

# Joint Scheduling and Resource Allocation in Uplink OFDM Systems

Jianwei Huang  
the Chinese University of Hong Kong  
Hong Kong, China

Vijay G. Subramanian  
Hamilton Institute, NUIM,  
Maynooth, Ireland

Randall Berry  
Northwestern University  
Evanston, IL USA

Rajeev Agrawal  
Motorola Inc.  
Arlington Heights, IL USA

**Abstract**—Orthogonal Frequency Division Multiplexing (OFDM) with dynamic scheduling and resource allocation is widely considered to be a key component of 4G cellular networks. However, scheduling and resource allocation in an OFDM system is complicated, especially in the uplink due to two reasons: (1) the discrete nature of channel assignments, and (2) the heterogeneity of the users' channel conditions, individual resource constraints and application requirements. We approach this problem using a gradient-based scheduling framework presented in previous work. Physical layer resources (bandwidth and power) are allocated to maximize the projection onto the gradient of a total system utility function which models application-layer Quality of Service (QoS). This is formulated as a convex optimization problem. We present an optimal solution using a dual decomposition. This solution has prohibitively high computational complexity but reveals guiding principles that we use to generate a family of lower complexity sub-optimal algorithms. We compare the performance of these algorithms via a realistic OFDM simulator.

## I. INTRODUCTION

This paper analyzes the uplink scheduling problem for OFDM systems. The specific problem is motivated by the WiMAX/802.16e standard<sup>1</sup> where there is a centralized scheduler that knows the QoS classes, queue-lengths and delays of the packets queued on each mobile device. The WiMAX/802.16e standard specifies reserved time-frequency slots for communicating this information to the scheduler and for conveying the scheduling decisions to the mobiles, both with low delays.

Using OFDM on the uplink of a cellular system with dynamic scheduling and resource allocation has only recently attracted significant attention. Thus the literature on this subject is still in a nascent state [13], [15].<sup>2</sup> This problem is precisely stated in Section II. We highlight two challenging aspects of this problem. First, the discrete nature of channel assignments in OFDM systems usually leads to hard integer programming problems. Second, the per-user power constraint

Part of this work was done while J. Huang and V. Subramanian were at Motorola. J. Huang is supported in part by Direct Grant of the Chinese University of Hong Kong under Grant 2050398.R. Berry was supported in part by the Motorola-Northwestern Center for Seamless Communications and NSF CAREER award CCR-0238382.

<sup>1</sup>LTE for 3GPP and 3GPP2 and the FLASH OFDM system from Qualcomm Flarion also fit the model we consider in this paper. Furthermore, this model is applicable for both FDD and TDD systems.

<sup>2</sup>The downlink version of this problem has received more attention, but as we discuss later, the uplink version of the problem introduces several new dimensions.

TABLE I  
KEY NOTATIONS

Notation	Physical Meaning
$N$	total number of carriers
$\mathcal{N}$	set of all carriers
$M$	total number of users
$\mathcal{M}$	set of all users
$w_i$	user $i$ 's (dynamic) weight
$e_{ij}$	normalized SINR on carrier $j$ for user $i$
$p_{ij}$	power allocated on carrier $j$ for user $i$
$x_{ij}$	fraction of carrier $j$ allocated to user $i$
$P_i$	maximum transmit power for user $i$

that arises in the uplink problem makes the problem even less tractable. We initially consider a mathematical abstraction in which multiple users can share one subcarrier/tone using orthogonalization (e.g. via time-sharing<sup>3</sup>), which relaxes the integer constraints. In Section III we derive an optimal solution to this relaxed problem using a dual decomposition. This provides insight into the structure of an optimal solution; however, due to the per-user power constraints determining this solution has high computational complexity. In Section IV we use the insights gained from the optimal solution to propose a family of sub-optimal algorithms that also take into account the integer constraint of one user per subcarrier/tone. Finally, in Section V we present numerical results for these algorithms using a realistic OFDM simulator.

## II. PROBLEM STATEMENT

We consider a model for uplink scheduling in an OFDM system that is based on our previous work on downlink scheduling in CDMA systems [3] and OFDM systems [4]. Specifically, in every scheduling epoch the scheduler seeks to maximize a (time-varying) weighted sum of the users' rates over a given (time-varying) rate-region. We begin by describing this rate-region. The key notations are listed in Table I; we use bold symbols to denote vectors of these quantities, e.g.,  $\mathbf{w} = \{w_i, \forall i\}$ ,  $\mathbf{e} = \{e_{ij}, \forall i, j\}$ ,  $\mathbf{p} = \{p_{ij}, \forall i, j\}$ , and  $\mathbf{x} = \{x_{ij}, \forall i, j\}$ .

We assume that the scheduler has the knowledge of the received Signal-to-noise ratio (SNR)  $e_{ij}$  per unit power for

<sup>3</sup>While super-position coding would yield an even larger capacity region, we do not use it as it is still not practical.

every user and tone.<sup>4</sup> We represent the time-varying channel quality vector at time  $t$  as  $\mathbf{e}_t$ . As in [4], this model can also incorporate various sub-channelization schemes where the resource allocation is performed in terms of subchannels (i.e., sets of tones). In this case,  $e_{ij}$  represents the channel condition for the sub-channel, e.g., the (geometric/arithmetic) average across the tones in the sub-channel. This model also applies if resource allocation is done with a granularity of multiple symbols in the time domain.

Let  $\mathcal{R}(\mathbf{e}_t)$  denote the feasible rate region at time  $t$ . We model this as

$$\mathcal{R}(\mathbf{e}_t) = \left\{ \mathbf{r} \in \mathbb{R}_+^M : \right. \\ \left. r_i = \sum_{j \in \mathcal{N}} x_{ij} \log \left( 1 + \frac{p_{ij} e_{ij}}{x_{ij}} \right), \forall i \in \mathcal{M} \right\}, \quad (1)$$

where  $(\mathbf{x}, \mathbf{p}) \in \mathcal{X}$  are chosen subject to

$$\sum_i x_{ij} \leq 1, \forall j \in \mathcal{N}, \quad (2)$$

$$\sum_j p_{ij} \leq P_i, \forall i \in \mathcal{M}, \quad (3)$$

and the set

$$\mathcal{X} := \left\{ (\mathbf{x}, \mathbf{p}) \geq \mathbf{0} : 0 \leq x_{ij} \leq 1, p_{ij} \leq \frac{x_{ij} s_{ij}}{e_{ij}} \forall i, j \right\}. \quad (4)$$

Here,  $s_{ij}$  is a maximum SNR constraint on tone  $j$  for user  $i$ . In practical OFDM systems,  $x_{ij}$  is constrained to be an integer, in which case we add the additional constraint  $x_{ij} \in \{0, 1\}$  for all  $i, j$ . Initially, we ignore this constraint; this corresponds to a system in which users can share each tone. If resource allocation is done blocks of OFDM symbols, then fractional values of  $x_{ij}$  can be implemented by time-sharing the symbols in a block.<sup>5</sup>

Next we formulate the scheduling and resource allocation problem. Our approach is based on the gradient-based scheduling framework presented in [2], [10], [11]. Each user  $i$  is assigned a utility function  $U_i(W_{i,t}, Q_{i,t})$  depending on their average throughput  $W_{i,t}$  up to time  $t$  and their queue-length  $Q_{i,t}$  at time  $t$ . This is used to quantify fairness and ensure stability of the queues. During each scheduling epoch  $t$ , the system objective is to choose a rate vector  $\mathbf{r}_t$  in  $\mathcal{R}(\mathbf{e}_t)$  that maximizes a (dynamic) weighted sum of the users' rates, where the weights are determined by the gradient of the sum utility across all users. More precisely, the scheduler seeks to maximize the projection of  $\mathbf{r}_t$  onto the gradient  $\nabla_{\mathbf{w}} U(\mathbf{W}_t, \mathbf{Q}_t) - \nabla_{\mathbf{q}} U(\mathbf{W}_t, \mathbf{Q}_t)$ , where  $U(\mathbf{W}_t, \mathbf{Q}_t) = \sum_{i=1}^K U_i(W_{i,t}, Q_{i,t})$ . We further assume that for each user  $i$ ,  $U_i(W_{i,t}, Q_{i,t}) = u_i(W_{i,t}) - \frac{d_i}{p} (Q_{i,t})^p$ , where  $u_i(W_{i,t})$  is a increasing concave function,  $d_i \geq 0$  is a QoS weight for user

$i$ 's queue length, and  $p > 1$  is a fairness parameter associated with the queue length. Hence, the scheduling and resource allocation decision is the solution to

$$\max_{\mathbf{r}_t \in \mathcal{R}(\mathbf{e}_t)} (\nabla_{\mathbf{w}} U(\mathbf{W}_t, \mathbf{Q}_t)^T - \nabla_{\mathbf{q}} U(\mathbf{W}_t, \mathbf{Q}_t)^T) \cdot \mathbf{r}_t = \\ \max_{\mathbf{r}_t \in \mathcal{R}(\mathbf{e}_t)} \sum_i \left( \frac{\partial u_i(W_{i,t})}{\partial W_{i,t}} + d_i (Q_{i,t})^{p-1} \right) r_{i,t}. \quad (5)$$

Several variations of the policy in (5) have been studied. If  $d_i = 0$  for all  $i \in \mathcal{M}$ , the resulting policy has been shown to yield utility maximizing solutions [2], [10], [11]. If  $u_i(\cdot) \equiv 0$  with  $d_i > 0$  for all  $i \in \mathcal{M}$  then this policy has been shown to be stabilizing in a variety of settings [5]–[7]. A specific choice of  $d_i$  for “usual” utility functions  $u_i(\cdot)$  has been shown to produce utility maximizing solutions subject to stability [9].

As a concrete example, one class of utility functions typically used (e.g. [1], [12]) for  $u_i(\cdot)$  is

$$u_i(W_{i,t}) = \begin{cases} \frac{c_i}{\alpha} (W_{i,t})^\alpha, & \alpha \leq 1, \alpha \neq 0 \\ c_i \log(W_{i,t}), & \alpha = 0, \end{cases} \quad (6)$$

where  $\alpha \leq 1$  is a fairness parameter and  $c_i \geq 0$  is a QoS weight. In this case, the objective in (5) becomes  $\sum_i (c_i (W_{i,t})^{\alpha-1} + d_i (Q_{i,t})^{p-1}) r_{i,t}$ . With zero queue weights  $d_i$  and equal throughput weights  $c_i$ , setting  $\alpha = 1$  results in a “maximum throughput” scheduling rule that maximizes the total throughput during each slot. For  $\alpha = 0$ , this results in the proportional fair rule [8].

The optimization in (5) can be written as

$$\max_{\mathbf{r}_t \in \mathcal{R}(\mathbf{e}_t)} \sum_i w_{i,t} r_{i,t}, \quad (7)$$

where  $w_{i,t} \geq 0$  is a time-varying weight assigned to the  $i$ th user at time  $t$ . In the above examples, these weights were given by the gradient of the utility function; however, other methods for generating these weights are also possible. We emphasize that (7) must be re-solved at each scheduling instant because of changes in both the channel state,  $\mathbf{e}_t$ , and the weights (e.g., the gradient of the utility).

### III. OPTIMAL SOLUTION

In this section we consider the optimal solution to (7) when  $\mathcal{R}(\mathbf{e}_t)$  is given by (1). This problem can be written as

$$\max_{(\mathbf{x}, \mathbf{p}) \in \mathcal{X}} \sum_{i \in \mathcal{M}} w_i \sum_{j \in \mathcal{N}} x_{ij} \log \left( 1 + \frac{p_{ij} e_{ij}}{x_{ij}} \right) \quad (\text{UL})$$

subject to the per carrier assignment constraints in (2) and the per user power constraints in (3), where  $\mathcal{X}$  is given in (4).

It can be shown that Problem UL has no duality gap and so we can solve it by considering a dual formulation. We associate dual variables  $\boldsymbol{\lambda} = (\lambda_i)_{i \in \mathcal{M}}$  with constraints (3) and  $\boldsymbol{\mu} = (\mu_j)_{j \in \mathcal{N}}$  with constraints (2), resulting in the Lagrangian,

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}, \mathbf{p}) := \sum_{i,j} w_i x_{ij} \log \left( 1 + \frac{p_{ij} e_{ij}}{x_{ij}} \right) \\ + \sum_i \lambda_i (P_i - \sum_j p_{ij}) + \sum_j \mu_j (1 - \sum_i x_{ij}). \quad (8)$$

<sup>4</sup>In both FDD and TDD systems this can be obtained using a combination of measurements made on the UL pilots as well as past transmissions from the mobiles.

<sup>5</sup>Likewise, if the number of channels are large enough so that the channel gains do not change dramatically among adjacent channels, then the fractional value of  $x_{ij}$  can also implemented by frequency sharing (e.g., [15]).

From duality theory, it follows that the optimal solution to Problem UL is given by

$$\min_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \geq \mathbf{0}} \max_{(\mathbf{x}, \mathbf{p}) \in \mathcal{X}} L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}, \mathbf{p}). \quad (9)$$

Next we solve this by first analytically solving for the optimal  $\mathbf{p}$  and  $\mathbf{x}$  given fixed values of the dual variables. We then show that the optimal  $\boldsymbol{\mu}$  is given by a performing a search for the maximum value of a per-user metric on each carrier. The final step is to numerically search for the optimal value of  $\boldsymbol{\lambda}$ .

Optimizing  $L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}, \mathbf{p})$  over  $\mathbf{p}$  given  $\mathbf{x}$ ,  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$ , we get

$$p_{ij}^* = \frac{x_{ij}}{e_{ij}} \min \left\{ \left( \frac{w_i e_{ij}}{\lambda_i} - 1 \right)^+, s_{ij} \right\}, \quad (10)$$

where  $\{x\}^+ = \max\{x, 0\}$ . Note that unless  $\sum_{j \in \mathcal{N}} \frac{x_{ij} s_{ij}}{e_{ij}} < P_i$ , it will always be that  $\sum_{j \in \mathcal{N}} p_{ij}^* = P_i$ . Assuming this is the case, (10) is the water-filling solution which takes into account the maximum SINR constraint. Substituting  $\mathbf{p}^*$  into  $L(\cdot, \cdot, \cdot, \cdot)$  yields

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}, \mathbf{p}^*) = \sum_{ij} x_{ij} (w_i h(\lambda_i, w_i e_{ij}, s_{ij}) - \mu_j) + \sum_j \mu_j + \sum_i \lambda_i P_i, \quad (11)$$

where we have used the function  $h(\cdot, \cdot, \cdot)$  from [3], namely,

$$h(a, b, c) = \begin{cases} 0 & \text{if } a \geq b; \\ \frac{a}{b} - 1 - \log \frac{a}{b} & \text{if } \frac{b}{1+c} \leq a < b; \\ \log(1+c) - \frac{b}{ac} & \text{if } a < \frac{b}{1+c}, \end{cases} \quad (12)$$

where  $a \geq 0$ ,  $b > 0$  and  $c \geq 0$ . Optimizing (11) over  $\mathbf{x}$  such that  $x_{ij} \in [0, 1]$  yields

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}^*, \mathbf{p}^*) = \sum_{ij} (w_i h(\lambda_i, w_i e_{ij}, s_{ij}) - \mu_j)^+ + \sum_j \mu_j + \sum_i \lambda_i P_i, \quad (13)$$

where the carrier allocation has the following structure

$$x_{ij}^*(\mu_j) = \begin{cases} 1, & \text{if } w_i h(\lambda_i, w_i e_{ij}, s_{ij}) > \mu_j; \\ [0, 1], & \text{if } w_i h(\lambda_i, w_i e_{ij}, s_{ij}) = \mu_j; \\ 0, & \text{if } w_i h(\lambda_i, w_i e_{ij}, s_{ij}) < \mu_j. \end{cases} \quad (14)$$

Since the cost function in (13) is separable, minimizing  $L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}^*, \mathbf{p}^*)$  to obtain the optimal  $\mu_j^*(\boldsymbol{\lambda})$  requires a simple sort per carrier similar as that in [3], namely,

$$\mu_j^*(\boldsymbol{\lambda}) = \max_i w_i h(\lambda_i, w_i e_{ij}, s_{ij}), \quad (15)$$

where  $\mu_{ij}(\cdot) := w_i h(\cdot, w_i e_{ij}, s_{ij})$ .

From (14) and (15), it is clear that  $x_{ij}^*(\mu_j^*(\boldsymbol{\lambda})) \equiv 0$  if  $i \notin \arg \max_{i \in \mathcal{M}} \mu_{ij}(\lambda_i)$ , i.e., there is a per subcarrier metric such that any user who does not maximize this metric on a given subcarrier will not be allocated the carrier. There will be ties when multiple users achieve the same value of  $\mu_j^*$  on carrier  $j$ . These can be broken arbitrarily to obtain the correct value for

$L(\boldsymbol{\lambda}, \boldsymbol{\mu}^*, \mathbf{x}^*, \mathbf{p}^*)$ . Now substituting  $\boldsymbol{\mu}^*$  into  $L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{x}^*, \mathbf{p}^*)$ , and noticing that  $\boldsymbol{\mu}, \mathbf{x}^*, \mathbf{p}^*$  are all functions of  $\boldsymbol{\lambda}$ , we have

$$L(\boldsymbol{\lambda}) := L(\boldsymbol{\lambda}, \boldsymbol{\mu}^*, \mathbf{x}^*, \mathbf{p}^*) = \sum_j \max_i \mu_{ij}(\lambda_i) + \sum_i \lambda_i P_i.$$

The solution to (9) is given by minimizing  $L(\boldsymbol{\lambda})$  over  $\boldsymbol{\lambda} \geq \mathbf{0}$ . For this we use a sub-gradient-based search, i.e.,

$$\lambda_i(t+1) = \left[ \lambda_i(t) - \kappa(t) \left( P_i - \sum_j p_{ij}^*(t) \right) \right]^+, \forall i \in \mathcal{M}.$$

The algorithm will converge when  $\kappa(t)$  is chosen sufficiently small [14]. The detailed algorithm is given in [16]. Given an optimal  $\boldsymbol{\lambda}$ , by duality,  $L(\boldsymbol{\lambda})$  is the optimal objective value to Problem UL. However, to implement this, the scheduler must specify the corresponding primal values of  $(\mathbf{x}, \mathbf{p})$ . Here, as in [4], more care is required. Specifically, when ties occur in (15), how the tie is resolved becomes important. Essentially, we need to inspect all possible ties in each of the channels, and find the feasible channel allocation that gives the maximum primal value among all ties.

In [4] we used a similar algorithm to solve a downlink OFDM scheduling problem. However, there are several major differences between the uplink and downlink setting which make this approach less appealing for implementation in the uplink setting. First, in the downlink case there is a single power constraint  $\sum_{i,j} p_{ij} \leq P$  for the base station instead of the per-user power constraints in (3). Hence, in the downlink case  $L(\boldsymbol{\lambda})$  is a function of only a single dual variable  $\lambda$ , which simplifies the numerical search for the optimal  $\lambda$ . This also makes it easier to break ties and to determine when to stop the algorithm.<sup>6</sup> Also, the uplink case can be more sensitive to how ties are resolved. For example, if two users,  $i$  and  $l$ , have the same weights ( $w_i = w_l$ ) and the same gains on channel  $j$  ( $e_{ij} = e_{lj}$ ), then allocating channel  $j$  to either user yields the same total weighted rate and the same total power usage in the downlink case. On the other hand, different allocations lead to different individual power consumptions in the uplink case, and thus may lead to different solutions.

Finally, the number of ties is typically much larger in the uplink case than in the downlink case. Consider a simple scenario with two users and two channels. Each user has the same gain over both channels, i.e.,  $e_{i1} = e_{i2} = e_i$  for  $i = 1, 2$ , and  $P = P_1 = P_2$ , where  $P$  is the total power constraint in the downlink case. Assume user 2 has a much better channel than user 1 so that in the downlink case, the unique optimal solution is to allocate both channels to user 2, and there is no tie. However, in the uplink case, it can be shown that at the optimal dual solution,  $\lambda_1$  and  $\lambda_2$  will satisfy

$$\mu_{1j}(\lambda_1) = \mu_{2j}(\lambda_2) \text{ for } j = 1, 2,$$

i.e., there is a tie in each channel and we have to compare four possible channel allocations to find the optimal solution. This

<sup>6</sup>In the downlink case the subgradients of  $L(\lambda)$  are scalars and so one can stop when the maximum subgradient is positive and the minimum subgradient is zero. In the uplink case the subgradients are vectors and so can not be well-ordered.

can be easily extended to  $M$  users and  $N$  channels, with each user having the same gain over all its channels. This results in  $M^N$  ties, independent of the variation in gains across users.

#### IV. SUBOPTIMAL ALGORITHMS

The algorithm in Section III yields the optimal solution to Problem UL in each scheduling interval, but due to the effects discussed above this is not computationally feasible for even a moderately sized system. We now present a family of sub-optimal algorithms (SOA's) that try to reduce this complexity while sacrificing little in optimality. These algorithms seek to exploit the problem structure revealed by the optimal algorithm. Furthermore, these sub-optimal algorithms all enforce an integer tone allocation during each scheduling interval. Additional heuristic algorithms are given in [16].

In the optimal algorithm, given the optimal  $\lambda^*$ , the optimal carrier allocation up to any ties is determined by sorting the users on each tone according to the metric  $\mu_{ij}(\lambda)$  as in (14). Given an optimal carrier allocation, the optimal power allocation is given by a per-user water-filling allocation as in (10). In each SOA, we use the same two phases with some modifications to reduce the complexity of computing  $\lambda^*$  and the optimal carrier allocation. Specifically, we begin with a *Carrier Allocation (CA)* phase in which we assign each sub-carrier to at most one user. Instead of using the metric given by the optimal  $\lambda$ , we consider metrics based on a constant power allocation over all carriers assigned to a user. We follow this with a *Power Allocation (PA)* phase in which each user's power is allocated across the assigned carriers using a waterfilling allocation as in the optimal algorithm. We describe these in more detail next.

##### A. Channel Allocation (CA) Phase

We consider a family of SOAs in which carriers are assigned sequentially in one pass based on a per user metric for each carrier, i.e. we iterate  $N$  times, where each iteration corresponds to the assignment of one carrier. Let  $\mathcal{K}_i(n)$  denote the set of carriers assigned to user  $i$  after the  $n$ th iteration. Let  $g_i(n)$  denote user  $i$ 's metric during the  $n$ th iteration and let  $l_i(n)$  be the carrier index that user  $i$  would like to be assigned if he is assigned the  $n$ th carrier. The resulting CA algorithm is given in Algorithm 1. Note that the user metrics are updated after each carrier is assigned.

We consider several variations of Algorithm 1 which correspond to different choices for Lines 4 and 5. The choices for Line 4 are:

(4A): Sort all of the carriers based on the best normalized SINR among the users, i.e., find a channel permutation  $\{\alpha_j\}$  such that  $\max_i e_{i\alpha_1} \geq \max_i e_{i\alpha_2} \geq \dots \geq \max_i e_{i\alpha_N}$ , and set  $l_i(n) = \alpha_n$  for each user  $i$ . Note this sort only needs to be performed once.

(4B): For each user  $i$ , set  $l_i(n)$  to be the carrier with the largest gain among all unassigned carriers, i.e.,  $l_i(n) = \arg \max_{j \in \mathcal{N} \setminus \cup_i \mathcal{K}_i(n-1)} e_{ij}$ . This requires  $M$  sorts (one per user); these also need to be performed only once (since each

---

#### Algorithm 1 CA Phase for SOAs

---

- 1: Initialization: set  $n = 0$  and  $\mathcal{K}_i(n) = \emptyset$  for each user  $i$ .
- 2: **while**  $n < N$  **do**
- 3:    $n = n + 1$ .
- 4:   Update carrier index  $l_i(n)$  for each user  $i$ .
- 5:   Update metric  $g_i(n)$  for each user  $i$ .
- 6:   Find  $i^*(n) = \arg \max_i g_i(n)$  (break ties arbitrarily).
- 7:   Assign the  $n$ th carrier to user  $i^*(n)$ :

$$\mathcal{K}_i(n) = \begin{cases} \mathcal{K}_i(n-1) \cup \{l_i(n)\}, & \text{if } i = i_n^*; \\ \mathcal{K}_i(n-1), & \text{otherwise.} \end{cases}$$

- 8: **end while**
- 

carrier assignment does not change a user's ordering of the remaining carriers) and can be done in parallel.

Let  $k_i(n) = |\mathcal{K}_i(n)|$ . The choices for Line 5 are:

(5A): Set  $g_i(n)$  to be the total increase in user  $i$ 's utility if assigned carrier  $l_i(n)$ , assuming constant power allocation over all assigned carriers, i.e.,

$$g_i(n) = w_i \left[ \sum_{j \in \mathcal{K}_i(n-1) \cup \{l_i(n)\}} \log \left( 1 + \frac{P_i e_{ij}}{k_i(n-1) + 1} \right) - \sum_{j \in \mathcal{K}_i(n-1)} \log \left( 1 + \frac{P_i e_{ij}}{k_i(n-1)} \right) \right].$$

(5B): Set  $g_i(n)$  to be user  $i$ 's gain from only carrier  $l_i(n)$ , assuming constant power allocation, i.e.

$$g_i(n) = w_i \log \left( 1 + \frac{P_i}{k_i(n-1) + 1} e_{i,l_i(n)} \right).$$

Compared with (5A), this metric ignores the change in user  $i$ 's utility due to the decrease in power allocated to any carriers in  $\mathcal{K}_i(n-1)$ .

##### B. Power Allocation (PA) phase

The objective of the power allocation phase is to optimally allocate each user's power over the carrier allocation  $x_{ij}^*$  determined in the CA phase. For each user  $i$ , the optimal power allocation,  $\mathbf{p}_i = (p_{ij}, j \in \mathcal{N})$  is the solution to:

$$\max_{\mathbf{p}_i \in \mathcal{P}_i} x_{ij}^* \log(1 + p_{ij} e_{ij}) \quad (\text{PA}_i)$$

where  $\mathcal{P}_i = \{\mathbf{p}_i \geq 0 : p_{ij} \leq \frac{s_{ij}}{e_{ij}}, \sum_{j \in \mathcal{N}} p_{ij} \leq P_i\}$ . If  $\sum_{j \in \mathcal{N}} x_{ij}^* \frac{s_{ij}}{e_{ij}} \leq P_i$ , then the solution to (PA <sub>$i$</sub> ) is  $p_{ij}^* = \frac{x_{ij}^* s_{ij}}{e_{ij}}$ . Otherwise, the optimal power allocation is again given by the waterfilling allocation in (10), where the (non-negative) constant  $\lambda_i$  is chosen such that  $\sum_{j \in \mathcal{N}} p_{ij}^* = P_i$ . It is possible to solve this problem in finite time; the details can be found in [16].

#### V. SIMULATION RESULTS

We report simulation results for the 4 versions of SOA as well as an "optimal" algorithm, which iterates to find the optimal  $\lambda$ ; as we discussed this algorithm results in many ties. To limit the complexity when ties occur, we inspect up to 128

TABLE II  
ALGORITHM PERFORMANCE FOR SCHEDULING EVERY 20 OFDM  
SYMBOLS (TOTAL RATE IN MBPS).

Algorithm	Utility	Log U	Total Rate	User Scheduled	
Optimal	994835	509.9	22.13	32.7	
SOA	4A & 5A	983539	505.9	22.23	30.7
	4A & 5B	973365	501.0	22.33	24.4
	4B & 5A	1024306	508.1	23.52	31.0
	4B & 5B	1007144	502.8	23.44	24.8
Base Line	534724	-1960.5	16.13	2.66	

ways of breaking the ties with an integer allocation and select the allocation among these with the largest weighted sum rate. We also give results for a base-line algorithm where each channel  $j$  is allocated to the user  $i$  with the highest  $e_{ij}$ , without considering the weights  $w_i$ 's and the power constraints.

All results are for a single OFDM cell with 40 users. Each user's channel gains are the product of a constant location-based term, picked using an empirically obtained distribution, and a fast fading term, generated using a block-fading model and a standard mobile delay-spread model with a delay spread of  $10\mu\text{sec}$ . The fast-fading component for each multi-path component is held fixed for 2msec and an independent value is generated for the next block, which corresponds to a 250MHz Doppler. The system bandwidth is 5MHz corresponding to 512 OFDM tones. Resource allocation is performed using adjacent groups of 8 tones.<sup>7</sup> The symbol duration is  $100\mu\text{sec}$  with a cyclic prefix of  $10\mu\text{sec}$ . All users are infinitely-backlogged with the same utility function of  $U_i(W_{i,t}) = (W_{i,t})^{0.75}/0.75$  and the same maximum power constraint of  $P_i = 2\text{W}$ . Each simulation run is for 1000 time blocks.

Table II gives the results of the algorithms (summed over all users) when scheduling decisions are made every 20 OFDM symbols. The Log U column denotes the logarithmic utility function, which provides a characterization of fairness among users. The "User Scheduled" column denotes the average number of users who receive positive rates within one scheduling interval. SOA with 4B & 5A gives the best results both in terms of utility and rate. This even performs better than the "optimal" algorithm, which is likely because only 128 ways to break ties are considered (this is typically not sufficient). The base-line algorithm always has poor performance.

Table III shows the performance of each algorithm when scheduling is performed every 80 OFDM symbols, with all other parameters the same as in Table II. It is clear that this coarser allocation leads to poorer performance, while SOA with 4B & 5A still gives the best performance. This shows the tradeoff between system performance and resource allocation frequency (and thus algorithm complexity).

## VI. CONCLUSIONS

We presented an optimization-based formulation for scheduling and resource allocation in the uplink of an OFDM

<sup>7</sup>This corresponds to the "Band AMC mode" of 802.16 d/e.

TABLE III  
ALGORITHM PERFORMANCE WITH SCHEDULING EVERY 80 OFDM  
SYMBOLS (TOTAL RATE IN MBPS)

Algorithm	Utility	Log U	Total Rate	User Scheduled	
Optimal	836853	498.6	17.78	32.8	
SOA	4A & 5A	840524	494.6	18.25	31.1
	4A & 5B	792350	486.3	17.20	24.6
	4B & 5A	857213	496.0	18.77	31.6
	4B & 5B	810850	487.6	17.78	25.2
	Base Line	389927	-2116.5	11.65	2.64

network. Compared to the downlink, we argued that the uplink was computationally more challenging due to the per-user power constraints. A (high complexity) optimal algorithm was given as well as a family of low complexity heuristics. The heuristics were shown to have good performance via simulations.

## REFERENCES

- [1] R. Agrawal, A. Bedekar, R. La, V. Subramanian, "A Class and Channel-Condition based Weighted Proportionally Fair Scheduler," *Proc. of ITC 2001*, Salvador, Brazil, Sept. 2001.
- [2] R. Agrawal and V. Subramanian, "Optimality of Certain Channel Aware Scheduling Policies," *Proc. of 2002 Allerton Conference*, Oct. 2002.
- [3] 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.
- [4] J. Huang, V. G. Subramanian, R. Agrawal and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," submitted to *IEEE Trans. on Wireless*, November 2007.
- [5] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queue with randomly varying connectivity", in *IEEE Transactions on Information Theory*, Vol. 39, pp. 466-478, March 1993.
- [6] R. Leelahakriengkrai and R. Agrawal, "Scheduling in Multimedia Wireless Networks," in *Proc. 17th Int. Teletraffic Congress*, Salvador de Bahia, Brazil, pp. 556-564, December, 2001.
- [7] M. Andrews, K. K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queueing system with asynchronously varying service rates," *Probability in Engineering and Information Sciences*, Volume 18, Number 2, pp.191-217, 2004.
- [8] A. Jalali, R. Padovani, R. Pankaj, "Data throughput of CDMA-HDR a high efficiency - high data rate personal communication wireless system," in *Proc. VTC '2000*, Spring, 2000.
- [9] A. L. Stolyar, "Maximizing Queueing Network Utility subject to Stability: Greedy Primal-Dual Algorithm," *Queueing Systems*, Vol. 50, pp. 401-457, 2005.
- [10] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Operations Research*, Vol. 53, No. 1, pp. 1225, 2005.
- [11] H. Kushner and P. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," in *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, 2002.
- [12] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, Vol. 8, No. 5, pp. 556-567, October 2000.
- [13] S. Pfletschinger, G. Muenz, and J. Speidel, "Efficient subcarrier allocation for multiple access in ofdm systems," in *7th International OFDM-Workshop 2002 (InOWo'02)*, 2002.
- [14] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.
- [15] W. Yu, R. Lui, and R. Cendrillon, "Dual optimization methods for multiuser orthogonal frequency division multiplex systems," in *Proceedings of IEEE Globecom*, vol. 1, 2004, pp. 225-229.
- [16] J. Huang, V. G. Subramanian, R. Berry and R. Agrawal, "Joint Scheduling and Resource Allocation in Uplink OFDM Systems," *journal version in preparation*, November 2007.