Optimal Resource Allocation in OFDMA Systems with Imperfect Channel Knowledge

Ian C. Wong, *Member, IEEE* and Brian L. Evans, *Senior Member, IEEE*

***Abstract*—Previous research efforts on OFDMA resource allocation have typically assumed the availability of perfect channel state information (CSI). Unfortunately, this is unrealistic, primarily due to channel estimation errors, and more importantly, channel feedback delay. In this paper, we develop optimal resource allocation algorithms for OFDMA systems assuming the availability of only partial (imperfect) CSI. We consider both continuous and discrete ergodic weighted sum rate maximization subject to total power constraints, and average bit error rate constraints for the discrete rate case. We approach these problems using a dual optimization framework, allowing us to solve these problems with $\mathcal{O}(MK)$ complexity per symbol for an OFDMA system with K used subcarriers and M active users, while achieving relative optimality gaps of less than 10^{-5} for continuous rates and less than 10^{-3} for discrete rates in simulations based on realistic parameters.**

I. Introduction

Next generation wireless systems, e.g. IEEE 802.16e [1] and 3GPP-Long Term Evolution (LTE) [2], consider Orthogonal Frequency Division Multiple Access (OFDMA) as the preferred physical layer multiple access scheme, especially for the downlink. OFDMA allows multiple users to transmit simultaneously on the different subcarriers per OFDM symbol. It is reasonable to assume that the channel response for each user is statistically independent, especially when there is considerable spatial separation among the users. Thus, we could potentially exploit this *multiuser diversity* through intelligent allocation of the subcarriers and power to each user, and increase the overall performance of the system.

The problem of assigning the subcarriers, rates, and powers to the different users in an OFDMA system has been an area of active research over the past several years (see e.g. [3] [4] [5] [6] [7] [8] [9]). These works considered various formulations (e.g. rate maximization and power minimization) while considering various notions of fairness (e.g. max-min fairness, proportional data rates and utility-based fairness). A common underlying assumption among these works is that the channel state information (CSI) of the users are known perfectly. This assumption is quite unrealistic due to channel estimation errors, and more importantly, channel feedback delay. Thus, in this paper, we focus on the case where only imperfect (partial) CSI is available.

The effect of imperfect CSI for rate maximization in wireless systems has been quite well studied for single-user wireless systems. In [10], adaptive trellis-coded modulation schemes using a single outdated channel estimate for single-carrier systems in flat-fading channels were proposed. In [11], uncoded adaptive modulation schemes using predicted CSI were developed, also for single-carrier flat-fading channels. In [12] and [13], the effect of channel estimation errors and channel feedback delay on adaptive modulation for OFDM systems in time and frequency selective channels was studied. It was concluded that the detrimental effect of outdated channel information is significant, and that using channel prediction [12] or using multiple channel estimates [13] is a viable way of overcoming this delay. In [14], power allocation methods for ergodic and outage capacity maximization in OFDM [14] were studied assuming that the partial CSI distribution information is available. Adaptive modulation in single-user single-carrier Multiple-Input, Multiple-Output (MIMO) systems [15] and MIMO-OFDM systems [16] assuming imperfect CSI have also been investigated. However, no work to the best of the authors's knowledge considered the multiuser OFDM case.

In the multiuser OFDM (or OFDMA) case, the difficulty arises from the fact that the exclusive subcarrier assignment restriction (i.e. only one user is allowed to transmit on each subcarrier) renders the problem to be combinatorial in nature. This difficulty is evident even under the perfect CSI assumption in the previous approaches, which advocated highly complex convex reformulations [3] [7] to solve the problem near-optimally; or lower complexity suboptimal greedy heuristics to solve the problem in a practical setting [4] [5] [8]. Fortunately, in our recent work on optimal resource allocation for ergodic rate maximization in OFDMA systems with perfect CSI [17], we have shown that using a dual optimization approach, the problem can be solved with just $\mathcal{O}(MK)$ complexity per symbol for an OFDMA system with M active users and K used subcarriers. Furthermore, our solution results in relative optimality gaps of less than 10^{-4} in typical scenarios, thereby supporting us to claim *practical optimality*. Using a similar dual optimization approach, we relax the assumption of perfect CSI and formulate and solve the problem assuming the availability of imperfect CSI. We use the statistics of this imperfect CSI to perform resource allocation for both continuous rate (capacity based) and discrete rate (adaptive modulation and coding based) maximization cases. We considered minimum mean square error (MMSE) OFDM channel prediction in this paper, but the framework can be easily extended to other estimation/prediction approaches as well. We show that by using the dual optimization framework, we can solve the imperfect CSI problem with relative optimality gaps of less than 10^{-5} for continuous rates and less than 10^{-3} for discrete rates in cases of practical interest.

This paper is organized as follows. Section II discusses the system model used in the paper, and the assumptions on the channel model. Section III discusses the optimal resource allocation algorithms for the continuous rate case

(ergodic (Shannon) capacity) assuming partial CSI. Section IV considers the more practically relevant case of allocation for discrete rates (adaptive modulation). Section V presents several numerical examples to solidify the claims in the paper.

II. System Model

A. OFDMA Signal Model

We consider a single $K_{\text{fft}}-$subcarrier with L_{cp} cyclic-prefix length OFDMA base station, with K used subcarriers and M active users indexed by the set $\mathcal{K} = \{1, \ldots, k, \ldots, K\}$ and $\mathcal{M} = \{1, \ldots, m, \ldots, M\}$ (typically $K \gg M$) respectively. We assume an average transmit power of $\bar{P} > 0$, bandwidth B, and noise density N_0. The received signal vector for the mth user at the nth OFDM symbol assuming perfect sample and symbol synchronization, and sufficient cyclic prefix length, is given as

$$\boldsymbol{y}_m[n] = \mathbf{G}_m[n]\mathbf{H}_m[n]\boldsymbol{x}_m[n] + \boldsymbol{w}_m[n] \tag{1}$$

where $\boldsymbol{y}_m[n]$ and $\boldsymbol{x}_m[n]$ are the K-length received and transmitted complex-valued signal vectors; $\mathbf{G}_m[n] = \text{diag}\left\{\sqrt{p_{m,1}[n]}, \ldots, \sqrt{p_{m,K}[n]}\right\}$ is the diagonal gain allocation matrix; $\boldsymbol{w}_m[n] \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \boldsymbol{I}_K)$ with noise variance $\sigma_w^2 = N_0 B/K$ is the spectrally and temporally white, zero-mean, circular-symmetric, complex Gaussian (ZMCSCG) noise vector; and $\mathbf{H}_m[n] = \text{diag}\{\boldsymbol{h}_m[n]\}$ with $\boldsymbol{h}_m[n] = [h_{m,1}[n], \ldots, h_{m,K}[n]]^T$ is the diagonal channel response matrix, where $h_{m,k}[n]$ are complex-valued frequency-domain wireless channel fading random processes for the mth user at the kth subcarrier. For a particular symbol n, $h_{m,k}[n]$ is the discrete-time Fourier transform of the N_t time-domain channel taps with time-delay τ_i and subcarrier spacing $\Delta f = F_s/K_{\text{fft}}$

$$h_{m,k}[n] = \sum_{i=1}^{N_t} g_{m,i}[n] e^{-j2\pi\tau_i k \Delta f}. \tag{2}$$

The $g_{m,i}[n]$ are the time-domain fading channel taps modeled as stationary and ergodic discrete-time random processes, with identical[1] normalized temporal autocorrelation function

$$r_m[\Delta] = \frac{1}{\sigma_{m,i}^2}\mathbb{E}\{g_{m,i}[n] g_{m,i}^*[n+\Delta]\}, \quad i = 1, \ldots, N_t \tag{3}$$

with tap powers $\sigma_{m,i}^2$, which we assume to be independent across the fading paths i and across users m. Since $g_{m,i}[n]$ is stationary, $h_{m,k}[n]$ is also stationary, and the distribution of $\boldsymbol{h}_m[n]$ is independent of symbol index n. In the subsequent discussion, we shall drop the symbol index n when the context is clear for notational brevity. Although the results of this paper are applicable to any fading distribution, we shall prescribe a particular fading distribution for the partial CSI in the next subsection for illustration purposes.

[1]We chose to make the *identical normalized temporal autocorrelation function* assumption to simplify the presentation and notation, but is not crucial to the development of our algorithms.

B. Partial Channel State Information

Assuming that the time domain channel taps are independent ZMCSCG random variables $g_{m,i} \sim \mathcal{CN}(0, \sigma_{m,i}^2)$, then from (2), we have

$$\begin{aligned} \boldsymbol{h}_m &\sim \mathcal{CN}(\mathbf{0}_K, \boldsymbol{\Sigma}_{\boldsymbol{h}_m}) \\ \boldsymbol{\Sigma}_{\boldsymbol{h}_m} &= \mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}_m}\mathbf{W}^H \end{aligned} \tag{4}$$

where $\mathbf{W}$ is the $K \times N_t$ DFT matrix with entries $[\mathbf{W}]_{k,i} = e^{-j2\pi\tau_i k\Delta f}, k = -(K-1)/2, \ldots, (K-1)/2; i = 1, \ldots, N_t$ and $\boldsymbol{\Sigma}_{\boldsymbol{g}_m} = \text{diag}\{\sigma_{m,1}^2, \ldots, \sigma_{m,N_t}^2\}$ is an $N_t \times N_t$ diagonal matrix of the time-domain path covariances[2]. Since we also assume that the fading for each user is independent, then the joint distribution of the stacked fading vector for all users $\boldsymbol{h} = [\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_M^T]^T$ is likewise a ZMCSCG random vector with distribution $\boldsymbol{h} \sim \mathcal{CN}(\mathbf{0}_{KM}, \boldsymbol{\Sigma}_{\boldsymbol{h}})$ where $\boldsymbol{\Sigma}_{\boldsymbol{h}}$ is the $KM \times KM$ block diagonal covariance matrix with $\boldsymbol{\Sigma}_{\boldsymbol{h}_m}$ as the diagonal block elements.

Suppose we wish to perform resource allocation for the m-th user with actual fading channel vector $\boldsymbol{h}_m$ at symbol index n, but only P symbols of delayed and noisy estimates of the channel D_t apart are available, which we denote as

$$\widetilde{\boldsymbol{h}}_m[n - pD_t] = \boldsymbol{h}_m[n - pD_t] + \boldsymbol{e}_m[n - pD_t] \tag{5}$$

where $\boldsymbol{e}_m[n - pD_t] \sim \mathcal{CN}(\mathbf{0}_K, \sigma_e^2 \mathbf{I}_K)$ is the spectrally and temporally white estimation error random vector with estimation error variance σ_e^2 which is uncorrelated with $\boldsymbol{h}_m[n-pD_t]$. This can effectively model a least-squares estimate of the channel using pilot tones with power σ_t^2, resulting in $\sigma_e^2 = \sigma_w^2/\sigma_t^2$. Stacking these into a PK-length vector, which we denote as $\widetilde{\boldsymbol{\mathfrak{h}}}_m = \left[\widetilde{\boldsymbol{h}}_m^T[n - D_t], \widetilde{\boldsymbol{h}}_m^T[n - 2D_t], \ldots, \widetilde{\boldsymbol{h}}_m^T[n - PD_t]\right]^T$, results in a ZMCSCG random vector with $PK \times PK$ block Hermitian-Toeplitz covariance matrix

$$\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\mathfrak{h}}}_m} = \mathbf{R}_m \otimes \boldsymbol{\Sigma}_{\boldsymbol{h}_m} + \sigma_e^2 \mathbf{I}_{PK} \tag{6}$$

where $\mathbf{R}_m$ is the Hermitian-symmetric and Toeplitz $P \times P$ temporal autocorrelation matrix with entries $[\mathbf{R}_m]_{i,j} = r_m[(i-j)D_t]$ and $\otimes$ is the Kronecker product. The conditional distribution of the desired channel $\boldsymbol{h}_m$ given $\widetilde{\boldsymbol{\mathfrak{h}}}_m$ is then $\boldsymbol{h}_m|\widetilde{\boldsymbol{\mathfrak{h}}}_m \sim \mathcal{CN}\left(\hat{\boldsymbol{h}}_m, \hat{\boldsymbol{\Sigma}}_m\right)$ where

$$\hat{\boldsymbol{h}}_m = \boldsymbol{\Sigma}_{\boldsymbol{h}_m \widetilde{\boldsymbol{\mathfrak{h}}}_m} \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\mathfrak{h}}}_m}^{-1} \widetilde{\boldsymbol{\mathfrak{h}}}_m \tag{7}$$

is the conditional mean estimator, which is also the MMSE predictor for the channel [18];

$$\hat{\boldsymbol{\Sigma}}_m = \boldsymbol{\Sigma}_{\boldsymbol{h}_m} - \boldsymbol{\Sigma}_{\boldsymbol{h}_m \widetilde{\boldsymbol{\mathfrak{h}}}_m} \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\mathfrak{h}}}_m}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{h}_m \widetilde{\boldsymbol{\mathfrak{h}}}_m}^H \tag{8}$$

is the conditional covariance, and is also the covariance matrix for the ZMCSCG prediction error vector we denote as $\hat{\boldsymbol{e}}_m$, and $\boldsymbol{\Sigma}_{\boldsymbol{h}_m \widetilde{\boldsymbol{\mathfrak{h}}}_m} = \mathbf{r}_m^T \otimes \boldsymbol{\Sigma}_{\boldsymbol{h}_m}$ with $\mathbf{r}_m^T = [r_m[D_t], \ldots, r_m[PD_t]]$ is the $K \times PK$ cross-covariance matrix. Interestingly, the pdf of $\boldsymbol{h}_m$ given the MMSE estimate $\hat{\boldsymbol{h}}_m$ is identical to $\boldsymbol{h}_m|\widetilde{\boldsymbol{\mathfrak{h}}}_m$, i.e.

$$\boldsymbol{h}_m|\hat{\boldsymbol{h}}_m \sim \mathcal{CN}\left(\hat{\boldsymbol{h}}_m, \hat{\boldsymbol{\Sigma}}_m\right) \tag{9}$$

[2]Following the convention in [1] and [2], we assume that the number of used subcarriers K is odd by including the null subcarrier at index 0 as part of the used subcarriers.

Note that $\hat{\boldsymbol{h}}_m$ is also ZMCSCG with covariance $\boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}_m} = \boldsymbol{\Sigma}_{\boldsymbol{h}_m} - \hat{\boldsymbol{\Sigma}}_m$. Thus, we can write

$$\boldsymbol{h}_m = \hat{\boldsymbol{h}}_m + \hat{\boldsymbol{e}}_m \tag{10}$$

which is also known as the *Statistician's Pythagorean Theorem* [19]. We shall use this equation to generate both the partial CSI and perfect CSI in the results section (Section V).

In ergodic capacity maximization with imperfect CSI, we require the marginal distribution for each subcarrier. The marginal fading distribution on subcarrier k conditioned on the estimated channels is a non-zero mean complex Gaussian random variable given as $h_{m,k}|\hat{h}_{m,k} \sim \mathcal{CN}(\hat{h}_{m,k}, \hat{\sigma}^2_{m,k})$ where $\hat{h}_{m,k}$ is the kth element in $\hat{\boldsymbol{h}}_m$ and $\hat{\sigma}^2_{m,k}$ is the kth diagonal element in $\hat{\boldsymbol{\Sigma}}_m$, which is essentially the prediction error variance for that subcarrier. Thus, the channel-to-noise ratio (CNR) $\gamma_{m,k} = |h_{m,k}|^2/\sigma^2_w$ conditioned on $\hat{\gamma}_{m,k} = |\hat{h}_{m,k}|^2/\sigma^2_w$ is in turn a non-central Chi-squared distributed random variable with two degrees of freedom with pdf [20, Eq. 2-1-118]

$$f_{\gamma_{m,k}}(\gamma_{m,k}|\hat{\gamma}_{m,k}) = \frac{1}{\rho_{m,k}} \exp\left(-\frac{\hat{\gamma}_{m,k}+\gamma_{m,k}}{\rho_{m,k}}\right) I_0\left(\frac{2}{\rho_{m,k}}\sqrt{\hat{\gamma}_{m,k}\gamma_{m,k}}\right) \tag{11}$$

where I_0 is the zeroth-order modified Bessel function of the first kind, and $\rho_{m,k} = \hat{\sigma}^2_{m,k}/\sigma^2_w$ is the ratio of the prediction error variance to the ambient noise variance.

III. Continuous Rate Maximization in OFDMA with Partial CSI

A. Problem Formulation

The capacity for user m and subcarrier k is given as

$$R_{m,k}(p_{m,k}\gamma_{m,k}) = \log_2(1 + p_{m,k}\gamma_{m,k}) \quad \text{bps/Hz} \tag{12}$$

Denote by $\boldsymbol{p} = [\boldsymbol{p}_1^T, \cdots, \boldsymbol{p}_K^T]^T$ the vector of powers to be determined, where $\boldsymbol{p}_k = [p_{1,k}, \cdots, p_{M,k}]^T$. The exclusive subcarrier assignment restriction can be written as $\boldsymbol{p}_k \in \boldsymbol{\mathcal{P}}_k \subset \mathbb{R}^M_+$, where

$$\boldsymbol{\mathcal{P}}_k \equiv \{p_{m,k} \geq 0 | p_{m,k}p_{m',k} = 0; \forall m \neq m'; m, m' \in \mathcal{M}\} \tag{13}$$

For notational convenience, we let $\boldsymbol{p} \in \boldsymbol{\mathcal{P}} \equiv \boldsymbol{\mathcal{P}}_1 \times \cdots \times \boldsymbol{\mathcal{P}}_K \subset \mathbb{R}^{MK}_+$ denote the space of allowable power vectors for all subcarriers.

We assume that we have knowledge of the imperfect CNR vector $\hat{\boldsymbol{\gamma}} = [\hat{\boldsymbol{\gamma}}_1^T, \ldots, \hat{\boldsymbol{\gamma}}_K^T]^T$, $\hat{\boldsymbol{\gamma}}_k = [\hat{\gamma}_{1,k}, \ldots, \hat{\gamma}_{M,k}]^T$; corresponding to an estimate of the actual CNR realization $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_K^T]^T$, $\boldsymbol{\gamma}_k = [\gamma_{1,k}, \ldots, \gamma_{M,k}]^T$. Thus, our power allocation vector $\boldsymbol{p}$ can only be a function of $\hat{\boldsymbol{\gamma}}$. The weighted expected sum rate maximization problem assuming partial CSI can be written as

$$\begin{aligned} f^* = \max_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \quad & \sum_{m\in\mathcal{M}} w_m \sum_{k\in\mathcal{K}} \mathbb{E}_{\gamma_{m,k}}\left\{ R_{m,k}\left(p_{m,k}\gamma_{m,k}\right) | \hat{\gamma}_{m,k}\right\} \\ \text{s.t.} \quad & \sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}} p_{m,k} \leq \bar{P} \end{aligned} \tag{14}$$

where w_m are positive constants such that $\sum_{m\in\mathcal{M}} w_m = 1$. Theoretically, varying these weights allows us to trace out the ergodic capacity region [21] assuming partial CSI; algorithmically, varying the weights allows us to prioritize the different users in the system and enforce certain notions of fairness, e.g. proportional fairnes and max-min fairness [9].

B. Dual Optimization Framework

Similar to [17], we begin our development by observing that the objective function in (14) is separable across the subcarriers, and is tied together only by the power constraint. We will approach this problem using dual optimization techniques [22]. The dual problem for (14) is defined as

$$g^* = \min_{\lambda\geq 0} \Theta(\lambda) \tag{15}$$

where the dual objective is given as

$$\begin{aligned} \Theta(\lambda) &= \max_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \sum_{m\in\mathcal{M}} w_m \sum_{k\in\mathcal{K}} \mathbb{E}\left\{ R_{m,k}\left(p_{m,k}\gamma_{m,k}\right) | \hat{\gamma}_{m,k}\right\} + \lambda\left(\bar{P} - \sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}} p_{m,k}\right) && \text{(16a)} \\ &= \lambda\bar{P} + \sum_{k\in\mathcal{K}} \max_{\boldsymbol{p}_k\in\boldsymbol{\mathcal{P}}_k} \sum_{m\in\mathcal{M}} w_m \mathbb{E}\left\{ R_{m,k}\left(p_{m,k}\gamma_{m,k}\right) | \hat{\gamma}_{m,k}\right\} - \lambda p_{m,k} && \text{(16b)} \\ &= \lambda\bar{P} + \sum_{k\in\mathcal{K}} \max_{m\in\mathcal{M}} \max_{p_{m,k}\geq 0} w_m \mathbb{E}\left\{ R_{m,k}\left(p_{m,k}\gamma_{m,k}\right) | \hat{\gamma}_{m,k}\right\} - \lambda p_{m,k} && \text{(16c)} \end{aligned}$$

where (16b) follows from the separability of the variables across subcarriers, and (16c) from the exclusive subcarrier assignment constraint. We have reduced the problem to a per-subcarrier optimization, and since $K \gg M$, we have also significantly decreased the computational burden.

We denote the optimal power allocation function for the innermost per-user and per-subcarrier problem in (16c) as $\hat{p}_{m,k}(\lambda)$, which can be found using simple differentiation, (see [14] for a similar derivation, albeit for a single-user system), and is given as

$$\hat{p}_{m,k}(\lambda) = \begin{cases} p_{m,k} : \mathbb{E}\left\{ \frac{\gamma_{m,k}}{1+p_{m,k}\gamma_{m,k}} \middle| \hat{\gamma}_{m,k}\right\} = \gamma_{0,m}, \\ \qquad\qquad \mathbb{E}\{\gamma_{m,k}|\hat{\gamma}_{m,k}\} \geq \gamma_{0,m} \\ 0 \qquad , \mathbb{E}\{\gamma_{m,k}|\hat{\gamma}_{m,k}\} < \gamma_{0,m} \end{cases} \tag{17}$$

where $\gamma_{0,m} = \frac{\lambda \ln 2}{w_m}$. This can be interpreted as a *multi-level water-filling* with cut-off CNR $\gamma_{0,m}$ similar to [17, Eq. 10], except that the cut-off is now based on the conditional mean of the CNR given its estimate, instead of the actual CNR. Using the pdf in (11), the conditional mean is simply $\mathbb{E}_{\gamma_{m,k}}\left\{\gamma_{m,k}|\hat{\gamma}_{m,k}\right\} = \hat{\gamma}_{m,k} + \rho_{m,k}$. Note that when we have perfect CSI, $\rho_{m,k} = 0$ and (17) actually reduces to the multi-level waterfilling equation for perfect CSI in [17, Eq. 10]. There is no closed form solution to (17), but it can be solved using numerical integration of the expectation, and a zero-finding procedure like bisection method [23] to find the power allocation.

Plugging (17) into (16c) and then in (15), we arrive at the following dual problem in the single variable λ

$$g^* = \min_{\lambda \geq 0} \lambda \bar{P} + \sum_{k \in \mathcal{K}} \max_{m \in \mathcal{M}} w_m \mathbb{E}\left\{ R_{m,k}\left(\hat{p}_{m,k}(\lambda)\gamma_{m,k}\right) \middle| \hat{\gamma}_{m,k} \right\} - \lambda \hat{p}_{m,k}(\lambda) \tag{18}$$

Using standard duality arguments (see e.g. [22, Prop. 5.1.2]), the objective in (18) can be shown to be convex in the single variable λ, but is actually not continuously differentiable due to the presence of the $\max$ function. Hence, powerful derivative-based minimization methods such as Newton's method cannot be used. Fortunately, we can use derivative-free single-dimensional line search methods that only need function evaluations, e.g. Golden-section or Fibonacci search [23] to find the optimum multiplier λ^*.

C. Optimal Subcarrier and Power Allocation

The optimal multiplier λ^* determines the optimal cutoff SNR $\gamma_{0,m}^* = \frac{\lambda^* \ln 2}{w_m}$. This can then be plugged back into the power allocation function (17) to arrive at the optimal subcarrier and power allocations:

$$m_k^* = \arg\max_{m \in \mathcal{M}} \mathbb{E}\left\{ w_m R_{m,k}\left(\hat{p}_{m,k}(\lambda^*)\gamma_{m,k}\right) \middle| \hat{\gamma}_{m,k} \right\} - \lambda^* \hat{p}_{m,k}(\lambda^*)$$

$$p_{m,k}^* = \begin{cases} \hat{p}_{m,k}(\lambda^*), & m = m_k^* \\ 0, & m \neq m_k^* \end{cases} \tag{19}$$

Note, however, that it is possible that the candidate power allocation does not satisfy the total power constraint, since the constraint is not enforced explicitly. Hence, our final power allocation values should be multiplied by a constant

$$\eta = \frac{\bar{P}}{\sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} p_{m,k}^*} \tag{20}$$

which we plug back into the objective in (14) to arrive at our computed primal optimal value

$$\hat{f}^* = \sum_{m \in \mathcal{M}} w_m \sum_{k \in \mathcal{K}} \mathbb{E}_{\gamma_{m,k}}\left\{ \log_2(1 + \eta p_{m,k}^* \gamma_{m,k}) \middle| \hat{\gamma}_{m,k} \right\} \tag{21}$$

Unfortunately, the above procedure is still highly computationally intensive, since for each candidate λ in the line search iterations, we need to compute MK power allocation values (17), each of which requires a zero-finding routine where a function value evaluation involves numerical integration to compute the expectation. Although both the line search and the zero-finding routines typically converge within very few iterations (< 10 in our experiments), the numerical integration procedure requires a lot more iterations (> 50), and hence is the main computational bottleneck. We shall overcome this problem using a closed-form approximation to the expectation in the power allocation function, as described in the next subsection.

D. Power Allocation Function Approximation

Our approach to approximating the expectation in (17) is to use a Gamma distribution to approximate the non-central Chi-squared distribution of $\gamma_{m,k}|\hat{\gamma}_{m,k}$ (11), which is known to approximate the body of this pdf quite well [24, p. 55]. This approximation is given by

$$f_{\gamma_{m,k}}(\gamma_{m,k}|\hat{\gamma}_{m,k}) \approx \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma_{m,k}^{\alpha-1} \exp\left(-\beta \gamma_{m,k}\right) \tag{22}$$

where $\alpha = (K_{m,k}+1)^2/(2K_{m,k}+1)$ is the Gamma pdf shape parameter with

$$K_{m,k} = \frac{\hat{\gamma}_{m,k}}{\rho_{m,k}} \tag{23}$$

as the specular to diffuse power ratio, equivalent to the $K-$factor in a Ricean pdf; and $\beta = \alpha/(\hat{\gamma}_{m,k} + \rho_{m,k})$ is the Gamma pdf rate parameter. Using this pdf, we can use [25, Section 3.383.10] to arrive at the following closed form approximation to the integral

$$\begin{aligned} \mathbb{E}\left\{ \frac{\gamma_{m,k}}{1 + p_{m,k}\gamma_{m,k}} \middle| \hat{\gamma}_{m,k} \right\} &\approx \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\gamma_{m,k}^\alpha}{1 + p_{m,k}\gamma_{m,k}} e^{-\beta\gamma_{m,k}} d\gamma_{m,k} \\ &= \frac{\alpha}{p_{m,k}} \left(\frac{\beta}{p_{m,k}}\right)^\alpha e^{\frac{\beta}{p_{m,k}}} \Gamma\left(-\alpha, \frac{\beta}{p_{m,k}}\right) \end{aligned} \tag{24}$$

where $\Gamma(a,x)$ is the incomplete Gamma function [25, Section 8.350]. Using (24) in (17) to solve for $p_{m,k}$, we are able to closely approximate the power allocation function $\hat{p}_{m,k}$. We plot the power allocation function using the Gamma pdf approximation and the actual Chi-squared pdf in Fig. 1 with $\gamma_0 = 1$ for various $\rho_{m,k} = \hat{\sigma}_{m,k}^2/\sigma_w^2$. Note that the approximation error is negligible, with a normalized mean-squared error of 5×10^{-5} and maximum error of 2.7×10^{-4}, while the computation of the approximation is almost $300\times$ faster than the actual using very crude computational time measurements in Matlab 7.2 (`tic-toc`).

E. Bound on the Relative Duality Gap

If we let $f^* > 0$ and $g^* > 0$ be the optimal values of the primal and dual problems given in (14) and (18), and let $\hat{f}^* > 0$ given in (21) be the computed feasible primal value, the relative duality (optimality) gap can be bounded as (see [17, Thm. 1] for a similar discussion and proof)

$$0 \leq \frac{g^* - f^*}{f^*} \leq \frac{g^* - \hat{f}^*}{\hat{f}^*} \tag{25}$$

The left inequality follows directly from the non-negativity of f^* and the weak duality theorem [22, Prop. 5.1.3. p. 495], and the right inequality follows from $\hat{f}^* \leq f^*$ since $\hat{f}^*$ is a feasible primal value and f^* is the optimal feasible primal value. In our numerical results, we show that the resulting optimality gaps using our algorithm are practically zero ($< 10^{-5}$). Thus, our approach can, for all practical purposes, be considered an optimal solution to the problem. This fortuitous phenomenon is brought about primarily by the separability of the problem, and furthermore by the fact that we have K

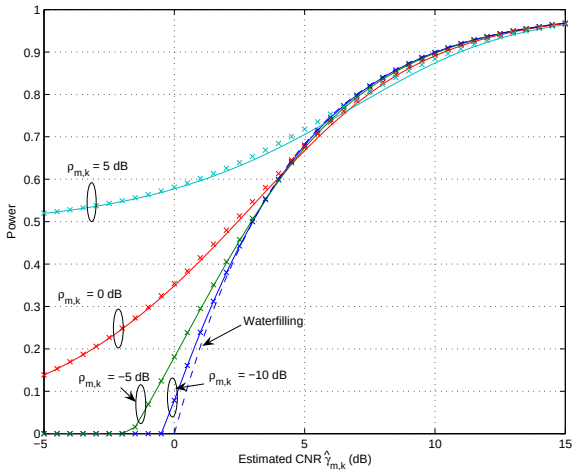

Fig. 1. Power allocation as a function of estimated CNR ($\hat{\gamma}$) with $\gamma_0 = 1$ for various $\rho_{m,k} = \hat{\sigma}^2_{m,k}/\sigma^2_w$. A solid line indicates the optimal power allocation, 'x's denote the approximation, and the dashed line correspond to the simple waterfilling allocation.

separable terms (which is typically large) and only a single constraint (average power constraint). This problem structure has been shown to be particularly suitable to dual optimization approaches, and has been noted in [26] (for the instantaneous optimization case), and more generally treated in [27].

F. Complexity Analysis

In each search iteration for λ in (18), we need to compute MK candidate power allocation functions given by (17) and (24). Each power allocation value calculation requires a zero-finding routine, e.g. bisection or Newton search [23], which we assume requires I_p function evaluations to converge. After determining the power allocation value, we then use it in the ergodic capacity integral in (18), which we assume requires I_c function evaluations to compute. Finally, assuming that we require I_λ line search iterations to solve for the optimum λ, the overall complexity is $\mathcal{O}(I_\lambda MK(I_p + I_c))$. Ignoring the constants I_λ, I_p, and I_c, the complexity is just $\mathcal{O}(MK)$.

IV. Discrete Rate Maximization in OFDMA with Partial CSI

In this section, we derive resource allocation algorithms for the practically relevant case of when only a discrete number of modulation and coding levels are available (i.e. adaptive modulation and coding).

A. Problem Formulation

1) Discrete rate function: In the discrete rate case, the data rate of the kth subcarrier for the mth user can be given by the following staircase function

$$R^d_{m,k}(p_{m,k}\gamma_{m,k}) = \begin{cases} 0, & p_{m,k}\gamma_{m,k} < \eta_0 \\ r_1, & \eta_0 \le p_{m,k}\gamma_{m,k} < \eta_1 \\ \vdots & \vdots \\ r_L, & \eta_{L-1} \le p_{m,k}\gamma_{m,k} < \eta_L \equiv \infty \end{cases} \tag{26}$$

where $\{r_l\}_{l\in\mathcal{L}}$, $\mathcal{L} = \{1,\ldots,L\}$ are the L available discrete information rates in increasing order, and $\{\eta_l\}_{l=0}^{L-1}$ are the SNR boundaries chosen in such a way that the information rate r_l is supportable subject to an instantaneous BER constraint. In the perfect CSI case, the candidate power allocation function that satisfies the BER constraint for each possible rate r_l is simply multi-level fading inversion (MFI), i.e. $p^{(l)}_{m,k} = \eta_l/\gamma_{m,k}$ [17]. This allows us to do away with having a BER function, since all that we require are the SNR rate region boundaries η_l which can be computed offline. However, with imperfect CSI, the average rate is given as

$$\bar{R}_{m,k} = \sum_{l\in\mathcal{L}} r_l P\left(\eta_{l-1} \le p_{m,k}\gamma_{m,k} < \eta_l | \hat{\gamma}_{m,k}\right) \tag{27}$$

Since we do not have the perfect CSI information $\gamma_{m,k}$, simply performing MFI on the imperfect CSI $\hat{\gamma}_{m,k}$ does not guarantee satisfaction of the BER constraint, and is illustrated in the results section (Section V). This necessitates a different approach of fulfilling the BER constraint.

2) Closed-form average BER function: With the imperfect CSI assumption, we require a BER function that can be expressed in terms of the SNR $p_{m,k}\gamma_{m,k}$ for a given r_l in order to enforce the *average* BER constraint. Suppose that we have this BER function for a given rate r_l denoted as $\text{BER}_l(p_{m,k}\gamma_{m,k})$, which could be derived using theoretical analysis or curve fitting from empirical data, the average BER constraint can be written as

$$\mathbb{E}_{\gamma_{m,k}}\left\{\text{BER}_l(p_{m,k}\gamma_{m,k})|\hat{\gamma}_{m,k}\right\} = \overline{\text{BER}} \tag{28}$$

Solving for $p_{m,k}$ in (28) for each $l \in \mathcal{L}$, we have L power allocation functions to choose from.

In order to simplify our development, we derive a closed-form expression for (28) assuming the fading distributions derived in Section II, and a representative BER prototype function that has been empirically shown to fit a lot of practical scenarios (see e.g. [28]). This prototype BER function is given by

$$\text{BER}_l(p_{m,k}\gamma_{m,k}) = a_l \exp\left(-b_l p_{m,k}\gamma_{m,k}\right) \tag{29}$$

where a_l and b_l are constants that are searched to fit the actual BER function for each r_l. For example, if we assume a Grey-coded square 2^{r_l}-QAM modulation scheme in AWGN, the BER function can be approximated to within 1-dB for $r_l \ge 2$ and BER $\le 10^{-3}$ with $a_l = 0.2$ and $b_l = 1.6/(2^{r_l} - 1)$ [28]. Using (29) in (28) with the pdf in (11), we have after some algebraic manipulation

$$\mathbb{E}\{\text{BER}_l(p_{m,k}\gamma_{m,k})|\hat{\gamma}_{m,k}\} = \hat{a}^{(l)}_{m,k}\int_0^\infty \exp\left(-x\left(\hat{b}^{(l)}_{m,k}p_{m,k}+1\right)\right) I_0\left(2\sqrt{K_{m,k}x}\right)dx \tag{30}$$

where $x = \gamma_{m,k}/\rho_{m,k}$, $K_{m,k} = \hat{\gamma}_{m,k}/\rho_{m,k}$, and

$$\begin{aligned} \hat{a}^{(l)}_{m,k} &= a_l \exp\left(-K_{m,k}\right) \\ \hat{b}^{(l)}_{m,k} &= b_l \rho_{m,k} \end{aligned} \tag{31}$$

Note that (30) can be interpreted as the Laplace transform of $I_0(2\sqrt{K_{m,k}x})$ with parameter $s = \hat{b}^{(l)}_{m,k}p_{m,k} + 1$, which is given in [29, Eq. 29.3.81]. Hence, a closed form expression

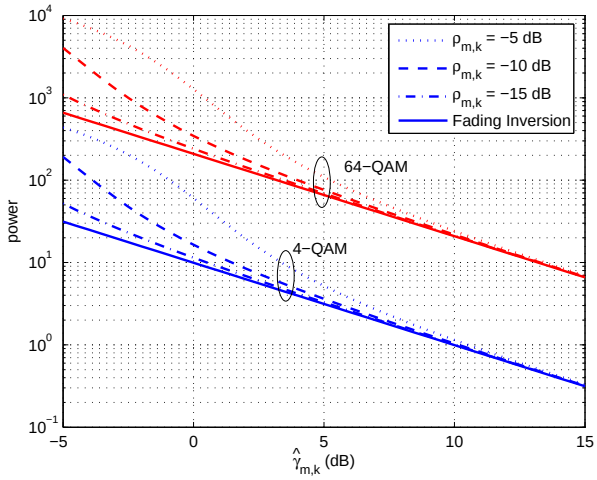

Fig. 2. Discrete rate power allocation as a function of estimated CNR ($\hat{\gamma}$) with $\gamma_0 = 1$ for various $\rho_{m,k}$.

for (30) can be written as Hence, a closed form expression for (30) can be written as

$$\mathbb{E}\{\text{BER}_l(p_{m,k}\gamma_{m,k})|\hat{\gamma}_{m,k}\} = \frac{\hat{a}_{m,k}^{(l)}}{\hat{b}_{m,k}^{(l)}p_{m,k}+1}\exp\left(\frac{K_{m,k}}{\hat{b}_{m,k}^{(l)}p_{m,k}+1}\right) \quad (32)$$

3) Closed-form power allocation function: Equating (32) with the target $\overline{\text{BER}}$, we arrive at the closed form expression for the candidate power allocation function given the estimated CNR $\hat{\gamma}_{m,k}$ and data rate r_l (see Appendix for derivation)

$$\hat{p}_{m,k}^{(l)} = \frac{1}{\hat{b}_{m,k}^{(l)}}\left(\frac{K_{m,k}}{W\left(\overline{\text{BER}}K_{m,k}/\hat{a}_{m,k}^{(l)}\right)}-1\right) \quad (33)$$

where $W(x)$ is the *Lambert-W* function, which is the solution to the transcendental equation $W(x)e^{W(x)} = x$. This function is ubiquitous in the physical sciences, and efficient algorithms have been extensively studied for its computation [30]. It is important to emphasize that (33) gives us the power allocation value that fulfills the average BER constraint when r_l is chosen as the rate for a particular user m and subcarrier k given imperfect CSI $\hat{\gamma}_{m,k}$. Fig. 2 shows the power allocation as a function of the estimated CNR $\hat{\gamma}_{m,k}$ for uncoded $4-$QAM and $64-$QAM for various $\rho_{m,k}$. We also plot the power allocation function when treating the $\hat{\gamma}_{m,k}$ as perfect, i.e. $p_{m,k}^{(l)} = \eta_l/\hat{\gamma}_{m,k}$ called multi-level fading inversion on imperfect CSI (Imperfect CSI-MFI). We can see that as $\rho_{m,k}$ decreases (prediction accuracy increases), the power allocation function approaches that of Imperfect CSI-MFI. On the other hand, a higher $\rho_{m,k}$ requires higher power in order to ensure the average BER requirement is met, especially for low estimated CNR $\hat{\gamma}_{m,k}$. Note also that the power allocation functions approach the Imperfect CSI-MFI value as $\hat{\gamma}_{m,k}$ becomes large, despite the value of $\rho_{m,k}$.

4) Closed-form average rate function: Using (33) in (27), the average rate given that r_l is chosen as the transmission rate can be written as

$$\begin{aligned}\bar{R}_{m,k}(r_l) &= \sum_{i\in\mathcal{L}} r_i P\left(\eta_{i-1} \le \hat{p}_{m,k}^{(l)}\gamma_{m,k} < \eta_i|\hat{\gamma}_{m,k}\right) \quad (34)\\ &= \sum_{i\in\mathcal{L}} r_i P\left(\frac{\eta_{i-1}}{\hat{p}_{m,k}^{(l)}} \le \gamma_{m,k} < \frac{\eta_i}{\hat{p}_{m,k}^{(l)}}\middle|\hat{\gamma}_{m,k}\right)\\ &= \sum_{i\in\mathcal{L}} r_i\left(F_{\gamma_{m,k}}\left(\frac{\eta_i}{\hat{p}_{m,k}^{(l)}}\middle|\hat{\gamma}_{m,k}\right) - F_{\gamma_{m,k}}\left(\frac{\eta_{i-1}}{\hat{p}_{m,k}^{(l)}}\middle|\hat{\gamma}_{m,k}\right)\right)\end{aligned}$$

From [20, Eq. 2.1-124], we have the following closed-form expression for the cdf of a non-central Chi-squared random variable

$$F_{\gamma_{m,k}}(x|\hat{\gamma}_{m,k}) = 1 - Q\left(\sqrt{\frac{2\hat{\gamma}_{m,k}}{\rho_{m,k}}}, \sqrt{\frac{2x}{\rho_{m,k}}}\right) \quad (35)$$

where $Q(a,b)$ is the Marcum-Q function. Using (35) in (34), we have a closed-form expression for the average rate for user m and subcarrier k given a choice of transmission rate r_l.

5) Optimization problem: Considering the above development, we can think of our decision variables in this case as a vector of rate allocation indices $\boldsymbol{l} = [\boldsymbol{l}_1^T, \ldots, \boldsymbol{l}_K^T]^T$ where $\boldsymbol{l}_k^T = [l_{1,k}, \ldots, l_{M,k}]^T$ and $l_{m,k} \in \{0, 1, \ldots, L\}$. The exclusive subcarrier assignment restriction can be written as $\boldsymbol{l}_k \in \boldsymbol{\mathcal{L}}_k$, where

$$\boldsymbol{\mathcal{L}}_k = \{l_{m,k} \in \{0,1,\ldots,L\}|l_{m,k}l_{m',k} = 0; \forall m \ne m'\} \quad (36)$$

For notational convenience, we let $\boldsymbol{l} \in \boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{L}}_1 \times \cdots \times \boldsymbol{\mathcal{L}}_K$ denote the space of allowable rate allocation indices for all subcarriers. Note that a decision of $l_{m,k} = 0$ means that neither rate nor power is transmitted on subcarrier k by user m. Thus, we can define $\bar{R}_{m,k}(r_0) \equiv 0$ and $\hat{p}_{m,k}^{(0)} \equiv 0$. The discrete weighted sum rate maximization problem with partial CSI is then formulated as

$$\begin{aligned} f_d^* = \max_{\boldsymbol{l}\in\boldsymbol{\mathcal{L}}} \quad & \sum_{m\in\mathcal{M}} w_m \sum_{k\in\mathcal{K}} \bar{R}_{m,k}(r_{l_{m,k}}) \\ \text{s.t.} \quad & \sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}} \hat{p}_{m,k}^{(l_{m,k})} \le \bar{P} \end{aligned} \quad (37)$$

where the power allocation function is given by (33) and the average rate by (34).

B. Dual Optimization Framework

Following a similar development as in Section III, the dual problem can be written as

$$g_d^* = \min_{\lambda\ge 0} \lambda\bar{P} + \sum_{k\in\mathcal{K}} \max_{m\in\mathcal{M}} \max_{l\in\mathcal{L}\cup\{0\}} \bar{R}_{m,k}(r_l) - \lambda\hat{p}_{m,k}^{(l)} \quad (38)$$

where we can once again use a univariate line-search method such as Golden-section search to compute for the optimum multiplier λ_d^*. Note that neither (33) nor (34) depend on λ. Hence, we can pre-compute these quantities for all $l \in \mathcal{L}$, $m \in \mathcal{M}$, and $k \in \mathcal{K}$ before running the line search iterations. Using λ_d^*, we arrive at the optimal rate allocation indices

$$l_{m,k}^* = \arg\max_{l\in\mathcal{L}} w_m\bar{R}_{m,k}(r_l) - \lambda_d^*\hat{p}_{m,k}^{(l)} \quad (39)$$

TABLE I
COMPLEXITY FOR THE OPTIMAL RESOURCE ALLOCATION ALGORITHMS WITH IMPERFECT CSI. M-NO. OF USERS, K-NO. OF SUBCARRIERS, I_λ-NO. OF LINE SEARCH ITERATIONS FOR DUAL PROBLEM.

Algorithm	Complexity
Continuous Rate Allocation	$\mathcal{O}(MKI_\lambda(I_p^{\text{a}}+ I_c^{\text{b}}))$
Discrete Rate Allocation	$\mathcal{O}(MK(I_\lambda + L^{\text{c}}))$

[a] No. of zero-finding iterations for the power allocation function (17)
[b] No. of function evaluations for numerical integration of the expected capacity (18)
[c] No. of discrete rate levels (26)

which in turn give us the optimal subcarrier, rate, and power allocation:

$$m_k^* = \arg\max_{m\in\mathcal{M}} w_m \bar{R}_{m,k}(r_{l^*_{m,k}}) - \lambda_d^* \hat{p}_{m,k}^{(l^*_{m,k})} \tag{40}$$

$$p_{m,k}^* = \begin{cases} \hat{p}_{m,k}^{(l^*_{m,k})}, & m = m_k^* \\ 0, & m \neq m_k^* \end{cases} \tag{41}$$

$$r_{m,k}^* = \begin{cases} r_{l^*_{m,k}}, & m = m_k^* \\ 0, & m \neq m_k^* \end{cases} \tag{42}$$

Finally, similar to the continuous rate case in Section III-E, the duality gap can be computed as in (25) to characterize how far away the solution is from the optimal. Furthermore, in cases where the power constraint is violated, we can employ a suitable heuristic to ensure that the constraint is met (c.f. (20)). One intuitive heuristic is to find a subcarrier k and user m with the least allocated rate but the most allocated power, and redistribute his power minus the constraint violation to the other subcarriers to improve their margins.

C. Complexity Analysis

Before running the line search iterations to compute for λ^* in (38), we need to compute MKL power allocation values (33) and average rate values (34) and store it in memory. This is followed by the search iterations which we assume to require I_λ, wherein each iteration requires $\mathcal{O}(MK)$ operations (38). The overall complexity order for the discrete rate resource allocation algorithm is thus $\mathcal{O}(MK(L+I_\lambda)$. Since L and I_λ are just constants independent of M and K, the complexity is $\mathcal{O}(MK)$. Note that in contrast to the continuous rate case, the ability to pre-compute the power and rate allocations, and the existence of closed-form solutions for these, actually result in the discrete rate allocation being less complex than the continuous rate allocation. This is fortunate because the discrete rate case is more practically relevant. Table I summarizes the complexity analysis for both continuous and discrete rate algorithms.

V. RESULTS AND DISCUSSION

We present several numerical examples to substantiate our theoretical claims. Our simulations are roughly based on a 3GPP-LTE downlink [2] system with parameters given in Table II.

We simulate the frequency-selective Rayleigh fading channel using the ITU-Vehicular A channel model [31]. We assume

TABLE II
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
Subcarriers (K_{fft})	64	Vehicular speed (V)	120 kph
Used Subcarriers (K)	33	Doppler frequency (F_d)	289 Hz
Bandwidth (B)	1.25 MHz	Prediction filter length (P)	4
Sampling Freq. (F_s)	1.92 MHz	Pilot spacing (D_t)	7
Carrier Freq. (F_c)	2.6 GHz	CP Length L_{cp}	6 samples

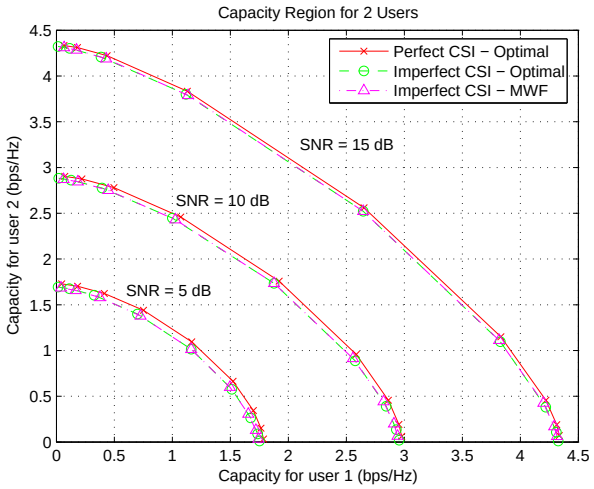

Fig. 3. 2-user capacity region for continuous rate optimal resource allocation with imperfect CSI. We also show the capacity region for optimal allocation with perfect CSI, and using multi-level waterfilling (MWF) on the imperfect CSI.

Clarke's U-shaped power spectrum [24] for each multipath tap, resulting in the temporal autocorrelation function $r_m[\Delta] = J_0(2\pi\Delta F_d D_t(K_{fft} + L_{cp})/F_s)$ where $J_0(x)$ is the zeroth-order Bessel function of the first kind [29, Ch. 9]. To simulate imperfect CSI, we generate IID realizations of $\hat{\boldsymbol{h}}_m$ and its prediction error vector $\hat{\boldsymbol{e}}_m$ as discussed in Section II-B. This allows us to also generate the "actual" channel $\boldsymbol{h}_m$ for the perfect CSI cases using (10).

In Fig. 3, we show the 2-user capacity region for continuous rate allocation with imperfect CSI (Imperfect CSI-Optimal) with 5000 channel realizations per data point. We also show the capacity region using optimal instantaneous rate resource allocation assuming perfect CSI (Perfect CSI-Optimal), which is essentially multi-level waterfilling (MWF) [17]; and the capacity region when we simply use MWF on the imperfect CSI (Imperfect CSI-MWF). Note that in all cases, rate maximization with imperfect CSI through channel prediction performs quite close to the case with perfect CSI. More important, Imperfect CSI-MWF performs similar to Imperfect CSI-Optimal. This can be explained by noticing that the optimal power allocation assuming imperfect CSI is almost equal to the waterfilling curve (see Fig. 1) except for very low estimated CNR. However, due to the effect of frequency and multiuser diversity, the subcarrier is typically assigned to the user with the highest CNR; thus, the power allocation is quite often almost identical to performing waterfilling on the imperfect CSI. A similar observation was also made in [14], albeit for the single user scenario.

Fig. 4 shows the discrete rate region for the optimal resource

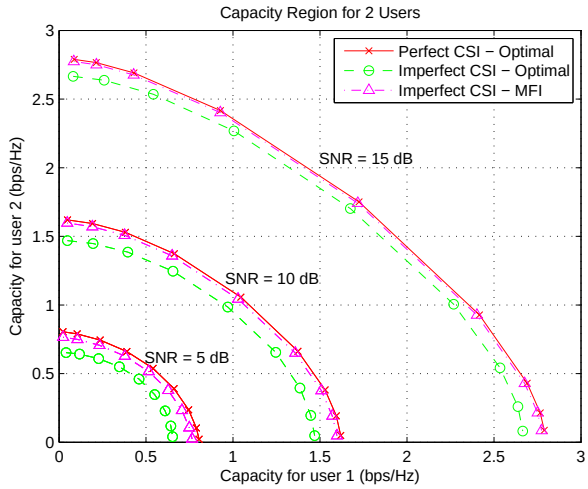

Fig. 4. 2-user capacity region for discrete rate optimal resource allocation with imperfect CSI. We also show the capacity region for optimal allocation with perfect CSI, and using multi-level fading inversion (MFI) on the imperfect CSI.

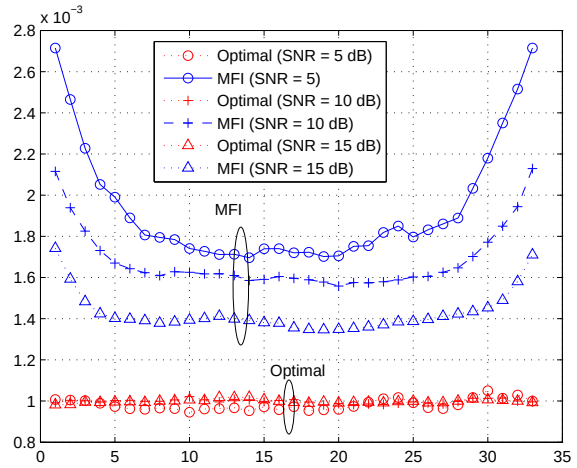

Fig. 5. Average BER for both users in each subcarrier for Imperfect CSI-Optimal and Imperfect CSI-MFI.

allocation algorithm assuming imperfect CSI (Imperfect CSI-Optimal). We also show the rate region achieved by using optimal resource allocation for discrete rates with perfect CSI (Perfect CSI-Optimal), which is essentially MFI [17], and by using MFI on the imperfect CSI (Imperfect CSI-MFI). Observe that due to the imperfect CSI assumption, Imperfect CSI-Optimal loses approximately 8% of the sum capacity when compared to Perfect CSI-Optimal. Observe also that Imperfect CSI-MFI results in a rate region that is quite close to the Perfect CSI-Optimal, and actually results in higher raw rates than the Imperfect CSI-Optimal. However, if we consider the average BER for each subcarrier shown in Fig. 5, Imperfect CSI-Optimal actually meets the average BER constraint of 10^{-3} (within $\pm 2\%$), but Imperfect CSI-MFI results in average BER violations of between $30-180\%$. Interestingly, the shape of the BER for Imperfect-CSI-Suboptimal closely resembles the shape of the prediction error variance $\hat{\sigma}^2_{m,k}$, shown in Fig. 6. This is intuitively satisfying, since a larger prediction error results in a larger mismatch between perfect and imperfect CSI, which is not taken into account by the Imperfect CSI-MFI algorithm. Thus, Imperfect CSI-MFI is equally aggressive in rate and power allocation even when the CSI prediction error is quite large. Our proposed Imperfect CSI-Optimal algorithm, on the other hand, is actually more conservative in rate and power allocation when the prediction MSE is large, thus allowing the average BER to be met. In a practical communications system, this would mean the difference of whether a packet is decoded successfully or not. Thus, using Imperfect CSI-MFI would result in unnecessary packet retransmissions and delays, and consequently decrease the throughput significantly. An explicit characterization in terms of throughput, however, is beyond the scope of this paper.

Table III shows the other relevant metrics of the optimal resource algorithms. The first column shows the average number of line-search iterations it took to converge to a tolerance of 10^{-4}. The second column shows the resulting relative duality

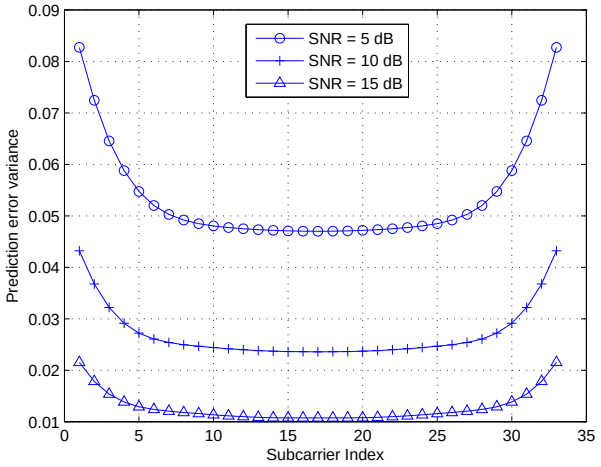

Fig. 6. Prediction error variance ($\hat{\sigma}^2_{m,k}$) for each subcarrier for different SNRs.

gaps given by (25). We can see that the duality gaps are virtually zero, and thus both algorithms can be considered optimal for all practical purposes.

VI. Conclusion

We have derived optimal resource allocation algorithms for ergodic continuous and discrete rate maximization assuming the availability of partial CSI. Using a dual optimization approach, we derived algorithms with complexity $\mathcal{O}(MK)$ per iteration and achieve relative duality gaps that are less than 10^{-5} for continuous rates and 10^{-3} for discrete rates in typical scenarios. Surprisingly, closed form expressions for the power allocation and average rates for the discrete rate case result in a lower complexity solution for discrete rates than for continuous rates.

TABLE III
RELEVANT METRICS FOR THE RESOURCE ALLOCATION ALGORITHMS

Metric	No. of Iterations (I_λ)			Relative Gap ($\times 10^{-4}$)		
SNR	5 dB	10 dB	15 dB	5 dB	10 dB	15 dB
Continuous Rates	8.599	8.501	8.686	.0840	.0568	.0412
Discrete Rates	21.33	21.15	21.12	71.48	7.707	5.662

APPENDIX A
DERIVATION OF (33)

Equating (32) to $\overline{\text{BER}}$ and after some algebraic manipulation, we have

$$\frac{K_{m,k}}{\hat{b}_{m,k}^{(l)} p_{m,k} + 1} \exp\left(\frac{K_{m,k}}{\hat{b}_{m,k}^{(l)} p_{m,k} + 1}\right) = \frac{K_{m,k}}{\hat{a}_{m,k}^{(l)}} \overline{\text{BER}} \quad (43)$$

Observe that this is in the form of the Lambert-W function $W(x)$ [30], which is the solution to $W(x)\exp(W(x)) = x$. Thus, we can write

$$W\left(\frac{K_{m,k}}{\hat{a}_{m,k}^{(l)}} \overline{\text{BER}}\right) = \frac{K_{m,k}}{\hat{b}_{m,k}^{(l)} p_{m,k} + 1} \quad (44)$$

which when solved for $p_{m,k}$ gives us (33).

REFERENCES

[1] *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE Std. 802.16e-2005, Feb. 2006.

[2] *3rd Generation Partnership Project, Technical Specification Group Radio Access Network; Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)*, 3GPP Std. TR 25.814 v. 7.0.0, 2006.

[3] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

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

[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 Trans. Broadcast.*, vol. 49, no. 4, pp. 362–370, Dec. 2003.

[6] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 171–178, Feb. 2003.

[7] L. Hoo, B. Halder, J. Tellado, and J. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 922–930, Jun. 2004.

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

[9] G. Song and Y. Li, "Cross-Layer Optimization for OFDM Wireless Networks Part II: Algorithm Development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[10] D. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, 1999.

[11] S. Falahati, A. Svensson, T. Ekman, and M. Sternad, "Adaptive modulation systems for predicted wireless channels," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 307–316, Feb. 2004.

[12] M. R. Souryal and R. L. Pickholtz, "Adaptive Modulation with Imperfect Channel Information in OFDM," in *Proc. IEEE Int. Conf. Comm.*, Jun. 2001, pp. 1861–1865.

[13] S. Ye, R. S. Blum, and J. L. J. Cimini, "Adaptive modulation for variable-rate OFDM systems with imperfect channel information," in *Proc. IEEE Vehicular Technology Conference*, vol. 2, Spring 2002, pp. 767–771.

[14] Y. Yao and G. Giannakis, "Rate-Maximizing Power Allocation in OFDM Based on Partial Channel Knowledge," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1073–1083, May 2005.

[15] S. Zhou and G. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?" *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, Jul. 2004.

[16] P. Xia, S. Zhou, and G. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 202–213, Jan. 2004.

[17] I. C. Wong and B. L. Evans, "Optimal OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates," *IEEE Trans. Wireless Commun.*, 2006 *submitted*.

[18] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 2002.

[19] L. L. Scharf and C. Demeure, *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*. Addison-Wesley Pub. Co., 1991.

[20] J. G. Proakis, *Digital communications*, 4th ed. McGraw-Hill, 2001.

[21] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels Part I.-Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[22] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.

[23] W. H. Press, *Numerical Recipes in C*. Cambridge University Press Cambridge, 1992.

[24] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Kluwer Academic, 2001.

[25] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products*, 6th ed. Academic Press, 2000.

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

[27] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

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

[29] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. U.S. Govt. Print. Off., 1972.

[30] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the Lambert-W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[31] *Selection procedures for the choice of radio transmission technologies for the UMTS*, ETSI Std. TR 101 112 v. 3.2.0, 1998.