A GENERIC APPROACH TO CROSS-LAYER ASSISTED RESOURCE ALLOCATION

Benno Zerlin and Josef A. Nossek

Institute for Circuit Theory and Signal Processing
Munich University of Technology
{zerlin,nossek}@nws.ei.tum.de

ABSTRACT

We address the problem of resource minimization subject to a set of constraints on the quality of service. Based upon a set of mathematical properties of the mode dependent mapping from resources to quality of service, an iterative procedure is derived to determine the mode of operation and the resource allocation in the multi-user system. The algorithm is proven to be optimum with respect a chosen metric on the system resources. Furthermore, it can be shown that every step of the iterative scheme is of extremely low complexity. Numerical results visulalize the great importance of cross-layer mode optimization in an exemplarily chosen setting.

1. INTRODUCTION

Several contributions have been made to investigate cross-layer optimization as a generically formulated problem. These investigations do not focus on a specific system model but rather try to formulate general properties of different cross-layer approaches. A large portion of these works focuses on bottom up approaches offering a maximum of service quality to upper layers at the price of a fix amount of resources. Among these works are the publications [1, 2] where the throughput in a multi-user system is maximized by means of optimum power allocation. The considerations within base upon a very abstract system that allows to transfer the resulting techniques to a variety of communication systems. While the results in [1] mainly apply to the bottom sublayers of communication systems more recent developments aim at the bottom-up optimization for a wider span of system layers. To this end e.g. [3] introduces a framework that bases on utility functions, i.e. a generalized QoS expression that is implied to be monotonic with respect to the data rate on the physical link. A central requirement within is the monotonicity of this relation. On the background of an OFDM system with infinitesimal granularity in the frequency domain the authors derive solutions to the problem of dynamic subcarrier allocation and an adaptive power allocation scheme for QoS maximization. Moreover the results are employed to derive statements on the convexity of the feasibility regions of data rates and the global optimum of the posed cross-layer optimum. The authors in [5] moreover employ the concept of utility functions to formulate a pricing problem for the optimization of network resources. Dispensing with the formulation of a single abstract model of the complete regarded portion of the system stack modular approaches evolved that focus on the passing of information between different layers. In [6] the authors propose to base the layer descriptions on efficient sets of Pareto optimum points of operation. With a central monotonicity constraint on the employed layer description this approach leads to a multiobjective optimization problem as the core of the bottom-up cross-layer design.

On the other hand a series of generic publications have dealt with top-down formulations of resource allocation problems. The authors in [7, 8, 9] discuss the problem of QoS constraint power allocation in cellular networks with single antenna CDMA links. Their PHY layer contribution includes statements about the capacity regions and establish an uplink-downlink duality. In a more generic setting the framework in [10] proves the optimality of a power allocation scheme in an SINR constraint setting for a standardized class of interference functions. The latter is defined through its positivity, its monotonicity and its scalability. This approach has been generalized to multiple complex valued resource variables per user, i.e. *multiple-input single-output* (MISO) systems, in [11]. Moreover the authors in [11] prove that the achievable QoS region for the regarded class of interference functions is a convex set. Hence the cross-layer program can be solved through methods of conventional convex optimization.

This article introduces a generic approach to top-down cross-layer optimization of communications systems. It aims at delivering a certain amount of *quality of service* (QoS) $\boldsymbol{Q}^{(\mathrm{rq})}$ to the users, while simultaneously minimizing the required system resources. To this end let us define the three relevant classes of system parameters explicitly:

- The matrix $\boldsymbol{Q} \in \mathbb{R}^{N_\mathrm{Q} \times K}$ contains the N_Q QoS parameters of the K users that are present in the system. These QoS parameters are the only relevant interface to upper layers and sublayers or the application itself. The different service demands of the users are characterized by requirements $\boldsymbol{Q}^{(\mathrm{rq})}$ upon these QoS parameters.

- The resources of the K users are denoted by a vector $\boldsymbol{P} \in \mathbb{R}^K$. Hence only a single resource parameter per user is considered.[1]

- The mode of operation $\boldsymbol{\mathcal{M}}$ contains all optimization parameters that are not considered as resources.

Without loss of generality let the ordering of users in $\boldsymbol{P}$, $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{Q}$ coincide. With these definitions the cross-layer problem can be formulated as the minimization of resources. For a given norm[2] in $\boldsymbol{P}$ we find the optimum resource allocation and the optimum mode of operation as the solution to the following optimization:

$$\{\boldsymbol{P}^*, \boldsymbol{\mathcal{M}}^*\} = \underset{\{\boldsymbol{P},\boldsymbol{\mathcal{M}}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \boldsymbol{Q} \geq \boldsymbol{Q}^{(\mathrm{rq})}, \qquad (2)$$

where the constraining inequalities hold component-wise. The presented technique is proven to be optimum for a wide class of systems. To this end Section 2 introduces 3 Propositions, which guarantee the applicability of the upcoming considerations. Based upon these Propositions Section 3 and 4 formulate a converging iterative approach that yields the optimum solution up to arbitrary accuracy. Key to solving the problem in this wide generality is the transformation of the original program into a conditioned version that is solved in Section 3. This equivalent problem can be shown to be independent of the shortterm channel properties. Moreover it is proven to be decoupled among the users which allows for an offline computation of the optimum mode of operation.

2. PROPOSITIONS

Let the representation of the K user system be given in the form:

$$\boldsymbol{Q} = \boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{P}), \quad \boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}} : \mathbb{R}^K_{+,0} \mapsto \mathbb{R}^{N_\mathrm{Q} \times K} \qquad (3)$$

where the N_Q different QoS values for each of the K users in $\boldsymbol{Q} \in \mathbb{R}^{N_\mathrm{Q} \times K}$ are defined through a mode $\boldsymbol{\mathcal{M}}$ dependent mapping of the system resources $\boldsymbol{P} \in \mathbb{R}^K_{+,0}$. This representation exists for all systems and for all choices of QoS and resources respectively. Yet not all of these representations can be given in closed or even in invertible explicit form. Preparing the derivation of a generic approach for cross-layer optimization of a wide class of communication systems some preconditions on $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$ will be made in the following paragraphs. To this end let us introduce the notation of *longterm* and *shortterm* parameters. It refers to the fading process of the physical channel and classifies all parameters that depend upon the instantaneous realization of this process as shortterm parameters. Characteristics of their probability density function and variables that are independent of the channel fading process are considered longterm parameters. On this background let the following propositions hold:

Proposition 1 *The mode of operation $\boldsymbol{\mathcal{M}}$ as well as all components of the QoS matrix $\boldsymbol{Q}$ are longterm parameters.*

Proposition 2 *The function $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$ is decomposable into three components as follows:*

$$\boldsymbol{\Upsilon}^{(1)} : \boldsymbol{P} \mapsto \tilde{\boldsymbol{P}} = \boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P}), \quad \tilde{\boldsymbol{P}}, \boldsymbol{P} \in \mathbb{R}^K_{0,+} \qquad (4)$$

$$\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}} : \tilde{\boldsymbol{P}} \mapsto \tilde{\boldsymbol{Q}} = \boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}(\tilde{\boldsymbol{P}}), \quad \tilde{\boldsymbol{Q}} \in \mathbb{R}^{N_\mathrm{Q} \times K} \qquad (5)$$

$$\boldsymbol{\Upsilon}^{(3)} : \tilde{\boldsymbol{Q}} \mapsto \boldsymbol{Q} = \boldsymbol{\Upsilon}^{(3)}(\tilde{\boldsymbol{Q}}, \boldsymbol{\pi}_{out}), \quad \boldsymbol{Q} \in \mathbb{R}^{N_\mathrm{Q} \times K}. \qquad (6)$$

Within $\tilde{\boldsymbol{P}} = \boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})$ gives a longterm description of the fading multiple access channel. The outage probability $\boldsymbol{\pi}_{out} \in [0;1]^K$ is defined through the shortterm pendant $\tilde{\boldsymbol{P}}^{(st)}$ of $\tilde{\boldsymbol{P}}$ as:

$$\boldsymbol{\pi}_{out} = Pr\left(\tilde{\boldsymbol{P}}^{(st)} < \tilde{\boldsymbol{P}}\right), \qquad (7)$$

where the inequality holds component wise.

Proposition 3 *The functions $\boldsymbol{\Upsilon}^{(1)}$, $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ and $\boldsymbol{\Upsilon}^{(3)}$ fulfill the following properties:*

1. *The function $\boldsymbol{\Upsilon}^{(1)}$ is independent of the mode of operation and is a monotonic function. It is strictly monotonically increasing on its diagonal[3] and is monotonically decreasing in all off-diagonal elements.*

2. *The function $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ is diagonal and independent of shortterm system parameters.*

3. *The function $\boldsymbol{\Upsilon}^{(3)}$ is independent of shortterm system parameters. Moreover conditioned on $\boldsymbol{\pi}_{out}$ a solution for its unique inversion $\boldsymbol{\Upsilon}^{(3),-1}$ exists with*

$$\boldsymbol{Q} = \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(3),-1}(\boldsymbol{Q}, \boldsymbol{\pi}_{out}), \boldsymbol{\pi}_{out}\right), \qquad (8)$$

Through the decomposability in Proposition 2 the parameters $\tilde{\boldsymbol{P}}$ and $\boldsymbol{\pi}_\mathrm{out}$ form a longterm description of the shortterm fading processes on the physical channel that is valid as long

[1] The extension to multiple resource parameters per user or the extension to integer resource metrics does in general not contradict the results of this paper. Yet the solution of the in this case matrix valued problem:

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \tilde{\boldsymbol{P}} \geq \tilde{\boldsymbol{P}}^{(\mathrm{rq})}, \qquad (1)$$

which occurs to be part of the resulting cross-layer optimization algorithm can not be implied for general choices of $\boldsymbol{P} \notin \mathbb{R}^K_{+,0}$. Hence the extension to multiple resources per user is left to the specific environment where additional constraints might apply to provide a solution to (1).

[2] In fact any order relation can be used to define this problem. For the remainder of this paper though we exemplarily employ the quasi order relations of different matrix norms. These norms are reflexive and transitive, i.e. a *quasi-relation*, and hence qualify in this context.

[3] Diagonality in this context refers to the user indices.

as the fading processes can be assumed stationary. For an information theoretic backup of this approach we refer to [13]. The outage probability therefore is defined as the probability that the shortterm representation $\tilde{\boldsymbol{P}}^{(\mathrm{st})}$ of $\tilde{\boldsymbol{P}}$ is smaller than $\tilde{\boldsymbol{P}}$.[4] The function $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ gives a description of the remaining system in non-outage cases, while $\boldsymbol{\Upsilon}^{(3)}$ determines how these outage events affect the specified QoS. Later reasonings will suggest the terminology *equivalent resources* for $\tilde{\boldsymbol{P}}$ and *equivalent QoS* for $\tilde{\boldsymbol{Q}}$ respectively. Fig. 1 visualizes the decomposition of $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$. With the definition of the three rep-

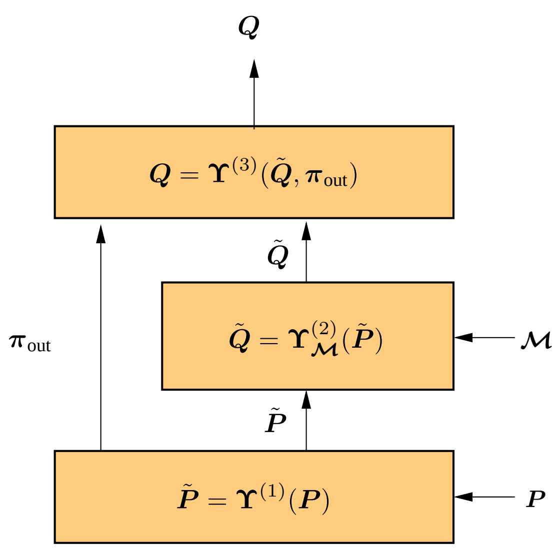

Fig. 1. Schematic representation of the system model

resentatives $\boldsymbol{\Upsilon}^{(1)}$, $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ and $\boldsymbol{\Upsilon}^{(3)}$ the decomposable function $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$ can be expressed in a concatenated form as:

$$\boldsymbol{Q} = \boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{P}) = \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}\left(\boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})\right), \boldsymbol{\pi}_{\text{out}}\right). \quad (9)$$

The components of this composition are subject to preconditions as presented in Proposition 3. According to the first clause within, the equivalent resource $\tilde{\boldsymbol{P}}$ of a user is strictly monotonically increasing with this user's resource while it is monotonically decreasing with the resources of all other users.[5] Through the second and third clause $\boldsymbol{\Upsilon}^{(1)}$ moreover is required to carry the complete regarded influence of shortterm system parameters. The required diagonal nature of $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ translates into user wise decoupled system cores that carry the complete dependence on the mode of operation. Hence the equivalent requirements of a user k in the vector $\tilde{\boldsymbol{Q}}^{(\mathrm{rq})}_k$ must be completely determined by the equivalent scalar resource $\tilde{P}_k$ of the same user k. This condition is somewhat canonical as parameters that influence the performance of all users are typically classified as resources rather than as mode parameters, e.g. the transmit power in multiple access schemes. Typically the function $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ will span the widest part of the cross-layer problem as the corresponding Clause 2 in Proposition 3 does not make any restrictions on its diagonal elements. Clearly $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ does not even have to be a continuous function neither does it need to be available in closed form. This fact comprises the central strength of the presented approach, because it can solve problems that are typically not accessible for conventional optimization techniques and their application in cross-layer design. The third representative $\boldsymbol{\Upsilon}^{(3)}$ includes the influence of the outage probability $\boldsymbol{\pi}_{\text{out}}$ and additionally allows for an invertible non-diagonal extension of the concatenation $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}\left(\boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})\right)$. This feature allows to include upper layer resource allocation schemes the parameters of which are not subject to the regarded cross-layer optimization themselves. Hence $\boldsymbol{\Upsilon}^{(3)}$ does not depend on the mode of operation $\boldsymbol{\mathcal{M}}$ and must be given in a description that is independent of shortterm system parameters. The extension $\boldsymbol{\Upsilon}^{(3)}$ is more important for the validity of the regarded QoS expressions than it is relevant to the optimization process with respect to $\boldsymbol{\mathcal{M}}$.

3. CONDITIONED OPTIMIZATION

The central problem of this work is to derive cross-layer resource allocation schemes for the above defined system class through the solution of the following optimization problem:

$$\{\boldsymbol{P}^*, \boldsymbol{\mathcal{M}}^*\} = \underset{\{\boldsymbol{P}, \boldsymbol{\mathcal{M}}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad (10)$$
$$\text{s.t.:} \quad \boldsymbol{Q} \geq \boldsymbol{Q}^{(\mathrm{rq})}, \quad \text{with} \quad \boldsymbol{Q} = \boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{P}).$$

To this end let us first derive the solution of this problem conditioned on the outage probability $\boldsymbol{\pi}_{\text{out}}$. Conditioning the solution to an a priori known value for $\boldsymbol{\pi}_{\text{out}}$ allows for the decomposition of the optimization task and thus shall be of a certain relevance to the optimum procedure in Section 4. Hence, the following paragraphs consider the optimization:

$$\{\boldsymbol{P}^*, \boldsymbol{\mathcal{M}}^*\} = \underset{\{\boldsymbol{P}, \boldsymbol{\mathcal{M}}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad (11)$$
$$\text{s.t.:} \quad \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}\left(\boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})\right), \boldsymbol{\pi}_{\text{out}}\right)\Big|_{\boldsymbol{\pi}_{\text{out}}} \geq \boldsymbol{Q}^{(\mathrm{rq})}.$$

As this program as well as the original task (10) is not accessible to conventional optimization techniques because the derivatives with respect to the possibly integer valued mode variable $\boldsymbol{\mathcal{M}}$ is not defined and $\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}$ additionally does not necessarily allow for a closed solution of the resulting Karush-Kuhn-Tucker conditions, let us regard the following theorem:

[4] This definition applies no matter if the strategy to adapt $\boldsymbol{P}$ operates on a longterm or a shortterm scale, i.e. no matter whether $\boldsymbol{P}$ itself is a longterm or shortterm parameter.

[5] The demanded property is inherent to a wide variety of multiple access schemes. It can as well be found in axiomatic approaches like [10, 11].

Theorem 1 *Let $\boldsymbol{\pi}_{out}$ be the outage probability that through $\boldsymbol{\Upsilon}^{(1)}$ corresponds to the equivalent resources $\tilde{\boldsymbol{P}}^{(rq)}$. Furthermore let the equivalent requirements $\tilde{\boldsymbol{Q}}^{(rq)}$ be defined through the inversion of $\boldsymbol{\Upsilon}^{(3)}$ conditioned on $\boldsymbol{\pi}_{out}$. Then the solution to the problem:*

$$\left\{\mathcal{M}_k^*, \tilde{P}_k^{(rq)}\right\} = \underset{\{\mathcal{M}_k, \tilde{P}_k\}}{\operatorname{argmin}} \ \tilde{P}_k \qquad \text{s.t.:} \quad \tilde{\boldsymbol{Q}}_k \geq \tilde{\boldsymbol{Q}}_k^{(rq)}, \quad (12)$$

is independent of the requirements[6] $\tilde{\boldsymbol{Q}}_\ell^{(rq)}$, $\forall \ell \neq k$ *and is independent of all shortterm system parameters. The union of the solutions $\mathcal{M}_k^*$ forms an optimizer $\boldsymbol{\mathcal{M}}^*$ to the cross-layer optimization program in* (11)*.*

The first two clauses of this Theorem are directly proven by Proposition 3. Due to the diagonality of $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}^{(2)}$ the constraints are diagonal too and hence the solution for the user k does not depend upon other users' equivalent QoS requirements. The same applies to the independence of the fading parameters. Key to the proof of the last clause of Theorem 1 is the monotonicity of the equivalent resources $\tilde{\boldsymbol{P}}$ with respect to all resources $\boldsymbol{P}$. This monotonicity has been part of Proposition 3 above. As the resources only through $\tilde{\boldsymbol{P}}$ and through $\boldsymbol{\pi}_{\text{out}}$ influence the QoS and furthermore the problem is conditioned on the outage probability this monotonicity suffices to show that a minimization of the resources $\boldsymbol{P}$ will inherently result in a minimization of all equivalent resources $\tilde{\boldsymbol{P}}$. Therefore the objective and the optimization with respect to $\boldsymbol{P}$ in (11) can be replaced by $\tilde{\boldsymbol{P}}$ without violating the validity of the solution $\mathcal{M}_k^*$. Employing the decomposability of $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$ and the invertibility of $\boldsymbol{\Upsilon}^{(3)}$ the constraints of (11) can equivalently be expressed through a set of requirements $\tilde{\boldsymbol{Q}}^{(\text{rq})}$ on the equivalent QoS $\tilde{\boldsymbol{Q}}$. This concludes the proof of the optimal nature of $\boldsymbol{\mathcal{M}}^*$ for (11).

The decoupled and equivalent formulation of the original program through the problems in (12) provides little advantage in terms of Lagrangian optimization. Still components of the optimization parameter $\mathcal{M}_k$ are discrete, the corresponding derivatives do not exist and the Karush-Kuhn-Tucker conditions can not be applied. But the achieved decoupling among users and the gained independence of the problem from the instantaneous channel realization render the accessibility to Lagrangian methods unnecessary:

Corollary 1 *The equivalent requirements $\tilde{\boldsymbol{Q}}_k^{(rq)}$ uniquely determine the solution to the optimization problems in* (12)*. In particular the program is independent of the state of operation and therefore the solution for any equivalent QoS requirements can be obtained offline and prior to operation.*

Explicitly the optimum mode of operation $\mathcal{M}_k^*$ can be precomputed for a sufficiently dense grid in $\tilde{\boldsymbol{Q}}_k^{(\text{rq})}$ offline by sampling the $N_{\text{Q}} \times K$ dimensional range of $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}^{(2)}$. The equivalent QoS $\tilde{\boldsymbol{Q}}$ to this end is computed for an arbitrarily large but finite number of system modes and for a suitable number of values for $\tilde{\boldsymbol{P}}_k$. Storing these offline computed solutions in an N_{Q} dimensional database allows for the offline determination of the optimum mode of operation for grid of equivalent requirements. Each grid point defines a feasibility region through its equivalent requirements. Searching this feasibility region for the mode that provides $\tilde{\boldsymbol{Q}}_k^{(\text{rq})}$ with a minimum $\tilde{P}_k^{(\text{rq})}$ is chosen as the optimum mode $\mathcal{M}_k^*$. During operation the solution to (12) thus can be obtained through a single table lookup. The problem (12) therefore can be solved very efficiently at an absolute minimum of computation cost. This makes the proposed procedure easily accessible for real-time implementations.

[6] $\tilde{\boldsymbol{Q}}_k$, $\mathcal{M}_k$ and $\tilde{P}_k$ denote the kth user's portion of $\tilde{\boldsymbol{Q}}$, $\boldsymbol{\mathcal{M}}$ and $\tilde{\boldsymbol{P}}$ respectively.

4. GENERAL CROSS-LAYER OPTIMIZATION

With the above Section a low complexity solution to the conditioned problem setting (12) is available. The upcoming considerations in this Section now focus on optimizing the mode of operation and the resource allocation among all K users employing the results form above. We target the problem of finding the mode of operation $\boldsymbol{\mathcal{M}}^*$ that fulfills a set of QoS requirements $\boldsymbol{Q}^{(\text{rq})}$ with a minimum amount of resources, cf. (10):

$$\{\boldsymbol{P}^*, \boldsymbol{\mathcal{M}}^*\} = \underset{\{\boldsymbol{P}, \boldsymbol{\mathcal{M}}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \qquad (13)$$
$$\text{s.t.:} \quad \boldsymbol{Q} \geq \boldsymbol{Q}^{(\text{rq})}, \quad \text{with} \quad \boldsymbol{Q} = \boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{P}),$$

where the function $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$ belongs to the above defined system class and fulfills the properties in Propositions 1-3 introduced in Section 2. Aiming at a generic solution that applies to all representatives of the above defined class the solution of (13) through the calculus of Lagrangian multipliers and the Kuhn-Tucker theorem renders impossible. Moreover neither the function $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}^{(2)}(\tilde{\boldsymbol{P}})$ nor the mode of operation $\boldsymbol{\mathcal{M}}$ necessarily is accessible to these methods.[7]

4.1. Iterative solution

We propose an iterative framework to solve (13) with arbitrary accuracy. An overview of this approach is given in Fig. 2. The approach is based on an iterative adaptation of the outage probability $\boldsymbol{\pi}_{\text{out}}$ which in the further context will be indexed by the iteration number i as $\boldsymbol{\pi}_{\text{out}}[i]$. The solutions within each iteration hence can base upon an assumption $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ on the outage probability that was obtained during the last iteration. The problem thus reduces to the condi-

[7] The function $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}^{(2)}(\tilde{\boldsymbol{P}})$ in non-trivial settings usually is too complex to provide an invertible system of equations from the Karush-Kuhn-Tucker conditions, whereas the mode of operation in many relevant applications contains integer variables. Hence the derivatives involved in the conventional solution of this problem are not defined.

tioned optimization investigated in Section 3. Given the outage probability $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ that corresponds to the optimum vector of equivalent resources $\tilde{\boldsymbol{P}}^*$ the program reads, cf. (11):

$$\{\boldsymbol{P}^*, \boldsymbol{\mathcal{M}}^*\} = \underset{\{\boldsymbol{P},\boldsymbol{\mathcal{M}}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \tag{14}$$
$$\text{s.t.:} \quad \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(2)}_{\boldsymbol{\mathcal{M}}}\left(\boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})\right), \hat{\boldsymbol{\pi}}_{\text{out}}[i]\right) \geq \boldsymbol{Q}^{(\text{rq})}.$$

As shown in Section 3 the propositions made can be used to decompose this optimization into three separate subproblems:

$$\tilde{\boldsymbol{Q}}^{(\text{rq})} = \boldsymbol{\Upsilon}^{(3),-1}(\boldsymbol{Q}^{(\text{rq})}, \hat{\boldsymbol{\pi}}_{\text{out}}[i]), \tag{15}$$

$$\left\{\mathcal{M}_k^*, \tilde{P}_k^{(\text{rq})}\right\} = \underset{\{\mathcal{M}_k,\tilde{P}_k\}}{\operatorname{argmin}} \tilde{P}_k \quad \text{s.t.:} \quad \tilde{\boldsymbol{Q}}_k \geq \tilde{\boldsymbol{Q}}_k^{(\text{rq})}, \forall k \tag{16}$$

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \tilde{\boldsymbol{P}} \geq \tilde{\boldsymbol{P}}^{(\text{rq})}. \tag{17}$$

For convenient reading we dropped the index $[i]$ in the notation of $\boldsymbol{P}^*$, $\tilde{\boldsymbol{P}}^{(\text{rq})}$ and $\tilde{\boldsymbol{Q}}^{(\text{rq})}$. While an efficient solution for the inverse of $\boldsymbol{\Upsilon}^{(3)}$ was a precondition on the regarded system class the solution to the problems (16) has been derived in Section 3. Moreover the solution to (17) for the vast majority of multiple access schemes is known or can be obtained through the monotonicity of $\boldsymbol{\Upsilon}^{(1)}$ as it was introduced in Section 2. The solution of the cross-layer optimization conditioned on $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ thus is known and can be obtained at very low computational cost. With the optimum pair of modes and resources $\{\boldsymbol{P}^*, \boldsymbol{\mathcal{M}}^*\}$ the resulting outage probability $\boldsymbol{\pi}_{\text{out}}[i]$ of the system can be determined through its definition in (7). Unless $\boldsymbol{\pi}_{\text{out}}[i] = \hat{\boldsymbol{\pi}}_{\text{out}}[i]$ the solutions obtained from the conditioned problem (16) is not a valid solution for the original task (13). The made assumption on $\boldsymbol{\pi}_{\text{out}}[i]$ has to be adapted and a new conditioned problem has to be solved. To this end we propose the following update rule:

$$\hat{\boldsymbol{\pi}}_{\text{out}}[i+1] = \boldsymbol{\pi}_{\text{out}}[i]. \tag{18}$$

The resulting structure of the iterative scheme is sketched in Fig. 2.

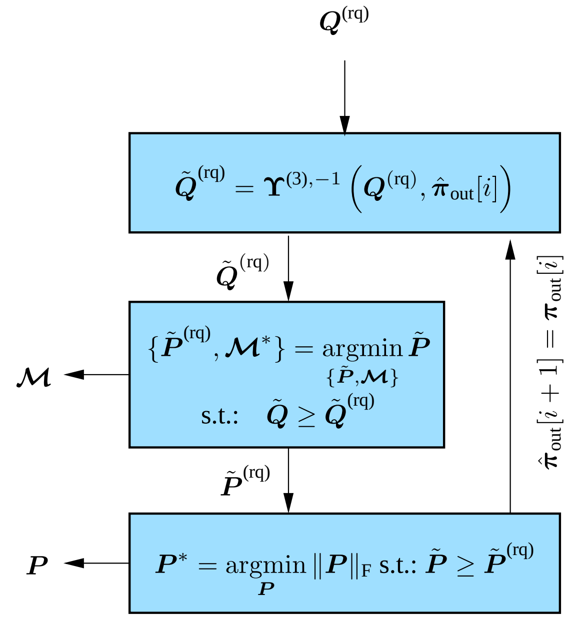

Fig. 2. Schematic representation of the iterative cross-layer optimization

4.2. Convergence

Through the solution of an $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ conditioned version of the original problem the cross-layer optimization can be solved through an iterative scheme as sketched in Fig. 2. Yet the obtained iteration is of no use if its convergence can not be proven. To this end let us state the following Theorem:

Theorem 2 *Let $\boldsymbol{\pi}_{out}[0] = \mathbf{0}$ be the initialization for the iterative scheme. Furthermore let $\boldsymbol{\pi}_{out}[i]$ be defined as the outage probability that results from $\boldsymbol{\mathcal{M}}^*$ and $\boldsymbol{P}^*$ as defined in* (14)*. Then the iteration $\hat{\boldsymbol{\pi}}_{out}[i] = \boldsymbol{\pi}_{out}[i-1]$ converges and $\boldsymbol{\pi}_{out}[i] = \boldsymbol{\pi}_{out}[i-1]$ holds with arbitrary accuracy for large i.*

To prove this convergence assume an arbitrary $\hat{\boldsymbol{\pi}}_{\text{out}}[i] = \boldsymbol{\pi}_{\text{out}}[i-1] > \hat{\boldsymbol{\pi}}_{\text{out}}[i-1]$ from the interval $[0;1]^K$. By the definition of the outage probability the increase in $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ necessarily results in an increase in all components of $\tilde{\boldsymbol{Q}}^{(\text{rq})}$, i.e. larger equivalent requirements with respect to the iteration $i-1$. This is due to the fact that an increased outage probability will enlarge the QoS decrease in $\boldsymbol{\Upsilon}^{(3)}$. Due to the inequality nature of the constraints in (16) this monotonicity with respect to $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ applies to the requirements on the equivalent resources $\tilde{\boldsymbol{P}}^{(\text{rq})}$ as well. To this end assume that an increase in $\tilde{\boldsymbol{Q}}_k^{(\text{rq})}$ from iteration $i-1$ to i yields a decrease of $\tilde{P}_k^{(\text{rq})}$ in the optimum solution of (16). Then this new value for $\tilde{P}_k^{(\text{rq})}$ would provide a smaller solution to the problem in iteration $i-1$. Thus the solution in iteration i can not be optimum, which is a contradiction. Hence an increase in outage probability $\hat{\boldsymbol{\pi}}_{\text{out}}[i] > \hat{\boldsymbol{\pi}}_{\text{out}}[i-1]$ always results in requirements $\tilde{\boldsymbol{P}}^{(\text{rq})}$ that are equal to or larger than the corresponding values in the previous iteration. Due to the positive semidefinite nature of the equivalent resources and the definition of the outage probability this inherently results in $\boldsymbol{\pi}_{\text{out}}[i] \geq \boldsymbol{\pi}_{\text{out}}[i-1]$. Through the update $\hat{\boldsymbol{\pi}}_{\text{out}}[i+1] = \boldsymbol{\pi}_{\text{out}}[i]$ a single increase in $\hat{\boldsymbol{\pi}}_{\text{out}}$ causes a monotonically increasing series $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$. Because $\boldsymbol{\pi}_{\text{out}}[i]$ is positive semidefinite, the initial choice $\hat{\boldsymbol{\pi}}_{\text{out}}[0] = 0$ results in $\hat{\boldsymbol{\pi}}_{\text{out}}[1] \geq \hat{\boldsymbol{\pi}}_{\text{out}}[0]$. In the case of equality $i = 1$ directly fulfills the condition for convergence $\boldsymbol{\pi}_{\text{out}}[i] = \boldsymbol{\pi}_{\text{out}}[i-1]$. In all other cases, the above derivation proves $\boldsymbol{\pi}_{\text{out}}[i]$ to be an monotonically increasing sequence. Hence the proof of convergence is obtained from the bounded nature of the probability integral, i.e. $\boldsymbol{\pi}_{\text{out}}[i] \in [0;1]^K$.

With the conclusion of this proof the iterative procedure of Section 4 is known to converge for all problem settings and systems that fall in the defined class. Yet this convergence does not prove the feasibility of the posed optimization task. As possible feasibility constraints on the system resources can always be included in the definition of the corresponding outage probability we propose to define the feasibility through the optimum mode of operation $\boldsymbol{\mathcal{M}}^*$. Hence a set of QoS requirements is considered feasible if for every iteration i a mode of operation $\boldsymbol{\mathcal{M}}^* = \bigcup_{k=1}^{K} \mathcal{M}_k^*$ exists among the finite number of system configurations such that:

$$\left\{\mathcal{M}_k^*, \tilde{P}_k^{(\text{rq})}\right\} = \underset{\{\mathcal{M}_k, \tilde{P}_k\}}{\operatorname{argmin}}\ \tilde{P}_k \quad \text{s.t.:} \quad \tilde{\boldsymbol{Q}}_k \geq \tilde{\boldsymbol{Q}}_k^{(\text{rq})}, \forall k. \quad (19)$$

For all feasible constellations though, the above results provide the means to solve the cross-layer problem (10) through the iterative consideration of K decoupled longterm problems. Based upon the results from the previous iteration, these problems are all conditioned on the outage probability. They have been subject to consideration in Section 3 where the optimum was proven to be independent of the state of operation and the solution therefore was obtained through a look-up in an offline generated table. Hence the presented iterative scheme provides the means to efficiently solve the top-down cross-layer optimization program (2) for all systems that fulfill the Propositions 1-3.

5. EXEMPLARY APPLICATION TO A SMART ANTENNA SYSTEM

To support the above elaborations and to visualize their applicability to HSDPA like MISO systems[8] this section performs numerical simulations of the proposed techniques and compares the resulting performance with relevant state-of-the-art approaches. To this end every simulation analyzes a total of 1000 longterm settings, each consisting of 5000 time slots. Each longterm environment faces independent user positions assuming an uniform distribution of users in the cell. The spatial channel model assumes 8 antenna elements at the transmitter and a $2°$ Laplacian angular spread. The coefficients in the corresponding Karhunen-Loewe representation are assumed to have circulary symmetric Gaussian distributions. The variances of these random variables are computed using the Hata pathloss model [16, 17, 18]. For further details of the used spatio-temporal channel model we refer to [19]. With these randomly generated vector valued channel coefficients an industrially employed FEC turbo code is used within an *hybrid automatic repeat request* (HARQ) protocol. The latter is used in its Chase combining (type I HARQ) form. We assumed a receiver noise level of -105 dBm and a maximum transmit power of 16 W, which together with a fix antenna gain of 18 dBi, results in an *effective isotropic radiated power* (EIRP) of 60 dBm.

5.1. QoS compliance

As the QoS compliant service of all users is a precondition to any performance enhancement in terms of transmit power savings or capacity increase in the upcoming sections, Fig. 3 verifies this compliance. Serving three users with QoS demands on throughput/delay of $[200$ kbps / 100 ms$]$, $[400$ kbps / 100 ms$]$ and $[600$ Kbps / 100 ms$]$ respectively, the simulation of 1000 user settings, i.e. location of the users in the cell, resulted in 1000 values for the throughput obtained as the ratio of the sum of correctly received information bits and the total transmit time. The different lines in the plot represent the corresponding histograms. The resulting performance was obtained using the proposed cross-layer resource allocation and solving the occurring downlink processing problem.[9] It can be seen from the results in Fig. 3 that the derived cross-layer scheme provides the means to assure the request QoS in the given MISO system. As the optimization can only access a finite number of modes of operation, the QoS all users are slightly oversatisfied. Let us remark, that the scheme precisely achieves the target throughput independent of the user positions, i.e. independent of the longterm channel properties. Yet different longterm scenarios will result in different transmit powers necessary to provide the QoS

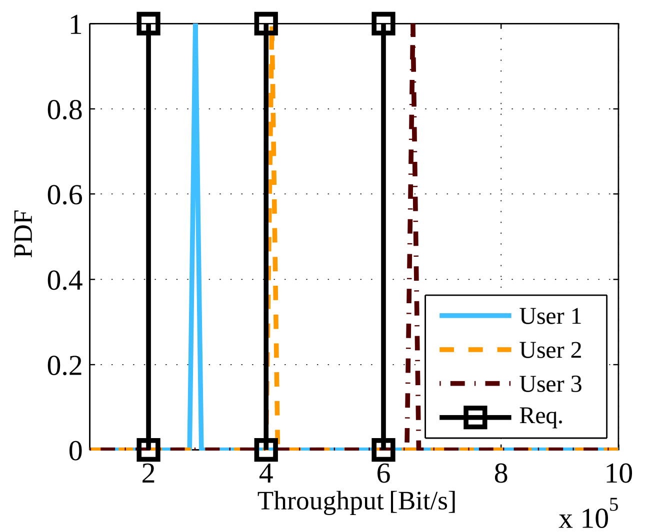

Fig. 3. Demonstration of the achieved QoS compliance with optimum scheme

[8] For details of the resulting form of $\boldsymbol{\Upsilon}_{\boldsymbol{\mathcal{M}}}$ and its decomposition we refer to works like [14] and its MISO extension in [15].

[9] Choosing the SINR as equivalent resource, this leads to a downlink processing problem investigated in e.g. [20].

compliance. These power savings are subject to the evaluations in Subsection 5.2.

5.2. Power advantage

In terms of the optimization (1) tackled within this paper the above analyzed QoS compliance shows that the posed constraints can be met with the scheme. Yet it remains to verify how large the power advantages are that can be obtained through the performed resource minimization. Hence this Section elaborates on the distribution of total transmit power that is necessary to provide the above QoS results in the described environment. As HSDPA does not yet command a cross-layer QoS management the only usefull reference is given by the use of the full available EIRP. Fig. 4 shows the cumulative distribution of the transmit power that is necessary in the 1000 different longterm settings. The plotted his-

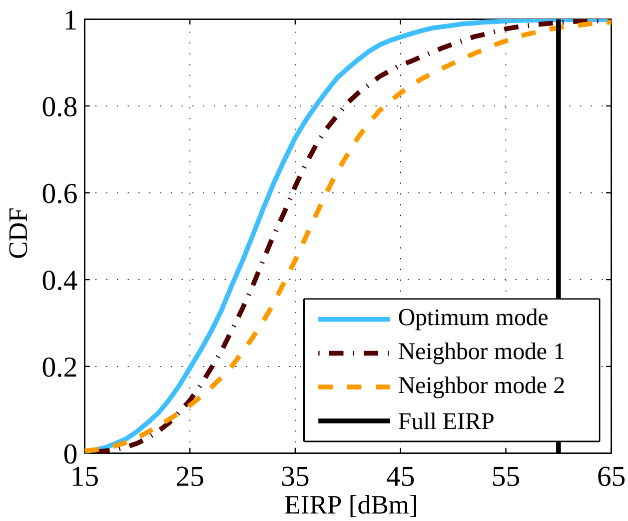

Fig. 4. Comparison of the necessary total transmit power for different modes

tograms reveal that a vast majority of longterm settings can be met with significantly less transmit power than the available 60 dBm. Precisely speaking the optimum scheme can fulfill the posed QoS constraints with less than 31 dBm in 50% of all cases. As this comparison with the maximum transmit power is a somehow coarse reference Fig. 4 also includes the histogram of transmit powers that are necessary to satisfy the QoS requirements with suboptimum mode choices. To this end higher modes neighboring the optimum choice were used to solve the downlink beamforming problem. These suboptimum modes inherently will pose higher SINR requirements in every iteration of the procedure. This will entail an increase of the corresponding outage probability. Through this iterative connection the loss in performance can significantly exceed the mere SINR gap known from conventional throughput considerations. Overall the distribution of transmit powers shows the high sensitivity of the MISO system with respect to mode mismatches.

6. CONCLUSION

The presented approach allows for the determination of the optimum resource allocation and the joint optimization of parameters in the lower layers of a communication system. Decomposing the functional description and conditioning the core optimization on the channel outage probability motivated an iterative procedure to solve the cross-layer problem. Each step can be shown to be of extremely low numerical complexity and the iterative algorithm is proven to converge. Due to its generic formulation in terms of a set of proposition the technique not only applies to one specific system set up, but is valid for an extremely large class of systems. This include the prominent HSDPA extension to UMTS to which the technique was applied yielding significant enhancements in power savings. The latter can be used for increasing the servable system load.

7. REFERENCES

[1] D. N. C. Tse, "Optimal power allocation over parallel Gaussian channels," in *Proc. Int. Symp. Information Theory, Ulm, Germany*, June 1997.

[2] D. N. C. Tse, *Transmitter directed, multiple receiver system using path diversity to equitably maximize throughput*, May 24 1999, patent filed.

[3] Guocong Song and Ye Geoffrey Li, "Cross-layer optimization for OFDM wireless networks - part I: theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, March 2005.

[4] Guocong Song and Ye Geoffrey Li, "Cross-layer optimization for OFDM wireless networks - part II: algorithm development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, March 2005.

[5] J. K. MacKie-Mason and H. R. Varian, "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1141–1149, September 1995.

[6] Johannes Brehmer and Wolfgang Utschick, "Physical layer characterization of MIMO systems by means of multiobjective optimization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005, Philadelphia, United States.

[7] Daniel Catrein, L. Imhof, and Rudolf Mathar, "Power Control, Capacity, and Duality of Up- and Downlink in

Cellular CDMA Systems," *IEEE Transactions on Communications*, vol. 52, no. 10, pp. 1777–1785, 2004.

[8] L. Imhof and Rudolf Mathar, "Capacity Regions and Optimal Power Allocation for CDMA Cellular Radio," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2011–2019, June 2005.

[9] L. Imhof and Rudolf Mathar, "The Geometry of the Capacity Region for CDMA Systems with General Power Constraints," *IEEE Transactions on Wireless Communications*, , no. 5, pp. 2040–2044, September 2005.

[10] Roy D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.

[11] Holger Boche, Martin Schubert, Slawomir Stanczak, and Marcin Wiczanowski, "An axiomatic approach to resource allocation and interference balancing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[12] Michel T. Ivrlac, *MIMO Systems - Models, Measures and Algorithms*, Ph.D. thesis, Munich University of Technology, 2005.

[13] L.H.Ozarow, S. Shamai, and A.D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[14] Benno Zerlin, Michel Ivrlač, Wolfgang Utschick, and Josef A. Nossek, "Joint optimization of radio parameters - a top-down approach," Accepted for publication in the EURASIP Signal Processing Journal, 2005.

[15] Benno Zerlin and Josef A. Nossek, "An Approach to Cross-Layer QoS Management in MISO Systems," in *Proceedings of the 5th IEEE International Symposium for Signal Processing and Information Theory*, December 2005, Athens, Greece.

[16] Masaharu Hata, "Empirical Formula for Propagation Path-Loss in Land Mobile Radio Services," in *Proceedings of the IEEE Vehicular Technology Conference*. IEEE, August 1980, vol. 29, pp. 317–325.

[17] Ysuhiro Oda, Koichi Tsunekawa, and Masaharu Hata, "Advanced LOS path-loss model in microcellular mobile communications," in *Proceedings of the Vehicular Technology Conference*. IEEE, November 2000, vol. 49, pp. 2121–2125.

[18] G1, "TS 03.30 8.3.0 - Cell Sizes and RF Budgets," Tech. Rep., Third Generation Partnership Project, 2000.

[19] Benno Zerlin, Michael Joham, Wolfgang Utschick, and Josef A. Nossek, "Covariance based linear precoding," Accepted for publication in the IEEE Journal of Selected Areas in Communications, 2005.

[20] Martin Schubert and Holger Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2324–2334, July 2005.