Research Article

CDIT-Based Constrained Resource Allocation for Mobile WiMAX Systems

Felix Brah, Jerome Louveaux, and Luc Vandendorpe

Ecole Polytechnique de Louvain, Université Catholique de Louvain, Place du Levant 2, 1348 Louvain-la-Neuve, Belgium

Correspondence should be addressed to Felix Brah, felix.brah@uclouvain.be

Adaptive resource allocation has been shown to provide substantial performance gain in OFDMA-based wireless systems, such as WiMAX, when full channel state information (CSI) is available at the transmitter. However, in some fading environments (e.g., fast fading), there may not be a feedback link sufficiently fast to convey the CSI to the transmitter. In this paper, we consider resource allocation strategies for downlink multiuser mobile WiMAX systems, where the transmitter has only the channel distribution information (CDI), but no knowledge of the instantaneous channel realization. We address the problem of subchannel assignment and power allocation, to maximize the ergodic weighted-sum rate under long-term fairness, minimum data rate requirement and power budget constraints. This problem is formulated as a nonlinear stochastic constrained optimization problem. We provide an analytical solution based on the Lagrange dual decomposition framework. The proposed method has a complexity of $\mathcal{O}(KM)$ for K users and M subchannels. Simulation results are provided to compare the performance of this method with other allocation schemes and to illustrate the tradeoff between maximized weighted-sum rate and the constraints.

1. Introduction

The mobile version of the Worldwide Interoperability for Microwave Access (mobile WiMAX) is one of the solutions in the competition for wireless broadband applications in challenging mobile environments [1, 2]. The mobile WiMAX air interface is based on orthogonal frequency division multiple access (OFDMA) for improved performances in multipath environments. One of the future aspects of OFDMA is the subchannelization which allows to group a total number of subcarriers into subsets of subcarriers called subchannels [3]. The major advantage of subchannelization is the provision of frequency diversity. A byproduct of the subchannelization is that the need for knowledge of radio channel quality is reduced from per-subcarrier to per-subchannel resolution and resources are allocated on per-subchannel basis. There are three types of subchannelizations, namely, adaptive modulation and coding (AMC), partially used subchannelization (PUSC), and fully used subchannelization (FUSC). With AMC, the subchannels are composed of contiguous groups of subcarriers. With both PUSC and FUSC, the subchannels are composed of distributed subcarriers. For PUSC, the set of used subcarriers, that is, data and pilots, is first partitioned into subchannels, and then pilot subcarriers are allocated within each subchannel. For FUSC, the pilot tones are common for all subchannels and are allocated first and then the remaining subcarriers are divided into data subchannels. In general, AMC is well suited for stationary, portable, and low mobility applications, whereas PUSC and FUSC are the best alternatives for mobile applications. We employ FUSC in this work. This method uses all the subchannels and employs full-channel diversity by distributing the allocated subcarriers to subchannels using a permutation mechanism. Thanks to the frequency diversity provided by the FUSC, the performance degradation due to fast fading in mobile environments is minimized.

Mobile WiMAX aimed at delivering broadband mobile services ranging from real-time interactive gaming, VoIP, and streaming media to nonreal-time web browsing and simple file transfers. Users have channels of different quality. With

classical best effort transmission, unfair resource allocation can lead to starvation of some users in bad channel conditions. Therefore, the achievement of fairness among users while satisfying users' minimum rate requirements is an important issue.

Most of the previous works on OFDMA resource allocation have considered only the case where instantaneous channel state (CSI) is available at the transmitter and various algorithms based on instantaneous CSI have been developed [4–14]. In [4], adaptive subcarriers assignment to minimize the total transmit power is investigated. The authors presented a heuristic algorithm, the so-called Hungarian algorithm, based on constructive assignment and iterative improvement. Following the Hungarian approach, [5] proposed an iterative algorithm for power minimization and bit loading. The algorithm is considered as suboptimal for adaptive modulation. To reduce the computational complexity, [6] proposed low complexity and computationally efficient bandwidth and power allocation algorithms to solve the problem of minimizing the total power consumption under bit error rate and transmission rate constraints. In [7], the performance of bandwidth-constrained power minimization and power minimization schemes in terms of outage probability and packet error rate under user data rate satisfaction are compared. It is shown that, in severe shadowing environment with both frequency selective and flat fading, the former scheme outperforms the later. Fairness issues in a wireline multiaccess channel have been taken into account in [8, 9]. The authors introduce the concept of balanced capacity to characterize the multiuser channel performance with total power constraints in [8] and they extend the concept to individual power constraints in [9]. This concept of balanced capacity is closely related to the one presented in [10] where a low complexity suboptimal algorithm that maximizes the sum capacity while maintaining proportional fairness among the users data rate is developed. In [11], suboptimal resource grids and power allocation algorithms to maximize the total throughput under user's data rate requirement are presented. Rate-power allocation algorithms for expected mutual information maximization based on partial channel knowledge have been developed in [13]. In [14], the authors investigated the impact of imperfect channel information on OFDMA-based systems under fairness and minimum rate constraints. Instantaneous resource allocation strategies are suitable for quasistatic or slow fading environments. However, when the channel variations are fast, the transmitter may not be able to adapt to the instantaneous channel realization. Hence, CSI-aware resource allocation is not suitable for environments with high mobility.

When the channel state can be accurately tracked at the receiver, the statistical channel model at the transmitter can be based on channel distribution information feedback from the receiver. We refer to knowledge of the channel distribution at the transmitter as CDIT. Power allocation for ergodic capacity maximization in relay networks based on CDIT under high SNR regime has been studied in [15].

This paper addresses CDIT-based resource allocation strategies for mobile WiMAX in all SNR regimes. The goal is to adaptively assign subchannels and distribute the total power to users with the objective to maximize the ergodic weighted-sum rate under tunable long-term fairness, minimum data rate requirements, and a total power constraint. This constrained optimization problem is formulated as an infinite dimensional stochastic problem. To efficiently solve the problem, we propose an analytical method based on the Lagrange dual decomposition framework. The remainder of this paper is organized as follows. In Section 2, the system model considered is described and the ergodic weighted-sum rate is derived. The problem of multiuser resource allocation based on CDIT is formulated in Section 3 and a solution guideline is given. In Section 4, some simulation results are presented. Finally, conclusions are drawn in Section 5.

2. System Model

Throughout the paper, we consider a single cell downlink WiMAX communication from a base station (BS) to K mobile user terminals, over a realistic frequency-selective fast fading channel with total bandwidth B. The BS splits up the downlink bandwith into different subchannels. Then the data to be transmitted to different mobile user terminals are amalgamated using downlink FUSC. After the downlink subchannelization, the resulting frequency domain OFDMA symbols are converted into time domain OFDMA symbols using inverse FFT. Then a cyclic prefix is added to each symbol to provide immunity against multipath propagation. Finally the signal undergoes frequency upconversion before it is transmitted from the base station to the user terminals. We assume that the user terminal has perfect CSI to perform coherent detection, but there is no fast feedback link to perfect the CSI to the base station. Hence, the base station has only channel distribution information (CDI), but no knowledge of the instantaneous channel realizations. Assuming that the receiver employs a maximum ratio combiner (MRC), the effective SNR of user k at mth subchannel is given by

$$\gamma_{k,m} = \frac{1}{\Gamma_k N \sigma_n^2} \sum_{n=0}^{N-1} g_{k,m}(n). \qquad (1)$$

In (1), N is the number of distributed subcarriers persubchannel, $g_{k,m}(n)$ is the channel gain of user k at subcarrier n of mth subchannel, which is the product of the distance attenuation and the fast fading gain, σ_n^2 is the noise variance, Γ_k is referred to as SNR gap related to the required bit error rate of user k (BER_k) and is approximated as $\Gamma_k \cong -\ln(5\,\mathrm{BER}_k/1.5)$ for QAM modulations [10]. We assume a Rayleigh channel model. Hence $\gamma_{k,m}$ is a *central chisquared* (χ^2) distributed random variable with two degrees of freedom and with probability density function (pdf)

$$T_{\gamma_{k,m}}(\gamma_{k,m}) = \frac{1}{\overline{\gamma}_{k,m}} e^{-\gamma_{k,m}/\overline{\gamma}_{k,m}}, \qquad (2)$$

where $\overline{\gamma}_{k,m}$ is the mean of the $\gamma_{k,m}$ distribution.

Each user is adaptively assigned a number of different subchannels to send and receive data. An indicator $\rho_{k,m}$ is

used to represent whether the mth subchannel is assigned to user k. Note that in a single cell OFDMA system, each subchannel can be assigned to at most one user at any time, that is, $\sum_{k=1}^{K} \rho_{k,m} \in \{0, 1\}$ for all m. Due to the consideration for the reduction of the signaling overhead in WIMAX, the power is equally distributed across subcarriers within each subchannel. We assume the duration of the transmission codewords is long enough to undergo all channel realizations. We further assume perfect CDIT, thereby allowing to take the expectation over the distribution. The ergodic weighted-sum rate of the multiuser system is defined as

$$U_\gamma = E_\gamma \left\{ \sum_{k=1}^{K} \frac{1}{\overline{R}_k^{\alpha_f}} \sum_{m=1}^{M} \rho_{k,m} \log_2 \left(1 + p_{k,m} \gamma_{k,m}\right) \right\}, \quad (3)$$

where $\gamma = [\gamma_1^T, \ldots, \gamma_K^T]^T$ with $\gamma_k = [\gamma_{k,1}, \ldots, \gamma_{k,M}]$, $p_{k,m}$ denotes the power allocated to the user k on subchannel m, $E_\gamma\{\cdot\}$ represents the statistical expectation with respect to γ, $\overline{R}_k$ is user k's average data rate so far at the allocation time, and α_f is a tunable fairness parameter. Setting α_f to 1 results in the *proportional fair* allocation. For $\alpha_f = 0$, this results in *maximum throughput* allocation. The average user rates $\overline{R}_k$ are updated according to

$$\overline{R}_k(t+1) = \left(1 - \frac{1}{\tau_c}\right)\overline{R}_k(t) + \frac{1}{\tau_c} r_k(t), \quad (4)$$

where $r_k(t)$ is the rate allocated to user k at time t and τ_c is the parameter that controls the latency of the system. This way, we consider both current rate as well as rates given to the users in the past, what is suitable for long-term fairness evaluation.

3. CDIT-Based Constrained Resource Allocation

3.1. Formulation of the Problem. The issue is how to adaptively assign the M subchannels to the K users and distribute the total power budget P_{tot} in order to maximize the ergodic weighted-sum rate (3) while satisfying user's minimum rate and system fairness requirements under a total power constraint. Mathematically, this constrained optimization problem is formulated as

$$f^* = \max_{\rho_{k,m}, p_{k,m}} E_\gamma \left\{ \sum_{k=1}^{K} \frac{1}{\overline{R}_k^{\alpha_f}} \sum_{m=1}^{M} \rho_{k,m} \log_2 \left(1 + p_{k,m} \gamma_{k,m}\right) \right\},$$

subject to

$$E_\gamma \left\{ \sum_{m=1}^{M} \rho_{k,m} \log_2 \left(1 + p_{k,m} \gamma_{k,m}\right) \right\} \geq R_k, \quad (5)$$

$$E_\gamma \left\{ \sum_{k=1}^{K} \sum_{m=1}^{M} \rho_{k,m} p_{k,m} \right\} \leq P_{\text{tot}}.$$

The first constraint in (5) is for the user's specific minimum data rate demand. We assume that appropriate admission control is performed such that the minimum data rates R_k

are feasible. The second constraint is system limitation on transmits powers.

Note that the optimization problem (5) involves both continuous variables $p_{k,m}$ and boolean variable $\rho_{k,m}$. Such an optimization problem is neither convex nor concave with respect to $(\rho_{k,m}, p_{k,m})$.

3.2. Solution Based on Lagrange Dual Decomposition. We can solve problem (5) using the Lagrange dual decomposition framework. Following the approach in [12], we relax $\rho_{k,m}$ to be $\rho_{k,m} \in [0, 1]$. Then $\rho_{k,m}$ can be regarded as time-sharing factor. Thanks to the linearity property of the expectation, the Lagrangian function of the primal problem (5) can be expressed as

$$
\begin{aligned}
&L_\gamma\left(p_{k,m}, \rho_{k,m}, \lambda_k, \mu\right) \\
&= \sum_{k=1}^{K} \sum_{m=1}^{M} E_{\gamma_{k,m}} \left\{ \frac{\rho_{k,m}}{\overline{R}_k^{\alpha_f}} \log_2 \left(1 + p_{k,m} \gamma_{k,m}\right) \right\} \\
&\quad + \sum_{k=1}^{K} \lambda_k \sum_{m=1}^{M} E_{\gamma_{k,m}} \left\{ \rho_{k,m} \log_2 \left(1 + p_{k,m} \gamma_{k,m}\right) \right\} \\
&\quad - \mu \sum_{k=1}^{K} \sum_{m=1}^{M} \rho_{k,m} p_{k,m} - \sum_{k=1}^{K} \lambda_k R_k + \mu P_{\text{tot}},
\end{aligned}
\quad (6)
$$

where λ_k and μ are Lagrangian multipliers. Let $p_{k,m}^*$ and $\rho_{k,m}^*$ denote the optimal solution of (6). We first investigate the problem for fixed values of λ_k and μ. By Karush-Kuhn-Tucker (KKT) first optimality condition [16], $\rho_{k,m}^*$ and $p_{k,m}^*$ should satisfy the following:

$$\left. \frac{\partial L_\gamma}{\partial p_{k,m}} \right|_{p_{k,m}=p_{k,m}^*} \begin{cases} < 0, & p_{k,m}^* = 0, \\ = 0, & p_{k,m}^* > 0, \end{cases} \quad (7)$$

$$\left. \frac{\partial L_\gamma}{\partial \rho_{k,m}} \right|_{\rho_{k,m}=\rho_{k,m}^*} \begin{cases} < 0, & \rho_{k,m}^* = 0, \\ = 0, & 0 < \rho_{k,m}^* < 1, \\ > 0, & \rho_{k,m}^* = 1. \end{cases} \quad (8)$$

For a nonzero power allocation and $\rho_{k,m}^* \in (0, 1)$, we obtain from (7) and (8)

$$\rho_{k,m}^* E_{\gamma_{k,m}} \left\{ \frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k}{\ln 2} \cdot \frac{\gamma_{k,m}}{1 + p_{k,m}^* \gamma_{k,m}} - \mu \right\} = 0, \quad (9)$$

$$E_{\gamma_{k,m}} \left\{ \frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k}{\ln 2} \cdot \ln\left(1 + p_{k,m}^* \gamma_{k,m}\right) - \mu p_{k,m}^* \right\} = 0. \quad (10)$$

We deduce from (9) that $p_{k,m}^*$ has to satisfy the following condition:

$$\rho_{k,m}^* E_{\gamma_{k,m}} \left\{ \frac{\gamma_{k,m}}{1 + p_{k,m}^* \gamma_{k,m}} - \frac{\mu \ln 2}{A_k} \right\} = 0, \quad (11)$$

where $A_k = \overline{R}_k^{(-\alpha_f)} + \lambda_k$.

When $\rho_{k,m}^* = 0$, the value of $p_{k,m}$ is undefined, and any value can be taken without any influence on the objective

function or on the constraints. On the other hand, for any other positive value, $\rho_{k,m}$ vanishes out of the expression and we get

$$E_{\gamma_{k,m}}\left\{\frac{\gamma_{k,m}}{1 + p_{k,m}^*\gamma_{k,m}} - \frac{\mu \ln 2}{A_k}\right\} = 0. \quad (12)$$

We can use the pdf of the SNR distribution (2) to transform (12) into

$$\int_0^\infty \left(\frac{\gamma_{k,m}}{1 + p_{k,m}^*\gamma_{k,m}} - \frac{\mu \ln 2}{A_k}\right) \cdot \frac{1}{\overline{\gamma}_{k,m}} e^{-\gamma_{k,m}/\overline{\gamma}_{k,m}} d\gamma_{k,m} = 0, \quad (13)$$

which is equivalent to

$$\frac{\mu \ln 2}{A_k} p_{k,m}^{*2} \overline{\gamma}_{k,m} - p_{k,m}^* \overline{\gamma}_{k,m}$$
$$+ \exp\left(\frac{1}{p_{k,m}^*\overline{\gamma}_{k,m}}\right) E_1\left(\frac{1}{p_{k,m}^*\overline{\gamma}_{k,m}}\right) = 0, \quad (14)$$

where

$$E_1(x) = \int_1^\infty \frac{e^{-tx}}{t} dt \quad (15)$$

is the exponential integral function of x [17].

Equation (10) is equivalent to

$$E_{\gamma_{k,m}}\left\{\frac{A_k}{\mu \ln 2} \ln\left(1 + p_{k,m}^*\gamma_{k,m}\right) - p_{k,m}^*\right\} = 0. \quad (16)$$

Using the pdf (2), (16) can be transformed into

$$\int_0^\infty \left(\frac{A_k}{\mu \ln 2} \ln\left(1 + p_{k,m}^*\gamma_{k,m}\right) - p_{k,m}^*\right)$$
$$\cdot \frac{1}{\overline{\gamma}_{k,m}} e^{-\gamma_{k,m}/\overline{\gamma}_{k,m}} d\gamma_{k,m} = 0, \quad (17)$$

which is finally equivalent to

$$\frac{A_k}{\mu \ln 2} \exp\left(\frac{1}{p_{k,m}^*\overline{\gamma}_{k,m}}\right) E_1\left(\frac{1}{p_{k,m}^*\overline{\gamma}_{k,m}}\right) - p_{k,m}^* = 0. \quad (18)$$

From (14) and (18), we derive

$$p_{k,m}^* = \left[\frac{A_k}{\mu \ln 2} - \frac{1}{\overline{\gamma}_{k,m}}\right]^+, \quad (19)$$

where $[x]^+ = \max(0, x)$. The expression (19) is in the form of *multilevel water-filling* power allocation with cut-off subchannel SNR $(\mu \ln 2)/A_k$ below which we do not transmit any power, and above which we transmit more power when $\overline{\gamma}_{k,m}$ is higher. The important difference is that, in contrast to the CSIT-based allocation where the $p_{k,m}^*$ depends on the instantaneous channels realizations, the optimal allocation here is dependent on the mean of the channel distribution, and thus needs to be computed only when the statistics of the channel has changed.

We have from (8) and (18) that

$$\rho_{k,m}^* = \begin{cases} 1, & \text{if } G_{k,m}(p_{k,m}^*) > 0, \\ \in (0, 1), & \text{if } G_{k,m}(p_{k,m}^*) = 0, \\ 0, & \text{if } G_{k,m}(p_{k,m}^*) < 0, \end{cases} \quad (20)$$

where

$$G_{k,m}(p_{k,m}^*) = \frac{A_k}{\mu \ln 2} \exp\left(\frac{1}{p_{k,m}^*\overline{\gamma}_{k,m}}\right) E_1\left(\frac{1}{p_{k,m}^*\overline{\gamma}_{k,m}}\right) - p_{k,m}^*. \quad (21)$$

Due to the exclusive subchannel assignment constraint in OFDMA, we can conclude that for each subchannel m, if $G_{k,m}$ are all different, then only the user with the largest $G_{k,m}$ can use that subchannel. In other words,

$$\rho_{k_m^*,m}^* = 1, \quad \rho_{k,m}^* = 0, \quad \forall k \neq k_m^*, \quad (22)$$

where

$$k_m^* = \arg\max_k G_{k,m}(p_{k,m}^*). \quad (23)$$

Substituting (19) into the Lagrange function (6) and thanks to the exclusive subchannel assignment constraint, we obtain the following per-subchannel dual problem:

$$g^* = \min_{\lambda_k, \mu} L_\gamma^*(\lambda_k, \mu), \quad (24)$$

where $L_\gamma^*(\lambda_k, \mu)$ is the dual function given by

$$L_\gamma^*(\lambda_k, \mu) = E_{\gamma_{k,m}}\left\{\overline{R}_k^{(-\alpha_f)} \log_2\left(\frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k}{\mu \ln 2} \overline{\gamma}_{k,m}\right)\right\}$$
$$+ \lambda_k E_{\gamma_{k,m}}\left\{\log_2\left(\frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k}{\mu \ln 2} \overline{\gamma}_{k,m}\right)\right\} \quad (25)$$
$$- \lambda_k R_k - \mu\left(\frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k}{\mu \ln 2} - \frac{1}{\overline{\gamma}_{k,m}}\right) + \mu P_{\text{tot}}.$$

Next we turn to the optimization of the dual function (25) over μ and λ_k. First we consider the optimization over λ_k for μ fixed to find. We differentiate (25) with respect to λ_k and set the derivative to 0 to obtain

$$\left.\frac{\partial L_\gamma^*}{\partial \lambda_k}\right|_{\lambda_k = \lambda_k^*} = \log_2\left(\frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k}{\mu \ln 2} \overline{\gamma}_{k,m}\right) - R_k = 0. \quad (26)$$

The optimum λ_k^* is derived from (26) as follows:

$$\lambda_k^*(\mu) = 2^{R_k} \frac{\mu \ln 2}{\overline{\gamma}_{k,m}} - \frac{1}{\overline{R}_k^{(\alpha_f)}}. \quad (27)$$

If some of the individual rate constraints are exceeded, the corresponding λ_k is equal to zero.

Substituting (27) into (25) we obtain

$$L_\gamma^*(\mu) = \max_{\lambda_k} L_\gamma^*(\lambda_k, \mu) = \frac{R_k}{\overline{R}_k^{(\alpha_f)}} - \frac{2^{R_k}\mu}{\overline{\gamma}_{k,m}} + \frac{\mu}{\overline{\gamma}_{k,m}} + \mu P_{\text{tot}}. \tag{28}$$

We next consider the optimization of $L_\gamma^*(\mu)$ over μ. The function $L_\gamma^*(\mu)$ can be shown to be a convex function of μ, which can then be minimized via a one-dimensional search with geometric convergence. The optimal values μ^* correspond to the ones that satisfy the total power constraint (with equality).

We can conclude that, if $G_{k,m}$ are all different, then a given subchannel m is exclusively assigned to the user $\tilde{k}_m^*$ satisfying

$$\tilde{k}_m^* = \arg\max_k G_{k,m}(\tilde{p}_{k,m}^*), \tag{29}$$

where $\tilde{p}_{k,m}^*$ is the optimal power allocation given by

$$\tilde{p}_{k,m}^* = \left[\frac{\overline{R}_k^{(-\alpha_f)} + \lambda_k^*}{\mu^* \ln 2} - \frac{1}{\tilde{\gamma}_{k,m}} \right]^+, \quad \text{if } k = \tilde{k}_m^*,$$
$$= 0, \quad \text{if } k \neq \tilde{k}_m^*. \tag{30}$$

3.3. Relative Duality Gap.

The relative duality (optimality) gap which indicates how far we are from the optimal value can be expressed as

$$d^* = \frac{g^* - f^*}{f^*} \geq 0, \tag{31}$$

where $f^* > 0$ and $g^* > 0$ given in (5) and (24) are the optimal values of the primal and dual problems. The inequality follows from the positivity of f^* and the weak duality theorem [18].

Without the minimum rate constraints in (5), problem (5) becomes a standard convex optimization problem, and then the duality gap is zero. Due to the nonlinearity of the minimum rate constraints, the convexity of problem (5) does not hold. However, the nonconvex optimization problem (5) for the investigated OFDMA-based WIMAX system fulfills the *time-sharing* condition as defined in [19]. Then when the power constraint is met tightly, that is, with equality, the duality gap is zero, and thus solving the dual problem (24) also solves the primal problem (5).

3.4. Instantaneous Resource Allocation Based on CSIT.

In order to assert the relevance of our approach, it was decided to compare it to the instantaneous allocation based on partial CSIT and to the instantaneous allocation based on perfect CSIT.

3.4.1. Resource Allocation Based on Partial CSIT.

Assuming that partial channel state information is available at the transmitter in the form of an estimate of the SNR, it has been shown that resources can be optimally allocated based on this partial CSIT (see, e.g., [13, 14]). Let $\gamma_{k,m}$ and $\hat{\gamma}_{k,m}$ denote the real and the estimated subchannel SNR. For Rayleigh fading channels, $\gamma_{k,m}$ conditioned on $\hat{\gamma}_{k,m}$ is a *noncentral chi-squared* distributed random variable with two degrees of freedom [13, 14]. Its probability density function (pdf) can be approximated to a *Gamma* function as

$$T_{\gamma_{k,m}}(\gamma_{k,m} \mid \hat{\gamma}_{k,m}) \approx \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma_{k,m}^{(\alpha-1)} e^{-\beta\gamma_{k,m}}. \tag{32}$$

In expression (32), $\alpha = (\hat{\gamma}_{k,m}\gamma_{e/n}^{-1} + 1)^2/(2\hat{\gamma}_{k,m}\gamma_{e/n}^{-1}+1)$ and $\beta = \alpha/(\hat{\gamma}_{k,m}+\gamma_{e/n})$ are the shape parameter and the rate parameter of the *Gamma* pdf, respectively, where $\gamma_{e/n}$ is the ratio of the subchannel estimation error variance to the background noise variance. $\Gamma(x)$ is the *Gamma* function of x.

Under the partial CSIT assumption, the optimization goal is to maximize the expected weighted-sum rate instead of the ergodic weighted-sum rate. In [14], the problem has been formulated as

$$\max_{p_{k,m}, \rho_{k,m}} E_\gamma \left\{ \sum_{k=1}^K \sum_{m=1}^M \frac{\rho_{k,m}}{\overline{R}_k^{\alpha_f}} \log_2(1 + p_{k,m}\gamma_{k,m}) \mid \hat{\gamma}_{k,m} \right\}, \tag{33}$$

subject to

$$E_\gamma \left\{ \sum_{m=1}^M \rho_{k,m} \log_2(1 + p_{k,m}\gamma_{k,m}) \mid \hat{\gamma}_{k,m} \right\} \geq R_k,$$
$$\sum_{k=1}^K \sum_{m=1}^M p_{k,m} \leq P_{\text{tot}}. \tag{34}$$

Using the pdf (32) and applying the KKT optimality conditions, it has been shown in [14] that the optimal power allocation $p_{k,m}^*$ is the solution of

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\gamma_{k,m}^\alpha}{1 + p_{k,m}^*\gamma_{k,m}} e^{-\beta\gamma_{k,m}} d\gamma_{k,m} - \frac{\mu\ln 2}{A_k} = 0. \tag{35}$$

Also by KKT optimality conditions, it has been shown in [14] that a given subchannel m is exclusively assigned to the user k_m^* satisfying

$$k_m^* = \arg\max_k A_k G_k(p_{k,m}^*), \tag{36}$$

where

$$G_k(p_{k,m}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \gamma_{k,m}^{(\alpha-1)} \log_2(1 + p_{k,m}\gamma_{k,m})$$
$$\times e^{-\beta\gamma_{k,m}} d\gamma_{k,m} - \mu p_{k,m}. \tag{37}$$

3.4.2. Resource Allocation Based on Perfect CSIT.

Under the *unrealistic* perfect CSIT assumption, instead of maximizing the ergodic or the expected weighted-sum rate, the optimization goal is to maximize the instantaneous weighted-sum rate

$$\max_{p_{k,m}, \rho_{k,m}} \sum_{k=1}^K \frac{1}{\overline{R}_k^{\alpha_f}} \sum_{m=1}^M \rho_{k,m} \log_2(1 + p_{k,m}\gamma_{k,m}), \tag{38}$$

subject to

$$\sum_{m=1}^{M} \rho_{k,m} \log_2 (1 + p_{k,m} \gamma_{k,m}) \geq R_k,$$

$$\sum_{k=1}^{K} \sum_{m=1}^{M} p_{k,m} \leq P_{\text{tot}}.$$

(39)

From the KKT optimality conditions, the optimal power allocation, solution of (39) is given by

$$\widetilde{p}_{k,m}^* = \left[\frac{A_k}{\mu \ln 2} - \frac{1}{\gamma_{k,m}} \right]^+.$$

(40)

This is a *multilevel water-filling* power allocation with cut-off subchannel SNR $(\mu \ln 2)/A_k$. The difference between (40) and (19) is that the power allocation in (40) depends on the instantaneous subchannel SNR $\gamma_{k,m}$ while the one in (19) depends on the mean of the SNR distribution $\overline{\gamma}_{k,m}$.

We also deduce from KKT optimality conditions [14] that a given subchannel m is exclusively assigned to the user $\widetilde{k}_m^*$ ($\rho_{\widetilde{k}_m^*,m} = 1$, $\rho_{k,m} = 0$ for $k \neq \widetilde{k}_m^*$) satisfying

$$\widetilde{k}_m^* = \max_k A_k (\log_2 (1 + p_{k,m}^* \widehat{\gamma}_{k,m}) - \mu p_{k,m}^*).$$

(41)

3.5. Feedback Reduction and Complexity Analysis. First, thanks to the subchannelization, the need for knowledge of radio channel quality in mobile WiMAX is reduced from per-subcarrier to per-subchannel resolution and resources are allocated on per-subchannel basis. Second, under CDIT-based allocation, instead of feeding back the instantaneous channel coefficients to the transmitter, the users simply feed back the mean of the subchannel SNR distribution. Putting these two facts together, the amount of feedback required for the resource allocation reduces significantly.

Using a dual decomposition framework, the optimization problem has been reduced to a per-subchannel optimization, and the computational complexity has been significantly decreased.

Since the optimal λ_k^* and μ^* depend on the mean of the subchannel SNR distribution and not on the instantaneous values, they need to be computed only when the statistic of the channel has changed. We need to run the line search to compute for μ^*. This is followed by the computation of KM values of multipliers λ_k^* (27) and power allocation values $\widetilde{p}_{k,m}^*$ (30). We assume the line search to require I_μ iterations. The computation of λ_k^* and $\widetilde{p}_{k,m}^*$ requires $\mathcal{O}(KM)$ operations. The overall complexity order for the CDIT-based resource allocation is thus $\mathcal{O}(KMI_\mu)$. Since I_μ is just constant independent of K and M, the complexity is $\mathcal{O}(KM)$. Once μ^*, λ_k^*, and $\widetilde{p}_{k,m}^*$ have been determined, we do not need to update them as long as the statistics of the fading channel remains the same.

Both expressions (19) and (40) are in the form of *multilevel water-filling* power allocation with cut-off subchannel SNR $(\mu \ln 2)/A_k$. Thus, the complexity in term of water filling is the same. The main difference between (19) and (40) is the amount of feedback required to perform resource allocation. Recall that, for the CSIT-based scheme,

the allocation is performed after each OFDM symbol period. Let N_s be the number of OFDM symbol periods after which the CDIT-based resource allocation is performed. Then a rough estimation tells us that the complexity of the CDIT-based allocation is reduced by $1/N_s$ compared to the perfect CSIT scheme.

3.6. Tradeoff Analysis. In the tradeoff analysis, we vary the constraints, and see the effect on the maximized weighted-sum rate. We define a relaxation coefficient η_q for the minimum rate constraint and replace the minimum rate requirement R_k by $(1 - \eta_q)R_k$ to form a perturbed problem. When $\eta_q = 0$, this reduces to the original problem (5). By increasing η_q, the minimum rate constraints are relaxed. We can also vary the fairness parameter α_f. Setting α_f to 0 results in the *maximum throughput* allocation. For $\alpha_f = 1$, this results in the *proportional fair* allocation. The relaxation of the constraints leads in general to an improvement of the optimal objective. The tradeoff curves are found by solving the perturbed problem for many values of η_q and α_f.

4. Simulation Results

To illustrate the performance of the proposed resource allocation method, we perform simulations for a three-users mobile WiMAX system with bandwidth divided into $M = 8$ subchannels. The subchannels are formed using the FUSC which is suitable for mobile applications. The FFT size of the OFDMA is 512 points. The performance is evaluated in multipath channel environments modeled as a tapped delay line with six taps as specified in the ITU M.1225 Vehicular A channel model [20]. We consider a scenario where user 2 is every time closer to the base station than users 1 and 3 and the relative mean SNR difference between user 2 and users 1 and 3 is -5 dB and -3 dB, respectively, while the minimum data rate demand of user 3 is higher than the one of users 2 and 1 ($R_3 > R_2 > R_1$). The target bit error rate is set to 10^{-3} (without channel coding). The performances are evaluated using simulations over 10 000 instances of independent channel realizations. For all the simulations, the total power is set to $P_{\text{tot}} = KM$.

In Figure 1, the performance of the proposed adaptive resource allocation is compared to those of optimal resource allocation based on perfect CSIT, resource allocation based on partial CSIT and a uniform power allocation. The result shows that the proposed adaptive resource allocation brings significant gain over resource allocation based on partial CSIT with higher estimation error. We can observe that when $\gamma_{e/n}$ is small, that means the effect of the estimation error is less dominant than the one of the background noise, the optimization under partial CSI is closed to the one under perfect CDIT. For very low estimation errors, the partial CSIT-based scheme outperforms the perfect CDIT scheme. The weighted-sum rate degrades quickly as the estimation error grows, especially for high SNRs. The highest weighted-sum rate is obtained with perfect CSIT but the
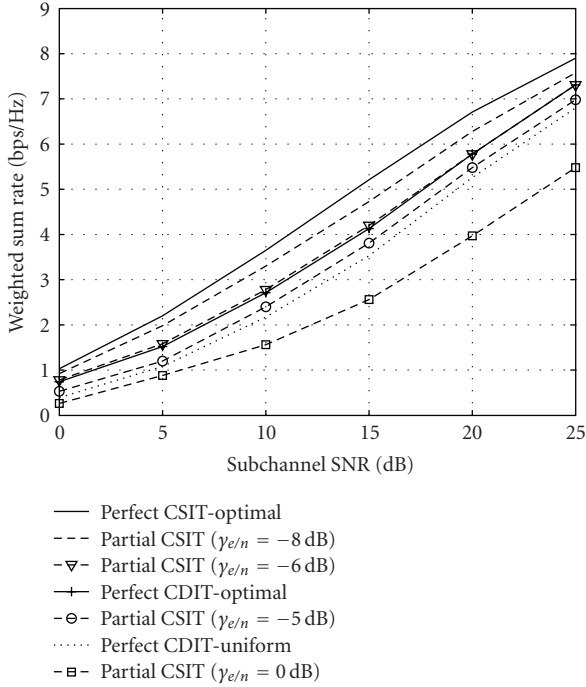
FIGURE 1: Maximized weighted-sum rate versus mean SNR for different resource allocation schemes and fairness parameter $\alpha_f = 0$.

FIGURE 3: Tradeoff between maximized weighted-sum rate and fairness requirement for different resource allocation schemes and mean SNR of 15 dB.
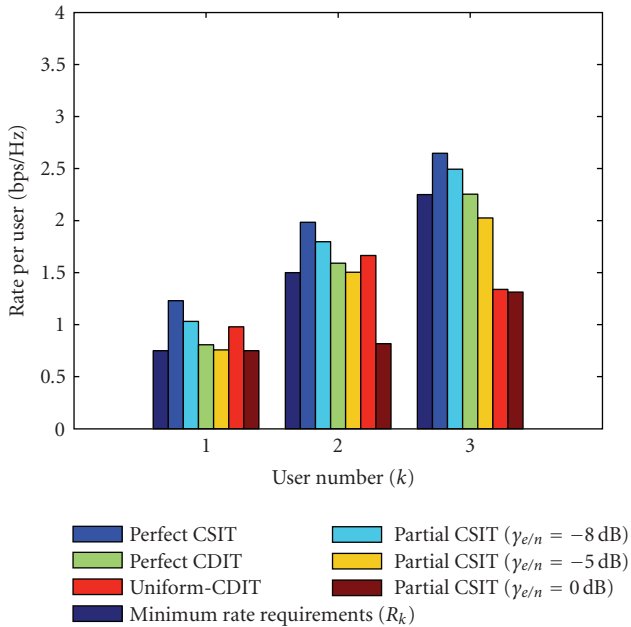

FIGURE 2: Users rates for different resource allocation schemes, mean SNR of 15 dB and fairness parameter $\alpha_f = 0$.

difference in terms of performance is not so significant compared to the difference of complexity between CDIT-based and CSIT-based allocation schemes. The proposed method outperforms the uniform power allocation.

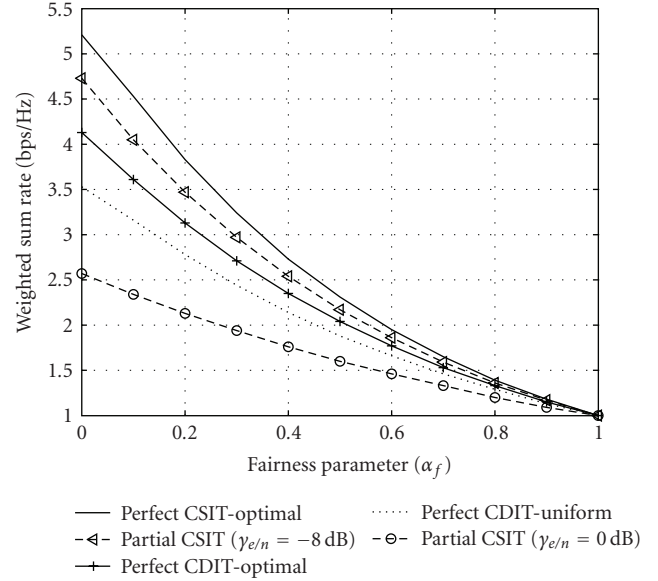Figure 2 shows the user's rate for different allocation schemes when the users minimum data rate demands are constrained to $R_3 = 2R_2 = 3R_1$ and the fairness parameter α_f is set to 0. We observe that under optimal allocation based on perfect CDIT, the need of all users in terms of data rate is satisfied. This is neither the case under allocation based on partial CSIT with high estimation error nor under uniform allocation where the high data rate demand of user 3 is not satisfied.

Figure 3 illustrates the tradeoff between the maximized weighted-sum rate and the fairness constraint when the minimum rate demand is relaxed to $(1 - \eta_q)R_k$ with $0 \leq \eta_q \leq 0.8$. The average user rates are updated according to (4) with $\tau_c = 20$. The maximum weighted-sum rate is achieved when $\alpha_f = 0$ which is very unfair. We can see that, as the fairness constraint is enforced, the weighted-sum rate decreases. For $\alpha_f = 1$, the allocation is strictly fair but inefficient in terms of sum rate. Looking at the solution obtained for different values of α_f, the system designer may then make a choice about the configuration he considers to be the most appropriate.

The tradeoff between reduced complexity and performance degradation of the proposed CDIT-based resource allocation in comparison with the perfect CSIT allocation is shown in Figure 4. Adapting the transmission strategy to the short-term channel statistics, that is, reducing N_s increases the performance but also increases the complexity. However, if the transmission is adapted to the long-term channel statistics, that is, for larger N_s, the complexity decreases significantly but with a penalty on the performance. For a CDIT-based allocation with a distribution taken over 16 OFDM symbol periods, the complexity is reduced by 93.75% while the performance degradation in terms of weighted-sum rate is less than 15%.
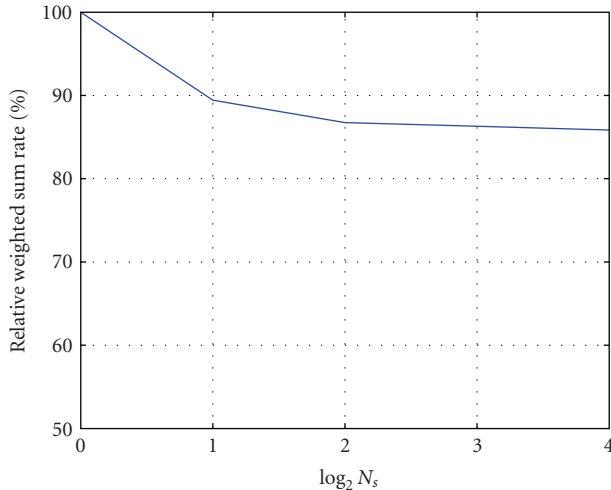
FIGURE 4: Tradeoff between reduced complexity and performance degradation of the CDIT-based allocation compared with the perfect CSIT.

5. Conclusion

In this paper, we have presented a resource allocation method that maximizes the ergodic weighted-sum rate of a multiuser mobile WiMAX while satisfying user's specific minimum rate demand and system fairness requirement for a given power budget. Though this is originally a nonlinear optimization problem, the problem can be reformulated as a Lagrangian dual problem. From this, a method has been proposed to efficiently solve the problem. The proposed method can find the optimal solution with significant lower computational complexity than the optimal CSIT-based allocation schemes. In fading environments, even with CDIT only, adaptive resource allocation strategies provide performance gain for OFDMA systems. Since user mobility is the principal driving force for mobile WiMAX, CDIT-based resource allocation strategies are of particular interest. These methods can be applied to other mobile OFDMA-based wireless systems such as Long Term Evolution (LTE) or High-Speed Downlink Packet Access (HSDPA).

Acknowledgment

References

[1] WiMAX Forum, "Mobile WiMAX—part I: a technical overview and performance evaluation," March 2006.

[2] WiMAX Forum, "A comparative analysis of mobile WiMAX deployment alternatives in the access networks," May 2007.

[3] S. Pietrzyk, *OFDMA for Broadband Wireless Access*, Artech House, London, UK, 2006.

[4] C. Y. Wong, C. Y. Tsui, R. S. Cheng, and K. B. Letaief, "A real-time sub-carrier allocation scheme for multiple access downlink OFDM transmission," in *Proceedings of the 50th IEEE Vehicular Technology Conference (VTC '99)*, vol. 2, pp. 1124–1128, Amsterdam, The Netherlands, September 1999.

[5] 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.

[6] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, 2003.

[7] N. Damji and T. Le-Ngoc, "Adaptive downlink resource allocation strategies for real-time data services in OFDM cellular systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, Article ID 17526, 11 pages, 2006.

[8] T. Sartenaer, L. Vandendorpe, and J. Louveaux, "Balanced capacity of wireline multiuser channels," *IEEE Transactions on Communications*, vol. 53, no. 12, pp. 2029–2042, 2005.

[9] T. Sartenaer, J. Louveaux, and L. Vandendorpe, "Balanced capacity of wireline multiple access channels with individual power constraints," *IEEE Transactions on Communications*, vol. 56, no. 6, pp. 925–936, 2008.

[10] I. C. Wong, Z. Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems," in *Proceedings of IEEE Workshop on Signal Processing Systems (SiPS '04)*, pp. 1–6, Austin, Tex, USA, October 2004.

[11] X. Zhang, E. Zhou, R. Zhu, S. Liu, and W. Wang, "Adaptive multiuser radio resource allocation for OFDMA systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 6, pp. 3846–3850, St. Louis, Mo, USA, November 2005.

[12] Z. Han, Z. Ji, and K. J. R. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1366–1376, 2005.

[13] Y. Yao and G. B. Giannakis, "Rate-maximizing power allocation in OFDM based on partial channel knowledge," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1073–1083, 2005.

[14] F. Brah, L. Vandendorpe, and J. Louveaux, "Constrained resource allocation in OFDMA downlink systems with partial CSIT," in *Proceedings of IEEE International Conference on Communications (ICC '08)*, pp. 4144–4148, Beijing, China, May 2008.

[15] C. T. K. Ng and A. J. Goldsmith, "Capacity and power allocation for transmitter and receiver cooperation in fading channels," in *Proceedings of IEEE International Conference on Communications (ICC '06)*, vol. 8, pp. 3741–3746, Istanbul, Turkey, July 2006.

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

[17] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, NY, USA, 1964.

[18] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Boston, Mass, USA, 2nd edition, 1999.

[19] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, 2006.

[20] Recommendation ITU-R M.1225, "Guidelines for evaluation of radio transmission technologies for IMT-2000," 1997.