Downlink Scheduling and Resource Allocation for OFDM Systems

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

***Abstract*—We consider scheduling and resource allocation for the downlink of a cellular OFDM system, with various practical considerations including integer carrier allocations, different subchannelization schemes, a maximum SNR constraint per tone, and "self-noise" due to channel estimation errors and phase noise. During each time-slot a subset of users must be scheduled for transmission, and the available tones and transmission power must be allocated among the selected users. Employing a gradient-based scheduling scheme presented in earlier papers reduces this to an optimization problem to be solved in each time-slot. Using dual decomposition techniques, we give an optimal algorithm for this problem when multiple users can time-share each carrier. We then give several low complexity heuristics that enforce an integer constraint on the carrier allocation. Simulations show that the algorithms presented all achieve similar performance under a wide range of scenarios, and that the performance gap between the optimal and suboptimal algorithms widens when per user SNR constraints or channel estimation errors are considered.**

I. Introduction

Channel-aware scheduling and resource allocation is essential in high-speed wireless data systems. In these systems, the users scheduled and the allocation of physical layer resources among them are dynamically adapted based on the users' channel conditions and quality of service (QoS) requirements. Many of the scheduling algorithms considered can be viewed as "gradient-based" algorithms, which select the transmission rate vector that maximizes the projection onto the gradient of the system's total utility [1]–[4]. Several such algorithms have been studied for time-division multiplexed (TDM) systems, such as the "proportionally fair rule" [4], [5] first proposed for CDMA 1xEVDO and is based on a logarithmic utility function of each user's throughput. A larger class of throughput-based utilities is considered in [2] where efficiency and fairness are allowed to be traded-off. The "Max Weight" policy (e.g. [7]–[9]) can also be viewed as a gradient-based policy, where the utility is now a function of a user's queue-size or delay.

J. Huang is with the Dept. 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. This work has been supported by the Competitive Earmarked Research Grants (Project Number 412308) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, the Direct Grant (Project Number C001-2050398) of The Chinese University of Hong Kong, SFI grant IN3/03/I346, the Motorola-Northwestern Center for Seamless Communications, and NSF CAREER award CCR-0238382. The work was partially presented at CISS 2006 and the 2006 Asilomar Conference on Signals, Systems and Computers.

In TDM systems, the scheduling and resource allocation decision is simple: schedule one user in a time-slot and choose the modulation and coding scheme for that user. In many current systems, multiple users may be multiplexed within a time-slot. Orthogonal Frequency Division Multiplexing (OFDM) is a common option for broadband wireless networks (e.g. IEEE 802.16/WiMAX [11] and 3GPP LTE [12]). This paper addresses gradient-based scheduling and resource allocation for the downlink in a cellular OFDM system. In this setting, in addition to determining which users are scheduled, the allocation of physical layer resources (e.g. transmission power and subcarriers) must be specified.

In prior work [10], we considered a related problem when code division multiple access (CDMA) is used to multiplex users within a time-slot, as in CDMA 1xEVDV and HSDPA. In [10], the physical layer resources are the number of spreading codes assigned to each user and the transmission power; allocating these according to a gradient-based policy requires maximizing the weighted sum rate in each time-slot, where the weights vary dynamically based on the gradient of the system utility. When the rate per code is given by the Shannon capacity formula, this maximization becomes a tractable convex optimization problem, enabling the development of low complexity near-optimal algorithms. Here, we follow a similar approach for an OFDM-based system. Compared to [10], the key difference is that we have more degrees of freedom (i.e., the tones) to allocate resources across. This enables exploiting both multi-user diversity and frequency diversity at a finer granularity, but also significantly increases the complexity of the optimization. Furthermore, we include the following considerations that are important in practical OFDM systems: 1) integer constraints on the tone allocation, i.e., a tone can be allocated to at most one user; 2) different subchannelization techniques in which resource allocation is performed at a larger granularity (i.e, groups of tones or symbols) in order to reduce the channel measurement and feedback overhead; 3) constraints on the maximum SNR (i.e., rate) per tone, which models a limitation on the available modulation and coding schemes; and 4) "self-noise" on tones due to channel estimation errors (e.g., [13]) or phase noise [24].

At the beginning of each scheduling interval, the gradient-based scheduling algorithm maximizes the weighted sum throughput over the current set of feasible rates. In Section II, we describe our model for this rate region, taking into account the preceding considerations. In Section III, we consider a dual formulation for the resulting optimization problem, which enables us to exploit the problem's structure and develop both optimal and simple sub-optimal algorithms with low complexity. In Section IV, we present simulation results when the scheduling weights are dynamically adjusted according to a gradient-based scheduling rule. We study the performance of the optimal and suboptimal algorithms under different choices of utility functions, subchannelization schemes, channel estimation errors and phase noise (self-noise), and SNR constraints. We conclude in Section V.

In terms of related work, a number of formulations for downlink OFDM resource allocation have been studied including [14]–[21]. In [15], [16], the goal is to minimize the total transmit power given target bit-rates for each user. In [16], the target bit-rates are determined by a fair queueing algorithm, which does not take into account the users' channel conditions. In [18]–[20], the focus is on maximizing the sum-rate given a minimum bit-rate per user; [17] also considers maximizing the sum-rate, but without any minimum bit-rate target. A special case of the problem we study that assumes a fixed set of weights, no constraints on the SNR per carrier, and no self-noise was considered in [14], [21]. In [14], a suboptimal algorithm with constant power per tone was shown in simulations to have little performance loss. Other heuristics that use a constant power per tone are given in [17]–[19]. We also consider such a heuristic in Section III-D. In [21], a similar dual-based algorithm to ours is considered and simulations are given which show that the duality gap of this problem quickly goes to zero as the number of tones increases. We will revisit the conclusions of [21] in Section III-B. Finally, in [22], the capacity region of a downlink broadcast channel with frequency-selective fading using a TDM scheme is given; the feasible rate region we consider, without any maximum SNR constraints, can be viewed as a special case of this region. None of these papers consider self-noise or per user SNR constraints, which we do here. Moreover, most of these papers optimize a static objective function, while we are interested in a dynamic setting where the objective changes over time according to a gradient-based algorithm. It is not *a priori* clear if a good heuristic for a static problem applied to each time-step, will be a good heuristic for the dynamic case since the optimality result in [1]–[4], [7]–[9] is predicated on solving the weighted-rate optimization problem exactly in each time-slot. Our simulation results show that the good performance of heuristics holds, at least for the utility maximization models considered in this paper. Finally, we note that in a companion paper [26] we use similar methods to solve a centralized uplink scheduling and resource allocation problem, the 'dual' of the downlink problem in a 802.16/WiMax setting.

II. Problem Formulation

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

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \nabla U(\boldsymbol{W}_t)^T \cdot \boldsymbol{r}_t = \max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i U_i'(W_{i,t}) r_{i,t}, \quad (1)$$

where $U_i'(\cdot)$ is the derivative of $U_i(\cdot)$. For example, one class of utility functions given in [2], [6] is

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

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

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

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

In general, we consider the problem of

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

where $w_{i,t} \ge 0$ is a time-varying weight assigned to the ith user at time t. In the above example these weights are given by the gradient of the utility; however, other methods for generating these weights (possibly depending upon queue-lengths and/or delays [7]–[9]) are also possible. We note that (4) must be re-solved at each scheduling instance because of changes in both the channel state and the weights (e.g., the gradients of the utilities). While the former changes are due to the time-varying nature of wireless channels, the latter changes are due to new arrivals and past service decisions.

A. *OFDM capacity regions*

The solution to (4) depends on the channel state dependent rate region $\mathcal{R}(\boldsymbol{e})$, where for simplicity we suppress the dependence on time. We consider a model appropriate for downlink OFDM systems; related models have been considered in [14], [22]. In this model, $\mathcal{R}(\boldsymbol{e})$ is parameterized by the allocation of tones to users and the allocation of power across tones. In a traditional OFDM system, at most one user may be assigned to any tone. Initially, as in [15], [16], we make the simplifying assumption that multiple users can share one tone using some orthogonalization technique (e.g. TDM).[1] In practice, if a scheduling interval contained multiple OFDM symbols, we can implement such sharing by giving a fraction of the symbols to each user; of course, each user will be constrained to use an integer number of symbols and the

[1] We focus on systems that do not use superposition coding and successive interference cancellation within a tone, as such techniques are generally considered too complex for practical systems.

required signaling overhead will increase.[2] We discuss the case where only one user can use a tone in Section III-C.

Let $\mathcal{N} = \{1, \ldots, N\}$ denote the set of tones. For each $j \in \mathcal{N}$ and user $i \in \mathcal{K}$, let e_{ij} be the received signal-to-noise ratio (SNR) per unit power. We denote the power allocated to user i on tone j by p_{ij} and the fraction of that tone allocated to user i by x_{ij}. The total power allocation must satisfy $\sum_{i,j} p_{ij} \leq P$, and the total allocation for each tone j must satisfy $\sum_i x_{ij} \leq 1$. For a given allocation, with perfect channel estimation, user i's feasible rate on tone j is $r_{ij} = x_{ij} B \log(1 + \frac{p_{ij} e_{ij}}{x_{ij}})$, which corresponds to the Shannon capacity of a Gaussian noise channel with bandwidth $x_{ij}B$ and received SNR $p_{ij}e_{ij}/x_{ij}$.[3] This SNR arises from viewing p_{ij} as the energy per time-slot user i uses on tone j; the corresponding transmission power becomes p_{ij}/x_{ij} when only a fraction x_{ij} of the tone is allocated. Without loss of generality we set $B = 1$ in the following.

In a realistic OFDM system, imperfect carrier synchronization and channel estimation may result in "self-noise" (e.g. [13], [24]). We follow a similar approach as in [13] to model self-noise. Let the received signal on the jth tone of user i be given by $y_{ij} = h_{ij}s_{ij} + n_{ij}$, where h_{ij}, s_{ij} and n_{ij} are the (complex) channel gain, transmitted signal and additive noise, respectively, with $n_{ij} \sim \mathcal{CN}(0, \sigma^2)$.[4] Assume that $h_{ij} = \tilde{h}_{ij} + h_{ij,\delta}$, where $\tilde{h}_{ij}$ is receiver i's estimate of h_{ij} and $h_{ij,\delta} \sim \mathcal{CN}(0, \delta_{ij}^2)$. After matched-filtering, the received signal will be $z_{ij} = \tilde{h}_{ij}^* y_{ij}$ resulting in an effective SNR of

$$\text{Eff-SNR} = \frac{\|\tilde{h}_{ij}\|^4 p_{ij}}{\sigma_{ij}^2 \|\tilde{h}_{ij}\|^2 + \delta_{ij}^2 p_{ij} \|\tilde{h}_{ij}\|^2} = \frac{p_{ij}\tilde{e}_{ij}}{1 + \beta_{ij} p_{ij} \tilde{e}_{ij}}, \quad (5)$$

where $p_{ij} = \mathrm{E}(\|s_{ij}\|^2)$, $\beta_{ij} = \frac{\delta_{ij}^2}{\|\tilde{h}_{ij}\|^2}$ and $\tilde{e}_{ij} = \frac{\|\tilde{h}_{ij}\|^2}{\sigma_{ij}^2}$.[5] Here, $\beta_{ij} p_{ij} \tilde{e}_{ij}$ is the self-noise term. As in the case without self-noise ($\beta_{ij} = 0$), the effective SNR is still increasing in p_{ij}. However, it now has a maximum of $1/\beta_{ij}$. For the sake of presentation, we assume that $\beta = \beta_{ij}$ for all i and j.[6] The analysis is almost identical if users have different β_{ij}'s.

With self-noise, user i's feasible rate on tone j becomes $r_{ij} = x_{ij} \log(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij} + \beta p_{ij} \tilde{e}_{ij}})$, where again x_{ij} models time-sharing of a tone. Under these assumptions, we have

$$\mathcal{R}(\boldsymbol{e}) = \Bigg\{ \boldsymbol{r} : r_i = \sum_j x_{ij} \log\left(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij} + \beta p_{ij}\tilde{e}_{ij}}\right), \sum_{i,j} p_{ij} \leq P, \sum_i x_{ij} \leq 1 \; \forall j, (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \Bigg\}, \quad (6)$$

where $\mathcal{X} := \prod_{j=1}^N \mathcal{X}_j$, and for all $j \in \mathcal{N}$,

$$\mathcal{X}_j := \left\{ (\boldsymbol{x}^j, \boldsymbol{p}^j) \geq \boldsymbol{0} : x_{ij} \leq 1, p_{ij} \leq \frac{x_{ij}\tilde{s}_{ij}}{\tilde{e}_{ij}} \; \forall i \right\}, \quad (7)$$

with $\boldsymbol{x}^j := (x_{ij}, \forall i \in \mathcal{K})$ and $\boldsymbol{p}^j := (p_{ij}, \forall i \in \mathcal{K})$. Here, $\tilde{s}_{ij} = \frac{\Gamma_{ij}}{1 - \Gamma_{ij}\beta}$, where $\Gamma_{ij} < 1/\beta$ is a maximum SNR constraint on tone j for user i, e.g., to model a constraint on the maximum rate per tone due to a limitation on the available modulation and coding schemes. At the cost of additional complexity, we could also include minimum rate constraints to model inelastic traffic, and maximum rate constraints to incorporate buffer sizes. While the former is a convex constraint, the latter is not (but it can still be proved that there is no duality gap). Similar techniques can be applied in these cases; however, in the interests of brevity we will not discuss the details here.

We assume that $\tilde{e}_{ij}$ is known by the scheduler for all i and j as is β (equivalently, the estimation error variance). In a frequency division duplex (FDD) system, this knowledge can be acquired by having the base station transmit pilot signals, from which the users can estimate their channel gains and feedback to the base station. In a time division duplex (TDD) system, these gains can also be acquired by having the users transmit uplink pilots; the base station can then exploit reciprocity to measure the channel gains. In both cases, this feedback information would need to be provided within the channel's coherence time.

B. Subchannelization

With many tones and users, providing pilots and/or feedback per tone can require excessive overhead; e.g., in IEEE 802.16e [11], a channel with bandwidth 1.25Mhz to 20Mhz is divided from 128 to 2048 tones. One way to reduce this overhead is to form *subchannels* from disjoint sets of tones. Feedback and resource allocation are then done at the granularity of subchannels, i.e., constant power is used and coding is done across the tones in each subchannel. Our model can be adapted to this setting by viewing $\mathcal{N}$ as the set of subchannels and $\tilde{e}_{ij}$ as the effective SNR per unit power for user i on the jth subchannel. Specifically, assuming that k tones are bundled into subchannel j, $\tilde{e}_{ij}$ is chosen so that the total rate for user i in this subchannel is approximately $kx_{ij} \log(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij} + \beta p_{ij}\tilde{e}_{ij}})$. Since $\log(1 + \frac{pe}{x + \beta pe})$ is a concave function of e, it follows from Jensen's inequality that the rate achieved over a sub-channel is upper bounded by setting $\tilde{e}_{ij}$ equal to the *arithmetic average* of the channel gains of tones belonging to subchannel j. We can also lower bound the rate using the strict convexity of $\log(l + \exp(y))$ for $y \in \Re$ (with $l > 0$) and Jensen's inequality. If $\beta = 0$, taking $y = \log\left(\frac{pe}{x}\right)$ and $l = 1$ we obtain a lower bound on the rate by setting $\tilde{e}_{ij}$ equal to the *geometric average* of the subchannel gains. When

[2] With a large number of tones, adjacent tones will have nearly identical gains, in which case this time-sharing can also be approximated by frequency sharing. As the number of tones increases, this approximation once again becomes tight.

[3] To better model the achievable rates in a practical system we can re-normalize e_{ij} by γe_{ij}, where $\gamma \in [0, 1]$ represents the system's "gap" from capacity. The simulations results in Section IV take an appropriate γ into consideration.

[4] We use the notation $x \sim \mathcal{CN}(0, b)$ to denote that x is a 0 mean, complex, circularly-symmetric Gaussian random variable with variance $b := \mathrm{E}(\|x\|^2)$.

[5] This is slightly different from the Eff-SNR in [13] in which the signal power is instead given by $\|h_{ij}\|^4 p_{ij}$; the following analysis works for such a model as well by a simple change of variables. For the problem at hand, (5) seems more reasonable in that the resource allocation will depend only on $\tilde{h}_{ij}$ and not on h_{ij}. We also note that (5) is shown in [23] to give an achievable lower bound on the capacity of this channel.

[6] Also, to simplify our presentation, we assume that β does not depend on the channel quality. If the main cause of self-noise is channel estimation error, this may not be the case. We discuss this more in Section. III.A.

$\beta > 0$ we take $y = -\log\left(1 + \frac{x}{\beta pe}\right)$ and $l = \beta$, apply Jensen's inequality followed by the arithmetic-mean geometric-mean inequality to obtain a lower bound on the rate by setting $\tilde{e}_{ij}$ equal to the *harmonic average* of the subchannel gains. The gap between the upper and lower bounds is quite small for reasonable values of pe. Indeed, for the range of SNRs achieved by scheduled users in our simulations, we do not see much difference.[7] From this point onwards we will use the terms tone/carrier/subchannel to mean the basic allocation unit; the specific distinctions will be clear from the context.

We consider the following three approaches for forming subchannels: (1) adjacent channelization, where adjacent tones are grouped into a subchannel; (2) interleaved channelization, where tones are (perfectly) interleaved to form subchannels; and (3) random channelization, where tones are randomly assigned to subchannels. In IEEE 802.16d/e [11], interleaved channelization is primarily used; the optional "band AMC mode" allows for adjacent channelization. Randomized channelization can model systems that employ frequency hopping as in the Flash OFDM system [25]. Using adjacent channelization enables the resource allocation to better exploit frequency diversity. Interleaved or random channelization reduces the variance of the effective SNR across subchannels for each user. One advantage to this is that when the variance is small, user i can simply feed back a single e_i value that will be representative of all its subchannels, further reducing overhead. Another advantage of random channelization is in managing inter-cell interference.

III. Optimal and Suboptimal Algorithms

From (4) and (6), the scheduling and resource allocation problem can be stated as:

$$\max_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}} V(\boldsymbol{x},\boldsymbol{p}) := \sum_i w_i \sum_j x_{ij} \log\left(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij}+\beta p_{ij}\tilde{e}_{ij}}\right)$$
$$\text{subject to:} \quad \sum_{i,j} p_{ij} \le P, \text{ and } \sum_i x_{ij} \le 1, \forall j \in \mathcal{N}. \tag{8}$$

Here, we are still assuming that multiple users can time-share each subchannel (i.e. no integer constraints on x_{ij}). We next give an algorithm to solve this problem via a dual decomposition. The resulting algorithm leads to a solution that has a computational complexity of $O(NK)$. We then consider the problem with integer constraints and propose three suboptimal algorithms, each also with complexity $O(NK)$.

A. Optimal Dual Solution

Consider the Lagrangian,

$$L(\boldsymbol{x},\boldsymbol{p},\lambda,\boldsymbol{\mu}) := \lambda P + \sum_{j=1}^{N} L_j(\boldsymbol{x}^j,\boldsymbol{p}^j,\lambda,\mu_j), \tag{9}$$

where

$$L_j(\boldsymbol{x}^j,\boldsymbol{p}^j,\lambda,\mu_j) := \mu_j + \sum_{i=1}^{K} w_i x_{ij} \log\left(1 + \frac{p_{ij}\tilde{e}_{ij}}{x_{ij}+\beta p_{ij}\tilde{e}_{ij}}\right) - \mu_j \sum_{i=1}^{K} x_{ij} - \lambda \sum_{i=1}^{K} p_{ij}, \tag{10}$$

and $\boldsymbol{\mu} = (\mu_j)_{j=1}^{N}$. The corresponding dual function is then

$$L(\lambda,\boldsymbol{\mu}) := \max_{(\boldsymbol{p},\boldsymbol{x})\in\mathcal{X}} L(\boldsymbol{x},\boldsymbol{p},\lambda,\boldsymbol{\mu}) = \lambda P + \sum_{j=1}^{N} \max_{(\boldsymbol{p}^j,\boldsymbol{x}^j)\in\mathcal{X}_j} L_j(\boldsymbol{x}^j,\boldsymbol{p}^j,\lambda,\mu_j).$$

By directly evaluating the Hessian of $x\log(1+\frac{p}{x+\beta p})$ it can be seen that this is jointly concave in (x,p). It then follows that Problem (8) is convex and satisfies Slater's condition. Hence, there is no duality gap and so $V^* := \min_{\lambda\ge 0,\boldsymbol{\mu}\ge \boldsymbol{0}} L(\lambda,\boldsymbol{\mu})$ is the optimal objective value [27].

Next we show that the maximization over $\boldsymbol{p}$ and $\boldsymbol{x}$ leads to a closed-form representation of $L(\lambda,\boldsymbol{\mu})$. We then show that minimizing $L(\lambda,\boldsymbol{\mu})$ over $\boldsymbol{\mu}$ only requires searching for the maximum of user dependent metrics for each tone j. The only numerical search needed is for the minimization over λ, which is just a one-dimensional search with low complexity.

1) Computing the Dual Function: First, we maximize $L_j(\boldsymbol{x}^j,\boldsymbol{p}^j,\lambda,\mu_j)$ over $\boldsymbol{p}^j$ given $\boldsymbol{x}^j$, μ_j and λ. The optimal solution is given by

$$p^*_{ij}(\boldsymbol{x},\lambda,\boldsymbol{\mu}) = \frac{x_{ij}}{\tilde{e}_{ij}}\left[q\left(\beta,\left(\frac{w_i\tilde{e}_{ij}}{\lambda}-1\right)^+\right) \wedge \tilde{s}_{ij}\right], \tag{11}$$

where $(x)^+ = \max(x,0)$, $a \wedge b = \min(a,b)$, and

$$q(\beta,z) := \begin{cases} z, & \text{if } \beta = 0,\\ \left(\frac{2\beta+1}{2\beta(\beta+1)}\right)\left(\sqrt{1+\frac{4\beta(\beta+1)}{(2\beta+1)^2}z}-1\right), & \text{if } \beta > 0. \end{cases}$$

Figure 1 shows p^*_{ij} in (11) as a function of $\tilde{e}_{ij}$ for three different values of $\beta = 0, 0.01, 0.1$. When $\beta = 0$, (11) becomes a "water-filling" type of solution in which $p^*_{ij}(\boldsymbol{x},\lambda,\boldsymbol{\mu})$ is non-decreasing in $\tilde{e}_{ij}$. For a fixed $\beta > 0$, this is not necessarily true, i.e., due to self-noise, less power may be allocated to "better" subchannels. The constant β case is applicable when the self-noise is due to phase noise as in [24]. On the other hand, when self-noise arises primarily from estimation errors, β may not be constant but could depend on the channel quality. The exact dependence will depend on the details of channel estimation. As an example, using the analysis in [23, Section IV] for the estimation error of a Gauss-Markov channel from a pilot with known power, we consider the cases when the pilot power is either constant or inversely proportional to channel quality subject to maximum and minimum power constraints (modeling power control). In both cases β is inversely proportional to channel condition for large e; e.g., $\beta(e) = 10/e$ in Fig. 1. It can be seen that the curve has a different shape and amplitude compared with the $\beta = 0$ case. For simplicity of presentation, we assume constant β's in the remainder of the paper.

[7] For example, our simulations show that for the optimal algorithm with $\beta = 0.01$, the differences between long term achievable utilities under arithmetic average and harmonic average approximations are 0.005%, 0.1%, and 0.4% under adjacent, interleaved and random subchannelizations, respectively.

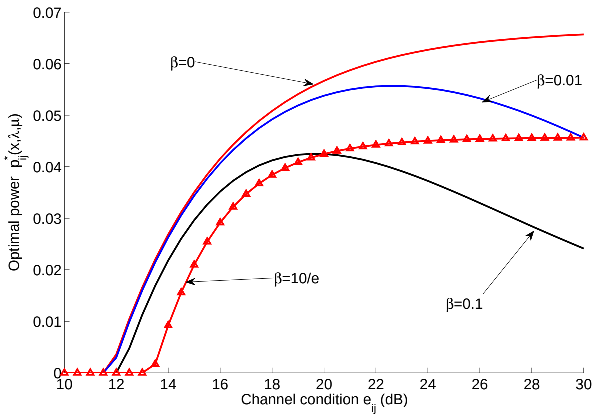

Fig. 1. Optimal power $p_{ij}^*(\boldsymbol{x}, \lambda, \boldsymbol{\mu})$ as a function of the channel condition e_{ij}. Here $x_{ij} = 1$, $w_i = 1$, and $\lambda = 15$.

Notice from (11) that the optimal value of p_{ij}^* is always a linear function of x_{ij}. Interestingly, substituting (11) into $L_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j)$ also results in a linear function of x_{ij}. Namely,

$$L_j(\boldsymbol{x}^j, \boldsymbol{p}^{j,*}, \lambda, \mu_j) = \sum_i x_{ij} \left(\mu_{ij}(\lambda) - \mu_j\right) + \mu_j,$$

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

$$h\left(\beta, \omega, \tilde{s}_{ij}\right) := \log\left(1 + \frac{q\left(\beta, (\omega-1)^+\right) \wedge \tilde{s}_{ij}}{1 + \beta\left(q\left(\beta, (\omega-1)^+\right) \wedge \tilde{s}_{ij}\right)}\right) - \frac{1}{\omega}\left(q\left(\beta, (\omega-1)^+\right) \wedge \tilde{s}_{ij}\right).$$

From this it follows that any choice

$$x_{ij}^*(\lambda, \boldsymbol{\mu}) \in \begin{cases} \{1\}, & \text{if } \mu_{ij}(\lambda) > \mu_j, \\ [0,1], & \text{if } \mu_{ij}(\lambda) = \mu_j, \\ \{0\}, & \text{if } \mu_{ij}(\lambda) < \mu_j \end{cases} \tag{12}$$

will maximize $L_j(\boldsymbol{x}^j, \boldsymbol{p}^{j,*}, \lambda, \mu_j)$. Hence, $L(\lambda, \boldsymbol{\mu}) := \lambda P + \sum_{j=1}^N L_j(\lambda, \mu_j)$, where

$$L_j(\lambda, \mu_j) := L_j(\boldsymbol{x}^{j,*}, \boldsymbol{p}^{j,*}, \lambda, \mu_j) = \sum_i \left(\mu_{ij}(\lambda) - \mu_j\right)^+ + \mu_j. \tag{13}$$

2) Optimizing the Dual Function over λ and $\boldsymbol{\mu}$: First we consider the optimization over $\boldsymbol{\mu}$.

Lemma 1: For all $\lambda \geq 0$,

$$L(\lambda) := \min_{\boldsymbol{\mu} \geq \boldsymbol{0}} L(\lambda, \boldsymbol{\mu}) = \lambda P + \sum_j \mu_j^*(\lambda), \tag{14}$$

where for every tone j, the minimizing value of μ_j^* is achieved by

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

The proof of Lemma 1 follows from a similar argument as in [10]. Note that (15) requires searching for the maximum value of the metrics μ_{ij} across all users for each tone j.

Since $L(\lambda)$ is the minimum of a convex function over a convex set, it is a convex function of λ; hence, it can be minimized using an iterated one dimensional search (e.g., the Golden Section method). Since there is no duality gap, at $\lambda^* = \arg\min_{\lambda \geq 0} L(\lambda)$, $L(\lambda^*)$ gives the optimal objective value of Problem (8).

B. Optimal primal variables with time-sharing

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

$$(\boldsymbol{x}^*(\lambda), \boldsymbol{p}^*(\lambda)) := \arg \max_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X}} L\left(\boldsymbol{x}, \boldsymbol{p}, \lambda, \boldsymbol{\mu}^*(\lambda)\right), \tag{16}$$

where $\boldsymbol{\mu}^*(\lambda)$ is given in (15). Problem (16) can be solved using the same procedure as in computing the dual function (i.e., (11) and (12)). Given that $\lambda = \lambda^*$, it follows from duality theory, that if the corresponding $(\boldsymbol{x}^*(\lambda^*), \boldsymbol{p}^*(\lambda^*))$ are primal feasible and satisfy complimentary slackness, then they are optimal primal values. In particular, if for each tone j there exists a unique user i that achieves the maximum in (15), then since there is no duality gap, allocating tone j only to that user must be a primal optimal solution.

However, if there are multiple users for a given tone whose metrics μ_{ij} are tied at the maximum value in (15), then there will be multiple primal values that satisfy (16), not all of which may be primal feasible (e.g., $\sum_i x_{ij}^* > 1$ or $\sum_{ij} p_{ij}^* > P$). Breaking these ties is necessary to find an optimal primal solution. A key point is that when ties occur at a given λ, $L(\lambda)$ is not differentiable at that λ. However, since $L(\lambda)$ is a convex function, subgradients exist.

Definition 1: Subgradients [28, pg. 214]: For a convex function $F(\boldsymbol{x}) : \Re^n \to \Re$ with domain $\mathcal{C} \subseteq \Re^n$ (a convex set), a vector $\boldsymbol{x}^* \in \Re^n$ is a *subgradient* of $F(\boldsymbol{x})$ at $\boldsymbol{x}$ if

$$F(\boldsymbol{z}) \geq F(\boldsymbol{x}) + \langle \boldsymbol{x}^*, \boldsymbol{z} - \boldsymbol{x} \rangle, \quad \forall\, \boldsymbol{z} \in \mathcal{C},$$

where $\langle \cdot, \cdot \rangle$ is the inner product in $\Re^n$. We denote the set of subgradients of $F(\boldsymbol{x})$ at $\boldsymbol{x}$ by $\partial F(\boldsymbol{x})$.

For an arbitrary λ, a subgradient (a scalar) for $L(\lambda)$ at λ can be found as follows:

Proposition 1: Let $(\hat{\boldsymbol{x}}(\lambda), \hat{\boldsymbol{p}}(\lambda))$ be a solution to (16) for a given λ that satisfies $\sum_i \hat{x}_{ij}(\lambda) \leq 1$ for all j and $\mu_j^*(\lambda)\left(1 - \sum_i \hat{x}_{ij}(\lambda)\right) = 0$ for each j, then $P - \sum_{i,j} \hat{p}_{ij}(\lambda)$ is a subgradient of $L(\lambda)$ at λ.

In the Appendix we prove that the conditions of Proposition 1 are, in fact, necessary and sufficient, i.e., every subgradient of $L(\lambda)$ at λ is characterized in the manner stated above.

For a given λ, let $\mathcal{A}_j := \{i | \mu_{ij}^*(\lambda) = \max_i \mu_{ij}^*(\lambda)\}$ be the set of users who achieve the maximum metric on tone j, and $|\mathcal{A}_j|$ be the size of set $\mathcal{A}_j$. From the previous analysis it follows that the set of all $\boldsymbol{x}(\lambda)$ that solve (16) are those that satisfy the following properties:

i) for $i \notin \mathcal{A}_j$, $x_{ij}^*(\lambda) = 0$;
ii) if $|\mathcal{A}_j| = 1$, then $x_{ij}^*(\lambda) = 1$ for $i \in \mathcal{A}_j$; and
iii) if $|\mathcal{A}_j| > 1$, then for all $i \in \mathcal{A}_j$, $x_{ij}^*(\lambda) \in [0,1]$ and $\sum_{i \in \mathcal{A}_j} x_{ij}^*(\lambda) = 1$.

In case (*iii*), we must break ties to determine the allocation for that tone. We refer to an allocation as an *extreme point* if it satisfies (*i*)-(*iii*) and $x_{ij}^*(\lambda) \in \{0,1\}$ for all i and j; such an allocation can be represented by a function $f : \mathcal{N} \to \mathcal{K}$, so that $f(j)$ indicates the user who is allocated channel j, i.e., $x^*_{f(j)j}(\lambda) = 1$. To satisfy (*i*)-(*iii*), it must be that $f(j) \in \mathcal{A}_j$ for all j. Let $\mathcal{B} = \{j : |\mathcal{A}_j| = 1\}$ and $\mathcal{B}^c = \mathcal{N} \setminus \mathcal{B}$. For each $j \in \mathcal{B}$, there are no ties, and so $f(j)$ is unique. For each tone $j \in \mathcal{B}^c$, there are $|\mathcal{A}_j|$ users in the tie, and so $f(j)$ can take $|\mathcal{A}_j|$ values. Hence, the total number of extreme points is $\prod_{j\in\mathcal{B}^c} |\mathcal{A}_j|$.

Each extreme point satisfies Proposition 1 and so provides a subgradient for $L(\lambda)$. From Proposition 1 it is also clear that all the subgradients of $L(\lambda)$ can be obtained as a convex combination of the values at the extreme points. Let

$$\tilde{p}_{ij} := \frac{1}{\tilde{e}_{ij}} \left[q\left(\beta, \left(\frac{w_i \tilde{e}_{ij}}{\lambda} - 1\right)^+\right) \wedge \tilde{s}_{ij}\right]. \tag{17}$$

Given an extreme point f, from (11), the resulting power allocation to tone j is given by $\tilde{p}_{ij}\mathbf{1}_{\{i=f(j)\}}$ where we use $\mathbf{1}_X$ to denote the (regular) indicator function of set X. Hence, the corresponding subgradient $d(f)$ is given by

$$d(f) = P - \sum_{j\in\mathcal{B}} \tilde{p}_{f(j)j} - \sum_{j\in\mathcal{B}^c} \tilde{p}_{f(j)j}. \tag{18}$$

Choosing different extreme points only effects the last term on the right of (18). It follows that the maximum subgradient of $L(\lambda)$ corresponds to the extreme points given by

$$\hat{f}(j) := \arg\min_{i\in\mathcal{A}(j)} \tilde{p}_{ij},\ \forall j. \tag{19}$$

Likewise, the minimum subgradient is given by extreme points that satisfy

$$\bar{f}(j) := \arg\max_{i\in\mathcal{A}(j)} \tilde{p}_{ij},\ \forall j. \tag{20}$$

At λ^*, the maximum subgradient given by an extreme point in (19) is always nonnegative, and the minimum subgradient given by an extreme point in (20) is always non-positive. If any of them is zero, then an optimal primal solution is found. Otherwise, we can find an optimal primal solution by time sharing as follows:

Proposition 2: There exists an optimal primal solution $(\boldsymbol{x}^*(\lambda^*), \boldsymbol{p}^*(\lambda^*))$, where $\boldsymbol{x}^*(\lambda^*)$ is given by time-sharing between the two extreme points in (19) and (20), and $\boldsymbol{p}^*(\lambda^*)$ is given by (11).

Proposition 2 implies that there is always an optimal primal solution for which at most two users time-share any tone. Moreover, each tone that is time-shared is shared in the same proportion. The optimal time-sharing factor can be found by the convex combination of the subgradients corresponding to (19) and (20) that is equal to zero.

The above steps give us an algorithm for finding the optimal solution to (8) in two stages:

1) First, find λ^* that minimizes $L(\lambda)$ as in Section III-A. This involves evaluating $L(\lambda)$ for a fixed value of λ as an inner loop, and a one-dimensional search over λ as an outer loop. The outer loop has a constant complexity that is independent of N and K[8]. The inner loop has a complexity of $O(NK)$ due to searching for the maximum of K metrics (15) on each of the N tones. Thus the total complexity of this stage is $O(NK)$.
2) Second, given λ^*, compute the maximum and minimum extreme points and find the optimal primal variables as in Proposition 2. The complexity of the second stage is also $O(NK)$, since for each tone we need to find the maximum and minimum in (19) and (20), respectively.

It follows from the above that the overall complexity of the optimal algorithm is $O(NK)$.

In our simulations, the actual complexity of the second stage is typically much smaller than $O(NK)$. This is due to the fact that "typically" only a few ties occur. For example, Table V in Section IV shows that for a system of 64 subchannels (grouped from 512 tones) and 40 users in a high mobility environment, there are on average only two extreme points at each scheduling interval (averaged over 3000 scheduling intervals) under either adjacent or random subchannelization. This arises because there is a tie on only one subchannel, and involving only two-users. The number of extreme points can be very large under interleaved subchannelization. This is because if two users are tied on one subchannel, it is very likely that they will also be tied on the other subchannels since all subchannels have roughly the same channel gain for the same user. However, if all the ties are due to the same two users, we can just allocate all subchannels with a tie to the same user and this will lead to either the largest or smallest subgradient. These observations are consistent with the work in [21], which argues that for a OFDM system with $\beta = 0$ in which no time-sharing is allowed, the duality gap as defined in [21] is small for a reasonable number of sub-channels (roughly 8 in their numerical examples). Problem (8) can be viewed as the dual of the dual problem studied in [21, eqn. (9)] and the duality gap in [21] can be viewed as a measure of the accuracy of approximating the OFDMA scheduling problem by the time-sharing version of it from (8). Note that when there is exactly one extreme point then the duality gap is clearly zero (since this will correspond to an integer solution). The arguments in [21] for a vanishing duality gap roughly corresponds to showing that the spread in the power consumption of different extreme points (i.e. the difference in the corresponding subgradient values) is typically small for a reasonable number of carriers. When this spread is small, one expects that fewer ties are occurring which is consistent with the above discussion. Furthermore, it can be seen that the arguments in [21] extend to the $\beta > 0$ case as well.

C. Single user per tone

In this section we consider the case where no time sharing is allowed, i.e., $x_{ij} \in \{0, 1\}$ for all i and j. Suppose we still find the optimal value of λ^* as in Section III-A. If there are no ties on any of the tones or if there is an extreme point with $\sum_{j\in\mathcal{N}} \tilde{p}_{f(j)j} = P$, the optimal primal solution given in Section III-B only allows one user per tone, and we are done. If not, Proposition 2 will no longer give a feasible primal solution, i.e., one that satisfies the integer constraints. In this case, a reasonable heuristic is to simply

[8] The computational complexity of a bi-section search is $O(\log(1/\epsilon))$, where ϵ is the relative error bound target for the search.

choose one extreme point allocation. In our simulations, we choose the extreme point corresponding to the subgradient with the smallest non-negative value; i.e., the extreme point f, for which $\sum_{j\in\mathcal{N}} \tilde{p}_{f(j)j}$ is closest to P, without exceeding it. Other rules for choosing an extreme point could also be used. Note that this requires searching over all extreme points, which has a worst-case complexity of $O(K^N)$ (if all users were tied on every tone). However, as discussed above, typically there are only two users tied on one tone and so this has almost constant complexity. If instead the largest or smallest subgradient was used, the worst-case complexity would again be $O(NK)$.

For a given extreme point f, the total transmit power $\sum_{j\in\mathcal{N}} \tilde{p}_{f(j)j}$ will be either greater or lesser than the constraint P (unless this point is optimal). We then need to re-optimize the power allocation for the given fixed feasible tone allocation $\boldsymbol{x}$ (i.e., $x_{ij} = 1$ if $i = f(j)$, otherwise x_{ij}=0), i.e., solve

$$\max_{\boldsymbol{p}:(\boldsymbol{p},\boldsymbol{x})\in\mathcal{X}} V(\boldsymbol{x},\boldsymbol{p}) \quad \text{s.t.} \sum_{i,j} p_{ij} \le P. \tag{21}$$

Let $L_{\boldsymbol{x}}(\lambda)$ be the dual function for this problem. Given $\tilde{\lambda} = \arg\min_{\lambda\ge 0} L_{\boldsymbol{x}}(\lambda)$, the optimal power allocation to (21) is given by (11) with $\lambda = \tilde{\lambda}$ and the given tone allocation $\boldsymbol{x}$. A simple one-dimensional search can again be used to find the optimal λ. This will have a complexity of $O(N)$ (to get within ϵ of the optimal) since each tone has at most one user.

When the self-noise term $\beta = 0$, we can actually find the optimal $\tilde{\lambda}$ in finite steps based on the following alternative characterization for the correct $\tilde{\lambda}$, the proof of which is based on a similar argument as in [10].

Proposition 3: Consider the case where $\beta = 0$. A given $\hat{\lambda}$ is the unique optimal solution to the dual problem $\min_{\lambda\ge 0} L_{\boldsymbol{x}}(\lambda)$ if and only if

$$\hat{\lambda} = \frac{\sum_{i,j} x_{ij} w_i 1_{\{\hat{\lambda}\in\mathcal{W}_{ij}\}}}{P - \sum_{i,j} \frac{\Gamma_{ij}}{e_{ij}} 1_{\{\hat{\lambda}\in\mathcal{Y}_{ij}\}} + \sum_{i,j} \frac{1}{e_{ij}} 1_{\{\hat{\lambda}\in\mathcal{W}_{ij}\}}}, \tag{22}$$

where $\mathcal{W}_{ij} = \left[\frac{x_{ij}w_i e_{ij}}{1+\Gamma_{ij}}, x_{ij} w_i e_{ij}\right)$, and $\mathcal{Y}_{ij} = \left[0, \frac{x_{ij}w_i e_{ij}}{1+\Gamma_{ij}}\right)$.

Proposition 3 suggests the following algorithm for finding $\tilde{\lambda}$. First check if the power constraint is violated when all users use maximum power on the allocated tones, i.e., if $\sum_{(i,j)} \frac{x_{ij}}{e_{ij}} \Gamma_{ij} > P$. If this is not true, the problem is solved. If this is true, we need to search for $\tilde{\lambda}$. Let $\boldsymbol{b}$ be a vector of length $2N$ containing the values of $x_{ij} w_i e_{ij}$ and $\frac{x_{ij}w_i e_{ij}}{1+\Gamma_{ij}}$ for all (i,j) such that $x_{ij} = 1$, sorted in descending order. Define two additional vectors $\boldsymbol{z}$ and $\boldsymbol{y}$ such that for any $k = 1, ..., 2N$, $z(k) = 1$ if $b(k) = \frac{x_{ij}w_i e_{ij}}{1+\Gamma_{ij}}$ for some (i,j) and 0 otherwise, and $y(k) = \{i,j\}$ if $b(k) = x_{ij} w_i e_{ij}$ or $\frac{x_{ij}w_i e_{ij}}{1+\Gamma_{ij}}$.

The complete λ search algorithm is given in Algorithm 1. The basic idea is to start from the largest λ, and calculate the right-hand side of (22). If it is less than the current value of λ, decrease λ and recalculate, until a fixed-point is found. It can be shown that the algorithm will stop in at most $2N$ steps with $\lambda(k) = \tilde{\lambda}$.

This algorithm requires a complete sort of the $2N$ values in the vector $\boldsymbol{b}$ and so has a complexity of $O(N \log N)$. Note that this is larger than the $O(N)$ complexity required by a one-dimensional search, but yields the exact optimal solution in finite time as opposed to an ϵ-optimal solution. Regardless of how the power allocation is determined, this algorithm still requires finding the optimal λ^*, which as complexity $O(NK)$ as before. Combining these observations, it follows that if the largest or smallest subgradients are used to break ties, the overall algorithm will have a complexity of $O(NK)$ or $O(NK + N\log N)$ depending on if the power is re-optimized using a one-dimensional search over λ or Algorithm 1, respectively.

Algorithm 1 Search Algorithm for Optimal λ under a Fixed Tone Assignment (with $\beta = 0$)

1) Initialization: $k = 0$, $G_{xw} = 0$, $G_{s/e} = 0$ and $G_{1/e} = 0$.
2) $k = k + 1$.
3) Let $\{i(k), j(k)\} = y(k)$.
4) If $z(k) = 0$, then $G_{xw} = G_{xw} + x_{i(k)j(k)} w_{i(k)}$ and $G_{1/e} = G_{1/e} + \frac{1}{e_{i(k)j(k)}}$.
5) If $z(k) = 1$, then $G_{s/e} = G_{s/e} + \frac{\Gamma_{i(k)j(k)}}{e_{i(k)j(k)}}$, $G_{xw} = G_{xw} - x_{i(k)j(k)} w_{i(k)}$, and $G_{1/e} = G_{1/e} - \frac{1}{e_{i(k)j(k)}}$.
6) Let $\lambda(k) = G_{xw} / \left(P - G_{s/e} + G_{1/e}\right)$.
7) If $\lambda(k) \le b(k)$ and $\lambda(k) \ge b(k+1)$, stop. Otherwise, go to step (2).

D. Single sort suboptimal algorithm

The optimal tone allocation is determined by assigning each tone j to the user with the largest metric $\mu_j^*(\lambda^*)$ on that tone (breaking any ties as discussed above). This requires iterating to find the optimal Lagrange multiplier λ^* in the first place. Now we introduce two sub-optimal algorithms that do not require finding the optimal λ^* iteratively. Instead, a carrier allocation is determined by a single sort on each tone based on some easily calculated metric. These heuristic algorithms are much faster than the previous algorithms, although their complexity is again $O(NK)$.

1) HEURISTIC 1: Each subchannel j is allocated to the user with the largest value of $w_i \bar{R}_{ij}$, where

$$\bar{R}_{ij} = \log\left[1 + \left(\Gamma_{ij} \wedge \left(\frac{\tilde{e}_{ij} P/N}{1 + \beta\tilde{e}_{ij} P/N}\right)\right)\right]$$

is the rate user i could achieve on subchannel j under a constant power allocation of P/N. Any ties are broken arbitrarily. Constant power is allocated on all subchannels. This metric was motivated in part by work in [14], [17] where a uniform power allocation (not necessarily over all tones) was shown to be nearly optimal.

2) HEURISTIC 2: Here subchannels are allocated as in HEURISTIC 1. However, after the allocation is determined, an optimal power allocation is performed as in Section III-C (instead of constant power allocation). It may turn out that no power is assigned to some subchannels.

IV. Simulation Study

We report simulation results based on a realistic OFDM simulator with parameters and assumptions commonly found in IEEE 802.16 standards [11]. We focus on the following three algorithms: the OPTIMAL algorithm which finds the optimal λ^* and then chooses a tone-allocation with one user per tone as described in Section III-C[9], and the two algorithms (HEURISTIC 1 and HEURISTIC 2) described in Section III-D.

We simulate a single OFDM cell with $M = 40$ users and a total transmission power of $P = 6$W at the base station. The channel gains e_{ij}'s are the products of a fixed location-based term for each user i and a frequency-selective fast time-scale/short-term fading term. The location-based components are picked using an empirically obtained distribution for many users in a large system. The fast time-scale fading term is generated using a block-fading model based upon the Doppler frequency (for the block-length in time) and a standard reference mobile delay-spread model (for variation in frequency). For a user's fast time-scale fading term, each multi-path component is held fixed for $2m\mathrm{sec}$ (i.e., a fading block length), which corresponds to a 250Hz Doppler frequency. The delay-spread is set to $1\mu\mathrm{sec}$. The users' channel conditions are averaged over the applicable channelization scheme and fed back to the scheduler at the base station.

We consider a system bandwidth of 5MHz consisting of 512 OFDM tones, grouped into 64 subchannels (8 tones per subchannel). The symbol duration is $100\mu\mathrm{sec}$ with a cyclic prefix of $10\mu\mathrm{sec}$, which roughly corresponds to 20 OFDM symbols per fading block (i.e., $2m\mathrm{sec}$). This is one of the allowed configurations in the IEEE 802.16 standards [11]. The resource allocation is done once per fading block. All the results are averaged over the last 2000 OFDM symbols out of 60000 OFDM symbols (i.e., 3000 fading blocks) by which time we can be reasonably confident that the system has reached stationarity. All users are infinitely back-logged and assigned a throughput-based utility as in (2) with parameter $c_i = 1$ and the same fairness parameters (α) across users. We solve problem (3) once at each scheduling instance.

We calculate the rate of user i on subchannel j as

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

where B is the subchannel bandwidth. Here 0.56 accounts for the "SNR gap" due to limited modulation and coding choices and 0.28 accounts for various factors such as hybrid ARQ transmission scheme and the overhead due to guard tones and control symbols, etc. While the scheduling is based on the geometric average for $\beta = 0$ and harmonic average for $\beta > 0$, the decoded rate is based on per tone channel conditions.[10]

The first set of simulation results are for a system with adjacent subchannelization, no self-noise ($\beta = 0$), and no per-user SNR constraints (i.e., $\tilde{s}_{ij} = \infty$ for all i and j). Table I shows the results for all three algorithms under different choices of the utility parameter α. The column "Utility" gives the average utility per user for each algorithm. The column "log U" shows the log utility per user; this gives some indication of the "fairness" of the resulting allocation (for $\alpha = 0$ this is the same as the utility). The column "Rate" is the average throughput per user, and the final column is the average number of users scheduled. For each choice of α, the three algorithms perform close to each other for each of these metrics. HEURISTIC 2 performs better than HEURISTIC 1, since the former re-optimizes the power allocation after tone allocation, and the latter just uses constant power allocation. When $\alpha = 1$ (maximum throughput), all three algorithms have almost identical performance.

TABLE I
PERFORMANCE FOR DIFFERENT CHOICES OF α (ADJACENT CHANNELIZATION, NO-SELF-NOISE, NO SNR CONSTRAINTS). RATE IS IN KBPS.

α	Algorithm	Utility	Log U	Rate	Num.
0	OPTIMAL	10.74	10.74	60.8	7.73
0	HEURISTIC 1	10.66	10.66	54.6	7.29
0	HEURISTIC 2	10.72	10.72	57.3	7.35
0.5	OPTIMAL	545.2	10.83	105.9	7.32
0.5	HEURISTIC 1	528.8	10.73	99.3	7.20
0.5	HEURISTIC 2	542.8	10.81	103.2	7.01
1	OPTIMAL	261677	6.79	261.7	2.58
1	HEURISTIC 1	261676	6.79	261.7	2.58
1	HEURISTIC 2	261676	6.77	261.7	2.58

TABLE II
PERFORMANCE OF DIFFERENT CHANNELIZATION SCHEMES ($\alpha = 0.5$, NO SELF-NOISE, NO SNR CONSTRAINTS). RATE IS IN KBPS.

Channelization	Algorithm	Utility	Log U	Rate	Num.
Adjacent	OPTIMAL	545.15	10.83	105.9	7.32
Adjacent	HEURISTIC 1	528.83	10.73	99.3	7.20
Adjacent	HEURISTIC 2	542.84	10.81	103.2	7.01
Interleaved	OPTIMAL	494.61	10.53	92.4	1.79
Interleaved	HEURISTIC 1	486.40	10.47	88.4	1.14
Interleave	HEURISTIC 2	487.02	10.48	87.8	1.15
Random	OPTIMAL	487.53	10.53	89.2	4.89
Random	HEURISTIC 1	479.07	10.46	84.2	4.39
Random	HEURISTIC 2	485.63	10.51	86.5	4.34

Next we consider the effect of different channelization schemes. Table II shows the performance of the three algorithms for the adjacent, random, and interleaved channelization schemes described in Section II-A. We set $\alpha = 0.5$, $\beta = 0$, and $\tilde{s}_{ij} = \infty$ for all i and j. Again, both HEURISTIC algorithms perform close to the OPTIMAL algorithm. For all three algorithms, the interleaved and random channelizations result in lower utility than the adjacent. This is likely due to higher frequency diversity with the adjacent subchannelization scheme. Indeed, for the channel model used here, in the interleaved case all subchannels can be shown to be almost identical, which explains why it typically schedules only one or two users.

Next we consider the case when the self-noise coefficient

[9] We have simulated both the algorithms in Section III-B and III-C, and found that they have identical performance under all parameter choices. This could be due to the fact that the gap in making the time-sharing assumption is small owing to there being very few significantly different extreme points at each scheduling interval as discussed at the end of Section III-B. We thus refer to the algorithm in Section III-C simply as the OPTIMAL algorithm.

[10] That is, the decoded rate is given by $\hat{r}_{ij} = 0.28\frac{B}{|\mathcal{N}_j|}x_{ij}\sum_{j_k\in\mathcal{N}_j}\log\left(1+\frac{0.56 p_{ij}\tilde{e}_{ij_k}}{x_{ij}+\beta p_{ij}\tilde{e}_{ij_k}}\right)$, where $\mathcal{N}_j$ is the set of tones in the jth subchannel and $\tilde{e}_{ij_k}$ is the SNR per unit power for tone j_k. This is reasonable if hybrid ARQ is used.

TABLE III
PERFORMANCE OF DIFFERENT CHANNELIZATION SCHEMES ($\alpha = 0.5$, SELF-NOISE $\beta = 0.01$, NO SNR CONSTRAINTS). RATE IS IN KBPS.

Channelization	Algorithm	Utility	Log U	Rate	Num.
Adjacent	OPTIMAL	512.20	10.82	82.5	7.52
Adjacent	HEURISTIC 1	489.32	10.70	73.7	7.40
Adjacent	HEURISTIC 2	504.00	10.78	77.2	7.22
Interleaved	OPTIMAL	467.00	10.51	73.5	1.98
Interleaved	HEURISTIC 1	453.16	10.43	66.8	1.26
Interleave	HEURISTIC 2	454.59	10.44	66.9	1.27
Random	OPTIMAL	460.53	10.51	71.6	5.60
Random	HEURISTIC 1	446.58	10.42	64.7	4.89
Random	HEURISTIC 2	453.51	10.48	66.1	4.85

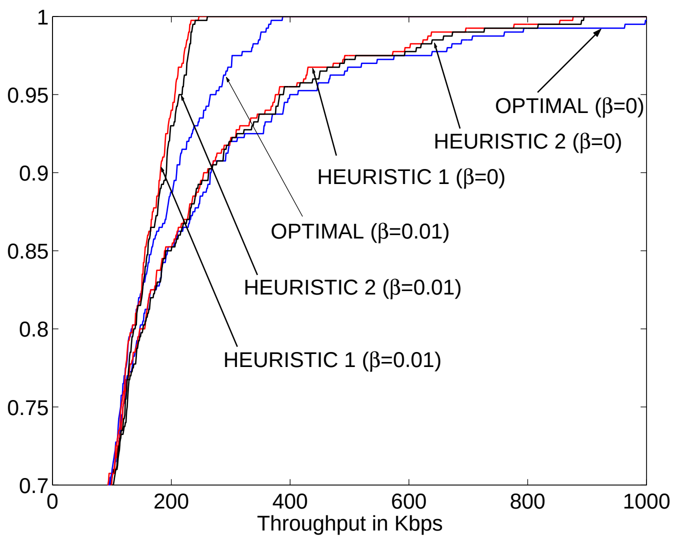

Fig. 2. Empirical CDF of users' throughputs (adjacent subchannelization, $\alpha = 0.5$, no per-user SNR constraints).

$\beta = 0.01$ in Table III. Here we assume $\alpha = 0.5$, and there is no per-user SNR constraint. With $\beta = 0.01$, the SNR values are upperbounded by $1/\beta = 20$dB. The performance gap between the three algorithms is slightly larger compared to the case without self-noise in Table II.

In Figure 2, we plot the throughput CDFs for all three algorithms, with self-noise ($\beta = 0.01$) and without ($\beta = 0$). Here adjacent channelization is used, $\alpha = 0.5$, and $\tilde{s}_{ij} = \infty$ for all i and j. It is clear that users achieve better throughput when there is no self-noise ($\beta = 0$). The OPTIMAL algorithm always achieves better rates compared with HEURISTIC ones under the same value of β. In fact, the OPTIMAL algorithm almost stochastically dominates the HEURISTIC ones (except at values very close to the lowest achievable rates) for both values of β.

In Table IV, we consider the effect of SNR constraints. In particular, we choose the SNR constraints to be ∞, 30dB, and 20dB, respectively, and the same across all users and all tones. We choose adjacent subchannelization with utility parameter $\alpha = 0.5$ and no self-noise. Comparing with the case with no SNR constraints, a constraint of 30dB does not change the results significantly, while a constraint of 20dB leads to substantial decrease in terms of achievable rates (13% for the OPTIMAL algorithm and 27% for HEURISTIC 1 algorithm).

TABLE IV
PERFORMANCE OF DIFFERENT SNR CONSTRAINTS (ADJACENT SUBCHANNELIZATION, $\alpha = 0.5$, NO SELF-NOISE). RATE IS IN KBPS.

SNR Con.	Algorithm	Utility	Log U	Rate	Num.
∞	OPTIMAL	545.15	10.83	105.9	7.32
∞	HEURISTIC 1	528.83	10.73	99.3	7.20
∞	HEURISTIC 2	542.84	10.81	103.2	7.01
30dB	OPTIMAL	542.78	10.83	102.97	7.33
30dB	HEURISTIC 1	519.81	10.72	91.87	7.25
30dB	HEURISTIC 2	535.89	10.81	96.35	7.10
20dB	OPTIMAL	522.48	10.82	88.11	7.40
20dB	HEURISTIC 1	483.50	10.66	72.60	7.09
20dB	HEURISTIC 2	505.81	10.77	78.61	6.92

TABLE V
DISTRIBUTION OF NUMBER OF TIES FOR THE ALGORITHM IN SECTION III-B (SELF-NOISE $\beta = 0.01$).

α	Channelization	2 ties	3 ties	4 ties	>= 5 ties
0	Adjacent	50.80%	0.09%	2.8%	0.14%
0	Random	51.58%	0.08%	2.69%	0.12%
0.5	Adjacent	56.81%	0.08%	4.81%	0.33%
0.5	Random	56.56%	0.08%	4.81%	0.33%
1	Adjacent	12.93%	0.03%	4.16%	3.83%
1	Random	8.79%	0.01%	2.17%	1.06%

Finally, in Table V, we investigate the average number of ties that occur under the algorithm described in Section III-B. We present the results for different α values under both adjacent and random subchannelizations with $\beta = 0.01$. For each row, we record the number of ties over 30000 scheduling time slots and calculate the percentages. Typically (around 99% of the time for most cases) very few ties occur (e.g., less than 5). Similar results were also observed for $\beta = 0$.

V. CONCLUSIONS

We have considered the problem of gradient-based scheduling and resource allocation for a downlink OFDM system, which essentially reduces to solving an optimization problem in each time-slot. We studied this problem for a general model that can accommodate various choices for user utility functions, different sub-channelization techniques, and self-noise due to imperfect channel estimates or phase noise. We first studied a relaxed version of this problem in which users can time-share each subchannel. Using duality theory we gave an optimal algorithm for solving this problem. This involves finding a maximum of a per user (closed form) metric for each sub-channel and a one-dimensional search of an optimal dual variable. More interestingly, a typical solution obtained by the optimal solution automatically yields an integer carrier allocation constraint (except on one or two tones). To enforce such a constraint on all tones, we further propose an algorithm that always picks an integer carrier allocation and re-optimizes the power allocation accordingly. The numerical performance of this algorithm is almost identical to the optimal solution of the relaxed problem. Finally, we propose two even simpler suboptimal algorithms that only perform a single sort on each of the tones and avoid any iterative calculations. Simulations once again show that the suboptimal algorithms achieve close to optimal performance under a wide range of scenarios, and the performance gap widens when per user SNR constraints or channel estimation errors are considered.

APPENDIX A
PROOF OF PROPOSITION 1

Before proceeding, define the (convex) indicator function of a convex set $\mathcal{C} \subseteq \Re^n$ to be

$$\delta(\boldsymbol{x}|\mathcal{C}) := \begin{cases} 0, & \text{if } \boldsymbol{x} \in \mathcal{C}, \\ +\infty, & \text{otherwise.} \end{cases} \quad (23)$$

From (14) and [29, Corollary 10.9, pp. 430-431] we have[11] $\partial L(\lambda) = \{P\} + \sum_{j=1}^K \partial \mu_j^*(\lambda)$. Hence, it suffices to characterize $\partial \mu_j^*(\lambda)$ for every j.

Let $\hat{L}_j(\lambda, \mu_j) := \mu_j - L_j(\lambda, \mu_j)$, where $L_j(\lambda, \mu_j)$ is defined in (13). Note that $\mu_j^*(\lambda) = \min_{\mu_j \geq 0}\{\mu_j - \hat{L}(\lambda, \mu_j)\}$. Applying [29, Theorem 10.13, pp. 433-434], it follows that

$$\partial \mu_j^*(\lambda) = \{y : (y, 0) \in \Big[(0,1) + \big(0, \partial\delta(\tilde{\mu}_i|\{\mu_i \geq 0\})\big) - \partial\hat{L}_j(\lambda, \tilde{\mu}_j)\Big]\Big\}, \quad (24)$$

for any $\tilde{\mu}_j \in \arg\min_{\mu_i \geq 0} \mu_j - \hat{L}_j(\lambda, \mu_j)$.

Similarly, let $\hat{L}_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j) = \mu_j - L_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j)$, where $L_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j)$ is defined in (10). It follows that $\hat{L}(\lambda, \mu_j) = \min_{(\boldsymbol{x}^j, \boldsymbol{p}^j) \in \mathcal{X}_j} \hat{L}_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j)$. From another application of [29, Theorem 10.13] we then have

$$\partial \hat{L}_j(\lambda, \mu_j) = \{(y, z) : (\mathbf{0}, \mathbf{0}, y, z) \in \left[\nabla \hat{L}_j(\tilde{\boldsymbol{x}}^j, \tilde{\boldsymbol{p}}^j, \lambda, \mu_j) + \left(\partial\delta\left((\tilde{\boldsymbol{x}}^j, \tilde{\boldsymbol{p}}^j)\Big|\mathcal{X}_j\right), 0, 0\right)\right]\right\}, \quad (25)$$

for any $(\tilde{\boldsymbol{x}}^j, \tilde{\boldsymbol{p}}^j) \in \arg\min_{(\boldsymbol{x}^j, \boldsymbol{p}^j) \in \mathcal{X}_j} \hat{L}_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j)$.

Note that $\frac{\partial \hat{L}_j(\lambda, \mu_j, \boldsymbol{x}^j, \boldsymbol{p}^j)}{\partial \lambda} = \sum_{i=1}^K p_{ij}$, $\frac{\partial \hat{L}_j(\lambda, \mu_j, \boldsymbol{x}^j, \boldsymbol{p}^j)}{\partial \mu_j} = \sum_{i=1}^K x_{ij}$, and finally that $(\hat{\boldsymbol{x}}(\lambda), \hat{\boldsymbol{p}}(\lambda))$ satisfies (16) if and only if $(\hat{\boldsymbol{x}}^j(\lambda), \hat{\boldsymbol{p}}^j(\lambda)) \in \arg\min_{(\boldsymbol{x}^j, \boldsymbol{p}^j) \in \mathcal{X}_j} \hat{L}_j(\boldsymbol{x}^j, \boldsymbol{p}^j, \lambda, \mu_j^*(\lambda))$ for all j. Hence, from (25)

$$\partial \hat{L}_j(\lambda, \mu_j^*(\lambda)) = \left\{ \left(\sum_{i=1}^K \hat{x}_{ij}(\lambda), \sum_{i=1}^K \hat{p}_{ij}(\lambda) \right) : (\hat{\boldsymbol{x}}^j(\lambda), \hat{\boldsymbol{p}}^j(\lambda)) \in \arg \min_{(\boldsymbol{x}^j, \boldsymbol{p}^j) \in \mathcal{X}_j} \hat{L}_j(\boldsymbol{x}^j, \boldsymbol{p}^j,) \right\}.$$

Next note that $\tilde{\mu}_j = \mu_j^*(\lambda)$ and that $\partial\delta(\tilde{\mu}_i|\{\mu_i \geq 0\}) = \{0\}1_{\{\tilde{\mu}_i > 0\}} + (-\infty, 0]1_{\{\tilde{\mu}_i = 0\}} + \emptyset 1_{\{\tilde{\mu}_i < 0\}}$, where $\emptyset$ is the empty set (see e.g. [28, pg. 226]). Using this in (24), it follows that we can write any subgradient of $\mu_j^*(\lambda)$ in terms of those elements of $\partial\hat{L}_j(\lambda, \mu_j^*(\lambda))$ for which the corresponding allocation $\hat{\boldsymbol{x}}^j(\lambda)$ satisfies

1) if $\mu_j^*(\lambda) > 0$, then $\sum_{i=1}^K \hat{x}_{ij}(\lambda) = 1$; and
2) if $\mu_j^*(\lambda) = 0$, then 0 must lie in $\left(-\infty, 1 - \sum_{i=1}^K \hat{x}_{ij}(\lambda)\right]$, i.e., $\sum_{i=1}^K \hat{x}_{ij}(\lambda) \leq 1$.

In either case, the value of the subgradient of $\mu_j^*(\lambda)$ is $P - \sum_{i=1}^K \hat{p}_{ij}(\lambda)$. ■

[11] The addition of sets $A, B \subseteq \Re^n$ is defined by $A + B := \{x + y : x \in A, y \in B\}$.

REFERENCES

[1] R. Agrawal and V. Subramanian, "Optimality of certain channel aware scheduling policies," *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, 2002.
[2] R. Agrawal, A. Bedekar, R. La, and V. Subramanian, "A class and channel-condition based weighted proportionally fair scheduler," *Proc. of ITC 2001*, Salvador, Brazil, Sept 2001.
[3] 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.
[4] H. Kushner and P. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," *40th Annual Allerton Conference on Communication, Control, and Computing*, 2002.
[5] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," *Proc. of IEEE Vehicular Technology Conference*, Spring, 2000.
[6] 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.
[7] A. L. Stolyar, "MaxWeight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic," *Annals of Applied Probability*, vol. 14, no. 1, pp. 1–53, 2004.
[8] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queue with randomly varying connectivity," *IEEE Trans. on Inform. Th.*, vol. 39, pp. 466–478, 1993.
[9] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, 2001.
[10] R. Agrawal, V. Subramanian, and R. Berry, "Joint scheduling and resource allocation in CDMA systems," in *Proc. WiOpt*, Cambridge, UK, March 2004. Journal version under submission.
[11] "IEEE 802.16e-2005 and IEEE Std 802.16-2004/Cor1-2005," `http://www.ieee802.org/16/`.
[12] "Long Term Evolution of the 3GPP radio technology," `http://www.3gpp.org/Highlights/LTE/LTE.htm`.
[13] H. Jin, R. Laroia, and T. Richardson, "Superposition by position," *2006 IEEE Information Theory Workshop*, March 2006.
[14] L. Hoo, B. Halder, J. Tellado, and J. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms," *IEEE Transactions on Communications*, vol. 52, no. 6, pp. 922–930, 2004.
[15] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun*, vol. 17, no. 10, pp. 1747–1758, 1999.
[16] Y. Zhang and K. Letaief, "Adaptive resource allocation and scheduling for multiuser packet-based OFDM networks," *2004 IEEE ICC*, vol. 5, 2004.
[17] J. Jang and K. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171–178, 2003.
[18] Y. Zhang and K. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Trans. on Wireless Communications*, vol. 3, no. 5, pp. 1566–1575, 2004.
[19] T. Chee, C. Lim, and J. Choi, "Adaptive power allocation with user prioritization for downlink orthogonal frequency division multiple access systems," *ICCS 2004*, pp. 210–214, 2004.
[20] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," *IEEE Globecom*, 2000.
[21] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA Downlink Systems," in *IEEE ISIT*, pp. 1394–1398, 2006.
[22] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels. I. Ergodic capacity," *IEEE Trans. on Information Theory*, vol. 47, no. 3, pp. 1083–1102, 2001.
[23] M. Medard, "The Effect Upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 935-946, May 2000.
[24] J. Lee, H. Lou and D. Toumpakaris, "Analysis of Phase Noise Effects on Time-Direction Differential OFDM Receivers," *IEEE GLOBECOM*, 2005
[25] *QUALCOMM Flarion Technologies, http://www.qualcomm.com/qft/.*
[26] J. Huang, V. Subramanian, R. Berry, and R. Agrawal, "Joint scheduling and resource allocation in OFDM systems: Algorithms and performance for the uplink," *Asilomar Conference*, 2007.
[27] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.

[28] R. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
[29] T. R. Rockafellar and R. J.-B. Wets, *Variational analysis*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 317, Springer-Verlag, Berlin, 1998.

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).

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 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'93-M'00) 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*.