Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2008, Article ID 437921, 14 pages
doi:10.1155/2008/437921

Research Article

Throughput Maximization under Rate Requirements for the OFDMA Downlink Channel with Limited Feedback

Gerhard Wunder,[1] Chan Zhou,[1] Hajo-Erich Bakker,[2] and Stephen Kaminski[2]

[1] *Fraunhofer German-Sino Lab for Mobile Communications, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Einstein-Ufer 37, 10587 Berlin, Germany*
[2] *Alcatel-Lucent Research & Innovation, Holderaeckerstrasse 35, 70499 Stuttgart, Germany*

Correspondence should be addressed to Gerhard Wunder, wunder@hhi.fhg.de

Received 1 May 2007; Revised 12 July 2007; Accepted 26 August 2007

Recommended by Arne Svensson

The purpose of this paper is to show the potential of UMTS long-term evolution using OFDM modulation by adopting a combined perspective on feedback channel design and resource allocation for OFDMA multiuser downlink channel. First, we provide an efficient feedback scheme that we call *mobility-dependent successive refinement* that enormously reduces the necessary feedback capacity demand. The main idea is not to report the complete frequency response all at once but in subsequent parts. Subsequent parts will be further refined in this process. After a predefined number of time slots, outdated parts are updated depending on the reported mobility class of the users. It is shown that this scheme requires very low feedback capacity and works even within the strict feedback capacity requirements of standard HSDPA. Then, by using this feedback scheme, we present a scheduling strategy which solves a weighted sum rate maximization problem for given rate requirements. This is a discrete optimization problem with nondifferentiable nonconvex objective due to the discrete properties of practical systems. In order to efficiently solve this problem, we present an algorithm which is motivated by a weight matching strategy stemming from a Lagrangian approach. We evaluate this algorithm and show that it outperforms a standard algorithm which is based on the well-known Hungarian algorithm both in achieved throughput, delay, and computational complexity.

1. INTRODUCTION

There are currently significant efforts to enhance the downlink capacity of the universal mobile telecommunications system (UMTS) within the long-term evolution (LTE) group of the 3GPP *evolved UMTS terrestrial radio access network* (E-UTRAN) standardization body. Recent contributions [1–3] show that alternatively using *orthogonal frequency division multiplex* (OFDM) as the downlink air interface yields superior performance and higher implementation-efficiency compared to standard wideband code division multiple access (WCDMA) and is therefore an attractive candidate for the UMTS cellular system. Furthermore, due to fine frequency resolution, OFDM offers flexible resource allocation schemes and the possibility of interference management in a multicell environment [4]. It is therefore self-evident that OFDM will be examined in the context of high-speed downlink packet access (HSDPA) where channel quality information (CQI) reports are used at node B in order to boost link capacity and to support packet-based multimedia services by proper scheduling of available resources. HSDPA employs a combination of time division multiple access (TDMA) and CDMA to enable fast scheduling in time and code domain. Furthermore, fast flexible link adaptation is achieved by adaptive modulation and variable forward error correction (FEC) coding. By contrast, for UMTS LTE a combination of TDMA and *orthogonal frequency division multiple access* (OFDMA) is used and link adaption is performed on subcarrier groups. Additionally, hybrid-ARQ with incremental redundancy transmission will be set up in both systems.

Since HSDPA does not support frequency-selective scheduling, only frequency-nonselective CQI needs to be reported by the user terminal, leading to a very low feedback rate. Obviously, the same channel information can in principle be used for the OFDM air interface taking advantage of the higher spectral efficiency. Moreover, by exploiting

frequency-selective channel information, the OFDM downlink capacity can be further drastically increased. However, in practice, one faces the difficulty that frequency-selective scheduling affords a much higher feedback rate if the feedback scheme is not properly designed which can serve as a severe argument against the use of this system concept. Also the interplay between limited uplink capacity, user mobility, and resource allocation is not regarded widening the gap between theoretical results and practical applications even further.

Additionally, resource allocation (subcarriers, modulation scheme, code rate, power) is completely different to standard HSDPA and more elaborate due to the huge number of degrees of freedom. There is a vast literature on different aspects of this problem. Wong et al. proposed an algorithm to minimize the total transmit power subject to a given set of user data rates [5]. Extensions of this algorithm have been given in [6–8]. The problem of maximizing the minimum of the users' data rate for a fixed transmit power budget has been considered in [8, 9]. Yin and Liu [10] presented an algorithm that maximizes the overall bit rate subject to a total power constraint *and* users' rate constraints. They proposed a subcarrier allocation method based on the so-called Hungarian assignment algorithm, which is optimal under the restriction that the number of subcarriers per user is fixed a priori.

In this paper we follow a somewhat different strand of work: a generic approach to performance optimization is to maximize a weighted sum of rates under a sum power constraint. This approach provides a convenient way to balance priorities of different services and, more general, to incorporate economical objectives in the scheduling policy by prioritizing more important clients [11]. Besides, supposing that the data packets can be stored in buffers awaiting their transmission, it was shown in [12] that the strategy maximizes the stability region if the weights are chosen to be the buffer lengths. Stability is here meant in the sense that all buffers stay finite as long as all bit arrival rate vectors are within the stability region. Moreover, an even further step is the consideration of user specific rate requirements [13]. Indeed, by guaranteeing minimum rates, QoS constraints can be regarded in the optimization model. However, the restriction of exclusive subcarrier allocation within the OFDMA concept complicates the analysis of the optimization problem significantly. Further, only certain rates are achievable, since a finite set of coding and modulation schemes can be used. Then the optimization problem results in a nonconvex problem over discrete sets rendering an optimal solution almost impossible.

Contributions

We consider the OFDMA multiuser downlink channel and provide strategies for feedback channel design and frequency-selective resource allocation. In particular, we show that frequency-selective resource scheduling is critical in terms of feedback capacity and present a design concept taking care of the limited uplink resources of a potential OFDM-based system. Our main idea is not to report the complete frequency response all at once but in parts depending on the mobility class of the users (we call this method *mobility-dependent successive refinement*). Each part reported has a life cycle in which the channel information remains valid apart from an error that can be estimated and considered at the base station. If its life cycle is outdated, the corresponding part has to be updated. Thus after all individual parts were reported, the frequency response is fully available with an inherent additional error that can be calculated for the mobility class.

Then we present a resource allocation scheme which uses an iterative algorithm to solve the weighted sum rate maximization problem for OFDMA, if quantized CQI is available following the above feedback scheme and additional certain rates have to be guaranteed. The algorithm is motivated by a weight-matching strategy stemming from a Lagrangian approach [14]. It can be motivated geometrically as the search for a suitable point on the convex hull of the achievable region. Further it is easy to implement and can be proven to converge very fast. Simulation results show that the scheduler based on this algorithm has excellent throughput performance compared to standard approaches. Finally, we sustain our claims with reference system simulations in terms of delay performance.

Organization

The rest of the paper is organized as follows: in Section 2 we describe the system and resource allocation model. Then, the design of the feedback channel is given in Section 3. In Section 4 we present our scheduling algorithm and the overall performance is evaluated in Section 5. Finally, we draw conclusions on the OFDM system design in Section 6.

2. SYSTEM MODEL

We consider a single-cell OFDM downlink scenario where base station communicates with M user terminals over K orthogonal subcarriers. Denote by $\mathcal{M} := \{1,\dots,M\}$ the set of users in the cell, and by $\mathcal{K} := \{1,\dots,K\}$ the set of available subcarriers. Assuming time-slotted transmission, in each transmit time interval (TTI) the information bits of each user m are mapped to a complex data block according to the selected transport format.[1] Following the OFDMA concept, the complex data of each user m is exclusively asserted to the subcarriers k belonging to a subset $\mathcal{S}_m \subseteq \mathcal{K}$. Clearly, by the OFDMA constraint we have $\mathcal{S}_m \cap \mathcal{S}_{m'} \equiv \varnothing, m \neq m'$. Writing $x_{m,k}$ for the complex data of user m on subcarrier k and neglecting both intersymbol and intercarrier interference, the corresponding received value $y_{m,k}$ is given by

$$y_{m,k} = h'_{m,k} x_{m,k} + n_{m,k}, \quad \forall m,k \in \mathcal{S}_m. \tag{1}$$

[1] While in practical systems the size of the complex data block is restricted which has some impact on the overall performance, here we ignore this impact and assume that the block size can be chosen arbitrarily.

Here, $n_{m,k} \sim \mathcal{N}_{\mathcal{C}}(0,1)$ is the additive white Gaussian noise (AWGN), that is, a circularly symmetric, complex Gaussian random variable, and $h'_{m,k}$ is the complex channel gain given by

$$h'_{m,k} = \sum_{l=1}^{L_m} h_m[l] e^{-2\pi j(l-1)(k-1)/k}, \quad (2)$$

where $h_m[l]$ is the lth tap of the channel impulse response and L_m is the length of channel impulse response of user m, respectively. According to 3GPP, the multipath fading channel can be modeled in three different categories, namely Pedestrian A/B, Vehicular A with a delay spread that is always smaller than the guard time of the OFDM symbol [15]. For example, in this paper frequently used Pedestrian B channel model has 29 taps modeled as random variables (but many with zero variance) such that $h'_{m,k} \sim \mathcal{N}_c(0,1)\ \forall m,k$, at a sampling rate of 7.86 MHz and corresponds to a channel with large frequency dispersion.

In our closed-loop concept, the complex channel gains $h'_{m,k}$ are estimated by the user terminals using reserved pilot subcarriers. Then, a proper CQI value of the estimated channel gains is generated and reported back to the base station through a feedback channel (note that it carries also necessary information for the hybrid-ARQ process used in Section 5). Usually a very low code rate and a small constellation size are used for the feedback channel (e.g., a $(20,5)$ code and BPSK modulation for HSDPA [16]) and it is reasonable to assume that the feedback channel can be considered error free. Finally, the CQI values are taken up by the scheduling entity in the base station that distributes the available resources among the users in terms of subcarrier allocation and adaptive modulation (bitloading).

Let $\Gamma : \mathbb{R}_+^K \to \mathbb{R}_+^K$ be some vector quantizer applied to the channel gains $|h'_{m,1}|, \ldots, |h'_{m,k}|, \forall m$. Denote the outcome of this mapping by $h_{m,1}, \ldots, h_{m,K}, \forall m$, which are equal to the reported channel gains due to the error free feedback channel. Then, given the power budget p_k on subcarrier k, the rate $r_{m,k}$ of user m on subcarrier k within the TTI can be calculated as

$$r_{m,k}(p_k, h_{m,k}) = N_s \cdot C_r(p_k, h_{m,k}) \cdot r_{\mathrm{mod}}(p_k, h_{m,k}) \quad (3)$$

if the subcarrier k is assigned to user m in this TTI. The number of OFDM symbols is given by $N_s \geq 1$ and we implicitly assumed that the channel is approximately constant over one TTI. The mapping $C_r(p_k, h_{m,k})$ is the asserted code rate and $r_{\mathrm{mod}}(p_k, h_{m,k})$ denotes the number of bits of the selected modulation scheme. Both terms depend on the channel state $h_{m,k}$ and the allocated power p_k. In order to determine an appropriate modulation scheme for given channel conditions, we used extensive link-level simulations to obtain the relationship between bit-error rate (BER) and signal-to-noise ratio (SNR $\hat{=}\ p_k h_{m,k}$) for the channels [17]. It turned out that in the low to medium mobility scenario (Pedestrian A/B, 3 km/h, and Vehicular A, 30 km/h), the required SNR levels are almost indistinguishable. Some of the SNR levels are given in Table 1 (low to medium mobility scenario) and Table 2 (high mobility scenario). In the following, all the reported channel gains and powers are arranged in vectors $\mathbf{h} \in \mathbb{R}_+^{MK}$ and $\mathbf{p} \in \mathbb{R}_+^K$, respectively.

Table 1: Required SNR Levels for 3GPP Pedestrian A/B, 3 km/h, and Vehicular A, 30 km/h, channel for given BER constraint.

BER	QPSK[db]	16 QAM[dB]	64 QAM[dB]
10^{-3}	9.8	16.6	22.7
10^{-5}	13.6	19.8	25.6

Table 2: Required SNR Levels for 3GPP Vehicular A, 120 km/h, channel for given BER constraint.

BER	QPSK[db]	16 QAM[dB]	64 QAM[dB]
10^{-3}	10.6	17.8	24
10^{-5}	13.6	21.5	27.9

Note that, since the selected transport format varies over the slots, control information has to be transmitted in parallel to users' data in the downlink channel containing user identifiers, the used coding and modulation scheme, and the overall subcarrier assignment. Note that there are several tradeoffs involved: while a smaller granularity in the downlink channel allows for more flexible scheduling strategies, it increases the amount of the necessary control information and, hence, decreases the available capacity for the user data. Furthermore, a large number of simultaneously supported users might yield a higher multiuser gain which in turn again affects the effective downlink capacity though.

3. FEEDBACK CHANNEL DESIGN

3.1. General concept

For feedback channel design in the frequency-selective case we introduce two fundamental principles: *mobility report and successive refinement of user-dependent frequency response.* Both principles are driven by the observation that complete channel information is not available at a time but if the channel is stationary enough, information can be gathered in a certain manner. By contrast, if the channel variations are too rapid, finer resolution of the frequency response cannot be obtained. Hence, throughput of a frequency-selective system distinctly decreases with the delay of feedback information. Figure 1 shows a sketch of the throughput decline related to the delay of feedback information, where the feedback rate is assumed to be unlimited. It can be observed that the stationary channels (Pedestrian A/B) provide much longer lifetime of feedback information. Hence, appealing to these principles, feedback channel information consists of two sections. The information in the first section describes the mobility class of users where mobility class is defined as the set of similar conditions of the variation of the frequency response. The information in the second section is a channel indicator. If mobility is high, no frequency-selective scheme will be used for this user and only a frequency-nonselective CQI will be reported as, for example, in HSDPA. On the other hand, if

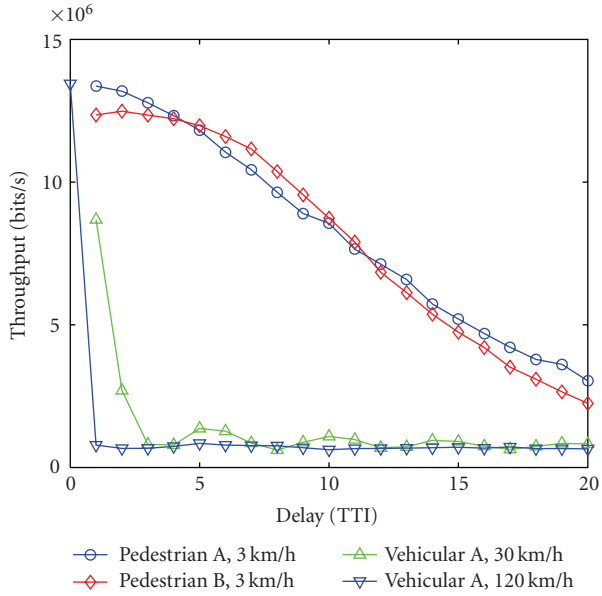

Figure 1: Throughput decline with respect to feedback delay (averaged transmit SNR equals 12 dB, perfect channel knowledge at transmitter and receiver, 5 users are simultaneously supported, code rate = 2/3). It is important to note that an inherent delay of 4 TTI (caused by the signal processing) is already considered in the simulation.

mobility is low, user proceeds in a different but predefined way as described next.

User report the channel gain as follows: the subcarriers are bundled together into groups. In the first TTI, the channel gains are reported in low resolution. In the next time slots, the subcarrier-groups with higher channel gain are further split into smaller groups and reported again so that base station has a finer resolution of the channel and so on. Due to mobility, the channel gain information of a group must be updated in a certain period of time dependent on the coherence time of the channel. Hence, if group information is outdated, the group information will be reported again limiting the maximum refinement. This process then repeats itself up to a predefined number of time slots (so-called restart period) when the frequency response will have significantly changed. The basic approach is depicted in Figure 2 where the scheme is tailored to the feedback channel used in HSDPA namely using effectively 5 bits.

3.2. Performance analysis

Suppose that the scheme is applied to independent channel realizations, then the following is true.

Theorem 3.1. *The feedback scheme is throughput optimal for large number of users, in the sense that the scheme achieves the same throughput up to a very small constant given by* (8)–(10) *compared to any other scheme using the same constellations per subcarrier but reports the channel gains for all subcarriers.*

Proof. First observe that with high probability, the event

$$\mathcal{A} := \Big\{ \log M + c_0 \log\log M > \max_{m\in\mathcal{M}} |h'_{m,k}|^2 > \log M - c_1 \log\log M,\ \forall k \Big\} \tag{4}$$

occurs where $c_0, c_1 > 0$ are real constants. It is worth mentioning that this result not only holds for Rayleigh fading but for a large class of fading distributions under very weak assumptions on the characteristic functions of the random taps [18]. Here, without loss of generality, we restrict our attention to Rayleigh fading, that is, $h'_{m,k} \sim \mathcal{N}_c(0,1)$. Then the probability of the event $\mathcal{A}$ can be lower bounded by [18]

$$\Pr(\mathcal{A}) \geq 1 - \frac{K}{\log M} \tag{5}$$

for large M, and, hence $\Pr(\mathcal{A}) \to 1$ as $M \to \infty$. We have now to establish that the maximum squared channel gain is tightly enclosed by (4) and is delivered by our feedback scheme up to a small constant so that the maximum throughput is indeed achieved.

Denote the subset of those users that attain their maximum gain on subcarrier k by $\mathcal{A}_k$ and abbreviate $f(M) := \log M - c_1 \log\log M$. Fix some subcarrier k_0 and consider the inequality

$$\Pr\Big(\max_{m\in\mathcal{M}} |h'_{m,k_0}|^2 \leq f(M)\Big) \leq \Pr\Big(\max_{m\in\mathcal{A}_{k_0}} |h'_{m,k_0}|^2 \leq f(M)\Big). \tag{6}$$

Since the maximum of each user's frequency response is unique (if not by the channel response itself then by the additional noise) and uniformly distributed over the subcarriers, a fixed percentage of the total number of users will belong to $\mathcal{A}_{k_0}$ with high probability for large M since the users provide M independent realizations. Hence the cardinality of $\mathcal{A}_{k_0}$ fulfills $|\mathcal{A}_{k_0}| \approx M/K \to \infty$ as $M \to \infty$. Since the $|h'_{m,k_0}|^2, m \in \mathcal{A}_{k_0}$, are stochastically lower bounded by chi-squared distributed random quantities the asymptotic gain is not affected yielding

$$\Pr\Big(\max_{m\in\mathcal{M}} |h'_{m,k_0}|^2 \leq f(M)\Big) \longrightarrow 0, \quad M \longrightarrow \infty. \tag{7}$$

Since only the minimum within groups is reported by our scheme, the latter argument bears great importance as it allows us to tightly lower bound the minimum within the subcarrier group that contains the maximum (which is by definition of our scheme the finest subcarrier group for each user). Let us analyze the preserved accuracy by calculating the decline within this group. The smallest cardinality is given by

$$N_{\mathrm{gr}} = \frac{N_{\mathrm{total}}}{N_{\mathrm{reports}}} \left(\frac{N_{\mathrm{refine}}}{N_{\mathrm{reports}}}\right)^{N_{\mathrm{update}}-1}, \tag{8}$$

where $N_{\mathrm{total}} \leq K$ denotes the total number of data subcarriers, N_{reports} the number of subcarrier groups per report, and N_{refine} the number of chosen subcarrier groups to be refined.

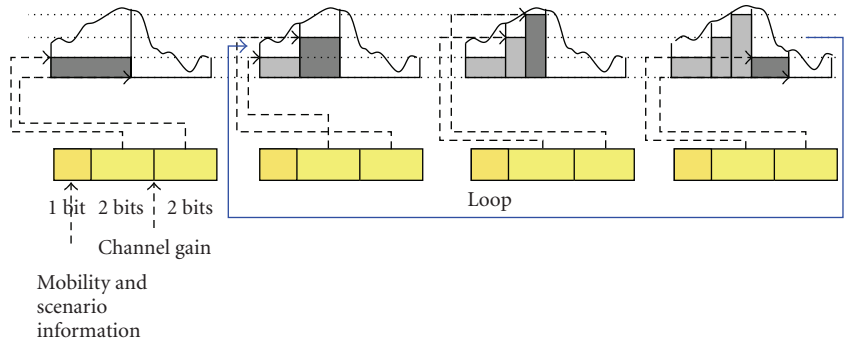

Figure 2: Illustration of successive refinement principle for feedback channel design.

Since only the users that belong to $\mathcal{A}_{k_0}$ need to be considered, we can nicely invoke [19, Theorem 2] stating that for some real ω, $\omega_0 := 2\pi k_0/K$

$$|h'_{m,k}| \geq \max_{k\in\mathcal{K}} |h'_{m,k}| \sqrt{\cos L(\omega-\omega_0)},$$
$$\omega_0 - \frac{\pi}{L} \leq \omega \leq \omega_0 + \frac{\pi}{L},\ m \in \mathcal{A}_{k_0}, \tag{9}$$

where $L = \max_{m\in\mathcal{M}} L_m$. Denoting the group of smallest cardinality by $\mathcal{S}_m^{k_0} \subseteq \mathcal{K}$, $m \in \mathcal{A}_{k_0}$, it follows that for $N_{\mathrm{gr}} < \lfloor K/2L \rfloor$ (9) will hold for all subcarriers within $\mathcal{S}_m^{k_0}$. This will indeed ensure that

$$\min_{k\in\mathcal{S}_m^{k_0}} |h'_{m,k}| \geq \sqrt{\cos\frac{\pi L N_{\mathrm{gr}}}{K}} \cdot \max_{k\in\mathcal{K}} |h'_{m,k}|, \quad m \in \mathcal{A}_{k_0}. \tag{10}$$

Since $\cos x \approx 1 - x^2$ for small x, the error will be small for large $K \gg L$. Further observing that it clearly holds

$$\Pr\left(\max_{m\in\mathcal{A}_{k_0}} |h'_{m,k_0}|^2 \geq \log M + c_0 \log\log M\right) \longrightarrow 0, \quad M \longrightarrow \infty \tag{11}$$

concludes the proof of the theorem. □

Theorem 3.1 characterizes the performance of the successive feedback scheme in terms of achievable throughput thereby, obviously, neglecting the impact of recurrent restart periods over time. In practice, the update period/restart period refers to a fraction/multiple of the channel coherent time $T_c = c/2\nu f_c$ where c denotes the speed of light, ν is the user speed, and f_c is the carrier frequency. A pedestrian user has T_c of 90 milliseconds. Hence, if the TTI length is 2 milliseconds, the deviation from the reported channel gain is less than 33% within 45 TTIs. In fact there is a tradeoff between the deviation and the number of refinement levels for each mobility class as shown in the simulations next.

3.3. Performance evaluation

In order to examine the throughput performance of the introduced feedback scheme, we use an opportunistic scheduler which assigns each subcarrier group to the user with best CQI value. We use physical parameters defined in [20] in order to evaluate the proposed system design. The transmission bandwidth is 5 MHz. The subcarriers 109 to 407 of the entire 512 subcarriers are occupied and used both for user data and feedforward control information. The number of subcarriers reserved for the feedforward channel is determined by the amount of the control information (assignment, user ID, modulation per subcarrier [group], code rate), the number of simultaneously supported users, and the employed coding scheme for the feedforward channel. For the feedforward scheme, many different approaches are thinkable. Here, we used an approach described in [17] but no effort has been made to optimize this approach. The TTI length is 2 milliseconds and the symbol rate is 27 symbols/TTI/subcarrier. In the sequel, always uniform power allocation is employed. If a subcarrier is asserted to a particular user, the complex data is modulated in either one of three constellations (QPSK, 16 QAM, 64 QAM, nothing at all) and one fixed coding scheme (2/3 code rate) is used. Perfect channel estimation is assumed throughout the paper and the required resources for pilot channels are neglected in the simulations. A detailed discussion of channel estimation schemes is beyond the scope of this paper (see, e.g., [21] for a discussion). Note that estimation errors can be easily incorporated since the transmitter performs bitloading based upon link-level simulations that can be repeated for different receiver structures. The feedback and feedforward link is assumed to be error free. *Furthermore, a delay interval of 4 TTIs between the CQI generation and transmission processing is considered in simulations*. The total number of users in the cell is set to 50 and no slow fading model is used.

The system throughput is measured as the amount of bits in data packets that are errorless received (over the air throughput). According to the current receive SNR and the used modulations on each subcarrier, a block error generator inserts erroneous blocks in the data stream. Since there is no standard error generation method in case of a dynamic frequency-selective transmission scheme, we use the simulation method given in [17] to generate the erroneous blocks.

Clearly, the better the scheduling works the more accurate the CQI reports represent the channel. Figure 3 shows the throughput improvement by increased feedback rate where the feedback scheme as described is used. In the scheme with 2 kbits/s feedback, 2 subcarrier groups are

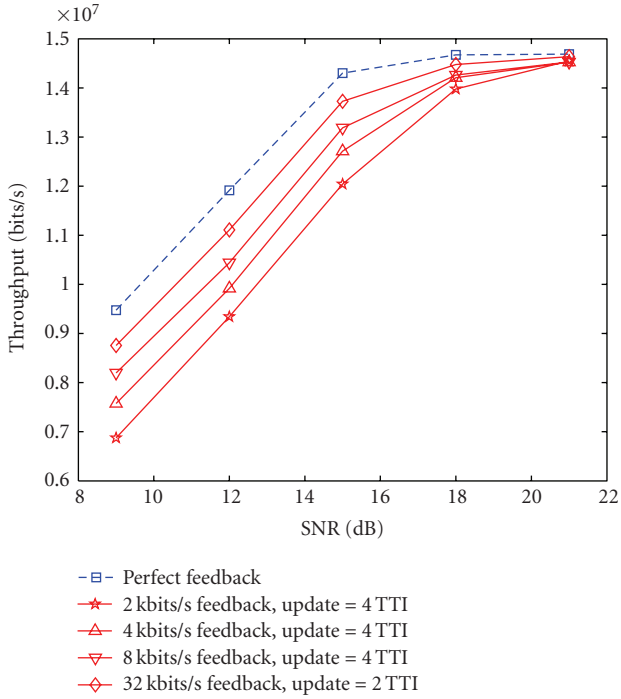

Figure 3: Throughput increase by improved feedback over average transmit SNR (5 users are simultaneously supported, Pedestrian B channel, 3 km/h, 24 subcarriers are reserved for feedforward control information).

reported in 4 levels per TTI. The channel gain of each subcarrier in the group must be higher than the reported level. Then the subcarrier group with higher level is split into 2 groups and reported in the next TTI. In the scheme with higher feedback rate, the number of reported groups per TTI is increased to 4, 8, and 32.

In our feedback scheme, the channel description is successively refined within a certain period of time. Obviously, the accuracy of the description largely depends on the period length. On the other hand, a long report period increases the delay of update information leading to a higher number of erroneous blocks. The throughput gain due to the improved feedback resolution and the loss caused by the delay is shown in Figure 4, where the throughput is maximized at an update period of 4 TTIs. Furthermore, the simultaneous support of several users provides multiuser gain. However, the necessary signaling information consisting of transmission modulation scheme, user identifier, subcarrier assignment has to be sent to the users through the downlink channel. The demand of the signaling information grows with the number of supported users and more subcarriers must be reserved for the feedforward channel instead of the data channel. Hence the achieved throughput gain is compensated by the increased signaling requirement. Figure 4 shows that the optimum is attained at 5 links with the present simulation setup. Note that, in order to improve the delay performance for delay-sensitive applications, a higher number of links can be applied at the cost of throughput loss.

The performance of frequency-selective and frequency-nonselective scheduling is presented in Figure 5. It was shown in [1] that even the frequency-nonselective OFDM system performs much better than the standard WCDMA system. Figure 5 shows that the frequency-selective scheduling yields much higher throughput for Pedestrian B, 3 km/h. The entire effective system throughput exceeds 10 MBit/s. The resulting block error rate is lower than 0.1. Note that for frequency-nonselective scheduling the required feedforward channel capacity is even neglected. The throughput gap between the frequency-selective and frequency-nonselective feedback schemes is also studied in [22].

4. SCHEDULER DESIGN

4.1. General concept

Users' QoS demands can be described by some appropriate utility functions that map the used resources into a real number. One typical class of utility functions is defined by the weighted sum of each user's rate, in which weight factors reflect different priority classes as, for example, used in HSDPA. If all weight factors are equal, the scheduler maximizes the total throughput. In addition, in order to meet strict requirements of real-time services, user specific rate demands have to be also considered. Heuristically, strict requirements also stem from retransmission requests of a running H-ARQ process which have to be treated in the very next time slot. Therefore, it is necessary to have additional individual minimum rate constraints in the utility maximization problem. Both is handled in the following scheduling scheme.

Arranging the (positive) weights and allocated rates for all user in vectors $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_M]^T$ and $\mathbf{R} = [R_1, \ldots, R_M]^T$, respectively, the resource allocation problem can be formulated as

$$\begin{aligned} &\text{maximize } \boldsymbol{\mu}^T \mathbf{R} \\ &\text{subject to } R_m \geq \overline{R}_m \quad \forall m \in \mathcal{M} \\ &\qquad \mathbf{R} \in \mathcal{C}_{\text{FDMA}}(\mathbf{h}, \mathbf{p}), \end{aligned} \tag{12}$$

where $\overline{\mathbf{R}} = [\overline{R}_1, \ldots, \overline{R}_M]^T$ are the required minimum rates. $\mathcal{C}_{\text{FDMA}}(\mathbf{h}, \mathbf{p})$ is the achievable OFDMA region for a fixed

$$\mathcal{C}_{\text{FDMA}}(\mathbf{h}, \mathbf{p}) \equiv \bigcup_{\substack{\sum_{m=1}^{M} \rho_{m,k} = 1 \ \forall k \\ \rho_{m,k} \in \{0,1\}}} \left\{ \mathbf{R} : R_m = \sum_{k=1}^{K} r_{m,k} \rho_{m,k} \right\}, \tag{13}$$

where the rates $r_{m,k}$ were defined in (3) and $\rho_{m,k} \in \{0, 1\}$ is the indicator if user m is mapped onto subcarrier k.

This problem is a nonlinear combinatorial problem that is difficult to solve directly, since there exist M^K subcarrier assignments to be checked. Thus, the computational demand for a brute-force solution is prohibitive.

In analogy to Lagrangian multipliers, we introduce in the following additional "soft" rewards $\widetilde{\boldsymbol{\mu}} = [\widetilde{\mu}_1, \ldots, \widetilde{\mu}_M]^T$ corresponding to the rate constraints. Note that since the problem is not defined on a convex set and the objective is not differentiable, it is not a convex-optimization problem. Nevertheless, the introduced formulation helps to find an excellent suboptimal solution.

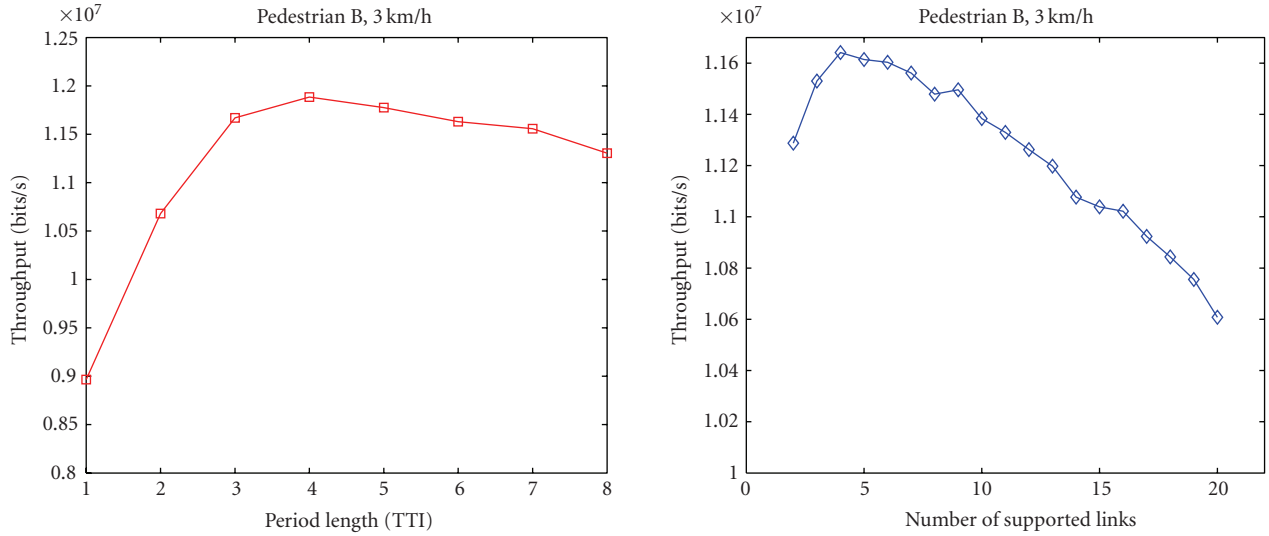

Figure 4: [a] Throughput with respect to update period (average transmit SNR equals 15 dB, 5 users are simultaneously supported). [b] Throughput with respect to simultaneously supported users (average transmit SNR equals 15 dB, feedback period equals 4 TTIs).

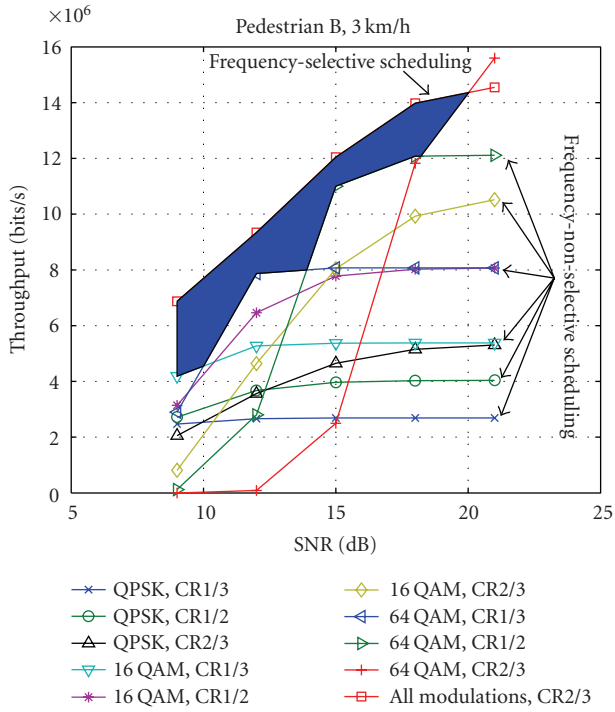

Figure 5: Throughput comparison of frequency-nonselective and frequency-selective scheduling over average transmit SNR (5 users are simultaneously supported and feedback period equals 4 TTIs).

Let us introduce the new problem with the additional "soft" rewards $\tilde{u}_m$,

$$\max_{\mathbf{R}\in\mathcal{C}_{\mathrm{FDMA}}(\mathbf{h},\mathbf{p})} \boldsymbol{\mu}^T\mathbf{R} + \tilde{\boldsymbol{\mu}}^T(\mathbf{R}-\overline{\mathbf{R}}). \tag{14}$$

Omitting the constant term $\tilde{\boldsymbol{\mu}}^T\overline{\mathbf{R}}$ in (14) and setting $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}+\tilde{\boldsymbol{\mu}}$ (14) can be rewritten as

$$\max_{\mathbf{R}\in\mathcal{C}_{\mathrm{FDMA}}(\mathbf{h},\mathbf{p})} \hat{\boldsymbol{\mu}}^T\mathbf{R}. \tag{15}$$

By varying the soft rewards $\tilde{\boldsymbol{\mu}}$, the convex hull of the set of all possible rate vectors is parameterized. If the solution to the original problem is a point on the convex hull of the achievable OFDMA region $\mathcal{C}_{\mathrm{FDMA}}(\mathbf{h},\mathbf{p})$, a set of soft rewards $\tilde{\boldsymbol{\mu}}$ has to be found such that the minimum rate constraints are met. Note, that the optimum may not lie on the convex hull and the reformulation will lead to a suboptimal solution. In this case, the obtained solution is the a point that lies on the convex hull and closest to the optimum. However, even for a moderate number of subcarriers, the said state is quite improbable.

The OFDM subcarriers constitute a set of orthogonal channels so the optimization problem (15) can be decomposed into a family of independent optimization problems

$$\max_{\mathbf{R}^{(k)}\in\mathcal{C}^{(k)}_{\mathrm{FDMA}}(\mathbf{h}_k,p_k)} \hat{\boldsymbol{\mu}}^T\mathbf{R}^{(k)} = \max_{n\in\mathcal{M}} \hat{\mu}_n r_{n,k}, \tag{16}$$

where $\mathbf{R}^{(k)}$ and $\mathcal{C}^{(k)}_{\mathrm{FDMA}}(\mathbf{h}_k,p_k)$ denote the rate vector and the achievable OFDMA region on subcarrier k, respectively, $\mathbf{h}_k = [h_{1,k},\ldots,h_{M,k}]^T$ is the vector of channel gains on subcarrier k. Assuming that the maximum $\max_{n\in\mathcal{M}}\hat{\mu}_n r_{n,k}$ is unique (which can be guaranteed by choosing $\hat{\boldsymbol{\mu}}$), the subcarrier and rate allocation can be calculated by a simple maximum search on each subcarrier. Hence the remaining task is to find a suitable vector of soft rate rewards $\tilde{\boldsymbol{\mu}}$ such that $\mathbf{R}(\hat{\boldsymbol{\mu}})$ maximizes $\boldsymbol{\mu}^T\mathbf{R}$ subject to the minimal rate constraints.

4.2. Scheduling algorithm

In the following, we introduce a simple iterative algorithm to obtain $\hat{\boldsymbol{\mu}}$ (see Algorithm 1). In the first step, the algorithm is initialized with $\hat{\boldsymbol{\mu}}^{(0)} = \boldsymbol{\mu}$. Note that step 0 is optional and will be introduced in the next subsection. Then in each iteration i, the rate rewards $\hat{\mu}_m^{(i-1)}$ are increased to $\hat{\mu}_m^{(i)}$ one after another such that the corresponding rate constraint $\overline{R}_m$ is met while the new reward $\hat{\mu}_m^{(i)}$ is the smallest possible

$$\begin{aligned} &\hat{\mu}_m \leq u, \quad \forall u \in \Theta, \\ &\Theta = \{u : R_m(\hat{\mu}_1, \ldots, \hat{\mu}_{m-1}, u, \ldots, \hat{\mu}_M) \geq \overline{R}_m\}. \end{aligned} \tag{17}$$

The search for $\hat{\mu}_m$ in step 3.1 can be done by simple bisection, since $R_m(\hat{\boldsymbol{\mu}})$ is monotone in $\hat{\mu}_m$. This fact is proven in the following Lemma.

Lemma 4.1. *For all m, if the mth component of $\hat{\boldsymbol{\mu}}$ is increased and the other components are held fixed, the rate $R_m(\hat{\boldsymbol{\mu}})$ remains the same or increases while $R_n(\hat{\boldsymbol{\mu}})$ remains the same or decreases for $n \neq m$.*

Proof. Denote the set of subcarriers assigned to user m as

$$\mathcal{S}_m = \left\{k : \hat{\mu}_m r_{m,k} = \max_{n \in \mathcal{M}} \hat{\mu}_n r_{n,k}\right\}. \tag{18}$$

The rates $R_m(\hat{\boldsymbol{\mu}})$ and $R_n(\hat{\boldsymbol{\mu}})$ only depend on the current subcarrier assignment. It is easy to show that in iteration $i+1$ an increase of $\hat{\mu}_m^{(i)}$ to $\hat{\mu}_m^{(i+1)}$ expands or preserves the set $\mathcal{S}_m$. More precisely, if there is any $k \in \mathcal{S}_m$ such that $\hat{\mu}_m^{(i)} r_{m,k} < \hat{\mu}_n r_{n,k} < \hat{\mu}_m^{(i+1)} r_{m,k}$, the rate of user m increases by $r_{m,k}$ while the rate of user n decreases by $r_{n,k}$. Otherwise the rates remain the same. □

To show the convergence of the algorithm, it is helpful to proof the *order preservingness* of the mapping defining the update of each step and hence the sequence $\{\hat{\boldsymbol{\mu}}^{(i)}\}$.

Lemma 4.2. *Let $\hat{\boldsymbol{\mu}}^{(i)} \leq \hat{\boldsymbol{\mu}}'^{(i)}$, where the inequality $\mathbf{a} \leq \mathbf{b}$ refers to component-wise smaller or equal. Then it follows $\hat{\boldsymbol{\mu}}^{(i+1)} \leq \hat{\boldsymbol{\mu}}'^{(i+1)}$.*

Proof. Observe user m and its rate reward $\hat{\mu}_m$ during iteration $i+1$. The subcarrier set allocated to user m after iteration $i+1$ is given by

$$\begin{aligned} &\mathcal{S}_m(\hat{\mu}_1'^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}'^{(i)}, \ldots, \hat{\mu}_M'^{(i)}) \\ &\quad = \{k : \hat{\mu}_m'^{(i+1)} r_{m,k} > \hat{\mu}_n'^{(i+1)} r_{n,k}, \forall n < m, \\ &\qquad \hat{\mu}_m'^{(i+1)} r_{m,k} > \hat{\mu}_n'^{(i)} r_{n,k}, \forall n > m\}. \end{aligned} \tag{19}$$

Due to the assumption, we have $\hat{\mu}_n'^{(i)} \geq \hat{\mu}_n^{(i)}$ for $n > m$. Additionally we assume

$$\hat{\mu}_n'^{(i+1)} \geq \hat{\mu}_n^{(i+1)} \tag{20}$$

for $n < m$, then for any subcarrier

$$k \in \mathcal{S}_m(\hat{\mu}_1'^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}'^{(i)}, \ldots, \hat{\mu}_M'^{(i)}), \tag{21}$$

it holds that

$$\begin{aligned} \hat{\mu}_m'^{(i+1)} r_{m,k} > \hat{\mu}_n'^{(i+1)} r_{n,k} \geq \hat{\mu}_n^{(i+1)} r_{n,k}, \quad \forall n < m \\ \hat{\mu}_m'^{(i+1)} r_{m,k} > \hat{\mu}_n'^{(i)} r_{n,k} \geq \hat{\mu}_n^{(i)} r_{n,k}, \quad \forall n > m. \end{aligned} \tag{22}$$

Hence,

$$\begin{aligned} &\mathcal{S}_m(\hat{\mu}_1'^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}'^{(i)}, \ldots, \hat{\mu}_M'^{(i)}) \\ &\quad \subseteq \mathcal{S}_m(\hat{\mu}_1^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}^{(i)}, \ldots, \hat{\mu}_M^{(i)}) \end{aligned} \tag{23}$$

and thus we get the following inequality for the rates:

$$\begin{aligned} &R_m(\hat{\mu}_1^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}^{(i)}, \ldots, \hat{\mu}_M^{(i)}) \\ &\quad \geq R_m(\hat{\mu}_1'^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}'^{(i)}, \ldots, \hat{\mu}_M'^{(i)}). \end{aligned} \tag{24}$$

According to the definition of the algorithm, we know that $R_m(\hat{\mu}_1'^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \hat{\mu}_{m+1}'^{(i)}, \ldots, \hat{\mu}_M'^{(i)})$ fulfills the rate constraint $\overline{R}_m$ and therefore also

$$R_m(\hat{\mu}_1^{(i+1)}, \ldots, \hat{\mu}_m'^{(i+1)}, \ldots, \hat{\mu}_M^{(i)}) \geq \overline{R}_m. \tag{25}$$

Recalling the criterion (17) of the update rule, we know that $\hat{\mu}_m^{(i+1)}$ must be the minimum of all possible μ that fulfill the inequality (25) so that $\hat{\mu}_m'^{(i+1)} \geq \hat{\mu}_m^{(i+1)}$ follows. This argument holds for the first user without the additional assumption (20) and the proof then can be extended inductively for users $n > 1$, which concludes the proof.

□

Now we are able to give the central theorem ensuring convergence of the algorithm.

Theorem 4.3. *The given algorithm converges to the componentwise smallest vector $\hat{\boldsymbol{\mu}}^*$, which is a feasible solution of the system such that $R_m(\hat{\boldsymbol{\mu}}^*) \geq \overline{R}_m$, $\forall m \in \mathcal{M}$.*

Proof. If $\mathbf{R}(\hat{\boldsymbol{\mu}}^*)$ fulfills all rate constraints, then $\hat{\boldsymbol{\mu}}^*$ is a fixed point of the algorithm $\hat{\boldsymbol{\mu}}^* = \hat{\boldsymbol{\mu}}^{(i)} = \hat{\boldsymbol{\mu}}^{(i+1)}$, $\forall i \in \mathbb{N}_+$. We also have $\hat{\boldsymbol{\mu}}^* \geq \boldsymbol{\mu}$ since $\tilde{\boldsymbol{\mu}} \in \mathbb{R}_+^{\mathbb{M}}$. Starting with $\hat{\boldsymbol{\mu}}^{(0)} = \boldsymbol{\mu}$, we know that $\{\hat{\boldsymbol{\mu}}^{(i)}\}$ is a componentwise monotone sequence $\hat{\boldsymbol{\mu}}^{(i+1)} \geq \hat{\boldsymbol{\mu}}^{(i)}$. Define a mapping U representing the update of the sequence $\{\hat{\boldsymbol{\mu}}^{(i)}\}$, it follows from Lemma 4.2 that for all i, $\hat{\boldsymbol{\mu}}^{(i)} = U^i(\hat{\boldsymbol{\mu}}^{(0)}) \leq U^i(\hat{\boldsymbol{\mu}}^*) = \hat{\boldsymbol{\mu}}^*$. Hence, $\{\hat{\boldsymbol{\mu}}^{(i)}\}$ is a monotone increasing sequence bounded from above and converges to the limiting fixed point $\hat{\boldsymbol{\mu}}^*$. This completes the proof. □

Next we analyze the obtained fixed point $\mathbf{R}(\hat{\boldsymbol{\mu}}^*)$. Given $\hat{\boldsymbol{\mu}}^*$ which is the fixed point of the algorithm, let $\tilde{\mathcal{S}}_m$ denote the set of carriers, which are assigned to user m at the fixed point $\hat{\boldsymbol{\mu}}^*$ of the algorithm, but were not allocated to according to the original weights $\boldsymbol{\mu}$

$$\tilde{\mathcal{S}}_m = \left\{k : \mu_m r_{m,k} < \max_{n \in \mathcal{M}} \mu_n r_{n,k}, \hat{\mu}_{m,k}^* r_{m,k} = \max_{n \in \mathcal{M}} \hat{\mu}_n^* r_{n,k}\right\}. \tag{26}$$

Denoting the optimal rate allocation not considering the minimal rate constraints as $\mathbf{R}_{\text{opt}}$, then the value of the objective function $f(\hat{\boldsymbol{\mu}}^*) \equiv \boldsymbol{\mu}^T \mathbf{R}(\hat{\boldsymbol{\mu}}^*)$ can be decomposed to the

(**0**) add a random noise matrix $\mathbf{\Delta}$ with uniformly distributed entries to the rate gain matrix: $\mathbf{r}' = \mathbf{r} + \mathbf{\Delta}$
(**1**) initialize weight vector $\hat{\boldsymbol{\mu}}^{(0)} = \boldsymbol{\mu}$
(**2**) calculate the subcarrier assignment $i(k) = \arg\max_{m\in\mathcal{M}} \hat{\mu}_m r'_{m,k} \forall k$ and the resulting rate allocation $R_m = \sum_{k\in\mathcal{K}, i(k)=m} r_{m,k}$
while rate constraints $\mathbf{R} \geq \overline{\mathbf{R}}$ not fulfilled **do**
  for $m = 1$ to M **do**
    if $R_m < \overline{R}_m$ **then**
      (**3.1**) increase $\hat{\mu}_m$ according to the criteria described in step (**2**) such that the rate constraint of user m is fulfilled
      (**3.2**) recalculate $i(k)$ and R_m
    end if
  end for
end while

Algorithm 1: Reward enhancement algorithm.

sum of this optimum value and an additional term stemming from the reassignment of carriers due to the modification of the rate rewards

$$f(\hat{\boldsymbol{\mu}}^*) = \boldsymbol{\mu}^T \mathbf{R}_{\text{opt}} + \sum_{m\in\mathcal{M}} \sum_{k\in\widetilde{\mathcal{S}}_m} \left(\mu_m r_{m,k} - \max_{n\in\mathcal{M}} \mu_n r_{n,k}\right). \quad (27)$$

Since each addend in the second term is negative due to the definition of $\widetilde{\mathcal{S}}_m$, any expansion of the set $\widetilde{\mathcal{S}}_m$ reduces the object value. Hence, each set size $\widetilde{\mathcal{S}}_m$ must be kept minimal while fulfilling the rate constraint R_m. Using Lemma 4.1, we can conclude that this is the case for the minimum value of $\hat{\boldsymbol{\mu}}$ already fulfilling the rate constraints.

4.3. Uniqueness and random noise addition

However, in some cases the minimum of $\widetilde{\mathcal{S}}_m$ cannot be achieved directly and the proposed algorithm has to be modified. This can be illustrated constructing the following example: assuming that there exist

$$\frac{r_{m,k}}{r_{m,j}} = \frac{r_{l,k}}{r_{l,j}}, \quad m \neq l, \quad (28)$$

$$\begin{aligned} \mu_l r_{l,k} &= \max_{n\in\mathcal{M}} \mu_n r_{n,k}, \\ \mu_l r_{l,j} &= \max_{n\in\mathcal{M}} \mu_n r_{n,j}, \\ \hat{\mu}_l^* r_{l,j} &= \max_{n\in\mathcal{M}, n\neq m} \hat{\mu}_n^* r_{n,j}. \end{aligned} \quad (29)$$

If $k \in \widetilde{\mathcal{S}}_m$ so that $\hat{\mu}_m^* r_{m,k} > \hat{\mu}_l^* r_{l,k}$, we get $\hat{\mu}_m^* r_{m,j} > \hat{\mu}_l^* r_{l,j}$ from (28) and further $j \in \widetilde{\mathcal{S}}_m$ from (29). If the set $\widetilde{\mathcal{S}}' = \widetilde{\mathcal{S}}/\{j\}$ which is the subset of $\widetilde{\mathcal{S}}$ without subcarrier j already meets the rate constraint, the selection of $\widetilde{\mathcal{S}}_m$ leads to a suboptimal solution. It is worth noting that the quantization and compression of the channel state information in feedback channel blur the distinctness between the rate profit $r_{m,k}$, therefore the aforementioned state occurs frequently. A simple workaround can cope with this effect. In order to avoid the leap in rate allocation we use modified rate profits

$$r'_{m,k} = r_{m,k} + \delta_{m,k}, \quad m \in \mathcal{M},\ k \in \mathcal{K}. \quad (30)$$

To this end, random noise $\delta_{m,k} \in \mathbb{R}_+$ is added to the original rate profits, where $\delta_{m,k}$ is uniformly distributed on the interval $(0, \triangle r)$, where $\triangle r$ is the minimum distance between all possible rate values. Thus the rate profits can be distinguished avoiding the occurrence of (28). Note that the user selection of the subcarriers is unchanged since for any $r_{m,k} > r_{l,k}$ we still have $r'_{m,k} > r'_{l,k}$. This effect can be illustrated geometrically and is depicted in Figure 6.

Geometrically, the objective is to depart a hyperplane with normal vector $\boldsymbol{\mu}$ as far as possible from the origin not leaving the achievable rate region $\mathcal{C}_{\text{FDMA}}$. In the upper example without random noise, the region has a big flat part with equal slope. In order to fulfill the rate constraint the normal vector of the plane is changed to $\hat{\mu}''$ so that $\mathbf{R}$ reaches the feasible region (filled region in the figure). Thus, the algorithm skips $\mathbf{R}^*$ and switches from $\mathbf{R}'$ directly to $\mathbf{R}''$ constituting a suboptimal point. In the second example, it can be seen that random noise makes the region more curved, avoiding the described problem. The algorithm now ends up in the optimum $\mathbf{R}^*$.

4.4. Performance evaluation

Using the same physical parameters for the evaluation of the control channel, we examine at first the throughput performance of the introduced scheduling algorithm.

Figure 7 illustrates the convergence process for an exemplary random channel with $K = 299$ subcarriers and $M = 5$ users. The complete system setting is the same as it is used in the previous throughput simulations. The channel state information is obtained through a feedback channel(2 kbits/s). In every TTI (2 milliseconds) 27 symbols are transmitted per subcarrier. The modulation is adapted to the different channel states on each subcarrier and can be chosen from QPSK, 16 QAM, 64 QAM. The averaged receive SNR is 15 dB, $\mu = [1, 1, 1, 1, 1]^T$. The required minimum rates are set to $\overline{\mathbf{R}} = [1000, 2000, 6000, 5000, 0]^T$ bits/TTI, where 0 means no minimum rate constraint. The algorithm stops at the point of complete convergence which is shown as the dashed vertical line in Figure 7. The number of iterations depends on the

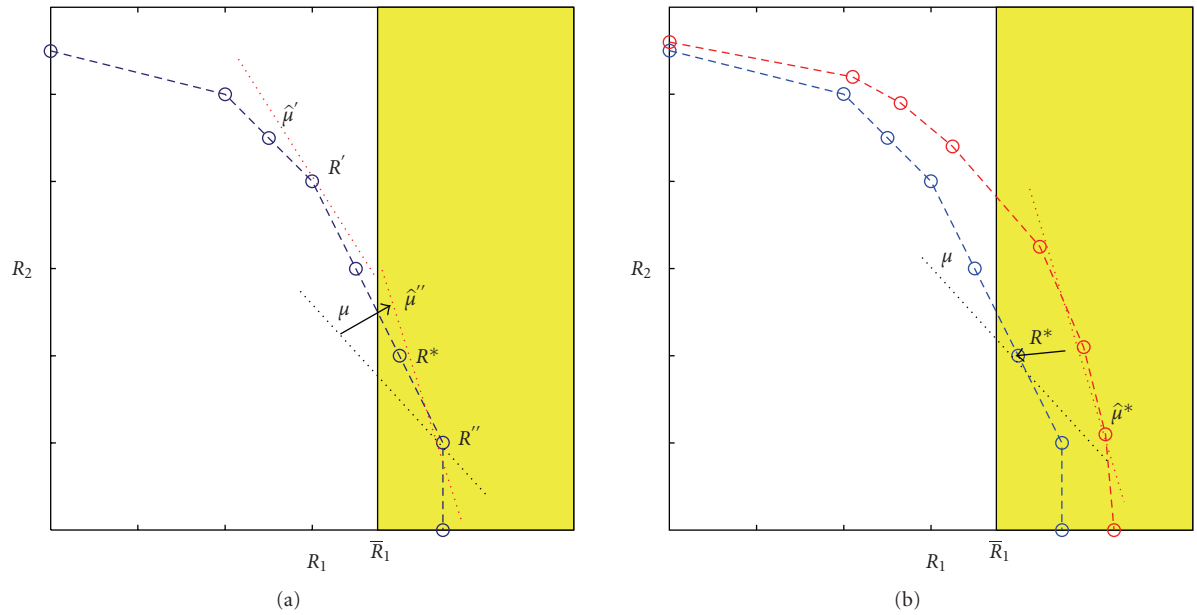

Figure 6: Fixed point of the algorithm without (left) and with (right) random noise.

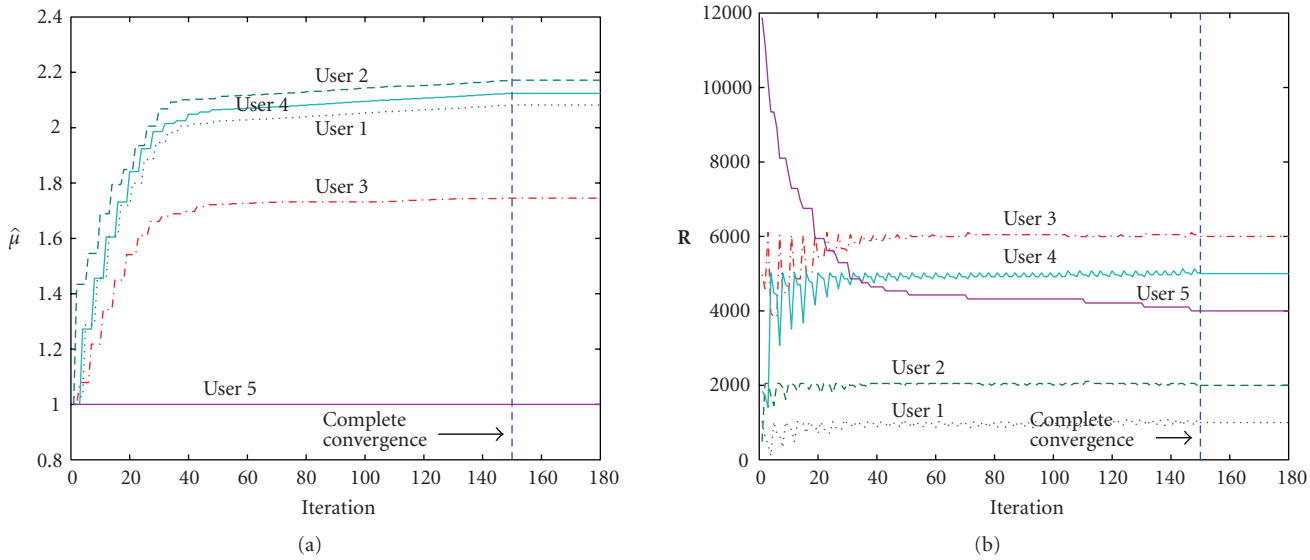

Figure 7: Convergence of $\hat{\boldsymbol{\mu}}$ (left) and $\mathbf{R}$ (right).

given channel rate profits and the rate constraints. For some channel states, the rate constraints are not achievable and the algorithm will not converge. To cope with these infeasible cases, we expand the algorithm by an additional break condition which consists of a maximum number of iterations. If the number of iteration steps is on a threshold, the iteration should be broken up and the user with the largest $\hat{\mu}$, who has also the worst channel condition, is removed from the scheduling list. Then the scheduling algorithm is initialized and started again. The removed user will not be served and the link is dropped in this TTI.

In order to evaluate the scheduler's performance, we also implemented the Hungarian assignment algorithm from [10] which solves a general resource assignment problem. Modeling the reward of certain resources as an $N \times N$ square matrix, of which each element represents the reward of assigning a "worker" (equal to a subcarrier) to a "job" (user), the Hungarian algorithm yields the optimal assignment that maximizes the total reward. Unfortunately, the complexity of the algorithm depends on the given reward matrix and increases very fast with the size of the matrix. The Hungarian algorithm realizes an optimal assignment strategy but, before starting the algorithm, the number of subcarriers each user is assigned must be determined a priori. This means that the scheduler must estimate the necessary number of subcarriers for each user in order to achieve the minimum

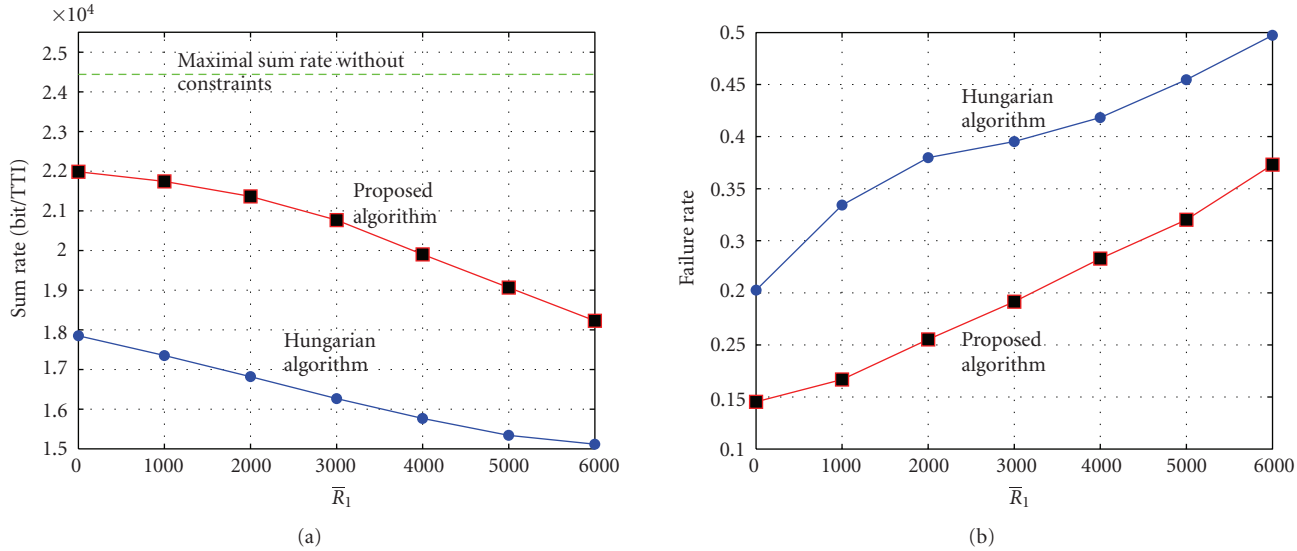

Figure 8: Sum rate (left) and failure rate (right) comparison between both algorithms.

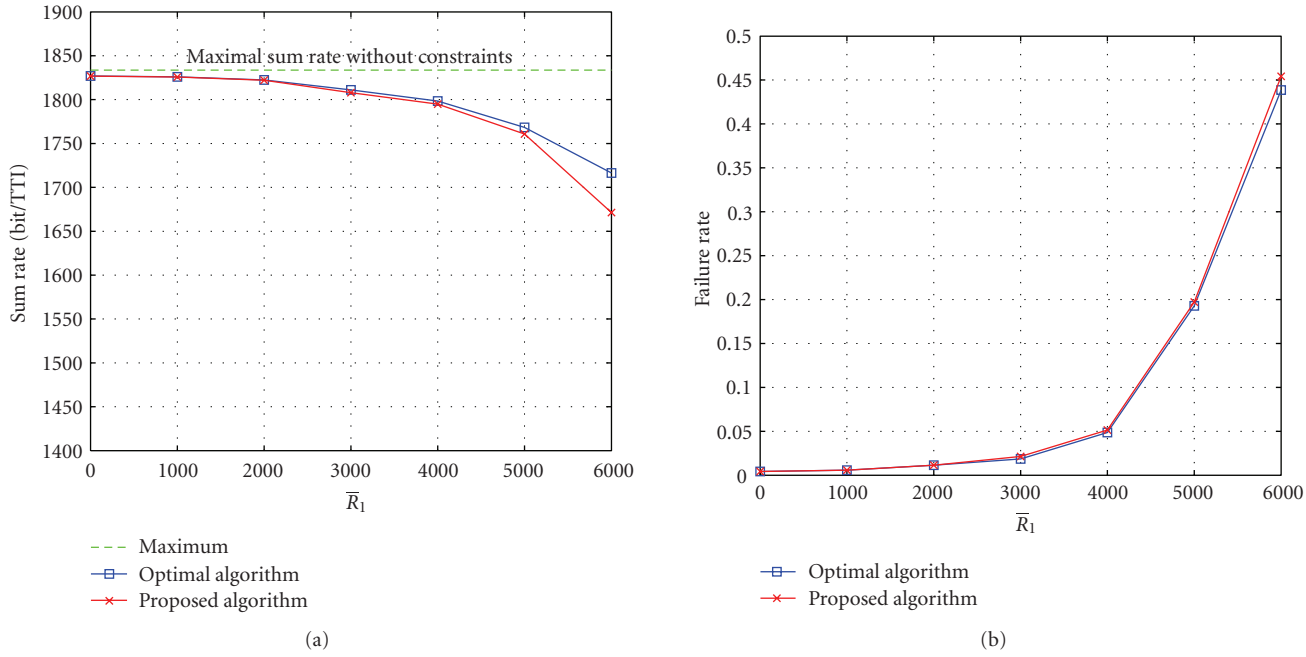

Figure 9: Sum rate (left) and failure rate (right) comparison with optimal solution.

rate requirement. Even though a rough estimate can be obtained by dividing the required rate by the average rate profit on each subcarrier as described in [10], we know that this estimation is quite imprecise in a frequency-selective channel with large frequency dispersion. Such an improper estimation impairs the algorithm even if the assignment algorithm itself is optimal. Figure 8 shows the comparison between the reference scheduler and the proposed scheduling algorithm. For the same system as in Figure 7, we set also $\mu = [1, 1, 1, 1, 1]^T$ which means that the sum rate of the system is maximized. Holding the minimum rate constraints of user 2–4 $[\overline{R}_2, \overline{R}_3, \overline{R}_4, \overline{R}_5]^T = [2000, 2000, 0, 0]^T$ bits/TTI and increasing the rate constraints of user 1 from 0 to 6000 bits/TTI, we can see a drop in sum rate in Figure 8. Defining a transmission failure in case that the minimum rate constraints are not met, the failure rate rise over the minimum rate $\overline{R}_1$ is depicted in the right part of Figure 8. It can be seen that the introduced algorithm clearly outperforms the reference scheduler for both measures.

A comparison with the optimal solution (with brute-force search) is shown in Figure 9. Due to the high computational demand we set $K = 16$ subcarriers and $M = 2$

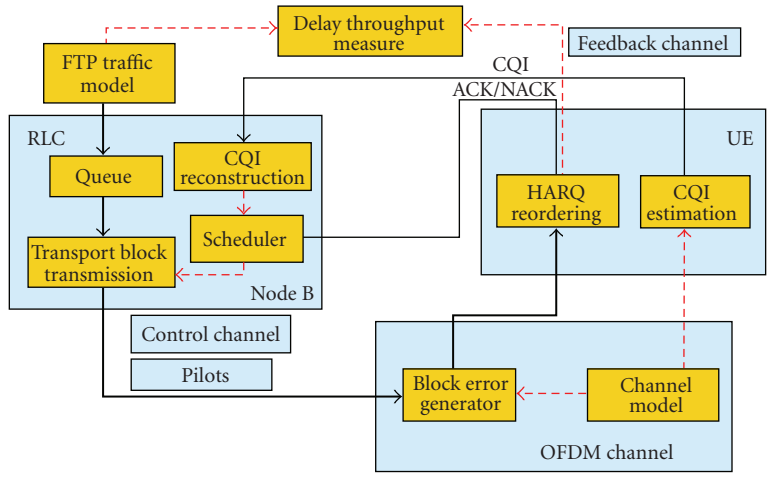

Figure 10: Simulation modules.

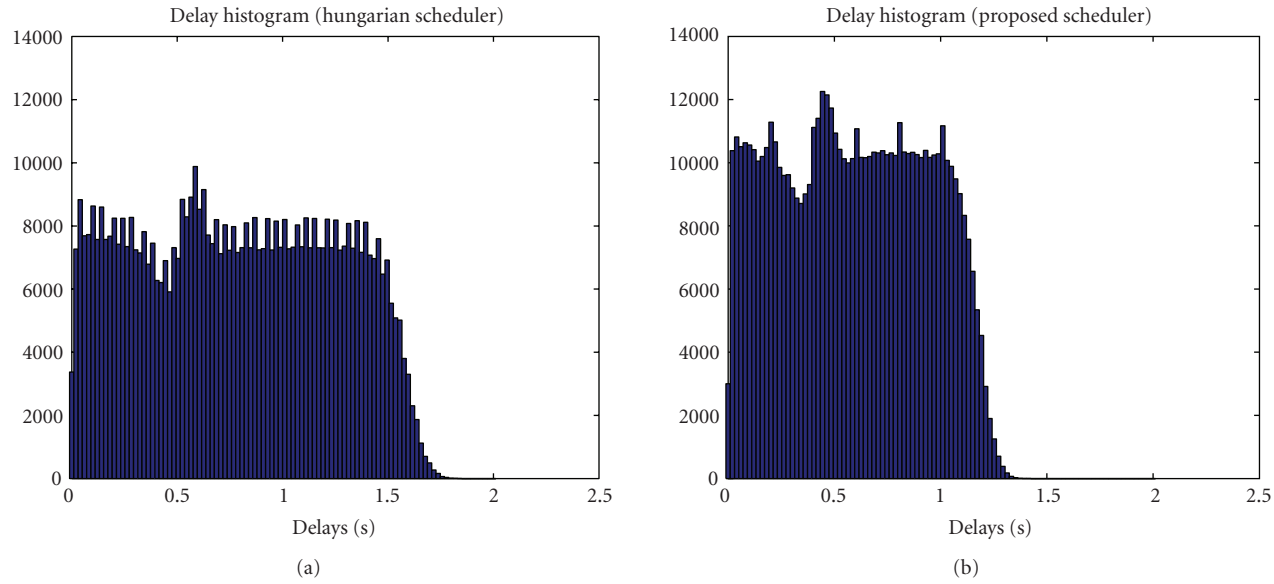

Figure 11: Delay histogram (total Tx power equals 43 dBm, the interference and noise power equals −47.46 dBm, feedback period equals 4 TTIs, maximal 5 users are simultaneously supported.

users. The rest of the system settings are the same as those in Figure 7. Increasing the rate constraint $\overline{R}_1$ from 0 to 1200 bits/TTI by fixed rate constraint $\overline{R}_2 = 400$ bits/TTI, we compare both algorithms in terms of sum rate and failure rate. The proposed algorithm causes only little performance loss in this simulation, as mentioned before, the performance loss will be further reduced in the system with higher number of subcarriers.

5. SYSTEM SIMULATIONS

We applied the simulation structure in Figure 10 to evaluate the entire system performance including the scheduler. An FTP traffic model was used in which the arrival page and packet size were fixed of 125 KBytes and 1500 Bytes and page reading time was 5 seconds.

In the base station, the amount of data to be transmitted for each user is stored separately in a queue backlog. A resource scheduler determines the transmit block size for each user based on queue states and feedback information per TTI. Using this transmission scheme and current channel conditions as input, a block error generator (the same that was used in Section 3.3) inserts erroneous blocks in the stream. The errorless transmission is confirmed with the H-ARQ signal and the block is removed from queue in base station. In the case of an erroneous transmission attempt, the block must be retransmitted in one of the next time slots. (There will be no packet loss in the system.)

The slow-fading performance is determined by the users' position that is described in a simple random walk model [23]. In the model the movement of users is restricted to the area of a single cell with 500 m diameter. Each user is

moving with a constant speed according to its mobility class and changes its direction with a statistical behavior. Thus, the mean path loss is calculated based on the distance between the user and base station. A slow fading shadowing model [24] is also applied which reflects the deviation from the mean path loss due to the specific shadowing. This shadowing deviation is determined by the density of solid shading objects that is specified in the simulation environment. twenty-five users were in the cell and were assumed to have the same channel profile (Pedestrian B, 3 km/h).

The delay performance of the system using Hungarian algorithm and the proposed algorithm is compared in Figure 11. We used the longest-queue-highest-possible-rate policy in μ for the proposed algorithm. The policy uses the current queue length as the weight factor $\boldsymbol{\mu}$ and is known to have good delay performance. The histograms show that the delay performance can be significantly improved by the new scheduler.

6. CONCLUSIONS

This paper addresses the conceptional evolution towards a new OFDM-based UMTS LTE concept. Practical constraints such as feedback capacity, feedforward demand, and user mobility strongly affect the overall performance. Hence, the linchpin, that is, the optimized feedback scheme, was devised to cope with these constraints and to facilitate optimum system performance. Further, we proposed a scheduling algorithm, which assigns subcarriers efficiently and is able to handle minimum rate constraints. This is a nonconvex discrete optimization problem with nondifferentiable objective. Nevertheless, based on a reward enhancement strategy, the algorithm is proven to converge to an excellent suboptimal solution, which often is the global optimum. Simulation results show that the proposed algorithm outperforms the well-known algorithm from [10] in terms of throughput and failure rate. Combining the algorithm with other scheduling policies, we verified by system simulations that it provides also excellent delay performance.

ACKNOWLEDGMENTS

Parts of this work were supported by the German Ministry for Education and Research under Grant FK 01 BU 350 (the *3GET* project). Parts were presented at the IEEE IST Summit in Dresden (Germany), June 2005 and the IEEE ICC 2007, Glasgow (GB).

REFERENCES

[1] Alcatel, "Further results on link level comparisons of WCDMA and OFDM transmission," 3GPP TSG RAN WG1 #37, R1-04-0571, May 2004.

[2] Ericsson, "Summary of downlink performance evaluation," TSG-RAN WG1 #49, R1-072444, May 2007.

[3] Alcatel-Lucent, "DL E-UTRA performance checkpoint," R1-071967, April 2007.

[4] Alcatel, "OFDM with interference control for improved HSDPA coverage," 3GPP TSG RAN WG1 #37, R1-04-0572, May 2004.

[5] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.

[6] M. Ergen, S. Coleri, and P. Varaiya, "Qos aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems," *IEEE Transactions on Broadcasting*, vol. 49, no. 4, pp. 362–370, 2003.

[7] S. Pietrzyk and G. J. M. Janssen, "Multiuser subcarrier allocation for QoS provision in the OFDMA systems," in *Proceedings of the 56th IEEE Vehicular Technology Conference (VTC '02)*, vol. 2, pp. 1077–1081, Vancouver, BC, Canada, September 2002.

[8] I. Kim, I.-S. Park, and Y. H. Lee, "Use of linear programming for dynamic subcarrier and bit allocation in multiuser OFDM," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1195–1207, 2006.

[9] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proceedings of the 51st IEEE Vehicular Technology Conference (VTC '00)*, vol. 2, pp. 1085–1089, Tokyo, Japan, May 2000.

[10] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," in *Proceedings of IEEE Global Telecommunication Conference (GLOBECOM '00)*, vol. 1, pp. 103–107, San Francisco, Calif, USA, November-December 2000.

[11] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," in *Proceedings of the 40th Annual Conference on Information Sciences and Systems (CISS '06)*, pp. 1272–1279, Princeton, NJ, USA, March 2006.

[12] E. Yeh and A. Cohen, "Throughput and delay optimal resource allocation in multiaccess fading channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '03)*, p. 245, Yokohama, Japan, June-July 2003.

[13] T. Michel and G. Wunder, "Minimum rates scheduling for OFDM broadcast channels," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, vol. 4, pp. 41–44, Toulouse, France, May 2006.

[14] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Nashua, NH, USA, 2003.

[15] 3GPP, "Technical specification group radio access network; user equipment, radio transmission and reception (FDD)," Technical Specification TS-25.101, v.6.3.0, Release 6, 3GPP, Valbonne, France, December 2004.

[16] 3GPP, "Technical specification group radio access network; multiplexing and channel coding (FDD)," Technical Specification TS-25.212, v.6.3.0, Release 6, 3GPP, Valbonne, France, 2004.

[17] C. Zhou, G. Wunder, H.-E. Bakker, and S. Kaminski, "OFDM-HSDPA: conceptual approach, simulation methodology and throughput performance," in *Proceedings of the 10th International OFDM Workshop*, Hamburg, Germany, August-September 2005.

[18] S. Litsyn and G. Wunder, "Generalized bounds on the crest-factor distribution of OFDM signals with applications to code design," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 992–1006, 2006.

[19] G. Wunder and H. Boche, "Peak value estimation of bandlimited signals from their samples, noise enhancement, and a local characterization in the neighborhood of an extremum," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 771–780, 2003.

[20] 3GPP, "Technical specification group radio access network; feasibility study for OFDM for UTRAN enhancement," Tech. Rep. TS-25.892, v.1.1.1, Release 6, 3GPP, Valbonne, France, May 2004.

[21] O. Edfors, M. Sandell, J.-J. van de Beek, B. Wilson, and P. Borjesson, "OFDM channel estimation by singular value decomposition," in *Proceedings of the 46th IEEE Vehicular Technology Conference (VTC '96)*, vol. 2, pp. 923–927, Atlanta, Ga, USA, April-May 1996.

[22] Alcatel-Lucent, "Incremental CQI feedback scheme and simulation results," 3GPP TSG RAN WG2 #58, R2-072670, June 2007.

[23] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002, special issue on Mobile Ad Hoc Networking.

[24] 3GPP, "Technical specification group radio access network; physical layer aspects for evolved universal terrestrial radio access (UTRA)," Tech. Rep. TS-25.814, v.7.1.0, Release 7, 3GPP, Valbonne, France, September 2006.