Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

ANDREAS EISENBLÄTTER AND HANS-FLORIAN GEERDES AND THORSTEN KOCH AND ALEXANDER MARTIN AND ROLAND WESSÄLY

UMTS Radio Network Evaluation and Optimization Beyond Snapshots

http://www.zib.de/Optimization/Projects/Telecom/Matheon-B4

ZIB-Report 04-15 (October 2004)

Andreas Eisenblätter · Hans-Florian Geerdes · Thorsten Koch · Alexander Martin · Roland Wessäly

UMTS Radio Network Evaluation and Optimization Beyond Snapshots

Abstract. This paper is concerned with UMTS radio network design. The task is to reconfigure antennas and the related cells as to improve network quality. In contrast to second generation GSM networks, *interference* plays a paramount role in the design of third generation radio networks. A recent compact characterization of a radio network through a linear equation system allows to assess cell load and interference. This characterization is based on user snapshots and is generalized here to average traffic intensities. This allows to overcome the notorious difficulties of snapshot-based network optimization approaches. A mixed-integer programming model for the network design problem is recalled that is based on user snapshots and it is contrasted with a new network design model based on the average coupling formulation. Exemplarily focusing on the particularly important problem of optimizing antenna tilts, computational results are presented for a fast local search algorithm and for the application of a MIP solver to both models. These results demonstrate that the new average-based approaches outperform state-of-the-art snapshot models for UMTS radio network optimization.

1. Introduction

The Universal Mobile Telecommunications System (UMTS) is a 3rd generation (3G) cellular system for mobile telecommunications. UMTS supports all services of the worldwide predominant GSM and GPRS networks and is more powerful, more flexible, and more radio spectrum efficient than its predecessors.

UMTS is a *Wideband Code Division Multiple Access* (WCDMA) system. Radio transmissions are generally not separated in the time or the frequency domain. Complex coding schemes are instead used to distinguish different radio transmissions at a receiver. The capability to properly detect the desired carrier signal, however, requires that this is not "too deeply" buried among interfering radio signals. Technically speaking, the carrier signal to interference ratio shall not drop below some threshold value. Interference, thus, needs to be carefully controlled during network planning and operation, because it is a limiting factor for network capacity.

Hans-Florian Geerdes, Thorsten Koch: Zuse Institute Berlin (ZIB)
e-mail: {geerdes,koch}@zib.de

Andreas Eisenblätter, Roland Wessäly: Atesio GmbH and ZIB
e-mail: {eisenblaetter,wessaely}@atesio.de

Alexander Martin: Darmstadt University of Technology (TUD)
e-mail: martin@mathematik.tu-darmstadt.de

This work has been supported by the DFG Research Center Matheon, "Mathematics for key technologies".

1.1. Overview

This paper is organized as follows. The rest of this section introduces the task of UMTS radio network optimization and addresses related work. Section 2 introduces the technical background in parallel with the notation used throughout the paper. Section 3 introduces a method to perform fast approximate analysis of network capacity. This depends on systems of linear equations describing the interference coupling among cells. In Section 4, we refine a mixed integer programming model for network optimization based on traffic snapshots [8,14]. A novel, alternative mathematical program, which overcomes some drawbacks of the previous model, is presented in Section 5. Section 6 contains computational results for realistic planning scenarios. These results are obtained by implementations based on the two mathematical models and from a local search procedure. We draw conclusions in Section 7.

1.2. Network Planning

Network operators are currently deploying UMTS in many countries across Europe and Asia. Unlike during the introduction of GSM in the 1990s, the new mobile telecommunication technology faces a well-developed market with GSM as a competing system. GSM already offers large areal coverage, high reliability, and acceptable prices. An incentive for the user to change technology towards UMTS is the availability of more services and higher communication data-rates in large areas with a competitive pricing. Even more complex planning tools than those for GSM are indispensable. This is particularly true for the radio network planning due to the change of the radio access technology from frequency and time multiplexing to code multiplexing.

As in the case of frequency planning for GSM networks, it is to be expected that the quality of the planning results can be largely improved if the radio network planner is assisted by automatic optimization functions. First commercial products in this domain are available, but it will certainly take a few more years to reach the maturity of GSM frequency planning. The paper contributes to this development with respect to network evaluation and network optimization.

A central part in the initial deployment and the subsequent expansions of a UMTS radio network is to decide about the location and configuration of the base stations including their antennas. Among others, the type of an antenna (and thus its radiation pattern), the mounting height, and its main radiation direction, the *azimuth* in the horizontal plane and the *tilt* in the vertical plane, have to be decided. This may look like a facility location problem at first sight, but it is not! The coverage area and the capacity of a base station may shrink significantly due to radio interference.

1.3. Tilt Optimization

All three optimization approaches proposed in this paper can be used to optimize all aspects of an antenna installation mentioned above. From the practical point of view, however, tilt optimization is of primary interest. Operators often deploy an UMTS networks in addition to existing 2G infrastructure. Since it is costly and increasingly difficult to acquire new sites, UMTS antennas are often installed in places where there is already a GSM antenna. The two types of antennas are sometimes assembled on the same mechanical structure, *dual band* antennas that operate for both GSM and UMTS are also in use. This basically fixes the antenna's height and azimuth. In addition, operators typically have standard antenna types. The one parameter that operators favor for optimization is tilt. We focus on optimizing this parameter in our computational studies in Section 6.

Antenna tilt comes in two flavors, mechanical and electrical tilt, which can be varied independently. The *mechanical tilt* is the angle by which the antenna is tilted out of the horizontal plane. For many UMTS antennas, the main radiation direction in the vertical plane can also be changed by electrical means. The resulting change is called *electrical tilt.*

Tilt optimization alone has a considerable leverage on network coverage as well as capacity, because it influences cell ranges and interference. In general, if an antenna is tilted down, the size of the corresponding cell and thereby the load that this antenna has to serve decrease. At the same time, the amount of interference in neighboring cells might be reduced.

1.4. Related Work

This work is largely based on the authors' participation in the EU-funded project Momentum. The interdisciplinary project had mathematicians, engineers, and telecommunication operators in the consortium, the scope was developing models and algorithms for simulation and automatic planning of UMTS radio networks. The first version of the snapshot model presented in Section 4 is formulated in [10, 14], refinements and detailed technical descriptions of the model parameters is given in [8]; a summary and first computational results are given in [11]. Due to the practical relevance of our topic we mention both works from the mathematical as well as from the engineering side. In [39] a survey charting the evolution of centralized planning for cellular systems is given, including the authors' contributions.

There is a variety of mathematical work on UMTS radio network planning. Optimization models similar to our snapshot model are suggested in [1, 2, 3] with computational results. The problems include site selection and base station configuration. Heuristic methods such as tabu or greedy search are used to solve instances. Integer programming methods for 3G network planning are presented in [29]. Some papers deal with subproblems: pilot power optimization under coverage constraints using mathematical programming is performed in [35, 38], power control and capacity issues with particular emphasis on network planning are treated in [6, 27]. To our knowledge, however, there is no work focusing on the network *tuning* aspect under technical constraints; tilt optimization falls into this category.

On the other hand, there is a vast body of work by engineers on problems in network planning. The first and exemplary landmark monography covering many technical aspects of UMTS networks and some practice-driven optimization and tuning rules is [26]. Optimization of network quality aspects not involving site selection and using meta-heuristics is treated in [17]. Network design with a special emphasis on optimizing cell geometry using evolutionary algorithms is presented in [22,23]. A mixed integer program for optimizing quality-of-service for users within a snapshot is developed in [36].

The latter paper also uses the dimension reduction technique that we present in Section 3. The dimension reduction had earlier been proposed for mono-service traffic in [31], and was extended to a service mix in [6,7]. We contribute a generalization to average user density maps and an optimization model based on these premises. Earlier papers addressing power control by considering systems of linear equations with positive solutions are [18,19,20].

2. Radio Network Design for WCDMA

We give an introduction to the properties of UMTS technology insofar as they are relevant to our paper. In mobile telecommunications, two transmission directions are distinguished: the *uplink* and the *downlink*. An uplink (UL) transmission takes place when a mobile transmits to a base station. A downlink (DL) transmission takes place from the base station to mobile users. The type of radio network we consider with UMTS is called a *cellular* network: there is an infrastructure of base stations. One or more antennas are installed at each base station. Each antenna serves users in a certain area (typically the area where this antenna provides the strongest signal), this area is called *cell*.

In a UMTS network, every antenna emits a special constant power *pilot* signal, which is used by the mobiles to measure which antenna is received best and for deciding which cell they associate to. In this paper, we consider up- and downlink transmissions on *dedicated channels*, which can be seen as point-to-point connections between a mobile and a base station. There are some other channels that are shared between the mobiles in a cell, but apart from the *pilot channel*, we do not model them explicitly. In UMTS networks rolled out today, transmissions in uplink and downlink do not interfere, because separate frequency bands are reserved for each direction (frequency domain duplex, FDD).

Examples for well-known mobile communication services are speech telephony or text messaging, some of the additional services that will be offered on the basis of UMTS are video telephony, downlink data streaming or web access. This diversification implies an increasing heterogeneity of the traffic in the network. An important difference between services is the requested data rate, especially as some of the new services require a substantially higher data rate than speech telephony, for example.

2.1. Preliminaries and Notation

For a complete account of the notation used in this paper please refer to Table 1. We consider an UMTS radio network with a set $\mathcal{N}$ of antennas (or cells) and aset $\mathcal{M}$ of mobile users in a traffic snapshot. We assume that each mobile is connected to exactly one antenna, namely, the one which has the strongest pilot signal. The practical situation is more complicated, since a mobile can be linked to more than one antenna at a time. We ignore this feature of UMTS called *soft-handover* for our presentation.

We denote a vector with components v_j in bold, $\mathbf{v}$. The notation $\operatorname{diag}(\mathbf{v})$ stands for a diagonal matrix of matching dimension with the elements of $\mathbf{v}$ on the main diagonal.

Table 1. Notation

$\mathcal{N}$		Set of antenna installations (cells) in the network
$\mathcal{I}$	$\mathcal{I} \supset \mathcal{N}$	Set of all possible installations for the network
i	$i \in \mathcal{N}$, $i \in \mathcal{I}$	Installation
$\mathcal{M}$		Set of mobiles
$\mathcal{M}_i$		Mobiles served by antenna installation i
$\mathcal{M}^\uparrow$		$\{m \in \mathcal{M} \mid \alpha_m^\uparrow > 0\}$
$\mathcal{M}^\downarrow$		$\{m \in \mathcal{M} \mid \alpha_m^\downarrow > 0\}$
m	$m \in \mathcal{M}$	Mobile
S		Set of services
s	$s \in S$	Service
A		Planning area
p	$p \in A$	Location in the planning area
A_i	$A_i \subset A$	Cell area (best server area) of installation i
T_s	$A \to \mathbb{R}_+$	Average spatial traffic distribution of service s
$p_m^\uparrow$	$\in \mathbb{R}_+$	Uplink transmit power from mobile $m \in \mathcal{M}^\uparrow$
$p_{im}^\downarrow$	$\in \mathbb{R}_+$	Downlink transmit power from installation i to mobile m
$\hat{p}_i^\downarrow$	$\in \mathbb{R}_+$	(Downlink) pilot transmit power from installation i
$\check{p}_i^\downarrow$	$\in \mathbb{R}_+$	(Downlink) common channels transmit power from installation i
$\bar{p}_i^\downarrow$	$\in \mathbb{R}_+$	Total transmit power of installation i
$\bar{p}_i^\uparrow$	$\in \mathbb{R}_+$	Total received power at installation i
$\Pi_i^{\max\downarrow}$	$\in \mathbb{R}_+$	Maximum total transmit power for installation i
$\gamma_{mi}^\uparrow$	$[0, 1]$	Uplink attenuation factor between mobile m and installation i
$\gamma_{im}^\downarrow$	$[0, 1]$	Downlink attenuation factor between installation i and mobile m
η_i, η_m	≥ 0	Noise at installation i/mobile m
$\alpha_m^\uparrow$, $\alpha_m^\downarrow$	$[0, 1]$	Uplink/downlink activity factor of mobile m
$\bar{\omega}_m$	$[0, 1]$	Orthogonality factor for mobile m
$\mu_m^\uparrow$	≥ 0	Uplink CIR target for mobile m
$\mu_m^\downarrow$	≥ 0	Downlink CIR target for mobile m
$\hat{\mu}_m^\downarrow$	≥ 0	Pilot $\mathrm{E_c/I_0}$ requirement for mobile m

2.1.1. Antenna installations. The degrees of freedom (antenna type, height, azimuth, tilt) when configuring an antenna and thereby determining the properties of the related cell have been mentioned above. For the antenna of a certain cell, we call a fixing of the parameter values in these dimensions an *installation*.

The set of all considered installations will be denoted by $\mathcal{I}$. An actual *network* or *network design* is a choice of all possible installations with the property that exactly one of the potential installations for each cell is chosen. In particular, it always holds that $\mathcal{N} \subset \mathcal{I}$. We will use the indices i and j for elements of both $\mathcal{N}$ and $\mathcal{I}$.

2.1.2. CIR inequality for a transmission. For a signal to be successfully decoded at the receiver with WCDMA technology, the ratio between the received strength of the desired signal and all interfering signals – including background noise exterior to the system – must exceed a specific threshold. This ratio is also called carrier-to-interference ratio (CIR), the threshold is called *CIR target*. For a transmission to take place, the following inequality must be satisfied:

$$\frac{\text{Strength of Desired Signal}}{\text{Noise} + \sum \text{Strength of Interfering Signals}} \geq \text{CIR target} \tag{1}$$

Code multiplexing technology introduces the possibility that the right hand side of this inequality is smaller than one, that is, the desired signal strength may be much weaker than the interference. We will now introduce the various parameters that play a role in (1).

2.1.3. CIR targets and activity. The CIR target is service specific. In order to successfully decode a higher data rate a higher CIR target must be satisfied. There is also a specific CIR target for the pilot channel, mainly depending on user equipment. We denote the CIR targets for a user $m \in \mathcal{M}$ for uplink, downlink, and pilot by $\mu_m^\uparrow$, $\mu_m^\downarrow$, and $\hat{\mu}_m^\downarrow$.

The way that users access a service over time also varies: in a speech conversation, each of the two users involved speaks 50 % of the time on average, and no data is transmitted in silence periods (discontinuous transmissions). This is taken into account in the form of *activity factors*. For each user $m \in \mathcal{M}$ we have two activity factors, $\alpha_m^\uparrow$ for the uplink and $\alpha_m^\downarrow$ for the downlink. The activity factor can be used to compute the *average* power that is sent over a link. There are services that cause traffic in only one direction, e. g. downlink data streaming. In this case the corresponding activity factor for the other direction is zero. (There is actually always some control traffic, but it is negligible.) In (1), we take this averaged value for all interfering signals, since we cannot tell whether the interfering links are currently in an active period or not. For the link in question, however, we need to reach the CIR target in active periods only, since there is no transmission in inactive periods. The activity is thus only taken into account in the denominator of (1), but not in the numerator.

2.1.4. Attenuation. The strength of a signal transmitted over a radio link is attenuated on its way to the receiver. The received power depends linearly on the output power at the sender. The attenuation on the radio channel (excluding transmission and reception equipment) is called *path loss*. Predicting the path loss in real-world settings is difficult and many methods ranging from rule-of-thumb formulas up to sophisticated raytracing methods using 3D building data and vegetation are available. We refer the interested reader to [16, 24, 5]. Besides this loss on the radio channel, some further losses and gains due to the cabling, hardware, and user equipment have to be considered. All this information is then summed up into *attenuation factors*. We have two factors for each pair $i \in \mathcal{N}$, $m \in \mathcal{M}$, an attenuation factor $\gamma^{\uparrow}_{im}$ for the uplink and $\gamma^{\downarrow}_{mi}$ for the downlink.

2.1.5. Complete CIR constraints. We denote the transmission power of a mobile $m \in \mathcal{M}$ in uplink by $p^{\uparrow}_m$. The received signal strength at installation $i \in \mathcal{N}$ is then $\gamma^{\uparrow}_{mi}\, p^{\uparrow}_m$. If we denote the received background noise at installation $i \in \mathcal{N}$ by η_i, the complete version of (1) for the uplink and the transmission from m to i reads

$$\frac{\gamma^{\uparrow}_{mi} p^{\uparrow}_m}{\eta_i + \sum_{n \neq m} \gamma^{\uparrow}_{ni}\, \alpha^{\uparrow}_n p^{\uparrow}_n} \geq \mu^{\uparrow}_m \,.$$

Writing

$$\bar{p}^{\uparrow}_i := \eta_i + \sum_{m \in \mathcal{M}} \gamma^{\uparrow}_{mi}\, \alpha^{\uparrow}_m\, p^{\uparrow}_m \tag{2}$$

for the *average* total received power at installation $i \in \mathcal{N}$, this simplifies to

$$\frac{\gamma^{\uparrow}_{mj}\, p^{\uparrow}_m}{\bar{p}^{\uparrow}_j - \gamma^{\uparrow}_{mj}\, \alpha^{\uparrow}_m\, p^{\uparrow}_m} \geq \mu^{\uparrow}_m \,. \tag{3}$$

In the *downlink*, the situation is more complicated. First of all, we have to take the pilot and common channels into account, whose transmission power we denote by $\hat{p}^{\downarrow}_i$ and $\check{p}^{\downarrow}_i$ for installation i. There is another UMTS feature to be considered here: each cell selects orthogonal transmission codes for the mobiles in its cell, which in theory do not interfere mutually (we neglect the use of secondary scrambling codes). However, due to reflections on the way the signals partly lose this property. Signals from other antennas do not have it at all. Hence, when summing up the interference, the interference from the same cell is reduced by an environment dependent *orthogonality factor* $\bar{\omega}_m \in [0, 1]$, with $\bar{\omega}_m = 0$ meaning perfect orthogonality and $\bar{\omega}_m = 1$ no orthogonality. We define the total *average* output power of installation i as

$$\bar{p}^{\downarrow}_i := \sum_{m \in M_i} \alpha^{\downarrow}_m\, p^{\downarrow}_{jm} + \hat{p}^{\downarrow}_i + \check{p}^{\downarrow}_i \;\;. \tag{4}$$

Writing η_m for the noise value at mobile m, we obtain the downlink version of (1) related to transmission from i to m:

$$\frac{\gamma^{\downarrow}_{jm}\, p^{\downarrow}_{jm}}{\gamma^{\downarrow}_{im}\, \bar{\omega}_{im} \left(\bar{p}^{\downarrow}_i - \alpha^{\downarrow}_m\, p^{\downarrow}_{im} \right) + \sum_{j \neq i} \gamma^{\downarrow}_{jm}\, \bar{p}^{\downarrow}_j + \eta_m} \geq \mu^{\downarrow}_m \,. \tag{5}$$

Note on downlink orthogonality. Strictly speaking, there is one control channel for which orthogonality to other channels in the cell does not apply, the synchronization channel (SYNC). We do not consider this detail for clarity of presentation. This is acceptable from the engineering point of view, because its transmission power is fairly low in comparison to the other common channels.

For the *pilot*, the situation is simpler. Technically speaking, the received chip energy of the pilot signal – called the pilot channel's $\mathrm{E_c}$ or CPICH $\mathrm{E_c}$ – relative to the total power spectral density $\mathrm{I_0}$ has to lie above a certain threshold. When computing the spectral density $\mathrm{I_0}$ as the denominator in the pilot version of (1), no benefit due to orthogonality applies and even the pilot's own contribution appears as interferer:

$$\frac{\gamma^{\downarrow}_{im}\hat{p}^{\downarrow}_{i}}{\eta_m + \sum_{i\in\mathcal{N}} \gamma^{\downarrow}_{im}\,\bar{p}^{\downarrow}_{i}} \geq \hat{\mu}^{\downarrow}_{m} \,. \tag{6}$$

The left hand fraction in this inequality is called the pilot channel's $\mathrm{E_c/I_0}$ (CPICH $\mathrm{E_c/I_0}$).

2.1.6. Power control. The quality of a radio channel can actually not be pinned to a constant factor γ. It varies strongly over time due to a collection of dynamical phenomena grouped under the terms *shadowing* and *fading*. In an UMTS system, these effects are largely made up for by *power control*. The receiver measures the signal strength in very short time intervals and notifies the sender of any changes, which then adjusts the transmission power accordingly to meet the CIR requirements (3) and (5) most of the time. We do not take this mechanism into account but assume *perfect power control*, that is, we suppose that the sender can adjust its transmission power exactly to the minimum level required to fulfill the CIR requirement.

2.2. Performance Metrics for UMTS Radio Networks

The "quality" or "performance" of a UMTS radio network is measured by various scales, see [21,26,25], for example. There is no single objective for optimizing a radio network, but a variety of factors need to be taken into account.

2.2.1. Coverage. The user associates to the network via the pilot channel. The quality of the pilot signal therefore determines the network's *coverage*. The first condition for coverage is that the pilot's absolute signal strength $\mathrm{E_c}$, the numerator in (6), is sufficiently large for the user's mobile equipment to reliably detect it. The areas in which this is the case are said to have $\mathrm{E_c}$ *coverage*.

At a location with sufficient $\mathrm{E_c}$ coverage, the pilot signal can only be decoded if the pilot's $\mathrm{E_c/I_0}$ is sufficient, that is, if (6) holds true; the locations where this is the case are said to have $\mathrm{E_c/I_0}$ *coverage*. Unlike $\mathrm{E_c}$ coverage, $\mathrm{E_c/I_0}$ coverage in a certain location depends on interference and, thus, on the traffic load within the network. While $\mathrm{E_c}$ problems are likely to occur at places that are too distant from any antenna, failure to reach the required $\mathrm{E_c/I_0}$ ratio is often a sign of too much interference from other cells.

2.2.2. Pilot Pollution. Mobiles continuously track the strength not only of the pilot signal of their own cell, but also the pilot signals from neighboring cells if these are received with a strength within a certain window (say, 5 dB) below the own cell's signal. This serves to allow a seamless *hand-off* in case the user moves and crosses cell borders. Due to hardware constraints, however, the number of pilots that can be tracked simultaneously is limited to only a few. Thus, if too many pilot signals are "audible" in a certain location, the location is said to suffer from *pilot pollution*.

2.2.3. Network load. Once that network coverage is given by sufficiently strong and clear pilot signals, users are capable of establishing a connection. The most important performance measure – and the one that is most involved to compute – essentially aims at determining how many users, or how much *load* the network can carry. Services differ in their demand for transmission capacity. Whether there is sufficient transmission capacity for all user requests and which link capacities are likely to be observed are evaluated by network simulations.

Monte-Carlo simulations are typically used for this analysis: User demand realizations called *traffic snapshots* are drawn, and each snapshot is analyzed. The analysis of independent snapshots is ideally continued until network performance indicators are statistically reliable [36]. In practice, however, the number of snapshots to be evaluated is often fixed in advance [26].

The details of analyzing a single snapshot differ. Each approach in essence has to decide at least which of the users are served by which cell(s), the power levels of each active link, and which users are out of coverage or unserved due to capacity or interference reasons. During this step, the limits on user and base station equipment power are considered, including the variations in CIR targets depending on service, equipment type, and speed, the effects of various forms of soft hand-over, etc. As a result, the transmit and receive signal power is determined for all cells. These power levels are in turn the basis for deriving $\mathrm{E_c/I_0}$ or service coverage maps, since they compose the $\mathrm{I_0}$.

The *downlink load* of an individual cell is measured relative to a full cell: it is the fraction of the maximum output power that the cell uses for transmission. The *uplink load*, on the other hand, is measured by the *noise rise* compared to an empty cell. The noise rise is the ratio of total received power at an antenna to the noise, which is always present. A noise rise of two (which means that the signals generated within the network reach the antenna with a strength equal to the noise) corresponds to 50 % uplink load; the load is at 100 % if the noise rise reaches infinity.

The average cell load is actually limited to values significantly below 100% to leave room for compensation of dynamic effects by the power control mechanisms and to ensure the system's stability. Typically the maximum average load should not exceed 50% in the uplink and 70% in the downlink. We denote the practical limit on the total transmit power in the downlink by $\Pi^{\max\downarrow}$.

3. Computing Network Load

This section introduces systems of linear equations that describe the up- and downlink load per cell and the coupling among the cells. The description is idealized since no transmit power or noise rise limitations are taken into account. For the sake of simplicity, we also assume that no mobile is in soft hand-over. With the necessary precautions, solving these systems nevertheless allows to get an estimate of network load without extensive simulations. We will use this method to quickly assess network load during local improvement methods. Moreover, these systems are at the core of an alternative network optimization model proposed in Section 5.

The systems were introduced and extended in [31,37,7,6] to speed up an central operation in Monte-Carlo simulation for the performance assessment of UMTS radio network (see Section 2.2). The original derivation for the equation systems is therefore based on user traffic in the form of traffic snapshots. We follow this approach to introduce the systems, but we then extend it to derive the systems based on *average* traffic load distribution or to combinations of average load and a traffic snapshot.

The central assumptions are that all users are served and that all CIR targets in the uplink (3) and downlink (5) are met at equality. We start from a network design with installations $i \in \mathcal{N}$ and a traffic snapshot with mobiles $m \in \mathcal{M}$, using the notation from Table 1.

3.1. Uplink.

Concerning the uplink at antenna installation j, recall that $\bar{p}^{\uparrow}_{j}$ is the total amount of received power including thermal and other noise. Under the above assumptions elementary transformation of the equality version of (3) allow to derive two quantities for every mobile m served by installation j: First, the transmission power $p^{\uparrow}_{m}$ of mobile m given the total received power $\bar{p}^{\uparrow}_{j}$ at the serving installation j. Second, the fraction of the total received power at the installation j originating in mobile m:

$$p^{\uparrow}_{m} = \frac{1}{\gamma^{\uparrow}_{mj}} \frac{\mu^{\uparrow}_{m}}{1+\alpha^{\uparrow}_{m}\mu^{\uparrow}_{m}} \bar{p}^{\uparrow}_{j} \tag{7}$$

$$\frac{\alpha^{\uparrow}_{m}\,\gamma^{\uparrow}_{mj}\,p^{\uparrow}_{m}}{\bar{p}^{\uparrow}_{j}} = \frac{\alpha^{\uparrow}_{m}\,\mu^{\uparrow}_{m}}{1+\alpha^{\uparrow}_{m}\,\mu^{\uparrow}_{m}} \tag{8}$$

It is convenient to define the *uplink user load* $l^{\uparrow}_{m}$ of a mobile m as the right hand side of (8):

$$l^{\uparrow}_{m} := \frac{\alpha^{\uparrow}_{m}\,\mu^{\uparrow}_{m}}{1+\alpha^{\uparrow}_{m}\,\mu^{\uparrow}_{m}} \tag{9}$$

We break down the contributions to the total received power $\bar{p}^{\uparrow}_{j}$ at installation j in dependence of all uplink connections (not just the ones served by j). Let

$\mathcal{M}_i \subseteq \mathcal{M}$ denote the set of all users served by installation i. Then (2) reads as

$$\bar{p}^{\uparrow}_j = \sum_{m\in\mathcal{M}_j} \gamma^{\uparrow}_{mj}\, \alpha^{\uparrow}_m\, p^{\uparrow}_m + \sum_{i\neq j}\sum_{m\in\mathcal{M}_i} \gamma^{\uparrow}_{mj}\, \alpha^{\uparrow}_m\, p^{\uparrow}_m + \eta_j\,.$$

Defining the installation *uplink coupling factors* $C^{\uparrow}_{ji}$ as

$$C^{\uparrow}_{ji} := \sum_{m\in M_i} \frac{\gamma^{\uparrow}_{mj}}{\gamma^{\uparrow}_{mi}}\, l^{\uparrow}_m\,,$$

and substituting (7), the individual uplink transmission powers can be written as

$$\bar{p}^{\uparrow}_j = C^{\uparrow}_{jj}\, \bar{p}^{\uparrow}_j + \sum_{i\neq j} C^{\uparrow}_{ji}\bar{p}^{\uparrow}_i + \eta_j\,. \tag{10}$$

Thus, the total power received at the installation is composed of three contributions, those from the own cell, those from other cells, and noise. The quantity $C^{\uparrow}_{jj}$ measures the contribution from the own users, and $C^{\uparrow}_{ji}$ scales the contribution from installation i. The matrix

$$C^{\uparrow} := \left(C^{\uparrow}_{ji}\right)_{1\leq i,j\leq|\mathcal{N}|}$$

is called the *uplink cell load coupling matrix.*

Collecting (10) for all installations and writing $\boldsymbol{\eta}^{\uparrow}$ for the vector of noise values, we obtain the desired system of linear equations governing the uplink cell reception powers:

$$\bar{\mathbf{p}}^{\uparrow} = C^{\uparrow}\bar{\mathbf{p}}^{\uparrow} + \boldsymbol{\eta}^{\uparrow}\,. \tag{11}$$

Under the assumptions stated in the beginning of this section, the solution of (11) is the uplink received powers at each installation. Necessary and sufficient conditions on $C^{\uparrow}$ for the existence of positive and bounded solutions to (11) are discussed in [6,7].

3.2. Downlink.

In the downlink case, we basically repeat what has just been done for the uplink. The starting point is the CIR constraint (5). Assuming that the constraint is met sharply for all mobiles it can be rewritten as

$$\frac{1+\bar{\omega}_m\, \alpha^{\downarrow}_m\, \mu^{\downarrow}_m}{\alpha^{\downarrow}_m\, \mu^{\downarrow}_m}\, \alpha^{\downarrow}_m\, p^{\downarrow}_{jm} = \bar{\omega}_m\, \bar{p}^{\downarrow}_j + \sum_{i\neq j} \frac{\gamma^{\downarrow}_{im}}{\gamma^{\downarrow}_{jm}}\, \bar{p}^{\downarrow}_i + \frac{\eta_m}{\gamma^{\downarrow}_{jm}}$$

We define the *downlink user load* of serving mobile m as:

$$l^{\downarrow}_m := \frac{\alpha^{\downarrow}_m\, \mu^{\downarrow}_m}{1+\bar{\omega}_m\, \alpha^{\downarrow}_m\, \mu^{\downarrow}_m}\,. \tag{12}$$

Recall that the break-down of the downlink transmission power at installation j is $\bar{p}^{\downarrow}_j$, as defined in (4). Similar to the uplink case, further notation is helpful to express the dependency of $\bar{p}^{\downarrow}_j$ on the downlink transmission power at all installations. We introduce the downlink *coupling factors*

$$C^{\downarrow}_{jj} := \sum_{m \in M_j} \bar{\omega}_m \, l^{\downarrow}_m \qquad \text{and} \qquad C^{\downarrow}_{ji} := \sum_{m \in M_j} \frac{\gamma^{\downarrow}_{im}}{\gamma^{\downarrow}_{jm}} \, l^{\downarrow}_m \quad (i \neq j)$$

for installations i and j as well as the installation's *traffic noise power*

$$p^{(\eta)}_j := \sum_{m \in M_j} \frac{\eta_m}{\gamma^{\downarrow}_{jm}} \, l^{\downarrow}_m \, .$$

The transmit powers at the installations satisfy the expression

$$\bar{p}^{\downarrow}_j = C^{\downarrow}_{jj} \, \bar{p}^{\downarrow}_j + \sum_{i \neq j} C^{\downarrow}_{ji} \, \bar{p}^{\downarrow}_i + p^{(\eta)}_j + \hat{p}^{\downarrow}_j + \check{p}^{\downarrow}_j \, . \tag{13}$$

The interpretation of $C^{\downarrow}_{jj}$, $C^{\downarrow}_{ji}$, and $p^{(\eta)}_j$ are as follows. The first term, $C^{\downarrow}_{jj}$, captures the effects due to intra-cell interference. The second term, $C^{\downarrow}_{ji}$, quantifies the fraction of transmission power spent on overcoming interference from other installations. The third quantity, $p^{(\eta)}_j$, states how much transmission power is needed to overcome the noise at the receivers if no intra-system interference were present. We call

$$C^{\downarrow} := \left(C^{\downarrow}_{ji}\right)_{1 \leq i,j \leq |\mathcal{N}|}$$

the *downlink load coupling matrix*.

Figure 1 contains a graphical representation of the downlink coupling matrix. The connections between each pair of antennas are shaded according to the absolute value of the sum of the two corresponding off-diagonal elements. Darker shades indicating higher values. The difference between Figs. 1(a) and 1(b) shows the decoupling effect that can be achieved by tilting.

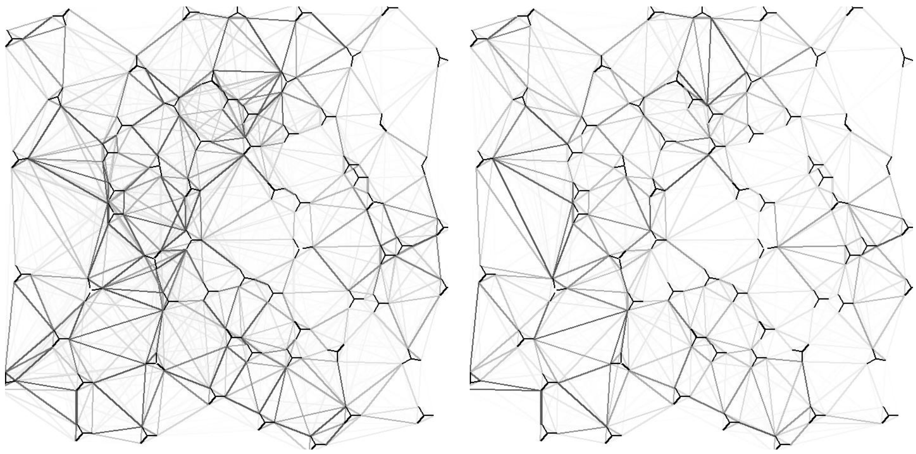
Again, equation (13) for all antenna installations in the network form a linear equation system that describes the downlink transmit power in each cell:

$$\bar{\mathbf{p}}^{\downarrow} = C^{\downarrow} \bar{\mathbf{p}}^{\downarrow} + \mathbf{p}^{(\eta)} + \hat{\mathbf{p}}^{\downarrow} + \check{\mathbf{p}}^{\downarrow} \tag{14}$$

The same qualifications as discussed in the uplink case hold.

3.3. Generalized Coupling Matrices.

We have to assess network load as part of network optimization, for example, in order to find out whether a tentative network design is capable of supporting the offered load. In this context, the above equation systems have two shortcomings:



(a) Original network (b) Optimized electrical tilts

Fig. 1. Effect of tilt adjusting on mutual interference, Berlin scenario

- The equation systems allow to determine the network load for one snapshot. In order to obtain statistically reliable results many snapshots, hundreds or even thousands, have to be evaluated. This is computationally prohibitive if the load evaluation shall be used within a local search approach and has to be executed during most steps.
- The results from the load calculation become meaningless once the network is in overload.

We address the first issue here. In order to obtain an estimation of the network load based on the analysis of only *one* equation system, we derive the coupling matrices and the downlink traffic noise power vector on the basis of *average* traffic load distributions. This approach is appealing because the computational complexity is radically diminished. The speed-up comes, however, at the expense of an estimation error. To our knowledge, there is no extensive research or upper bound for this error. It is reported [30] that for a mono-service situation the cell load calculation on the basis of the average load distribution leads to a systematic underestimation of user blocking, but the corresponding error is estimated as less than 10%. A more rigorous error analysis is in demand to provide the cell load estimates with the necessary accuracy estimates.

We denote by A the total planning area and by $A_i \subset A$ the best server area of cell i. Let S be the set of services under consideration, let T_s be the service-specific spatial traffic distributions, and let $T_s(p)$ denote the average traffic intensity of service s at location p (for some specific point in time, e. g., during the busy hour). We assume that the traffic intensity measures the quantity of concurrent (packet) calls at location p. The position p specifies a two-dimensional coordinate at some reference height, e. g., 1.5 m, or a three-dimensional coordinate.

For the sake of a simpler exposition, we assume that each service s has reprasentative CIR targets $\mu_s^\uparrow, \mu_s^\downarrow$ and activity factors $\alpha_s^\uparrow, \alpha_s^\downarrow$ in up- and downlink; we also assume a location-specific noise η_p at a mobile in position p. Instead of using $\gamma_{mj}^\uparrow, \gamma_{jm}^\downarrow, \bar{\omega}_m$ for the connection between installation j and a mobile at location p, we simply write $\gamma_{jp}^\uparrow, \gamma_{jp}^\downarrow, \bar{\omega}_p$. With these conventions, the snapshot-based definitions of coupling matrices and downlink traffic noise power can be easily generalized to traffic load distributions.

The basic load definitions (9) and (12) are replaced by

$$l_p^\downarrow := \sum_{s \in S} \frac{\alpha_s^\downarrow \, \mu_s^\downarrow}{1 + \bar{\omega}_p \, \alpha_s^\downarrow \, \mu_s^\downarrow} \, T_s(p) \, .$$

The uplink coupling matrix is given by

$$C_{jj}^\uparrow := \int_{p \in A_j} l_p^\uparrow \, dp \, , \qquad C_{ji}^\uparrow := \int_{p \in A_i} \frac{\gamma_{jp}^\uparrow}{\gamma_{ip}^\uparrow} \, l_p^\uparrow \, dp \, ,$$

the downlink coupling matrix and the traffic noise power are given by

$$C_{jj}^\downarrow := \int_{p \in A_j} \bar{\omega}_p \, l_p^\downarrow \, dp \, , \quad C_{ji}^\downarrow := \int_{p \in A_j} \frac{\gamma_{ip}^\downarrow}{\gamma_{jp}^\downarrow} \, l_p^\downarrow \, dp \, , \quad p_j^{(\eta)} := \int_{p \in A_j} \frac{\eta_p}{\gamma_{jp}^\uparrow} \, l_p^\downarrow \, dp \, . \tag{15}$$

The above definitions can be generalized further in two directions. First, the definitions also make sense if some part of the traffic is taken according to average load distributions and another part from a snapshot. This can be helpful for Monte-Carlo simulations: service usage well approximated by average behavior can be treated through average load maps, while high data-rate users with bursty traffic can be analyzed at snapshot-level. Second, the best server information can be replaced by assignment probability information, indicating the probability with which some point is served from an installation.

4. Mixed-Integer Programming Model Based on Snapshots

In this section we sketch how to model a snapshot based MIP for the UMTS planning problem. Details on this and a full account of the model can be found in [10, 14]. Besides the set of potential antenna installations for each cell of the radio network under consideration, the input to the model consists of a set $\mathcal{D}$ of snapshots, each snapshot containing many users. Every user belongs to a specific snapshot $d \in \mathcal{D}$. The model allows to select a radio network design by picking one antenna configuration for each cell. Depending on this design choice, the quality of the current solution is assessed in each snapshot. To this end, a decision has to be made for each user whether it is dropped (left without service) or served, and if so, by which cell. For the served users power levels for each link have to be determined such that all CIR requirements are met. Our objective is to maximize the number of served users.

4.1. Outline of the Model

Binary variables $z_i, i \in \mathcal{I}$ are used to indicate which installations are used in the network. It is required that exactly one potential installation is chosen for each cell. Binary variables $x_{mi}, m \in \mathcal{M}, i \in \mathcal{I}$, for each pair of mobile m and installation i indicate whether or not m is served by i. We require $x_{mi} \leq z_i$ to ensure that only selected installations serve mobiles.

Bounded continuous variables are used for the uplink power level $p^{\uparrow}_{m}$ of mobile m, for the total received power $\bar{p}^{\uparrow}_{id}$ at installation i in snapshot d, for the downlink transmission power level $p^{\downarrow}_{im}$, and for the for the total downlink transmission power $\bar{p}^{\downarrow}_{id}$ of installation i in snapshot d. The transmission powers of the pilot $\hat{p}^{\downarrow}_{i}$ and the other common channels $\check{p}^{\downarrow}_{i}$ are fixed.

If an installation serves a mobile, the required carrier-to-interference ratios for up- and downlink have to be fulfilled. To ensure this, we multiply the right hand side of the CIR inequalities in the uplink (3) and in the downlink (5) by x_{mi}. The resulting quadratic inequalities is linearized using a *big-M* formulation. We show this exemplarily for the downlink. Defining the denominator of the modified downlink CIR inequality (5) as

$$\phi(m,i) = \bar{\omega}_{im}\, \gamma^{\downarrow}_{im}\, (\bar{p}^{\downarrow}_{i\rho(m)} - \alpha^{\downarrow}_{m}\, p^{\downarrow}_{im}) + \sum_{\substack{j \in \mathcal{I} \\ j \neq i}} \gamma^{\downarrow}_{jm}\, \bar{p}^{\downarrow}_{j\rho(m)}$$

we can write

$$\gamma^{\downarrow}_{im}\, p^{\downarrow}_{im} \geq \mu^{\downarrow}_{m} \phi(m,i)\, x_{mi} + \mu^{\downarrow}_{m} \eta_m\, x_{mi}\,.$$

Setting *big-M* to

$$\Theta_{im} = \bar{\omega}_{im}\, \gamma^{\downarrow}_{im}\, \Pi^{\max\downarrow}_{i} + \sum_{\substack{j \in \mathcal{I} \\ j \neq i}} \gamma^{\downarrow}_{jm}\, \Pi^{\max\downarrow}_{j} + \eta_m$$

we obtain a linearized version of (5) suitable for the MIP model

$$\frac{\gamma^{\downarrow}_{im}}{\mu^{\downarrow}_{m}}\, p^{\downarrow}_{im} - \phi(m,i) - \Theta_{im}\, x_{mi} \geq \eta_m - \Theta_{im}\,. \tag{16}$$

The η_m on the right-hand side cancels out when Θ is expanded.

Mixed integer rounding (MIR) cuts, see for example [33,28], can be added to the model in order to tighten the linear relaxation of the model. The following valid inequality is derived from the modified downlink CIR constraint:

$$x_{mi} - \Big(1/\mu^{\downarrow}_{m} + \bar{\omega}_{im} \alpha^{\uparrow}_{m}\Big) \frac{\gamma^{\downarrow}_{im}}{\eta_m}\, p^{\downarrow}_{im} \;\leq\; 0\,.$$

Note that $\bar{p}^{\uparrow}_{i}$ is dropped from the constraint. This is feasible, because its value has to be non-negative and its coefficient is one. The uplink case can be dealt with in a similar manner.

4.2. Assessment

We conducted extensive computational experiments [9, 11] with snapshot-based models using several different objectives. In the cited publications the goal was to minimize network cost while maintaining acceptable quality. For the computational results in this paper, we essentially fixed the network cost and tried to optimize the quality of the network. All our computational studies, including the ones presented in Section 6, reveal some shortcomings of the snapshot model.

The most serious issue is their poor solvability due to the very high dynamic range of the input data and excessive symmetries. The high dynamic range is due to the necessity to deal with attenuation values in linear scale. Relevant attenuation coefficients γ vary between 60 dB and 160 dB. They apply to output powers $p_m^{\uparrow}$ of mobiles in the range of -50 dBm up to 21 dBm. This results in a dynamic range of received powers of 171 dB; these are 17 orders of magnitude. With these dynamics, we are on the very border of precision for double precision IEEE floating point arithmetic. The symmetries are caused by alternative but relatively similar installations and mobiles located close to each other. A significant and bad consequence of the MIP's poor solvability in our studies is that only very few snapshots could be handled in an optimization run; often only a single snapshot. This is simply not enough to obtain significant optimization results, since the whole idea of snapshot evaluations is based on the hope of convergence for a high number of snapshots.

Furthermore, the model has no notion of a coverage area. Thus, coverage is often neglected in very low traffic areas. A possible remedy would be the introduction of artificial snapshots to cover these areas. Last but not least, the number of variables in the model grows with the number of potential installations times the number of mobiles. Even though we apply elaborate preprocessing to remove impossible combinations, this severely limits the scalability of the model.

To overcome these problems of poor numerics and scalability, we present a new optimization model in the next section. That model is based on the compact evaluation method presented in Section 3 and works on the basis of average load rather than snapshots.

5. Optimizing the Coupling Matrix

The performance properties of a particular network design on a given snapshot are to a large extent described by the coupling matrices $C^{\uparrow}$ and $C^{\downarrow}$ derived in Section 3. The generalized average coupling matrices described in Section 3.3 provide an estimation of the network properties that is independent of single snapshots. This leads to the idea of looking at radio network design as a problem of designing a "good" average coupling matrix.

At the core of the model developed in this section resides a linear description of the coupling matrix for the potential network designs. In addition, we impose coverage constraints inspired by the performance metrics described in Section 2.2. Our presentation will focus only on the downlink and on the average-load based description with best server areas as presented in Section 3.3, an extension to the uplink direction is obvious.

Again, we describe any possible network design by an incidence vector $\mathbf{z} \in \{0,1\}^{|\mathcal{I}|}$. For a feasible network design, we have to choose exactly one installation for each cell. We call the set of network designs that fulfill this condition $\mathcal{F}$.

5.1. Determining the Entries of the Coupling Matrix

Recall the entries of the average coupling matrix as given in (15). All entries in row i of the coupling matrix are computed by integrating over the area A_i served by cell i. For determining the elements of the coupling matrix, it is therefore crucial to determine these best server areas.

5.1.1. Server of Single Points. Consider a point p in the planning area. We introduce a decision variable $c_i^{(p)}$ that expresses whether or not p is served by installation i. This is the case if and only if a) installation i is selected and b) no installation with a stronger signal at p is selected. In the example in Fig.2, p is served by i if and only if installation j (and not k) is selected for the right-hand cell. Assuming that the pilot powers of all antennas are set to the same value, this depends only on the attenuation values $\gamma^{\downarrow}$. We denote by $D_i^{(p)}$ the set of all installations that *dominate* i at p:

$$D_i^{(p)} = \left\{ j \in \mathcal{I} \, : \, \gamma_{jp}^{\downarrow} > \gamma_{ip}^{\downarrow} \right\}$$

In the example, only k dominates i at p, so $D_i^{(p)} = \{k\}$. We have the relation

$$c_i^{(p)} = 1 \Leftrightarrow z_i = 1 \wedge z_j = 0 \; \forall \; j \in D_i^{(p)} \, .$$

This relation can be expressed by linear inequalities as follows:

$$\begin{aligned} c_i^{(p)} &\geq z_i - \sum_{g \in D_i^{(p)}} z_g \\ c_i^{(p)} &\leq z_i \\ c_i^{(p)} &\leq 1 - z_g \qquad\qquad \forall \; g \in D_i^{(p)} \end{aligned} \tag{17}$$

5.1.2. Partitioning the Planning Area. The above construction can be carried out for all points in the planning area. The area served by installation i is then the set of all points $p \in A$ with $c_i^{(p)} = 1$. However, this linear description of the area served by i may be very large On the other hand, the description for a single point p depends only on the set $D_i^{(p)}$ of installations that dominate i at p: all points with the same set of installations dominating i will lead to the same inequalities (17). We aggregate these points into one set. This yields a partition of the planning area A according to the installations that dominate i. For a given

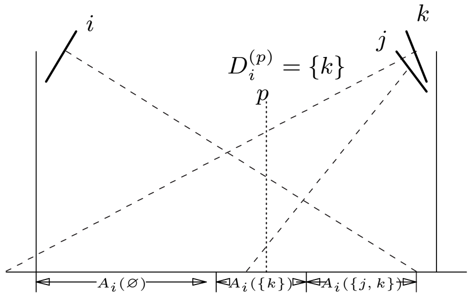


Fig. 2. Example

set $D \subset \mathcal{I}$, $i \notin D$ of installations, we define $A_i(D)$ as the set of points in which exactly the installations in D dominate i:

$$A_i(D) := \left\{ p \in A \,:\, D_i^{(p)} = D \right\}$$
$$= \left\{ p \in A \,:\, \gamma_{jp}^{\downarrow} > \gamma_{ip}^{\downarrow} \text{ for } j \in D,\, \gamma_{jp}^{\downarrow} \leq \gamma_{ip}^{\downarrow} \text{ for } j \notin D \;\right\}$$

These sets do not actually for a partitioning of the area, since there might be many points receiving the same signal strength from different installations. This can be resolved by breaking ties arbitrarily (which is on all accounts valid if the set of such points is very small). However, a more refined version of this model should include soft handover and thereby deal with regions of similar signal strength from several installations in a different way. Assuming that there are no "ties", we have

$$A = \biguplus_{D \subset \mathcal{I}} A_i(D)\,.$$

The example partitioning generated by installation i is shown in Fig.2. Either all points in $A_i(D)$ are served by i or none, depending on whether or not any installation in D is selected. We thus define a binary variable $c_i^{(D)} \in \{0,1\}$ stating wether or not this is the case, and couple it to the decision variables $\mathbf{z}$ in analogy to (17). For a given network design $\mathbf{z} \in \mathcal{F}$ we then know that

$$A_i(\mathbf{z}) = \biguplus_{\substack{D \subset \mathcal{I}\,: \\ c_i^{(D)} = 1}} A_i(D)\,. \tag{18}$$

We obtain a mathematical programming model that describes the best server area of any network design. This can be used to calculate both the main diagonal entries of the downlink coupling matrix and the (interference dependent) off-diagonal entries.

5.1.3. Calculating Main Diagonal Entries and Noise Load. The main diagonal entries of $C^{\downarrow}$ are given as

$$C^{\downarrow}_{jj} := \int_{p \in A_j} \bar{\omega}_p \, l^{\downarrow}_p \, dp \, .$$

Using the partitioning (18) of A_i, we can calculate this value for any network design $\mathbf{z} \in \mathcal{F}$ by summing the contributions to $C^{\downarrow}_{jj}$ on each set:

$$C^{\downarrow}_{ii}(\mathbf{z}) = \sum_{D \subset \mathcal{I}} \left(\int_{p \in A_i(D)} \bar{\omega}_p \, l^{\downarrow}_p \, dp \right) c^{(D)}_i \tag{19}$$

The noise load $p^{(\eta)}_i$ can analogously be determined:

$$p^{(\eta)}_i(\mathbf{z}) = \sum_{D \subset \mathcal{I}} \left(\int_{p \in A_i(D)} \frac{\eta_p}{\gamma^{\uparrow}_{ip}} \, l^{\downarrow}_p \, dp \right) c^{(D)}_i$$

5.1.4. Calculating Off-Diagonal Entries. The same principle can be used to determine the off-diagonal elements of $C^{\downarrow}$. However, there is one more difficulty. The main diagonal elements only depend on the service area A_i of installation i, whereas the off-diagonal element $C^{\downarrow}_{ij}$ are also influenced by the settings for installation j. We thus introduce another dependent binary variable $c^{(D)}_{ij}$ that specifies whether or not any contribution to $C^{\downarrow}_{ij}$ is generated on $A_i(D)$. This is the case if and only if $A_i(D)$ belongs to the service area of installation i and installation j is selected

$$c^{(D)}_{ij} = c^{(D)}_i \cdot z_j$$

This product of binary variables can be transformed to three linear inequalities in the canonical way:

$$\begin{aligned} c^{(D)}_{ij} &\leq z_j \\ c^{(D)}_{ij} &\leq c^{(D)}_i \\ c^{(D)}_{ij} &\geq c^{(D)}_i + z_j - 1 \end{aligned}$$

The entry $C^{\downarrow}_{ij}$ is now determined for any network design $\mathbf{z} \in \mathcal{F}$ in analogy to (19):

$$C^{\downarrow}_{ij}(\mathbf{z}) = \sum_{D \subset \mathcal{I}} \left(\int_{p \in A_i(D)} \frac{\gamma^{\downarrow}_{jp}}{\gamma^{\downarrow}_{ip}} \, l^{\downarrow}_p \, dp \right) c^{(D)}_{ij}$$

Technically speaking the matrix $C^{\downarrow}(\mathbf{z})$ as defined here has dimension $|\mathcal{I}| \times |\mathcal{I}|$. The network's actual coupling matrix $C^{\downarrow}$ of dimension $|\mathcal{N}| \times |\mathcal{N}|$ can of course be obtained by deleting the all-zero rows and columns of installations that have not been chosen. The analogue applies to $\mathbf{p}^{(\boldsymbol{\eta})}(\mathbf{z})$ and $\mathbf{p}^{(\boldsymbol{\eta})}$.

5.1.5. Coverage. There is another condition for a point to be served by installation i: the pilot signal has to be received with sufficient absolute strength ($\mathrm{E_c}$ coverage) and signal quality ($\mathrm{E_c/I_0}$ coverage) for the user to register with the network. We only treat the first point here. Assuming a fixed pilot power level, this means that the attenuation $\gamma^{\downarrow}_{ip}$ to any point p served by in question must not fall below a threshold value $\bar{\gamma}$. This condition can easily be added to the definition of $A_i(D)$. For imposing additional coverage constraints (e. g. that 99% of the area A must be covered by the network), we can then use the area covered by the cell of installation i:

$$|A_i(\mathbf{z})| = \sum_{D \subset \mathcal{I}} |A_i(D)| \; c_i^{(D)}$$

The total area covered by the network is $|A(\mathbf{z})| = \sum_{i\in\mathcal{I}} |A_i(\mathbf{z})|$. The *load* covered by the individual cells can be read off the main diagonal elements of the matrix.

5.2. Objective and Additional Constraints

The model presented in the previous paragraphs can be used to design the downlink coupling matrix $C^{\downarrow}$ in a variety of settings. For the following, minimization of the total downlink load is emphasized, that is, minimizing the sum $\mathbf{1}^T\bar{\mathbf{p}}^{\downarrow}$ of the components of the downlink power vector. This vector cannot simply be determined by designing the matrix, since it is the solution of the (scaled) downlink coupling equation (14). Solving (14) when designing the matrix falls out of the scope of a linear optimization model. However, at the same time it should be ensured that other performance measures are not compromised when minimizing downlink load. We present one way to formulate a full optimization model for network design based on the matrix entries, others are likely to be investigated in the future.

In the case of network tuning, there is an original network design $\mathbf{z}^0$ on which we want to improve. The associated power vector $\bar{\mathbf{p}}^{\downarrow}(\mathbf{z}^0)$ can be computed for this network design as described in Section 3. We propose to use the following optimization model for network design:

$$\begin{aligned}
\min_{\mathbf{z}\in\mathcal{F}} \quad & \mathbf{1}^T \left(C^{\downarrow}(\mathbf{z}) \;\; \bar{\mathbf{p}}^{\downarrow}(\mathbf{z}^0) + \mathbf{p}^{(\boldsymbol{\eta})}(\mathbf{z}) \right) \\
s.t. \quad & |A(\mathbf{z})| \geq |A(\mathbf{z}^0)| && (20) \\
& \mathbf{1}^T \mathrm{diag}\left(C^{\downarrow}(\mathbf{z}) \right) \geq \mathbf{1}^T \mathrm{diag}\left(C^{\downarrow}(\mathbf{z}^0) \right) && (21)
\end{aligned}$$

The basic idea is to approximate the quantities that cannot be computed in a linear model by the referring values of the reference network. This is valid as long as the optimization result is "similar" to the reference network. In the objective function, we use the downlink power vector $\bar{\mathbf{p}}^{\downarrow}(\mathbf{z}^0)$ in the right hand side of the fix point form (14). Constraint (20) ensures that $\mathrm{E_c}$ coverage does not diminish compared to the reference network. An analogue to this condition is stated in (21), but the covered area is here weighted with the user load. This makes sure that coverage in highly frequented areas is not traded for coverage in areas with little traffic.

When producing good solutions in the sense of this objective function, different issues play a role:

- Interference reduction is rewarded: cells try to minimize their influence on the neighbors
- Noise load is reduced: cells try to improve the attenuation to their own mobiles
- Load is shifted away from overloaded cells to emptier cells

These points (which are to some extent contradictory) suggest that network designs that improve on the objective function also have better performance properties. Computational experiments (cf. Section 6) confirm this.

However, there are some disadvantages of this approach. First, this optimization model depends largely on the input network. Different input networks will lead to different optimization results. When using this model iteratively, no stabilization of the result could be observed as of yet. Second, if the used approximation is not valid on the entire set $\mathcal{F}$, the objective function might be misleading and results of less quality than the reference network design are produced. This has also been observed in computational experiments.

5.3. Model Size

The full model as described above is of exponential size in the input length, since the contribution variables are indexed with all subsets $D \subseteq \mathcal{I}$. However, the relevant sets of dominating installations are comparatively small in practice, and it is crucial to determine these sets beforehand. This is done in a preprocessing step. The area A is usually described in the finite form of a pixel grid. We determine for each element $p \in A$ and for all $i \in \mathcal{I}$ with $\gamma^{\downarrow}_{ip} \geq \bar{\gamma}$ the dominator set $D^{(p)}_i$. There are at most $|A| \cdot |\mathcal{I}|$ such sets, leading to a maximum of $|A| \cdot |\mathcal{I}|$ variables $c^{(D)}_i$. However, this number is in practice considerably reduced by the $\mathrm{E_c}$-coverage condition, since it typically admits only cells in a regional vicinity of p. The constraint that one installation per cell must be selected can be used to further reduce the number of relevant dominator sets in an obvious way. The potential interference sources per pixel is at most $|\mathcal{I}| \cdot (|\mathcal{I}| + 1)/2$. The number of variables $c^{(D)}_{ij}$ is thus $O(|A| \cdot |\mathcal{I}|^3)$. At the expense of accuracy in determining the matrix entries, variables $c^{(D)}_i$ and $c^{(D)}_{ij}$ with very small contributions $\int_{p\in A_i(D)} \bar{\omega}_p \, l^{\downarrow}_p \, dp$ and $\int_{p\in A_i(D)} (\gamma^{\downarrow}_{jp}/\gamma^{\downarrow}_{ip}) \, l^{\downarrow}_p \, dp$, respectively, can be deleted from the model.

Example model sizes for the planning scenarios used for computation are listed in Table 2. In all cases, three potential installation configurations are allowed for each central cell; cells at the border of the scenarios were not varied. The installation configurations differ only in tilt (mechanical and electrical). All dominator sets and hence all variables $c^{(D)}_i$ are included in the model. Variables

Table 2. Size of Matrix Design Models

Scenario	Cells	Options	Binary Vars	Constraints	Nonzeroes
The Hague	30	3	123,426	212,861	822,921
Berlin	108	3	101,756	159,782	627,409
Lisbon	110	3	108,929	182,133	724,842

$c_{ij}^{(D)}$ are deleted if their potential contribution to the referring off-diagonal element is less than 10^{-4}. This ensures that main diagonal elements and noise load are computed with the MIP solver's precision. The off-diagonal elements have in all tests been computed with an absolute error no larger than 10^{-3}.

In general, the model sizes are acceptable and still tractable with standard MIP solvers. A little surprising is the large size of the model for The Hague, which has the smallest number of cells. This is due to some peculiarities in the data; radio signals propagate with little attenuation across the scenario, so mutual coupling is very high.

6. Computational Results

This section describes the computational results from the models described in Sections 4 and 5. The results are analyzed with the methods outlined in Section 3. We focus on adjusting the *tilts* of the antennas in a given radio network starting from reference networks based on GSM settings. The potential impact of adjusting tilts is described in Section 1.3, a graphical illustration can be found in Fig. 1 on page 13.

6.1. The Scenarios

Three real-world scenarios developed within the MOMENTUM project [32] are used as test cases. We give only a brief description of the scenarios, the details can be found in [34,15,12,13].

6.1.1. General description. The scenarios cover the downtown areas of The Hague, Berlin, and Lisbon; some key information can be found in Table 3. The total (downlink) load in the scenario as given in the table is $\int_A l_p^{\downarrow}\, dp$. The locations of base stations are called *sites*, there are commonly three antennas and thereby three cells associated to each base station.

Table 3. The scenarios

Name	Sites	Cells	Avg. Users	Total DL-Load	Area km^2
The Hague	12	36	673	20.38	16
Berlin	65	193	2328	70.05	56
Lisbon	60	164	3149	77.02	21

The smallest scenario is The Hague. In the original setting, there are 76 potential site locations; since it is a low traffic scenario, we select a subset of 12 sites and 36 cells for the UMTS network with a modified set covering model and conducted the optimization on the basis of this network. The traffic in the larger Berlin scenario, which has 65 sites and 193 cells, is about 90% of the value for Lisbon. However, the latter scenario with 60 sites and 164 cells has less infrastructure for a higher traffic load. This high load is concentrated on an area of $21\,\mathrm{km}^2$, less than half of the Berlin's scenario area ($56\,\mathrm{km}^2$). This makes Lisbon a scenario with high traffic concentration.

The *traffic mix* of services is not listed in the table, it is comparable for all the scenarios: about 55 % speech telephony users, 17 % data streaming users, and 7 % video telephony users are present, the residual active users being distributed over a variety of data services. While speech and video telephony users are relatively symmetric in their resource requests in up- and downlink, the noticeable presence of demanding data streaming users, which have only a negligible effect in the uplink, makes the downlink the limiting direction.

6.1.2. Networks and reference network design. The radio networks considered for optimization are provided by the telecommunication operators within the MOMENTUM consortium. In the reference network designs, tilts are not adjusted specifically, but a global electrical tilt of 6° and mechanical tilt of 0° are used. Detailed analysis reveals that for The Hague and Berlin these networks perform reasonable well in most cells. This is not true for the high load areas, especially in Lisbon. The allowed variations are limited to electrical tilts in Berlin and The Hague, in Lisbon also mechanical tilts can be adjusted. The results listed below show that by adjusting tilts the load can be better distributed among cells and more traffic can be handled.

6.2. The Algorithms

We optimize the original networks with three different algorithms:

1. (*Snapshot*) A snapshot based MIP model as mentioned in Section 4. Due to the described computational difficulties, only a single snapshot is used.
2. (*Matrix*) A MIP model for designing the downlink coupling matrix as described in Section 5. The power and scaling vector of the reference network is used as a starting point. The network in The Hague is treated in its entirety, while in the larger scenarios tilt variations are allowed only in subareas that are problematic and highly loaded in the reference network design.
3. (*Local Search*) A local search algorithm that repeatedly enumerates all possible installations for each cell. For each variation, the relevant performance metrics are computed with the methods presented in Section 3, the variation that performs best on a weighting of the metric is chosen to proceed. This is repeated in passes over all cells, until no further improvement can be found for any cell within a pass.

The MIPs resulting from the snapshot and matrix based models are solved using CPLEX 9.0 with a limit of about 2 days running time on contemporary PCs. The local search procedure takes at most a few hours to terminate.

6.3. Results

Table 4 lists relevant performance metrics as described in Section 2.2 for the networks that are computed by the different algorithms as well as the reference networks.

6.3.1. Performance metrics in the table. Lower values indicate better performance in all cases. The columns labeled E_c and E_c/I_0 give the percentage of the total area with insufficient E_c and E_c/I_0 coverage, respectively. The column *Pilot* holds the percentage of the area with pilot pollution. A location is here defined as suffering of pilot pollution if more than three of the strongest pilot signals are within a 5 dB range. The *Load* columns show the averaged load in percent of all cells in up- and downlink, respectively. The *Coupling* values (Cpl.), computed on the basis of the coupling matrices, give an indication of which percentage of the cells' transmission power is on average needed due to interference. For the downlink, the limiting direction here, we also list the number of *overloaded cells* (Ovl. cells) and the *total tranmission power* (TP) of all cells.

Table 4. Computational results

Scenario	E_c	E_c/I_0	Pilot	Uplink Load	Uplink Cpl.	Downlink Ovl.	Downlink Load	Downlink Cpl.	TP
Algorithm	A%	A%	A%	ϕ %	ϕ %	cells	ϕ %	ϕ %	dBm
The Hague									
Original	0.00	10.39	22.76	23.78	50.23	9	45.34	61.88	326
Snapshot	0.00	3.02	18.19	22.84	48.90	7	41.11	59.53	295
Matrix	0.00	0.86	10.82	20.07	44.74	3	34.12	50.30	245
Local Search	0.00	0.60	11.46	20.27	44.57	3	35.37	51.98	254
Berlin									
Original	0.68	2.95	8.78	12.35	40.43	12	24.46	30.99	942
Snapshot	0.68	2.95	8.78	12.35	40.44	12	24.47	30.99	942
Matrix	1.55	0.50	5.61	11.22	36.26	6	21.85	25.60	841
Local Search	0.37	0.39	4.35	11.45	35.86	6	22.77	27.17	877
Lisbon									
Original	1.55	48.59	4.33	16.24	34.70	32	33.48	38.49	1096
Snapshot	1.61	47.99	4.15	16.08	34.28	33	33.45	38.09	1095
Matrix	1.81	38.37	3.71	14.97	33.38	18	29.23	32.49	980
Local Search	1.48	27.55	3.32	15.45	31.82	14	27.42	31.66	897

6.3.2. Assessment of results. The Matrix and Local Search algorithm were able to substantially improve on the performance of the reference networks in all cases. The number of overloaded cells could be at least halved and the interference was reduced by about 18%. Hardly any improvement was achieved with the snapshot model. The explanation for the comparatively weak results is the small number of snapshots that could be considered; one snapshot was apparently not enough to yield stable results performing well on the stochastical average. In the Lisbon scenario, the Local Search noticeably outperforms the Matrix design algorithm. This is one example of poor performance of the objective proposed in Section 5.2.

7. Conclusion

In this paper, we have shown a way to overcome the notorious problems of scalability and statistical reliability of network design and optimization models that mimic traditional network quality assessment schemes working on user snapshots. The results described are a step ahead towards automatic planning of well performing UMTS networks for real-world problems.

Our main contributions are a generalization of the dimension reduction method for the power control problem in interference limited systems to stochastical average load and a new mathematical optimization formulation based on this.

The generalization to average user load is not sophisticated, but it allows for the first time to rapidly compute a good estimate of the cell load in UMTS networks. This enables a fast local search procedure. Moreover, we have shown how to use our approach in a new optimization formulation that substantially differs from all optimization models proposed for the problem and eliminates their main difficulties. We carried out automatic optimization on real-world data adjusting antenna tilts, which is at the time a relevant issue for telecommunication operators. The results show that our approaches can be efficiently implemented and outperform the "classical" snapshot methods of network design.

Based on these developments, there is a variety of questions to pursue in the future. First of all, a careful assessment of the estimation errors of our average-based load calculation is needed. The core of the matrix design optimization model seems better suited to the problem than the previously known approaches. It seems promising to employ it with different objectives and constraints and extend it in several ways. Even a formulation incorporating soft hand-over is conceivable. Soft hand-over is still ignored in all known automatic planning methods that are not based on network simulation.

References

1. E. Amaldi, A. Capone, and F. Malucelli. Planning UMTS base station location: Optimization models with power control and algorithms. *IEEE Transactions on Wireless Communications*, 2(5):939–952, Sept. 2003.

2. E. Amaldi, A. Capone, F. Malucelli, and F. Signori. UMTS radio planning: Optimizing base station configuration. In *Proc. of IEEE VTC Fall 2002*, volume 2, pages 768–772, 2002.
3. E. Amaldi, A. Capone, F. Malucelli, and F. Signori. Optimizing base station location and configuration in UMTS networks. In *Proc. of INOC 2003*, pages 13–18, 2003.
4. A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Classics in Applied Mathematics. 9. Philadelphia, PA: SIAM,. xx, 340 p., 1994.
5. H. L. Bertoni. *Radio Progation for Modern Wirless Applications*. Prentice-Hall, 2000.
6. D. Catrein, L. Imhof, and R. Mathar. Power control, capacity, and duality of up- and downlink in cellular CDMA systems. Technical report, RWTH Aachen, 2003.
7. D. Catrein and R. Mathar. On the existence and efficient computation of feasible power control for cdma cellular radio. In *Proc. of ATNAC 2003*, Melbourne, Australia, Dec. 2003.
8. A. Eisenblätter, E. R. Fledderus, A. Fügenschuh, H.-F. Geerdes, B. Heideck, D. Junglas, T. Koch, T. Kürner, and A. Martin. Mathematical methods for automatic optimisation of UMTS radio networks. Technical Report D4.3, Momentum IST-2000-28088, 2003.
9. A. Eisenblätter, A. Fügenschuh, H.-F. Geerdes, D. Junglas, T. Koch, and A. Martin. Integer programming methods for UMTS radio network planning. In *Proc. of WiOpt'04*, Cambridge, UK, 2004.
10. A. Eisenblätter, A. Fügenschuh, T. Koch, A. M. C. A. Koster, A. Martin, T. Pfender, O. Wegel, and R. Wessäly. Modelling feasible network configurations for UMTS. Technical Report ZR-02-16, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Germany, 2002.
11. A. Eisenblätter, H.-F. Geerdes, D. Junglas, T. Koch, T. Kürner, and A. Martin. Final report on automatic planning and optimisation. Technical Report D4.7, IST-2000-28088 Momentum, 2003.
12. A. Eisenblätter, H.-F. Geerdes, U. Türke, and T. Koch. Momentum data scenarios for radio network planning and simulation (extended abstract). Technical Report ZR-04-07, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Germany, 2004.
13. A. Eisenblätter, H.-F. Geerdes, U. Türke, and T. Koch. Momentum data scenarios for radio network planning and simulation (extended abstract). In *Proc. of WiOpt'04*, Cambridge, UK, 2004.
14. A. Eisenblätter, T. Koch, A. Martin, T. Achterberg, A. Fügenschuh, A. Koster, O. Wegel, and R. Wessäly. Modelling feasible network configurations for UMTS. In G. Anandalingam and S. Raghavan, editors, *Telecommunications Network Design and Management*. Kluwer, 2002.
15. H.-F. Geerdes, E. Lamers, P. Lourenço, E. Meijerink, U. Türke, S. Verwijmeren, and T. Kürner. Evaluation of reference and public scenarios. Technical Report D5.3, IST-2000-28088 Momentum, 2003.
16. N. Geng and W. Wiesbeck. *Planungsmethoden für die Mobilkommunikation*. Springer, 1998.
17. A. Gerdenitsch, S. Jakl, M. Toeltsch, and T. Neubauer. Intelligent algorithms for system capacity optimization of UMTS FDD networks. In *Proc. IEEE 4th International Conference on 3G Mobile Communication Technology*, pages 222–226, London, 2002.
18. S. A. Grandhi, R. Vijayan, and D. J. Goodman. Centralized power control in cellular radio systems. *IEEE Transactions on Vehicular Technology*, 42(4):466–468, 1993.
19. S. A. Grandhi, R. D. Yates, and D. J. Goodman. Resource allocation for cellular radio systems. *IEEE Transactions on Vehicular Technology*, 46(3):581–587, 1997.
20. S. V. Hanly. An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity. *IEEE Journal on Sleected Areas in Communications*, 13(2):1332–1340, Sept. 1995.
21. H. Holma and A. Toskala. *WCDMA for UMTS*. Wiley, 2001.
22. S. B. Jamaa, Z. Altman, J. Picard, B. Fourestie, and J. Mourlon. Manual and automatic design for UMTS networks. In *Proc. of WiOpt'03*, Sophia Antipolis, France, Mar. 2003. INRIA Press.
23. A. Jedidi, A. Caminada, and G. Finke. 2-objective optimization of cells overlap and geometry with evolutionary algorithms. In *Proc. of EvoWorkshops 2004*, volume 3005 of *Lecture Notes in Computer Science*, pages 130–139, Coimbra, Portugal, Apr. 2004. Springer.
24. T. Kürner. Propagation models for macro-cells. In *Digital Mobile Radio towards Future Generation Systems*, pages 134–148. COST Telecom Secretariat, 1999.

25. J. Laiho, A. Wacker, and C. Johnson. Radio network planning for 3G. Technical Report TD (02) 061, COST 273, Helsinki, Finland, 2002.
26. J. Laiho, A. Wacker, and T. Novosad, editors. *Radio Network Planning and Optimization for UMTS*. Wiley, 2001.
27. K. Leibnitz. *Analytical Modeling of Power Control and its Impact on Wideband CDMA Capacity and Planning*. PhD thesis, University of Würzburg, 2003.
28. H. Marchand and L. A. Wolsey. Aggregation and mixed integer rounding to solve MIPs. *Operations Research*, 49:363–371, 2001.
29. R. Mathar and M. Schmeink. Optimal base station positioning and channel assignment for 3G mobile networks by integer programming. *Ann. of Operations Research*, (107):225–236, 2001.
30. E. Meijerink, E. R. Fledderus, O. C. Mantel, U. Türke, T. Winter, and A. Serrador. Characterisation of the impact of UMTS multi-service traffic on the air interface. Technical Report D2.3, IST-2000-28088 Momentum, 2003. (Internal).
31. L. Mendo and J. M. Hernando. On dimension reduction for the power control problem. *IEEE Transactions on Communications*, 49(2):243–248, Feb. 2001.
32. Momentum Project, IST-2000-28088. Momentum public UMTS planning scenarios. Available online at `http://momentum.zib.de/data.php`, 2003.
33. G. L. Nemhauser and L. A. Wolsey. A recursive procedure to generate all cuts for 0-1 mixed integer programs. *Mathematical Programming*, 46:379–390, 1990.
34. B. Rakoczi, E. R. Fledderus, B. Heideck, P. Lourenço, and T. Kürner. Reference scenarios. Technical Report D5.2, IST-2000-28088 Momentum, 2003.
35. I. Siomina and D. Yuan. Pilot power optimization in WCDMA networks. In *Proc. of WiOpt'04*, Cambridge, UK, 2004.
36. U. Türke, R. Perera, E. Lamers, T. Winter, and C. Görg. An advanced approach for QoS analysis in UMTS radio network planning. In *Proc. of the 18th International Teletraffic Congress*, pages 91–100. VDE, 2003.
37. U. Türke, R. Perreira, E. Lamers, T. Winter, and C. Görg. Snapshot based simulation techniques for UMTS network planning. In *Proc. of the IST-Mobile Summit*, Aveiro, Portugal, 2003.
38. P. Värbrandt and D. Yuan. A mathematical programming approach for pilot power optimization in WCDMA networks. In *Proc. of ATNAC 2003*, Melbourne, Australia, Dec. 2003.
39. R. M. Whitaker and S. Hurley. Evolution of planning for wireless communication systems. In *Proc. of HICSS'03*, Big Island, Hawaii, Jan. 2003. IEEE.