OPTIMUM SCHEDULING FOR ORTHOGONAL MULTIPLE ACCESS OVER FADING CHANNELS USING QUANTIZED CHANNEL STATE INFORMATION

Antonio G. Marques, Georgios B. Giannakis†, and F. Javier Ramos**

*Dept. of Signal Theory and Communications, Rey Juan Carlos University (SPAIN)
†Dept. of Electrical and Computer Engineering, University of Minnesota (USA)

ABSTRACT

The efficiency of multi-access communications over wireless fading links benefits from channel-adaptive allocation of the available bandwidth and power resources. Different from most existing approaches that allocate resources based on perfect channel state information (P-CSI), this work optimizes channel scheduling and resource allocation over orthogonal fading channels when user terminals and the scheduler rely on quantized channel state information (Q-CSI). The novel unifying approach optimizes an average transmit-performance criterion subject to average quality of service requirements. The resultant optimal policy per fading realization either allocates the entire channel to a single (winner) user, or, to a small group of winner users whose percentage of shared resources is found by solving a linear program. Both alternatives become possible by smoothing the allocation scheme. The smooth policy is asymptotically optimal and incurs reduced computational complexity.

1. INTRODUCTION

The importance of channel-adaptive allocation of bandwidth and power resources in wireless multi-access connections over fading links has been well documented from both information theoretic and practical communication perspectives. Per fading realization, parameters including rate, power and percentages of time frames (or system subcarriers) are adjusted across users to optimize utility measures of performance quantified by bit error rate (BER), weighted sum-rate or power efficiency, under quality of service (QoS) constraints such as prescribed BER, delay, maximum power or minimum rate requirements. To carry out such constrained optimization tasks, most existing approaches assume that perfect CSI (P-CSI) is available wherever needed [3], [5], [6]. However, it is well appreciated that errors in estimating the channel, feedback delay, and the asymmetry between forward and reverse links render acquisition of deterministically perfect CSI at transmitters (P-CSIT) impossible in most wireless scenarios [2]. This has motivated scheduling and resource allocation schemes based on perfect CSI at the receivers (P-CSIR) but only quantized (Q-) CSIT that can be pragmatically obtained through finite-rate feedback from the receiver; see e.g., [4].

This work goes one step further to pursue optimal scheduling and resource allocation for orthogonal multi-access transmissions over fading links when only Q-CSI is available both at receiver(s) and transmitter(s). The unifying approach minimizes an average power cost (or in a dual formulation maximizes utility functions of average rates) subject to average QoS constraints on rate (respectively power) related constraints. This setup is particularly suited for systems where the receiver does not have accurate channel estimates (e.g., when differential (de-)modulation is employed), or, in sensor networks where scheduling and resource allocation are decided at the fusion center which can only acquire Q-CSI sent by the sensors.

The rest of the paper is organized as follows. After modelling preliminaries, the general problem is formulated in Section 2, and the optimal solution is characterized in Section 3. A smooth policy that reduces complexity and guarantees asymptotic optimality is developed in Section 4, followed by numerical tests described in Section 5. Concluding remarks are offered in Section 6.[1]

2. PRELIMINARIES AND PROBLEM STATEMENT

Consider a wireless network with M user terminals, indexed by $m \in \{1, \ldots, M\}$, transmitting over K flat-fading orthogonal channels, indexed by $k \in \{1, \ldots, K\}$, to a common

Work in this paper was supported by the US ARL under the CTA Program, Cooperative Agreement No. DAAD19-01-2-0011; by the US-DoD ARO grant No. W911NF-05-1-0283; by the Spanish Govt grant No. TEC2005-06766- C03-01/TCM; and by the C.A. Madrid grant No. P-TIC-000223-0505.

[1] *N*otation: Boldface upper (lower) case letters are used for matrix (column vectors); $(\cdot)^T$ denotes transpose; $[\cdot]_{k,l}$ the (k,l)th entry of a matrix, and $[\cdot]_k$ the (k)th column (entry) of a matrix (vector); $\cdot$ stands for entrywise matrix product and also to denote differentiation; $\mathbf{1}$ and $\mathbf{0}$ are the all-one and all-zero matrices. Calligraphic letters are used for sets with $|\mathcal{X}|$ denoting cardinality of the set $\mathcal{X}$. For a random scalar (matrix) variable x ($\mathbf{X}$), the univariate (multivariate) probability density function (pdf) is denoted by $f_x(x)$ (respectively $f_{\mathbf{X}}(\mathbf{X})$) and its cumulative distribution function (cdf) by $F_x(x)$ (respectively $F_{(\mathbf{X})}(\mathbf{X})$). Finally, $\wedge$ denotes the "and" logic operator, x^* the optimal value of variable x; and, $\mathbf{I}_{\{\cdot\}}$ the indicator function ($\mathbf{I}_{\{x\}} = 1$ if x is true and zero otherwise).

destination, e.g., base station or access point. Zero-mean additive white Gaussian noise (AWGN) with unit variance is assumed at the receiver. With $g_{m,k}$ denoting the kth channel's instantaneous gain (magnitude square of the fading coefficient) between the mth user and the destination, the overall channel is described by the $M \times N$ matrix $\mathbf{G}$ for which $[\mathbf{G}]_{m,k} := g_{m,k}$. The range of values each $g_{m,k}$ takes is divided into non-overlapping regions; and instead of $g_{m,k}$ itself, destination and transmitters have available only the binary codeword indexing the region $g_{m,k}$ falls into. With $j_{m,k}$ representing the corresponding region index, the $M \times N$ matrix $\mathbf{J}$ with entries $[\mathbf{J}]_{m,k} := j_{m,k}$ constitutes the Q-CSI of the overall system. Since $g_{m,k}$ is random, $j_{m,k}$ is also a discrete random variable; and likewise $\mathbf{J}$ is random taking matrix values from the set $\mathcal{J}$ with finite cardinality $|\mathcal{J}|$.

As in [4] or [6], users at the outset can be scheduled to access *simultaneously but orthogonally* (in time or frequency) any of the K channels. The channel scheduling policy is described by the matrix $\mathbf{W}$ whose nonnegative entry $[\mathbf{W}]_{m,k}$ corresponds to the percentage of the kth channel scheduled for the mth user. Clearly, it holds that $\sum_{m=1}^{M}[\mathbf{W}]_{m,k} \in [0,1]$ $\forall k$. The system power and rate resources are collected in the $K \times M$ matrices $\mathbf{P}$ and $\mathbf{R}$. Each of the corresponding entries $[\mathbf{P}]_{m,k}$ and $[\mathbf{R}]_{m,k}$ represent, respectively, the *nominal* power and rate the mth user terminal would be allocated if it were the only terminal scheduled to transmit over the kth channel. Since scheduling and allocation are to be adapted based on Q-CSI, matrices $\mathbf{W}$, $\mathbf{P}$ and $\mathbf{R}$ will be dependent on $\mathbf{J}$ and each can take at most $|\mathcal{J}|$ different values. Under BER or capacity constraints, rate and power variables are coupled. This power-rate coupling will be represented by the function Υ (respectively Υ^{-1} for the rate-power coupling) and will relate $[\mathbf{P}]_{m,k}$ with $[\mathbf{R}]_{m,k}$ over the same Q-CSI region $\mathcal{R}$. (We will write $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ to exemplify this dependence.)

2.1. Problem Formulation

Given the Q-CSI matrix $\mathbf{J}$ and prescribed QoS requirements, the goal is to find $\mathbf{W}(\mathbf{J})$, $\mathbf{P}(\mathbf{J})$ and $\mathbf{R}(\mathbf{J})$ so that the *overall* average *weighted* performance is optimized. (Overall here refers to performance of all users and weighted refers to different user priorities effected through the weight vector $\boldsymbol{\mu} := [\mu_1, \ldots, \mu_M]^T$ with nonnegative entries.) Depending on desirable objectives, the problem can be formulated either as constrained utility maximization of the average weighted sum-rate subject to average power constraints; or, as a constrained minimization of the average weighted power subject to average rate constraints. Although focus will be placed here on power minimization, both problems can be tackled in parallel by dual substitutions; namely, after interchanging the roles of $\mathbf{R}$ and $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ by $\mathbf{P}$ and $\Upsilon^{-1}_{\mathcal{R}([\mathbf{J}]_{m,k})}$, respectively.

Specifically, the weighted average transmit-power will be minimized subject to individual minimum average rate constraints collected in the vector $\check{\mathbf{r}} := [\check{r}_1, \ldots, \check{r}_M]^T$. Per Q-CSI realization $\mathbf{J}$, the overall weighted transmit-power is given by $\sum_{m=1}^{M}[\boldsymbol{\mu}]_m \sum_{k=1}^{K}([\mathbf{P}(\mathbf{J})]_{m,k}[\mathbf{W}(\mathbf{J})]_{m,k})$; while the mth user's transmit-rate is $\sum_{k=1}^{K}([\mathbf{R}(\mathbf{J})]_{m,k}[\mathbf{W}(\mathbf{J})]_{m,k})$. Using the probability mass function $\Pr\{\mathbf{J}\}$, these expressions can be used to obtain the average transmit-power and transmit-rate. For a given channel quantizer, i.e., with $\mathcal{R}$ fixed, and the fading pdf assumed known, $\Pr\{\mathbf{J}\}$ can be obtained as $\Pr\{\mathbf{J}\} = \int_{\mathcal{R}(\mathbf{J})} f_{\mathbf{G}}(\mathbf{G})d\mathbf{G}$. Since $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ links $\mathbf{R}$ with $\mathbf{P}$, it suffices to optimize only over one of them. Note also that the binomial $[\mathbf{R}(\mathbf{J})]_{m,k}[\mathbf{W}(\mathbf{J})]_{m,k}$ is not jointly convex with respect to (w.r.t.) $\mathbf{R}(\mathbf{J})$ and $\mathbf{W}(\mathbf{J})$. For this reason, we will instead consider the auxiliary variable $[\tilde{\mathbf{R}}(\mathbf{J})]_{m,k} := [\mathbf{R}(\mathbf{J})]_{m,k}[\mathbf{W}(\mathbf{J})]_{m,k}$ and seek allocation and scheduling matrices solving the following optimization problem:

$$\begin{cases} \min_{\tilde{\mathbf{R}}(\mathbf{J})\geq \mathbf{0}, \mathbf{W}(\mathbf{J})\geq \mathbf{0}} \sum_{\forall \mathbf{J}\in\mathcal{J}} \Big(\sum_{m=1}^{M}[\boldsymbol{\mu}]_m \sum_{k=1}^{K} \\ \qquad \Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}\left(\frac{[\tilde{\mathbf{R}}(\mathbf{J})]_{m,k}}{[\mathbf{W}(\mathbf{J})]_{m,k}}\right)[\mathbf{W}(\mathbf{J})]_{m,k}\Big)\Pr\{\mathbf{J}\} \\ \text{s. to}: \sum_{\forall \mathbf{J}\in\mathcal{J}} \left(\sum_{k=1}^{K}[\tilde{\mathbf{R}}(\mathbf{J})]_{m,k}\right)\Pr\{\mathbf{J}\} \geq [\check{\mathbf{r}}]_m, \quad \forall m \\ \qquad \sum_{m=1}^{M}[\mathbf{W}(\mathbf{J})]_{m,k} \leq 1, \quad \forall k, \forall \mathbf{J}\,. \end{cases} \tag{1}$$

If $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ is a convex function, then problem (1) is convex. Throughout this paper it will be assumed that:

(as) *The power-rate function* $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ *is strictly convex.*

Note that (as) holds generally true for orthogonal access but not when multiuser interference is present. For example, if QoS requirements impose a maximum instantaneous BER $\check{\epsilon}_{\max}$ and symbols are drawn from QAM constellations, then $\epsilon_{\max} = 0.2\exp(-g_{m,k}^{\min}([\mathbf{J}]_{m,k})p_{m,k}/(2^{r_{m,k}}-1))$, where $g_{m,k}^{\min}([\mathbf{J}]_{m,k}) := \min_{g_{m,k}}\{g_{m,k} \in \mathcal{R}([\mathbf{J}]_{m,k})\}$; see [4]. Hence, $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ can be written as $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}(x) = ((2^x-1)\ln(0.2/\check{\epsilon}_{\max}))/g_{m,k}^{\min}([\mathbf{J}]_{m,k})$, which is certainly convex.

Since $\mathcal{R}$ is involved in specifying $\Pr\{\mathbf{J}\}$ and $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$, the choice of $\mathcal{R}$ affects the optimum allocation. Selecting the quantization regions to optimize (1) is thus of interest but goes beyond the scope of this paper. Near-optimal channel quantizers for orthogonal frequency-division multiple access (OFDMA) can be found in [4].

3. OPTIMUM RESOURCE ALLOCATION

In this section, the optimum $\mathbf{W}$, $\mathbf{P}$ and $\mathbf{R}$ matrices will be characterized as a function of $\mathbf{J}$ and the optimum multipliers of the constrained optimization problem in (1). As with other algorithms relying on ensemble fading statistics, the main burden is associated with finding the optimum multipliers – a task carried out off-line. Once those are available, the on-line scheme per fading realization is very simple.

Let $\boldsymbol{\lambda}^R$ denote the $M \times 1$ vector formed by the Lagrange multipliers corresponding to each rate constraint. Applying the Karush-Kuhn-Tucker (KKT) conditions [1] to (1), the following can be proved[2] (recall $\dot{x}$ denotes derivative of x).

Proposition 1 *The optimum rate allocation is given by:*
(i) $[\tilde{\mathbf{R}}^*(\mathbf{J})]_{m,k} = 0$*, if either* $[\mathbf{W}^*(\mathbf{J})]_{m,k} = 0$ *or* $[\boldsymbol{\lambda}^{R*}]_m < \dot{\Upsilon}_{\mathcal{R}([\mathbf{J}]_{m,k})}\left(\frac{[\tilde{\mathbf{R}}^*(\mathbf{J})]_{m,k}}{[\mathbf{W}^*(\mathbf{J})]_{m,k}}\right)$;
(ii) otherwise, the optimum rate allocation is

$$[\tilde{\mathbf{R}}^*(\mathbf{J})]_{m,k} = \dot{\Upsilon}^{-1}_{\mathcal{R}([\mathbf{J}]_{m,k})}\left(\frac{[\boldsymbol{\lambda}^{R*}]_m}{[\boldsymbol{\mu}]_m}\right)[\mathbf{W}^*(\mathbf{J})]_{m,k} \quad (2)$$

where $\dot{\Upsilon}^{-1}_{\mathcal{R}([\mathbf{J}]_{m,k})}$ *denotes the inverse function of* $\dot{\Upsilon}_{\mathcal{R}([\mathbf{J}]_{m,k})}$.

Given the relationship of the auxiliary $\tilde{\mathbf{R}}$ with $\mathbf{R}$, the optimum transmit-rate can be obtained as

$$[\mathbf{R}^*(\mathbf{J})]_{m,k} = \dot{\Upsilon}^{-1}_{\mathcal{R}([\mathbf{J}]_{m,k})}\left(\frac{[\boldsymbol{\lambda}^{R*}]_m}{[\boldsymbol{\mu}]_m}\right). \quad (3)$$

To find the optimum $\mathbf{W}$, define first the functional

$$\begin{aligned}[\mathbf{C}_{\mathbf{W}}(\mathbf{J})]_{m,k} \;&:=\; [\boldsymbol{\mu}]_m \Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}([\mathbf{R}^*(\mathbf{J})]_{m,k}) \\ &-\; [\boldsymbol{\lambda}^{R*}]_m[\mathbf{R}^*(\mathbf{J})]_{m,k} \quad (4)\end{aligned}$$

which represents the cost of scheduling channel k to user m when the Q-CSI is $\mathbf{J}$, i.e., the cost of selecting $[\mathbf{W}(\mathbf{J})]_{m,k} = 1$. (This cost emerges also in the unconstrained Lagrangian of (1), call it $\mathcal{L}$.) With $\wedge$ denoting the "and" operator, define also vector $[\mathbf{c}^*_{\mathbf{W}}(\mathbf{J}, \boldsymbol{\lambda}^{\mathbf{R}}))]_k := \min_m\{[\mathbf{C}_{\mathbf{W}}\ (\mathbf{J}, \boldsymbol{\lambda}^{R}))]_{m,k}\}_{m=1}^{M}$, and the set $\mathcal{M}(\mathbf{J}, k) := \{m : [\mathbf{C}_{\mathbf{W}}(\mathbf{J}, \boldsymbol{\lambda}^{\mathbf{R}}))]_{m,k} = [\mathbf{c}_{\mathbf{W}}{}^*(\mathbf{J}, \boldsymbol{\lambda}^{\mathbf{R}}))]_k \wedge ([\mathbf{c}^*_{\mathbf{W}}(\mathbf{J}, \boldsymbol{\lambda}^{\mathbf{R}}))]_k < 0)\}$. Based on these notational conventions, it can be shown that:

Proposition 2 *The optimum scheduling* $\mathbf{W}^*(\mathbf{J})$ *satisfies:*
(i) If $[\mathbf{W}^*(\mathbf{J})]_{m,k} > 0$*, then* $m \in \mathcal{M}(\mathbf{J}, k)$.
(ii) If $|\mathcal{M}(\mathbf{J}, k)| > 0$*, then* $\sum_{m\in\mathcal{M}(\mathbf{J},k)}[\mathbf{W}^*(\mathbf{J})]_{m,k} = 1$.
(iii) If $|\mathcal{M}(\mathbf{J}, k)| = 0$*, then* $[\mathbf{W}^*(\mathbf{J})]_{m,k} = 0\ \forall m$.

In words, the optimal scheduler assigns the channel only to user(s) with minimum negative cost, in most cases to a single user. This can be viewed as a greedy policy because only one user with minimum cost is selected to transmit per Q-CSI realization, while others defer. Note that with P-CSIR, the optimum scheduling over orthogonal channels is also greedy, whether based on P-CSIT [5] or Q-CSIT [4]).
Case 1 (Single winner): When the minimum cost is attained by only one user, $\mathbf{W}^*$ in Proposition 2 can be written using the indicator function, as

$$[\mathbf{W}^*(\mathbf{J})]_{m,k} = \mathbf{I}_{\{m\in\mathcal{M}(\mathbf{J},k)\}}\ . \quad (5)$$

Since $[\mathbf{C}_{\mathbf{W}}(\mathbf{J})]_{m,k}$ is a function of the channel quantizer, fading realization, priority weight and Lagrange multiplier, for most CSI realizations the costs corresponding to different users m are distinct and the emerging winner is unique.
Case 2 (Multiple winners): The event of having different users attaining the minimum cost will be henceforth referred to as a "tie". The main difficulty with a tie is that Proposition 2-*(iii)* does not specify how the channel should be split among winner users because any arbitrary allocation minimizes $\mathcal{L}$. On the other hand, only a subset (for most realizations one) of them is the actual solution to the original primal problem. To find the optimum schedule in this case, let first define the matrix of single-winner access as $[\mathbf{W}_{one}(\mathbf{J})]_{m,k} := [\mathbf{W}(\mathbf{J})^*]_{m,k}$ in (5) for all $(\mathbf{J}, k)$ so that $|\mathcal{M}(\mathbf{J}, k)| = 1$, and $[\mathbf{W}_{one}(\mathbf{J})]_{m,k} := 0$ otherwise; the matrix of multiple-winner access as $[\mathbf{W}_{tie}(\mathbf{J})]_{m,k} = 0$ if $|\mathcal{M}(\mathbf{J}, k)| \leq 1$ or if $|\mathcal{M}(\mathbf{J}, k)| > 1$ but $m \notin |\mathcal{M}(\mathbf{J}, k)|$, and $[\mathbf{W}_{tie}(\mathbf{J})]_{m,k} \in [0, 1]$ otherwise; the set of multiple-winner scheduling matrices as $\mathcal{W}_{tie} := \{\mathbf{W}_{tie}(\mathbf{J}) \mid \forall\mathbf{J}$; the average single-winner transmit-rate vector as $[\bar{\mathbf{r}}_{one}]_m := \sum_{\forall\mathbf{J}}\left(\sum_{k=1}^{K}[\mathbf{R}^*(\mathbf{J})]_{m,k}[\mathbf{W}_{one}(\mathbf{J})]_{m,k}\right)\Pr\{\mathbf{J}\}$; and $\check{\mathbf{r}}_{tie} := \check{\mathbf{r}} - \bar{\mathbf{r}}_{one}$. Using these definitions, the optimum schedule $\mathbf{W}_{tie}(\mathbf{J})$ for all $(\mathbf{J}, k)$ so that $|\mathcal{M}(\mathbf{J}, k)| > 1$, can be found as the solution of the following linear program:

$$\begin{cases}\min_{\mathbf{W}_{tie}(\mathbf{J})\in\mathcal{W}_{tie}}\ \sum_{\forall\mathbf{J}}\left(\sum_{k=1}^{K}\sum_{m=1}^{M}[\boldsymbol{\mu}]_m\right. \\ \qquad\left.\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}\left([\mathbf{R}^*(\mathbf{J})]_{m,k}\right)[\mathbf{W}_{tie}(\mathbf{J})]_{m,k}\right)\Pr\{\mathbf{J}\} \\ \text{s. to}: \sum_{\forall\mathbf{J}}\left(\sum_{k=1}^{K}[\mathbf{R}^*(\mathbf{J})]_{m,k}[\mathbf{W}_{tie}(\mathbf{J})]_{m,k}\right)\Pr\{\mathbf{J}\} \\ \qquad\qquad = [\check{\mathbf{r}}_{tie}]_m,\ \ \forall m, \\ \qquad\sum_{m=1}^{M}[\mathbf{W}_{tie}(\mathbf{J})]_{m,k} = 1,\ \forall(\mathbf{J}, k):\ |\mathcal{M}(\mathbf{J}, k)| > 1.\end{cases} \quad (6)$$

Note that in the optimization process, only the $\mathbf{J}$ for which a tie occurs are considered and for those only the non-zero entries of $\mathbf{W}_{tie}(\mathbf{J})$ are optimized.

Among all schedules minimizing the Lagrangian when a tie occurs (second constraint), the optimal one for the primal problem is the one for which the average rate constraints are satisfied with equality. It is worth noticing that although (6) applies in general to any scheduling scheme for orthogonal multi-access ystems, neither [3], [5] (P-CSIR and P-CSIT) nor [4] (P-CSIR and Q-CSIT) consider (6). This is because based on P-CSIR, the set of fading realizations $\mathbf{G}$ for which a tie occurs has Lebesgue measure zero. Therefore, any arbitrary channel scheduling among tied users is equally optimum. Indeed, the contribution of any specific $\mathbf{G}$ to the average performance when integrated over the channel pdf is zero. But when dealing with Q-CSI, neither the probability of a Q-CSI realization $\mathbf{J}$ nor the contribution to the average cost are negligible. And this precisely necessitates solving (6) to obtain the optimum schedule. Intuitively, as the number of regions and channels increases sharing a channel becomes less likely, which in turn brings the solution closer to the P-CSI case and the effect of neglecting (6) becomes less harmful.

4. OPTIMAL LAGRANGE MULTIPLIERS

The optimum scheduling and rate allocation scheme of the previous section requires the calculation of the optimum multiplier vector $\boldsymbol{\lambda}^{R*}$. Since the rate constraints in (1) are al-

[2]Proofs of all propositions are omitted due to space limitation s.

ways active, the KKT conditions imply that when $\boldsymbol{\lambda}^R = \boldsymbol{\lambda}^{R*}$ those constraints are satisfied as equality. As $\boldsymbol{\lambda}^{R*}$ cannot be obtained analytically from this condition, numerical search is needed to find $\boldsymbol{\lambda}^{R*}$. This is possible through the dual function [cf. (1)][3]

$$D(\boldsymbol{\lambda}^R) := \inf_{\tilde{\mathbf{R}}(\mathbf{J})\geq\mathbf{0},\mathbf{W}(\mathbf{J})\geq\mathbf{0}} \mathcal{L}(\boldsymbol{\lambda}^R, \tilde{\mathbf{R}}(\mathbf{J}), \mathbf{W}(\mathbf{J}))$$
$$= \mathcal{L}(\boldsymbol{\lambda}^R, \mathbf{R}^*(\mathbf{J},\boldsymbol{\lambda}^R)\cdot\mathbf{W}^*(\mathbf{J},\boldsymbol{\lambda}^R), \mathbf{W}^*(\mathbf{J},\boldsymbol{\lambda}^R)) \quad (7)$$

which is concave. Based on (7), the dual problem of (1) is

$$\max_{\boldsymbol{\lambda}^R\geq\mathbf{0}} D(\boldsymbol{\lambda}^R). \quad (8)$$

Thanks to convexity, the duality gap is zero and the value of $\boldsymbol{\lambda}^R$ optimizing (8) can be used to find the optimum primal solution. A major challenge in obtaining $\boldsymbol{\lambda}^{R*}$ using subgradient iterations (gradient iterations are impossible because $D(\boldsymbol{\lambda}^R)$ is non-differentiable w.r.t. $[\boldsymbol{\lambda}^R]_m$) is that the mth entry of the subgradient vector $[\partial D(\boldsymbol{\lambda}^R)]_m := [\check{\mathbf{r}}]_m - \sum_{\forall\mathbf{J}}\sum_{\forall k}[\mathbf{R}^*(\mathbf{J}, \boldsymbol{\lambda}^R)]_{m,k}\,[\mathbf{W}^*(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k}\Pr\{\mathbf{J}\}$ is not Lipschitz continuous because $\mathbf{W}^*(\mathbf{J},\boldsymbol{\lambda}^R)$ in (7) is discontinuous for some $\boldsymbol{\lambda}^R$. Our approach to repairing Lipschitz continuity is to smooth the scheduling function. Smoothing ensures continuity or differentiability and has been successfully applied to different optimization problems; see e.g., [7].

Since scheduling discontinuities appear in the transition from a tie to a single-user, the idea is to relax the condition for scheduling the kth channel only when $m \in \mathcal{M}(\mathbf{J},k)$. This is possible through the set $\mathcal{M}^s(\mathbf{J},k) := \{m : ([\mathbf{C}_\mathbf{W}(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k} - [\mathbf{c}_\mathbf{W}{}^*(\mathbf{J},\boldsymbol{\lambda}^R)]_k < \varepsilon) \wedge ([\mathbf{c}_\mathbf{W}{}^*(\mathbf{J},\boldsymbol{\lambda}^R)]_k < 0)\}$, where ε is a small positive number. Based on $\mathcal{M}^s(\mathbf{J},k)$, consider the following suboptimal but smooth scheduling matrix

$$[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k} := \mathbf{I}_{\{m\in\mathcal{M}^s(\mathbf{J},k)\}} \quad (9)$$
$$\times \frac{\left(1-\frac{[\mathbf{C}_\mathbf{W}(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k}-[\mathbf{c}_\mathbf{W}{}^*(\mathbf{J},\boldsymbol{\lambda}^R)]_k}{\varepsilon}\right)^2}{\sum_{m\in\mathcal{M}^s(\mathbf{J},k)}\left(1-\frac{[\mathbf{C}_\mathbf{W}(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k}-[\mathbf{c}_\mathbf{W}{}^*(\mathbf{J},\boldsymbol{\lambda}^R)]_k}{\varepsilon}\right)^2}$$

which is inspired by the quadratic ε-smoothing of [7]. Note that $[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k}$ schedules channel k to users m whose cost is not minimum but ε-close to the minimum.

Proposition 3 *The smooth scheduler $\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)$ satisfies:*
(i) If $[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k} > 0$, then $m \in \mathcal{M}^s(\mathbf{J},k)$ and $[\mathbf{C}_\mathbf{W}(\mathbf{J}, \boldsymbol{\lambda}^R)]_{m,k} < [\mathbf{c}_\mathbf{W}{}^*(\mathbf{J},\boldsymbol{\lambda}^R)]_k + \varepsilon$.
(ii) If $|\mathcal{M}^s(\mathbf{J},k)| > 0$, $\sum_{m\in\mathcal{M}^s(\mathbf{J},k)}[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k} = 1$.
(iii) If $|\mathcal{M}(\mathbf{J},k)| = 0$, then $[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k} = 0\ \forall m$.
(iv) $[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k}$ is a continuous function of $\boldsymbol{\lambda}^R$.

Properties *(i)-(iii)* of $\mathbf{W}^s$ are similar to those of $\mathbf{W}^*$ stated in Proposition 2, while *(iv)* ensures continuity. Based on Proposition 3, the following result can be established.

Proposition 4 *If $D^s(\boldsymbol{\lambda}^R) := \mathcal{L}(\boldsymbol{\lambda}^R, \mathbf{R}^*(\mathbf{J},\boldsymbol{\lambda}^R)\cdot\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R), \mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R))$ and $[\partial^s D(\boldsymbol{\lambda}^R)]_m := [\check{\mathbf{r}}]_m - \sum_{\forall\mathbf{J}}\sum_{\forall k}[\mathbf{R}^*(\mathbf{J}, \boldsymbol{\lambda}^R)]_{m,k}\,[\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R)]_{m,k}\Pr\{\mathbf{J}\}$ denote smooth versions of the dual function and its subgradient, then:*
(i) For all $\boldsymbol{\lambda}^R$, it holds that $D^s(\boldsymbol{\lambda}^R) < D(\boldsymbol{\lambda}^R) + K\varepsilon$; and
(ii) $[\partial^s D(\boldsymbol{\lambda}^R)]_m$ is Lipschitz continuous.

Proposition 4 guarantees that $\partial D^s(\boldsymbol{\lambda}^R)$ is a Lipschitz continuous $(K\varepsilon)$-subgradient of $D(\boldsymbol{\lambda}^R)$ [1, Sec. 6.3.2].

Proposition 5 *(i) For a sufficiently small stepsize β, the iteration $\boldsymbol{\lambda}^{R(i)} = \boldsymbol{\lambda}^{R(i-1)} + \beta\partial^s D(\boldsymbol{\lambda}^{R(i-1)})$ converges, i.e., $\boldsymbol{\lambda}^{R(i)} \to \boldsymbol{\lambda}^{Rs}$; and*
(ii) The dual function satisfies $|D(\boldsymbol{\lambda}^{Rs})$-$D(\boldsymbol{\lambda}^{R*})| < 2K\varepsilon$.

Proposition 5 implies that the $2K\varepsilon$-optimal value of $\boldsymbol{\lambda}^R$ can be found by implementing the following algorithm[4]:

Algorithm 1 ***(S1.0)*** *Initialization: set vectors $\boldsymbol{\delta_1}$, $\boldsymbol{\delta_2}$ to small positive values; $\boldsymbol{\lambda}^{R(0)} = \boldsymbol{\delta_1}$, and the iteration index $i = 1$.*
(S1.1) *Resource allocation update: per Q-CSI realization $\mathbf{J}$, use $\boldsymbol{\lambda}^{R(i-1)}$ to obtain $\mathbf{R}(\mathbf{J})^{(i)}$ and $\mathbf{P}(\mathbf{J})^{(i)}$ based on (3) and $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$; and $\mathbf{W}^s(\mathbf{J})^{(i)}$ using (9).*
(S1.2) *Dual update: use (S1.1) to find $\partial^s D(\boldsymbol{\lambda}^{R(i-1)})$. Stop if $|\partial^s D(\boldsymbol{\lambda}^{R(i-1)})| < \boldsymbol{\delta_2}$; update $\boldsymbol{\lambda}^{R(i)}$ as in Proposition 5, and set $i = i+1$; otherwise, go to (S1.1).*

Due to the average formulation in (1), Algorithm 1 entails computing the average rate and power which require the channel pdf. It must be run off line only when channel statistics or the users' QoS requirements change. Once $\boldsymbol{\lambda}^R$ is known, the (ε-) optimum allocation per $\mathbf{J}$ is found online using (3), $\Upsilon_{\mathcal{R}([\mathbf{J}]_{m,k})}$ and (9).

5. NUMERICAL EXAMPLES

To test the algorithms developed, we simulated uncorrelated complex Gaussian fading channels per user and quantized uniformly each channel gain $g_{m,k}$ to 5 regions. Symbols were drawn from QAM constellations and the corresponding BER was approximated as $0.2\exp(-g_{m,k}p_{m,k}/(2^{r_{m,k}}-1))$.

Test Case 1 (Convergence comparison): A time-division multiple access (TDMA) system was simulated with $K = 32$ channels to serve $M = 4$ users with minimum rate requirements $\check{\mathbf{r}} = [10.5, 18, 25, 31]$ while ensuring instantaneous BER not exceeding 10^{-3} with SNR=6dB at the access point. Figure 1 depicts average individual rates versus iterations with stepsize $\beta = 10^{-3}$ for: (i) the optimal channel allocation developed in Section 3; and (ii) the allocation based on the smooth scheduling policy developed in Section 4. The trajectories confirm that while optimal scheduling does not always

[3] Throughout this section, dependence on $\boldsymbol{\lambda}^R$ will be made explicit wherever it contributes to clarity.

[4] In practice, the gap w.r.t. $D(\boldsymbol{\lambda}^{R*})$ is much smaller than 2ε. This holds because $\mathbf{W}^s(\mathbf{J},\boldsymbol{\lambda}^R) \neq \mathbf{W}^*(\mathbf{J},\boldsymbol{\lambda}^R)$ only if $|\mathcal{M}^s(\mathbf{J},k)| > 1$, which is a rare event; hence, on average, the bound in Proposition 4-*(i)* is very loose.

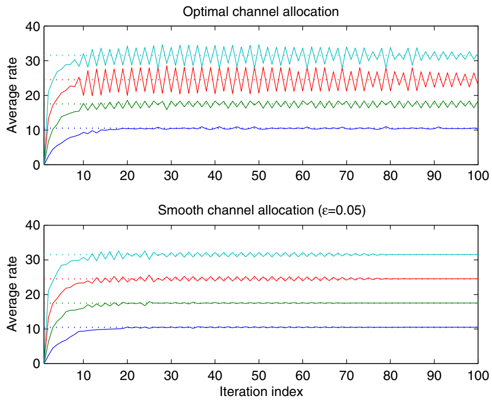

Fig. 1. Average transmit-rates in bits per channel use. Optimal non-smooth policy (top) and smooth policy (bottom).

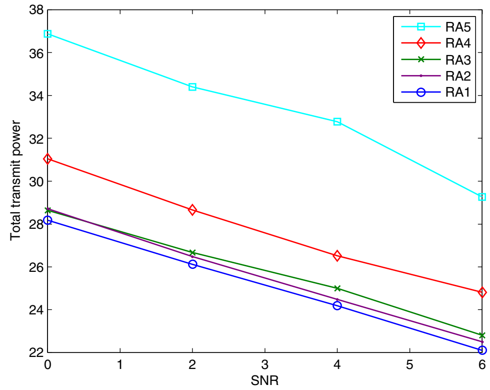

Fig. 2. Comparison of various resource allocation schemes on the basis of average transmit-power [dB].

satisfy the constraints and rate allocation hovers around its optimum, the smooth policy converges in a finite number of iterations with negligible loss in performance. (Behavior of transmit-powers is similar to that for transmit-rates.)

Test Case 2 (Performance comparison): An OFDMA system was simulated here with $K = 32$ subcarriers to serve $M = 3$ users with $\check{\mathbf{r}} = [30, 30, 30]$ and average BER=10^{-3} per user channel generated as before to have eight exponentially decaying gains. Figure 2 compares the *overall* average transmit-power for different SNR values. Results for five different resource allocation (RA) policies are depicted: (i) the benchmark allocation obtained when P-CSI is available (RA1); (ii) the optimum Q-CSIT based policy (with P-CSIR) of [4] (RA2); (iii) the smooth policy developed with the optimum channel quantizer of [4] (RA3); (iv) this paper's smooth policy with a random quantizer (RA4); and (v) a policy based on Q-CSI which adapts $\mathbf{R}$ but fixes the channel scheduling matrix $\mathbf{W}$. Interestingly, if the quantizer is optimum the novel scheme (RA3) performs very close to the optimum P-CSIT and Q-CSIT one, while the penalty of using suboptimum quantization in (RA4) is about 3dB. Finally, it is worth stressing the significant power savings of RA3 and RA4 relative to a suboptimum/heuristic scheme (RA5).

6. CONCLUDING SUMMARY

This paper developed optimal scheduling and resource allocation policies for orthogonal multi-access transmissions over fading channels when both transmitter(s) and receiver(s) have to rely only on quantized CSI. The differences relative to their counterparts based on perfect CSI at the receiver(s), show up in channel scheduling. For most channel realizations the optimum scheduling amounts to a single user (winner) accessing the channel, while for a smaller set of realizations a few users share the resources. Optimal allocation in the sharing case is obtained as the solution of a linear program. The resultant online iterations hover around the optimum allocation and solving the linear program increases complexity relative to the single winner-takes-all case. Albeit suboptimum, a smooth scheduler was found to mitigate these challenges at reduced complexity and asymptotically optimal performance.

7. REFERENCES

[1] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[2] A. Lapidoth and S. Shamai, "Fading Channels: How Perfect Need 'Perfect Side Information' Be?," *IEEE Trans. on Info. Theory*, pp. 1118-1134, May 2002.

[3] L. Li and A. J. Goldsmith, "Capacity and Optimal Resource Allocation for Fading Broadcast Channels–Part I: Ergodic capacity," *IEEE Trans. on Info. Theory*, vol. 47, no.3, pp. 1083-1102, Mar. 2001.

[4] A. G. Marques, G. B. Giannakis, F. F. Digham, and F. J. Ramos, "Power-Efficient Wireless OFDMA using Limited-Rate Feedback," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 2, pp. 685-696, Feb. 2008.

[5] X. Wang and G. B. Giannakis, "Power-Efficient Resource Allocation in Time-Division Multiple Access over Fading Channels," *IEEE Trans. on Info. Theory*, vol. 54, Mar. 2008.

[6] C.Y. Wong, R.S. Cheng, K.B. Lataief, R.D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation,"*IEEE J. on Selected Ar. in Commun.*, vol. 17, no. 10, pp.1747–1758, Oct. 1999.

[7] S. A. Zenios, M. C. Pinar, and R. S. Dembo, "A Smooth Penalty Function Algorithm for Network Structured Problems", *European J. of Operational Research*, vol. 83, pp. 220–236, May 1995.