# Joint Scheduling and Resource Allocation in Uplink OFDM Systems

Jianwei Huang, Vijay G. Subramanian, Rajeev Agrawal, and Randall Berry

***Abstract*—Orthogonal Frequency Division Multiplexing (OFDM) with dynamic scheduling and resource allocation is widely considered to be a key component of most emerging broadband wireless access networks such as WiMAX and LTE (Long Term Evolution) for 3GPP. However, scheduling and resource allocation in an OFDM system is complicated, especially in the uplink due to two reasons: (1) the discrete nature of subchannel assignments, and (2) the heterogeneity of the users' subchannel 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 and is solved using a dual decomposition approach. This solution has prohibitively high computational complexity but reveals guiding principles that we use to generate several lower complexity sub-optimal algorithms. We analyze the complexity and compare the performance of these algorithms via a realistic OFDM simulator.**

## I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is a promising technology for future broadband wireless networks, due to many of its advantages such as robustness against intersymbol interference and multipath fading as well as low complexity receiver equalization. It is the core technology for a number of wireless data systems, such as IEEE 802.16 (WiMAX), IEEE 802.11a/g (Wireless LANs), and LTE for 3GPP. In this paper, we analyze an uplink scheduling and resource allocation problem for OFDM wireless access networks. The specific problem is motivated by the WiMAX/802.16e standard, where there is a centralized scheduler that knows the QoS classes, and can estimate the queue-lengths on each mobile device. The WiMAX/802.16e standard specifies mechanisms for communicating this information to the scheduler and for conveying the scheduling decisions to the mobiles, both with low delays.[1]

Our approach to the uplink OFDM scheduling problem is motivated by our previous work on downlink scheduling in CDMA systems [3] and OFDM systems [4]. As in [3], [4], we consider the use of a *gradient-based* scheduling framework, which is described in detail in Section II along with our system model. In this framework, the gradient of a time-varying utility function is used to guide the resource allocation decisions and provide long-term Quality of Service (QoS) guarantees. In an uplink OFDM system, this rule reduces to maximizing a weighted sum-rate during each scheduling interval, where the weights are time-varying. The optimization variables are the assignment of OFDM subchannels to the users and the allocation of each user's power across the assigned subchannels. We highlight two challenging aspects of this problem. First, the (well-known) discrete nature of subchannel assignments in OFDM systems usually leads to hard integer programming problems. Second, the per-user power constraint that arises in the uplink makes the problem even less tractable. We initially relax the integer constraints and consider a mathematical abstraction in which multiple users can share one subchannel/tone using orthogonalization (e.g. via time-sharing[2]). In Section III we derive an optimal solution to this relaxed problem using a dual decomposition approach. This provides insight into the structure of an optimal solution; however, due to the per-user power constraints the resulting algorithm has a very high computational complexity. In Section V 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 subchannel. Finally, in Section VI we present numerical results for these algorithms using a realistic OFDM simulator.

Most initial work on scheduling and resource allocation in OFDM systems focused on the downlink case. The optimality conditions derived and algorithms proposed for the downlink case, however, can not be directly applied to the uplink due to differences in terms of network resource constraints; a more detailed discussion on this is provided in Section IV. Recently, uplink OFDM resource allocation has received some attention, including [21]–[28]. In [21], a resource allocation problem was formulated in the framework of Nash Bargaining, and an iterative algorithm was proposed with relatively high complexity. The authors of [22] proposed a heuristic algorithm

J. Huang is with the Department of Information Engineering, The Chinese University of Hong Kong, e-mail: jwhuang@ie.cuhk.edu.hk. V. G Subramanian is with the Hamilton Institute, National University of Ireland, email: vijay.subramanian@nuim.ie. R. Agrawal is with the Advanced Networks and Performance Dept., Motorola Inc., e-mail: Rajeev.Agrawal@motorola.com. R. Berry is with the Dept. of EECS, Northwestern University, e-mail: rberry@ece.northwestern.edu.

Part of this work was done while J. Huang and V. G. Subramanian were at Motorola. J. Huang is supported by the Competitive Earmarked Research Grants (Project Number 412308) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, and the Direct Grant (Project Number C001-2050398) of The Chinese University of Hong Kong. V. Subramanian is supported by SFI grant 03/IN3/I346. R. Berry was supported in part by the Motorola-Northwestern Center for Seamless Communications and NSF CAREER award CCR-0238382. The work was partially presented at the 2007 Asilomar Conference on Signals, Systems and Computers.

[1] Our model is also appropriate for LTE [29] for 3GPP, UMB [30] for 3GPP2 and the FLASH OFDM system [19] from Qualcomm Flarion.

[2] While super-position coding would yield an even larger (and more tractable) capacity region, we do not consider it as it is still not practical.

that tries to minimize each user's transmission power while satisfying the individual rate constraints. In [23], the author considered the sum-rate maximization problem, which is a special case of the problem considered here with equal weights. The algorithm derived in [23] assumes Rayleigh fading on each subchannel; we do not make such an assumption here. In [24], an uplink problem with multiple antennas at the base station was considered; this enables spatial multiplexing of subchannels among multiple users. Here, we focus on single antenna systems where at most one user can be assigned per sub-channel.

The work in [25]–[28] is closer to our model. The authors in [25] also considered a weighted rate maximization problem in the uplink case, but assumed static weights. They proposed two algorithms, which are similar to one of the algorithms proposed in this paper. We propose several other algorithms that outperform those in [25] with similar or slightly higher complexity. Paper [26] generalized the results in [25] by considering the utility maximization in one time-slot, where the utility is a function of the instantaneous rate in this time-slot. Another work that focused on per time-slot fairness is [27]. [28] proposed a heuristic algorithm based on Lagrangian relaxation, which has high complexity due to a subgradient search of the dual variables.

None of the prior work mentioned above considered long-term total utility maximization, where utility is a function of long-term throughput and/or queue length. The gradient-based scheduler which we consider here easily accommodates this. For elastic data applications, long-term QoS evaluation is more reasonable than the short-term QoS evaluation during each time slot. It not only more faithfully reflects users' actual perceived performance, but also gives the system more flexibility in terms of exploiting multi-user diversity. Another contribution of our work is the formulation and solution of the relaxed optimization problem, which captures network performance when time sharing is allowed. Solving this problem provides both an upperbound on the actual system performance as well as the intuition we use to design good heuristic algorithms.

## II. Problem Statement

We consider the problem of scheduling and resource allocation for the uplink of a single OFDM cell in which a set $\mathcal{M} = \{1, \ldots, M\}$ of users are transmitting to the same base station. Each user $i \in \mathcal{M}$ has a total transmission power constraint $P_i$. The total frequency band is divided into a set $\mathcal{N} = \{1, \ldots, N\}$ of subchannels (e.g., frequency bands), and a user $i$ can transmit over a subset of the subchannels, with transmit power $p_{ij}$ over subchannel $j$. Let $x_{ij}$ denote the fraction of subchannel $j$ allocated to user $i$, where the total allocation across all users should be no larger than 1, i.e., $\sum_i x_{ij} \leq 1$.

Time is divided into equal length slots. At every scheduling epoch (i.e., the beginning of every time slot), the scheduler seeks to maximize a (time-varying) weighted sum of the users' rates over a given (time-varying) rate-region. We will describe the feasible rate region in details next. The key notation is listed in Table I. We use bold symbols to denote vectors and matrices of these quantities, e.g., $\boldsymbol{w} = \{w_i, \forall i\}$, $\boldsymbol{e} = \{e_{ij}, \forall i, j\}$, $\boldsymbol{p} = \{p_{ij}, \forall i, j\}$, and $\boldsymbol{x} = \{x_{ij}, \forall i, j\}$.

TABLE I
KEY NOTATIONS

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

We assume that the scheduler has knowledge of the normalized received signal-to-interference plus noise ratio (SINR) per unit transmit power, $e_{ij}$, for each user $i$ and each subchannel $j$.[3] We represent the time-varying subchannel quality vector at time $t$ as $\boldsymbol{e}_t$. As in [4], this model can also incorporate various subchannelization schemes where the resource allocation is performed in terms of subchannels (i.e., sets of tones). In this case, $e_{ij}$ represents the quality indicator for the subchannel, e.g., the (harmonic/geometric/arithmetic) average across the tones in the subchannel. This model also applies if resource allocation is done with a granularity of multiple symbols in the time domain.

Let $\mathcal{R}(\boldsymbol{e}_t)$ denote the feasible rate region at time $t$, i.e.,

$$\mathcal{R}(\boldsymbol{e}_t) = \{\boldsymbol{r} \in \Re_+^M : r_i = \sum_{j \in \mathcal{N}} x_{ij} \log\left(1 + \tfrac{p_{ij} e_{ij}}{x_{ij}}, \right), \forall i \in \mathcal{M}\}, \tag{1}$$

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

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

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

and the set

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

Here, $r_i$ is the achievable rate of user $i$. By continuity, we assume that that $x_{ij} \log(1 + \frac{p_{ij} e_{ij}}{x_{ij}}) = 0$ if $x_{ij} = 0$. The value of $s_{ij}$ is a maximum SINR constraint on subchannel $j$ for user $i$. This could arise, for example, when users have limited choices of modulation and coding schemes. With such limitations allocating a power that is too high will not increase the goodput, and would result in wastage.[4]

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

[3] In both FDD and TDD systems this can be obtained using a combination of measurements made on the uplink pilots as well as past transmissions from the mobiles.

[4] The rate region corresponds to a Gaussian multiple-access channel with time sharing as given in [17, Section 15.3.6, pg. 547].

resource allocation is done on blocks of OFDM symbols, then fractional values of $x_{ij}$ can be implemented by time-sharing the symbols in a block.[5] We will bring such integer constraints back to the problem in Sections IV and V.

Next we formulate the scheduling and resource allocation problem. Our approach is based on the gradient-based scheduling framework presented in [2], [10], [12]. 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 $\boldsymbol{r}_t$ in $\mathcal{R}(\boldsymbol{e}_t)$ 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 , i.e., we solve

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} (\nabla_{\boldsymbol{w}} U(\boldsymbol{W}_t, \boldsymbol{Q}_t)^T - \nabla_{\boldsymbol{q}} U(\boldsymbol{W}_t, \boldsymbol{Q}_t)^T) \cdot \boldsymbol{r}_t,$$

where $U(\boldsymbol{W}_t, \boldsymbol{Q}_t) = \sum_{i=1}^{K} U_i(W_{i,t}, Q_{i,t})$. We further assume that for each user $i$ we have $U_i(W_{i,t}, Q_{i,t}) = u_i(W_{i,t}) - \frac{d_i}{p}(Q_{i,t})^p$, where $u_i(W_{i,t})$ is a increasing concave function used to represent elastic data applications [18], $d_i \geq 0$ is a QoS weight for user $i$'s queue length, and $p > 1$ is a fairness parameter associated with the queue length. With this the scheduling and resource allocation decision is the solution to

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i \left( \frac{\partial u_i(W_{i,t})}{\partial W_{i,t}} + d_i (Q_{i,t})^{p-1} \right) r_{i,t}. \qquad (5)$$

The broad class of policies in (5) can be tuned to yield good operating points by a proper choice of parameters. If $d_i = 0$ for all $i \in \mathcal{M}$, the resulting policy has been shown to yield utility maximizing solutions [2], [10], [12]. If $u_i(\cdot) \equiv 0$ with $d_i > 0$ for all $i \in \mathcal{M}$, then the 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]. The weights can also be adapted [11] so as to maximize sum utility subject to (feasible) minimum throughput constraints. As a concrete example, for $u_i(\cdot)$ we consider the class of isoelastic utility functions (e.g. [1], [13]) given by

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

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 $\sum_i \left( c_i (W_{i,t})^{\alpha-1} + d_i (Q_{i,t})^{p-1} \right) r_{i,t}$. 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 $\alpha = 0$, one gets the *proportionally fair* rule [8].

More generally, the optimization in (5) can be written as

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

where $w_{i,t} \geq 0$ is a time-varying weight assigned to the $i$th user at time $t$. Our focus in this paper is on solving such a problem for an uplink OFDM system, i.e., when $\mathcal{R}(\boldsymbol{e}_t)$ is given by (1). We emphasize that (7) must be re-solved at each scheduling instant because of changes in both the subchannel state, $\boldsymbol{e}_t$, and the weights (e.g., the gradient of the utility). We also note that while in the above examples, the weights $w_{i,t}$ were given by the gradient of the utility function, our algorithms also apply if other methods for generating these weights are used.

## III. Optimal Solution with Fractional Allocations

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

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

subject to the per subchannel 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 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,

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

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

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

We solve this problem by the following steps. First, we analytically find the optimal $\boldsymbol{p}$ and $\boldsymbol{x}$ given fixed values of the dual variables. We then show that the optimal $\boldsymbol{\mu}$ is given by a search for the maximum value of a per-user metric on each subchannel. The final step is to numerically search for the optimal value of $\boldsymbol{\lambda}$.

The value of $\boldsymbol{p}$ which maximizes $L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{p})$ given $\boldsymbol{x}$, $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ is given by

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

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 a water-filling solution which takes into account the maximum SINR constraint. Substituting $\boldsymbol{p}^*$ into $L(\cdot,\cdot,\cdot,\cdot)$ yields

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}) = \sum_{ij} x_{ij} \left( w_i h\left(\lambda_i, w_i e_{ij}, s_{ij}\right) - \mu_j \right) + \sum_j \mu_j + \sum_i \lambda_i P_i, \qquad (11)$$

[5]Likewise, if the number of subchannels are large enough so that the subchannel gains do not change dramatically among adjacent subchannels, then the fractional value of $x_{ij}$ can also implemented by frequency sharing (e.g., [31]).

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

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

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

$$L(\boldsymbol{\lambda},\boldsymbol{\mu})=\sum_j \mu_j+\sum_i \lambda_i P_i + \sum_{ij}\left(w_i h\left(\lambda_i, w_i e_{ij}, s_{ij}\right)-\mu_j\right)^+, \tag{13}$$

where the optimal subchannel allocation has the following structure

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

Since the cost function in (13) is separable, by defining $\mu_{ij}(\cdot):=w_i h\left(\cdot, w_i e_{ij}, s_{ij}\right)$ we can minimize $L(\boldsymbol{\lambda},\boldsymbol{\mu})$ over $\boldsymbol{\mu}$ for a given $\lambda$ by a simple sort per subchannel similar to that in [3], i.e. the optimal $\mu_j^*(\boldsymbol{\lambda})$ is given by

$$\mu_j^*(\boldsymbol{\lambda})=\max_i \mu_{ij}\left(\lambda_i\right). \tag{15}$$

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

$$L(\boldsymbol{\lambda})=\sum_j \max_i \mu_{ij}(\lambda_i)+\sum_i \lambda_i P_i.$$

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 appropriately [20]. There are many variants of subgradient search algorithms with different convergence properties; we choose to use the one described in [20, Exercise 6.3.2]. Given an optimal $\boldsymbol{\lambda}^*$, by duality, $L(\boldsymbol{\lambda}^*)$ is the optimal objective value to Problem UL.

However, to implement this algorithm, the scheduler must specify the corresponding optimal primal values of $(\boldsymbol{x}^*,\boldsymbol{p}^*)$. Here, as in [4], more care is required. Specifically, when ties occur in (15), it is often needed to split the subchannel among several users (i.e., allowing fractional values of $\boldsymbol{x}^*$). Following a similar approach as in [3] (for the downlink), the optimal fractional values can be found by solving a linear program whose size increases with the number of users and tones involved in each tie. As we discuss below, this number can be quite large in the uplink setting. Moreover, as noted earlier, one is typically interested in an integer allocation in practice, i.e. each subchannel can be allocated to at most one user. We consider this problem next.

## IV. Integer Subchannel Allocation Based on Optimal Algorithm

We now address the problem:

$$\max_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X},x_{ij}\in\{0,1\},\forall i,j}\sum_{i\in\mathcal{M}} w_i\sum_{j\in\mathcal{N}} x_{ij}\log\left(1+\frac{p_{ij}e_{ij}}{x_{ij}}\right). \tag{UL-Integer}$$

Initially, consider the following heuristic for Problem UL-Integer: (*i*) Solve Problem UL as in the previous section, and (*ii*) "break" any ties on all subchannels, i.e., whenever there is a fractional $x_{ij}^*$ value, choose one user in the tie and allocate subchannel $j$ to that user only. Clearly, if there are no ties to break, this algorithm gives the optimal solution to Problem UL-Integer. After all ties are broken, we can then re-optimize the power allocation for each user so that each individual power constraint is tight; this can be accomplished using a finite-time *water-filling* algorithm as in [4].

In [4] we used a similar algorithm to solve a downlink OFDM scheduling problem (i.e., we first solved the dual problem by allowing fractional subchannel allocation, and then broke any ties to find an integer subchannel allocation). However, there are several major differences between the uplink and downlink settings that make this approach less appealing for implementation in the uplink case. 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 $L(\lambda)$ is a function of only a single dual variable $\lambda$, which simplifies the numerical search for the optimal $\lambda$. In the uplink setting, the convergence of the subgradient search is too slow to be useful for scheduling on a fast time-scale.

Second, even if the optimal $\boldsymbol{\lambda}$ can be found, breaking ties is more difficult than in the downlink case. One reason for this is that having a single dual variable provides simple tie-breaking rules for the downlink based on the (scalar) subgradients of $L(\lambda)$ (see [4]); in the uplink case, the subgradients are vectors, complicating such an approach. 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 subchannel $j$ $(e_{ij}=e_{lj})$, then allocating subchannel $j$ 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 totally 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 subchannels. Each user has the same gain over both subchannels, 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 subchannel than user 1 so that in the downlink case, the unique optimal solution is to allocate both subchannels 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 $\mu_{1j}(\lambda_1) = \mu_{2j}(\lambda_2)$ for $j = 1$ and 2, i.e., there is a tie in each subchannel and we have to compare four possible subchannel allocations to determine how to break the tie. This can be easily extended to $M$ users and $N$ subchannels, with each user having the same gain over all its subchannels. This results in $M^N$ ties, independent of the variation in gains across users.

## V. Low Complexity Suboptimal Algorithms

Due to the issues discussed in the previous section, finding the optimal dual solution to Problem UL and breaking any ties to get an integer allocation is not computationally feasible, even for 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 performance. These algorithms seek to exploit the problem structure revealed by the optimal algorithm. Furthermore, all of these sub-optimal algorithms enforce an integer tone allocation during each scheduling interval.

In the optimal algorithm, given the optimal $\boldsymbol{\lambda}^*$, the optimal subchannel 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 subchannel 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 $\boldsymbol{\lambda}^*$ and the optimal subchannel allocation. Specifically, we begin with a *subChannel Allocation (CA)* phase in which we assign each sub-subchannel to at most one user. We consider two different SOAs that implement the CA phase differently. In SOA1, instead of using the metric given by the optimal $\boldsymbol{\lambda}$, we consider metrics based on a constant power allocation over all subchannels assigned to a user. In SOA2, we find the subchannel allocation, once again through a dual based approach, but here we first determine the number of subchannels assigned to each user and then match specific subchannels and users. After the subchannel allocation is done in both SOAs, we perform a *Power Allocation (PA)* phase in which each user's power is allocated across the assigned subchannels using a water-filling allocation as in the optimal algorithm.

### *A. CA in SOA1: Progressive subchannel Allocation based on Metric Sorting*

In this family of SOAs, subchannels are assigned sequentially in one pass based on a per user metric for each subchannel, i.e., we iterate $N$ times, where each iteration corresponds to the assignment of one subchannel. Let $\mathcal{K}_i(n)$ denote the set of subchannels 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 subchannel index that user $i$ would like to be assigned if he/she is assigned the $n$th subchannel. The resulting CA algorithm is given in Algorithm 1. Note that all the user metrics are updated after each subchannel is assigned.

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

**Algorithm 1** CA Phase for SOA1

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

*(4A):* Sort the subchannels based on the best channel condition among *all* users. This involves two steps. First, for each subchannel $j$, find the best channel condition among all users and denote it by $\tilde{\mu}_j := \max_i e_{ij}$. Second, find a subchannel permutation $\{\alpha_j\}_{j\in\mathcal{N}}$ such that $\tilde{\mu}_{\alpha_1} \geq \tilde{\mu}_{\alpha_2} \geq \cdots \geq \tilde{\mu}_{\alpha_N}$, and set $l_i(n) = \alpha_n$ for each user $i$ at the $n$th iteration. Each $\max$ operation has complexity of $O(M)$, and the sorting operation has a complexity of $O(N\log(N))$. The total complexity is $O(NM + N\log N)$. We note that this is a one-time "pre-processing" that needs to done before the CA phase starts. During the subchannel allocation iterations, the users just choose the subchannel index from the sorted list.

*(4B):* Sort the subchannels based on the channel conditions for each *individual* user. For each user $i$ at the $n$th iteration, set $l_i(n)$ to be the subchannel index with the largest gain among all unassigned subchannels, 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 subchannel assignment does not change a user's ordering of the remaining subchannels) and can be done in parallel. The total complexity of the $M$ sorting operations is $O(MN\log N)$, which is higher than that in *(4A)*.

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 subchannel $l_i(n)$, assuming constant power allocation over all assigned subchannels, i.e.,

$$g_i(n) = w_i\Bigg[\sum_{j\in\mathcal{K}_i(n-1)\cup\{l_i(n)\}} \log\left(1 + \frac{P_i e_{ij}}{k_i(n-1)+1}\right) - \sum_{j\in\mathcal{K}_i(n-1)} \log\left(1 + \frac{P_i e_{ij}}{k_i(n-1)}\right)\Bigg]. \quad (16)$$

*(5B):* Set $g_i(n)$ to be user $i$'s gain from only subchannel $l_i(n)$, 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 subchannels in $\mathcal{K}_i(n-1)$.

The complexity of either of these choices over $N$ iterations is $O(NM)$, and so the total complexity for the CA phase [6] is $O(NM + N\log N)$ (if $(4A)$ is chosen) or $O(MN\log N)$ (if $(4B)$ is chosen).

### *B. CA in SOA2: subchannel Number Assignment & subchannel User Matching*

SOA2 implements the CA phase through two steps: subchannel number assignment (CNA) and subchannel user matching (CUM). The algorithm is summarized in Algorithm 2.

**Algorithm 2** CA Phase of SOA2

1: subChannel Number Assignment (CNA) step: determine the number of subchannels $n_i$ allocated to each user $i$ such that $\sum_{i\in\mathcal{M}} n_i \leq N$.

2: subChannel User Matching (CUM) step: determine the subchannel assignment $x_{ij} \in \{0,1\}$ for all users $i$ and subchannels $j$, such that $\sum_{j\in\mathcal{N}} x_{ij} = n_i$.

*1) subChannel Number Assignment (CNA):* In the CNA step, we determine the number of subchannels $n_i$ assigned to each user $i \in \mathcal{M}$. The assignment is calculated based on the approximation that each user sees a flat wide-band fading subchannel. Notice that here we do not specify which subchannel is allocated to which user; such a mapping will be determined in the CUM step. The CNA step is further divided into two stages: a basic assignment stage and an assignment improvement stage.

*Stage 1, Basic Assignment*: Here, the assignment is based on the normalized SINR averaged over all subchannels. Specifically, we model each user $i$ as having a normalized SINR $\bar{e}_i = \frac{1}{N}\sum_{j\in\mathcal{N}} e_{ij}$, and then determine a subchannel number assignment $n_i$ for all $i$ by solving:

$$\begin{aligned} \max_{\{n_i\geq 0, i\in\mathcal{M}\}} & \sum_{i\in\mathcal{M}} w_i n_i \log\left(1+\frac{P_i}{n_i}\bar{e}_i\right) \\ \text{subject to: } & \sum_{i\in\mathcal{M}} n_i \leq N. \end{aligned} \quad \text{(SOA2-CNA)}$$

It can be shown that Problem SOA2-CNA is a standard concave maximization problem over a convex set with a unique and possibly non-integer solution; we use a dual relaxation method to find this solution. Consider the Lagrangian

$$L(\boldsymbol{n},\lambda) := \sum_{i\in\mathcal{M}} w_i n_i \log\left(1+\frac{P_i}{n_i}\bar{e}_i\right) - \lambda\left(\sum_{i\in\mathcal{M}} n_i - N\right).$$

Optimizing $L(\boldsymbol{n},\lambda)$ over $\boldsymbol{n} = (n_i, i\in\mathcal{M})$ for a given $\lambda$ is equivalent to solving the following $M$ subproblems,

$$n_i(\lambda) = \arg\max_{n_i}\left(w_i n_i \log\left(1+\frac{P_i}{n_i}\bar{e}_i\right) - \lambda n_i\right), \forall i. \quad (17)$$

Problem (17) can be solved by a simple line search over the range of $(0, N]$. Substituting the corresponding results into the Lagrangian yields

$$\begin{aligned} L(\lambda) := & \sum_{i\in\mathcal{M}} w_i n_i^*(\lambda) \log\left(1+\frac{P_i}{n_i^*(\lambda)}\bar{e}_i\right) \\ & - \lambda\left(\sum_{i\in\mathcal{M}} n_i^*(\lambda) - N\right), \end{aligned}$$

which is a convex function of $\lambda$ [20]. The optimal value

$$\lambda^* = \arg\min_{\lambda\geq 0} L(\lambda) \quad (18)$$

can be found by a line section search over $[0, \max_i(w_i \log(1+P_i\bar{e}_i))]$[7]. For a given search precision, the maximum number of iterations needed to solve either (17) or (18) is fixed.[8]. Hence, the worst case complexity of the solving each subproblem is independent of $M$ or $N$. Since there are $M$ subproblems in (17), it follows that the complexity of the basic assignment step is $O(M)$. If the resultant channel allocations contain non-integer values, we will approximate with an integer solution that satisfies $\sum_{i\in\mathcal{M}} n_i = N$.[9] Since each user is allocated only a subset of the subchannels, the normalized SINR $\bar{e}_i = \frac{1}{N}\sum_{j\in\mathcal{N}} e_{ij}$ is typically a pessimistic estimate of the averaged subchannel conditions over the allocated subset. This motivates us to consider the following assignment improvement stage of CNA.

*Stage 2, Assignment Improvement*: Here, assignment is performed by means of iterative calculations using the normalized SINR averaged over the best subchannel subset. Specifically, we iteratively solve the following variation of Problem SOA2-CNA:

$$\begin{aligned} \max_{\{n_i(t)\geq 0, i\in\mathcal{M}\}} & \sum_{i\in\mathcal{M}} w_i n_i(t) \log\left(1+\frac{P_i}{n_i(t)}\bar{e}_i(t)\right) \\ \text{subject to: } & \sum_{i\in\mathcal{M}} n_i(t) \leq N, \end{aligned} \quad \text{(SOA2-CNA-t)}$$

for $t = 1, 2, \ldots$. During the $t$-th iteration, $\bar{e}_i(t)$ is a refined estimate of the normalized SINR based on the best $n_i(t-1)$ subchannels of user $i$. The iteration stops when the subchannel allocation converges or the maximum number of iterations allowed is reached. An integer approximation will be performed if needed.

The complete algorithm for the CNA phase of SOA2 is given in Algorithm 3. In order to perform the assignment

[6] We note that SOA1 with $(4B)$ and $(5B)$ is similar to the algorithms proposed in [25]. In Section VI we show that other variations of SOA1 ($(4B)$ and $(5A)$) and SOA2 can achieve better performance with similar or slightly higher complexity.

[7] The upperbound of the search interval is obtained by examining the first order optimality condition of (17).

[8] For example, if we use bi-section search to solve (17) and stop when the relative error of the solution is less than $N/2^{10}$, then we only need a maximum of ten search iterations.

[9] One possible integer approximation is the following. Assume $n_i^*$ is the unique optimal solution of Problem SOA2-CNA. First, sort users in the descending order of the mantissa of $n_i^*$, $fr(n_i^*) = n_i^* - \lfloor n_i^* \rfloor$. That is, find a user permutation subset $\{\alpha_k, 1 \leq k \leq N\}$ such that $fr(n_{\alpha_1}^*) \geq fr(n_{\alpha_2}^*) \geq \cdots \geq fr(n_{\alpha_M}^*)$. Second, calculate the number of unallocated subchannels, $N^A = N - \sum_i \lfloor n_i^* \rfloor$. Finally, adjust users with large mantissas such that all the subchannels are allocated, i.e., $\tilde{n}_{\alpha_i}^* = \lfloor n_i^* \rfloor + 1$ for all $1 \leq i \leq N^A$. The resulting $\{\tilde{n}_i^*\}_{i\in\mathcal{M}}$ are the integer approximation.

improvement, we need to perform $M$ sorting operations, with a total complexity $O(MN\log(N))$. Note that this only needs to be done once. Step 4 of each iteration has complexity of $O(M)$ due to solving $M$ subproblems for a fixed dual variable. The maximum number of iterations is fixed and thus is independent of $N$ or $M$. The integer approximation stage requires a sorting with the complexity of $O(M\log(M))$. So the total complexity for the CNA phase of SOA2 is $O(MN\log(N)+M\log(M))$.

**Algorithm 3** CNA Phase of SOA2

1: Initialization: integer MaxIte$>0$, $t=0$, $n_i(0)=N$ and $n_i(1)=N/2$ for each user $i$.
2: **while** $(n_i(t+1)\neq n_i(t)$ for some $i)$ & $(t<$MaxIte$)$ **do**
3: $\quad t=t+1$.
4: $\quad$ For each user $i$, $\bar{e}_i(t)=$ average gain of user $i$'s best $n_i(t-1)$ subchannels.
5: $\quad$ Solve Problem (SOA2-CNA-t) to determine the optimal $n_i(t)$ for each user $i$.
6: **end while**
7: let $n_i^*=n_i(t)$ for each user $i$.

*2) subChannel User Matching (CUM) Step:* After the CNA step, we know how many subchannels are to be allocated to each user. However, we still need to determine which specific subchannels are assigned to which user. This is accomplished in the CUM step by finding a subchannel assignment that maximizes the weighted-sum rate assuming each user employs a flat power allocation, i.e. we solve the problem:

$$\begin{aligned}\max_{x_{ij}\in\{0,1\}} & \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{N}} x_{ij}w_i\log\left(1+\frac{P_i}{n_i^*}e_{ij}\right)\\ \text{subject to: } & \sum_{j\in\mathcal{N}} x_{ij}=n_i^*,\forall i\in\mathcal{M}, \qquad \text{(SOA2-CUM)}\\ & \sum_{i\in\mathcal{M}} x_{ij}=1,\forall j\in\mathcal{N},\end{aligned}$$

where $\boldsymbol{n}^*=(n_i^*, i\in\mathcal{M})$ is the integer subchannel allocation obtained in the CNA step.

Problem SOA2-CUM is an integer *Assignment Problem* whose *optimal* solution can be found by using the *Hungarian Algorithm* [15].[10] To use the Hungarian algorithm in our problem, we need to perform “virtual user splitting” as explained next. For user $i$, let $r_{ij}=w_i\log\left(1+\frac{P_i}{n_i^*}e_{ij}\right)$, and let

$$\boldsymbol{r}_i=[r_{i1},r_{i2},\cdots,r_{iN}]$$

be user $i$'s achievable rates over all possible subchannels. We can then form a $M\times N$ matrix $\boldsymbol{R}=\left[\boldsymbol{r}_1^T,\boldsymbol{r}_2^T,\cdots,\boldsymbol{r}_M^T\right]^T$[11]. Next, we split each user $i$ into $n_i^*$ virtual users by adding $n_i^*-1$ copies of the row vector $\boldsymbol{r}_i$ to the matrix $\boldsymbol{R}$. This expands $\boldsymbol{R}$ into a $N\times N$ square matrix. Solving Problem SOA2-CUM is then equivalent to finding a *permutation matrix* $\boldsymbol{C}^*=[c_{ij}]_{N\times N}$ such that

$$\boldsymbol{C}^*=\arg\min_{\boldsymbol{C}\in\mathcal{C}} -\boldsymbol{C}\cdot\boldsymbol{R}:=\arg\min_{\boldsymbol{C}\in\mathcal{C}} -\sum_{i=1}^N\sum_{j=1}^N c_{ij}r_{ij}. \qquad (19)$$

Here $\mathcal{C}$ is the set of permutation matrices, i.e., for any $\boldsymbol{C}\in\mathcal{C}$, we have $c_{ij}\in\{0,1\}$, $\sum_i c_{ij}=1$ and $\sum_j c_{ij}=1$ for all $i$ and $j$. This problem can be solved by the standard Hungarian algorithm which has a computational complexity of $O\left(N^3\right)$, where $N$ is the total number of subchannels. The detailed algorithm can be found in [15]. After obtaining $\boldsymbol{C}^*$, we can calculate the corresponding subchannel allocation $\boldsymbol{x}^*$. For example, if $c_{kj}^*=1$ and virtual user $k$ corresponds to the actual user $i$, then we know $x_{ij}^*=1$, i.e., subchannel $j$ is allocated to user $i$ only.

### *C. Power Allocation (PA) phase with Single-user Water-filling*

After the subchannel allocation (CA) phase in either SOA1 and SOA2, each user still needs to allocate its transmission power across all subchannels that are allocated to it. For user $i$, this can be formulated as the following problem

$$\max_{\boldsymbol{p}_i\in\mathcal{P}_i}\sum_j x_{ij}^*\log\left(1+p_{ij}e_{ij}\right), \qquad (\text{PA}_i)$$

where

$$\mathcal{P}_i=\left\{\boldsymbol{p}_i\geq 0: p_{ij}\leq\frac{s_{ij}}{e_{ij}}, \sum_{j\in\mathcal{N}}p_{ij}\leq P_i\right\}.$$

This can be done by using a water-filling type of power allocation, taking into account the per subchannel SINR constraint. Specifically, if $\sum_{j\in\mathcal{N}}x_{ij}^*\frac{s_{ij}}{e_{ij}}\leq P_i$, we let $p_{ij}^*=\frac{s_{ij}}{e_{ij}}$. Otherwise, the optimal power allocation is determined by

$$p_{ij}^*=\left(\frac{x_{ij}^*}{\nu_i}-\frac{1}{e_{ij}}\right)^+\wedge\frac{s_{ij}}{e_{ij}},$$

where the constant $\nu_i$ is chosen so that $\sum_{j\in\mathcal{N}}p_{ij}^*=P_i$. Note that it is possible that some subchannel is allocated to user $i$ but gets zero transmission power due to its poor subchannel gain. The optimal value of $\nu_i$ can be found through a simple line search, with a constant worst case complexity given a fixed search precision as in our discussion of (17). Note that this typically leads to an *approximate* value of the $v_i$, although in practice it will be good enough if the search precision is set small enough. Alternatively, we can use a finite-time algorithm (with a maximum of $2n_i^*$ steps) to calculate the exact value of $v_i$ as in [3], [4]. The key difference the search only needs to be done once in the downlink case [3], [4], but here we need to run the algorithm up to $M$ times, once for each user which is allocated a subchannel. Since $\sum n_i^*=N$ , it follows that the total worst case computational complexity for this approach is $O(MN)$.

[10] A similar idea has been used to solve various downlink OFDM resource allocation problems (e.g., [16]) as well as to find user coalitions for Nash Bargaining in an uplink OFDM system in [21].

[11] Here we assume that each user is allocated at least one subchannel. If only $\tilde{M}<M$ users are allocated positive amount of channels, we can replace $M$ by $\tilde{M}$ in the discussions.

TABLE II
WORST CASE COMPUTATIONAL COMPLEXITY OF SUBOPTIMAL ALGORITHMS

| Suboptimal Algorithms | | | Worst Case Complexity |
|---|---|---|---|
| SOA1 | subChannel Allocation | 4A & 5A | $O(NM + N \log N)$ |
| | | 4A & 5B | $O(NM + N \log N)$ |
| | | 4B & 5A | $O(MN \log N)$ |
| | | 4B & 5B | $O(MN \log N)$ |
| | Power Allocation | | $O(MN)$ |
| | **Total** | | $\boldsymbol{O}(MN \log N)$ |
| SOA2 | subChannel Allocation | CNA | $O(MN \log N + M \log M)$ |
| | | CUM | $O(N^3)$ |
| | Power Allocation | | $O(MN)$ |
| | **Total** | | $\boldsymbol{O}(N^3 + MN \log N + M \log M)$ |

### D. Complexity Summary of Suboptimal Algorithms

The worst case computational complexities of SOA1 and SOA2 are summarized in Table II. Notice that both algorithms have polynomial time complexities, which is much better than the tie-breaking integer allocation algorithm proposed in Section IV with a worst case complexity of $O(M^N)$.

## VI. SIMULATION RESULTS

We report simulation results for the following four algorithms:

1) Integer-Dual: integer subchannel allocation (with tie breaking) based on optimal algorithm as in Section IV and power control as in Section V-C. Since in general many ties exist after solving Problem UL, it is too time-consuming to compare all possible ties and to break them with the best one. To limit the run time when ties occur, we inspect up to 128 ways of breaking the ties with an integer allocation and select the allocation among these with the largest weighted sum rate (before reallocating the power).
2) SOA1: subchannel allocation as in Section V-A and power control as in Section V-C. There are four versions of SOA1, depending on how steps 4 and 5 in Algorithm 1 are implemented; we present results for each.
3) SOA2: subchannel allocation as in Section V-B (with up to 10 iterations) and power control as in Section V-C.
4) Base-line: each subchannel $j$ is allocated to the user $i$ with the highest $e_{ij}$, without considering the weights $w_i$'s and the power constraints. Each user's power is then allocated as in Section V-C.

We consider a system bandwidth of 5MHz consisting of 512 OFDM tones, grouped into 64 subchannels (8 adjacent tones per subchannel, i.e., corresponding to the "Band AMC mode" of 802.16 d/e.). Each user's subchannel gains are the product of a constant location-based term, picked using an measurement-based 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 $10\mu$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 250Hz Doppler. The empirical cumulative distribution function (CDF)

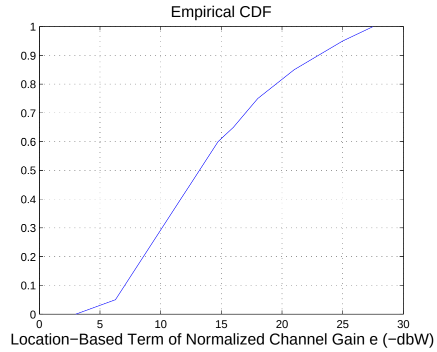


Fig. 1. Empirical CDF of the location-based term in users' normalized channel condition $e_{ij}$'s for all user $i$ and subchannel $j$

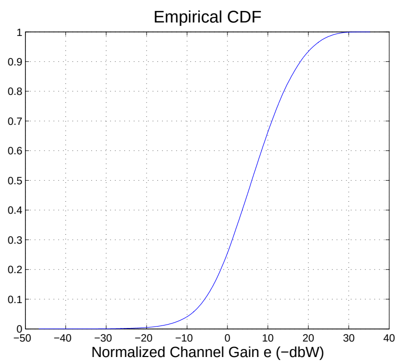


Fig. 2. Empirical CDF of users' normalized channel condition $e_{ij}$'s for all user $i$ and subchannel $j$

of the location-based term of users' location-based term of the $e_{ij}$'s is given in Fig. 1, and the empirical CDF of the net $e_{ij}$'s (i.e., including the fast-fading component) is given in Fig. 2. Since the units of $e_{ij}$ are in 1/watt, the x-axis in both figures is measured in terms of -dbW. The symbol duration is $100\mu$sec with a cyclic prefix of $10\mu$sec, which roughly corresponds to 20 OFDM symbols per fading block (i.e., $2m$sec). This is one of the allowed configurations in the IEEE 802.16 standards [14]. The resource allocation is done once per fading block. All the results are averaged over the last 1000 fading blocks. For the sake of illustration we assume that all users are infinitely back-logged and assigned the same utility function of $U_i(W_{i,t}) = (W_{i,t})^{\alpha}/\alpha$, where $W_{i,t}$ is the long term average throughput of user $i$ up to time $t$.[12] All users have the same maximum power constraint of $P_i = 2$W. We solve Problem (UL) once at each scheduling instance. Here, we calculate the achievable rate of user $i$ on subchannel $j$ as

$$r_{ij} = Bx_{ij} \log\left(1 + \frac{p_{ij}e_{ij}}{x_{ij}}\right),$$

where $B$ is the subchannel bandwidth.

In all tables, the "Utility" column denotes $\sum_i U_i(W_{i,t})$ where $t$ is the end of simulation time. The "Log U" column

[12]For simplicity here we do not consider queue length in the utility function. Incorporating queue length will only change the utility gradients at each scheduling epoch. The proposed algorithms can be used without any modification.

TABLE III
ALGORITHM PERFORMANCE FOR SCHEDULING EVERY 20 OFDM SYMBOLS, $\alpha = 0.5$

| Algorithms | | Utility | Log U | Rate | User # | Opt. Ratio |
|---|---|---|---|---|---|---|
| Integer-Dual | | 53922 | 514.0 | 21.56 | 37.5 | 0.9412 |
| SOA 1 | 4A & 5A | 52494 | 510.7 | 22.86 | 34.6 | N/A |
| | 4A & 5B | 51697 | 509.2 | 20.22 | 28.1 | N/A |
| | 4B & 5A | 54165 | 513.3 | 22.25 | 35.0 | N/A |
| | 4B & 5B | 53156 | 511.4 | 21.43 | 28.6 | N/A |
| SOA 2 | | 54316 | 513.6 | 22.33 | 35.1 | N/A |
| Base Line | | 21406 | -1960.5 | 16.13 | 2.66 | N/A |

TABLE IV
ALGORITHM PERFORMANCE WITH SCHEDULING EVERY 80 OFDM SYMBOLS, $\alpha = 0.5$

| Algorithms | | Utility | Log U | Rate | User # | Opt. Ratio |
|---|---|---|---|---|---|---|
| Integer-Dual | | 47843 | 503.58 | 17.26 | 37.1 | 0.9318 |
| SOA 1 | 4A & 5A | 46851 | 500.4 | 16.98 | 34.5 | N/A |
| | 4A & 5B | 42901 | 492.5 | 14.38 | 25.3 | N/A |
| | 4B & 5A | 47716 | 502.0 | 17.62 | 35.1 | N/A |
| | 4B & 5B | 43999 | 494.6 | 15.14 | 26.3 | N/A |
| SOA2 | | 47682 | 502.1 | 17.55 | 35.1 | N/A |
| Base Line | | 15887 | -2116.5 | 11.65 | 2.64 | N/A |

denotes the logarithmic utility function, $\sum_i \log(W_{i,t})$, which provides a characterization of fairness among users. The “Rate” column denotes the total rate achieved by users in Mbps. The “User #” column denotes the average number of users who receive positive rates within one scheduling interval. The “Opt. Ratio” for the Integer-Dual algorithm denotes the ratio between the objective value of Problem UL-Integer achieved with the algorithm and the maximum objective value of Problem UL (i.e., by looking at the optimal dual value) averaged over all scheduling instances. This gives us an idea of how good the performance of the Integer-Dual algorithm is. Note that the optimality ratio is a conservative estimation of the algorithm performance, since the optimal objective value of the relaxed Problem UL typically is strictly larger than the optimal objective function of Problem UL-Integer (i.e., the ratio may never reach 1 even if we solve Problem UL-Integer optimally). Furthermore, we emphasize that this value is based on the average performance of solving problem UL in each scheduling instance; this does not translate directly into a bound on the *average utility* under a optimal scheduling rule and under the given policy as the trajectory of the scheduling weights will be different under the two policies. Since other suboptimal algorithms do not find the optimal dual value of Problem UL at each scheduling instance, the “Optimality Ratio” does not apply to them.

Table III shows results for all the algorithms (summed over all users) when scheduling decisions are made every 20 OFDM symbols (i.e., a fading block of 2msec). The utility parameter is $\alpha = 0.5$. In Table III, SOA1 (with 4B & 5A) and SOA2 achieve the best performance in terms of total utility. Their performance is even better than the Integer-Dual approach, which was obtained based on the optimal value of the relaxed problem. This is likely because only 128 ways to break ties are considered which is typically not sufficient. Since the Integer-Dual algorithm achieves an optimality ratio of 0.9412, this suggests that SOA1 and SOA2 achieve very close to optimal performance as well. The base-line algorithm always has poor performance.

Table IV shows the performance of each algorithm when scheduling is performed every 80 OFDM symbols[13], with all other parameters the same as in Table III. It is clear that this coarser allocation leads to poorer performance compared with Table III. This shows the trade-off between system performance and resource allocation frequency (and thus algorithm complexity).

TABLE V
ALGORITHM PERFORMANCE FOR SCHEDULING EVERY 20 OFDM SYMBOLS, $\alpha = 0$

| Algorithm | | Utility | Log U | Rate | User # | Opt. Ratio |
|---|---|---|---|---|---|---|
| Integer-Dual | | 515 | 515 | 19.80 | 39.0 | 0.9715 |
| SOA 1 | 4A & 5A | 511 | 511 | 18.47 | 36.9 | N/A |
| | 4A & 5B | 509 | 509 | 21.56 | 28.6 | N/A |
| | 4B & 5A | 515 | 515 | 20.04 | 37.4 | N/A |
| | 4B & 5B | 512 | 512 | 17.80 | 29.3 | N/A |
| SOA 2 | | 515 | 515 | 19.91 | 37.4 | N/A |
| Base Line | | -1961 | -1961 | 16.13 | 2.66 | N/A |

Tables V and VI[14] give the results of the algorithms when the utility parameter $\alpha$ is equal to 0 (proportional fair allocation) and 1 (maximum rate allocation), respectively. In each case scheduling decisions are made every 20 OFDM symbols, It is clear that there is a trade-off between fairness (measured by Log U) and efficiency (measured by total rate). The value of $\alpha = 0$ gives a throughput allocation that is the fairest and the least efficient, while $\alpha = 1$ is the most efficient and least fair. Once again we note that all of the heuristics have good performance with SOA1 (with 4B & 5A) and SOA2 achieving the best performances in terms of total utility.

In each case simulated, all of the SOA’s have good performance with SOA1 (with 4A & 5B) and SOA2 consistently achieving the best performance in terms of total utility. From

TABLE VI
ALGORITHM PERFORMANCE FOR SCHEDULING EVERY 20 OFDM SYMBOLS, $\alpha = 1$

| Algorithm | | Utility | Log U | Rate | User # | Opt. Ratio |
|---|---|---|---|---|---|---|
| Integer-Dual | | 23.24e6 | 472.56 | 23.37 | 20.43 | 0.82541 |
| SOA 1 | 4A & 5A | 23.19e6 | 448.99 | 22.28 | 22.6 | N/A |
| | 4A & 5B | 23.11e6 | -136.03 | 23.20 | 15.6 | N/A |
| | 4B & 5A | 24.31e6 | 444.02 | 24.42 | 32.7 | N/A |
| | 4B & 5B | 23.95e6 | -195.10 | 24.05 | 15.6 | N/A |
| SOA 2 | | 24.46e6 | 372.95 | 24.57 | 21.8 | N/A |
| Base Line | | 16.08e6 | -1961 | 16.13 | 2.66 | N/A |

[13]Here the scheduling and resource decisions on made based on the first 20 OFDM symbols.

[14]In the “Utility” column of Table VI, $xe6$ means $x \times 10^6$.

the analysis in Section V-D, we note that these have slightly higher complexity than some of the other SOA's. Hence if lower complexity is desired, this can be provided with only a slight loss in performance. We also note that in each case the SOAs and the integer-dual algorithm schedule a large number of users on average in each time-slot. A potential cost from this is that it may increase the needed signaling overhead. One way to reduce this cost is to add a penalty term to our objective which increases with the number of users scheduled.

## VII. Conclusions

We presented an optimization-based formulation for scheduling and resource allocation in the uplink OFDM access network. Compared to the downlink, we showed that the uplink was computationally more challenging due in part 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 for a range of different user utilities and scheduling time-scales. Two algorithms from this family consistently achieved the best performance, but had a slightly higher complexity than some of the other algorithms, enabling complexity to be traded off with performance within this family of algorithms.

## References


[1] R. Agrawal, A. Bedekar, R. La, V. Subramanian, "A Class and Channel-Condition based Weighted Proportionally Fair Scheduler," in *Proc. ITC*, Salvador, Brazil, Dec. 2001.

[2] R. Agrawal and V. Subramanian, "Optimality of Certain Channel Aware Scheduling Policies," in *Proc. of Allerton Conference*, Oct. 2002.

[3] R. Agrawal, V. Subramanian and R. Berry, "Joint Scheduling and Resource Allocation in CDMA Systems," in *Proc. of WiOpt '04*, Cambridge, UK, March 2004.

[4] J. Huang, V. G. Subramanian, R. Agrawal and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. on Wireless*, to appear.

[5] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queue with randomly varying connectivity", *IEEE Trans. IT*, vol. 39, pp. 466-478, March 1993.

[6] R. Leelahakriengkrai and R. Agrawal, "Scheduling in Multimedia Wireless Networks," in *Proc. ITC*, Salvador, Brazil, pp. 556–564, Dec. 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 Informational Sciences*, vol. 18, No. 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. IEEE VTC*, Spring 2000, vol. 3, pp. 1854–1858.

[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. 12–25, 2005.

[11] M. Andrews, L. Qian, and A. Stolyar, "Optimal utility based multi-user throughput allocation subject to throughput constraints," in *Proc. IEEE INFOCOM*, March 2005, vol. 4, pp. 2415–2424.

[12] H. Kushner and P. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," in *Proc. Allerton Conference*, Oct. 2002.

[13] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Networking*, vol. 8, Issue 5, pp. 556–567, Oct. 2000.

[14] "IEEE 802.16e-2005 and IEEE Std 802.16-2004/Cor1-2005," 2005, Available: http://www.ieee802.org/16/.

[15] E. D. Nering and A. W. Tucker, *Linear Programs and Related Problems*. Academic Press Inc., 1993.

[16] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," in *Proc. IEEE Globecom*, vol. 1, pp. 103–107, Dec. 2000.

[17] T. M. Cover and J. A. Thomas, "Elements of information theory," Second edition, Wiley-Interscience, Hoboken, NJ, 2006.

[18] S. Shenker, "Fundamental design issues for the future Internet," *IEEE JSAC*, vol. 13, no. 7, pp. 1176–1188, 1995.

[19] FLASH-OFDM Technology, http://www.qualcomm.com/technology/flash-ofdm

[20] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.

[21] Z. Han, Z. Ji, and K. Liu, "Fair Multiuser Channel Allocation for OFDMA Networks Using Nash Bargaining Solutions and Coalitions," *IEEE Trans. Comm.*, vol. 53, no. 8, pp. 1366–1376, 2005.

[22] S. Pfletschinger, G. Muenz, and J. Speidel, "Efficient subcarrier allocation for multiple access in OFDM systems," in *Proc. International OFDM-Workshop*, 2002.

[23] Y. Ma, "Constrained Rate-Maximization Scheduling for Uplink OFDMA," in *Proc. IEEE MILCOM*, pp. 1–7, Oct. 2007.

[24] B. Da and C. Ko, "Dynamic subcarrier sharing algorithms for uplink OFDMA resource allocation," in *Proc. 6th ICICS*, pp. 1–5, 2007.

[25] K. Kim, Y. Han, and S. Kim, "Joint subcarrier and power allocation in uplink OFDMA systems," *IEEE Comms. Letters*, vol. 9, no. 6, pp. 526–528, 2005.

[26] C. Ng and C. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Trans. Wireless Comm.*, vol. 7, no. 5 Part 1, pp. 1667–1675, 2008.

[27] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Trans. Wireless Comm.*, vol. 7, no. 5 Part 1, pp. 1507–1511, 2008.

[28] K. Kwon, Y. Han, and S. Kim, "Efficient Subcarrier and Power Allocation Algorithm in OFDMA Uplink System," *IEICE Trans. Comm.*, vol. 90, no. 2, pp. 368–371, 2007.

[29] "3GPP Long Term Evolution," 2008, Available: http://www.3gpp.org/Highlights/LTE/LTE.htm .

[30] "3GPP2 Ultra Mobile Broadband," 2007, Available: http://www.3gpp2.org/ .

[31] W. Yu, R. Lui, and R. Cendrillon, "Dual optimization methods for multiuser orthogonal frequency division multiplex systems," in *Proc. IEEE Globecom*, vol. 1, 2004, pp. 225–229.


**Jianwei Huang** (S'01-M'06) is an Assistant Professor in Information Engineering Department at the Chinese University of Hong Kong. He received the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Northwestern University in 2003 and 2005, respectively. From 2005 to 2007, He worked as a Postdoctoral Research Associate in the Department of Electrical Engineering at Princeton University. In 2004 and 2005, he worked in the Mathematics of Communication Networks Group at Motorola (Arlington Heights, IL, USA) both as a full time summer intern and a part time researcher. His main research interests lie in the area of modeling and performance analysis of communication networks, including cognitive radio networks, OFDM and CDMA systems, wireless medium access control, multimedia communications, network economics, and applications of optimization theory and game theory.

Dr. Huang is an Associate Editor of Journal of Computer & Electrical Engineering, the Lead Guest Editor of the IEEE Journal of Selected Areas in Communications special issue on "Game Theory in Communication Systems", the Lead Guest Editor of the Journal of Advances in Multimedia special issue on "Collaboration and Optimization in Multimedia Communications", and a Guest Editor of the Journal of Advances in Multimedia special issue on "Cross-layer Optimized Wireless Multimedia Communications", the Technical Program Committee Co-Chair of the International Conference on Game Theory for Networks (GameNets'09). Dr. Huang is the recipient of a 2001 Walter P. Murphy Fellowship at Northwestern University and a 1999 Chinese National Excellent Student Award.

**Vijay G. Subramanian** (S'???-M'??) received the B.Tech. degree from the Depart- ment of Electrical Engineering, Indian Institute of Technology, Madras, in 1993, the M.S. degree in electrical engineering from the Indian Institute of Science, Bangalore, in 1995, and the Ph.D. degree in electrical engineering from the Uni- versity of Illinois at Urbana-Champaign, Urbana, in 1999. From 1999 to 2006, he was with the Networks Business, Motorola, Arlington Heights, IL. Since May 2006 he is a Research Fellow at the Hamilton Institute, NUIM, Ireland. His research interests include information theory, communications, communication networks, wireless networks, queueing theory, and applied probability and stochastic processes.

**Rajeev Agrawal** (S'???-M'??) is a Fellow of the Technical Staff at Motorola where his responsibilities include the architecture, design and optimization of Motorola's next generation wireless systems.

Prior to joining Motorola in 1999, Rajeev was Professor of Electrical and Computer Engineering and Computer Science departments at the University of Wisconsin - Madison. He also spent a sabbatical year at IBM TJ Watson Research, British Telecom Labs, and INRIA-Sophia Antipolis.

Rajeev received his M.S. (1987) and Ph.D. (1988) degrees in electrical engineering-systems from the University of Michigan, Ann Arbor and his B.Tech. (1985) degree in electrical engineering from the Indian Institute of Technology, Kanpur.

**Randall A. Berry** (S'????-M'???) received the B.S. degree in Electrical Engineering from the University of Missouri-Rolla in 1993 and the M.S. and PhD degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 1996 and 2000, respectively. In September 2000, he joined the faculty of Northwestern University, where he is currently an Associate Professor in the Department of Electrical Engineering and Computer Science. In 1998 he was on the technical staff at MIT Lincoln Laboratory in the Advanced Networks Group, where he worked on optical network protocols. His current research interests include wireless communication, data networks and information theory.

Dr. Berry is the recipient of a 2003 NSF CAREER award. He is currently serving on the editorial boards of the *IEEE Transactions on Information Theory* and the *IEEE Transactions on Wireless Communications*.