Optimality of Dual Methods for Discrete Multiuser Multicarrier Resource Allocation Problems

Simon Görtzen and Anke Schmeink, *Member, IEEE*

***Abstract*—Dual methods based on Lagrangian relaxation are the state of the art to solve multiuser multicarrier resource allocation problems. This applies to concave utility functions as well as to practical systems employing adaptive modulation, in which users' data rates can be described by step functions. We show that this discrete resource allocation problem can be formulated as an integer linear program belonging to the class of multiple-choice knapsack problems. As a knapsack problem with additional constraints, this problem is NP-hard, but facilitates approximation algorithms based on Lagrangian relaxation. We show that these dual methods can be described as rounding methods. As an immediate result, we conclude that prior claims of optimality, based on a vanishing duality gap, are insufficient. To answer the question of optimality of dual methods for discrete multicarrier resource allocation problems, we present bounds on the absolute integrality gap for three exemplary downlink resource allocation problems with different objectives when employing rounding methods. The obtained bounds are asymptotically optimal in the sense that the relative performance loss vanishes as the number of subcarriers tends to infinity. The exemplary problems considered in this work are sum rate maximization, sum power minimization and max-min fairness.**

***Index Terms*—Resource allocation, adaptive modulation, orthogonal frequency division multiple access (OFDMA), duality theory, combinatorial optimization.**

I. Introduction

THE problem of resource allocation in multicarrier communication systems has been widely studied. For the single-user case, the optimal solution is achieved by classical bit-loading [1]. The problem becomes much more complex when multiple users have to be serviced, as subcarriers have to be allocated to users in a way that maximizes a specific objective function which depends on the user's data rate and power consumption on each subcarrier. Additional constraints regarding power consumption and/or data rate requirements further complicate the problem. In [2], the authors present dual methods for non-convex multicarrier resource allocation problems (RAPs) with both concave as well as non-differentiable, "discrete" utility functions. These methods are based on prior advances in optimization theory [3], [4] and have been investigated further for RAPs by various authors. In general, the main focus is on concave utility functions [5]–[8] but discrete utility functions are also considered [9]–[12], usually in addition to the concave case. In [13], a discrete RAP is analysed with a Minimum Cost Network Flow model and dual methods are applied. This list is not intended to be exhaustive but showcases the popularity of dual methods in this field. To date, dual methods are applied as a powerful tool to solve arbitrary RAPs, which are often discrete due to their application in practical systems. The performance of these methods is attributed to the fact that the duality gap vanishes when the number of subcarriers tends to infinity, which can be interpreted as allowing arbitrary time-sharing within the system. In this paper, we show that the vanishing duality gap of discrete RAPs alone is not sufficient to guarantee near-optimal performance of dual methods.

To the best of the authors' knowledge, these discrete problem formulations have never been analysed in detail from the perspective of linear and integer linear optimization. In this paper, we show that all multicarrier RAPs with discrete utility functions can be formulated as multiple-choice knapsack problems (MCKPs), which form a special class of NP-hard integer linear programs (ILPs). These combinatorial problems have been studied intensively and applied in various fields, including but not limited to operations research, VLSI design, and data compression. As an immediate conclusion, arbitrary time-sharing corresponds to a linear relaxation of the integer problem. Furthermore, the dual problems of the discrete and the relaxed formulation are identical, which means that all dual methods based on [2] can be shown to be rounding algorithms based on "discretizing" the linear programming solution. In combinatorial optimization, this is a common approach to obtain feasible approximative solutions, for example to be used in branch-and-bound algorithms. Interestingly, the performance of these rounding algorithms does not depend on the duality gap between the primal and dual problem, but on the rounding strategy and the costs induced by rounding towards a feasible solution. This implies that, at least for the discrete case, the optimality arguments of [2] and subsequent research have to be re-evaluated, as a small duality gap does not imply optimal performance of rounding algorithms.

In this work, we analyse three formulations of the RAP commonly encountered in wireless multiuser multicarrier systems, for example in the OFDMA downlink setting. These problems are sum rate maximization, sum power minimization and max-min fairness. All of them have been previously studied and solved by dual methods for concave utility functions. Dual methods for the sum rate maximization problem (SRMP) and the sum power minimization problem (SPMP) with concave utility functions are derived in [5], while problems of proportional fairness are considered in [14]. The latter can be

Manuscript received April 13, 2012; revised June 28, 2012; accepted August 6, 2012. The associate editor coordinating the review of this paper and approving it for publication was G. Yue.

This work was supported by the UMIC research cluster at RWTH Aachen University.

The authors are with the UMIC Research Centre, RWTH Aachen University, Aachen, Germany (e-mail: {goertzen, schmeink}@umic.rwth-aachen.de).

Digital Object Identifier 10.1109/TWC.2012.091812.120513

formulated as a max-min fairness problem (MMFP). In the case of a practical system with a finite number of modulation and coding schemes (MCSs), it is obviously advantageous to solve the discrete formulation instead of the concave one. Even if the utility step functions can be approximated by a concave function, one suffers quantization losses from rounding towards the nearest MCS. As this happens on every subcarrier, the performance loss scales with the number of subcarriers in the system. Furthermore, the discrete formulation is not limited to a certain shape or structure of the power-rate pairs of each subcarrier. Finally, the integer linear formulation and the dual methods discussed in this paper rely only on basic arithmetics, whereas dual methods for concave utility functions usually include at least some potentially complex logarithmic functions as well as the need to differentiate.

The remainder of this paper is structured as follows: In Section II, the discrete RAP is formally formulated. We show that discrete RAPs belong to the class of MCKPs. In this section, we also introduce the three exemplary RAPs on which we focus our analysis. Section III deals with the Lagrangian dual problems corresponding to the discrete RAPs. In this section, we present the connection between dual and rounding methods and show that the duality gap is not a sufficient measure of performance. We follow this up with an analysis of the integrality gap and feasibility of rounding solutions in Section IV, which is where we present bounds on the integrality gap as well as results on the asymptotic optimality of dual methods. Section V concludes the paper.

II. Resource Allocation Problems

We consider a wireless communication system with K users and N subcarriers, in which every subcarrier can only be used by at most one user. Let $p_{k,n}$ denote the transmit power spent for user k on subcarrier n. Then, user k's data rate on subcarrier n is given as $r_{k,n} = u_{k,n}(p_{k,n})$ for a rate utility function $u_{k,n}$ incorporating channel gain information and other factors. In the case of concave utility functions, these can usually be described by

$$u_{k,n}(p_{k,n}) = c \log_2 \left(1 + \frac{p_{k,n}}{\Gamma}\right) \tag{1}$$

with appropriately chosen fitting factors c and Γ. This logarithmic formulation is based on the Shannon bound but can also be used to approximate the data rates of practical systems [15]. However, in the case of a finite number of MCSs $m = 1, \ldots, M$, the utility functions are monotonically increasing non-negative step functions of the form

$$u_{k,n}(p_{k,n}) = \begin{cases} r_{k,1,n}, & \text{if } p_{k,1,n} \le p_{k,n} < p_{k,2,n}, \\ r_{k,2,n}, & \text{if } p_{k,2,n} \le p_{k,n} < p_{k,3,n}, \\ & \vdots \\ r_{k,M,n}, & \text{if } p_{k,M,n} \le p_{k,n}. \end{cases} \tag{2}$$

In this formulation, the channel gain information is incorporated into the power values $p_{k,m,n}$, $m = 1, \ldots, M$. As it is possible to not utilize a subcarrier at all, we assume that $(p_{k,1,n}, r_{k,1,n}) = (0, 0)$ holds. The power-rate pairs $(p_{k,m,n}, r_{k,m,n})$, $m = 1, \ldots, M$, denote the discrete set of optimal operating points of the system. A visualization is given in Fig. 1, in which a step utility function $u_{k,n}$ and the corresponding power-rate pairs $(p_{k,m,n}, r_{k,m,n})$ are depicted. For comparison, the dashed line shows a concave logarithmic utility function which has been fitted to the power-rate pairs of $u_{k,n}$. However, the results of this paper apply to arbitrary step utility functions and do not depend on the structure of the power-rate pairs.

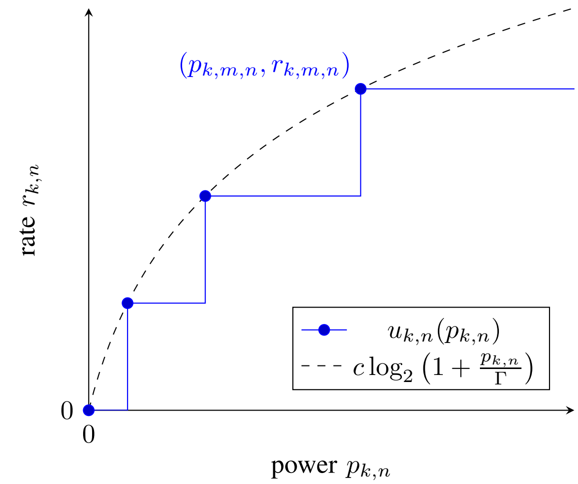

Fig. 1. Utility function $u_{k,n}$ and corresponding power-rate pairs. The dashed line shows a fitted logarithmic utility function.

We now introduce binary decision variables $x_{k,m,n} \in \{0, 1\}$ which denote if user k is using MCS m on subcarrier n. If this is the case, $x_{k,m,n}$ is one, and zero otherwise. Because every subcarrier can only be used by exactly one user with a unique MCS, we obtain the set of constraints

$$\sum_{k=1}^{K} \sum_{m=1}^{M} x_{k,m,n} = 1, \quad n = 1, \ldots, N. \tag{3}$$

A general RAP is a linear optimization problem of the form

$$\max \quad \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N} c_{k,m,n} x_{k,m,n} \tag{4}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N} a^{(i)}_{k,m,n} x_{k,m,n} \le b^{(i)}, i = 1, \ldots, s, \tag{5}$$

$$\sum_{k=1}^{K} \sum_{m=1}^{M} x_{k,m,n} = 1, \qquad n = 1, \ldots, N, \tag{6}$$

$$x_{k,m,n} \ge 0, \qquad \forall\, k, m, n. \tag{7}$$

Parameter $c_{k,m,n}$ describes the profit gained from allocating the user-MCS pair (k, m) to subcarrier n. Positive values can be used to formulate a utility maximization problem, while negative values refer to costs that have to be minimized, as, for example, in power minimization problems. Similarly, the parameters $a^{(i)}_{k,m,n}$ and $b^{(i)}$ in each of the s inequalities (5) can be used to describe system constraints like power budgets and data rate demands. Note that the constraint $x_{k,m,n} \le 1$ does not have to be enforced explicitly as it is implied by (6). The domain of the RAP is $\mathcal{D} = \mathbb{Z}^q$ with $q = KMN$, which makes it an ILP. Clearly, we can also formulate the RAP in matrix-vector notation. We denote the vector $(x_{k,m,n})_{k,m,n}$ by $\boldsymbol{x}$ without specifying an enumeration order for reasons of

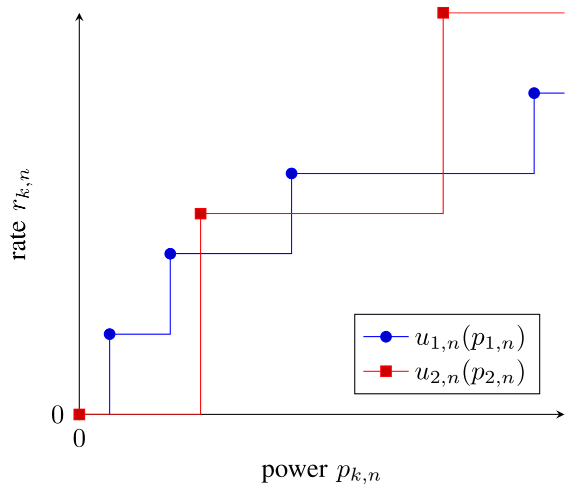

Fig. 2. Two users competing for resources on subcarrier n. No user dominates the other due to the different utility functions $u_{1,n}$ and $u_{2,n}$.

simplicity. The RAP then reads

$$\text{maximize} \quad \boldsymbol{c}^T\boldsymbol{x} \tag{8}$$
$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \tag{9}$$
$$\boldsymbol{C}\boldsymbol{x} = \boldsymbol{1}, \tag{10}$$
$$\boldsymbol{x} \ge \boldsymbol{0}, \tag{11}$$

in which the nth row of $\boldsymbol{C}$ is one in positions corresponding to the set $\{x_{k,m,n} \,|\, k = 1, \dots, K,\ m = 1, \dots, M\}$, and zero elsewhere. Here, comparisons between vectors are componentwise, $\boldsymbol{0} = (0, \dots, 0)^T$, and $\boldsymbol{1} = (1, \dots, 1)^T$.

In general, all users are competing for resources on each subcarrier. On these subcarriers, varying channel gains influence the values of $p_{k,m,n}$. Furthermore, depending on the system and problem formulation, the utility $r_{k,m,n}$ gained from utilizing MCS m on subcarrier n varies between users. Fig. 2 shows an exemplary situation of two users competing for resources on a subcarrier. In this case, the second user has a weaker channel but a higher utility than the first user.

We present three RAPs in the following subsections: Sum rate maximization and max-min fairness under a global power constraint as well as sum power minimization under individual rate constraints.

A. Sum Rate Maximization

Let the power-rate pairs $(p_{k,m,n}, r_{k,m,n})$ be as defined in (2) and a global power budget constraint p be given. Then, the SRMP is formulated as follows:

$$\max \quad \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} r_{k,m,n}x_{k,m,n} \tag{12}$$
$$\text{s.t.} \quad \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{k,m,n}x_{k,m,n} \le p, \tag{13}$$
$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m,n} = 1, \qquad n = 1, \dots, N, \tag{14}$$
$$x_{k,m,n} \ge 0, \qquad \forall\, k, m, n. \tag{15}$$

To introduce an additional degree of fairness, some formulations of the SRMP include weights $w_1, \dots, w_K$ and objective values $w_k r_{k,m,n}$. Because we allow arbitrary values for $r_{k,m,n}$, this so-called weighted SRMP is covered by the above formulation.

B. Sum Power Minimization

The second problem formulation is the SPMP under rate demand constraints r_k for each user:

$$\min \quad \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{k,m,n}x_{k,m,n} \tag{16}$$
$$\text{s.t.} \quad \sum_{m=1}^{M}\sum_{n=1}^{N} r_{k,m,n}x_{k,m,n} \ge r_k, \; k = 1, \dots, K, \tag{17}$$
$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m,n} = 1, \qquad n = 1, \dots, N, \tag{18}$$
$$x_{k,m,n} \ge 0, \qquad \forall\, k, m, n. \tag{19}$$

The so-called weighted SPMP replaces $p_{k,m,n}$ in the objective function with $w_k p_{k,m,n}$ for given weights $w_1, \dots, w_K$. As before, this formulation is covered by an appropriate transformation of variables.

C. Max-Min Fairness

The last problem is the MMFP, which strives to distribute data rate as evenly as possible among users under a global power budget constraint p:

$$\max \quad \min_{1 \le k \le K} \left\{ \sum_{m=1}^{M}\sum_{n=1}^{N} r_{k,m,n}x_{k,m,n} \right\} \tag{20}$$
$$\text{s.t.} \quad \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{k,m,n}x_{k,m,n} \le p, \tag{21}$$
$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m,n} = 1, \qquad n = 1, \dots, N, \tag{22}$$
$$x_{k,m,n} \ge 0, \qquad \forall\, k, m, n. \tag{23}$$

The objective function in (20) is not linear. However, by introducing an auxiliary variable z, the MMFP can be given in an equivalent linear formulation

$$\max \quad z \tag{24}$$
$$\text{s.t.} \quad \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{k,m,n}x_{k,m,n} \le p, \tag{25}$$
$$\sum_{m=1}^{M}\sum_{n=1}^{N} r_{k,m,n}x_{k,m,n} \ge z, \qquad k = 1, \dots, K, \tag{26}$$
$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m,n} = 1, \qquad n = 1, \dots, N, \tag{27}$$
$$x_{k,m,n} \ge 0, \qquad \forall\, k, m, n. \tag{28}$$

Because there is only a discrete range of possible values for $r_{k,m,n}$, it can be assumed w.l.o.g. that z takes only integer values, which means that the MMFP is an ILP. Strictly speaking, the introduction of z violates the prior definition of an RAP, but it is obvious that the MMFP is a problem which is not only closely related but also important for practical multicarrier systems.

A related problem is the proportional fairness problem, in which the data rates of all users have to be proportional to given factors $\phi_1 : \phi_2 : \cdots : \phi_K$. The problem of proportional fairness can be formulated as an MMFP by replacing $r_{k,m,n}$ in the objective function with $\phi_k^{-1} r_{k,m,n}$.

In the general RAP, we use indices k and m to distinguish between users and MCSs, respectively. However, this distinction is not needed from a formal point of view and user-MCS pairs can be indexed by a single variable $j = 1, \ldots, J$. Doing so results in the following problem formulation:

$$\text{maximize} \quad \sum_{j=1}^{J} \sum_{n=1}^{N} c_{j,n} x_{j,n} \tag{29}$$

$$\text{subject to} \quad \sum_{j=1}^{J} \sum_{n=1}^{N} a_{j,n}^{(i)} x_{j,n} \le b^{(i)}, \quad i = 1, \ldots, s, \tag{30}$$

$$\sum_{j=1}^{J} x_{j,n} = 1, \qquad n = 1, \ldots, N, \tag{31}$$

$$x_{j,n} \ge 0, \qquad \forall\, j, n. \tag{32}$$

The above problem is known in combinatorial optimization as the MCKP if $s = 1$, and the multidimensional MCKP for $s \ge 2$. The interpretation is to maximize the overall profit in a knapsack with weight constraint(s) $b^{(i)}$, $i = 1, \ldots, s$, by filling it with items of weight(s) $a_{j,n}^{(i)}$ and profit $c_{j,n}$. As an additional restriction, each item belongs to a class of items $n = 1, \ldots, N$, and exactly one item of each class has to be put into the knapsack. This restriction is called uniqueness or multiple-choice constraint. More details can be found in [16]. In summary, we obtain the following result.

Proposition 1. *The RAP in Section II is an MCKP with s additional inequality constraints. Thus, it is multidimensional for $s \ge 2$. The SRMP is a classical MCKP, while the SPMP is a multidimensional MCKP with K inequality constraints. In its original formulation, the MMFP is a one-dimensional MCKP with a max-min objective. In its linear formulation, it has $K + 1$ inequality constraints.*

III. Dual Methods and Rounding

As shown above, RAPs are ILPs. As MCKPs they are NP-hard [16]. Their discrete nature means that optimization techniques based on continuity or convexity are not directly applicable. In [2], a dual method based on Lagrangian relaxation is proposed. The approach is to optimally and efficiently solve the dual problem, and follow this up with a method to obtain a solution to the original RAP. The main steps of this procedure can be broadly summarized as

1) Formulate the discrete RAP.
2) Formulate and solve the dual problem.
3) From the optimal solution of the dual, obtain a solution to the RAP.

If sensible, additional optimization steps can be performed at any point. In this case, questions of performance and time complexity have to be taken into account.

Note that for an ILP of the form (8) with variable $\boldsymbol{x} \in \mathcal{D} = \mathbb{Z}^q$ and $\boldsymbol{c} \in \mathbb{R}^q$, $\boldsymbol{A} \in \mathbb{R}^{r \times q}$, $\boldsymbol{b} \in \mathbb{R}^r$, $\boldsymbol{C} \in \mathbb{R}^{s \times q}$ and $\boldsymbol{d} \in \mathbb{R}^s$, the Lagrangian dual problem is

$$\text{minimize} \quad \boldsymbol{b}^T \boldsymbol{\lambda} + \boldsymbol{d}^T \boldsymbol{\mu} \tag{33}$$

$$\text{subject to} \quad -\boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{\lambda} + \boldsymbol{C}^T \boldsymbol{\mu} \ge 0, \tag{34}$$

$$\boldsymbol{\lambda} \ge \boldsymbol{0}. \tag{35}$$

The problem is optimized over the dual variables $\boldsymbol{\lambda} \in \mathbb{R}^r$ and $\boldsymbol{\mu} \in \mathbb{R}^s$ corresponding to inequality and equality constraints, respectively. Note that this dual problem is also the dual problem D of the linear program P obtained by extending the domain of $\boldsymbol{x}$ to $\mathcal{D} = \mathbb{R}^q$. In that case, the duality theorem of linear programming implies that strong duality holds [17]. Denoting the optimal objective values of P and D by p^* and d^*, respectively, strong duality translates to $p^* = d^*$. In general, the difference $d^* - p^*$ describes the duality gap between the dual and the primal problem. In the following, we formulate the dual problems corresponding to the resource allocations problems above.

A. Sum Rate Maximization

For the SRMP, we introduce Lagrangian multipliers $\lambda \in \mathbb{R}$ and $\boldsymbol{\mu} \in \mathbb{R}^N$. The corresponding Lagrangian dual problem is

$$\text{minimize} \quad \lambda p + \sum_{n=1}^{N} \mu_n \tag{36}$$

$$\text{subject to} \quad -r_{k,m,n} + \lambda p_{k,m,n} + \mu_n \ge 0 \quad \forall\, k, m, n, \tag{37}$$

$$\lambda \ge 0. \tag{38}$$

This formulation can be simplified by implicitly moving the constraints (37) into the objective function. The constraints in (37) are equivalent to

$$\mu_n \ge r_{k,m,n} - \lambda p_{k,m,n} \qquad \forall\, k, m, n, \tag{39}$$

$$\Leftrightarrow \quad \mu_n \ge \max_{k,m} \{r_{k,m,n} - \lambda p_{k,m,n}\} \quad \forall\, n. \tag{40}$$

From (36), it is clearly optimal to chose each μ_n as small as possible, i.e., $\mu_n = \max_{k,m}\{r_{k,m,n} - \lambda p_{k,m,n}\}$ for $n = 1, \ldots, N$. Thus, the dualized SRMP is

$$\text{minimize} \quad \lambda p + \sum_{n=1}^{N} \max_{k,m} \{r_{k,m,n} - \lambda p_{k,m,n}\} \tag{41}$$

$$\text{subject to} \quad \lambda \ge 0. \tag{42}$$

Despite not being linear, this formulation is advantageous for different reasons. Most importantly, it reduces the Lagrangian dual problem to an optimization problem with a single variable λ, which can be efficiently solved with bisection methods. Furthermore, the computation of μ_n provides an optimality criterion for the user-MCS pairs of each subcarrier, which we analyse in detail later.

B. Sum Power Minimization

For the SPMP, we introduce Lagrangian multipliers $\boldsymbol{\lambda} \in \mathbb{R}^K$ and $\boldsymbol{\mu} \in \mathbb{R}^N$. This yields the corresponding Lagrangian dual problem

$$\text{maximize} \quad \sum_{k=1}^{K} \lambda_k r_k - \sum_{n=1}^{N} \mu_n \tag{43}$$

$$\text{subject to} \quad p_{k,m,n} - \lambda_k r_{k,m,n} + \mu_n \ge 0 \quad \forall\, k, m, n, \tag{44}$$

$$\boldsymbol{\lambda} \ge \boldsymbol{0}. \tag{45}$$

Analogously to the SRMP, it is optimal to chose each μ_n as small as possible while satisfying (44). With $\mu_n = \max_{k,m}\{\lambda_k r_{k,m,n} - p_{k,m,n}\}$ we obtain the dualized SPMP as

$$\text{maximize} \quad \sum_{k=1}^{K} \lambda_k r_k - \sum_{n=1}^{N} \max_{k,m} \{\lambda_k r_{k,m,n} - p_{k,m,n}\} \quad (46)$$

$$\text{subject to} \quad \boldsymbol{\lambda} \geq \mathbf{0}. \quad (47)$$

C. Max-Min Fairness

For the MMFP, we introduce Lagrangian multipliers $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_K) \in \mathbb{R}^{K+1}$ and $\boldsymbol{\mu} \in \mathbb{R}^N$. Dualizing the linear formulation in (24) yields the corresponding Lagrangian dual problem

$$\text{minimize} \quad \lambda_0 p + \sum_{n=1}^{N} \mu_n \quad (48)$$

$$\text{subject to} \quad -1 + \sum_{k=1}^{K} \lambda_k = 0, \quad (49)$$

$$\lambda_0 p_{k,m,n} - \lambda_k r_{k,m,n} + \mu_n \geq 0, \quad (50)$$

$$\boldsymbol{\lambda} \geq \mathbf{0}. \quad (51)$$

With $\mu_n = \max_{k,m}\{\lambda_k r_{k,m,n} - \lambda_0 p_{k,m,n}\}$, the dualized MMFP can be written as

$$\text{minimize} \quad \lambda_0 p + \sum_{n=1}^{N} \max_{k,m} \{\lambda_k r_{k,m,n} - \lambda_0 p_{k,m,n}\} \quad (52)$$

$$\text{subject to} \quad \sum_{k=1}^{K} \lambda_k = 1, \quad (53)$$

$$\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_K) \geq 0. \quad (54)$$

In general, due to the multiple-choice-constraint, the dualized RAP is a minimization problem that includes $\sum_{n=1}^{N} \mu_n$ in the objective. Choosing each μ_n, $n = 1, \ldots, N$, as small as possible leads to an optimality criterion

$$\mu_n = \max_{k,m} \{\alpha_k(\boldsymbol{\lambda}) r_{k,m,n} - \beta_k(\boldsymbol{\lambda}) p_{k,m,n}\} \quad (55)$$

for a set of functions α_k and β_k, $k = 1, \ldots, K$. For each $\boldsymbol{\lambda}$, this allows to solve the sub-problem of determining μ_n for each subcarrier n. The dual methods proposed by [2] are based on this property, which is usually called separability across subcarriers. See Fig. 3 and Fig. 4 for a geometric visualization of the optimality criterion.

Definition 1 (Dual Method)**.** *Let P be an RAP with multiple-choice constraint such that the dual problem D satisfies the separability property described by* (55)*. Denote by $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ the optimal solution of D. A dual method is an approach to obtain a binary vector $\boldsymbol{x} = (x_{k,m,n})_{k,m,n}$ satisfying the multiple-choice constraint such that $x_{k,m,n} = 1$ implies that the user-MCS pair (k, m) is optimal on subcarrier n with respect to* (55)*, i.e.,*

$$\mu_n^* = \alpha_k(\boldsymbol{\lambda}^*) r_{k,m,n} - \beta_k(\boldsymbol{\lambda}^*) p_{k,m,n}. \quad (56)$$

We call a dual method feasible if $\boldsymbol{x}$ is feasible, i.e., $\boldsymbol{x}$ satisfies the constraints $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$.

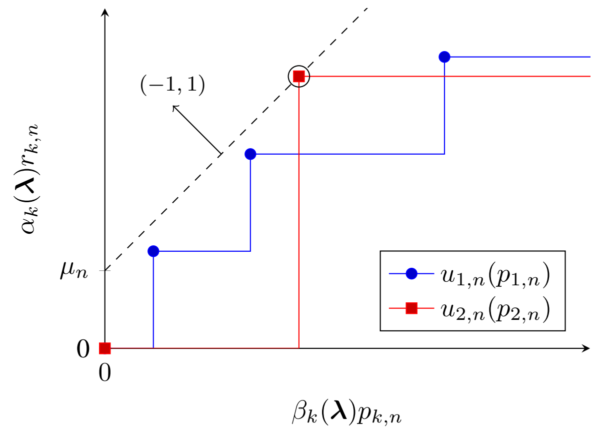

Fig. 3. This figure of the utility functions of two users shows a geometric interpretation of optimality criterion (55). As before, rate is plotted against power, but the axes have been scaled by $\alpha_k(\boldsymbol{\lambda})$ and $\beta_k(\boldsymbol{\lambda})$, respectively. Normalized this way, a user-MCS pair satisfies (55) if it is maximal with respect to the dashed line with normal $(-1, 1)$. In this case, only the single encircled pair does.

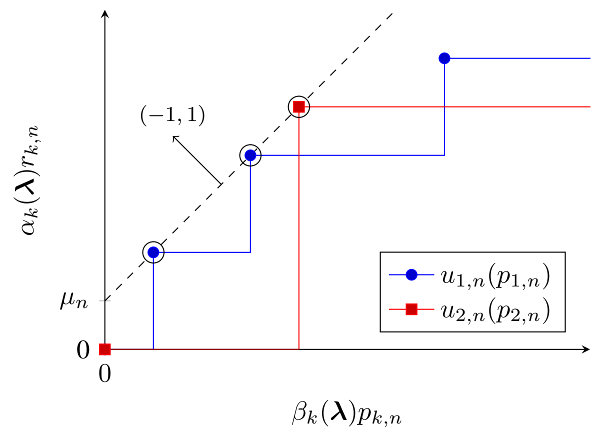

Fig. 4. Different from the situation in Fig. 3, multiple user-MCS pairs are maximal with respect to the dashed line with normal $(-1, 1)$. The three encircled pairs all satisfy optimality criterion (55).

In Fig. 3, assuming a dual optimal $\boldsymbol{\lambda}$, a dual method as defined above assigns a value of $x_{k,m,n} = 1$ to the user-MCS pair on the dashed line, allocating subcarrier n to user k to be used with MCS m. For the situation given in Fig. 4, three user-MCS pairs satisfy (55). A dual method consists of choosing one of these pairs for resource allocation on this subcarrier. Applying dual methods to continuous and discrete RAPs is suggested by [2]. The performance of the above method is attributed to the fact that the duality gap of any RAP tends to zero for $N \to \infty$. This vanishing duality gap is given as the reason for the dual method to be dual optimal. We claim that this reasoning is insufficient. Our claim is proved as follows:

Irrespective of the number of subcarriers N, a discrete RAP is never convex. The two main ways to obtain a convex problem are linear and Lagrangian relaxation. However, as noted before, the dual problems of ILPs and their linear relaxation coincide. Denote by P the linear relaxation and by D its dual problem, which is also the dual to the discrete RAP. P and D are linear problems with zero duality gap, which means that solving D is equivalent to solving P. This means that a dual method that takes the optimal solution of D

into account is equivalent to a method that takes the optimal solution of P into account. Approaches to obtain a discrete solution $\boldsymbol{y} \in \mathbb{Z}^q$ from a linearly relaxed one $\boldsymbol{x}^* \in \mathbb{R}^q$ are commonly described as rounding methods.

In the case that the optimal solution to the relaxed problems happens to be binary, it is also the optimal solution to the primal problem. In general, this will not be the case. In fact, for general multidimensional MCKP, it is not always possible to guarantee the feasibility of $\boldsymbol{y}$. Furthermore, it might not be possible to give a tight bound for the integrality gap, which describes the performance loss induced through rounding.

Definition 2 (Integrality Gap)**.** *Let P be a linearly relaxed RAP with domain $\mathcal{D} = \mathbb{R}^q$ and $\boldsymbol{x}^*$ its optimal solution. Let $\boldsymbol{y} \in \mathbb{Z}^q$ be the result of a rounding method. If $\boldsymbol{y}$ is feasible, the integrality gap is defined as their difference in objective value, i.e.,*

$$\boldsymbol{c}^T\boldsymbol{x}^* - \boldsymbol{c}^T\boldsymbol{y} = \boldsymbol{c}^T(\boldsymbol{x}^* - \boldsymbol{y}). \tag{57}$$

Proposition 2. *The integrality gap is an upper bound on the duality gap between the discrete RAP and its dual.*

Proof: Let $\boldsymbol{y}^*$ denote the optimal solution to the discrete RAP and d^* the objective value of the optimal solution to its dual problem. With the notation of Definition 2, it holds that

$$d^* - \boldsymbol{c}^T\boldsymbol{y}^* = \boldsymbol{c}^T\boldsymbol{x}^* - \boldsymbol{c}^T\boldsymbol{y}^* \leq \boldsymbol{c}^T\boldsymbol{x}^* - \boldsymbol{c}^T\boldsymbol{y}, \tag{58}$$

in which the first transformations follow from strong duality between P and D, and the inequality follows from the optimality of $\boldsymbol{y}^*$. ■

We conclude that the duality gap, however small, is not sufficient to guarantee the performance of any rounding method. To analyse the performance of duality methods, we describe the rounding process in the primal domain in detail.

Proposition 3. *Let $\boldsymbol{x} = (x_{k,m,n})_{k,m,n}$ denote the output of a dual method. Denote by $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ the optimal solution of P and D, respectively. For each subcarrier n, the rounding process implied by a dual method depends on the number of user-MCS pairs fulfilling the optimality condition* (56)*.*

1) *Let J denote the number of user-MCS pairs $(k_1, m_1), \ldots, (k_J, m_J)$, fulfilling* (56)*. Then, $x_{k_j,m_j,n} = 1$ holds for exactly one $j \in \{1, \ldots, J\}$. The linearly relaxed variables $x^*_{k_j,m_j,n}$ take arbitrary values between zero and one and sum up to one.*
2) *In the case that $J = 1$ holds, there is exactly one user-MCS pair (k, m) fulfilling* (56)*, and it holds that $x_{k,m,n} = x^*_{k,m,n} = 1$.*

Proof: For a primal-dual solution $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ of a linear program, the complimentary slackness condition holds. This means that for $\lambda_i > 0$, $\boldsymbol{x}^*$ satisfies the i-th inequality constraint with equality. As the primal problem is the dual of its dual, the same holds for the inequality constraints of the dual problem. These are given by (34) *and $\boldsymbol{x}^*$ is the corresponding dual variable. It follows that for $x^*_{k,m,n} > 0$, the constraint $\mu_n \geq \alpha_k(\boldsymbol{\lambda}^*)r_{k,m,n} - \beta_k(\boldsymbol{\lambda}^*)p_{k,m,n}$ is fulfilled with equality, which is equivalent to the user-MCS pair (k, m) fulfilling* (56)*. Thus, $x^*_{k,m,n} > 0$ implies that $(k, m) = (k_j, m_j)$ for a $j \in \{1, \ldots, J\}$. As there has to be at least one non-zero $x_{k,m,n}$ on each subcarrier, the proposition holds.* ■

We conclude that dual methods are rounding methods in which the user-MCS pair (k, m) that gets rounded up to one is picked from the set of pairs (k_j, m_j), $j = 1, \ldots, J$, fulfilling (56). In the case of $J \geq 2$ this choice is non-trivial for different reasons. On the one hand, iterative algorithms to solve the dual problem might not be able to identify all the pairs (k_j, m_j), $j = 1, \ldots, J$, for numerical reasons. On the other hand, the choice affects both the feasibility as well as the performance of the obtained solution.

IV. Integrality Gap Bounds and Feasibility

In this section, we assume that the RAP's dual problem has been solved and the optimal solution of the linearly relaxed primal problem $\boldsymbol{x}^*$ has been obtained. The following proposition gives bounds on the integrality gap of rounding methods for the RAPs discussed in this paper.

Proposition 4. *For each of the RAPs discussed above, there exists an upper bound B for the integrality gap of feasible roundings $\hat{\boldsymbol{x}}$. These bounds are:*

1) *$B = \max_{k,m,n}\{r_{k,m,n}\}$ for the SRMP,*
2) *$B = K \cdot \max_{k,m,n}\{p_{k,m,n}\}$ for the SPMP, and*
3) *$B = (K + 1)\max_{k,m,n}\{r_{k,m,n}\}$ for the MMFP.*

Proof: We cite a theorem which is known from its application to the simplex algorithm. From [18, Ch. I.2, Th. 3.5], if a linear program has a finite optimal value, it has an optimal basic feasible solution. A solution is defined as basic if the columns which correspond to its non-zero components are linearly independent. Geometrically, a basic solution lies on a simplex of the polyhedral solution space defined by the constraint set. Let $\boldsymbol{x}^*$ be a basic solution to the linearly relaxed RAP. Note that $\boldsymbol{x}^*$ optimally solves a linear program with $N + C$ constraints. For the SRMP, it holds that $C = 1$, for the SPMP, it holds that $C = K$ and for the MMFP, it holds that $C = K + 1$. As a set of linearly independent columns of dimension $N + C$ can have at most cardinality $N + C$, this limits the number of non-zero components of $\boldsymbol{x}^*$ to $N + C$ as well.

Let N' denote the number of subcarriers n for which $\boldsymbol{x}^*$ has non-binary components, i.e., there exist user-MCS pairs (k, m) with $x^*_{k,m,n} \in (0, 1)$. Then, each of the remaining $N - N'$ subcarriers n has exactly one user-MCS pair (k, m) with $x^*_{k,m,n} = 1$. Let $r = \#\{(k, m, n) \mid x^*_{k,m,n} \in (0, 1)\}$ denote the number of non-binary components of $\boldsymbol{x}^*$. Then, r can be computed by subtracting the number of components equal to one from the number of non-zero components, i.e., $r \leq N + C - (N - N') = N' + C$.

Let $\hat{\boldsymbol{x}}$ be a feasible rounding of $\boldsymbol{x}^*$. In order to round, exactly one component in each of the N' subcarriers with non-binary components is rounded up to one, and the remaining components are rounded down to zero. For the SRMP, this means that at most $C = 1$ component is rounded down to zero. As this is the only potential source of sum rate loss compared to $\boldsymbol{x}^*$, the integrality gap is bounded by $\max_{k,m,n}\{r_{k,m,n}\}$.

For the MMFP, up to $C = K + 1$ components are rounded down to zero. In the worst case, this procedure disadvantages only a single user, which reduces the max-min rate by the corresponding $r_{k,m,n}$ for every one of those components. Thus, the integrality gap is bounded by $(K + 1)\max_{k,m,n}\{r_{k,m,n}\}$.

For the SPMP, we suffer a loss in performance when spending more power than necessary. Therefore, the number of components that are rounded up to one influences the integrality gap. From above, this number is defined as N'. In order to bound N', note that each subcarrier with non-binary components contributes at least two non-binary components, as a single non-binary value can never satisfy the multiple-choice constraint (19). *It follows that $2N' \leq r \leq N' + C$. Subtracting N' from both sides yields $N' \leq C$, which, in the case of the SPMP, corresponds to $N' \leq K$. We conclude that at most K components are rounded up to one. This procedure increases the sum power by the corresponding $p_{k,m,n}$ for each of those components. Thus, the integrality gap is bounded by $K \cdot \max_{k,m,n}\{p_{k,m,n}\}$.* ■

All bounds of Proposition 4 are independent of N. Thus, any feasible rounding $\hat{\boldsymbol{x}}$ is asymptotically optimal in the sense that the relative integrality gap vanishes for $N \to \infty$ as long as the objective of the RAP scales with N, which is given for any system with sufficient subcarrier usage. In a next step, we cover the existence of feasible roundings in general.

A. Sum Rate Maximization and Max-Min Fairness

In the case of both the SRMP and the MMFP, it is always possible to obtain a feasible rounding. As it suffices to satisfy the power budget constraint, this is achieved by picking the item with lowest $p_{k,m,n}$ from each subcarrier with non-binary variables. Let $\boldsymbol{x}^*$ denote the optimal solution to the linearly relaxed primal problem. Define

$$\hat{x}_{\hat{k},\hat{m},n} = \begin{cases} 1, & (\hat{k},\hat{m}) = \arg\min\limits_{(k,m)} \left\{ p_{k,m,n} \,\middle|\, x^*_{k,m,n} > 0 \right\}, \\ 0, & \text{else.} \end{cases} \tag{59}$$

Then $\hat{\boldsymbol{x}}$ is a feasible rounding for the discrete RAP. The power budget constraint is satisfied due to

$$\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{k,m,n}\hat{x}_{k,m,n} \tag{60}$$

$$= \sum_{n=1}^{N} \min_{(k,m)} \left\{ p_{k,m,n} \,\middle|\, x^*_{k,m,n} > 0 \right\} \tag{61}$$

$$\leq \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{k,m,n}x^*_{k,m,n} \leq p. \tag{62}$$

B. Sum Power Minimization

For the SPMP, it is not straightforward to obtain a feasible rounding from $\boldsymbol{x}^*$ due to the problem's multiobjective nature. While it is generally not a problem to efficiently convert a relaxed solution to a feasible discrete resource allocation by swapping MCSs between subcarriers as needed, there is no general rounding procedure which guarantees that the rate demand of every user is fulfilled.

Proposition 5. *There exists no rounding method for the SPMP that guarantees feasibility.*

Proof: This is a property of general multidimensional MCKPs, which also applies to the SPMP. As a counterexample, consider two users sharing a single subcarrier such that both rate demands are satisfied with equality. A rounding method has to allocate the subcarrier to one of those users, which means that the data rate of the other user decreases below the allowed threshold. A minimal numerical example is given by $r_{1,m,n} = r_{2,m,n} = 2$ and $r_1 = r_2 = 1$. The optimal solution to this problem is $\boldsymbol{x}^* = (0.5, 0.5)$, but both $\hat{\boldsymbol{x}} = (1, 0)$ and $\hat{\boldsymbol{x}} = (0, 1)$ are infeasible. This shows that the discrete problem cannot be feasibly solved by rounding. ■

The above also holds for dual methods as they are effectively rounding methods. While still applicable to obtain heuristic solutions to the SPMP, asymptotic optimality results rely on bounds for the integrality gap, which can only be given if feasibility is guaranteed. To the best of the authors' knowledge, this has not been done for the SPMP or comparable problems. However, Proposition 4 gives an asymptotically optimal bound on the integrality gap for the case that a feasible rounding exists.

Practical Comments

Formulating the discrete RAP as an ILP has multiple practical advantages compared to problem formulations with concave utility functions. For a communication system with a finite number of MCSs, the discrete formulation does not require fitting and is the most accurate as it does not suffer performance loss due to quantization. Furthermore, to evaluate both the primal as well as the dual problem of the discrete formulation only basic arithmetics are required.

In addition to these advantages, the theoretical insights presented in this paper are relevant in practice as well. One practical example is a reduction of the time needed to compute a solution. From the results of Section III, how to identify and deal with multiple user-MCS pairs fulfilling (56) is of crucial importance for any kind of computation. Depending on the terminating conditions and numerical implementation, it is easily possible that one ends up with suboptimal choices. Assuming an iterative method is applied to solve the dual problem, a high number of iterations might be necessary to achieve the desired accuracy. Despite the fact that the dual problem is relatively easy to solve, it still demands a local search with additional computations, thus making each step of the iteration costly. The following facts can be used to reduce the number of iterations:

1) For each problem, there is a fixed limit to the number of competing user-MCS pairs, where two pairs are competing if they fulfill (56) on the same subcarrier. This knowledge can be used to identify competing pairs long before the dual algorithm terminates. As can be derived from the results in Section IV, specifically the bounds derived in the proof of Prop. 4, the number of competing user-MCS pairs is at most $C \in \{1, K, K+1\}$ depending on the problem.
2) Once each subcarrier is assigned to a user, the resulting RAP is easy to solve by assigning MCSs in a bit-loading fashion. Therefore, there is no need to chose between competing user-MCS pairs belonging to the same user.
3) The dual algorithm provides upper bounds for the optimal solution in each iteration, which can be used as a termination criterion once a sufficient primal solution is found via 1) and 2).

V. Conclusions

Dual methods have proven to be one of the most efficient ways to approximately solve practical RAPs encountered in wireless communications and many other fields. However, when dealing with non-convex problems, one has to take care not to take optimality for granted. In this paper, we focused on multiuser multicarrier RAPs with discrete utility functions and analyzed three exemplary problems in detail. This covered a wide range of problems encountered in practical systems employing a finite number of MCSs. The discrete nature of these problems allowed for a formulation as MCKPs, a well-known problem class in combinatorial optimization. Next, we computed the dual problems corresponding to the SRMP, the SPMP and the MMFP. We showed that all of them are separable across the subcarriers, and thus efficiently solvable. However, the approximate solutions obtained this way were shown to be equivalent to rounded solutions of the linearly relaxed primal problem. To give a performance guarantee of these methods, the integrality gap had to be bounded. We demonstrated that optimality arguments based on the duality gap are insufficient by showing that the integrality gap is always larger than the duality gap. We next derived bounds for the integrality gap of feasible roundings for the RAPs above. Furthermore, the existence of feasible roundings was shown for the SRMP and the MMFP. Both asymptotically optimal bounds of the discussed methods as well as structural properties to be utilised in practical computations were presented. We believe that this paper contributes to the understanding and analysis of dual methods for general RAPs, and discrete RAPs in particular.

References

[1] J. Campello, "Optimal discrete bit loading for multicarrier modulation systems," in *Proc. 1998 IEEE International Symposium on Information Theory*, p. 193.

[2] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.

[3] J. P. Aubin and I. Ekeland, "Estimates of the duality gap in nonconvex optimization," *Mathematics of Operations Research*, vol. 1, no. 3, pp. 225–245, 1976.

[4] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, 1996.

[5] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. 2006 IEEE International Symposium on Information Theory*, pp. 1394–1398.

[6] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 288–296, Jan. 2009.

[7] C. Liu, A. Schmeink, and R. Mathar, "Dual optimal resource allocation for heterogeneous transmission in OFDMA systems," in *2009 IEEE Globecom*.

[8] X. Wang and G. Giannakis, "Resource allocation for wireless multiuser OFDM networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4359–4372, July 2011.

[9] I. Wong and B. Evans, "Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 962–971, Mar. 2008.

[10] ——, *Resource Allocation in Multiuser Multicarrier Wireless Systems*. Springer, 2008.

[11] ——, "Optimal resource allocation in the OFDMA downlink with imperfect channel knowledge," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 232–241, Jan. 2009.

[12] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems–part II: joint chunk, power and bit allocation," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 499–509, Feb. 2012.

[13] A. Zaki and A. Fapojuwo, "Optimal and efficient graph-based resource allocation algorithms for multiservice frame-based OFDMA networks," *IEEE Trans. Mobile Computing*, vol. 10, no. 8, pp. 1175–1186, Aug. 2011.

[14] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[15] S. T. Chung and A. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[16] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*. Springer, 2004.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*. Wiley-Interscience, 1988.

Simon Görtzen received his Diploma degree in mathematics with a minor in computer science from RWTH Aachen University, Germany, in 2008. He is a PhD student at the UMIC Research Centre at RWTH Aachen University, pursuing a degree in electrical engineering and information technology. His research interests are OFDMA systems and resource allocation problems, convex optimization techniques and algorithm design.

Anke Schmeink received her Diploma degree in mathematics with a minor in medicine and the PhD degree in electrical engineering and information technology from RWTH Aachen University, Germany, in 2002 and 2006, respectively. She worked as a research scientist for Philips Research before joining RWTH Aachen University as Assistant Professor in 2008. Anke Schmeink spent several research visits at the University of Melbourne, Australia, and at the University of York, England. She is a member of the Young Academy at the North Rhine-Westphalia Academy of Science. Anke Schmeink received the E-plus best dissertation award 2006, the Vodafone young scientist award 2007 and the Helene Lange young scientist award 2009. Her research interests are in information theory, systematic design of communication systems and bio-inspired signal processing.