the first time, DGMs are used to study cell planning problems. The advantage of these models is that they are independent

of geographical data and no field strength calculation is needed. Thus we can concentrate on investigating algorithms. A number of test problems were presented and simulations were made for the purpose of algorithm study.

Application of genetic algorithms (GAs) to cell planning problems. GAs are particularly well-suited to multi-objective optimization because of their ability to find multiple Pareto optimal solutions in a single run. The elitist NSGA is applied to the cell planning problems. Various GA parameters are investigated in order to improve the algorithm performance. Besides, real-valued GAs are applied to the cell dimensioning problem, in comparison with binary GAs. Simulation results demonstrate that real-valued GAs are more convenient and may achieve better performance for continuous parameter optimization.

Hierarchical approach. A hierarchical approach is proposed to solve the ABSPAD that is intractable due to the computational complexity. A geographical cost is introduced in order to avoid unsuitable locations such as those in water, and a breadth-first search method is used to prevent the automatic process from stopping at local optima. Simulation results show that the hierarchical approach can not only reduce the computational time but also obtain better solutions.

Growth planning algorithm. The growth planning problem has also been studied. A novel heuristic algorithm based on the technique of cell splitting is presented to solve this problem. Experiments were made for automatic growth planning. It is demonstrated that this algorithm is able to find optimal sites for new cells and optimize parameters of both original cells and additional cells, such that the growing service demand can be met effectively while offering minimum disturbance to the existing network.

As a whole, the present work is largely original and should be of great interest to the mobile communications system design community and operations research society, as well as other engineering design fields.

7.2 FUTURE WORK

Several promising areas for further work have arisen from this research concerning multi-objective genetic algorithms (MOGAs), multi-objective decision making, and cellular system design.

The MOGA should be further improved and made more efficient. This is an important factor when dealing with large-scale problems. Parallelization of the algorithm could provide great enhancement. Since the objective functions are computationally intensive for cell planning problems, the computational time would be significantly reduced if individuals can be evaluated in parallel. Performance assessment of multi-objective optimization algorithms is also an interesting topic. In

general, three criteria should be taken into account: i) minimal distance to the global Pareto optimal front, ii) good distribution of the solutions found, iii) maximum spread of the obtained non-dominated front.

The method for integrated multi-objective search and decision making has been demonstrated to be superior to the other methods. However, it requires increased interaction between the decision maker and the optimizer. In case of a human decision maker, the need for constant human intervention may be actually a disadvantage. Moreover, if more than three objectives are involved, the difficulty of showing the tradeoff surface becomes apparent. This may heavily influence the human decision maker. Therefore, we suggest an expert system is developed to integrate the experience of human decision makers, such that man-machine interaction is not mandatory.

Although this thesis has investigated many aspects of cell planning, a lot still remains to be done in this area. Besides macrocells, experiments for microcells and indoor picocells should be made. For the purpose of investigation of various algorithms, a benchmark test problem suite needs to be built in consideration of different design scenarios. Further development can be done for cell dimensioning taking into account more parameters, such as the antenna height, type, azimuth, and downtilt, in addition to the transmitted power. As the mobile communications system evolves, the methodology and algorithms presented in this work can be applied to 3G systems (e.g., UMTS) and 3G beyond systems.

Appendix A

Abbreviations and acronyms

2D	two-dimensional
3D	three-dimensional
3G	third-generation
ABSPAD	automatic base station placement and dimensioning
ACI	adjacent channel interference
BER	bit error rate
BS	base station
C/I	carrier-to-interference ratio
CAP	channel assignment problem
CCH	control channel
CCI	co-channel interference
CDMA	code division multiple access
COST	European Cooperation in the Field of Scientific and Technical Research
CPU	central processing unit
DCA	dynamic channel assignment
DGM	disk graph model
EA	evolutionary algorithm
EIRP	effective isotropic radiated power
EP	evolutionary programming
erf	error function
ESs	Evolutionsstrategies, or evolution strategies
ETSI	European Telecommunications Standard Institute
FAP	frequency assignment problem
FCA	fixed channel assignment
FCH	frequency channel
FDMA	frequency division multiple access
GA	genetic algorithm
GIS	geographic information systems
GOS	grade of service

GSM	global system for mobile communications
GTD	geometrical theory of diffraction
HCA	hybrid channel assignment
IEEE	the Institute of Electrical and Electronics Engineers
iff	if and only if
LOS	line-of-sight
MOGA	multi-objective genetic algorithm
MOP	multi-objective optimization problem
MS	mobile station
MSC	mobile switching center
NLOS	non-line-of-sight
NP	nondeterministic polynomial
NSGA	non-dominated sorting genetic algorithm
PC	personal computer
PDA	personal digital assistant
pdf	probability density function
PL	path loss
RAM	random access memory
Rx	receiver
SA	simulated annealing
SBX	simulated binary crossover
SOP	single objective optimization problem
SP	sequential packing
SPEA	strength Pareto evolutionary algorithm
TCH	traffic channel
TDMA	time division multiple access
TSP	travelling salesman problem
TRX	transceiver
Tx	transmitter
UDGM	unit disk graph model
UMTS	universal mobile telecommunications system
UTD	uniform geometrical theory of diffraction

Bibliography

- [Aarts(1989)] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons.
- [Anderson(1994)] H. R. Anderson, and J. P. McGeehan, "Optimizing microcell base station locations using simulated annealing techniques," in *Proceedings of the 44th IEEE Vehicular Technology Conference (VTC'94)*, pp. 858–862.
- [Bäck(1991)] M. Bäck and A. Frisell, *Cell Planning CME20*, Training Document, Ericsson Radio System AB.
- [Bäck(1993)] T. Bäck, "Optimal mutation rates in genetic algorithms," in *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp. 2–8, Morgan Kaufmann.
- [Badsberg(1995)] M. Badsberg, J. B. Andersen, and P. Mogensen, "Exploitation of the terrain profile in the Hata model," *COST 231 TD(95)9*.
- [Bagchi(1999)] T. P. Bagchi, *Multiobjective Scheduling by Genetic Algorithms*, Kluwer Academic Publishers.
- [Becker(1999)] T. C. Becker, F. Küchen, and W. Wiesbeck, "A new network planning approach for digital audio broadcasting," in *IEEE Transactions on Vehicular Technology*, vol. 48, no. 2, pp. 619–626.
- [Beckmann(1999)] D. Beckmann and U. Killat, "A new strategy for the application of genetic algorithms to the channel assignment problem," in *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1261–1269.
- [Berg(1995)] J. E. Berg, "A recursive method for street microcell path loss calculations," in *Proceedings of the 6th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'95)*, pp. 140–143.
- [Box(1978)] F. Box, "A heuristic technique for assigning frequencies to radio nets," in *IEEE Transactions on Vehicular Technology*, vol. 27, no. 2, pp. 57–64.
- [Calégari(1997)] P. Calégari, F. Guidec, P. Kuonen, and D. Wagner, "Genetic approach to radio network optimization for mobile systems,"

- in *Proceedings of the 47th IEEE Vehicular Technology Conference (VTC'97)*, pp. 755–759.
- [Cátedra(1999)] M. F. Cátedra and J. Pérez-Arriaga, *Cell Planning for Wireless Communications*, Artech House.
- [Chamaret(1997)] B. Chamaret, S. Josselin, P. Kuonen, M. Pizarroso, B. Salas-Manzanedo, S. Ubeda, and D. Wagner, “Radio network optimization with maximum independent set search,” in *Proceedings of the 47th IEEE Vehicular Technology Conference (VTC'97)*, pp. 770–774.
- [Chan(1994)] P. T. H. Chan, M. Palaniswami, and D. Everitt, “Neural network-based dynamic channel assignment for cellular mobile communication systems,” in *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 279–288.
- [Cheung(1994)] J. C. S. Cheung, M. A. Beach, and J. P. McGeehan, “Network planning for third-generation mobile radio systems,” in *IEEE Communications Magazine*, vol. 32, no. 11, pp. 54–59.
- [Cichon(1994)] D. J. Cichon, (in German) *Strahlenoptische Modellierung der Wellenausbreitung in urbanen Mikro- und Pikofunkzellen*, Ph.D. thesis, Faculty of Electrical Engineering, University of Karlsruhe, Germany.
- [Clark(1990)] B. N. Clark, C. J. Colbourn, and D. S. Johnson, “Unit disk graphs,” in *Discrete Mathematics*, vol. 86, pp. 165–177.
- [Coello(1999)] C. A. C. Coello, “An updated survey of evolutionary multiobjective optimization techniques: state of the art and future trends,” in *1999 Congress on Evolutionary Computation (CEC99)*, pp. 3–13.
- [COST 231(1991)] COST 231, “Urban transmission loss models for mobile radio in the 900 and 1800 MHz bands (revision 2),” *COST 231 TD(90)119, Rev. 2*.
- [COST 231(1999)] COST 231, “Digital mobile radio towards future generation systems,” *COST 231 Final Report*.
- [Cvetković(2000)] D. Cvetković, *Evolutionary Multi-objective Decision Support Systems for Conceptual Design*, Ph.D. thesis, University of Plymouth, UK.
- [Davis(1991)] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand.
- [De Jong(1975)] K. A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*, Ph.D. thesis, University of Michigan, USA.

- [Deb(1989)] K. Deb and D. E. Goldberg, "An investigation of niche and species formation in genetic function optimization," in *Proceedings of the Third International Conference on Genetic Algorithms*, pp. 42–50.
- [Deb(1995)] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," in *Complex Systems*, vol. 9, pp. 115–148.
- [Deb(1999)] K. Deb, "Evolutionary algorithms for multi-criterion optimization in engineering design," in *Evolutionary Algorithms in Engineering and Computer Science*, Chap. 8, pp. 135–161, John Wiley & Sons.
- [Deb(2000a)] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pp. 849–858.
- [Deb(2000b)] K. Deb and T. Goyal, "Controlled Elitist Non-dominated Sorting Genetic Algorithms for Better Convergence," KanGAL report 200004, Indian Institute of Technology, Kanpur, India.
- [Deygout(1966)] J. Deygout, "Multiple knife-edge diffraction of microwaves," in *IEEE Transactions on Antennas and Propagation*, vol. 14, No. 4, pp. 480–489.
- [Didascalou(2000)] D. Didascalou, *Ray-Optical Wave Propagation Modelling in Arbitrarily Shaped Tunnels*, Ph.D. thesis, Faculty of Electrical Engineering, University of Karlsruhe, Germany.
- [Döttling(1997)] M. Döttling, F. Küchen, and W. Wiesbeck, "Deterministic modeling of the street canyon effect in urban micro and pico cells," in *Proceedings of the IEEE International Conference on Communications (ICC'97)*, pp. 36–40.
- [Duque-Anton(1993)] M. Duque-Anton, D. Kunz, and B. Rueber, "Channel assignment for cellular radio using simulated annealing," in *IEEE Transactions on Vehicular Technology*, vol. 42, no. 1, pp. 14–21.
- [Epstein(1953)] J. Epstein and D. Peterson, "An experimental study of wave propagation at 850 MHz," in *Proceedings of IEEE*, vol. 41, No. 5, pp. 595–611.
- [ETSI 03.30(2000)] ETSI, "Radio network planning aspects," *GSM Recommendations 03.30 Version 8.3.0 Release1999*.

- [ETSI 05.05(2000)] ETSI, "Radio transmission and reception," *GSM Recommendations 05.05 Version 8.5.0 Release1999*.
- [Fleury(1996)] B. H. Fleury and P. E. Leuthold, "Radio wave propagation in mobile communications: an overview of European research," in *IEEE Communications Magazine*, vol. 34, no. 2, pp. 70–81.
- [Fogel(1995)] D. B. Fogel, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, IEEE Press.
- [Fonseca(1993)] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multiobjective optimization: formulation, discussion and generalization," in *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp. 416–423, Morgan Kaufmann.
- [Fonseca(1995a)] C. M. Fonseca, *Multiobjective Genetic Algorithms with Application to Control Engineering Problems*, Ph.D. thesis, The University of Sheffield, UK.
- [Fonseca(1995b)] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," in *Evolutionary Computation*, vol. 3, no. 1, pp. 1–16.
- [Fonseca(1998a)] C. M. Fonseca and P. J. Fleming, "Multiobjective optimization and multiple constraint handling with evolutionary algorithms – Part I: a unified formulation," in *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, vol. 28, no. 1, pp. 26–37.
- [Fonseca(1998b)] C. M. Fonseca and P. J. Fleming, "Multiobjective optimization and multiple constraint handling with evolutionary algorithms – Part II: application example," in *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, vol. 28, no. 1, pp. 38–47.
- [Funabiki(1992)] N. Funabiki and Y. Takefuji, "A neural network parallel algorithm for channel assignments in cellular radio networks," in *IEEE Transactions on Vehicular Technology*, vol. 41, no. 4, pp. 430–437.
- [Gamst(1982)] A. Gamst and W. Rave, "On frequency assignment in mobile automatic telephone systems," in *Proceedings of IEEE Globecom '82*, pp. 309–315.
- [Gamst(1986)] A. Gamst, "Some lower bounds for a class of frequency assignment problems," in *IEEE Transactions on Vehicular Technology*, vol. 35, no. 1, pp. 8–14.

- [Garey(1979)] M. R. Garey and D. S. Johnson, *Computers and Intractability: a Guide to the Theory of NP-Completeness*, W. H. Freeman and Company.
- [Geng(1998)] N. Geng and W. Wiesbeck, (in German) *Planungsmethoden für die Mobilkommunikation: Funknetzplanung unter realen physikalischen Ausbreitungsbedingungen*, Springer-Verlag.
- [Goldberg(1989)] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.
- [Green(1990)] E. Green, "Radio link design for microcellular systems," in *British Telecom Technology Journal*, vol. 8, no. 1, pp. 85–96.
- [Haas(1997)] Z. J. Haas, J. H. Winters, and D. S. Johnson, "Simulation results of the capacity of cellular systems," in *IEEE Transactions on Vehicular Technology*, vol. 46, no. 4, pp. 805–817.
- [Hale(1980)] W. K. Hale, "Frequency assignment: theory and applications," in *Proceedings of IEEE*, vol. 68, No. 12, pp. 1497–1514.
- [Hancock(1994)] P. J. B. Hancock, "An empirical comparison of selection methods in evolutionary algorithms," in *Evolutionary Computing: AISB Workshop*, pp. 80–94, Springer-Verlag.
- [Hao(1997)] Q. Hao, B.-H. Soong, E. Gunawan, J.-T. Ong, C.-B. Soh, and Z. Li, "A low-cost cellular mobile communication system: a hierarchical optimization network resource planning approach," in *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 7, pp. 1315–1326.
- [Hata(1980)] M. Hata, "Empirical formula for propagation loss in land mobile radio services," in *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325.
- [Haupt(1998)] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, John Wiley & Sons.
- [Hifi(1997)] M. Hifi, "A genetic algorithm-based heuristic for solving the weighted maximum independent set and some equivalent problems," in *Journal of the Operational Research Society*, vol. 48, no. 6, pp. 612–622.
- [Holland(1975)] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press.
- [Holma(2000)] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, John Wiley & Sons.

- [Horn(1994)] J. Horn, N. Nafpliotis, and D. E. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization," in *Proceedings of the First IEEE Conference on Evolutionary Computation (ICEC'94)*, pp. 82–87.
- [Horn(1997)] J. Horn, "Multicriterion decision making," in *Handbook of Evolutionary Computation*, vol. 1, F1.9:1–F1.9:15, IOP Publishing Ltd. and Oxford University Press.
- [Huang(2000a)] X. Huang, U. Behr, and W. Wiesbeck, "A new approach to automatic base station placement in mobile networks," in *Proceedings of International Zurich Seminar on Broadband Communications*, pp. 301–306.
- [Huang(2000b)] X. Huang, U. Behr, and W. Wiesbeck, "Improved algorithm for automatic cell planning in mobile networks," in *Proceedings of European Wireless 2000*, pp. 91–96.
- [Huang(2000c)] X. Huang, U. Behr, and W. Wiesbeck, "Automatic base station placement and dimensioning for mobile network planning," in *Proceedings of the 52th IEEE Vehicular Technology Conference (VTC2000-Fall)*, pp. 1544–1549.
- [Huang(2000d)] X. Huang, U. Behr, and W. Wiesbeck, "Automatic cell planning for a low-cost and spectrum efficient wireless network," in *Proceedings of IEEE GLOBECOM 2000*, pp. 276–282.
- [Hurley(2000)] S. Hurley, "Automatic base station selection and configuration in mobile networks," in *Proceedings of the 52th IEEE Vehicular Technology Conference (VTC2000-Fall)*, pp. 2585–2592.
- [Ibbetson(1997)] L. J. Ibbetson, and L. B. Lopes, "An automatic base site placement algorithm," in *Proceedings of the 47th IEEE Vehicular Technology Conference (VTC'97)*, pp. 760–764.
- [Ikegami(1984)] F. Ikegami, S. Yoshida, T. Takeuchi, and M. Umehira, "Propagation factors controlling mean field strength on urban streets," in *IEEE Transactions on Antennas and Propagation*, vol. 32, No. 12, pp. 822–829.
- [Ingber(2000)] L. Ingber, *The Simulated Annealing Archive*, <http://www.ingber.com/>.
- [Johri(1994)] P. K. Johri, "An insight into dynamic channel assignment in cellular mobile communications systems," in *European Journal of Operational Research*, vol. 74, no. 1, pp. 70–77.

- [Katzela(1996)] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey," in *IEEE Personal Communications*, vol. 11, no. 7, pp. 10–31.
- [Kim(1997)] J. S. Kim, S. H. Park, P. W. Dowd, and N. M. Nasrabadi, "Cellular radio channel assignment using a modified Hopfield network," in *IEEE Transactions on Vehicular Technology*, vol. 46, no. 4, pp. 957–967.
- [Kirkpatrick(1983)] S. Kirkpatrick, C. D. Gellatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," in *Science*, vol. 220, no. 4598, pp. 671–680.
- [Koski(1984)] J. Koski, "Multicriterion optimization in structural design," in *New Directions in Optimum Structural Design*, pp. 483–503, John Wiley & Sons.
- [Koster(1999)] A. M. C. A. Koster, *Frequency Assignment – Models and Algorithms*, Ph.D. thesis, Maastricht University, The Netherlands.
- [Kouyoumijan(1974)] R. G. Kouyoumijan and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," in *Proceedings of IEEE*, vol. 62, No. 11, pp. 1448–1461.
- [Küchen(1996)] F. Küchen, T. C. Becker, and W. Wiesbeck, "Optimizing the coverage area of single frequency networks," in *Proceedings of the International Broadcasting Convention*, pp. 236–241.
- [Küchen(1998)] F. Küchen, (in German) *Auf Wellenausbreitungsmodellen basierende Planung terrestrischer COFDM-Gleichwellennetze für den mobilen Empfang*, Ph.D. thesis, Faculty of Electrical Engineering, University of Karlsruhe, Germany.
- [Kunz(1991)] D. Kunz, "Channel assignment for cellular radio using neural networks," in *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 188–193.
- [Kürner(1993)] T. Kürner, D. J. Cichon, and W. Wiesbeck, "Concepts and results for 3D digital terrain based wave propagation models – an overview," in *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 7, pp. 1002–1012.
- [Kürner(1998)] T. Kürner, "Analysis of cell assignment probability predictions," *COST 259 TD(98)5*.

- [Lai(1996)] W. K. Lai and G. G. Coghill, "Channel assignment through evolutionary optimization," in *IEEE Transactions on Vehicular Technology*, vol. 45, no. 1, pp. 91–96.
- [Lee(1989)] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill.
- [Lee(1993)] W. C. Y. Lee, *Mobile Communications Design Fundamentals*, John Wiley & Sons.
- [Lee(1998)] J. S. Lee and L. E. Miller, *CDMA Systems Engineering Handbook*, Artech House.
- [Lebherz(1992)] M. Lebherz, W. Wiesbeck, and W. Krank, "A versatile wave propagation model for the VHF/UHF range considering three dimensional terrain," in *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 10, pp. 1121–1131.
- [Luebbers(1984)] R. J. Luebbers, "Finite conductivity uniform GTD versus knife edge diffraction in prediction of propagation path loss," in *IEEE Transactions on Antennas and Propagation*, vol. 32, No. 1, pp. 70–76.
- [MacDonald(1979)] V. H. MacDonald, "The cellular concept," in *The Bell System Technical Journal*, vol. 58, no. 1, pp. 15–41.
- [Mamdani(1974)] E. H. Mamdani, "Applications of fuzzy algorithms for simple dynamic plant," in *Proceedings of IEE*, vol. 121, no. 12, pp. 1585–1588.
- [Man(1999)] K. F. Man, K. S. Tang, and S. Kwong, *Genetic Algorithms: Concepts and Designs*, Springer-Verlag.
- [Marathe(1995)] M. V. Marathe, H. Breu, H. B. Hunt, S. S. Ravi, and D. J. Rosenkrantz, "Simple heuristics for unit disk graphs," in *Networks*, vol. 25, pp. 59–68.
- [Mathar(1993)] R. Mathar and J. Mattfeldt, "Channel assignment in cellular radio networks," in *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 647–656.
- [McKee(1999)] T. A. McKee and F. R. McMorris, *Topics in Intersection Graph Theory*, SIAM.
- [Michalewicz(1996)] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag.
- [Minoli(1993)] D. Minoli, *Broadband Network Analysis and Design*, Artech House.

- [Molina(1999)] A. Molina, G. E. Athanasiadou, and A. R. Nix, "The automatic location of base-stations for optimised cellular coverage: a new combinatorial approach," in *Proceedings of the 49th IEEE Vehicular Technology Conference (VTC'99)*, pp. 606–610.
- [Ngo(1998)] C. Y. Ngo and V. O. K. Li, "Fixed channel assignment in cellular radio networks using a modified genetic algorithm," in *IEEE Transactions on Vehicular Technology*, vol. 47, no. 1, pp. 163–172.
- [Perez(1998)] R. Perez, *Wireless Communications Design Handbook: Aspects of Noise, Interference, and Environmental Concerns*, vol. 2, Academic Press.
- [Prasad(1993)] R. Prasad and A. Kegel, "Effects of Rician and log-normal shadowed signals on spectrum efficiency in microcellular radio," in *IEEE Transactions on Vehicular Technology*, vol. 42, no. 3, pp. 274–281.
- [Prasad(1998)] R. Prasad, *Universal Wireless Personal Communications*, Artech House.
- [Quellmalz(1993)] A. Quellmalz, (in German) *Graphenorientierte Planung von Sendernetzen*, Ph.D. thesis, Faculty of Electrical Engineering, University of Karlsruhe, Germany.
- [Rappaport(1996)] T. S. Rappaport, *Wireless Communications: Principles and Practice*, IEEE Press and Prentice Hall.
- [Saunders(1999)] S. R. Saunders, *Antennas and Propagation for Wireless Communication Systems*, John Wiley & Sons.
- [Schaffer(1985)] J. D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Proceedings of the First International Conference on Genetic Algorithms*, pp. 93–100.
- [Sherali(1996)] H. D. Sherali, C. M. Pendyala, and T. S. Rappaport, "Optimal location of transmitters for micro-cellular radio communication system design," in *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, pp. 662–673.
- [Siqueira(1997)] G. L. Siqueira, E. A. Vasquez, R. A. Gomes, C. B. Sampaio, and M. A. Socorro, "Optimization of base station antenna position based on propagation measurements on dense urban microcells," in *Proceedings of the 47th IEEE Vehicular Technology Conference (VTC'97)*, pp. 1133–1137.

- [Sivarajan(1989)] K. N. Sivarajan, R. J. McEliece, and J. W. Ketchum, "Channel assignment in cellular radio," in *Proceedings of the 39th IEEE Vehicular Technology Conference (VTC '89)*, pp. 846–850.
- [Smith(1997)] K. Smith and M. Palaniswami, "Static and dynamic channel assignment using neural networks," in *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 2, pp. 238–249.
- [Smith(1998)] D. H. Smith, S. Hurley, and S. U. Thiel, "Improving heuristics for the frequency assignment problem," in *European Journal of Operational Research*, vol. 107, no. 1, pp. 76–86.
- [Srinivas(1994)] N. Srinivas, and K. Deb, "Multi-objective function optimization using non-dominated sorting genetic algorithms," in *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248.
- [Steele(1992)] R. Steele, *Mobile Radio Communications*, Pentech Press and IEEE Press.
- [Steuer(1986)] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation, and Application*, John Wiley & Sons.
- [Sung(1997)] C. W. Sung and W. S. Wong, "Sequential packing algorithm for channel assignment under cochannel and adjacent-channel interference constraint," in *IEEE Transactions on Vehicular Technology*, vol. 46, no. 3, pp. 676–686.
- [Tutschku(1997)] K. Tutschku, K. Leibnitz, and P. Tran-Gia, "ICEPT-An integrated cellular network planning tool," in *Proceedings of the 47th IEEE Vehicular Technology Conference (VTC '97)*, pp. 765–769.
- [Tutschku(1998)] K. Tutschku and P. Tran-Gia, "Spatial traffic estimation and characterization for mobile communication network design," in *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 804–811.
- [Vasudevan(1999)] M. Vasudevan, K. Rasmussen, C. Yu, and W. Egner, "An algorithmic approach to growth planning in a wireless network," in *Proceedings of the 50th IEEE Vehicular Technology Conference (VTC '99-Fall)*, pp. 2029–2033.
- [Wagen(1993)] J. Wagen and K. Rizk, "Simulation of radio wave propagation in urban microcellular environments," in *Proceedings of IEEE International Conference on Universal Personal Communications (ICUPC '93)*, pp. 595–599.

- [Walfisch(1988)] J. Walfisch and H. L. Bertoni, "A theoretical model of UHF propagation in urban environments," in *IEEE Transactions on Antennas and Propagation*, vol. 36, No. 12, pp. 1788–1796.
- [Wall(1999)] M. Wall, *GAlib: a C++ Library of Genetic Algorithm Components*, <http://lancet.mit.edu/ga/>.
- [Wang(1994)] L.-X. Wang, *Adaptive Fuzzy Systems and Control: Design and Stability Analysis*, Prentice Hall.
- [Wiart(1993)] J. Wiart, A. Marquis, and M. Juy, "Analytical microcell path loss model at 2.2GHz," in *Proceedings of the 4th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'93)*, pp. 30–34.
- [Yacoub(1999)] M. D. Yacoub, "Cell design principles," in *The Mobile Communications Handbook*, Chap. 21, pp. 21:1–15, CRC Press, Springer-Verlag, and IEEE Press.
- [Zitzler(1999a)] E. Zitzler, *Evolutionary Algorithms for Multiobjective Optimizations: Methods and Applications*, Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.
- [Zitzler(1999b)] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," in *IEEE Transactions on Evolutionary Computation*, vol. 3, No. 4, pp. 257–271.
- [Zwick(2000)] T. Zwick, (in German) *Die Modellierung von richtungsaufgelösten Mehrwegegebäudefunkkanälen durch markierte Poisson-Prozesse*, Ph.D. thesis, Faculty of Electrical Engineering, University of Karlsruhe, Germany.

Lebenslauf

Persönliche Daten

Name: Huang
Vorname: Xuemin
Geburtsdatum: 5. Februar 1971
Geburtsort: Linchuan, Jiangxi, China
Staatsangeh.: chinesisch
Familienstand: verheiratet

Schulbildung

1977-1982 Vierte Grundschule, Fuzhou, Jiangxi, China
1982-1985 Unterstufe der vierten Mittelschule, Fuzhou, Jiangxi, China
1985-1988 Oberstufe der ersten Mittelschule, Fuzhou, Jiangxi, China

Studium und Berufsweg

1988-1992 Studium der Informatik an der *National University of Defense Technology*, Changsha, China
Vertiefungsrichtung: Computeranwendung
Diplom: *Bachelor of Engineering*

1992-1995 Studium der Informationstechnik an der *East China Normal University*, Shanghai, China
Vertiefungsrichtung: Intelligentes System
Diplom: *Master of Engineering*

1995-1997 Wissenschaftlicher Mitarbeiter am Institut für Bildverarbeitung und Mustererkennung der *Shanghai Jiao Tong University*, Shanghai, China
Mitarbeit in Forschung und Lehre
Arbeitsgebiet: Digitale Bildanalyse, Bildkommunikation, Multimedia-Technik

seit Nov. 1997 Doktorand des Institut für Höchstfrequenztechnik und Elektronik der Universität Karlsruhe und der LS telcom AG, Lichtenau, Deutschland
Forschungsschwerpunkte: Planungswerkzeuge für Mobilkommunikationsnetze, Wellenausbreitungsmodellierung, künstliche Intelligenz, Optimierungsalgorithmen