Tradeoff Between Coverage And Capacity In Dynamic Optimization of 3G Cellular Networks

G. Hampel, K. L. Clarkson, J. D. Hobby, P. A. Polakos

ABSTRACT

For 3G cellular networks, capacity becomes an additional objective along with coverage when characterizing the performance of high-data-rate services. Since network capacity depends on the degree traffic load is balanced over all cells, changing traffic patterns demand dynamic network reconfiguration to maintain good performance. We study the competitive character network coverage and capacity have in such a network optimization process, concentrating on a four-cell sample network that captures the essence of layout irregularities found in network designs of real markets. Optimizing the downlink through antenna tilt, pilot power fraction and cell power adjustment, we find that each set of variables provides a distinct solution space with a different coverage-capacity tradeoff behavior. This allows selecting one over the other based on the particular application.

I. INTRODUCTION

With the introduction of high-data-rate services, network capacity has become a crucial objective for the performance characterization of 3G access networks [1][2]. It therefore must be considered along with network coverage as an objective during network planning [3][4][5]. Since network capacity strongly depends on how well the actual traffic load is balanced over all cells, dynamic load balancing routines can be applied in response to changing traffic patterns after a network has gone into service [3][6][7]. Such load balancing operations, in turn, have to include coverage as on additional objective.

For GSM/GPRS and TDMA IS-136, cell load variations due to changing traffic patterns can be rebalanced by changing the number of channels per cell, and a subsequent recomputation of the frequency plan helps to mitigate the interference problems this introduces [6][7][8]. This approach is not feasible for CDMA networks because the wide carrier bandwidth prevents load balancing by swapping carriers among cells, and a frequency plan to mitigate interference does not exist. Further, interference affects both coverage and capacity, coupling both objectives tightly together. Network optimization is typically done by moving cell boundaries and soft handoff areas via the adjustment of cell locations, antenna configurations and power levels, each of which affects both objectives simultaneously [9]-[13]. Since the functional dependence between objectives and adjustment parameters is complicated, automated optimization approaches have been proposed by various authors [5][9][10][12].

In this paper, we illustrate the general problem of simultaneously optimizing network coverage and capacity as it occurs when performing load-balancing operations in 3G cellular networks. The study is done on a four-cell model network with a quadratic layout, which is incompatible with the optimal hexagonal pattern and therefore represents the layout irregularities found in real markets due to cell placement constraints. The small size of the model network makes it an ideal candidate for a systematic study, since network performance changes can easily be attributed to particular configuration adjustments. We further focus on the downlink for the performance analysis, assuming it will be the weaker link in the presence of asymmetrical load as expected for 3G data services [14]. We consider optimizing antenna tilt, cell power level, and pilot power fraction, and we find that under general circumstances, coverage and capacity cannot be optimized simultaneously. We must introduce a performance tradeoff representation in order to capture the complete set of network configurations that have optimum performance but form different compromises between the two objectives. We will demonstrate how this tradeoff representation can be used to differentiate and rank potential tuning parameters with respect to their application-specific impact on overall network performance.

In the following, we first present the definitions of network coverage and capacity as used in this context, and then we introduce the performance tradeoff representation. Next, we describe the performance modeling for the four-cell network and then we present and discuss the tradeoffs obtained from optimization of the model network with respect to a number of variables. Finally, we discuss the different impact these variables have on overall network performance when used to load-balance the network to traffic pattern variations.

II. TRADEOFF BETWEEN NETWORK CAPACITY AND COVERAGE

The outcome of a network performance analysis and optimization can be very dependent on the particular choice of objective functions. In this paper, we focus on performance enhancements in live networks. Since real markets are characterized by irregular network layout, complex propagation patterns, and inhomogeneous traffic distributions that give each cell its own shape, performance and traffic load, coverage and capacity have to be defined in terms of network properties rather than cell properties.
This means that both objectives need to be linked to the actual distribution, i.e. capacity has to capture the degree of load balancing among cells while coverage should be weighted by the local traffic density to capture the network-wide fraction of users that receive adequate service.

We use the following definitions for network coverage and capacity, and we simplify things by restricting ourselves to one service. Since we focus on dynamic adjustment, the definitions are based on a time snapshot or a rather short time interval where fluctuations due to the statistical nature of user arrival can be neglected:

- *Network capacity* is the maximum covered traffic of a known distribution that the network can serve with given resources. This definition automatically captures the degree of load balance between actual traffic and available resources, as well as interference-dependent resource requirements in 3G (W)-CDMA systems.
- *Network coverage* is given by the fraction of traffic-weighted area that can receive the offered service with minimum guaranteed quality of service, where the traffic load is set by the capacity definition.

Note that the two objectives are interdependent and they are defined with respect to the maximum acceptable instantaneous traffic load that can be handled without access blocking.

With these definitions, every network configuration or network design is characterized by a performance pair. Network configurations that result from optimization can be compared with each other in a two-dimensional plot with network coverage and network capacity as components. The performance analysis of all potential network configurations for a given set of tuning parameters produces a cloud of points (Fig.1). A perfect network optimization algorithm that tries to maximize the coverage and capacity objectives would naturally select all those configurations whose performance points form the upper right boundary of this cloud (points A-G in Fig.1). Since these are optimum solutions to the optimization process that differ solely in how they trade off network coverage and capacity, we call these points "Tradeoff Points" and the curve they form a "Tradeoff Curve". The length of the tradeoff curve and its shape are measures for how well network coverage and capacity can be simultaneously optimized.

A simple optimization process for these two objectives, however, proves insufficient, i.e. the tradeoff points found do not necessarily represent the complete set of optimum solutions. Since network coverage depends on interference, reducing traffic load to values below network capacity will result in higher coverage. The overall performance under such circumstances can potentially be better than provided by lower-lying tradeoff points evaluated at maximum traffic load. To illustrate this phenomenon, we will include network performance points for less than maximum traffic load into Fig.1 by changing its ordinate from "Capacity" to "Carried Traffic". For the three tradeoff points, A, C, and E, we have added the performance as it varies with traffic load (black lines, Fig.1), which we will refer to as "Load Curve". The load curve through point E provides better performance than the tradeoff points F and G. The optimum configurations are therefore only in the tradeoff section between points A-E. We will investigate whether the four-cell network suggests that such situations can also occur in real networks.

III. FOUR-CELL NETWORK PERFORMANCE MODELING

The model network consists of four cell sectors at the corners of a square (Fig.2). We will optimize the model network once for homogenous traffic and once for an example of an inhomogenous traffic distribution. The first case is used to study optimization of imperfect network layouts, and the second case serves to study the performance sensitivity to traffic pattern variations and the potential gain that can be achieved through network reoptimization. In the simulation, performance will be evaluated over a uniform grid of $m \times m$ points, where the traffic of the i^{th} point at location (x_i,y_i) has traffic T_i. The two traffic distributions are as follows (Fig.2):

Homogeneous distribution: $T_i = T_0$

Inhomogeneous distribution: $T_i = T_0 \cdot (1 + x_i - y_i), \quad x_i, y_i \in [0,1]$

Here, the dimension of the square layout has been set to one, and T_0 represents the total number of users per area.

To provide a clear picture of the coverage-versus-capacity tradeoff in 3G-network optimization, we will simplify the performance modeling so that it only includes the essential features. We will focus on the downlink using one common pilot channel and a power-controlled dedicated traffic channel with constant data rate. Power control is assumed to be perfect, and network access is assumed to be limited by power amplifier (PA) overload, as can be expected for many 3G applications.

The vertical antenna pattern of each cell sector is approximated by a $\cos^2$- pattern for the main lobe, without side lobes and finite back-lobe suppression, and the azimuthal pattern does not have any structure. The antenna height is set in relative terms by the down angle from the antenna to the center of the network.

The common pilot channel coverage at location *i* is given by $(Ec/Io)_i$ of the strongest server:

$$\text{Location } i \text{ has pilot coverage } \Leftrightarrow \max_k (Ec/Io)_{ik} \geq q_{Pl},$$

where *Ec* denotes the received common pilot channel signal strength of the k^{th} server, q_{Pl} an absolute pilot threshold, and *Io* the receiver RSSI. *Io* can be expressed as

$$Io = N + IF_{Ext} + \sum_{k=1}^{4} b_k \cdot Ec_k / g_k^{PL} = h + \sum_{k=1}^{4} b_k \cdot Ec_k / g_k^{PL},$$

where

N is thermal noise,

IF_{ext} is interference from other surrounding cells,

b_k is PA load fraction of *k*th sector,

g_k^{PL} is the pilot power fraction of the *k*th sector.

Since only four cell sectors are modeled here, the interference contribution of surrounding cells has to be included too. Since we have no spatial distribution for it, we just add it to the thermal noise floor and treat both terms as one common background interference term η.

Ec_{ik} is linked to the pilot EIRP, Pc_k, of server *k* via the path loss L_{ik}, which will be approximated by a power law

$$Pc_k = Ec_{ik} \cdot L_{ik}, \quad L_{ik} = r^k / r_0 .$$

The PA load fraction b_k results from the traffic-channel load analysis and will be solved self-consistently for the entire network. (See below.)

We further assume that any of the four pilots can enter the active set as soon as its pilot signal strength exceeds the same absolute threshold as introduced for the coverage condition: $(Ec/Io)_{ik} \geq q_{Pl}$. We neglect additional relative thresholds. These approximations are not very serious since the soft handoff areas in the given scenarios are rather small.

The dedicated traffic channel will be modeled via SIR estimation. The SIR at location *i* with respect to server *k* is

$$(S/I)_{ik} = S_{ik} / \left(Io - a \cdot b_k \cdot Ec / g_k^{PL}\right),$$

where S_{ik} denotes the received traffic-channel power level, and a is an orthogonality factor with respect to same-cell channelization codes. In simplex mode, the power control algorithm gives $(S/I)_{ik} = 1/PC_{dwn}$,

where PC_{dwn} denotes the downlink pole capacity for this service, including Ebno, processing gain and activity factor.

In soft handoff, we assume that the UE or mobile can perform maximum ratio combining, where the channel estimation is given by the pilot channel. The resulting power control equation is

$$(S/I)_i = \left(\sum_k \sqrt{S_{ik}}\sqrt{Ec_{ik}}\right)^2 \Big/ \sum_k I_{ik} Ec_{ik} = 1/PC_{dwn} ,$$

where the sum is taken over all pilots in the active set. To solve the power control equation in soft handoff, we assume that all contributing traffic channel power levels, P_{ik}, are synchronized:

$$P_{ik} = S_{ik} \cdot L_{ik} = P_{il} \quad \forall l,k \ .$$

Downlink traffic channel coverage is achieved as long as power control stays below the maximum bound: $P_{ik} \le P_{\max}$. In the following analyses, however, coverage is always limited by the common pilot channel.

The PA load fraction $\underline{b}$ can be obtained from the sum of pilot power and the power provided to all traffic channels, the latter weighted with respect to the local traffic density

$$b_k = \left(Pc_k + \sum_i \frac{T_i}{T_0} \cdot P_{ik} \right) \Big/ Ptot_k$$

where $Ptot_{\text{k}}$ denotes the total PA power of the k^{th} sector. Here, the sum is taken over all locations that have active-set membership with respect to the k^{th} server.

During the course of the simulation, we slowly ramp up the total traffic T_0, and for each value of T_0, we self-consistently evaluate network coverage and the load vector, $\underline{b}$. We evaluate over a dense grid starting with common pilot channel coverage and traffic channel coverage based on initial default values for $\underline{b}$. The resulting network coverage is given by

Network coverage = $\sum_i T_i \cdot Cov_i / T_0$,

where $Cov_i = 1$, if location or grid point i has pilot and traffic-channel coverage, and zero otherwise.

The resulting PA load fraction vector serves for the next iteration until convergence has been achieved. The final values for $\underline{b}$ will be used to estimate the next T_0 value. The T_0 -

ramp-up is performed until one of the sectors reaches maximum load, i.e. $b_k = 1$. At that point, network capacity is set as follows:

$$\text{Network Capacity} = \sum_i T_i \cdot Cov_i = T_0 \cdot \text{Network Coverage}\ .$$

The simple connection to network coverage is caused by the fact that coverage is defined based on traffic rather than area.

IV. OPTIMIZATION AND OPTIMIZATION VARIABLES

We will optimize the model network for both traffic distributions and with respect to three different types of parameters:

- Mechanical antenna tilt
- Pilot power fraction (maximum PA power stays the same)
- Max PA power level (pilot power fraction stays the same).

For each of these parameters, we will determine the entire tradeoff curve within a reasonable window for coverage and capacity. Since all three of these tuning parameters essentially cause an area breathing of the associated cell sector, we expect them to create similar tradeoff curves. For every tradeoff configuration, we will evaluate the corresponding load curves using the same formalism as above but with lower bounds on $\underline{b}$. Since only four variables have to be considered at a time, the optimization can be done by brute-force.

V. OPTIMIZATION FOR HOMOGENEOUS TRAFFIC

The optimization for homogeneous traffic starts out from an initial design, in which all cells are set to maximum power, the antennas are tilted so as to point to the center of the network and the pilot fraction is set to 10% of total power. The detailed set of parameters is summarized in Table1. This initial network configuration is symmetrical and has a symmetrical traffic distribution, resulting in the network performance shown in Fig. 3 (open square). The resulting three tradeoff curves (also plotted in Fig.3) show the following features:

- All three tradeoff curves have rather distinct shapes, setting them apart in their potential impact on optimization. Adjustment for power, for instance, gains far less performance than could be achieved via antenna tilt or pilot fraction. Tilt adjustment can simultaneously improve coverage and capacity with respect to the initial configurations. This is achieved by increasing all tilt values from 1.5^0 to 3^0 (Point T2 in Fig.3). Pilot adjustment is the tuning operation best able to increase coverage.
- All three curves seem to indicate a rather large tradeoff between coverage and capacity: A few percent of coverage can be traded for 20-30% capacity.

- The shape of the tradeoff curves allows selecting some of the points as potential candidates and eliminating others. The tilt curve, for instance, is very flat above the kink (Point T1). Tradeoff configurations in this range of the curve should not be considered for implementation since too much coverage has to be sacrificed for only small capacity gains.
- Slope variations as seen in the tilt curve are caused by different types of network configurations. The high-capacity branch of the tilt curve has all antennas tilted equally, while the high-coverage branch has an asymmetric network configuration with two antennas tilted higher than the other two. The pilot curve does not have a kink because the corresponding configurations are all symmetrical (Fig. 3, right-hand side).

Fig. 3 also shows load curves for two points of the tilt tradeoff curve (T1 & T2) and the pilot tradeoff curve (P1 & P2). The power curve has been omitted.

- The load curves generally give a performance tradeoff comparable to the tradeoff curves generated through optimization.
- The high-coverage branch of the tilt tradeoff curve shows worse performance than the load curve through point T2. This means that all tradeoff points of this curve with higher coverage are suboptimal, i.e. they lead to worse performance than the configuration T2 when run at correspondingly lower traffic load. The remaining tilt tradeoff curve comprises optimum solutions, but only the range between T1 and T2 provides a reasonable tradeoff between both objectives.
- The pilot tradeoff curve entirely consists of optimal configuration since it proves better than the individual load curves. However, the potential tradeoff between two points always has to be compared with respect to same value of carried traffic. Choosing for instance configuration P2 over P1 would lead to a coverage gain of less than 1% (evaluated at carried traffic of ~300 users).

The small improvements achievable through optimization and the rather short tradeoff range are due to the fact that network traffic and network layout have the same symmetry, a situation which rarely occurs in real markets.

VI. OPTIMIZATION FOR INHOMOGENEOUS TRAFFIC

When the network traffic pattern is changed from a homogenous distribution to an inhomogeneous one, the initial performance should degrade. This development can be confirmed by overlaying the load curves of the initial configuration for both traffic distributions (load curves through points P1 and P1' in Fig.4a). The corresponding capacity degradation is about 26% (Fig.4a, gray downward pointing arrow). Coverage also deteriorates, although only slightly, since the load curve for inhomogeneous traffic lies left of the original load curve.

Reoptimization for inhomogeneous traffic using adjustment of pilot fraction produces a new tradeoff curve (Fig.4a, open dots). A comparison to the old tradeoff curve, obtained from optimization for homogeneous traffic, shows the following:

- The new tradeoff curve is slightly flatter, and all of its points are optimal solutions. However, the new tradeoff curve lies below the old curve indicating that the load balancing operation has effectively lost performance.
- Choosing a configuration from the new curve (point P10) that has same coverage as the initial configuration for homogeneous traffic (point P1) results in a capacity loss of 7% (arrow between both points). This value, however, has to be compared to the capacity loss of 26% when no load balancing is performed. If desired, a configuration could be chosen with the same capacity as before but 4% less coverage (horizontal arrow in Fig.4a).

The overall performance loss of the load-balancing operation is due to a complex combination of power control, channelization code-orthogonality and cell size variation introduced by the optimization efforts. The latter has shifted the soft-handoff zone toward the heavily loaded cell. Off-loaded users, now fully supported by the neighbor cells, cause more interference to the congested cell since their channel power is not orthogonal anymore to its own channelization codes. This effect is more important for 3G1X which has enhanced orthogonality due to a lower chip rate than for UMTS. The shift of the soft handoff areas is illustrated on the right-hand side of Fig 4a for configurations P1 and P10.

We have repeated the same optimization procedure using antenna tilt instead of pilot fraction (Fig.4b) with the initial configuration based on the tradeoff point T2. As in the prior case, the capacity loss caused by the change of the traffic distribution is about 26%. The new tradeoff curve obtained through network reoptimization provides more optimum solutions than the tilt tradeoff curve for homogeneous traffic before (Fig.4b only shows optimal solutions for comparison purposes). The spatial separation between both tilt tradeoff curves, however, is larger than the separation between the pilot tradeoff curves, indicating that pilot fraction can better regain performance through load balancing than tilt can. Choosing a new tilt configuration (Point T20) with same coverage as before shows an overall capacity decrease of 10% opposed to 7% when optimizing with respect to pilot, which confirms this notion. The absolute capacity of both T20, however, is still 2.5% higher than that of P10.

We conclude that antenna tilt can provide better performance gains in a limited tradeoff range where coverage constraints are not too striking and interference reduction can be achieved through uniform tilt enhancement of all cells. If coverage is the performance-driving factor, pilot fraction adjustment becomes superior since it cannot only shrink but also expand the size of a cell. For very inhomogeneous traffic distributions, load balancing will require an increasingly larger area expansion for some of the cells and make pilot fraction to the preferable tuning knob. Since power level cannot reduce overall interference as effectively as antenna tilt does, and it does not allow enhancing the cell size beyond the limits set by its maximum level, it is the tuning parameter least suited for load-balancing operations.

VII. CONCLUSION

We have shown that load balancing of 3G cellular networks needs to consider the tradeoff between the objectives of network coverage and capacity, particularly for data services where both impact the overall quality of service. It is important to include the CDMA-generic load dependence of coverage, which also has a tradeoff character and provides a reference for the optimization results. We have shown that a complete tradeoff analysis can provide a variety of configurations for load-balancing operations with different compromises between coverage and capacity. This allows selecting the best potential adjustment parameter for the particular application. We found that adjustment of tilt can be superior in a limited tradeoff range, while pilot adjustment is better when cells have to be expanded in area. Further investigation including reverse link and packet data servers are underway.

Table. 1: Parameter for Model Network

Length of network layout:	Normalized to one
Antenna height:	Defined through 1.5 deg down angle to network center.
Antenna vertical beam width:	6 deg, $\cos^2$-pattern, no sidelobes
Antenna backlobe suppression:	-25 dB
Antenna horizontal beam width:	Uniform over 90 deg.
Maximum EIRP per cell:	Set to 0dB received signal strength at network center.
Ext. Interference:	12dB above max EIRP reference.
Overhead channels:	Pilot only.
Pilot threshold:	-15dB
Pathloss slope:	40 dB/decade
DL pole capacity per cell:	100
Orthogonality factor:	0.5
Initial power level:	Max power
Initial pilot fraction:	10% of total EIRP.
Initial tilt:	1.5 deg.

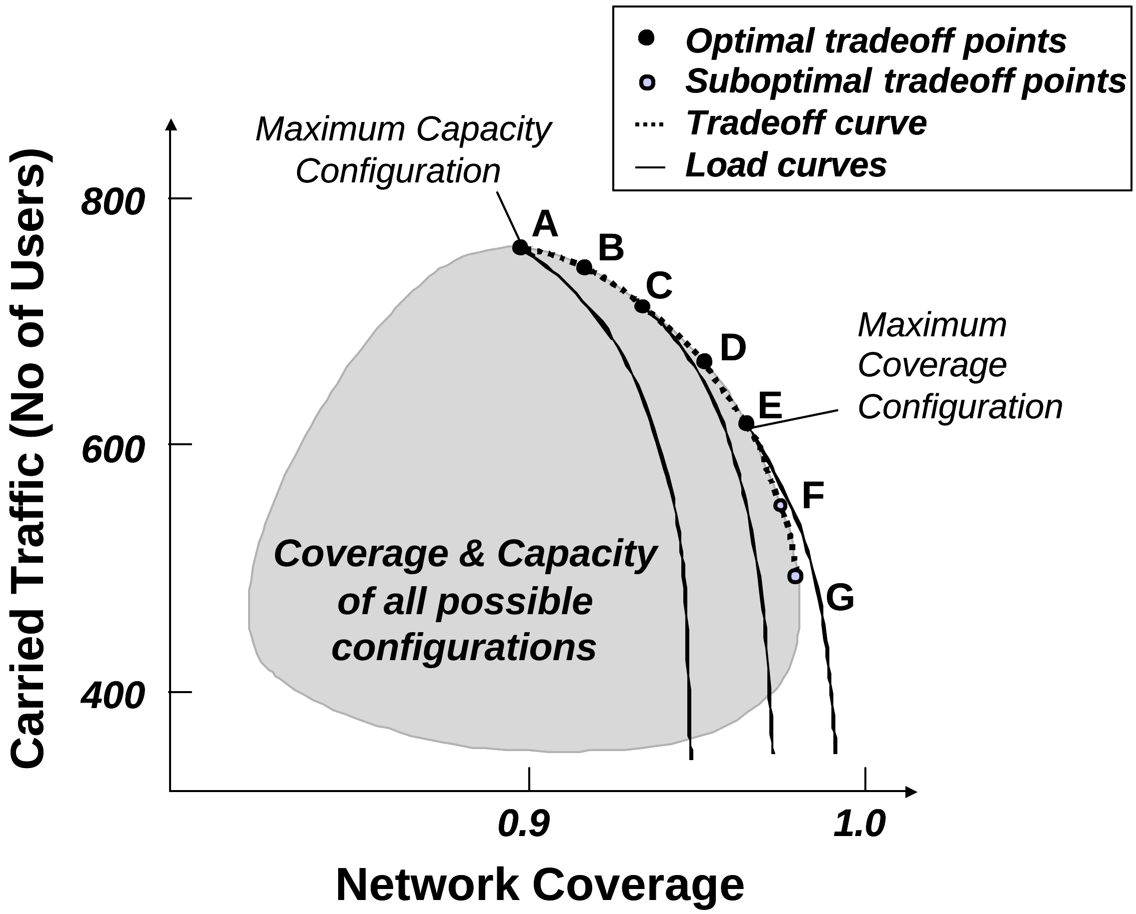

Fig. 1:

A network performance plot depicting network coverage and carried traffic for all network configurations (shaded area). The Tradeoff Curve (dashed line) represents the performance of all solutions of the optimization process. The back lines represent load curves for three tradeoff points.

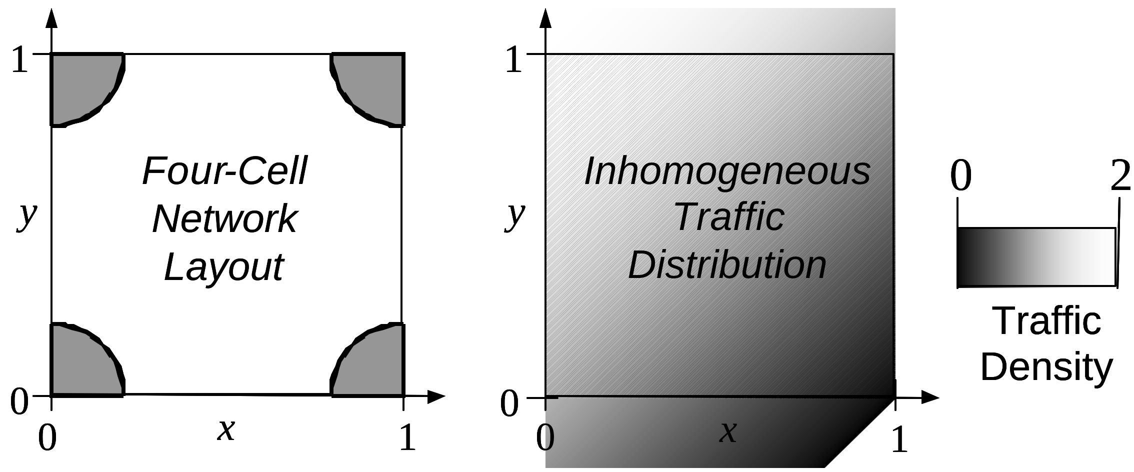

Fig. 2

A layout for a four-cell model network with an inhomogeneous traffic distribution. The traffic density increases diagonally across the network area.

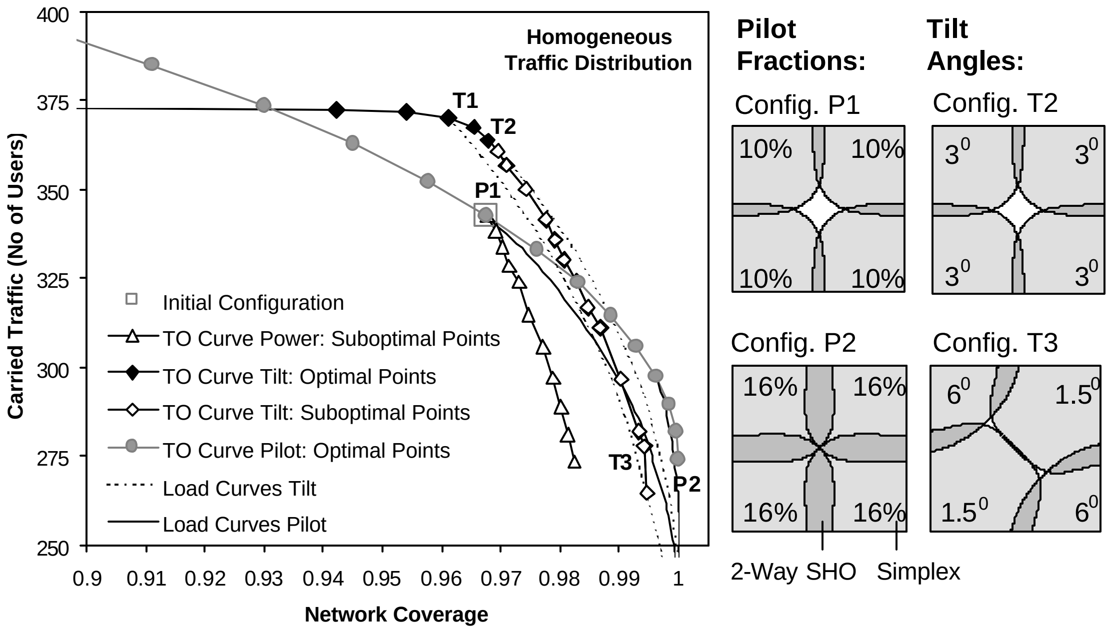

Fig. 3

Tradeoff curves for a homogeneous traffic distribution with respect to adjustment of antenna tilt (diamond), pilot fraction (dot) and power level (triangle). The open square shows the performance of the initial configuration. The full dots represent optimal- and the open dots sub-optimal solutions, the thin lines represent load curves. On the right-hand side, the soft-handoff areas and corresponding network configurations are shown for four tradeoff points P1, P2, T2 and T3.

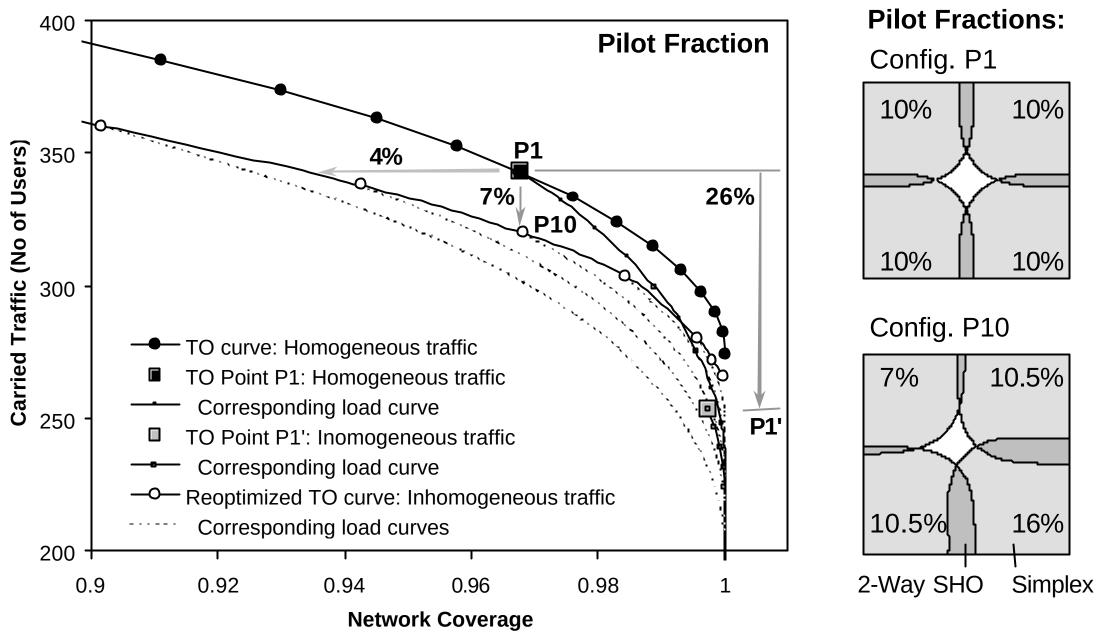

Fig. 4a

Load balancing via adjustment of the pilot fraction: The black dots show the tradeoff curve for homogeneous traffic, the open dots for inhomogeneous traffic. The capacity of configuration P1 deteriorates by 26% when the traffic changes from homogeneous to inhomogeneous (P1'). After load balancing, a tradeoff can be chosen with 7% less capacity or with 4% less coverage than before. The right side shows the soft-handoff areas of configurations P1 and P10.

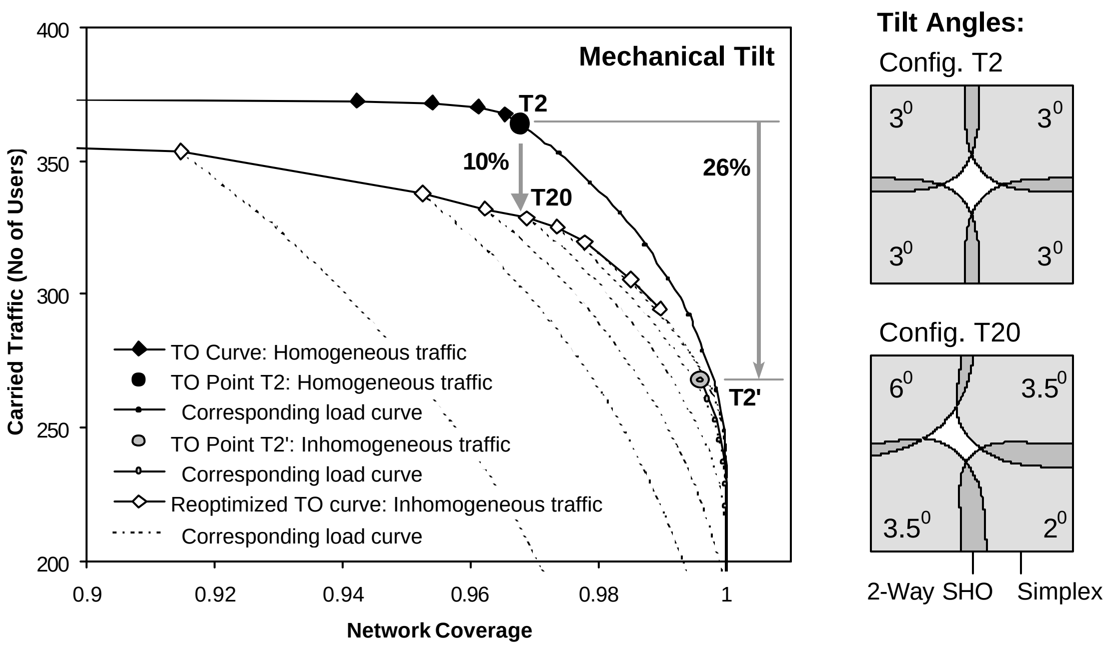

Fig. 4b:
Same plot as Fig. 4a but for optimization with respect to tilt. The capacity deterioration through change of traffic distribution is about the same as for optimization with respect to pilot fraction. The cell footprints and soft handoff areas for the reoptimized network, however, are reshaped.

REFERENCES:

[1] T. Ojanperä and R. Prasad, "WCDMA: Towards IP mobility and mobile internet", *Artech House Publishers*, 2001.

[2] H. Holma and A. Toskala, "WCDMA for UMTS: Radio access for third generation mobile communications", John Wiley & Sons, Ltd, 2000.

[3] S.T.S. Chia, "Design and optimization for cellular access network", *Electron.& Commun. Eng. J.*, vol. 8, no.6, pp.269-277, Dec. 1996.

[4] S. Hurley, "Planning effective cellular mobile radio networks", *IEEE Trans. Veh. Technol.*, vol.51, no.2, March 2002.

[5] 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 J. Select. Areas Commun.*, vol. 15, no. 7, pp. 1315-1326, Sept. 1997.

[6] C. Y. Lee and H. G. Kang, "Cell planning with capacity expansion in mobile communications: A tabu search approach", in *IEEE Trans. Veh. Technol.*, vol.49, no.5, pp. 1678-1691, Sept.2000.

[7] S. K. Das, S. K. Sen and R. Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment", *Wireless Networks*, vol.3, pp. 333-347, Aug. 1997.

[8] S. K. das, S. K. Sen and r. Jayaram, "A novel load balancing scheme for the tele-trffic hot spot problem in cellular networks", *Wireless Networks*, vol. 4, pp.325-340, July 98.

[9] C. Y. Lee, "A dynamic sectorization of microcells for balanced traffic in CDMA: Genetic algorithms approach", *IEEE Trans. Veh. Technol.*, vol.51, no.1, Jan 2002.

[10] K. Takeo and S. Sato, "Evaluation of a CDMA cell design algorithm considering non-uniformity of traffic and base station locations", IEICE Trans. Fundamentals, vol. E81, no.7, pp.1367-1377, July 1998.

[11] D. H. Kim, D. D. Lee, H. J. Kim, and K. C. Whang, "Capacity analysis of macro/mircocellular CDMA with power ratio control and tilted antenna", *IEEE Trans. Veh. Technol*, vol. 49, no. 1, Jan. 2000.

[12] R. G. Akl, M. V. Hegde, M. Naraghi-Pour and P. S. Min, "Multicell CDMA network design", IEEE Trans. Veh. Technol., vol. 50, no. 4, pp. 711-722, May 2001.

[13] J.-S. Wu, J.-K. Chung and C.-C. Wen, "Hot-spot traffic relief with a tilted antenna in CDMA cellular networks", *IEEE Trans. Veh. Technol*, vol. 47, no. 1, pp. 1-9, Feb. 1998.

[14] M. Frodigh, S. Parkvall, C. Roobol, P. Johansson, and P. Larsson, *"Future-Generation Wireless Networks"*, IEEE Pers. Commun., vol. 8, no. 5, pp. 10-17, Oct. 2001.