# Performance Prediction for OFDMA Systems with Dynamic Power and Subcarrier Allocation☆

James Gross[a], Michael Reyer[b]

[a] *UMIC Research Centre*
*RWTH Aachen University, Germany*
[b] *Institute for Theoretical Information Technology*
*RWTH Aachen University, Germany*

*Key words:* OFDMA, admission control, performance prediction, dynamic resource allocation

## Abstract

*It is well known that channel-dependent OFDMA resource assignment algorithms provide a significant performance improvement compared to static (i.e., channel-unaware) approaches. Such dynamic algorithms constantly adapt resource assignments to current channel states according to some objective function. Due to these dynamics, it is difficult to predict the resulting performance for such schemes given a certain scenario (characterized by the number of terminals in the cell and their average channel gains). In this paper we provide a novel, analytical framework for performance prediction, which takes dynamic power and subcarrier allocation into account. The analysis is based on fundamental transformations of the channel gains caused by the dynamic subcarrier allocations. This insight allows for deriving probability functions of the achieved rate per subcarrier which ultimately yields expressions for the expected minimal rates as well as outage probabilities for certain rate demands. Hence, the methods presented in this paper for performance prediction can be employed for admission control in systems with dynamic resource allocation. We illustrate the applicability of our derivations with respect to the capacity of 802.16e systems for Voice-over-IP and video streams. The results demonstrate a significant improvement compared to state-of-the art approaches but also reveal room for improvement of this approach compared to the optimal system performance.*

## 1. Introduction

Over the last decade orthogonal frequency division multiple access (OFDMA) has become one of the major multiple access schemes for broadband wireless systems. Today, OFDMA is already part of the IEEE 802.16e [1] standard for metropolitan area networking while it is also going to be the dominant multiple access technology in 3GPP's Long Term Evolution (LTE) systems for cellular networking [2]. Finally, there is some evidence that OFDMA might also be implemented in future wireless local area networks, i.e., post IEEE 802.11n systems (standardization activity has just started in IEEE 802.11's task group 'ac').

There are several reasons for this success of OFDMA as multiple access scheme. First of all, the proliferation of orthogonal frequency division multiplexing (OFDM) as favorable transmission scheme for broadband wireless links has made OFDMA the „natural" choice for multiple access. Almost all broadband wireless systems, either on the market or under standardization, are based on OFDM due to its resilience to frequency-selective fading paired with a relatively low implementation complexity [3]. OFDMA allows for fine-grained scheduling of multiple different terminals which is particularly important for packet-oriented wireless networks. In addition, mobile terminals benefit from frequency diversity if uncorrelated parts of the OFDM bandwidth are scheduled for their data transmission.

However, over the last decade one of the strongest arguments made by research for OFDMA lies in exploiting multi-user diversity. OFDM splits the given system bandwidth into small parallel communication channels referred to as subcarriers.

☆This work has been supported by the UMIC Research Centre, RWTH Aachen University.
*Email addresses:* gross@umic.rwth-aachen.de (James Gross), reyer@ti.rwth-aachen.de (Michael Reyer)

Any single link in a broadband OFDM system experiences a varying channel gain for these numerous subcarriers (stemming from frequency-selectivity). In addition, if multiple different receivers (i.e., links) are considered (for example, in the down-link), the different receivers experience very different gains for any given subcarrier (as channel gain is uncorrelated in space over the distances usually separating multiple terminals). So, physical layer efficiency can be improved significantly by allocating suitable subcarriers to terminals in a frequency division multiplexing fashion. Such channel-dependent OFDMA resource assignment is now under discussion for about ten years [4, 5]. Research mainly focuses on assignment algorithms to be applied for the down-link of wireless OFDMA cells. Many different algorithms have been proposed in the past, addressing mainly the trade-off between complexity (i.e., run time of the algorithm) and physical layer efficiency. Apart from that, some contributions have also investigated the associated overhead (collecting the channel state information as well as signaling the dynamic assignments). IEEE 802.16e as well as 3GPP's LTE enable the application of channel-dependent OFDMA resource assignments by providing the protocol requirements in the system.

While channel-dependent OFDMA resource assignments do provide a significant performance gain, these schemes also cause new problems. One particularly important issue deals with the predictability of system performance which is strongly related to admission control: Assume that a certain algorithm at the base station assigns subcarriers to terminals for each down-link phase in order to improve some metric (for example, proportional fair throughput or overall throughput). Based on stochastic simulations, it is relatively easy to determine the average performance obtained from any such algorithm for some chosen scenario (number of terminals in the cell, distance of the terminals to the base station etc.). However, it is much harder to *predict* the obtained performance of any assignment algorithm for some given scenario due to the channel-dependent nature of these algorithms. Unfortunately, some form of performance prediction is required, for example for admission control. To see this, assume that a certain number of Voice-over-IP (VoIP) calls are currently admitted to a wireless OFDMA cell. Next, five new calls arrive and request admission to the cell. Some method is clearly required to judge if the channel-dependent OFDMA resource assignments can still fulfill the quality-of-service requirements of all the voice streams if the new ones are admitted to the cell. In a more general setting, we need a framework to determine if an OFDMA cell can support a given set of transmission requests (characterized by quality-of-service demands as well as the average channel gains between the base station and the requesting terminal) with a given amount of resources (subcarriers, power, modulation and coding schemes). Notice that due to the stochastic nature of the wireless channel the framework of interest is of stochastic nature.

In this paper we present a novel framework for performance prediction in OFDMA cells which feature dynamic subcarrier *and* power assignments. The framework allows the prediction of achievable rates given a certain amount of resources and a certain set of requests. In contrast to previous work [6–9] (see also Section 6 for a detailed discussion), our major contribution is that we provide an *analytical* framework for performance prediction of *channel-dependent* OFDMA resource assignments (including the assignment of subcarriers *and* power). The analytical framework is based on a fundamental insight into such OFDMA resource assignment algorithms regarding the way they modify the fading statistics of *assigned* subcarriers, as presented by previous work of ours in [9]. We provide a closed form expression for these resulting fading statistics (which we refer to as OFDMA channel transformations). From this analytical core and by means of power allocation as in [10] a lower performance bound for a set of OFDMA subcarrier assignment optimization problems is derived. This lower bound can be used to perform admission control with respect to different policies, as illustrated for VoIP and video streams in this paper. To the best of our knowledge, this approach is novel.

The remaining paper is structured as follows. In Section 2 we present the basic system model including the physical layer and state the problem of interest. Section 3 introduces the essential steps that are applied for performance prediction. As we pursue an iterative approach, the individual steps depend on each other which has implications for certain parameter settings in each step. Hence, there is room for improving the prediction quality which is discussed in Section 4. The proposed method for performance prediction is then evaluated numerically and compared to optimal solutions in Section 5. Section 6 discusses the state-of-the art. Finally, we conclude the paper in Section 7.

## 2. System Model

We consider a single cell of a wireless system with a base station serving $J$ terminals. The system is centrally organized, all data transmissions within the cell are controlled by the base station. Time is split into frames while each frame features a down-link and an up-link phase. In the following we focus on the down-link phase and denote its duration by $T_{\mathrm{dl}}$. Via a backbone the base station receives data destined for the terminals in the cell. Upon transmission, this data is queued separately for each terminal. Prior to each down-link phase, the base station schedules (some or all of the) currently queued data for transmission during the next down-link phase. Denote these scheduled data amounts for terminal $j$ by $\sigma_j$. Afterwards a resource assignment unit tries to match system resources with the set of requested data amounts $\{\sigma_j\}$ as good as possible. In the following, we first describe these system resources in Section 2.1 and focus afterwards on the resource assignment and link layer in Section 2.2. Finally, we present the problem statement in Section 2.3.

## 2.1. Physical Layer

Data is transmitted via an OFDM system of total bandwidth $B$ [Hz] at a center frequency $f_c$ [Hz]. A maximum transmit power of $P_{\max}$ can be utilized in this frequency band. The bandwidth is split into $N$ subcarriers on which information is transmitted in parallel by digital symbols of length $T_s = 1/\Delta f = N/B$. We refer to all $N$ subcarrier symbols transmitted simultaneously as an OFDM symbol. In order to mitigate intersymbol interference, a guard period of length $T_g$ is added to each OFDM symbol. Per down-link phase, the system features $S = \lfloor T_{dl} / \left( T_s + T_g \right) \rfloor$ OFDM symbols for data transmission.

For each subcarrier $n$ and terminal $j$ the gain $g_{j,n}$ varies over time and frequency, i.e., each subcarrier/terminal gain depends on a constant component (path loss, denoted by $\rho_j$) and a random, time- and frequency-variant fading component. We assume this gain to be exponentially distributed (with mean $\rho_j$). Matrix $\mathbf{G}$ groups all subcarrier/terminal gains. $\mathbf{G}$ is assumed to stay constant for the duration of one down-link phase but varies over longer time spans. Based on the transmit power $p_n$ per subcarrier and the variance $\sigma^2$ of the white Gaussian noise process per subcarrier, we obtain the signal-to-noise ratio (SNR) of subcarrier $n$ and terminal $j$ per down-link phase by:

$$\gamma_{j,n} = \frac{p_n \cdot g_{j,n}}{\sigma^2} \ . \tag{1}$$

We assume $M$ different modulation types to be featured by the transmission system. Modulation type $m$ represents $b_m$ bits per symbol (for example $b_m = 2$ for QPSK). The base station can adapt the modulation type for each subcarrier separately (referred to as adaptive modulation). However, any down-link transmission is constrained by a terminal specific target bit error rate $\beta_j$. Hence, a modulation type $m$ is applied if the current SNR for this terminal is within the range $\Gamma_{j,m} \leq \gamma < \Gamma_{j,m+1}$. The SNR range delimiters are determined by modulation specific bit error probability functions, as for example presented in [11]. Denote by $F_j(\gamma)$ the function returning the amount of bits that can be transmitted per OFDM symbol to terminal $j$ at an SNR of $\gamma$. Notice that for any SNR below $\Gamma_{j,1}$ no modulation is applied, i.e., $F_j\left(\gamma < \Gamma_{j,1}\right) = 0$. Also, for modulation $M$ the range has no upper limit.

## 2.2. Medium Access Control Layer and Resource Assignment

During each down-link phase frequency division multiplexing is applied. Thus, during the resource assignment step disjoint subsets of subcarriers are assigned to terminals based on perfect knowledge of subcarrier gains. Subcarrier assignments are valid throughout the entire down-link phase. The assignment of a subcarrier/terminal pair is denoted by the binary variable $x_{j,n}$. In addition to the subcarrier/terminal assignments, the resource assignment also contains a variable transmit power $p_n$ per subcarrier.

As the scheduler requests the resource assignment unit to transmit the data amount $\sigma_j$ per terminal during the next down-link phase, the resource assignment unit first identifies the minimum requested data amount $\sigma_j^\star$ and computes then for all terminals the proportion factor:

$$\alpha_j = \sigma_j / \sigma_j^\star \ . \tag{2}$$

Based on these quantities the resource assignment unit solves the following rate-adaptive optimization problem [12, 13]:

$$\begin{aligned} \max \quad & \epsilon \\ \text{s. t.} \quad & \sum_j x_{j,n} \leq 1 \quad \forall\, n \\ & S \cdot \sum_n F_j\left(\frac{p_n \cdot g_{j,n}}{\sigma^2}\right) \cdot x_{j,n} \geq \alpha_j \cdot \epsilon \quad \forall\, j \\ & \sum_n p_n \leq P_{\max} \end{aligned} \tag{3}$$

In this maximization of the minimum rate the factor $\alpha_j$ assures a proportional scaling of the achieved minimum rate in case that the scheduler requests different data amounts per terminals.

As the system operates on a discrete set of modulation types, the rate-adaptive optimization problem can be restated as integer program such that only binary decision variables are taken into account:

$$\begin{aligned} \max \quad & \epsilon \\ \text{s. t.} \quad & \sum_{j,m} y_{j,n,m} \leq 1 \quad \forall\, n \\ & S \cdot \sum_{n,m} b_m \cdot y_{j,n,m} \geq \alpha_j \cdot \epsilon \quad \forall\, j \\ & \sum_{j,n,m} p_{j,n,m} \cdot y_{j,n,m} \leq P_{\max} \end{aligned} \tag{4}$$

In this problem formulation the decision variable $y_{j,n,m}$ models the decision to assign subcarrier $n$ with modulation type $m$ to terminal $j$ (or not). This assignment requires a transmit power of $p_{j,n,m}$ which directly results from (1) and the fact that per terminal $j$ specific SNR ranges $\Gamma_{j,m}$ are employed for each modulation type depending on the terminal specific bit error rate constraint $\beta_j$. The first constraint ensures that each subcarrier is assigned at most once to a specific terminal/modulation combination. The second constraint ensures that the assigned rate per terminal is above the lower limit $\epsilon$ while the last constraint limits the transmit power to $P_{\max}$.

Note that both problems (3) and 4 are NP-hard [5]. Different approaches have been suggested how to solve these problems by faster heuristics [13, 14]. In the following, we do not consider the specific algorithm further. We focus on the IP formulation of the rate adaptive problem as given in (4) and simply assume that it can be solved prior to each down-link phase optimally or with negligible performance degradation (but likely with a high computational overhead leading to execution times close to the length of a down-link phase). Finally, some control channel is needed to inform the terminals of their assignments (subcarriers and modulation types). We assume the existence of such a separate, error-free control channel conveying this information to the terminals.

### 2.3. Problem Statement

We address the problem of performance prediction and admission control for OFDMA down-link. When scheduling per terminal a data amount of $\sigma_j$ the scheduler needs to estimate previously $\epsilon$. The efficiency of the system depends strongly on the interplay between scheduler and resource allocation unit. If the scheduler requests too much data to be transmitted per down-link phase, most of the scheduled data will not be transmitted completely. On the other hand, if the scheduler requests too few data to be transmitted, the queued data will suffer from additional delays. Note that an iterative approach between scheduler and resource allocation unit would characterize the possible system performance per down-link phase quite well. However, it is unlikely to be implemented as a single iteration step would require optimization problem (4) to be solved which is quite time-consuming. Therefore, solving it several times in a row is prohibitively costly.

Clearly, the considered performance estimate of $\epsilon$ depends on the number of terminals $J$, the two sets of their respective average channel gains $\{\rho_j\}$ and their target bit error rates $\{\beta_j\}$ as well as the resources available for transmission (i.e., the total transmit power $P_{\max}$, the total bandwidth $B$, the amount of subcarriers $N$, the length of the down-link phase $T_{\text{dl}}$ as well as the set of available modulation types). Notice that such an estimate resembles on a larger time scale one key component to admission control for (the down-link of) OFDMA systems. Flows destined for terminals within the cell request admission and quality-of-service requirements are specified, given for example by their average rate requirement, delay tolerance, packet loss rates etc. In order to decide about admission (with respect to the already admitted flows and their respective quality-of-service requirements) the scheduler needs to determine the impact of admitting the new flow on $\epsilon$.

Any framework enabling the scheduler to estimate $\epsilon$ is of probabilistic nature. That is, given a characterization of the load and the available resources, we can only hope to derive probabilities that the solution to problem (4) results in some (required or estimated) $\epsilon$ – unless we restrict ourselves to an instance of matrix $\mathbf{G}$ (in which case we are rather trying to design a suitable algorithm for assigning resources than dealing with analytical performance predictions). Hence, when scheduling data portions $\sigma_j$ for each terminal, these decisions have to be based on some probability that the resource assignment unit will not be able to transmit the requested capacity. In the following we refer to such „scheduling misses“ simply as outages.

## 3. OFDMA Channel Transformations and Power Loading

Ideally, we could directly derive an analytical framework from the rate adaptive IP formulation in (4). However, as this problem is NP-hard, the general (optimal) solution to it depends on exhaustive search. It is therefore not analytically tractable. Still, we are interested in deriving some analytical framework and take therefore the following approach. We split the assignment of subcarriers and power into two subsequent steps. For the first step – the subcarrier assignment presented in Section 3 – we pick a suboptimal algorithm which can be captured by analysis. The analysis of this algorithm allows the derivation of the channel gain distributions *after* the assignment depending on the input distributions of the subcarrier gains. This can be interpreted as a transformation of the gain distributions by the channel-dependent OFDMA resource allocation. Based on these modified channel gains, in the second step we employ an optimal power loading algorithm to determine the power allocation per subcarrier, as presented in Section 3.2. These two steps together yield a lower bound on the estimate for $\epsilon$, in fact we can even derive rate probability mass functions for each terminal individually as finally shown in Section 3.3. Note that both steps depend on some parameterization and there is room for improving the predicted system performance by adjusting the parameters accordingly. In this section though, we only present the basic schemes and their mathematical analysis while the next section covers the prediction improvement by an appropriate (iterative) parameterization.

### 3.1. First Step: Assignment Algorithm

The underlying algorithm considered for the first task, the subcarrier assignment, takes the channel state information matrix **G** as input and works as the following. Initially, each terminal considered for the next down-link phase is allocated a certain number of subcarriers $l_j$. Note that this only fixes the amount of subcarriers obtained by each terminal, the assignments are still open. Given the allocation of subcarriers, the algorithm iterates in a round robin fashion over the set of terminals each time assigning each terminal the best subcarrier out of the set of remaining ones. After assigning a subcarrier, the subcarrier is taken out of the list of available ones. If for some round a terminal reaches its allocated number of subcarriers, the terminal is no longer considered in later rounds. A more formal description of the algorithm is presented below as Algorithm 1.

```
    Given:      Subcarrier gains G
                Set of subcarrier allocations {l_j}
    Initialize: ∀ j ∈ J : X_j = ∅
                N = {1, ··· , N}
while (N ≠ ∅) do
    for (j ∈ J) do
        if (l_j > |X_j|) then
            ñ ← argmax_{n∈N}{g_{j,n}}
            X_j ← X_j ∪ {ñ}
            N ← N \ {ñ}
        end
    end
end
return {X_1, . . . , X_J}
```

Algorithm 1: Scheme of the approximation algorithm.

#### 3.1.1. Basic Analysis

Two sets of variables determine the resulting performance obtained per terminal: the set of allocated subcarriers, i.e., $\{l_j\}$ and the order with which the terminals are served. In the following let us assume that both are fixed. Let us consider terminal $j$. Recall that we assume the subcarrier gain of an arbitrarily selected subcarrier for terminal $j$ to be exponentially distributed with mean $\rho_j$. Denote the random channel gain of an arbitrarily selected subcarrier for terminal $j$ by $g_j$, then the density function for this random variable is given by:

$$f_{g_j}(x) = \frac{1}{\rho_j} \cdot e^{\frac{-x}{\rho_j}} \tag{5}$$

with the corresponding distribution function:

$$P\left[g_j \leq x\right] = F_{g_j}(x) = 1 - e^{\frac{-x}{\rho_j}} \ . \tag{6}$$

However, the $l_j$ subcarriers assigned by the algorithm mentioned above are not arbitrarily chosen. Instead, they are the best subcarriers chosen out of the remaining set of subcarriers per round.

Denote by $A_{j,(1)}, \ldots, A_{j,(l_j)}$ the amount of remaining subcarriers before the algorithm selects the respective best subcarrier for terminal $j$ per round. Furthermore, denote by $\tilde{g}_{j,(1)}, \ldots, \tilde{g}_{j,(l_j)}$ the random channel gains of these $l_j$ best subcarriers chosen per round out of the set of remaining subcarriers. We can characterize their density function and distribution function by applying results from order statistics [15].

Let $X_1, X_2, \ldots, X_N$ be $N$ independent and identically distributed random variables. Denote by $\tilde{X}_{(N)}$ the largest of these $N$ random variables. Then from order statistics it is well known that the density function of $\tilde{X}_{(N)}$ is given by:

$$f_{\tilde{X}_{(N)}}(x) = N \cdot P(X \leq x)^{N-1} \cdot f_X(x) \tag{7}$$

and correspondingly we have for the distribution function:

$$P\left(\tilde{X}_{(N)} \leq x\right) = P(X \leq x)^N \ . \tag{8}$$

Based on Equations (5), (6) and (7), we obtain the density function of the random gain of the best subcarrier selected for terminal $j$ during round $k$ out of the set of $A_{j,(k)}$ remaining subcarriers by:

$$f_{\tilde{g}_{j,(k)}}(x) = \frac{A_{j,(k)}}{\rho_j} \cdot \left(1 - e^{\frac{-x}{\rho_j}}\right)^{A_{j,(k)}-1} \cdot e^{\frac{-x}{\rho_j}} \ . \tag{9}$$

Correspondingly, we obtain for the distribution function from Equations (5), (6) and (8):

$$F_{\tilde{g}_{j,(k)}}(x) = \left(1 - e^{\frac{-x}{\rho_j}}\right)^{A_{j,(k)}} \ . \tag{10}$$

Note that these equations are based on the assumption of independence and identical distribution of the subcarrier gains. While the property of identical distribution is likely to exist, independence strongly depends on the spacing between subcarriers and the coherence frequency of the propagation environment. We restrict our analysis to cases where independence can be assumed due to a coherence bandwidth which equals roughly the bandwidth of a subcarrier.

Equations (9) and (10) formulate explicitly the effect of dynamic resource allocation in OFDMA systems. The fading statistic of subcarriers is thus *transformed* by the resource allocation. We refer to this as OFDMA channel transformations. The resulting analytical density and distribution can be used for any probabilistic bound, for example on the rate or the power consumption. It is clear that regarding any such objective function, there exist more and less suitable choices for the set of allocations $\{l_j\}$ and the selection order. However, the fact that we have analytical expressions for the stochastic behavior of the channel gains allows us to study some insights of channel-dependent OFDMA resource assignments before continuing the specific analysis with respect to optimization problem (3).

#### *3.1.2. Example Numerical Results*

Consider the following system with a total of $N = 48$ subcarriers and $J = 6$ terminals. Subcarrier allocation is kept simple, each terminal is allocated $l_j = 8$ subcarriers. All terminals have an average channel gain of $\rho_j = 1$ (no path loss as well as assuming them to be positioned along a circle around the base station). For each down-link phase subcarriers are assigned according to the approximation algorithm. From Equation (9) we derive the density functions for the first subcarrier assigned to each terminal (so, during the first round) in Figure 1. We observe from the figure a much better gain PDF than for the exponentially distributed gain of the basic subcarriers. Furthermore, these best subcarriers all have almost the same gain PDF as the basic subcarrier set from which to choose from (48 down to 42 subcarriers) is comparable large. On the other hand, considering in Figure 2 the density functions of the subcarriers chosen in the last round we observe that even these have significantly better gain densities compared to the exponentially distributed subcarrier gain (except for the very last subcarrier chosen). This demonstrates the impact of exploiting multi-user diversity.

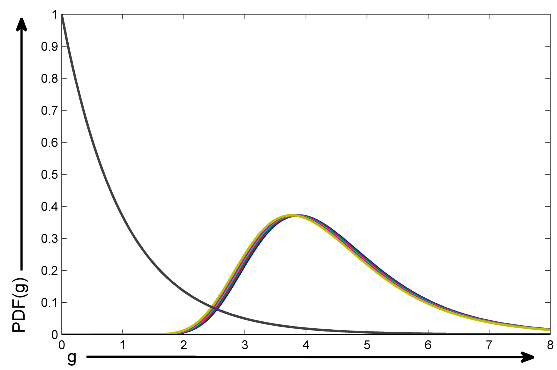


**Figure 1. Probability density functions of the first (i.e., best) subcarrier selected for each terminal (48 subcarriers, 6 terminals considered).**

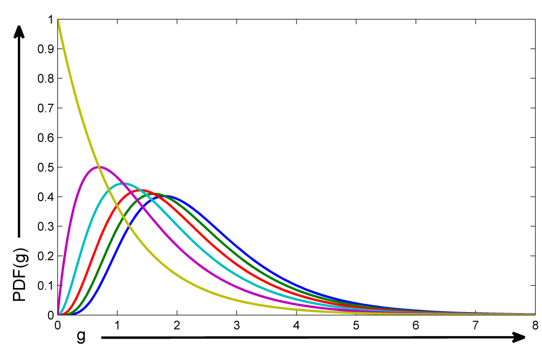


**Figure 2. Probability density functions of the last (i.e., worst) subcarrier selected for each terminal (48 subcarriers, 6 terminals considered).**

## 3.2. Second Step: Power Loading

Recall that the first step above only provides a vector of gain distribution functions $\vec{\mathbf{f}} = \left\{ f_{\tilde{g}_{j,(k)}}(x) \;|\forall j \in J,\; \forall k \in l_j \right\}$ for each terminal $j$ with respect to its allocated amount of $l_j$ subcarriers. However, we are interested in predicting the performance of an OFDMA system that allocates subcarriers and power dynamically. Hence, we still require some power assignment per subcarrier. In the context of OFDMA systems this power assignment process is often referred to as *power loading*[1].

Theoretically, we could directly perform a power loading on the vector of distribution functions $\vec{\mathbf{f}}$, effectively solving a stochastic optimization problem. Instead, we take the following straightforward approach: From the vector of channel gain distribution functions, we derive a vector of expected channel gains divided by the noise variance:

$$\vec{\mathbf{c}} = \left\{ c_{j,(k)} = E\left[\tilde{g}_{j,(k)}\right] / \sigma^2 \;|\forall j \in J,\; \forall k \in l_j \right\}. \tag{11}$$

Based on this vector of normalized average channel gains we are interested in the solution to the following variant of the rate-adaptive problem where the gain per subcarrier is already fixed (taking the values of $\vec{\mathbf{c}}$ and hence no subcarrier assignments are performed):

$$\begin{aligned} \max \quad & \epsilon \\ \text{s. t.} \quad & S \cdot \sum_{k \in l_j} F_j\left(p_{j,(k)} \cdot c_{j,(k)}\right) \geq \alpha_j \cdot \epsilon \quad \forall\ j \\ & \sum_j \sum_{k \in l_j} p_{j,(k)} \leq P_{\max} \end{aligned} \tag{12}$$

This yields a power assignment vector:

$$\vec{\mathbf{p}} = \left\{ p_{j,(k)} \;|\forall j \in J,\; \forall k \in l_j \right\}$$

which is then used to derive rate probability functions (see Section 3.3). Notice that the optimization problem above is equivalent to the problem of maximizing rates subject to proportional fairness[2] and power constraints, which is addressed in [10, 17–19]. The solution to the above optimization problem faces two problems: First of all, the function $F_j(\gamma)$ is assumed to be a step function as modulation types are used within certain SNR ranges (see Section 2.1). Second, the above optimization problem is much easier computed by iterating over a „dual“ version of the problem referred to as margin-adaptive one. In the following, we discuss these issues. Mathematical details on the algorithmic approach can be found in [10]. The full derivations of the following results are presented there.

### *3.2.1. Iterated Power Loading*

Let us assume in the following that we have obtained a continuous relationship $\psi_j(\gamma)$ between achievable rate and SNR which approximates $F_j(\gamma)$. Details how to obtain such a function are discussed below in Section 3.2.2. For such a continuous „power-rate“ function the so called margin-adaptive optimization problem is much easier to compute. This problem asks for the minimal power that is required to transmit a fixed amount of bits to each terminal in the OFDMA system. Taking the notation from Problem (12), the margin-adaptive problem is given by:

$$\begin{aligned} \min \quad & \sum_j \sum_{k \in l_j} p_{j,(k)} \\ \text{s. t.} \quad & S \cdot \sum_{k \in l_j} \psi_j\left(p_{j,(k)} \cdot c_{j,(k)}\right) = \alpha_j \cdot \epsilon \quad \forall\ j \end{aligned} \tag{13}$$

Note that for the optimal solution $\epsilon^*$ of problem (12) (assuming a continuous power-rate function) equality holds for the first inequality constraint, i.e., no terminal $j$ has a larger rate than $\alpha_j \cdot \epsilon$. Using that fact enables to determine if an $\epsilon$ can be achieved in the rate-adaptive problem (12) by comparing the minimum power $p^*(\epsilon)$ of the corresponding margin-adaptive problem (13) to the maximum power $P_{\max}$. If $p^*(\epsilon) \leq P_{\max}$ holds, the optimal solution $\epsilon^*$ of the rate-adaptive problem (with fixed subcarrier assignment) is obviously greater or equal to $\epsilon$. This fact is formally described in the following proposition.

[1] In information theory it is also referred to as water-filling [16].

[2] Proportional fairness refers here to the terminal rates, which are proportional with factors $\alpha_j$.

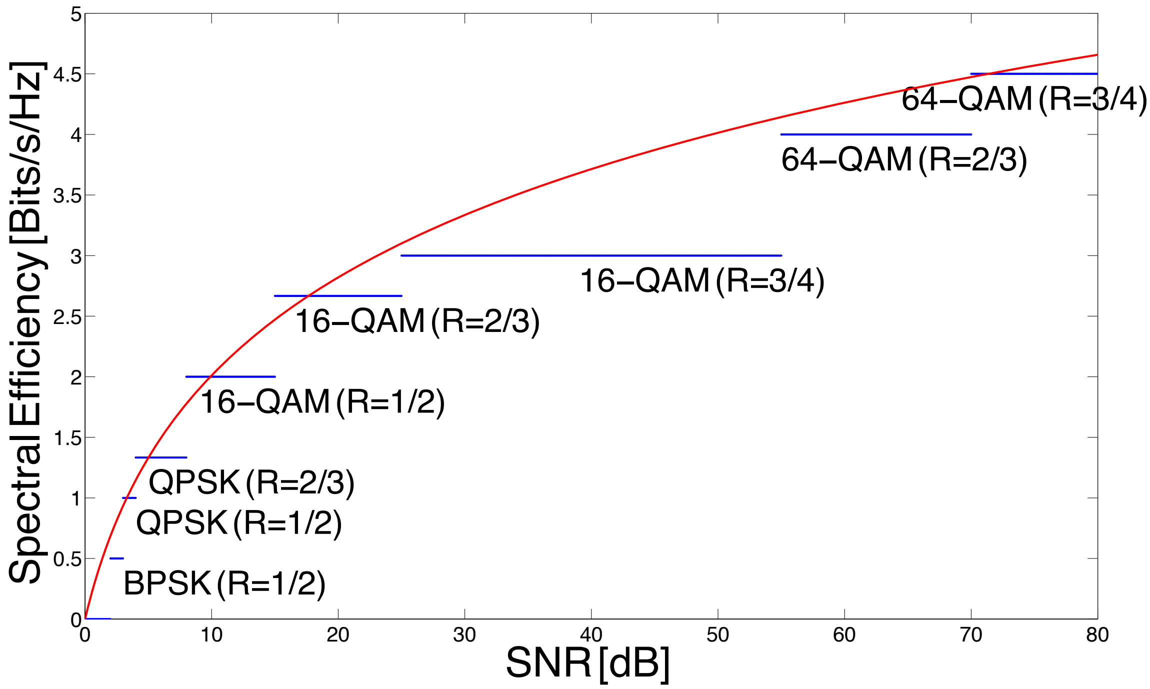


**Figure 3. Power-rate function, cf. [20].**

**Proposition 1.** *The optimum of the rate-adaptive problem $\epsilon^*$ and the margin-adaptive problem subject to proportionality constraints (and a fixed assignment) satisfy:*

$$\epsilon^* = \max\{\epsilon \in \mathbb{R}_+ \mid p^*(\epsilon) \le P_{\max}\}.$$

Hence, a solution to the proportional rate-adaptive problem may be determined by a nested algorithm where in the outer loop $\epsilon$ is increased and in the inner loop $p^*(\epsilon)$ is computed until the maximum power $P_{\max}$ is reached.

As no subcarrier assignments are performed in this step, determining $p^*(\epsilon)$ decomposes into $J$ separate subproblems as:

$$p^*(\epsilon) = \sum_j p_j^*(\alpha_j \cdot \epsilon). \tag{14}$$

Each $p_j^*(\alpha_j \cdot \epsilon)$ requires to determine the minimum power per allocated subcarrier to terminal $j$ such that the sum rate achieved on all $l_j$ subcarriers equals $\alpha_j \cdot \epsilon$. This problem is known as (margin-adaptive) generalized water-filling, which can be computed efficiently for certain classes of power-rate functions. One example class will be discussed in the following.

*3.2.2. A Concrete Class of Power-Rate Functions*

As an approximation to actually implementable power-rate functions and its inverse we use:

$$\begin{aligned} \psi(\gamma) &= \log_2\left(1 + \frac{\gamma}{a}\right), \quad \gamma \ge 0, \\ \psi^{-1}(r) &= a \cdot (2^r - 1), \qquad r \ge 0 \end{aligned} \tag{15}$$

as depicted in Figure 3. It is widely used in the literature (e.g., [21]) and well motivated by achievable rates using PSK and QAM modulation schemes, see [22]. Parameter $a > 0$ depends on the maximal acceptable BER and the available combinations of modulation and coding schemes. We have approximated those parameters for terminal $j$, rates $b_m$, and corresponding SNR range delimiters $\Gamma_{j,m}$ by minimizing the mean squared error. Denote the average channel gains allocated to each terminal $j$ in the following by vector:

$$\vec{\mathbf{c}}_j = (c_{j,(1)}, \ldots, c_{j,(l_j)}).$$

Applying the above introduced power-rate function on Equation (14) yields the following solution for the margin-adaptive problem:

$$
\begin{aligned}
p^*(\epsilon) &= \sum_j p_j^*(\alpha_j \cdot \epsilon) \\
&= \sum_j a \cdot l_j \left[ \frac{1}{\mathrm{gMean}\left(\vec{\mathbf{c}}_j\right)} 2^{\frac{\alpha_j \cdot \epsilon}{l_j}} - \frac{1}{\mathrm{hMean}\left(\vec{\mathbf{c}}_j\right)} \right]
\end{aligned}
\tag{16}
$$

In Equation (16)[3] the functions gMean(·) and hMean(·) are the geometric mean as well as the harmonic mean.

Formula (16) can be efficiently computed such that the nested algorithm derived from Proposition 1 is applicable. Let us focus now on the outer iteration in the following. We first give a lower and upper bound for $\epsilon^*$ which is of great help in the numerical computations. To determine the bounds given in Proposition 2 we need the optimal solution of the single-user rate-adaptive problem for fixed transmit power $p$, which is given by:

$$
r_j^*(p) = l_j \cdot \log_2 \left( \mathrm{gMean}\left(\vec{\mathbf{c}}_j\right) \left[ \frac{p}{a \cdot l_j} + \frac{1}{\mathrm{hMean}\left(\vec{\mathbf{c}}_j\right)} \right] \right). \tag{17}
$$

**Proposition 2.** *For a multi-user rate-adaptive problem with proportionality constraints (and fixed subcarrier assignments):*

$$
\min_j \frac{r_j^*\left(\delta_j \cdot P_{\max}\right)}{\alpha_j} \leq \epsilon^* \leq \max_j \frac{r_j^*\left(\delta_j \cdot P_{\max}\right)}{\alpha_j}
$$

*holds for all* $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J) \in \mathbb{R}_+^J$ *with* $\sum_j \delta_j = 1$.

In this proposition the maximal transmit power $P_{\max}$ is divided arbitrarily onto the terminals by $\boldsymbol{\delta}$. Then the relative rates (i.e., $r_j^*\left(\delta_j \cdot P_{\max}\right)/\alpha_j$) for each terminal $j$ are calculated by Equation (17). From Proposition 2 we can conclude that the optimum $\epsilon^*$ is bounded by the minimal and maximal relative rate over all rates of all terminals $J$. Note that for the optimal solution all relative rates are equal. Hence, we need a good choice for an initial power division $\boldsymbol{\delta}$. Obviously, the relative power for terminal $j$ and given $\epsilon$ is:

$$
\Delta_j(\epsilon) = \frac{p_j^*(\alpha_j \cdot \epsilon)}{\sum_l p_l^*(\alpha_l \cdot \epsilon)}, \tag{18}
$$

such that $\left(\Delta_j(\epsilon^*)\right)$ will produce the optimal power division and therefore result in the optimal $\epsilon^*$.

Algorithm 2 numerically computes the optimal parameter $\epsilon^*$ of a given multi-user rate-adaptive problem with proportionality constraints. It stops if the relative error of the current power $p_{\min}$ is less than $\eta$. First, $\epsilon^*$ is limited by the bounds of Proposition 2 with $\delta_j = \Delta_j(1)$. After that, the corresponding powers $p_{\min}$ and $p_{\max}$ are computed according to (16). The overall transmit power is convex in the rates because it is a linear combination of convex functions $\psi^{-1}$. Consequently, its inverse $\psi$ is concave. The rate is proportional to $\epsilon$, such that the function representing $\epsilon$ given a power is concave. This ensures that the updates of $\epsilon_{\min}$ stay below the optimal level $\epsilon^*$ within each loop. The updates converge strictly monotonic increasing to $\epsilon^*$ corresponding to the desired power $P_{\max}$. Figure 4 illustrates the first three steps of the algorithm. The complexity of the algorithm is $O(J)$ assuming the number of iterations to be constant. This is a reasonable assumption as simulations have shown that the average number of iterations stays below four for $\nu = 10^{-4}$ and number of subcarriers up to $N = 2048$. Note that the evaluation of the geometric as well as the harmonic mean have complexity $O(1)$ as they are constant within the algorithm because the subcarrier gains $\vec{\mathbf{c}}$ are constant.

With help of the optimal $\epsilon^*$ the corresponding optimal powers $p_{j,(k)}$ for terminal $j$ and its $k$-th subcarrier are finally obtained by:

$$
p_{j,(k)} = a \left[ \frac{1}{\mathrm{gMean}\left(\vec{\mathbf{c}}_j\right)} 2^{\frac{\alpha_j \cdot \epsilon^*}{l_j}} - \frac{1}{c_{j,(k)}} \right].
$$

[3] The above formulation assumes that for all terminals $j$ all allocated subcarriers of vector $\vec{\mathbf{c}}_j$ are utilized for data transmission. In fact, depending on the channel gains, some subcarriers might be „switched" off as their channel gains are too low. To include these cases, the above formulation would have to be altered, apart from that the algorithm is the same.

**Given**: Power $P_{\max}$
Subcarrier gains $\vec{\mathbf{c}}$
Proportion factors $(\alpha_j)$
Relative error $\eta$

$\epsilon_{\min} \leftarrow$ GETMINALPHA() {set $\delta_j = \Delta_j(1)$, cf. (18)}
$\epsilon_{\max} \leftarrow$ GETMAXALPHA()
$p_{\min} \leftarrow$ GETPOWER($\epsilon_{\min}$) {cf. (16)}
$p_{\max} \leftarrow$ GETPOWER($\epsilon_{\max}$)
**while** $\left(1 - \frac{p_{min}}{P_{\max}} > \eta\right)$ **do**
    $\epsilon_{\min} \leftarrow \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) \frac{P_{\max} - p_{\min}}{p_{\max} - p_{\min}}$
    $p_{\min} \leftarrow$ GETPOWER($\epsilon_{\min}$)
**end**
**return** $\epsilon^* = \epsilon_{\min}$

Algorithm 2: Iterated power loading algorithm.

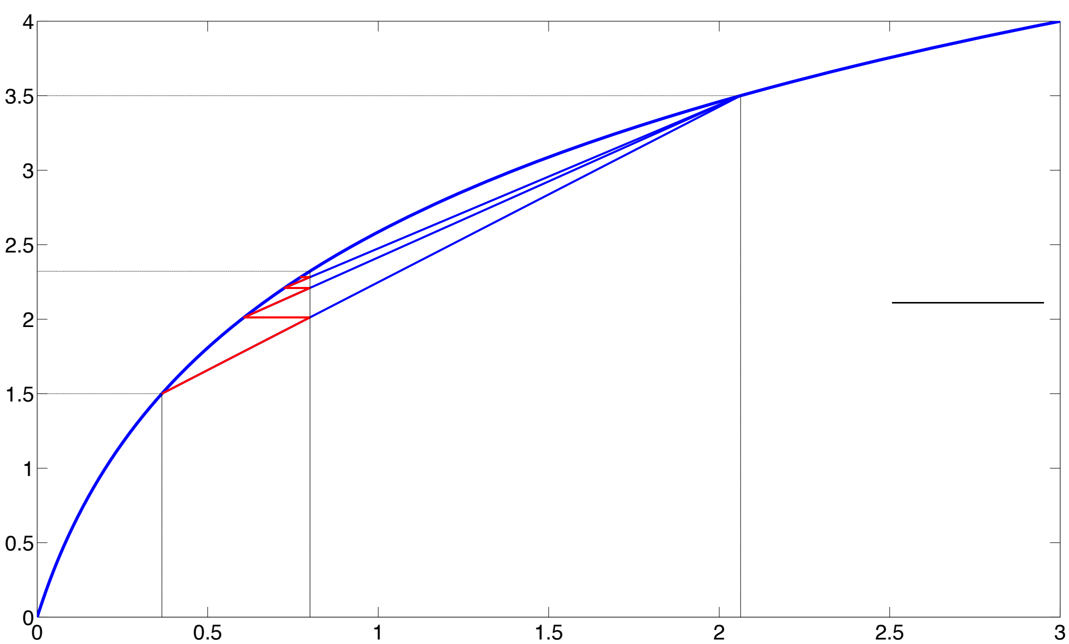


**Figure 4. Visualization of the iterated power loading algorithm.**

### 3.3. Deriving Probabilistic Rate Guarantees

After performing the subcarrier assignment of the first step and the power loading of the second step, we obtain a set of subcarrier gain probability density functions from vector $\vec{\mathbf{f}}$ and a vector of power assignments $\vec{\mathbf{p}} = \left\{p_{j,(k)} \,|\forall j \in J,\ \forall k \in l_j\right\}$. Regardless of the way they have been derived, we can obtain probabilistic expressions for the achievable rate per terminal which ultimately serve as performance estimate of the considered OFDMA system. Recall from Section 2.1 that there exist $M$ different modulation types in the system and for each terminal $j$ there exists a target bit error rate $\beta_j$, resulting in SNR thresholds $\Gamma_{j,m}$. Based on the continuous distribution function for the subcarrier gains, we can derive (approximately) discrete probability mass functions for the (random) amount of data that can be transmitted on an assigned subcarrier per down-link phase. Denote by $z_{j,(k)}$ the random amount of bits that can be transmitted during one down-link phase to terminal $j$ on the $k$-th allocated subcarrier. We can characterize its probability mass function by:

$$P\left[z_{j,(k)} = S \cdot b_m\right] = F_{\tilde{g}_{j,(k)}}\left(\frac{\Gamma_{j,m+1} \cdot \sigma^2}{p_{j,(k)}}\right) - F_{\tilde{g}_{j,(k)}}\left(\frac{\Gamma_{j,m} \cdot \sigma^2}{p_{j,(k)}}\right). \tag{19}$$

Essentially, in this equation the probability is calculated that the k-th allocated subcarrier regarding terminal $j$ can be employed by modulation type $m$ (where $b_m$ is the amount of bits that can be transmitted per symbol by this modulation type). Due to the required bit error probability, this modulation type can only be applied for an SNR larger than $\Gamma_{j,m}$ and lower than $\Gamma_{j,m+1}$. The required, corresponding channel gains for these lower and upper SNR bounds are equal to the expressions in the brackets of the probability function in Equation (19). From Equation (19) we obtain a PMF for $z_{j,(k)}$:

$$p\left(z_{j,(k)}\right) = \left[P\left(z_{j,(k)} = S \cdot b_0\right), \ldots, P\left(z_{j,(k)} = S \cdot b_M\right)\right]. \tag{20}$$

As terminal $j$ receives a total of $l_j$ subcarriers, we proceed with considering the random amount of data that can be transmitted over all these subcarriers, denoted by $Z_j$. It is given by the sum of the single subcarrier rate random variables $z_{j,(k)}$.

$$Z_j = \sum_{k=1}^{l_j} z_{j,(k)} \tag{21}$$

Under the assumption of independent random variables $z_{j,(k)}$ we obtain the PMF for $Z_j$ by the convolution of the PMFs for all $z_{j,(k)}$ – denoting the convolution operator by $\bigodot$. Note that in fact the random variables $z_{j,(k)}$ are not independent as they are obtained from a selection order which imposes a dependency. However, we still treat the random variables $z_{j,(k)}$ to be independent at the cost of obtaining only an approximation for the probability mass function of the overall amount of data transmitted in one down-link phase per terminal:

$$p\left(Z_j\right) = \bigodot_{k=1}^{l_j} p\left(z_{j,(k)}\right). \tag{22}$$

Finally, we remark that the same framework (i.e., Equations (21) and (22)) can also be used to determine the rate PMF in case that no time-varying subcarrier assignments are performed (i.e., a static or channel-independent subcarrier assignment where we still adapt the modulation type with or without dynamic power allocation). To see this, let us again assume a certain set of subcarrier allocations $\{l_j\}$. As channel-dependent subcarrier assignment is not performed, subcarrier allocations are either done contiguously or interleaved. In either case, we can obtain a PMF for the data amount that can be transmitted on a single subcarrier from the fact that the channel gains are exponentially distributed (ignoring in the contiguous case the correlation in frequency). Thus, we obtain the rate PMF per subcarrier from Equation (19) but using the exponential distribution function $F_{g_j}$ from Equation (6).

## 4. Improving Prediction Quality

In this section the basic principles of the evaluation steps described in Section 3 are extended to improve the prediction quality. There is such a potential as the first and second step (subcarrier assignment and power loading) are based on approximations and are therefore bound to be suboptimal. Options of improvement lie within the subcarrier allocation $l_j$, which particularly leads to different subcarrier gain vectors $\vec{\mathbf{c}}_j$, and in the variation of the proportion factors $\alpha_j$ during the power loading step. In the following we present for both steps ways to iteratively improve the prediction quality. These steps are based on the continuous approximation of the power-rate function given in Equation (15).

### 4.1. Subcarrier Allocation

In this section a full description of the subcarrier allocation is given, as applied below in the numerical evaluation. Our approach is a revision of the algorithm BABS (Bandwidth Assignment Based on SNR) presented in [20] which determines the amount of allocated subcarriers $l_j$ for each terminal $j$. In BABS for each terminal all their subcarrier gains are assumed to be equal. Initially, each terminal gets one subcarrier. The rest of the subcarriers are greedily assigned to the terminals step by step. In each allocation step an additional subcarrier is given to the terminal for which the power consumption of transmitting data with a rate of $\sigma_j/T_{\text{dl}}$ is reduced most by utilizing one more subcarrier. This is formalized in Algorithm 3.

**Given**: Average gains $(\rho_j)$
Rate requirements $(\sigma_j/T_{\text{dl}})$
Rate-power functions $\left(\psi_j^{-1}\right)$
$(l_1, \ldots, l_J) \leftarrow (1, \ldots, 1)$
**while** $\sum_j l_j < N$ **do**
  $i \leftarrow \operatorname{argmin}_j \frac{l_j+1}{\rho_j} \psi_j^{-1}\left(\frac{\sigma_j}{T_{\text{dl}} \cdot (l_j+1)}\right) - \frac{l_j}{\rho_j} \psi_j^{-1}\left(\frac{\sigma_j}{T_{\text{dl}} \cdot l_j}\right)$
  $l_i \leftarrow l_i + 1$
**end**
**return** $(l_j)$

Algorithm 3: BABS, see [20].

Given a subcarrier allocation obtained by Algorithm 3, Algorithm 1 could then be applied yielding the vector of subcarrier gain PDFs. However, the allocation of subcarriers and deriving the corresponding vector of channel gain PDFs can be significantly accelerated by combining Algorithm 3 and Algorithm 1. Additionally, it can be considerably improved by replacing the average channel gains $\rho_j$ (which is constant for each subcarrier) with the normalized average channel gains $c_j$ which are calculated by means of the order statistic choosing the best out of the remaining subcarriers and calculating the resulting expected channel gain divided by the noise variance (which varies), cf. Equation (11). The resulting allocation algorithm is given below as Algorithm 4.

**Given**: Average gains $(\rho_j)$
Rate requirements $(\sigma_j/T_{\text{dl}})$
Rate-power functions $\left(\psi_j^{-1}\right)$
$(l_1, \ldots, l_J) \leftarrow (1, \ldots, 1)$
$A_{j,(1)} \leftarrow N + 1 - j \quad \forall\, j \in J$
$M \leftarrow N - J$ {Current number of available subcarriers}
$h_j \leftarrow 1/\textsc{GetNormAvgCG}(A_{j,(1)}, \rho_j)\ \forall\, j \in J$, see (11)
**while** $\sum_j l_j < N$ **do**
  $c_j \leftarrow \textsc{GetNormAvgCG}(M, \rho_j)\ \forall\, j \in J$, see (11)
  $i \leftarrow \operatorname{argmin}_j \left(h_j + \frac{1}{c_j}\right) \cdot \psi_j^{-1}\left(\frac{\sigma_j}{T_{\text{dl}} \cdot (l_j+1)}\right) - h_j \cdot \psi_j^{-1}\left(\frac{\sigma_j}{T_{\text{dl}} \cdot l_j}\right)$
  $l_i \leftarrow l_i + 1$
  $A_{i,(l_i)} \leftarrow M$
  $M \leftarrow M - 1$
  $h_i \leftarrow h_i + \frac{1}{c_i}$
**end**
**return** $((l_j), (A_{j,(k)}))$

Algorithm 4: Subcarrier allocation algorithm.

In the initialization part every terminal gets one subcarrier starting with the first terminal, such that in the beginning of the while-loop $M = N - J$ subcarriers are still available. In practice the assignment is ordered by the terminals' average gains, starting with the worst one. Then the auxiliary variables $h_j$ are initialized with the reciprocal of the normalized average channel gains. In the while-loop the current normalized average channel gains $c_j$ are calculated for $M$ remaining subcarriers. After that, like in BABS, the terminal which is reducing the resulting transmit power most will be allocated another subcarrier. Note that all the so far achieved normalized channel gains are considered as $h_j$ denotes the sum of their reciprocals. After updating the parameters of terminal $i$ this procedure is repeated until every subcarrier is allocated. The amount of subcarriers per terminal $l_j$ and the amount of remaining subcarriers while choosing the $k$-th one $A_{j,(k)}$ are returned for all terminals and their subcarriers.

### 4.2. Parameter Optimization

During the power loading process, the only parameters that directly influence the algorithm are the proportion factors $\alpha_j$. The adaptation of the proportion factors is performed in such a way that first the sum of the proportion factors stays constant and second the maximal outage of a terminal is reduced. The constant sum of proportion factors enables an easy comparison of the results. The goal of the adaptation is to equalize the outages, i.e., the minimal and maximal outage are the same, which is always obtainable. In practice the adaptation is stopped additionally, if the desired outage either is reached or cannot be achieved. The outage can be calculated from Equation (21) as:

$$o_j = P\left(Z_j < \alpha_j \cdot \epsilon\right). \tag{23}$$

Updating the outages includes, beside Equation (23), the power loading of Section 3.2 which delivers in particular the update of $\epsilon$ and the derivation of the probabilistic rate guarantees discussed in Section 3.3. The outline of this procedure is presented in Algorithm 5.

**Given**: Proportion factors $(\alpha_j)$
$\epsilon$
Number of subcarriers $(l_j)$
Remaining subcarriers $(A_{j,(k)})$
Outages $(o_j)$
**while** $\min_j\{o_j\} \neq \max_j\{o_j\}$ **do**
  $(\alpha_j) \leftarrow$ ADAPTPROPORTIONS$((\alpha_j), (o_j))$
  $(o_j, \epsilon) \leftarrow$ UPDATEOUTAGES$((\alpha_j), (l_j), (A_{j,(k)}), \epsilon)$
**end**
**return** $(\alpha_j)$

Algorithm 5: Parameter optimization.

## 5. Numerical Evaluation

In this section we initially investigate the quality of our performance prediction in comparison to the optimal solution of problem (4). After performing this evaluation, we draw our attention to the problem of admission control in the downlink of an OFDMA (WiMAX-like) system. We consider here two different applications types: Voice-over-IP streams and video streams. In both cases we are interested in the maximum number of flows that can be served due to flow-specific quality-of-service constraints.

### 5.1. Basic Evaluation Scenario

We consider data transmissions in IEEE 802.16e-2005 [1] systems as basic scenario. This standard defines a set of OFDMA systems with multiple „allocation modes". We consider here a total system bandwidth of $B = 10$ MHz. Each subcarrier has a bandwidth of 11.16 kHz. Hence, the system bandwidth is split into $N = 865$ net subcarriers (the total number equals 1024 but 159 of them are used as guard bands) out of which 96 are used as pilots. The symbol time results to $T_s = 89.6$ $\mu$s while the considered setting for the guard time equals $T_g = 11.2$ $\mu$s, yielding a total length of an OFDM symbol of 100.8 $\mu$s. We focus explicitly on the AMC (adaptive modulation and coding) transmission mode, therefore the net subcarriers are divided into bands of 36 subcarriers each which are again subdivided into bins of 9 subcarriers (one pilot and 8 data subcarriers). There are 24 bands and 96 bins. A bin is the smallest share of bandwidth that can be allocated separately. In the following we assume that a bin is assigned to a certain terminal throughout the whole down-link phase (in fact, the standard allows for time sharing of bins, where the bin can be reassigned after 6 OFDM symbols). Furthermore, we assume a frame length of 5 ms out of which $T_{dl} = 2.5$ ms are available for the down-link transmission (assuming TDD). This results to $S = 24$ symbols available per bin per down-link phase. From these 24 symbols we assume 20 to be available for data transmission (the remaining ones reserved for control signaling). The adaptive modulation system of 802.16e features BPSK, QPSK, 16-QAM and 64-QAM. A maximum transmit power of $P_{max} = 40$ W is available for the considered bandwidth.

Terminals in the cell are located with an equal distance to the base station. Hence, for all considered scenarios we have the same basic channel gain $\rho_j$ for all terminals (for different scenarios we vary the distances between base station and terminals resulting in a different average channel gain). Our considered path loss model corresponds to the UTRA/EUTRA simulation case 1 with an path loss exponent of $\alpha = 3.7$ and a reference propagation loss (over one meter) of $K = 35.3$ dB. We do not take a shadowing model into account. Fading components are generated due to an exponential distribution. We do not consider correlation in time or in frequency of the channel gains. Also, terminals are assumed to be stationary.

As traffic we assume two different types of applications. In the first case we consider the transmission of Voice-over-IP flows encoded according to the G.711 standard. Such streams generate a packet every $d = 20$ ms of size 80 Byte (resulting in an application layer rate of 32 kBit/s). Adding to this RTP, UDP and IP packet overheads yields a final size of $\sigma = 120$ Byte. We assume them to be encoded according to a rate 3/4 convolutional encoder. Finally, 802.16e adds a MAC overhead of $O = 10$ Byte (6 Byte header plus a CRC32), which yields a total MAC packet size of 170 Byte (which equals a required physical layer rate of 68 kBit/s). Packet transmissions are subject to a packet error rate of $p_j^{\text{Err}} = 0.01$. Due to the usage of a 3/4 convolutional code this yields a target bit error rate of $\beta_j = 0.0052$ (following the derivations in [23] and assuming hard-decision Viterbi decoding with independent errors). The second type of application that we consider is video streaming. In this case we consider constant bit rate streams with a payload packet size of 455 Byte while the packet interarrival time is set to $d = 20$ ms (which yields an application layer rate of 180 kBit/s). Assuming the same overhead by RTP, UDP and IP as well as an convolutional coder of rate 1/2, this yields (together withe the MAC header of 802.16e) an overall MAC packet size of 1000 Byte.

## 5.2. Methodology

Initially, we are interested in the quality of our derived performance prediction framework for OFDMA systems. As metric we consider the average minimum rate per terminal per down-link phase. This metric is strongly correlated with the optimal solution of the optimization problem (4) and therefore serves as predictor for $\epsilon$ (assuming that $\alpha_j = 1$ holds in optimization problem 4 for all terminals). Therefore we use these as synonyms throughout this section. Apart from that we also consider the outage probability of a VoIP packet (of size 170 Byte) as metric. The following schemes are compared with each other regarding these two metrics.

- The optimal solution to problem (4), setting $\alpha_j = 1$ for all terminals $j$. In order to determine it, we generate channel gains for a large number of down-link phases and formulate the corresponding integer programming problem. This is then passed to the IP solver CPLEX [24]. The resulting optimal assignments are further processed for statistical purposes. Hence, the optimal solution to problem (4) is determined by stochastic simulations. From these simulations we obtain the average down-link rate as well as the outage probability per terminal. We refer to this in the following as the **IP Optimum DynPow** indicating a system with dynamic power assignments. Notice that very close performance to the IP optimum can be achieved for example by applying linear relaxation techniques to simplify the problem [13].

- The analytical lower bound for channel-dependent OFDMA assignment as derived in Section 3. Hence, for the bound we evaluate the minimal average down-link rate:

$$\epsilon = \min_j E[Z_j]$$

  and outage probability $o_j$ by analysis, see Equations (21) and (23). We refer to this as the **Lower Bound DynPow** in the following. The evaluation is accomplished using Matlab.

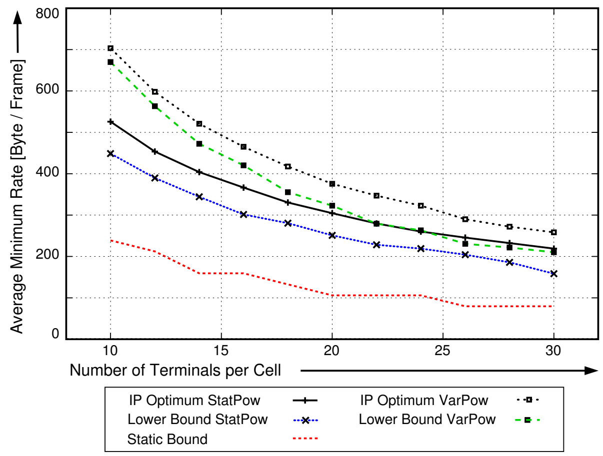


**Figure 5. Minimum average rate per frame per terminal for an increasing number of terminals (average SNR of $10$ dB).**

- In our previous work [9] we have considered an OFDMA system with dynamic subcarrier assignments but *static* power assignments. For comparison purposes we include the performance predictions for this type of system as well. We refer to the analytical estimate as **Lower Bound StatPow** and the corresponding simulated optimal system performance as **IP Optimum StatPow**. Details on the derivations of this specific type of problem can be found in [9]. Notice that these derivations depend on a slightly different assignment algorithm in comparison to the one presented in Section 3.1.

- Finally, the above framework can also be used to predict the performance if no channel-dependent subcarrier and power assignments are performed. Notice that we still assume the usage of adaptive modulation as channel-dependent means. But terminals receive a certain block of subcarriers and reuse them throughout all down-link phases. Hence, the performance metrics of this scheme are evaluated by analysis. We refer to this scheme as **Static Bound**. The static approach essentially resembles what has been suggested for performance prediction and admission control in previous work [6–8].

## 5.3. Bound Evaluation

Regarding the above metrics we have evaluated the five different schemes in a setting where the number of terminals in the cell increases from initially 10 terminals up to 30 terminals and all terminals have a fixed distance of about 570 meters to the base station. In Figure 5 we present the resulting minimum average rate $\bar{\epsilon}$ per down-link phase. From this figure we see that the lower bound approximates the performance of the IP optimum with a gap equal to about 80 Byte per frame in case of dynamic subcarrier assignments but static power assignment. If dynamic power assignment is featured by the system, the gap between the optimal system performance and the estimate varies. For a low number of terminals in the cell, the gap is small while for a larger number of terminals in the system the gap is larger. As mentioned, this behavior is not present for the system with dynamic subcarrier assignments and static power assignment. In general, the prediction of the optimal system performance is very difficult due to the NP-hardness of the corresponding optimization problem. Recall that some form of performance prediction is required in such OFDMA for admission control, for example. The only alternative to using the lower bound for the channel-dependent OFDMA resource assignment is to base the analysis on the average channel gains $\rho_j$ which leads to the estimate of $\epsilon$ represented by the static approach. Taking these two arguments into account, the proposed lower bound for channel-dependent OFDMA resource assignment is a significant improvement. Correspondingly, in Figure 6 we observe the outage probability of the five different schemes (now considering up to 40 terminals in the cell). Essentially, the same statements can be deduced from this graph. The lower bound provides a much better estimate of the outage probability than what is provided by the static approach (not taking adaptive subcarrier assignments into account). However, there remains some gap between the lower bound and the IP optimum in both cases of power assignments.

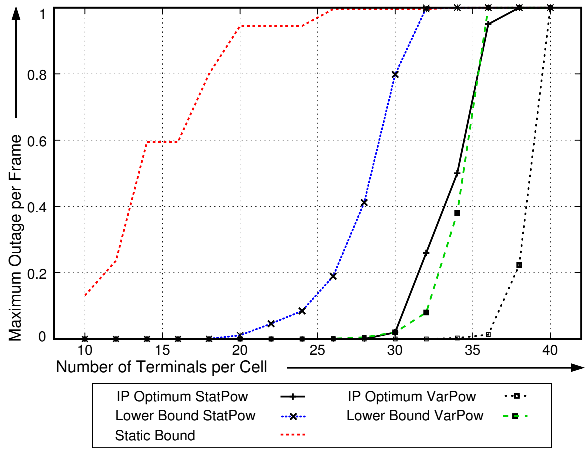


**Figure 6. Maximum outage probability per frame for a packet of size $170$ Byte over an increasing number of terminals (average SNR of $10$ dB).**

## 5.4. Application to Admission Control

Next, we investigate the application of the analytical framework to admission control. We illustrate this by considering the down-link transmission of VoIP streams (parameters as described in Section 5.1) initially. Each VoIP stream has a (constant) bit rate requirement on the physical layer (including all packet overheads) of 68 kBit/s. However, the stream is divided into packets with an (average) interarrival time of 20 ms. We assume in the following that packets are transmitted in one piece during a single down-link phase. Hence, the question comes up how many such packets can be transmitted in a single down-link frame (each one to a different terminal). Once we have determined this number, the total down-link VoIP capacity of the system is simply this number times four, as we consider a frame time of 5 ms (and packet interarrival time is 20 ms).

As quality-of-service requirement we consider two different regimes. In the first case we simply assume that every stream requires its packets to be transmitted successfully *on average*. That is, we can admit as many VoIP streams to a single down-link phase as we have for all admitted flows:

$$E\left[Z_j\right] \geq 170 \text{ Byte}. \tag{24}$$

We refer to this admission control scheme as „average rate QoS criterion“. In this case we do not care about outage probabilities. In contrast, the second admission control scheme requires for each VoIP packet transmission a maximum outage probability of 0.05, i.e., we admit as many streams to each down-link phase of the system as for each of them we have:

$$P\left[Z_j \leq 170 \text{ Byte}\right] \leq 0.05. \tag{25}$$

In Figure 7 we show the corresponding VoIP capacity when considering the average rate QoS criterion. The figure shows the maximum number of VoIP packets that can be supported by a single down-link phase versus an increasing average SNR of all terminals[4]. We show the resulting admission capacity of the static bound, the two lower bounds and the two (simulated) optimal capacities of the system. Clearly, the static approach is limited by the fact that it „ignores“ multi-user diversity. Compared to the IP optimum with dynamic subcarrier assignments but static power assignment, admission control by such a scheme misses capacity by about 50 to 75%. In contrast, the lower bound for dynamic subcarrier but static power assignments achieves a much better capacity estimate still with some gap in comparison to the optimum one. If the system also features dynamic power assignments, the resulting system performance is larger. Furthermore, we observe a less volatile performance prediction in case of the system with dynamic power assignment in comparison to the system with static power assignments (where the prediction gap is small for small SNR but rather large for large SNR).

[4] This corresponds to a decreasing homogeneous distance of all terminals to the base station. The corresponding SNR value refers to the average that a system with fixed power assignments - splitting the overall transmit power evenly on all subcarriers - achieves.

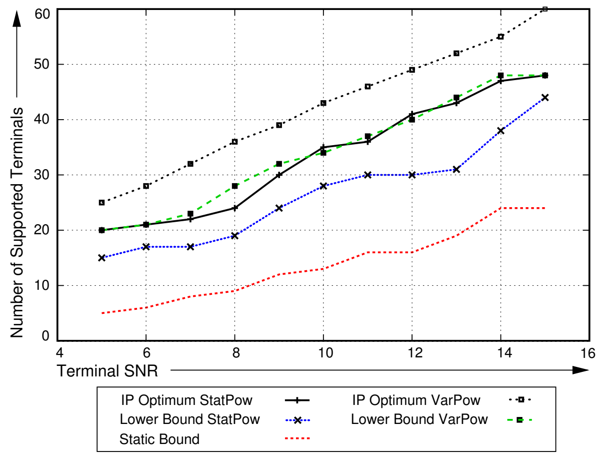


**Figure 7. Maximum number of VoIP packets that can be supported by a single down-link frame according to an average rate QoS criterion over an increasing average SNR per terminal.**

Next, we consider the resulting VoIP capacity if the outage admission regime is considered. Here the full strength of the analytical framework can be observed as outage behavior has rarely been investigated in the literature. Figure 8 shows the corresponding VoIP capacity of a single down-link phase in this case again for an increasing average SNR per terminal. In contrast to Figure 7 we observe that the lower bound provides now a slightly better estimate of the IP optimum than in case of the average QoS criterion, at least for the system with static power assignments. In case of dynamic power assignments, predicting the system performance is more difficult and results in a larger gap between the optimum and the predicted system performance.

To reveal the origin of the additional gap between the optimal system performance and the predicted one we show in Figure 9 plots of the channel gain probability density functions as predicted by our analytical method (see Section 3) and two observed histograms of the channel gains as chosen by the IP solver CPLEX in case of simulating the optimal system performance. In particular we consider a system with 30 terminals and an average SNR per subcarrier of 10 dB. For this case, we first discuss the analytical channel gains predicted by our framework. In Figure 9 we show the corresponding PDFs for terminal 1 and terminal 24. Due to the selection order the resulting PDFs differ obviously. In this case, terminal 24 has a less suitable PDF than terminal 1. Both PDFs though feature a mean which is around 3. Turning our attention to the observed channel gains as chosen by CPLEX, we present in Figure 9 the corresponding histogram for terminal 1 as well as the histogram over all terminals. Except for statistical variances due to a lower sample size in case of terminal 1, both histograms have almost the same features (i.e., mean and range). However, the histograms indicate a more suitable distribution of the gains in case of the optimal solution: Less mass is in the lower channel gains regions while more mass is in higher gain regions resulting in a larger mean (4 compared to 3). This leads to the better system performance in general. It reflects the fact that the optimal assignment scheme exploits multi-user diversity in a better way (algorithmically this means, for example, that multiple subcarrier states for different terminals are considered jointly before deciding about a subcarrier assignment instead of greedily assigning the currently best subcarrier to a terminal as in our assignment algorithm).

Finally, we consider in Figure 10 the performance prediction results for much larger packet sizes, as encountered usually in video streams. Here, we still apply an outage regime but with a larger maximum outage probability of 0.1. Larger packet sizes obviously require more resources to be assigned per stream/packet. This has a significant effect on the performance prediction results in comparison to the packet sizes considered for VoIP. For the video case we observe now a much better quality of the performance prediction in both system settings, assuming a static as well as a dynamic power assignment. This is present over the entire considered SNR range, only for very large SNRs occurs a larger performance gap in case of the system with dynamic power assignment. We suspect that this is due to a lower impact of multi-user diversity on the optimal system performance.

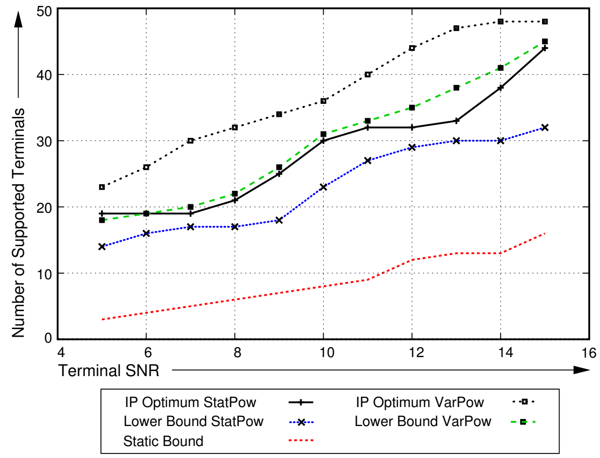


**Figure 8. Maximum number of VoIP packets that can be supported by a single down-link frame according to an outage-based QoS criterion over an increasing average SNR per terminal (maximum outage probability of $0.05$ is assumed).**

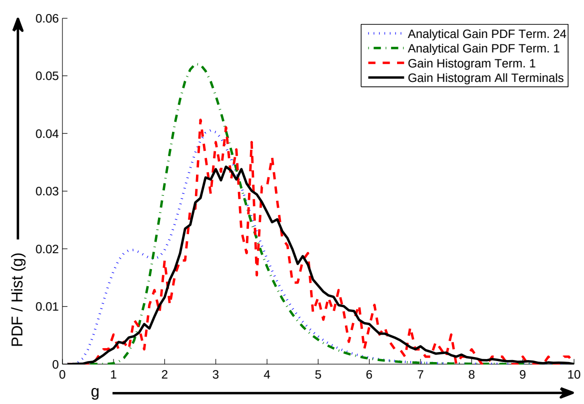


**Figure 9. Analytically derived gain PDFs for selected terminals versus observed gain histograms for the optimal solution (considering a cell setup with $30$ terminals and an average SNR of $10$ dB).**

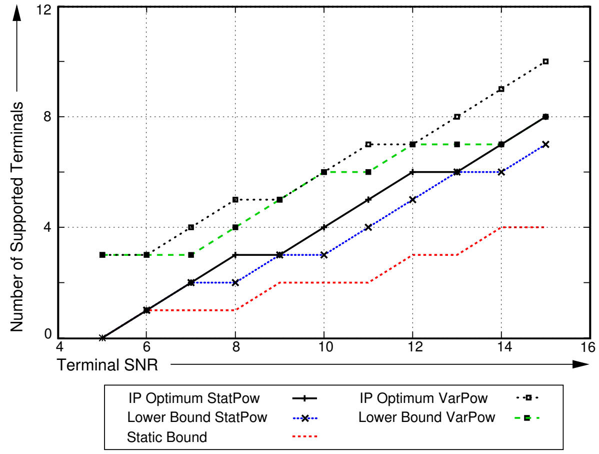


**Figure 10. Maximum number of video packets that can be supported by a single down-link frame according to an outage-based QoS criterion over an increasing average SNR per terminal (maximum outage probability of $0.1$ is assumed).**

## 6. Related Work

The issue of channel-dependent OFDMA resource assignment (in the down-link) is well investigated where typical studies [4, 13, 20] focus on suboptimal algorithms for either minimizing the transmit power or maximizing the rate per terminal. The concept of proportional fairness is used to share the medium between different classes which in addition reduces the complexity. Solutions to such problems are discussed in [10, 17–19]. Numerical evaluation, based on the simulation of numerous down-link phases, show that on average channel-dependent schemes outperform static ones significantly. Neither do these contributions address the problem of admission control nor do they provide an analytical framework for expected performance of the proposed algorithms.

Admission control for OFDMA networks is addressed by [6–8]. In [6] the authors investigate Poisson arrivals to a base station with finite buffers. Based on a queuing model, blocking probabilities are derived which an admission control scheme can base its operation on. However, the authors lack an analytical framework for the expected rate obtained from channel-dependent OFDMA resource assignments. Instead, they base their proposed admission control scheme on the average channel gain, i.e., average subcarrier rate. This clearly underestimates the true system capacity as shown above. This shortcoming applies also to [7, 8] where different admission control strategies are studied for two distinct systems and scenarios, still the achievable rate per subcarrier/subchannel is always derived from the average channel gain, underestimating the system's capacity significantly. In [8] the authors investigate admission control schemes for OFDMA cells under the power minimization regime. Per call, the required resources are planned according to the average channel gain, not taking multi-user diversity into account. Instead, the authors focus on the effect of mobility of terminals receiving multimedia streams for which some system resources have to be reserved for hand-over. In contrast to these three investigations, [25] studies a queuing analysis framework for static and channel-dependent OFDMA resource assignments based on weighted fair queuing. While the provided queuing-theoretic framework is quite deep in general, the authors do not determine analytical expressions for the instantaneous terminal rate obtained from channel-dependent OFDMA resource assignments. In order to still perform a queuing-theoretic analysis, they obtain average terminal rates (under channel-dependent OFDMA resource assignments) from simulations. While for the specific considered case this serves very well, it is certainly not practical to simulate each and every possible situation in a wireless OFDMA cell that might have impact on the average terminal rates. Furthermore, no outage based admission control can be performed from this approach which is another significant disadvantage. A recent paper [26] addresses admission control in OFDMA system with dynamic resource allocation. Here, the authors focus on determining the amount of terminals that can be served during a single down-link phase given the instanteous subcarrier gain matrix $\mathbf{G}$. However, the work does not provide a longer-term prediction of the system capacity as presented in our work.

## 7. Conclusions

In this paper we derived an analytical framework for performance prediction in OFDMA systems with dynamic resource allocation. The framework suites two particular approaches: a fixed power assignment per subcarrier and a dynamic power assignment (where the later system setup features a larger performance improvement but also a larger complexity of the resource allocation problem). Using this framework, it is possible to predict the transmission success of scheduling a specific set of packets, i.e., the probability that the resource allocation unit will be able to transmit the entire set of packets completely. Our numerical evaluations show that the gap between the optimal system performance and the predicted one is only about $10 - 20\%$ which is therefore a significant improvement compared to state-of-the art. Furthermore, the framework provides a key component to admission control in such systems as it enables to predict the dependency between system performance and system load (in terms of packet sizes, overall terminals, and average channel gains) accurately with respect to many different quality-of-service criteria.

## References


[1] IEEE, IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems (February 2006).

[2] E. Dahlmann, S. Parkvall, J. Skld, P. Beming, 3G Evolution: HSPA and LTE for Mobile Broadband, Academic Press, 2008.

[3] H. Liu, G. Li, OFDM-Based Broadband Wireless Networks, John Wiley & Sons, Inc., 2005.

[4] C. Wong, R. Cheng, K. Letaief, R. Murch, Multiuser OFDM with adaptive subcarrier, bit and power allocation, IEEE J. Select. Areas Commun. 17 (10) (1999) 1747–1758.

[5] M. Bohge, J. Gross, M. Meyer, A. Wolisz, Dynamic resource allocation in OFDM systems: An overview of cross-layer optimization principles and techniques, IEEE Network 21 (1) (2007) 53–59.

[6] G. Li, H. Liu, Dynamic resource allocation with finite buffer constraint in broadband OFDMA networks, in: Proc. of IEEE Wireless Communications and Networking Conference (WCNC), 2003.

[7] K. Lee, V. Leung, Capacity planning for group-mobility users in OFDMA wireless networks, EURASIP Journal on Wireless Communication and Networking (2006) 1–12.

[8] C. Qin, G. Yu, Z. Zhang, H. Jia, A. Huang, Power reservation-based admission control scheme for IEEE 802.16e OFDMA systems, in: Proc. of IEEE Wireless Communications and Networking Conference (WCNC), 2007.

[9] J. Gross, Admission control based on OFDMA channel transformations, in: Proc. of 10th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2009.

[10] R. Mathar, M. Reyer, Iterated water-filling for OFDMA rate and power allocation with proportionality constraints, in: Proc. IEEE Conference on Global Communications (GLOBECOM), Washington D.C., USA, 2007, pp. 3230–3234. doi:10.1109/GLOCOM.2007.612.

[11] K. Cho, D. Yoon, On the general BER expressions of one- and two-dimensional amplitude modulations, IEEE Trans. Commun. 50 (7) (2002) 1074–1080.

[12] W. Rhee, J. Cioffi, Increase in capacity of multiuser OFDM system using dynamic subchannel allocation, in: Proc. Vehicular Technology Conference (VTC), 2000, pp. 1085–1089.

[13] M. Ergen, S. Coleri, P. Varaiya, QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems, IEEE Trans. Broadcast. 49 (4) (2003) 362–370.

[14] I. Kim, H. Lee, B. Kim, Y. Lee, On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM, in: Proc. of the Global Telecommunications Conference, 2001.

[15] S. Ross, Introduction to Probability Models, Academic Press, 2002.

[16] T. Cover, J. Thomas, Elements of Information Theory, J. Wiley & Sons, Inc., 1991.

[17] Z. Shen, J. G. Andrews, B. L. Evans, Optimal power allocation in multiuser OFDM systems, in: Proc. IEEE Conference on Global Communications (GLOBECOM), San Francisco, California, USA, 2003, pp. 337–341.

[18] I. C. Wong, Z. Shen, B. L. Evans, J. G. Andrews, A low complexity algorithm for proportional resource allocation in OFDMA systems, in: IEEE Workshop on Signal Processing Systems (SIPS), Austin, Texas, USA, 2004, pp. 1–6. doi:10.1109/SIPS.2004.1363015.

[19] Z. Shen, J. G. Andrews, B. L. Evans, Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints, IEEE Trans. Wireless Commun. 4 (6) (2005) 2726–2737.

[20] D. Kivanc, G. Li, H. Liu, Computationally efficient bandwidth allocation and power control for OFDMA, IEEE Trans. Wireless Commun. 2 (6) (2003) 1150–1158.

[21] Z. Mao, X. Wang, J. Lin, Fast optimal radio resource allocation in OFDMA system based on branch-and-bound method, in: Proc. IEEE International Conference on Signal Processing and Communications (ICSPC), Victoria, British Columbia, Canada, 2005, pp. 348–351. doi:10.1109/PACRIM.2005.1517297.

[22] J. Campello, Practical bit loading for DMT, in: Proc. IEEE International Conference on Communications (ICC), Vol. 2, Vancouver, British Columbia, Canada, 1999, pp. 801–805.

[23] M. Pursley, D. Taipale, Error probabilities for spread-spectrum packet radio with convolutional codes and Viterbi decoding, IEEE Trans. Commun. COM-35 (1) (1987) 1–12.

[24] S. ILOG, ILOG CPLEX 9.0 - User's Manual, Paris, France (2004).

[25] D. Niyato, E. Hossain, Adaptive fair subcarrier/rate allocation in multirate OFDMA networks: Radio link level queuing performance analysis, IEEE Trans. Veh. Technol. 55 (6) (2006) 1897–1907.

[26] S. Bashar, Z. Ding, Admission control and resource allocation in a heterogeneous ofdma wireless network, IEEE Trans. Wireless Commun. 8 (8) (2009) 4200–4210.