Optimal Scheduling for OFDMA Systems

Rajeev Agrawal
Motorola Inc.
Arlington Heights, IL USA

Randall Berry
Northwestern Univ.
Evanston, IL USA

Jianwei Huang
Princeton Univ.
Princeton, NJ USA

Vijay Subramanian
Hamilton Institute
Maynooth, Ireland

Abstract— **We consider a scheduling and resource allocation problem for the downlink of an OFDMA-based wireless network, where the channel estimation error is modeled by a self-noise term in the decoding process. During each time-slot this involves selecting a subset of users for transmission, determining the assignment of available subcarriers to selected users, and for each subcarrier determining the transmission power and the coding and modulation scheme used. We address this in the context of a utility-based scheduling scheme presented in earlier papers. This results in an optimization problem, which is convex for a reasonable model of the feasible rates. By exploiting the structure of this problem, we develop optimal and sub-optimal algorithms for its solution. We provide simulation results comparing different algorithms and parameter settings.**

I. INTRODUCTION

Dynamic "channel-aware" scheduling and resource allocation is an essential component of most recent wireless data systems. A number of *gradient-based* scheduling and resource allocation algorithms have been considered, which attempt to maximize the projection onto the gradient of a system utility, see e.g. [2]–[5]. The utility is used to quantify fairness and other QoS considerations. This paper addresses gradient-based scheduling and resource allocation for a system using a combination of Time Division Multiplexing (TDM) and Orthogonal Frequency Division Multiplexing (OFDM). An example of such a system is IEEE 802.16 (WiMAX). In this setting, the problem is to determine which users are scheduled, as well as the allocation of transmission power and OFDM subcarriers[1] among the scheduled users.

In prior work [6], we considered gradient-based scheduling and resource allocation when code division multiple access (CDMA) was used to multiplex users within a time-slot, as in CDMA 1xEVDV. In [7], we extended the approach in [6] to an OFDM downlink where the transmitters and receivers have accurate channel information. In this paper, we consider the practical case where there are channel estimation errors, which are modeled as a self-noise term when calculating the achievable rate [13]. This significantly changes the resource allocation problem. However, we show that the general approach in [6], [7] can still be applied in this case.

As in [7], within each time-slot the gradient-based policy requires maximizing the weighted sum throughput over the set of feasible rates. The set of feasible rates depends on what subchannelization is used, the current channel state information, and the resource allocation decisions. We also allow constraints on the maximum SINR or rate per subcarrier, which can model a limitation on the available modulation order. When the rate per sub-carrier is given via the Shannon capacity formula and users are allowed to time-share each sub-carrier, as in [7] this becomes a tractable convex optimization problem.[2] A special case of this problem (without self-noise) for a fixed set of weights and no constraints on the SINR per carrier was considered in [8]; there a suboptimal algorithm with constant power per sub-carrier was given and shown in simulations to have little performance loss. Here, we consider a dual formulation for this problem, which enables us characterize some structural properties and leads to both optimal and reduced complexity sub-optimal algorithms. We also present simulation results in a system where the scheduling weights are dynamically adjusted according to a gradient-based scheduling rule. Some other related work on OFDM optimization can be found in [9]–[11]; in these papers self noise is not considered.

II. PROBLEM FORMULATION

We consider the downlink of a single cell in an OFDM system with K users. Time is divided into TDM time-slots that contain an integer number of OFDM symbols. In each time-slot, the scheduling and resource allocation decision can be viewed as selecting a rate vector $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{K,t})$ from the current feasible rate region $\mathcal{R}(\boldsymbol{e}_t) \subset \mathbb{R}_+^K$, where $\boldsymbol{e}_t$ indicates the time-varying channel state information available at the scheduler. This decision is made according to the gradient-based scheduling framework in [2], [4]. Namely, an $\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)$ is selected that has the maximum projection onto the gradient of a system utility function $\nabla U(\boldsymbol{W}_t)$, where $U(\boldsymbol{W}_t) = \sum_{i=1}^K U_i(W_{i,t})$, and $U_i(W_{i,t})$ is an increasing concave utility function of user i's average throughput, $W_{i,t}$, up to time t. In other words, the scheduling and resource allocation decision is the solution to

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \nabla U(\boldsymbol{W}_t)^T \cdot \boldsymbol{r}_t = \max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i \dot{U}_i(W_{i,t}) r_{i,t}. \quad (1)$$

For example, one class of utility functions given in [4] is

$$U_i(W_{i,t}) = \begin{cases} \frac{c_i}{\alpha}(W_{i,t})^\alpha, & \alpha \le 1, \ \alpha \ne 0, \\ c_i \log(W_{i,t}), & \alpha = 0, \end{cases} \quad (2)$$

[1]In the following we use the terms subcarrier and tone synonymously.

[2]We focus on systems that do not use superposition coding and successive interference cancellation within a sub-carrier.

where $\alpha \leq 1$ is a fairness parameter and c_i is a QoS weight. In this case, (1) becomes

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i c_i (W_{i,t})^{\alpha-1} r_{i,t}. \tag{3}$$

With equal class weights, setting $\alpha = 1$ results in a scheduling rule that maximizes the total throughput during each slot. For $\alpha = 0$, this results in the proportional fair rule.

In general, we consider the problem of

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i w_{i,t} r_{i,t}, \tag{4}$$

where $w_{i,t} \geq 0$ is a time-varying weight assigned to the ith user at time t. We note that (4) must be re-solved at each scheduling instant because of changes in both the channel state and the weights (e.g., the gradient of the utility).

A. OFDM capacity regions

Solving (4) depends on the state dependent capacity region $\mathcal{R}(\boldsymbol{e})$.[3] We focus on a model appropriate for downlink OFDM systems; similar models have been considered in [8], [12]. In this model, $\mathcal{R}(\boldsymbol{e})$ is parameterized by the allocation of subcarriers to users and the allocation of power across subcarriers. In a traditional OFDM system, at most one user may be assigned to any subcarrier. Here, as in [9], [11], we make the simplifying assumption that multiple users can share one tone using some orthogonalization technique (e.g. via TDM) but not super-position coding. In practice, if a scheduling interval contained multiple OFDM symbols, we could implement such sharing by giving a fraction of the symbols to each user; of course, each user would be constrained to use an integer number of symbols and the required signaling overhead would increase. If the number of tones is large[4], then this approximation gets tighter. Given a solution to this problem, we can obtain a feasible solution allowing only one user per tone by applying an appropriate projection. For the simulations in Section IV, we assign only one user per tone.

Let $\mathcal{N} = \{1, \ldots, N\}$ denote the set of tones. For each tone j and user i, let e_{ij} denote the received signal to noise ratio (SNR) per unit power. We denote the power allocated to user i on subchannel j by p_{ij} and the fraction of that subchannel allocated to user i by x_{ij}. These must satisfy a total power constraint, $\sum_{i,j} p_{ij} \leq P$, and for all subcarriers j, $\sum_i x_{ij} \leq 1$, i.e., the total fraction of each sub-carrier allocated must be no greater than one. For a given allocation, with perfect channel estimation, user i's feasible rate on subcarrier j is given by $r_{ij} = x_{ij} B \log(1 + \frac{p_{ij} e_{ij}}{x_{ij}})$. This corresponds to the Shannon capacity of a Gaussian noise channel with bandwidth $x_{ij} B$ and received SNR $\frac{p_{ij} e_{ij}}{x_{ij}}$.[5] This SNR arises from viewing p_{ij} as the energy per time-slot user i is allowed to use on subcarrier j; the corresponding transmission power becomes $\frac{p_{ij}}{x_{ij}}$ when

---

[3]To simplify notation we have suppressed the time-dependence.

[4]Most proposals consider about a thousand subcarriers.

[5]As in [7], to better model the achievable rates in a practical system we can re-normalize e_{ij} by γe_{ij}, where $\gamma \in [0, 1]$ represents the system's "gap" from capacity.

only a fraction x_{ij} of the tone is allocated. Without loss of generality we set $B = 1$.

It has recently been suggested that for realistic systems the OFDM model must have a self-interference term. Following a similar model as in [13], we assume the received signal on the j's subchannel of user i is

$$y_{ij} = h_{ij} s_{ij} + n_{ij}.$$

Assume that the receiver estimates channel h_{ij} with $h_{ij} + h_{ij,\delta}$, where $h_{ij,\delta}$ is the channel estimation error. As a result, the processed received signal is

$$z_{ij} = h_{ij}^* h_{ij} s_{ij} + h_{ij}^* n_{ij} + h_{ij,\delta}^* (h_{ij} s_{ij} + n_{ij}),$$

and the effective SNR is

$$\text{Eff-SNR} = \frac{\|h_{ij}\|^2 \|s_{ij}\|^2}{2\sigma_{ij}^2 \left(1 + \frac{\|h_{ij,\delta}\|^2}{\|h_{ij}\|^2}\right) + \|h_{ij,\delta}\|^2 \|s_{ij}\|^2},$$

where the variance of noise n_{ij} is assumed to be $2\sigma_{ij}^2$. Defining $g_{ij} = \|h_{ij}\|^2$, $a_{ij} = \frac{\|h_{ij,\delta}\|^2}{\|h_{ij}\|^2}$, $e_{ij} = \frac{\|h_{ij}\|^2}{2\sigma_{ij}^2}$ and $\|s_{ij}\|^2 = p_{ij}$, we get the effective SNR to be

$$\text{Eff-SNR} = \frac{p_{ij} e_{ij}}{1 + a_{ij} + a_{ij} e_{ij} p_{ij}}. \tag{5}$$

Defining $\tilde{e}_{ij} := e_{ij}/(1 + a_{ij})$, then (5) can be written as

$$\text{Eff-SNR} = \frac{p_{ij} \tilde{e}_{ij}}{1 + a_{ij} p_{ij} \tilde{e}_{ij}}.$$

Compared to the case without self-noise, the effective SNR is still increasing in p_{ij}. However, it now has a maximum of $1/a_{ij}$. For the sake of presentation, we assume that $a = a_{ij}$ for all i and j. The analysis is almost identical if users have different a_{ij}'s.

Taking time sharing into consideration, user i's feasible rate on subcarrier j is now given by

$$r_{ij} = x_{ij} B \log \left(1 + \frac{p_{ij} \tilde{e}_{ij}}{x_{ij} + a p_{ij} \tilde{e}_{ij}}\right). \tag{6}$$

The achievable rate region is then

$$\mathcal{R}(\boldsymbol{e}) = \Bigg\{ \boldsymbol{r} : \ r_i = \sum_j x_{ij} \log \left(1 + \frac{p_{ij} \tilde{e}_{ij}}{x_{ij} + a p_{ij} \tilde{e}_{ij}}\right), \\ \sum_{ij} p_{ij} \leq P, \ \sum_i x_{ij} \leq 1 \ \forall j, \ (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X}, \Bigg\}, \tag{7}$$

where[6]

$$\mathcal{X} := \left\{ (\boldsymbol{x}, \boldsymbol{p}) \geq \boldsymbol{0} : x_{ij} \leq 1, p_{ij} \leq \frac{x_{ij} s_{ij}}{e_{ij}} \ \forall i, j \right\}. \tag{8}$$

Here, s_{ij} is a maximum SINR constraint on tone j for user i, e.g. to model a constraint on the maximum rate per tone due to a limitation on the available modulation order.

We assume that the joint channel state $\boldsymbol{e}$ is known by the scheduler for all users and tones with an estimation error quantified by a. With many tones and users, providing pilots

---

[6]We use boldfaced symbols to denote the vector of all the corresponding values across users/tones, e.g. $\boldsymbol{x}$ is the vector of all x_{ij}'s.

and/or feedback per tone can require excessive overhead. One approach for reducing this overhead is by forming *subchannels* from disjoint sets of tones. Feedback and resource allocation is then done at the granularity of these subchannels. The above model can be adapted to this setting by viewing $\mathcal{N}$ as the set of subchannels and e_{ij} as the effective SNR per unit power for user i within the jth subchannel. Specifically, assuming that k tones are bundled into subchannel j, e_{ij} is chosen so that the total rate for user i in this subchannel is approximately $kx_{ij} \log(1 + \frac{p_{ij}e_{ij}}{x_{ij}+ap_{ij}e_{ij}})$. For our simulations, we set e_{ij} to be the geometric average of the SNR per unit power of all the tones in a subchannel, which provides a (provable) lower bound of the achievable rate.

Subchannels can be formed in various ways; in our simulations, the following three approaches are considered: (1) adjacent channelization, where adjacent tones are grouped into sub-channels; (2) interleaved channelization, where tones are interleaved to form subchannels; and (3) random channelization, where tones are randomly assigned to subchannels. In IEEE 802.16d/e, interleaved channelization is primarily used; the optional "band AMC mode" allows for adjacent channelization. Randomized channelization can model systems that employ frequency hopping as in the Flash OFDM system.

III. Optimal and Suboptimal Algorithms

From (4) and (7), the scheduling and resource allocation problem can be stated as:

$$\max_{(\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X}} V(\boldsymbol{x},\boldsymbol{p}) := \sum_i w_i \sum_j x_{ij} \log\left(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij}+ap_{ij}\tilde{e}_{ij}}\right)$$

$$\text{subject to:} \quad \sum_{i,j} p_{ij} \le P, \text{ and } \sum_i x_{ij} \le 1, \forall j \in \mathcal{N}. \tag{9}$$

A. Optimal Dual Solution

We first solve this problem using duality methods. It can be shown that (9) is convex and Slater's condition holds, so there is no duality gap and the optimal solution is characterized by the Karush-Khun-Tucker conditions [1].

Consider the Lagrangian,

$$L(\boldsymbol{x},\boldsymbol{p},\lambda,\boldsymbol{\mu}) := \sum_i w_i \sum_j x_{ij} \log\left(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij}+ap_{ij}\tilde{e}_{ij}}\right)$$

$$+ \lambda\left(P - \sum_{i,j} p_{ij}\right) + \sum_j \mu_j\left(1 - \sum_i x_{ij}\right).$$

First we optimize over $\boldsymbol{p}$ given $\boldsymbol{x}$, $\boldsymbol{\mu}$ and λ. If there is no self-noise ($a = 0$), we obtain a "water-filling" type of solution,[7]

$$p_{ij}^* = \frac{x_{ij}}{\tilde{e}_{ij}}\left[\left(\frac{w_i\tilde{e}_{ij}}{\lambda} - 1\right)^+ \wedge s_{ij}\right]. \tag{10}$$

When $a > 0$, we obtain

$$p_{ij}^* = \frac{x_{ij}}{\tilde{e}_{ij}}\left[q\left(a,\left(\frac{w_i\tilde{e}_{ij}}{\lambda} - 1\right)^+\right) \wedge s_{ij}\right], \tag{11}$$

[7]The notation $(x)^+ = \max(x,0)$ and $x \wedge y = \min(x,y)$.

where

$$q(a,z) = \left(\frac{2a+1}{2a(a+1)}\right)\left(\sqrt{1 + \frac{4a(a+1)}{(2a+1)^2}z} - 1\right).$$

Notice that the optimal power allocation is no longer increasing in $\tilde{e}_{ij}$ as in the case when $a = 0$. On the other hand, the optimal value of p_{ij}^* is still a linear function of x_{ij}, which means the resulting Lagrangian is also a linear function of x_{ij}.

Substituting (11) into $L(\boldsymbol{x},\boldsymbol{p},\lambda,\boldsymbol{\mu})$, we have

$$L(\boldsymbol{x},\boldsymbol{p}^*,\lambda,\boldsymbol{\mu}) = \sum_{ij} x_{ij}\left(\mu_{ij}(\lambda) - \mu_j\right) + \sum_j \mu_j + \lambda P,$$

where $\mu_{ij}(\lambda) := w_i h\left(a, \frac{w_i\tilde{e}_{ij}}{\lambda}, s_{ij}\right)$, and

$$h(a,\omega,s_{ij}) := \log\left(1 + \frac{q\left(a,(\omega-1)^+\right) \wedge s_{ij}}{1 + q\left(a,(\omega-1)^+\right) \wedge s_{ij}}\right)$$

$$- \frac{1}{\omega}\left(q\left(a,(\omega-1)^+\right) \wedge s_{ij}\right).$$

Optimizing $L(\boldsymbol{x},\boldsymbol{p}^*,\lambda,\boldsymbol{\mu})$ over $\boldsymbol{x}$ yields the corresponding dual function

$$L(\lambda,\boldsymbol{\mu}) := L(\boldsymbol{x}^*,\boldsymbol{p}^*,\lambda,\boldsymbol{\mu})$$

$$= \sum_{ij}\left(\mu_{ij}(\lambda) - \mu_j\right)^+ + \sum_j \mu_j + \lambda P.$$

Since there is no duality gap, it follows that minimizing this over λ and $\boldsymbol{\mu}$ yields an optimal solution to (9). First considering the optimal $\boldsymbol{\mu}$, and we have as in [7]:

Lemma 3.1: For all $\lambda \ge 0$,

$$L(\lambda) := \min_{\boldsymbol{\mu} \ge \boldsymbol{0}} L(\lambda,\boldsymbol{\mu}) = \lambda P + \sum_j \mu_j^*(\lambda),$$

where for all j, the minimizing value of μ_j is

$$\mu_j^*(\lambda) = \max_i \mu_{ij}(\lambda). \tag{12}$$

Note that (12) requires a sort of all the users according to the metrics μ_{ij} for each sub-channel j. It can be shown that $L(\lambda)$ is a convex function of λ; hence it can be minimized using an iterated one dimensional search, like the Golden Section method. At the minimizing value λ^*, $L(\lambda^*)$ gives the optimal solution to (9).

B. Optimal primal variables with time-sharing

Next we turn to finding optimal values of the primal variables $(\boldsymbol{x},\boldsymbol{p})$. For a given $\lambda \ge 0$, let

$$(\boldsymbol{x}^*,\boldsymbol{p}^*) := \arg\max_{(\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X}} L\left(\boldsymbol{x},\boldsymbol{p},\lambda,\boldsymbol{\mu}^*(\lambda)\right), \tag{13}$$

which can be solved using the same procedure as in deriving the dual function. Given that $\lambda = \lambda^*$, it follows from duality theory that if the corresponding $(\boldsymbol{x}^*,\boldsymbol{p}^*)$ are primal feasible and satisfy complimentary slackness, then they are optimal. However, in (12) there can be multiple users in a given subchannel whose metrics μ_{ij} are tied at the maximum value. In this case, there will be multiple primal values that satisfy (13), not all of which may be feasible. Thus, breaking these

ties to settle on a specific primal solution is necessary to find the optimal solution. A key point is that when ties occur at a given λ, $L(\lambda)$ is not differentiable at that λ. However, since $L(\lambda)$ is a convex function, subgradients exists. Such ties can be broken using the subgradient information as in [7]. Details are omitted due to space constraints.

C. Single user per tone

Next we consider the case where the final allocation is restricted to at most one user per subchannel. We can still find the optimal λ^*. If there are no ties as discussed above, the optimal solution will only allow one user/tone. If there are ties, a reasonable heuristic is to simply choose one extreme point allocation. In our simulations, we choose the extreme point corresponding to the subgradient with the smallest non-negative value; i.e. the extreme point f, for which $\sum_{j \in \mathcal{N}} \tilde{p}_{f(j)j}$ is closest to P, without exceeding it. Other rules for choosing an extreme point could also be used.

For a given extreme point, the total power constraint using the powers $\tilde{p}_{f(j)j}$ will be over-shot or under-shot (unless this point is optimal). We then re-optimize the power allocation for the given fixed tone allocation $\boldsymbol{x}$, i.e. we solve

$$\max_{\boldsymbol{p}:(\boldsymbol{p},\boldsymbol{x})\in\mathcal{X}} V(\boldsymbol{n},\boldsymbol{p}) \quad \text{s.t.} \sum_{ij} p_{ij} \le P. \quad (14)$$

Let $L_{\boldsymbol{x}}(\lambda)$ be the dual function for this problem. Given that $\tilde{\lambda} = \arg\min_{\lambda \ge 0} L_{\boldsymbol{x}}(\lambda)$, the optimal power allocation to (14) is given by (10) with $\lambda = \tilde{\lambda}$ and the given tone allocation. A bisection search can again be used to find the optimal λ.

D. Single sort suboptimal algorithm

The optimal sub-carrier allocation is determined by assigning each tone j to the user with the largest metric $\mu_j^*(\lambda^*)$ on that tone (breaking any ties as discussed above). This requires iterating to find the optimal Lagrange multiplier λ^*. We give a sub-optimal algorithm that is based instead on a single sort of the users on each tone according to a different metric. Here, we consider using the metric $w_i \bar{R}_{ij}$, where $\bar{R}_{ij}$ is the rate that user i could achieve on this channel under a constant power allocation, i.e., $\bar{R}_{ij} = \log[1 + (s_{ij} \wedge (\frac{\tilde{e}_{ij}P/N}{1+a\tilde{e}_{ij}P/N}))]$. The tone is allocated to the user with the largest metric, with ties broken arbitrarily. After the tone allocation is made, constant power is allocated on all subchannels. This metric was motivated in part by prior work in [8], [10] where a uniform power allocation was shown to be nearly optimal.

IV. SIMULATION STUDY

In this section we report simulation results for the algorithm (OPTIMAL) that finds the optimal λ^* and then chooses a tone-allocation with one user per tone as described in Section III-C. We also consider the sub-optimal algorithm (HEURISTIC) described in Section III-D. We simulate a single cell with $M = 40$ users. The channel gains e_{ij} are the product of a fixed location-based term for each user i and a frequency-selective fast fading term. The location-based components were picked using an empirically obtained distribution for many users in a large system. The fast-fading term was generated using a

block-fading model based upon the Doppler frequency (for the block-length in time) and a standard reference mobile delay-spread model (for variation in frequency). For a user's fast-fading term, each multi-path component was held fixed for $2msec$ and an independent value was generated for the next block, corresponding to a 250MHz Doppler frequency. The delay-spread was $1\mu sec$. The user's channel conditions averaged over the applicable channelization scheme are fed back to the scheduler for all the channels.

We considered a system bandwidth of 5MHz corresponding to 512 OFDM tones. We group these into 64 subchannels (8 tones per subchannel). The symbol duration was $100\mu sec$ with a cyclic prefix of $10\mu sec$. This roughly corresponds to 20 OFDM symbols per fading block. The resource allocation is done once per fading block. All the results are averaged over the last 2000 OFDM symbols out of 60000 OFDM symbols (i.e., 3000 fading blocks), at this time the system has reached a stationary operation point. All users were infinitely back-logged and assigned a throughput-based utility as in (2) with parameter $c_i = 1$ and the same fairness parameters (α) for each simulation. To account for realistic network conditions, we calculate the achievable rate of user i on subchannel j as

$$r_{ij} = 0.28 \frac{B}{S} x_{ij} \log \left(1 + \frac{0.56 p_{ij} \tilde{e}_{ij}}{x_{ij} + a p_{ij} \tilde{e}_{ij}} \right),$$

where B is the subchannel bandwidth and S is the symbol length. The scheduling is based on the geometric average of the subchannel gains; the decoded rate is based on per tone channel conditions. The total power constraint is fixed at $P = 6$W. There are no per-user SINR constraints (i.e., $s_{ij} = \infty$).

The first set of simulation results are for a system with adjacent subchannelization and no self-noise ($a = 0$). Table I shows results for both algorithms for different choices of the utility parameter α. The column "Utility" gives the average utility per user for each algorithm. The column "log U" shows the log utility per user; this gives some indication of the "fairness" of the resulting allocation (for $\alpha = 0$ this is the same as the utility). The column "Rate" is the average throughput per user, and the final column is the average number of users scheduled. For each choice of α, the two algorithms perform close to each other for each of these metrics; when $\alpha = 1$ (maximum throughput) they have identical performance.

Next we consider the effect of different channelization schemes. Table II shows the performance of the two algorithms for the adjacent (Adj.), randomized (Ran.), and Interleaved

TABLE II

PERFORMANCE OF DIFFERENT CHANNELIZATION SCHEMES ($\alpha = 0.5$, NO SELF-NOISE).

Chan.	Algorithm	Utility	Log U	Rate (kbps)	Num.
Adj.	OPTIMAL	545.15	10.83	105.9	7.32
Adj.	HEURISTIC	528.83	10.73	99.3	7.20
Int.	OPTIMAL	494.61	10.53	92.4	1.79
Int.	HEURISTIC	486.40	10.47	88.4	1.14
Ran.	OPTIMAL	487.53	10.53	89.2	4.89
Ran.	HEURISTIC	479.07	10.46	84.2	4.39

TABLE III

PERFORMANCE OF DIFFERENT CHANNELIZATION SCHEMES ($\alpha = 0.5$, SELF-NOISE $a = 0.01$).

Chan.	Algorithm	Utility	Log U	Rate (kbps)	Num.
Adj.	OPTIMAL	512.17	10.82	82.5	7.51
Adj.	HEURISTIC	489.30	10.70	73.7	7.39
Int.	OPTIMAL	466.38	10.53	73.4	1.86
Int.	HEURISTIC	452.30	10.45	66.1	1.16
Ran.	OPTIMAL	458.53	10.51	70.0	5.03
Ran.	HEURISTIC	444.69	10.43	63.1	4.46

Fig. 1.   Empirical CDF of users' throughputs.

(Int.) channelization schemes described in Section II-A. The other parameters are the same as in Table I, with $\alpha = 0.5$. Again, HEURISTIC performs nearly the same as OPTIMAL and in the interleaved case even achieves a slightly higher utility. For both algorithms, the random channelization results in lower utility than the adjacent, and the interleaved results in yet lower utility. This is likely due to the decreased frequency diversity with each scheme. Indeed, for the channel model used here, in the interleaved case all subchannels can be shown to be very nearly identical[8], which explains why both schemes schedule only one or two users.

Next, in Table III, we consider the case when the self-noise coefficient $a = 0.01$ and the other parameters are the same as in Table II. The performance gap between the two algorithms is slightly larger compared to the case without self-noise. In Figure 1, we plot the throughput CDF for both algorithms, with self-noise ($a = 0.01$) and without ($a = 0$). Here adjacent channelization is used and $\alpha = 0.5$. It is clear that users achieve better throughput when there is no self-noise ($a = 0$). The OPTIMAL algorithm always achieves better rates compared with HEURISTIC under the same value of a.

V. CONCLUSIONS

We considered scheduling and resource allocation for the downlink of OFDM systems, where the channel estimation error is represented by a self-noise term. We formulated an optimal scheduling and resource allocation problem, which was shown to be a convex problem. Using a dual formulation, we characterized the optimal solution, and used this to develop optimal and sub-optimal algorithms. The algorithms can be applied across different channelization schemes and accommodate per user SINR constraints. We presented simulation

results showing a simpler sub-optimal algorithm, in which users are sorted once per tone based on a uniform power allocation, yields reasonable performance. The performance gap between the optimal and suboptimal algorithms increases with the self-noise coefficient.

REFERENCES

[1] D. Bertsekas, *Nonlinear Programming*, 2nd ed.   Belmont, Massachusetts: Athena Scientific, 1999.
[2] R. Agrawal and V. Subramanian, "Optimality of Certain Channel Aware Scheduling Policies," *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, Oct. 2002.
[3] H. Kushner and P. Whiting, "Asymptotic Properties of Proportional-Fair Sharing Algorithms," *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, Oct. 2002.
[4] R. Agrawal, A. Bedekar, R. La, and V. Subramanian, "A Class and Channel-Condition based Weighted Proportionally Fair Scheduler," *Proc. of ITC 2001*, Salvador, Brazil, Sept. 2001.
[5] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multi-user throughput allocation," *Operations Research*, 53(1), 2005, pp.12-25.
[6] R. Agrawal, V. Subramanian and R. Berry, "Joint Scheduling and Resource Allocation in CDMA Systems," *Proc. of WiOpt '04*, Cambridge, UK, March 24-26, 2004.
[7] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink Scheduling and Resource Allocation for OFDM Systems," *Conference on Information Sciences and Systems (CISS)*, March 2006
[8] L. Hoo, B. Halder, J. Tellado, and J. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channels: Asymptotic FDMA Capacity Region and Algorithms," *IEEE Trans. on Communications*, vol. 52, no. 6, pp. 922-930, June 2004.
[9] C. Y. Wong, R. S. Cheng, K. B. Letaief and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation," *IEEE Journal on Selected Areas in Communications* vol. 17, no. 10, Oct. 1999.
[10] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM System," *IEEE Journal on Selected Areas in Communications* vol. 21, no. 2, pp. 171-178, Feb. 2003.
[11] Y. J. Zhang and K. B. Letaief, "Adaptive Resource Allocation and Scheduling for Multiuser Packet-based OFDM Networks," in *Proc. of IEEE ICC*, pp. 2949-2953, June 2004.
[12] L. Li and A. Goldsmith, "Optimal Resource Allocation for Fading Broadcast Channels- Part I: Ergodic Capacity," *IEEE Trans. on Information Theory*, vol. 47, no. 3, pp. 1083-1102, March 2001.
[13] H. Jin, R. Laroia and T. Richardson, "Superposition by Position," preprint, 2006

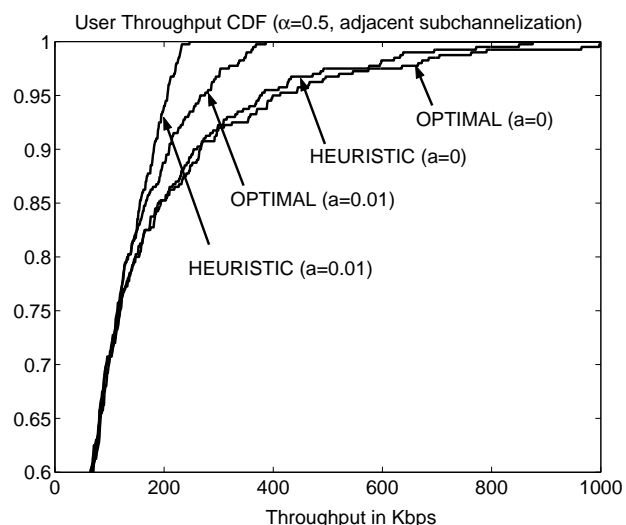[8]Using an arithmetic mean instead of the geometric mean one can prove that the sub-channels are exactly identical.