On Resource Allocation in Fading Multiple Access Channels – An Efficient Approximate Projection Approach

Ali ParandehGheibi, Atilla Eryilmaz, Asuman Ozdaglar, and Muriel Médard

Abstract

We consider the problem of rate and power allocation in a multiple-access channel. Our objective is to obtain rate and power allocation policies that maximize a general concave utility function of average transmission rates on the information theoretic capacity region of the multiple-access channel. Our policies does not require queue-length information. We consider several different scenarios. First, we address the utility maximization problem in a non-fading channel to obtain the optimal operating rates, and present an iterative gradient projection algorithm that uses approximate projection. By exploiting the polymatroid structure of the capacity region, we show that the approximate projection can be implemented in time polynomial in the number of users. Second, we consider resource allocation in a fading channel. Optimal rate and power allocation policies are presented for the case that power control is possible and channel statistics are available. For the case that transmission power is fixed and channel statistics are unknown, we propose a greedy rate allocation policy and provide bounds on the performance difference of this policy and the optimal policy in terms of channel variations and structure of the utility function. We present numerical results that demonstrate superior convergence rate performance for the greedy policy compared to queue-length based policies. In order to reduce the computational complexity of the greedy policy, we present approximate rate allocation policies which track the greedy policy within a certain neighborhood that is characterized in terms of the speed of fading.

Index Terms

Multiple access, resource allocation, power control, utility maximization, fading channel, rate splitting.

I. INTRODUCTION

Dynamic allocation of communication resources such as bandwidth or transmission power is a central issue in multiple access channels in view of the time varying nature of the channel and the interference effects. Most of the existing literature focuses on specific communication schemes such as TDMA (time-division multiple access) [1], CDMA (code-division multiple access) [2], [3], and OFDM (Orthogonal Frequency Division Multiplexing) [4] systems. An exception is the work by Tse *et al.* [5], which consider the notion of *throughput capacity* for the fading channel with Channel State Information (CSI). The throughput capacity is the notion of Shannon capacity applied to the fading channel, where the codeword length can be arbitrarily long to average over the fading of the channel.

A. ParandehGheibi is with the Laboratory for Information and Decision Systems, Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge MA, 02139 (e-mail: parandeh@mit.edu)

A. Eryilmaz is with the Electrical and Computer Engineering, Ohio State University, OH, 43210 (e-mail: eryilmaz@ece.osu.edu)

A. Ozdaglar and M. Médard are with the Laboratory for Information and Decision Systems, Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge MA, 02139 (e-mails: asuman@mit.edu, medard@mit.edu)

Tse *et al.* [5] consider allocation of rate and power to maximize a linear utility function of the transmission rates over the throughput region, which characterizes the points on the boundary of the throughput capacity region.

In this paper, we consider the problem of rate and power allocation in a multiple access channel with perfect CSI. Contrary to the linear case in [5], we consider maximizing a general utility function of transmission rates over the throughput capacity region. Such a general concave utility function allows us to capture different performance metrics such as fairness or delay (cf. Shenker [6], Srikant [7]). Our contributions can be summarized as follows.

We first consider a non-fading multiple-access channel where we introduce a gradient projection algorithm for the problem of maximizing a concave utility function of transmission rates over the capacity region. We establish the convergence of the method to the optimal rate allocation. Since the capacity region of the multiple-access channel is described by a number of constraints exponential in the number of users, the projection operation used in the method can be computationally expensive. To reduce the computational complexity, we introduce a new method that utilizes *approximate projections*. By exploiting the polymatroid structure of the capacity region, we show that the approximate projection operation can be implemented in time polynomial in number of users by using submodular function minimization algorithms. Moreover, we present a more efficient algorithm for the approximate projection problem which relies on rate-splitting [8]. This algorithm also provides the extra information that allows the receiver to decode the message by successive cancelation.

Second, we consider a fading multiple access channel and study the case where channel statistics are known and transmission power can be controlled at the transmitters. Owing to strict convexity properties of the capacity region along the boundary, we show that the resource allocation problem for a general concave utility is equivalent to another problem with a linear utility. Hence, the *optimal* resource allocation policies are obtained by applying the results in [5] for the linear utility. Given a general utility function, the conditional gradient method is used to obtain the corresponding linear utility.

If the transmitters do not have the power control feature and channel statistics are not known, the throughput capacity region is a polyhedron and the strictly convexity properties of the region do not hold any more. Hence, the previous approach is not applicable. In this case, we consider a greedy policy, which maximizes the utility function for any given channel state. This policy is suboptimal, however, we can bound the performance difference between the optimal and the greedy policies. We show that this bound is tight in the sense that it goes to zero either as the utility function tends to a linear function of the rates or as the channel variations vanish.

The greedy policy requires exact solution of a nonlinear program in each time slot, which makes it computationally intractable. To alleviate this problem, we present approximate rate allocation policies based on the gradient projection method with approximate projection and study its tracking capabilities when the channel conditions vary over time. In our algorithm, the solution is updated in every time slot in a direction to increase the utility function at that time slot. But, since the channel may vary between time-slots, the level of these temporal channel variations become critical to the performance. We explicitly quantify the impact of the speed of fading on the performance of the policy, both for the worst-case and the average speed of fading. Our results also capture the effect of the degree of concavity of the utility functions on the average performance.

An important literature relevant to our work appears in the context of cross-layer design, where joint scheduling-routing-flow control algorithms have been proposed and shown to achieve utility maximization for concave utility functions while guaranteeing network stability (e.g. [9], [10], [11], [12]). The common idea behind these schemes is to use properly maintained queues to make dynamic decisions about new packet generation as well as rate

allocation.

Some of these works ([10], [11]) explicitly address the fading channel conditions, and show that the associated policies can achieve rates arbitrarily close to the optimal based on a design parameter choice. However, the rate allocation with these schemes requires that a large optimization problem requiring global queue-length information be solved over a complex feasible set in every time slot. Clearly, this may not always be possible owing to the limitations of the available information, the processing power, or the complexity intrinsic to the feasible set. Requirement for queue-length information may impose much more overhead on the system than channel state information. On the other hand, even in the absence of fading, the interference constraints among nearby nodes' transmissions may make the feasible set so complex that the optimal rate allocation problem becomes NP-hard (see [13]). Moreover, the convergence results of queue-length based policies ([10], [11]) are asymptotic, and our simulation results show that such policies may suffer from poor convergence rate. In fact, duration of a communication session may not be sufficient for these algorithms to approach the optimal solution while suboptimal policies such as the greedy policy seems to have superior performance when communication time is limited, even though the greedy policy does not use queue-length information.

In the absence of fading, several works have proposed and analyzed approximate randomized and/or distributed rate allocation algorithms for various interference models to reduce the computational of the centralized optimization problem of the rate allocation policy ([14], [9], [15], [13], [16], [17]). The effect of these algorithms on the utility achieved is investigated in [13], [18]. However, no similar work exists for fading channel conditions, where the changes in the fading conditions coupled with the inability to solve the optimization problem instantaneously make the solution much more challenging.

Other than the papers cited above, our work is also related to the work of Vishwanath *et al.* [19] which builds on [5] and takes a similar approach to the resource allocation problem for linear utility functions. Other works address different criteria for resource allocation including minimizing delay by a queue-length based approach [20], minimizing the weighted sum of transmission powers [21], and considering Quality of Service (QoS) constraints [22]. In contrast to this literature, we consider the utility maximization framework for general concave utility functions.

The remainder of this paper is organized as follows: In Section II, we introduce the model and describe the capacity region of a fading multiple-access channel. In Section III, we consider the utility maximization problem in a non-fading channel and present the gradient projection method with approximate projection. In Section IV, we address the resource allocation problem with power control and known channel statistics. In Section V, we consider the same problem without power control and knowledge of channel statistics. We present the greedy policy and approximate rate allocation policies and study their tracking behavior. Section VI provides the simulation results, and we give our concluding remarks in Section VII.

Regarding the notation, we denote by x_i the i-th component of a vector $\boldsymbol{x}$. We denote the nonnegative orthant by $\mathbb{R}_+^n$, i.e., $\mathbb{R}_+^n = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \geq 0\}$. We write $\boldsymbol{x}'$ to denote the transpose of a vector $\boldsymbol{x}$. We use the notation $\mathbf{Pr}(\cdot)$ for the probability of an event in the Borel σ-algebra on $\mathbb{R}^n$. The exact projection operation on a closed convex set is denoted by $\mathcal{P}$, i.e., for any closed convex set $X \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$, we have $\mathcal{P}(\boldsymbol{x}) = \operatorname{argmin}_{\boldsymbol{y} \in X} \|\boldsymbol{x} - \boldsymbol{y}\|$, where $\|\cdot\|$ denotes the Euclidean norm.

II. SYSTEM MODEL

We consider M transmitters sharing the same media to communicate to a single receiver. We model the channel as a Gaussian multiple access channel with flat fading effects,

$$Y(n) = \sum_{i=1}^{M} \sqrt{H_i(n)} X_i(n) + Z(n), \tag{1}$$

where $X_i(n)$ and $H_i(n)$ are the transmitted waveform and the fading process of the i-th transmitter, respectively, and $Z(n)$ is properly bandlimited Gaussian noise with variance N_0. We assume that the fading processes of all transmitters are jointly stationary and ergodic, and the stationary distribution of the fading process has continuous density. We assume that all the transmitters and the receiver have instant access to channel state information. In practice, the receiver measures the channels and feeds back the channel information to the transmitters. The implicit assumption in this model is that the channel variations are much slower than the data rate, so that the channel can be measured accurately at the receiver and the amount of feedback bits is negligible compared to that of transmitting information.

Definition 1: The *temporal variation* in fading is modeled as follows:

$$|H_i(n+1) - H_i(n)| = V_n^i, \quad \text{for all } n, \; i = 1, \ldots, M, \tag{2}$$

where the V_n^is are nonnegative random variables independent across time slots for each i. We assume that for each i, the random variables V_n^i are uniformly bounded from above by $\hat{v}^i$, which we refer to as the *maximum speed of fading*. Under slow fading conditions, the distribution of V_n^i is expected to be more concentrated around zero.

Consider the non-fading case where the channel state vector is fixed. The capacity region of the Gaussian multiple-access channel with no power control is described as follows [23],

$$C_g(\boldsymbol{P}, \boldsymbol{H}) = \left\{ \boldsymbol{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq C\left(\sum_{i \in S} H_i P_i, N_0 \right), \text{for all } S \subseteq \mathcal{M} = \{1, \ldots, M\} \right\}, \tag{3}$$

where P_i and R_i are the i-th transmitter's power and rate, respectively. $C(P, N)$ denotes Shannon's formula for the capacity of the AWGN channel given by

$$C(P, N) = \frac{1}{2} \log\left(1 + \frac{P}{N}\right) \quad \text{nats.} \tag{4}$$

For a multiple-access channel with fading, but fixed transmission powers P_i, the *throughput* capacity region is given by averaging the instantaneous capacity regions with respect to the fading process [24],

$$C_a(\boldsymbol{P}) = \left\{ \boldsymbol{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq \mathbb{E}_{\boldsymbol{H}}\left[C\left(\sum_{i \in S} H_i P_i, N_0 \right) \right], \text{for all } S \subseteq \mathcal{M} \right\}, \tag{5}$$

where $\boldsymbol{H}$ is a random vector with the stationary distribution of the fading process.

A power control policy $\boldsymbol{\pi}$ is a function that maps any given fading state $\boldsymbol{h}$ to the powers allocated to the transmitters $\boldsymbol{\pi}(\boldsymbol{h}) = (\boldsymbol{\pi}_1(\boldsymbol{h}), \ldots, \boldsymbol{\pi}_M(\boldsymbol{h}))$. Similarly, we can define the rate allocation policy, $\mathcal{R}$, as a function that maps the fading state $\boldsymbol{h}$ to the transmission rates, $\mathcal{R}(\boldsymbol{h})$. For any given power-control policy $\boldsymbol{\pi}$, the capacity region follows from (5) as

$$C_f(\boldsymbol{\pi}) = \left\{ \boldsymbol{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq \mathbb{E}_{\boldsymbol{H}}\left[C\left(\sum_{i \in S} H_i \boldsymbol{\pi}_i(\boldsymbol{H}), N_0 \right) \right], \text{for all } S \subseteq \mathcal{M} \right\}. \tag{6}$$

Tse *et al.* [5] have shown that the throughput capacity of a multiple access fading channel is given by

$$C(\bar{\boldsymbol{P}}) = \bigcup_{\boldsymbol{\pi} \in \mathcal{G}} C_f(\boldsymbol{\pi}), \tag{7}$$

where $\mathcal{G} = \{\boldsymbol{\pi} : \mathbb{E}_{\boldsymbol{H}}[\pi_i(\boldsymbol{H})] \leq \bar{P}_i, \text{for all } i\}$ is the set of all power control policies satisfying the average power constraint. Let us define the notion of boundary or dominant face for any of the capacity regions defined above.

Definition 2: The *dominant face* or *boundary* of a capacity region, denoted by $\mathcal{F}(\cdot)$, is defined as the set of all M-tuples in the capacity region such that no component can be increased without decreasing others while remaining in the capacity region.

III. RATE ALLOCATION IN A NON-FADING CHANNEL

In this section, we address the problem of finding the optimal operation rates in a non-fading multiple-access channel. Without loss of generality, we fix the channel state vector to unity throughout this section, and denote the capacity region by a simpler notation $C_g(\boldsymbol{P})$ instead of $C_g(\boldsymbol{P}, \mathbf{1})$, where $\boldsymbol{P} > 0$ denotes the transmission power. Consider the following utility maximization problem for a M-user channel.

$$\begin{aligned} \text{maximize} \quad & u(\boldsymbol{R}) \\ \text{subject to} \quad & \boldsymbol{R} \in C_g(\boldsymbol{P}), \end{aligned} \tag{8}$$

where R_i and P_i are i-th user rate and power, respectively. The utility function $u(\boldsymbol{R})$ is assumed to satisfy the following conditions.

Assumption 1: The following conditions hold:

(a) The utility function $u : \mathbb{R}_+^M \to \mathbb{R}$ is concave with respect to vector $\boldsymbol{R}$.

(b) $u(\boldsymbol{R})$ is monotonically increasing with respect to R_i, for $i = 1, \ldots, M$.

Assumption 2: There exists a scalar B such that

$$\|\boldsymbol{g}\| \leq B, \quad \text{for all } \boldsymbol{g} \in \partial u(\boldsymbol{R}) \text{ and all } \boldsymbol{R},$$

where $\partial u(\boldsymbol{R})$ denotes the subdifferential of u at $\boldsymbol{R}$, i.e., the set of all subgradients [1] of u at $\boldsymbol{R}$.

Note that Assumption 2 is standard in the analysis of subgradient methods for non-differentiable optimization problems [25]. The maximization problem in (8) is a convex program and the optimal solution can be obtained by several optimization methods such as the gradient projection method. The gradient projection method with exact projection is typically used for problems where the projection operation is simple, i.e., for problems with simple constraint sets such as the non-negative orthant or a simplex. However, the constraint set in (8) is defined by exponentially many constraints, making the projection problem computationally intractable. To alleviate this problem, we use an approximate projection, which is obtained by successively projecting on violated constraints.

Definition 3: Let $X = \{\boldsymbol{x} \in \mathbb{R}^n | A\boldsymbol{x} \leq \boldsymbol{b}\}$, where A has non-negative entries. Let $\boldsymbol{y} \in \mathbb{R}^n$ violate the constraint $\boldsymbol{a}_i' \boldsymbol{x} \leq b_i$, for $i \in \{i_1, \ldots, i_l\}$. The approximate projection of $\boldsymbol{y}$ on X, denoted by $\tilde{\mathcal{P}}$, is given by

$$\tilde{\mathcal{P}}(\boldsymbol{y}) = \mathcal{P}_{i_1}(\ldots (\mathcal{P}_{i_{l-1}}(\mathcal{P}_{i_l}(\boldsymbol{y})))),$$

where $\mathcal{P}_{i_k}$ denotes the exact projection on the hyperplane $\{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{a}_{i_k}' \boldsymbol{x} = b_{i_k}\}$.

---

[1]The vector $\boldsymbol{g}$ is a subgradient of a concave function $f : D \to \mathbb{R}$ at $\boldsymbol{x}_0$, if and only if $f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \leq \boldsymbol{g}'(\boldsymbol{x} - \boldsymbol{x}_0)$ for all $\boldsymbol{x} \in D$.
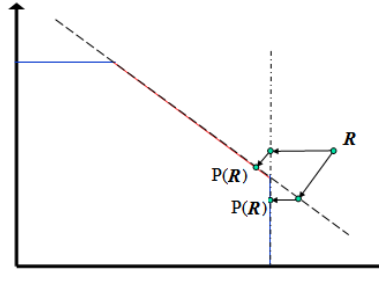
Fig. 1.   Approximate projection of $\boldsymbol{R}$ on a two-user MAC capacity region

An example of approximate projection on a two-user multiple-access capacity region is illustrated in Figure 1. As shown in the figure, the result of approximate projection is not necessarily unique. In the following, when we write $\tilde{\mathcal{P}}$, it refers to an approximate projection for an arbitrary order of projections on the violated hyperplanes. Although the approximate projection is not unique, it is pseudo-nonexpansive as claimed in the following Lemma.

Lemma 1: The approximate projection $\tilde{\mathcal{P}}$ given by Definition 3 has the following properties:

(i) For any $\boldsymbol{y} \in \mathbb{R}^n$, $\tilde{\mathcal{P}}(\boldsymbol{y})$ is feasible with respect to set X, i.e., $\tilde{\mathcal{P}}(\boldsymbol{y}) \in X$.

(ii) $\tilde{\mathcal{P}}$ is pseudo-nonexpansive, i.e.,

$$\|\tilde{\mathcal{P}}(\boldsymbol{y}) - \tilde{\boldsymbol{y}}\| \leq \|\boldsymbol{y} - \tilde{\boldsymbol{y}}\|, \quad \text{for all } \tilde{\boldsymbol{y}} \in X. \tag{9}$$

Proof: For part (i), it is straightforward to see that $\mathcal{P}_i(\boldsymbol{y})$ is given by (cf. [26] Sec. 2.1.1)

$$\mathcal{P}_i(\boldsymbol{y}) = \boldsymbol{y} - \frac{\boldsymbol{a}_i' \boldsymbol{y} - b_i}{\|\boldsymbol{a}_i\|} \boldsymbol{a}_i.$$

Since $\boldsymbol{a}_i$ has only non-negative entries, all components of $\boldsymbol{y}$ are decreased after projection and hence, the constraint i will not be violated in the subsequent projections. This shows that given an infeasible vector $\boldsymbol{y} \in \mathbb{R}^n$, the approximate projection operation given in Definition 3 yields a feasible vector with respect to set X.

Part (ii) can be verified by using the nonexpansiveness property of projection on a closed convex set (See Proposition 2.1.3 of [26]) for l times. Since $\tilde{\boldsymbol{y}}$ is a fixed point of $\mathcal{P}_i$ for all i, we have

$$
\begin{aligned}
\|\tilde{\mathcal{P}}(\boldsymbol{y}) - \tilde{\boldsymbol{y}}\| &= \|\mathcal{P}_{i_1}(\dots(\mathcal{P}_{i_l}(\boldsymbol{y}))) - \mathcal{P}_{i_1}(\dots(\mathcal{P}_{i_l}(\tilde{\boldsymbol{y}})))\| \\
&\leq \|\mathcal{P}_{i_2}(\dots(\mathcal{P}_{i_l}(\boldsymbol{y}))) - \mathcal{P}_{i_2}(\dots(\mathcal{P}_{i_l}(\tilde{\boldsymbol{y}})))\| \\
&\vdots \\
&\leq \|\boldsymbol{y} - \tilde{\boldsymbol{y}}\|.
\end{aligned}
\tag{10}
$$

$\blacksquare$

Here, we present the gradient projection method with approximate projection to solve the problem in (8). The k-th iteration of the gradient projection method with approximate projection is given by

$$\boldsymbol{R}^{k+1} = \tilde{\mathcal{P}}(\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k), \quad \boldsymbol{g}^k \in \partial u(\boldsymbol{R}^k), \tag{11}$$

where $\boldsymbol{g}^k$ is a subgradient of u at $\boldsymbol{R}^k$, and α^k denotes the stepsize. Figure 2 demonstrates gradient projection iterations for a two-user multiple access channel. The following theorem provides a sufficient condition which can be used to establish convergence of (11) to the optimal solution.
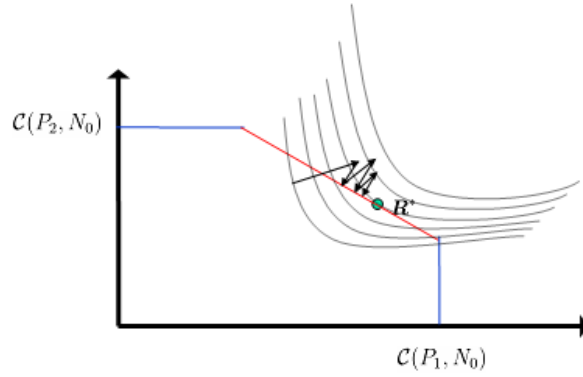
Fig. 2.   Gradient projection method with approximate projection on a two-user MAC region

Theorem 1: Let Assumptions 1 and 2 hold, and $\boldsymbol{R}^*$ be an optimal solution of problem (8). Also, let the sequence $\{\boldsymbol{R}^k\}$ be generated by the iteration in (11). If the stepsize α^k satisfies

$$0 < \alpha^k < \frac{2\left(u(\boldsymbol{R}^*) - u(\boldsymbol{R}^k)\right)}{\|\boldsymbol{g}^k\|^2}, \tag{12}$$

then

$$\|\boldsymbol{R}^{k+1} - \boldsymbol{R}^*\| < \|\boldsymbol{R}^k - \boldsymbol{R}^*\|. \tag{13}$$

Proof: We have

$$\|\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k - \boldsymbol{R}^*\|^2 = \|\boldsymbol{R}^k - \boldsymbol{R}^*\|^2 + 2\alpha^k(\boldsymbol{R}^k - \boldsymbol{R}^*)'\boldsymbol{g}^k + (\alpha^k)^2\|\boldsymbol{g}^k\|^2.$$

By concavity of $u(\cdot)$, we have

$$(\boldsymbol{R}^* - \boldsymbol{R}^k)'\boldsymbol{g}^k \geq u(\boldsymbol{R}^*) - u(\boldsymbol{R}^k). \tag{14}$$

Hence,

$$\|\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k - \boldsymbol{R}^*\|^2 \leq \|\boldsymbol{R}^k - \boldsymbol{R}^*\|^2 - \alpha^k\left[2\left(u(\boldsymbol{R}^*) - u(\boldsymbol{R}^k)\right) - (\alpha^k)\|\boldsymbol{g}^k\|^2\right].$$

If the stepsize satisfies (12), the above relation yields the following

$$\|\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k - \boldsymbol{R}^*\| < \|\boldsymbol{R}^k - \boldsymbol{R}^*\|.$$

Now by applying pseudo-nonexpansiveness of the approximate projection we have

$$\|\boldsymbol{R}^{k+1} - \boldsymbol{R}^*\| = \|\tilde{\mathcal{P}}(\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k) - \boldsymbol{R}^*\| \leq \|\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k - \boldsymbol{R}^*\| < \|\boldsymbol{R}^k - \boldsymbol{R}^*\|.$$

∎

Theorem 2: Let Assumptions 1 and 2 hold. Also, let the sequence $\{\boldsymbol{R}^k\}$ be generated by the iteration in (11). If the stepsize α^k satisfies (12), then $\{\boldsymbol{R}^k\}$ converges to an optimal solution $\boldsymbol{R}^*$.

Proof: See Proposition 8.2.7 of [25]. ∎

The convergence analysis for this method can be extended for different stepsize selection rules. For instance, we can employ diminishing stepsize, i.e.,

$$\alpha^k \to 0, \qquad \sum_{k=0}^{\infty} \alpha^k = \infty,$$

or more complicated dynamic stepsize selection rules such as the *path-based incremental target level* algorithm proposed by Brännlund [27] which guarantees convergence to the optimal solution [25], and has better convergence rate compared to the diminishing stepsize rule.

A. Complexity of the Projection Problem

Even though the approximate projection is simply obtained by successive projection on the violated constraints, it requires to find the violated constraints among exponentially many constraints describing the constraint set. In this part, we exploit the special structure of the capacity region so that each gradient projection step in (11) can be performed in polynomial time in M.

Definition 4: Let $f : 2^{\mathcal{M}} \to \mathbb{R}$ be a function defined over all subsets of $\mathcal{M}$. The function f is *submodular* if

$$f(S \cup T) + f(S \cap T) \le f(S) + f(T), \quad \text{for all } S, T \in 2^{\mathcal{M}}. \tag{15}$$

Lemma 2: Define $f_C(S) : 2^{\mathcal{M}} \to \mathbb{R}$ as follows:

$$f_C(S) = C\big(\sum_{i \in S} P_i, N_0\big), \quad \text{for all } S \subseteq \mathcal{M}. \tag{16}$$

If $P_i > 0$ for all $i \in \mathcal{M}$, then $f_C(S)$ is submodular. Moreover, the inequality (15) holds with equality if and only if $S \subseteq T$, or $T \subseteq S$.

Proof: The proof is simply by plugging the definition of $f_C(\cdot)$ in inequality (15). In particular,

$$
\begin{aligned}
f_C(S) + f_C(T) - f(S \cup T) - f(S \cap T) &= \frac{1}{2} \log \left[\frac{(N_0 + \sum_{i \in S} P_i)(N_0 + \sum_{i \in T} P_i)}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\
&= \frac{1}{2} \log \left[1 + \frac{\sum_{(i,j) \in (S \setminus T) \times (T \setminus S)} P_i P_j}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\
&\ge 0.
\end{aligned}
\tag{17}
$$

Since $P_i > 0$, the above inequality holds with equality if and only if $S \setminus T = \emptyset$, or $T \setminus S = \emptyset$. This condition is equivalent to either S or T contains the other. ∎

Theorem 3: For any $\bar{\boldsymbol{R}} \in \mathbb{R}_+^M$, finding the most violated capacity constraint in (3) can be written as a *submodular function minimization* (SFM) problem, that is unconstrained minimization of a submodular function over all $S \subseteq \mathcal{M}$.

Proof: We can rewrite the capacity constraints of $C_g(\boldsymbol{P})$ as

$$f_C(S) - \sum_{i \in S} R_i \ge 0, \quad \text{for all } S \subseteq \mathcal{M}. \tag{18}$$

Thus, the most violated constraint at $\bar{\boldsymbol{R}}$ corresponds to

$$S^* = \arg\min_{S \in 2^{\mathcal{M}}} \quad f_C(S) - \sum_{i \in S} \overline{R}_i.$$

By Lemma 2 f_C is a submodular function. Since summation of a submodular and a linear function is also submodular, the problem above is of the form of submodular function minimization. ∎

It was first shown by Grötschel *et al.* [28] that an SFM problem can be solved in polynomial time. The are several fully combinatorial strongly polynomial algorithms in the literature. The best known algorithm for SFM proposed by Orlin [29] has running time $O(M^6)$. Note that approximate projection does not require any specific order for successive projections. Hence, finding the most violated constraint is not necessary for approximate projection. In view of this fact, a more efficient algorithm based on rate-splitting is presented in Appendix I, to find a violated constraint. It is shown in Theorem 11 that the rate-splitting-based algorithm runs in $O(M^2 \log M)$ time, where M is the number of users.

Although a violated constraint can be obtained in polynomial time, it does not guarantee that the approximate projection can be performed in polynomial time. Because it is possible to have exponentially many constraints

violated at some point and hence the total running time of the projection would be exponential in M. However, we show that for a small enough stepsize in the gradient projection iteration (11), no more than M constraints can be violated at each iteration. Let us first define the notions of expansion and distance for a polyhedra.

Definition 5: Let Q be a polyhedron described by a set of linear inequalities, i.e.,

$$Q = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \le \boldsymbol{b}\}. \tag{19}$$

Define the *expansion* of Q by δ, denoted by $\mathcal{E}_\delta(Q)$, as the polyhedron obtained by relaxing all the constraints in (19), i.e., $\mathcal{E}_\delta(Q) = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \le \boldsymbol{b} + \delta\boldsymbol{1}\}$, where $\boldsymbol{1}$ is the vector of all ones.

Definition 6: Let X and Y be two polyhedra described by a set of linear constraints. Let $\mathcal{E}_d(X)$ be an expansion of X by d as defined in Definition 5. The distance $d_H(X, Y)$ between X and Y is defined as the minimum scalar d such that $X \subseteq \mathcal{E}_d(Y)$ and $Y \subseteq \mathcal{E}_d(X)$.

Lemma 3: Let f_C be as defined in (16). There exists a positive scalar δ satisfying

$$\delta \le \frac{1}{2}(f_C(S) + f_C(T) - f_C(S \cap T) - f_C(S \cup T)), \qquad \text{for all } S, T \in 2^{\mathcal{M}}, \quad S \cap T \ne S, T, \tag{20}$$

such that any point in the relaxed capacity region of an M-user multiple-access channel, $\mathcal{E}_\delta(C_g(\boldsymbol{P}))$, violates no more than M constraints of $C_g(\boldsymbol{P})$ defined in (3).

Proof: Existence of a positive scalar δ satisfying (20) follows directly from Lemma 2, using the fact that neither S nor T contains the other one.

Suppose for some $\boldsymbol{R} \in \mathcal{E}_\delta(C_g(\boldsymbol{P}))$, there are $M + 1$ violated constraints of $C_g(\boldsymbol{P})$. Since it is not possible to have $M + 1$ non-empty nested sets in $2^{\mathcal{M}}$, there are at least two violated constraints corresponding to some sets $S, T \in 2^{\mathcal{M}}$ where $S \cap T \ne S, T$, and

$$-\sum_{i \in S} R_i < -f_C(S), \tag{21}$$

$$-\sum_{i \in T} R_i < -f_C(T). \tag{22}$$

Since $\boldsymbol{R}$ is feasible in the relaxed region,

$$\sum_{i \in S \cap T} R_i \le f_C(S \cap T) + \delta, \tag{23}$$

$$\sum_{i \in S \cup T} R_i \le f_C(S \cup T) + \delta. \tag{24}$$

Note that if $S \cap T = \emptyset$, (23) reduces to $0 \le \delta$, which is a valid inequality.

By summing the above inequalities we conclude

$$\delta > \frac{1}{2}(f_C(S) + f_C(T) - f_C(S \cap T) - f_C(S \cup T)), \tag{25}$$

which is a contradiction. ∎

Theorem 4: Let Assumptions 1 and 2 hold. Let $P_1 \le P_2 \le \ldots \le P_M$ be the transmission powers.

If the stepsize α^k in the k-th iteration (11) satisfies

$$\alpha^k \le \frac{1}{4B\sqrt{M}} \log\left[1 + \frac{P_1 P_2}{(N_0 + \sum_{i=3}^{M} P_i)(N_0 + \sum_{i=1}^{M} P_i)}\right], \tag{26}$$

then at most M constraints of the capacity region $C_g(\boldsymbol{P})$ can be violated at each iteration step.

Proof: We first show that inequality in (20) holds for the following choice of δ:

$$\delta = \frac{1}{4} \log \left[1 + \frac{P_1 P_2}{(N_0 + \sum_{i=3}^{M} P_i)(N_0 + \sum_{i=1}^{M} P_i)} \right].$$ (27)

In order to verify this, rewrite the right hand side of (20) as

$$\frac{1}{4} \log \left[\frac{(N_0 + \sum_{i \in S} P_i)(N_0 + \sum_{i \in T} P_i)}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right]$$

$$= \frac{1}{4} \log \left[1 + \frac{\sum_{(i,j) \in (S \setminus T) \times (T \setminus S)} P_i P_j}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right]$$

$$\geq \frac{1}{4} \log \left[1 + \frac{P_1 P_2}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right]$$

$$\geq \frac{1}{4} \log \left[1 + \frac{P_1 P_2}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i=1}^{M} P_i)} \right]$$

$$\geq \frac{1}{4} \log \left[1 + \frac{P_1 P_2}{(N_0 + \sum_{i=3}^{M} P_i)(N_0 + \sum_{i=1}^{M} P_i)} \right].$$

The inequalities can be justified by using the monotonicity of the logarithm function and the fact that $(S \setminus T) \times (T \setminus S)$ is non-empty because $S \cap T \neq S, T$.

Now, let $\boldsymbol{R}^k$ be feasible in the capacity region, $C_g(\boldsymbol{P})$. For every $S \subseteq \mathcal{M}$, we have

$$\sum_{i \in S} (R_i^k + \alpha^k g_i^k) = \sum_{i \in S} R_i^k + \alpha^k \|g^k\| \sum_{i \in S} \frac{g_i^k}{\|g^k\|}$$

$$\leq f(S) + \frac{\delta}{B\sqrt{M}} B \sum_{i \in S} \frac{g_i^k}{\|g^k\|}$$

$$\leq f(S) + \delta,$$ (28)

where the first inequality follows from Assumption 1(b), Assumption 2, and Eq. (26). The second inequality holds because for any unit vector $\boldsymbol{d} \in \mathbb{R}^M$, it is true that

$$\sum_{i \in S} d_i \leq \sum_{i \in S} |d_i| \leq \sqrt{M}.$$ (29)

Thus, if α^k satisfies (26) then $(\boldsymbol{R}^k + \alpha^k \boldsymbol{g}^k) \in \mathcal{E}_\delta(C_g(\boldsymbol{P}))$, for some δ for which (20) holds. Therefore, by Lemma 3 the number of violated constraints does not exceed M. ∎

In view of the fact that a violated constraint can be identified in $O(M^2 \log M)$ time (see the Algorithm in Appendix I), Theorem 4 implies that, for small enough stepsize, the approximate projection can be implemented in $O(M^3 \log M)$ time.

In section V, we will develop algorithms that use the gradient projection method for dynamic rate allocation in a time varying channel.

IV. DYNAMIC RATE AND POWER ALLOCATION IN FADING CHANNEL WITH KNOWN CHANNEL STATISTICS

In this section, we assume that the channel statistics are known. Our goal is to find feasible rate and power allocation policies denoted by $\mathcal{R}^*$ and $\boldsymbol{\pi}^*$, respectively, such that $\mathcal{R}^*(\boldsymbol{H}) \in C_g\big(\boldsymbol{\pi}^*(\boldsymbol{H}), \boldsymbol{H}\big)$, and $\boldsymbol{\pi}^* \in \mathcal{G}$. Moreover,

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}^*(\boldsymbol{H})] = \boldsymbol{R}^* \in \operatorname{argmax} \quad u(\boldsymbol{R}), \qquad \text{subject to} \quad \boldsymbol{R} \in C(\bar{\boldsymbol{P}}),$$ (30)

where $u(\cdot)$ is a given utility function and is assumed to be differentiable and satisfy Assumption 1.

For the case of a linear utility function, i.e., $u(\boldsymbol{R}) = \boldsymbol{\mu}'\boldsymbol{R}$ for some $\boldsymbol{\mu} \in \mathbb{R}_+^M$, Tse *et al.* [5] have shown that the optimal rate and power allocation policies are given by the optimal solution to a linear program, i.e.,

$$(\mathcal{R}^*(\boldsymbol{h}), \boldsymbol{\pi}^*(\boldsymbol{h})) = \arg\max_{\boldsymbol{r}, \boldsymbol{p}} (\boldsymbol{\mu}'\boldsymbol{r} - \boldsymbol{\lambda}'\boldsymbol{p}), \quad \text{subject to} \quad \boldsymbol{r} \in C_g(\boldsymbol{p}, \boldsymbol{h}), \tag{31}$$

where $\boldsymbol{h}$ is the channel state realization, and $\boldsymbol{\lambda} \in \mathbb{R}_+^M$ is a Lagrange multiplier satisfying the average power constraint, i.e., $\boldsymbol{\lambda}$ is the unique solution of the following equations

$$\int_0^\infty \frac{1}{h} \int_{\frac{2\lambda_i(N_0+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k \left(\frac{2\lambda_k h(N_0+z)}{2\lambda_i(N_0+z) + (\mu_k - \mu_i)h} \right) f_i(h)\mathrm{d}h\mathrm{d}z = \bar{P}_i, \tag{32}$$

where F_k and f_k are, respectively, the cumulative distribution function (CDF) and the probability density function (PDF) of the stationary distribution of the channel state process for transmitter k.

Exploiting the polymatroid structure of the capacity region, problem (31) can be solved by a simple greedy algorithm (see Lemma 3.2 of [5]). It is also shown in [5] that, for positive $\boldsymbol{\mu}$, the optimal solution, $\boldsymbol{R}^*$, to the problem in (30) is *uniquely* obtained. Given the distribution of channel state process, denoted by F_k and f_k, we have

$$R_i^*(\boldsymbol{\mu}) = \int_0^\infty \frac{1}{2(N_0+z)} \int_{\frac{2\lambda_i(N_0+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k \left(\frac{2\lambda_k h(N_0+z)}{2\lambda_i(N_0+z) + (\mu_k - \mu_i)h} \right) f_i(h)\mathrm{d}h\mathrm{d}z. \tag{33}$$

The uniqueness of $\boldsymbol{R}^*$ follows from the fact that the stationary distribution of the channel state process has a continuous density [5]. It is worth mentioning that (33) parametrically describes the *boundary* of the capacity region which is precisely defined in Definition 2. Thus, there is a one-to-one correspondence between the boundary of $C(\bar{\boldsymbol{P}})$ and the positive vectors $\boldsymbol{\mu}$ with unit norm.

Now consider a general concave utility function satisfying Assumption 1. It is straightforward to show that $\boldsymbol{R}^*$, the optimal solution to (30), is unique. Moreover, by Assumption 1(b) it lies on the boundary of the throughput region. Now suppose that $\boldsymbol{R}^*$ is given by some genie. We can choose $\boldsymbol{\mu}^* = \nabla u(\boldsymbol{R}^*)$ and $\tilde{u}(\boldsymbol{R}) = (\boldsymbol{\mu}^*)'\boldsymbol{R}$, as a replacement for the nonlinear utility. By checking the optimality conditions, it can be seen that $\boldsymbol{R}^*$ is also the optimal solution of the problem in (30), i.e.,

$$\boldsymbol{R}^* = \arg\max \quad (\boldsymbol{\mu}^*)'\boldsymbol{R} \quad \text{subject to} \quad \boldsymbol{R} \in C(\bar{\boldsymbol{P}}). \tag{34}$$

Thus, we can employ the greedy rate and power allocation policies in (31) for the linear utility function $\tilde{u}(\cdot)$, and achieve the optimal average rate for the nonlinear utility function $u(\cdot)$. Therefore, the problem of optimal resource allocation reduces to computing the vector $\boldsymbol{R}^*$. Note that the throughput capacity region is not characterized by a finite set of constraints, so standard optimization methods such as gradient projection or interior-point methods are not applicable in this case. However, the closed-form solution to maximization of a linear function on the throughput region is given by (33). This naturally leads us to the conditional gradient method [26] to compute $\boldsymbol{R}^*$. The k-th iteration of the method is given by

$$\boldsymbol{R}^{k+1} = \boldsymbol{R}^k + \alpha^k(\bar{\boldsymbol{R}}^k - \boldsymbol{R}^k), \tag{35}$$

where α^k is the stepsize and $\bar{\boldsymbol{R}}^k$ is obtained as

$$\bar{\boldsymbol{R}}^k \in \arg\max_{\boldsymbol{R} \in C(\bar{\boldsymbol{P}})} \left(\nabla u(\boldsymbol{R}^k)'(\boldsymbol{R} - \boldsymbol{R}^k) \right), \tag{36}$$

where $\nabla u(\boldsymbol{R}^k)$ denotes the gradient vector of $u(\cdot)$ at $\boldsymbol{R}^k$. Since the utility function is monotonically increasing by Assumption 1(b), the gradient vector is always positive and, hence, the unique optimal solution to the above sub-problem is obtained by (33), in which $\boldsymbol{\mu}$ is replaced by $\nabla u(\boldsymbol{R}^k)$. By concavity of the utility function and convexity of the capacity region, the iteration (35) will converge to the optimal solution of (30) for appropriate stepsize selection rules such as the Armijo rule or limited maximization rule (cf. [26] pp. 220-222).

Note that our goal is to determine rate and power allocation policies. Finding $\boldsymbol{R}^*$ allows us to determine such policies by the greedy policy in (31) for $\boldsymbol{\mu}^* = \nabla u(\boldsymbol{R}^*)$. It is worth mentioning that all the computations for obtaining $\boldsymbol{R}^*$ are performed once in the setup of the communication session. Here, the convergence rate of the conditional gradient method is generally not of critical importance.

V. Dynamic Rate Allocation without Knowledge of Channel Statistics

In this part we assume that the channel statistics are not known and that the transmission powers are fixed to $\boldsymbol{P}$. In practice, this scenario occurs when the transmission power may be limited owing to environmental limitations such as human presence, or limitations of the hardware.

The capacity region of the fading multiple access channel for this scenario is a polyhedron given by (5). Similarly to the previous case, the goal is to find an optimal rate allocation policy $\mathcal{R}^*(\cdot)$ with respect to a given utility function, which we formally define next.

Definition 7: [Optimal Policy] The optimal rate allocation policy denoted by $\mathcal{R}^*(\cdot)$ is a mapping that satisfies $\mathcal{R}^*(\boldsymbol{H}) \in C_g(\boldsymbol{P}, \boldsymbol{H})$ for all $\boldsymbol{H}$, such that

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}^*(\boldsymbol{H})] = \boldsymbol{R}^* \in \quad \underset{}{\text{argmax}} \quad u(\boldsymbol{R})$$
$$\text{subject to} \quad \boldsymbol{R} \in C_a(\boldsymbol{P}). \tag{37}$$

It is worth noting that the approach used to find the optimal resource allocation policies for the case with known channel statistics does not apply to this scenario, because $C_g(\boldsymbol{P}, \boldsymbol{h})$ is a polyhedron and hence, unlike in Section IV the uniqueness of the optimal solution, $\boldsymbol{R}^*$ for any positive vector $\boldsymbol{\mu}$ does not hold anymore.

Here we present a greedy rate allocation policy and compare its performance with the unknown optimal policy. The performance of a particular rate allocation policy is defined as the utility function evaluated at the average rate achieved by that policy.

Definition 8: [Greedy Policy] A *greedy* rate allocation policy, denoted by $\bar{\mathcal{R}}$, is given by

$$\bar{\mathcal{R}}(\boldsymbol{H}) = \quad \underset{}{\text{argmax}} \quad u(\boldsymbol{R})$$
$$\text{subject to} \quad \boldsymbol{R} \in C_g(\boldsymbol{P}, \boldsymbol{H}) \tag{38}$$

i.e., for each channel state, the greedy policy chooses the rate vector that maximizes the utility function over the corresponding capacity region.

The utility function $u(\boldsymbol{R})$ is assumed to satisfy the following conditions.

Assumption 3: For every $\delta > 0$, let $\mathcal{N}_\delta = \left\{ \boldsymbol{H} : d_H(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})) \leq \delta \right\}$. The following conditions hold:

(a) There exists a scalar $B(\delta)$ such that for all $\boldsymbol{H} \in \mathcal{N}_\delta$,

$$|u(\boldsymbol{R}_1) - u(\boldsymbol{R}_2)| \leq B(\delta) \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|, \qquad \text{for all} \quad \boldsymbol{R}_i, \|\boldsymbol{R}_i\| \geq D_\delta, i = 1, 2,$$

where

$$D_\delta = \inf_{\boldsymbol{H} \in \mathcal{N}_\delta} \sup_{\boldsymbol{R} \in C_g(\boldsymbol{P}, \boldsymbol{H})} \|\boldsymbol{R}\|. \tag{39}$$

(b) There exists a scalar $A(\delta)$ such that for all $\boldsymbol{H} \in \mathcal{N}_\delta$,

$$|u(\bar{\mathcal{R}}(\boldsymbol{H})) - u(\boldsymbol{R})| \geq A(\delta)\|\bar{\mathcal{R}}(\boldsymbol{H}) - \boldsymbol{R}\|^2, \quad \text{for all } \boldsymbol{R} \in C_g(\boldsymbol{P}, \boldsymbol{H}).$$

Assumption 3(a) is a weakened version of Assumption 2, which imposes a bound on subgradients of the utility function. This assumption only requires bound on the subgradient in a neighborhood of the optimal solution and away from the origin, which is satisfied by a larger class of functions. Assumption 3(b) is a strong concavity type assumption. In fact, strong concavity of the utility implies Assumption 3(b), but it is not necessary. The scalar $A(\delta)$ becomes small as the utility tends to have a linear structure with level sets tangent to the dominant face of the capacity region. Assumption 3 holds for a large class of utility functions including the well known α-fair functions given by

$$f_\alpha(x) = \begin{cases} \frac{x^{1-\alpha}}{1-\alpha}, & \alpha \neq 1 \\ \log(x), & \alpha = 1, \end{cases} \tag{40}$$

which do not satisfy Assumption 2.

Note that the greedy policy is not necessarily optimal for general concave utility functions. Consider the following relations

$$\begin{aligned} \mathbb{E}_{\boldsymbol{H}}\big[u\big(\mathcal{R}^*(\boldsymbol{H})\big)\big] &\leq \mathbb{E}_{\boldsymbol{H}}\big[u\big(\bar{\mathcal{R}}(\boldsymbol{H})\big)\big] \\ &\leq u\big(\mathbb{E}_{\boldsymbol{H}}\big[\bar{\mathcal{R}}(\boldsymbol{H})\big]\big) \\ &\leq u\big(\mathbb{E}_{\boldsymbol{H}}\big[\mathcal{R}^*(\boldsymbol{H})\big]\big), \end{aligned} \tag{41}$$

where the first and third inequality follow from the feasibility of the optimal and the greedy policy for any channel state, and the second inequality follows from Jensen's inequality by concavity of the utility function.

In the case of a linear utility function we have $u\big(\mathbb{E}_{\boldsymbol{H}}\big[\mathcal{R}^*(\boldsymbol{H})\big]\big) = \mathbb{E}_{\boldsymbol{H}}\big[u\big(\mathcal{R}^*(\boldsymbol{H})\big)\big]$, so equality holds throughout in (41) and $\bar{\mathcal{R}}(\cdot)$ is indeed the optimal rate allocation policy. For nonlinear utility functions, the greedy policy can be strictly suboptimal.

However, the greedy policy is not arbitrarily worse than the optimal one. In view of (41), we can bound the performance difference, $u(\boldsymbol{R}^*) - u\big(\mathbb{E}_{\boldsymbol{H}}\big[\bar{\mathcal{R}}(\boldsymbol{H})\big]\big)$, by bounding $\big|u\big(\mathbb{E}_{\boldsymbol{H}}\big[\mathcal{R}^*(\boldsymbol{H})\big]\big) - u\big(\mathbb{E}_{\boldsymbol{H}}\big[\bar{\mathcal{R}}(\boldsymbol{H})\big]\big)\big|$ or $\big|u\big(\mathbb{E}_{\boldsymbol{H}}\big[\mathcal{R}^*(\boldsymbol{H})\big]\big) - \mathbb{E}_{\boldsymbol{H}}\big[u\big(\mathcal{R}^*(\boldsymbol{H})\big)\big]\big|$ from above. We show that the first bound goes to zero as the channel variations become small and the second bound vanishes as the utility function tends to have a more linear structure.

Before stating the main theorems, let us introduce some useful lemmas. The first lemma asserts that the optimal and greedy policies assign rates on the dominant face of the capacity region.

Lemma 4: Let $u(\cdot)$ satisfy Assumption 1(b). Also, let $\mathcal{R}^*(\cdot)$ and $\bar{\mathcal{R}}(\cdot)$ be optimal and greedy rate allocation policies as in Definitions 7 and 8, respectively. Then,

(a) $\bar{\mathcal{R}}(\boldsymbol{H}) \in \mathcal{F}\big(C_g(\boldsymbol{P}, \boldsymbol{H})\big), \quad$ for all $\boldsymbol{H}$.

(b) $\mathbf{Pr}\big\{\boldsymbol{H} : \mathcal{R}^*(\boldsymbol{H}) \in \mathcal{F}\big(C_g(\boldsymbol{P}, \boldsymbol{H})\big)\big\} = 1.$

where $\mathcal{F}(\cdot)$ denotes the dominant face of a capacity region (cf. Definition 2).

Proof: Part (a) is direct consequence of Assumption 1(b) and Definition 2. If the optimal solution to the utility maximization problem is not on the dominant face, there exists a user i such that we can increase its rate and keep all other user's rates fixed while staying in the capacity region. Thus, we are able to increase the utility by Assumption 1(b), which leads to a contradiction.

For part (b), first note that with the same argument as above we have

$$\boldsymbol{R}^* = \mathbb{E}_{\boldsymbol{H}}[\mathcal{R}^*(\boldsymbol{H})] \ \in \ \mathcal{F}(C_a(\boldsymbol{P})). \tag{42}$$

From Definition 2 and the definition of throughput capacity region in (5), we have

$$\mathbb{E}_{\boldsymbol{H}}\Big[\sum_{i=1}^M \mathcal{R}_i^*(\boldsymbol{H})\Big] = \mathbb{E}_{\boldsymbol{H}}\Big[C\Big(\sum_{i=1}^M H_i P_i, N_0\Big)\Big]. \tag{43}$$

Thus, $\sum_{i=1}^M \mathcal{R}_i^*(\boldsymbol{H}) = C\big(\sum_{i=1}^M H_i P_i, N_0\big)$, with probability one, because $C\big(\sum_{i=1}^M H_i P_i, N_0\big) - \sum_{i=1}^M \mathcal{R}_i^*(\boldsymbol{H}) \geq 0$, for all $\boldsymbol{H}$. Therefore, by definition of MAC capacity region in (3) we conclude $\mathcal{R}^*(\boldsymbol{H}) \in \mathcal{F}\big(C_g(\boldsymbol{P}, \boldsymbol{H})\big)$, with probability one. ∎

The following lemma extends Chebyshev's inequality for capacity regions. It states that, with high probability, the time varying capacity region does not deviate much from its mean.

Lemma 5: Let $\boldsymbol{H}$ be a random vector with the stationary distribution of the channel state process, mean $\bar{\boldsymbol{H}}$ and covariance matrix K. Then

$$\mathbf{Pr}\Big\{d_H\left(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})\right) > \delta\Big\} \leq \frac{\sigma_H^2}{\delta^2}, \tag{44}$$

where σ_H^2 is defined as

$$\sigma_H^2 \triangleq \frac{1}{4} \sum_{S \subseteq \{1, \dots, M\}} \boldsymbol{\Gamma}_S' K \boldsymbol{\Gamma}_S \left(1 + \left[(1 + \boldsymbol{\Gamma}_S' \bar{\boldsymbol{H}})(\sqrt{2\log(1 + \boldsymbol{\Gamma}_S' \bar{\boldsymbol{H}})} - \frac{\sqrt{\boldsymbol{\Gamma}_S' K \boldsymbol{\Gamma}_S}}{2})\right]^2\right), \tag{45}$$

where

$$(\boldsymbol{\Gamma}_S)_i = \begin{cases} \frac{P_i}{N_0}, & i \in S \\ 0, & \text{otherwise.} \end{cases} \tag{46}$$

Proof: See Appendix II. ∎

The system parameter σ_H^2 in Lemma 5 is proportional to channel variations, and we expect it to vanish for very small channel variations. The following lemma ensures that the distance between the optimal solutions of the utility maximization problem over two regions is small, provided that the regions are close to each other.

Lemma 6: Let the utility function, $u : \mathbb{R}^M \to \mathbb{R}$, satisfy Assumptions 1 and 3. Also, let $\boldsymbol{R}_1^*$ and $\boldsymbol{R}_2^*$ be the optimal solution of maximizing the utility over $C_a(\boldsymbol{P})$ and $C_g(\boldsymbol{P}, \boldsymbol{H})$, respectively. If

$$d_H\big(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})\big) \leq \delta,$$

then we have

$$\|\boldsymbol{R}_1^* - \boldsymbol{R}_2^*\| \leq \delta^{\frac{1}{2}} \left[\delta^{\frac{1}{2}} + \left(\frac{B(\delta)}{A(\delta)}\right)^{\frac{1}{2}}\right]. \tag{47}$$

Proof: See Appendix III. ∎

The following theorem combines the results of the above two lemmas to obtain a bound on the performance difference of the greedy and the optimal policy.

Theorem 5: Let $u : \mathbb{R}^M \to \mathbb{R}_+$ satisfy Assumptions 1 and 3. Also, let $\mathcal{R}^*(\cdot)$ and $\bar{\mathcal{R}}(\cdot)$ be optimal and greedy rate allocation policies as in Definitions 7 and 8, respectively. Then for every $\delta \in [\sigma_H^2, \infty)$,

$$u(\boldsymbol{R}^*) - u\big(\mathbb{E}_{\boldsymbol{H}}[\bar{\mathcal{R}}(\boldsymbol{H})]\big) \leq \frac{\sigma_H^2}{\delta^2} u(\boldsymbol{R}^*) + \Big(1 - \frac{\sigma_H^2}{\delta^2}\Big) B(\delta) \Big[\delta^{\frac{1}{2}} + \Big(\frac{B(\delta)}{A(\delta)}\Big)^{\frac{1}{2}}\Big] \delta^{\frac{1}{2}}, \tag{48}$$

where $\boldsymbol{R}^* = \mathbb{E}_{\boldsymbol{H}}[\mathcal{R}^*(\boldsymbol{H})]$, and $A(\delta)$ and $B(\delta)$ are positive scalars defined in Assumption 3.

Proof: Pick any $\delta \in [\sigma_H^2, \infty)$. Define the event, $\mathcal{V}$ as

$$\mathcal{V} = \left\{ \boldsymbol{H} : d_H\big(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})\big) \leq \delta \right\}.$$

By Lemma 5, the probability of this event is at least $1 - \frac{\sigma_H^2}{\delta^2}$. Using Jensen's inequality as in (41) we can bound the left-hand side of (48) as follows

$$
\begin{aligned}
u(\boldsymbol{R}^*) - u\big(\mathbb{E}_{\boldsymbol{H}}[\bar{\mathcal{R}}(\boldsymbol{H})]\big) &\leq u(\boldsymbol{R}^*) - \mathbb{E}_{\boldsymbol{H}}\big[u(\bar{\mathcal{R}}(\boldsymbol{H}))\big] \\
&= u(\boldsymbol{R}^*) - (1 - \frac{\sigma_H^2}{\delta^2})\mathbb{E}_{\boldsymbol{H}}\Big[u(\bar{\mathcal{R}}(\boldsymbol{H}))\Big|\mathcal{V}\Big] \\
&\quad - \mathbf{Pr}(\mathcal{V}^c)\mathbb{E}_{\boldsymbol{H}}\Big[u(\bar{\mathcal{R}}(\boldsymbol{H}))\Big|\mathcal{V}^c\Big] \\
&\leq \frac{\sigma_H^2}{\delta^2}u(\boldsymbol{R}^*) + (1 - \frac{\sigma_H^2}{\delta^2})\Big(u(\boldsymbol{R}^*) - \mathbb{E}_{\boldsymbol{H}}\big[u(\bar{\mathcal{R}}(\boldsymbol{H}))|\mathcal{V}\big]\Big) \\
&\leq \frac{\sigma_H^2}{\delta^2}u(\boldsymbol{R}^*) + (1 - \frac{\sigma_H^2}{\delta^2})\Big|\mathbb{E}_{\boldsymbol{H}}\big[u(\boldsymbol{R}^*) - u(\bar{\mathcal{R}}(\boldsymbol{H}))|\mathcal{V}\big]\Big| \\
&\leq \frac{\sigma_H^2}{\delta^2}u(\boldsymbol{R}^*) + (1 - \frac{\sigma_H^2}{\delta^2})\mathbb{E}_{\boldsymbol{H}}\Big[\big|u(\boldsymbol{R}^*) - u(\bar{\mathcal{R}}(\boldsymbol{H}))\big|\Big|\mathcal{V}\Big]. \quad (49)
\end{aligned}
$$

In the above relations, the first inequality follows from the fact that $\mathbf{Pr}(\mathcal{V}) \geq 1 - \frac{\sigma_H^2}{\delta^2}$, and the second inequality holds because of the non-negativity of $u(\boldsymbol{R})$.

On the other hand, by incorporating Lemma 4 in Assumption 3(a) we have

$$|u(\boldsymbol{R}^*) - u(\bar{\mathcal{R}}(\boldsymbol{H}))| \leq B(\delta)\|\bar{\mathcal{R}}(\boldsymbol{H}) - \boldsymbol{R}^*\|.$$

Now by Assumption 3 we can employ Lemma 6 to conclude the following from the above relation:

$$|u(\boldsymbol{R}^*) - u(\bar{\mathcal{R}}(\boldsymbol{H}))| \leq B(\delta)\Big(\delta^{\frac{1}{2}} + \Big(\frac{B(\delta)}{A(\delta)}\Big)^{\frac{1}{2}}\Big)\delta^{\frac{1}{2}}, \quad \text{for all } \boldsymbol{H}, d_H\big(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})\big) \leq \delta,$$

which implies

$$\mathbb{E}_{\boldsymbol{H}}\Big[\big|u(\boldsymbol{R}^*) - u(\bar{\mathcal{R}}(\boldsymbol{H}))\big|\Big|\mathcal{V}\Big] \leq B(\delta)\Big(\delta^{\frac{1}{2}} + \Big(\frac{B(\delta)}{A(\delta)}\Big)^{\frac{1}{2}}\Big)\delta^{\frac{1}{2}}. \quad (50)$$

The desired result follows immediately from substituting (50) in (49).

∎

Theorem 5 provides a bound parameterized by δ. For very small channel variations, σ_H becomes small. Therefore, the parameter δ can be picked small enough such that the bound in (48) tends to zero. Figure 3 illustrates the behavior of right hand side of Eq. (48) as a function of δ for different values of σ_H. For each value of σ_H, the upper bound is minimized for a specific choice of δ, which is illustrated by a dot in Figure 3. As demonstrated in the figure, for smaller channel variations, a smaller gap is achieved and the parameter δ that minimizes the bound decreases.

The next theorem provides another bound demonstrating the impact of the structure of the utility function on the performance of the greedy policy.

Theorem 6: Let Assumption 1 hold for the twice differentiable function $u : \mathbb{R}^M \to \mathbb{R}_+$. Also, let $\mathcal{R}^*(\cdot)$ and $\bar{\mathcal{R}}(\cdot)$ be the optimal and the greedy rate allocation policies, defined in Definitions 7 and 8, respectively. Then for every $\epsilon \in (0, 1]$,

$$u(\boldsymbol{R}^*) - u\big(\mathbb{E}_{\boldsymbol{H}}[\bar{\mathcal{R}}(\boldsymbol{H})]\big) \leq \epsilon u(\boldsymbol{R}^*) + \frac{1}{2}(1 - \epsilon)r(\epsilon)^2\Omega, \quad (51)$$
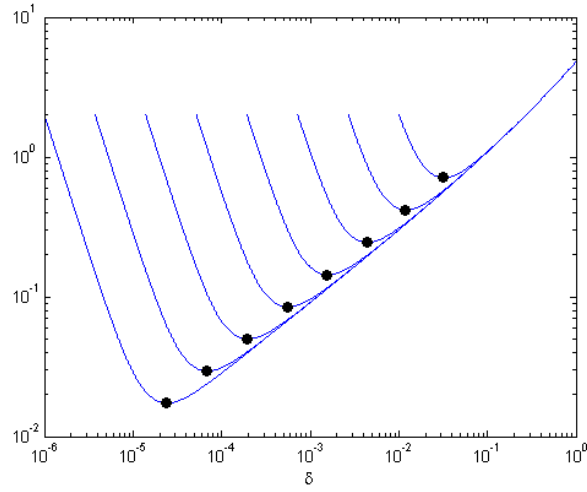
Fig. 3. Parametric upper bound on performance difference between greedy and optimal policies as in right hand side of (48) for different channel variations, σ_H, as a function of δ

where $\boldsymbol{R}^* = \mathbb{E}_{\boldsymbol{H}}[\mathcal{R}^*(\boldsymbol{H})]$, and Ω satisfies the following

$$\lambda_{\max}\big(-\nabla^2 u(\boldsymbol{\xi})\big) \leq \Omega, \qquad \text{for all } \boldsymbol{\xi}, \ \|\boldsymbol{\xi} - \boldsymbol{R}^*\| \leq r(\epsilon), \tag{52}$$

in which ∇^2 denotes the Hessian of u, and $r(\epsilon)$ is given by

$$r(\epsilon) = \sqrt{M}\frac{\sigma_H}{\sqrt{\epsilon}} + \left[\sum_{i=1}^{M} \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(\frac{(1+H_iP_i)(1+\sum_{j\neq i}H_jP_j)}{1+\sum_{j=1}^{M}H_jP_j}\right)\right]^2\right]^{\frac{1}{2}}. \tag{53}$$

Proof: Similarly to the proof of Theorem 5, for any $\epsilon \in (0,1]$ define the event $\mathcal{V}$ as

$$\mathcal{V} = \left\{\boldsymbol{H} : d_H(C_g(\boldsymbol{P},\boldsymbol{H}), C_a(\boldsymbol{P})) \leq \frac{\sigma_H}{\sqrt{\epsilon}}\right\}. \tag{54}$$

By Lemma 5, this event has probability at least $1 - \epsilon$. Lemma 4 asserts that the optimal policy almost surely allocate rate vectors on the dominant face of $C_g(\boldsymbol{P}, \boldsymbol{H})$. Therefore, for almost all $\boldsymbol{H} \in \mathcal{V}$, the optimal policy satisfies the following

$$\mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1+\frac{H_iP_i}{1+\sum_{j\neq i}H_jP_j}\right)\right] - \frac{\sigma_H}{\sqrt{\epsilon}} \leq \mathcal{R}_i^*(\boldsymbol{H}) \leq \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1+H_iP_i\right)\right] + \frac{\sigma_H}{\sqrt{\epsilon}}. \tag{55}$$

Thus, for almost all $\boldsymbol{H} \in \mathcal{V}$, we have

$$|\mathcal{R}_i^*(\boldsymbol{H}) - R_i^*| \leq \frac{\sigma_H}{\sqrt{\epsilon}} + \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(\frac{(1+H_iP_i)(1+\sum_{j\neq i}H_jP_j)}{1+\sum_{j=1}^{M}H_jP_j}\right)\right].$$

Therefore,

$$\begin{aligned}
\|\mathcal{R}^*(\boldsymbol{H}) - \boldsymbol{R}^*\| &\leq \sqrt{M}\frac{\sigma_H}{\sqrt{\epsilon}} + \left[\sum_{i=1}^{M}\mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(\frac{(1+H_iP_i)(1+\sum_{j\neq i}H_jP_j)}{1+\sum_{j=1}^{M}H_jP_j}\right)\right]^2\right]^{\frac{1}{2}} \\
&= r(\epsilon), \quad \text{for almost all } \boldsymbol{H} \in \mathcal{V}.
\end{aligned} \tag{56}$$

Now let us write the Taylor expansion of $u(\cdot)$ at $\boldsymbol{R}^*$ in the direction of $\boldsymbol{R}$,

$$
\begin{aligned}
u(\boldsymbol{R}) &= u(\boldsymbol{R}^*) + \nabla u(\boldsymbol{R}^*)'(\boldsymbol{R} - \boldsymbol{R}^*) - \frac{1}{2}(\boldsymbol{R} - \boldsymbol{R}^*)'(-\nabla^2 u(\xi))(\boldsymbol{R} - \boldsymbol{R}^*) \\
&\geq u(\boldsymbol{R}^*) + \nabla u(\boldsymbol{R}^*)'(\boldsymbol{R} - \boldsymbol{R}^*) - \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}^*\|^2 \lambda_{\max}(-\nabla^2 u(\xi)) \\
&\qquad\qquad\qquad \text{for some } \xi,\ \|\xi - \boldsymbol{R}^*\| \leq \|\boldsymbol{R} - \boldsymbol{R}^*\|.
\end{aligned}
\tag{57}
$$

In the above relation, let $\boldsymbol{R} = \mathcal{R}^*(\boldsymbol{H})$ for all $\boldsymbol{H} \in \mathcal{V}$. The utility function is concave, so its Hessian is negative definite and we can combine (56) with the above relation to write

$$
u(\mathcal{R}^*(\boldsymbol{H})) \geq u(\boldsymbol{R}^*) + \nabla u(\boldsymbol{R}^*)'(\mathcal{R}^*(\boldsymbol{H}) - \boldsymbol{R}^*) - \frac{1}{2}r(\epsilon)^2\Omega, \quad \text{for almost all } \boldsymbol{H} \in \mathcal{V}.
\tag{58}
$$

Taking the expectation conditioned on $\mathcal{V}$, and using the fact that $\mathcal{R}^*(\boldsymbol{H}) \in \mathcal{F}(C_g(\boldsymbol{P}, \boldsymbol{H}))$ we have the following

$$
\mathbb{E}_{\boldsymbol{H}}\big[u(\mathcal{R}^*(\boldsymbol{H}))\big|\mathcal{V}\big] \geq u(\boldsymbol{R}^*) - \frac{1}{2}r(\epsilon)^2\Omega.
\tag{59}
$$

Hence, we conclude

$$
\begin{aligned}
u(\boldsymbol{R}^*) - u(\mathbb{E}_{\boldsymbol{H}}(\bar{\mathcal{R}}(\boldsymbol{H}))) &\leq u(\boldsymbol{R}^*) - \mathbb{E}_{\boldsymbol{H}}[u(\mathcal{R}^*(\boldsymbol{H}))] \\
&\leq u(\boldsymbol{R}^*) - (1 - \epsilon)\mathbb{E}_{\boldsymbol{H}}\Big[u(\mathcal{R}^*(\boldsymbol{H}))\Big|\mathcal{V}\Big] \\
&\qquad - \mathbf{Pr}(\mathcal{V}^c)\mathbb{E}_{\boldsymbol{H}}\Big[u(\mathcal{R}^*(\boldsymbol{H}))\Big|\mathcal{V}^c\Big] \\
&\leq u(\boldsymbol{R}^*) - (1 - \epsilon)\big(u(\boldsymbol{R}^*) - \frac{1}{2}r(\epsilon)^2\Omega\big) \\
&= \epsilon u(\boldsymbol{R}^*) + \frac{1}{2}(1 - \epsilon)r(\epsilon)^2\Omega.
\end{aligned}
$$

where the first inequality is verified by (41), and the third inequality follows from non-negativity of the utility function and the inequality in (59). ∎

Similarly to Theorem 5, Theorem 6 provides a bound parameterized by ϵ. As the utility function tends to have a more linear structure, Ω tends to zero. For instance, Ω is proportional to α for a weighted sum α-fair utility function. Hence, we can choose ϵ small such that the right hand side of (51) goes to zero. The behavior of this upper bound for different values of Ω is similar to the one plotted in Figure 3.

In summary, the performance difference between the greedy and the optimal policy is bounded from above by the minimum of the bounds provided by Theorem 5 and Theorem 6.

Even though the greedy policy can perform closely to the optimal policy, it requires solving a nonlinear program in each time slot. For each channel state, finding even a near-optimal solution of the problem in (38) requires a large number of iterations, making the online evaluation of the greedy policy impractical. In the following section, we introduce an alternative rate allocation policy, which implements a single gradient projection iteration of the form (11) per time slot.

A. Approximate Rate Allocation Policy

In this part, we assume that the channel state information is available at each time slot n, and the computational resources are limited such that a single iteration of the gradient projection method in (11) can be implemented in each time slot. In order to simplify the notation in this part and avoid unnecessary technical details, we consider a stronger version of Assumption 3(b).

Assumption 4: Let $\boldsymbol{R}^{\dagger} = argmax_{\boldsymbol{R} \in C_g(\boldsymbol{P}, \boldsymbol{H})} u(\boldsymbol{R})$. *Then there exists a positive scalar A such that*

$$|u(\boldsymbol{R}^{\dagger}) - u(\boldsymbol{R})| \geq A \|\boldsymbol{R}^{\dagger} - \boldsymbol{R}\|^2, \quad \text{for all } \boldsymbol{R} \in C_g(\boldsymbol{P}, \boldsymbol{H}).$$

Definition 9: [Approximate Policy] Given some fixed integer $k \geq 1$, we define the *approximate* rate allocation policy, $\widetilde{\mathcal{R}}$, as follows:

$$\widetilde{\mathcal{R}}\big(\boldsymbol{H}(n)\big) \triangleq \begin{cases} \bar{\mathcal{R}}\big(\boldsymbol{H}(0)\big), & n = 0 \\ \tilde{\boldsymbol{R}}_{t(n)}^{\tau}, & n \geq 1, \end{cases} \tag{60}$$

where

$$\tau = \underset{0 \leq j < k-1}{\text{argmax}} \ u(\tilde{\boldsymbol{R}}_{t(n)}^j), \quad t(n) = \left\lfloor \frac{n-1}{k} \right\rfloor, \tag{61}$$

and $\tilde{\boldsymbol{R}}_{t(n)}^j \in \mathbb{R}^M$ is given by the following gradient projection iterations:

$$\begin{aligned} \tilde{\boldsymbol{R}}_{t(n)}^0 &= \tilde{P}_{t(n)}\left[\widetilde{\mathcal{R}}\Big(\boldsymbol{H}\big(kt(n)\big)\Big)\right], \\ \tilde{\boldsymbol{R}}_{t(n)}^{j+1} &= \tilde{P}_{t(n)}\left[\tilde{\boldsymbol{R}}_{t(n)}^j + \alpha^j \tilde{\boldsymbol{g}}_{t(n)}^j\right], \quad j = 1, \ldots, k-1, \end{aligned} \tag{62}$$

where $\tilde{\boldsymbol{g}}_{t(n)}^j$ is a subgradient of $u(\cdot)$ at $\tilde{\boldsymbol{R}}_{t(n)}^j$, α^j denotes the stepsize and $\tilde{P}_{t(n)}$ is the approximate projection on $C_g\big(\boldsymbol{P}, \boldsymbol{H}(kt(n))\big)$.

For $k = 1$, (62) reduces to taking only one gradient projection iteration at each time slot. For $k > 1$, the proposed rate allocation policy essentially allows the channel state to change for a block of k consecutive time slots, and then takes k iterations of the gradient projection method with the approximate projection. We will show below that this method tracks the greedy policy closely. Hence, this yields an efficient method that on average requires only one iteration step per time slot. Note that to compute the policy at time slot n, we are using the channel state information at time slots $kt, k(t-1), \ldots$. Hence, in practice the channel measurements need to be done only every k time slots.

There is a tradeoff in choosing system parameter k, because taking only one gradient projection step may not be sufficient to get close enough to the greedy policy's operating point. Moreover, for large k the new operating point of the greedy policy can be far from the previous one, and k iterations may be insufficient.

Before stating the main result, let us introduce some useful lemmas. In the following lemma, we translate the model in Definition 1 for temporal variations in channel state into changes in the corresponding capacity regions.

Lemma 7: Let $\big\{[H_i(n)]_{i=1,\ldots,M}\big\}$ be the fading process that satisfies condition in (2). We have

$$d_H\Big(C_g\big(\boldsymbol{P}, \boldsymbol{H}(n+1)\big), C_g\big(\boldsymbol{P}, \boldsymbol{H}(n)\big)\Big) \leq W_n, \tag{63}$$

where $\{W_n\}$ are non-negative independent identically distributed random variables bounded from above by $\hat{w} = \frac{1}{2}\sum_{i=1}^M \hat{v}^i P_i$, where $\hat{v}^i$ is a uniform upper bound on the sequence of random variables $\{V_n^i\}$ and P_i is the i-th user's transmission power.

Proof: By Definition 6 we have

$$d_H\Big(C_g\big(\boldsymbol{P}, \boldsymbol{H}(n+1)\big), C_g\big(\boldsymbol{P}, \boldsymbol{H}(n)\big)\Big)$$

$$= \max_{S \subseteq \mathcal{M}} \frac{1}{2}\left| \log\left(1 + \frac{\sum_{i \in S}(H_i(n+1) - H_i(n))P_i}{1 + \sum_{i \in S} H_i(n)P_i}\right)\right|$$

$$\leq \max_{S \subseteq \mathcal{M}} \frac{\sum_{i \in S}|H_i(n+1) - H_i(n)|P_i}{2(1 + \sum_{i \in S} H_i(n)P_i)}$$

$$\leq \frac{1}{2}\sum_{i=1}^{M} |H_i(n+1) - H_i(n)|P_i = \frac{1}{2}\sum_{i=1}^{M} V_n^i P_i. \tag{64}$$

Therefore, (63) is true for $W_n = \frac{1}{2}\sum_{i=1}^{M} V_n^i P_i$. Since the random variables V_n^i are i.i.d. and bounded above by $\hat{v}_n^i$, the random variables W_n are i.i.d. and bounded from above by $\frac{1}{2}\sum_{i=1}^{M} \hat{v}^i P_i$.

∎

The following useful lemma by Nedić and Bertsekas [30] addresses the convergence rate of the gradient projection method with constant stepsize.

Lemma 8: Let rate allocation policies $\bar{\mathcal{R}}$ and $\widetilde{\mathcal{R}}$ be given by Definition 8 and Definition 9, respectively. Also, let Assumptions 1, 2 and 4 hold and the stepsize α^n be fixed to some positive constant α. Then for a positive scalar ϵ we have

$$u\Big(\widetilde{\mathcal{R}}\big(\boldsymbol{H}(n)\big)\Big) \geq u\Big(\bar{\mathcal{R}}\big(\boldsymbol{H}(kt)\big)\Big) - \frac{\alpha B^2 + \epsilon}{2}, \tag{65}$$

if k satisfies

$$k \geq \left\lfloor \frac{\|\tilde{\boldsymbol{R}}_t^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(kt)\big)\|^2}{\alpha \epsilon} \right\rfloor. \tag{66}$$

Proof: See Proposition 2.3 of [30]. ∎

We next state our main result, which shows that the approximate rate allocation policy given by Definition 9 tracks the greedy policy within a neighborhood which is quantified as a function of the maximum speed of fading, the parameters of the utility function, and the transmission powers.

Theorem 7: Let Assumptions 1, 2 and 4 hold and the rate allocation policies $\bar{\mathcal{R}}$ and $\widetilde{\mathcal{R}}$ be given by Definition 8 and Definition 9, respectively. Choose the system parameters k and α for the approximate policy in Definition 9 as

$$k = \left\lfloor (\frac{2B}{Aw'})^{\frac{2}{3}} \right\rfloor, \qquad \alpha = \left(\frac{16A}{B^2}\right)^{\frac{1}{3}} w'^{\frac{2}{3}},$$

where $w' = \hat{w}^{\frac{1}{2}}\big(\hat{w}^{\frac{1}{2}} + (\frac{B}{A})^{\frac{1}{2}}\big)$, $\hat{w}$ is the upper bound on W_n as defined in Lemma 7, A and B are constants given in Assumptions 4 and 2. Then, we have

$$\|\widetilde{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(n)\big)\| \leq 2\theta = 2\Big(\frac{2B}{A}\Big)^{\frac{2}{3}} w'^{\frac{1}{3}}. \tag{67}$$

Proof: First, we show that

$$\|\widetilde{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(kt)\big)\| \leq \theta = \Big(\frac{2B}{A}\Big)^{\frac{2}{3}} w'^{\frac{1}{3}}, \tag{68}$$

where $t = \lfloor \frac{n-1}{k} \rfloor$. The proof is by induction on t. For $t = 0$ the claim is trivially true. Now suppose that (68) is true for some positive t. Hence, it also holds for $n = k(t+1)$ by induction hypothesis, i.e.,

$$\|\tilde{\boldsymbol{R}}_{t+1}^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(kt)\big)\| \leq \theta. \tag{69}$$

On the other hand, by Lemma 7 implies that for every n,

$$d_H\Big(C_g\big(\boldsymbol{P}, \boldsymbol{H}(n+1)\big), C_g\big(\boldsymbol{P}, \boldsymbol{H}(n)\big)\Big) \leq \hat{w}.$$

Thus, by Lemma 6 and the triangle inequality we have

$$\|\bar{\mathcal{R}}\big(\boldsymbol{H}(k(t+1))\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(kt)\big)\| \le kw' \le \theta. \tag{70}$$

Therefore, by another triangle inequality we conclude from (69) and (70) that

$$\|\tilde{\boldsymbol{R}}_{t+1}^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(k(t+1))\big)\| \le 2\theta. \tag{71}$$

After plugging the corresponding values of α and θ, it is straightforward to show that (66) holds for $\epsilon = \alpha B^2$. Thus, we can apply Lemma 8 to show

$$\left| u\Big(\widetilde{\mathcal{R}}\big(\boldsymbol{H}(n)\big)\Big) - u\Big(\bar{\mathcal{R}}\big(\boldsymbol{H}(k(t+1))\big)\Big) \right| \le \alpha B^2. \tag{72}$$

By Assumption 4 we can write

$$\|\widetilde{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(k(t+1))\big)\| \le \Big(\frac{\alpha B^2}{A}\Big)^{\frac{1}{2}} = \theta. \tag{73}$$

Therefore, the proof of (68) is complete by induction.

Again by applying Lemma 6 and Lemma 7 we have

$$\|\bar{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(kt)\big)\| \le kw' \le \theta, \tag{74}$$

and the desired result directly follows from (68) and (74) by the triangle inequality. ∎

Theorem 7 provides a bound on the size of the tracking neighborhood as a function of the maximum speed of fading, denoted by $\hat{w}$, which may be too conservative. It is of interest to provide a rate allocation policy and a bound on the size of its tracking neighborhood as a function of the average speed of fading. The next section addresses this issue.

B. Improved Approximate Rate Allocation Policy

In this section, we design an efficient rate allocation policy that tracks the greedy policy within a neighborhood characterized by the average speed of fading which is typically much smaller than the maximum speed of fading. We consider policies which can implement one gradient projection iteration per time slot.

Unlike the approximate policy given by (60) which uses the channel state information once in every k time slots, we present an algorithm which uses the channel state information in all time slots. Roughly speaking, this method takes a fixed number of gradient projection iterations only after the change in the channel state has reached a certain threshold.

Definition 10: [Improved Approximate Policy] Let $\{W_n\}$ be the sequence of non-negative random variables as defined in Lemma 7, and γ be a positive constant. Define the sequence $\{T_i\}$ as

$$
\begin{aligned}
T_0 &= 0, \\
T_{i+1} &= \min\Big\{t \mid \sum_{n=T_i}^{t-1} W_n \ge \gamma\Big\}.
\end{aligned}
\tag{75}
$$

Define the *improved approximate* rate allocation policy, $\widehat{\mathcal{R}}$, with parameters γ and k, as follows:

$$\widehat{\mathcal{R}}\big(\boldsymbol{H}(n)\big) \triangleq \begin{cases} \bar{\mathcal{R}}\big(\boldsymbol{H}(0)\big), & n = 0 \\ \hat{\boldsymbol{R}}_{t(n)}^\tau, & n \ge 1, \end{cases} \tag{76}$$
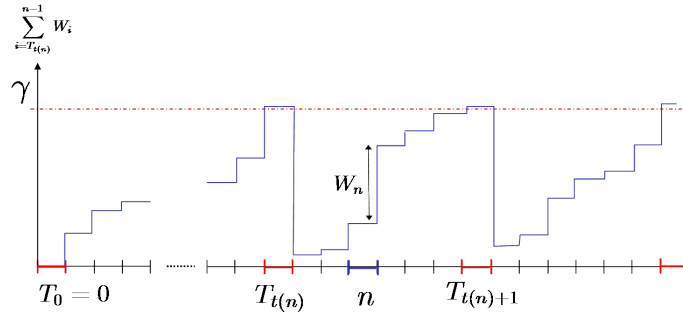
Fig. 4. The improved approximate policy takes k gradient projection iterations at time $T_{t(n)}$, which is the time that the random walk generated by the random variables W_n reach the threshold γ.

where

$$t(n) \;=\; \max\{i \mid T_i < n\}, \tag{77}$$

$$\tau \;=\; \operatorname*{arg\,max}_{0 \le j < k-1} \; u\!\left(\hat{\boldsymbol{R}}_{t(n)}^{j}\right), \tag{78}$$

and $\hat{\boldsymbol{R}}_{t(n)}^{j} \in \mathbb{R}^M$ is given by the following gradient projection iterations

$$\hat{\boldsymbol{R}}_{t(n)}^{0} \;=\; \tilde{P}_{t(n)}\left[\widehat{\mathcal{R}}\big(\boldsymbol{H}(T_{t(n)})\big)\right],$$

$$\hat{\boldsymbol{R}}_{t(n)}^{j+1} \;=\; \tilde{P}_{t(n)}\left[\hat{\boldsymbol{R}}_{t(n)}^{j} + \alpha^j \hat{\boldsymbol{g}}_{t(n)}^{j}\right], \quad j = 1, \ldots, k-1, \tag{79}$$

where $\hat{\boldsymbol{g}}_{t(n)}^{j}$ is a subgradient of $u(\cdot)$ at $\hat{\boldsymbol{R}}_{t(n)}^{j}$, α^j denotes the stepsize and $\tilde{P}_{t(n)}$ is the approximate projection on $C_g(\boldsymbol{P}, \boldsymbol{H}(T_{t(n)}))$.

Figure 4 depicts a particular realization of the random walk generated by W_n, and the operation of the improved approximate policy.

Theorem 8: Let $t(n)$ be as defined in (77), and let $\bar{w} = \mathbb{E}[W_n]$. If $k = \frac{\gamma}{\bar{w}}$, then we have

$$\lim_{n \to \infty} \frac{n}{t(n)k} = 1, \quad \text{with probability 1.} \tag{80}$$

Proof: The sequence $\{T_i\}$ is obtained as the random walk generated by the W_n crosses the threshold level γ. Since the random variables W_n are positive, we can think of the threshold crossing as a renewal process, denoted by $N(\cdot)$, with inter-arrivals W_n.

We can rewrite the limit as follows

$$\lim_{n \to \infty} \frac{n - N\big(t(n)\gamma\big) + N\big(t(n)\gamma\big)}{t(n)k} = \lim_{n \to \infty} \frac{n - N\big(t(n)\gamma\big)}{t(n)k} + \bar{w}\frac{N\big(t(n)\gamma\big)}{t(n)\gamma}. \tag{81}$$

Since the random walk will hit the threshold with probability 1, the first term goes to zero with probability 1. Also, by Strong law for renewal processes the second terms goes to 1 with probability 1 (see [31], p.60). ∎

Theorem 8 essentially guarantees that the number of gradient projection iterations is the same as the number of channel measurements in the long run with probability 1.

Theorem 9: Let Assumptions 1, 2 and 4 hold and the rate allocation policies $\bar{\mathcal{R}}$ and $\widehat{\mathcal{R}}$ be given by Definition 8 and Definition 10, respectively. Also, let $k = \lfloor \frac{\gamma}{\bar{w}} \rfloor$, and fix the stepsize to $\alpha = \frac{A\gamma^2}{B^2}$ in (79), where $\gamma = c(\frac{B}{A})^{\frac{3}{4}}\bar{w}^{\frac{1}{4}}$, and $c \ge 1$ is a constant satisfying the following equation

$$\frac{(c^2 - 1)^8}{2^8 c^4} = \hat{w}. \tag{82}$$

Then

$$\|\widehat{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(n)\big)\| \le 2\gamma + \Big(\frac{\gamma B}{A}\Big)^{\frac{1}{2}}. \tag{83}$$

Proof: We follow the line of proof of Theorem 7. First, by induction on t we show that

$$\|\widehat{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_t)\big)\| \le \gamma, \tag{84}$$

where t is defined in (77). The base is trivial. Similar to (69), by induction hypothesis we have

$$\|\hat{\boldsymbol{R}}_{t+1}^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_t)\big)\| \le \gamma. \tag{85}$$

By definition of T_i in (75) we can write

$$d_H\Big(C_g\big(\boldsymbol{P}, \boldsymbol{H}(T_{t+1})\big), C_g\big(\boldsymbol{P}, \boldsymbol{H}(T_t)\big)\Big) \le \gamma. \tag{86}$$

Thus, by Lemma 6, we have

$$\|\bar{\mathcal{R}}\big(\boldsymbol{H}(T_{t+1})\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_t)\big)\| \le \gamma^{\frac{1}{2}}\Big(\gamma^{\frac{1}{2}} + \big(\frac{B}{A}\big)^{\frac{1}{2}}\Big). \tag{87}$$

Therefore, by combining (85) and (87) by triangle inequality we obtain

$$\|\hat{\boldsymbol{R}}_{t+1}^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_{t+1})\big)\| \le 2\gamma + \Big(\frac{\gamma B}{A}\Big)^{\frac{1}{2}}. \tag{88}$$

Using the fact that $\bar{w} \le \hat{w} = \frac{(c^2-1)^8}{2^8 c^4}$, after a few steps of straightforward manipulations we can show that

$$\|\hat{\boldsymbol{R}}_{t+1}^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_{t+1})\big)\|^2 \le \Big(2\gamma + \big(\frac{\gamma B}{A}\big)^{\frac{1}{2}}\Big)^2 \le c^4 \frac{\gamma B}{A}. \tag{89}$$

Now by plugging the values of α and γ in terms of system parameters in (66), we can verify that

$$k = \Big\lfloor \frac{\gamma}{\bar{w}} \Big\rfloor = \Big\lfloor \frac{c^4 \frac{\gamma B}{A}}{A \frac{\gamma^2}{B^2} A \gamma^2} \Big\rfloor \ge \Big\lfloor \frac{\|\hat{\boldsymbol{R}}_{t+1}^0 - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_{t+1})\big)\|^2}{\alpha\epsilon} \Big\rfloor. \tag{90}$$

Hence, we can apply Lemma 8 for $\epsilon = A\gamma^2$, and conclude

$$\Big| u\Big(\widehat{\mathcal{R}}\big(\boldsymbol{H}(n)\big)\Big) - u\Big(\bar{\mathcal{R}}\big(\boldsymbol{H}(T_{t+1})\big)\Big) \Big| \le \alpha B^2. \tag{91}$$

By exploiting Assumption 4 we have

$$\|\widehat{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_{t+1})\big)\| \le \Big(\frac{\alpha B^2}{A}\Big)^{\frac{1}{2}} = \gamma. \tag{92}$$

Therefore, the proof of (84) is complete by induction. Similarly to (87) we have

$$\|\bar{\mathcal{R}}\big(\boldsymbol{H}(n)\big) - \bar{\mathcal{R}}\big(\boldsymbol{H}(T_t)\big)\| \le \gamma^{\frac{1}{2}}\Big(\gamma^{\frac{1}{2}} + \big(\frac{B}{A}\big)^{\frac{1}{2}}\Big), \tag{93}$$

and (83) follows immediately from (84) and (93) by invoking triangle inequality. ∎

Theorem 8 and Theorem 9 guarantee that the presented rate allocation policy tracks the greedy policy within a small neighborhood while only one gradient projection iteration is computed per time slot, with probability 1. The neighborhood is characterized in terms of the average behavior of temporal channel variations and vanishes as the fading speed decreases.
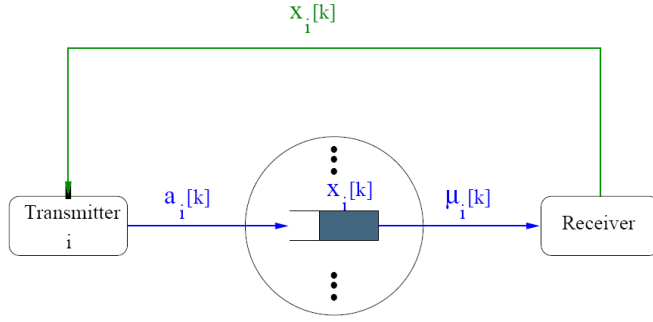
Fig. 5.   Structure of the i-th transmitter and the receiver for the queue-length-based policy [10].
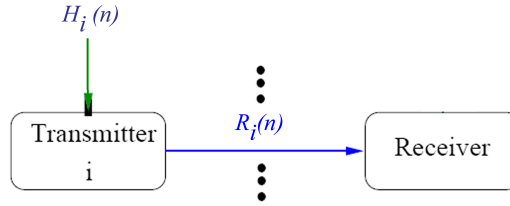

Fig. 6.   Structure of the i-th transmitter and the receiver for the presented policies.

VI. SIMULATION RESULTS AND DISCUSSION

In this section, we provide simulation results to complement our analytical results and make a comparison with other fair resource allocation algorithms. We focus on the case with no power control or knowledge of channel statistics. We also make reasonable assumption that the channel state processes are generated by independent identical finite state Markov chains. We consider weighted α-fair function as the utility function, i.e.,

$$u(\boldsymbol{R}) = \sum_i w_i f_\alpha(R_i), \tag{94}$$

where $f_\alpha(\cdot)$ is given by equation (40).

We consider two different scenarios to compare the performance of the greedy policy with the queue-based rate allocation policy by Eryilmaz and Srikant [10]. This policy, parameterized by some parameter K, uses queue length information to allocate the rates arbitrarily close to the optimal policy by choosing K large enough. As illustrated in Figure 6, $x_i(n)$ denotes the queue-length of the i-th user. At time slot n, the scheduler chooses the service rate vector $\boldsymbol{\mu}(n)$ based on a max-weight policy, i.e.,

$$\boldsymbol{\mu}(n) = \quad \underset{}{\text{argmax}} \quad \sum_{i=1}^{M} x_i(n) R_i$$
$$\text{subject to} \quad \boldsymbol{R} \in C_g(\boldsymbol{P}, \boldsymbol{H}(n)) \tag{95}$$

The congestion controller proposed in [10] leads to a fair allocation of the rates for a given α-fair utility function. In particular, the data generation rate for the i-th transmitter, denoted by $a_i(n)$ is a random variable satisfying the following conditions:

$$\mathbb{E}\big[a_i(n) \,|x_i(n)\big] = \min\left\{K\Big(\frac{w_i}{x_i(n)}\Big)^{\frac{1}{\alpha}}, D\right\},$$
$$\mathbb{E}\big[a_i^2(n) \,|x_i(n)\big] \leq U < \infty, \quad \text{for all } x_i(n), \tag{96}$$
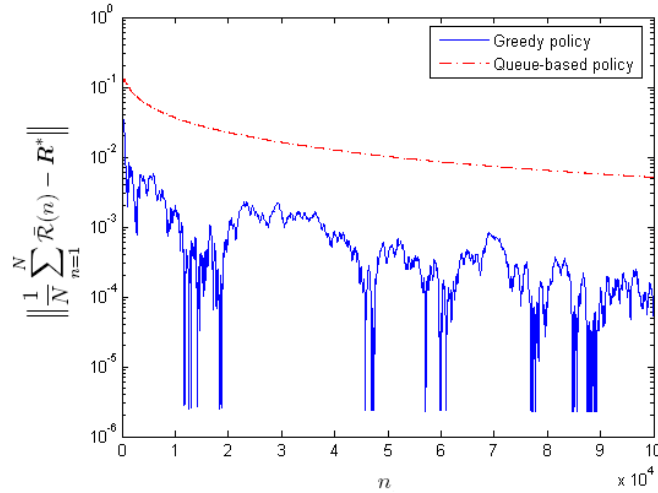
Fig. 7.   Performance comparison of greedy and queue-based policies for a communication session with limited duration, for $\frac{\sigma_H}{H} = 1.22$.
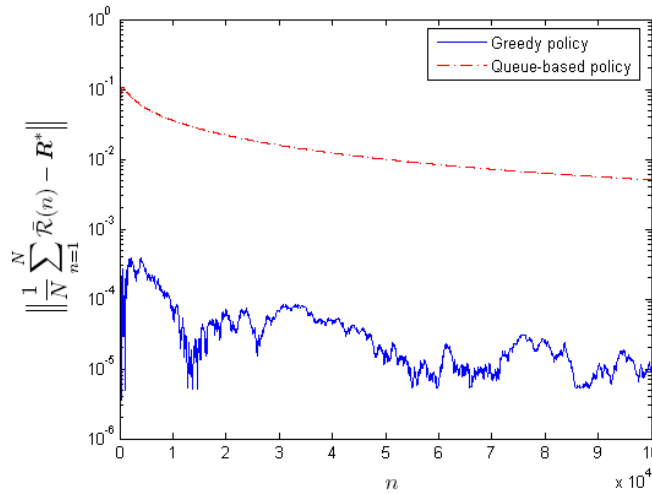

Fig. 8.   Performance comparison of greedy and queue-based policies for a communication session with limited duration, for $\frac{\sigma_H}{H} = 0.13$.

where α, D and U are positive constants.

In the first scenario, we compare the average achieved rate of the policies for a communication session with limited duration. Figure 7 depicts the distance between empirical average rate achieved by the greedy or the queue-length based policy, and $\boldsymbol{R}^*$, the maximizer of the utility function over the throughput region. In this case, the utility function is given by (94) with $\alpha = 2$ and $w_1 = 1.5w_2 = 1.5$, and the corresponding optimal solution is $\boldsymbol{R}^* = (0.60, 0.49)$. As observed in Figure 7, the greedy policy outperforms the queue-length based policy a communication session with limited duration. It is worth noting that there is a tradeoff in choosing the parameter K of the queue-length based policy. In order to guarantee achieving close to optimal rates by queue-based policy, the parameter K should be chosen large which results in large expected queue length and lower convergence rate. On the other hand, if K takes a small value to improve the convergence rate, the achieved rate of the queue based policy converges to a larger neighborhood of the $\boldsymbol{R}^*$.

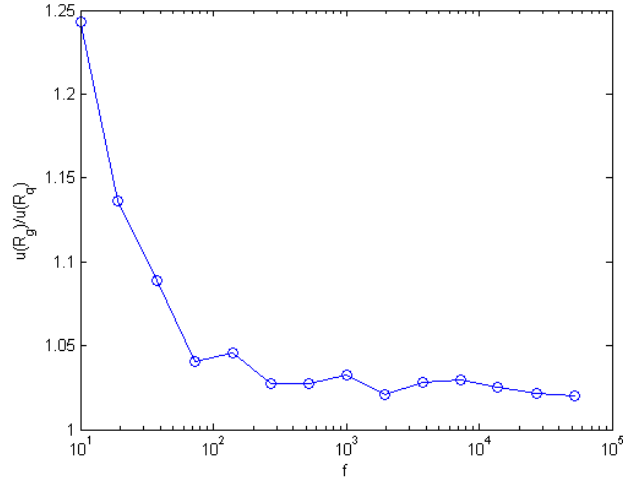As established in Theorem 5, the performance of the greedy policy improves by decreasing the channel variations.

Fig. 9. Performance comparison of greedy and queue-based policies for file upload scenario with respect to file size $f = f_1 = f_2$. R_g and R_q are expected upload rate of the greedy and the queue-length based policy, respectively.

Figure 8 demonstrates the improvement in performance of the greedy policy when $\frac{\sigma_H}{H}$ decreases from 1.22 to 0.13. We also observe in Figure 8 that the queue-length based policy is not sensitive to channel variations, and its performance does not improve by decreasing the channel variations. It is worth mentioning that the greedy policy as observed in the simulation results performs significantly better than the bounds provided by Theorems 5 and 6. These upper bounds characterize the behavior of the greedy policy in terms of channel variations and structure of the utility function, but they are not necessarily tight.

Second, we consider a file upload scenario where each user transmitting a file with finite size to the base station in a rateless manner. Let $\mathcal{T}_i$ be the i-th user's completion time of the file upload session for a file of size f_i. Define the average upload rate for the i-th user as $\frac{f_i}{\mathcal{T}_i}$. We can measure the performance of each policy for this scenario by evaluating the utility function at the average upload rate. Figure 9 demonstrates the utility difference of the greedy and the queue-based policy for different file sizes. We can observe that for small file sizes the greedy policy outperforms the queue-based policy significantly, and this difference decreases by increasing the file size. We can interpret this behavior as follows. The files are first buffered into the queues based on the queue lengths and the weighted α-fair utility, while the queues are emptied by a max-weight scheduler. Once the files are all buffered in the queues, the queues are empties with the same rate which is not fair because it does not give any priority to the users based on their utility. For larger file size, the duration for which the entire file is emptied into the queue is negligible compared to the total transmission time, and the average upload rate converges to a near-optimal rate.

VII. Conclusion

We addressed the problem of optimal resource allocation in a fading multiple access channel from an information theoretic point of view. We formulated the problem as a utility maximization problem for a more general class of utility functions.

We considered several different scenarios. First, we considered the problem of optimal rate allocation in a non-fading channel. We presented the notion of approximate projection for the gradient projection method to solve the rate allocation problem in polynomial time in the number of users.

Second, we studied rate and power allocation in a fading channel with known channel statistics. In this case, the optimal rate and power allocation policies are obtained by greedily maximizing a properly defined linear utility function. If for the fading channel power control and channel statistics are not available, the greedy policy is not optimal for nonlinear utility functions. However, we showed that its performance in terms of the utility is not arbitrarily worse compared to the optimal policy, by bounding their performance difference. The provided bound tends to zero as the channel variations become small or the utility function behaves more linearly.

The greedy policy may itself be computationally expensive. A computationally efficient algorithm can be employed to allocate rates close to the ones allocated by the greedy policy. Two different rate allocation policies are presented which only take one iteration of the gradient projection method with approximate projection at each time slot. It is shown that these policies track the greedy policy within a neighborhood which is characterized by average speed of fading as well as fading speed in the worst case.

APPENDIX I
ALGORITHM FOR FINDING A VIOLATED CONSTRAINT

In this section, we present an alternative algorithm based on rate-splitting idea to identify a violated constraint for an infeasible point. For a feasible point, the algorithm provides information for decoding by successive cancellation. We first introduce some definitions.

Definition 11: The quadruple $(M, \boldsymbol{P}, \boldsymbol{R}, N_0)$ is called a *configuration* for an M-user multiple-access channel, where $\boldsymbol{R} = (R_1, \ldots, R_M)$ is the rate tuple, $\boldsymbol{P} = (P_1, \ldots, P_M)$ represents the received power and N_0 is the noise variance. For any given configuration, the *elevation*, $\boldsymbol{\delta} \in \mathbb{R}^M$, is defined as the unique vector satisfying

$$R_i = C(P_i, N_0 + \delta_i), \quad i = 1, \ldots, M. \tag{97}$$

Intuitively, we can think of message i as rectangles of height P_i, raised above the noise level by δ_i. In fact, δ_i is the amount of additional Gaussian interference that message i can tolerate. Note that if the rate vector corresponding to a configuration is feasible its elevation vector is non-negative. However, that is not sufficient for feasibility check.

Definition 12: The configuration $(M, \boldsymbol{P}, \boldsymbol{R}, N_0)$ is *single-user codable*, if after possible re-indexing,

$$\delta_{i+1} \geq \delta_i + P_i, i = 0, 1, \ldots, M - 1, \tag{98}$$

where we have defined $\delta_0 = P_0 = 0$ for convention.

By the graphical representation described earlier, a configuration is single-user codable if the none of the messages are overlapping. Figure 10(a) gives an example of graphical representing for a message with power P_i and elevation δ_i. Figures 10(b) and 10(c) illustrate overlapping and non-overlapping configurations, respectively.

Definition 13: The quadruple $(m, \boldsymbol{p}, \boldsymbol{r}, N_0)$ is a *spin-off* of $(M, \boldsymbol{P}, \boldsymbol{R}, N_0)$ if there exists a surjective mapping $\phi : \{1, \ldots, m\} \to \{1, \ldots, M\}$ such that for all $i \in \{1, \ldots, M\}$ we have

$$P_i \geq \sum_{j \in \phi^{-1}(i)} p_j,$$

$$R_i \leq \sum_{j \in \phi^{-1}(i)} r_j.$$

where $\phi^{-1}(i)$ is the set of all $j \in \{1, \ldots, m\}$ that map into i by means of ϕ.

Definition 14: A *hyper-user* with power $\bar{P}$, rate $\bar{R}$, is obtained by merging d actual users with powers $(P_{i_1}, \ldots, P_{i_d})$ and rates $(R_{i_1}, \ldots, R_{i_d})$, i.e,

$$\bar{P} = \sum_{k=1}^{d} P_{i_k}, \quad \bar{R} = \sum_{k=1}^{d} R_{i_k}. \tag{99}$$

Theorem 10: For any M-user achievable configuration $(M, \boldsymbol{P}, \boldsymbol{R}, N_0)$, there exists a spin-off $(m, \boldsymbol{p}, \boldsymbol{r}, N_0)$ which is single user codable.

Proof: See Theorem 1 of [8]. ∎

Here, we give a brief sketch of the proof to give intuition about the algorithm. The proof is by induction on M. For a given configuration, if none of the messages are overlapping then the spin-off is trivially equal to the configuration. Otherwise, merge the two overlapping users into a *hyper-user* of rate and power equal the sum rate and sum power of the overlapping users, respectively. Now the problem is reduced to rate splitting for $(M-1)$ users. This proof suggests a recursive algorithm for rate-splitting that gives the actual spin-off for a given configuration.

It follows directly from the proof of Proposition 10 that this recursive algorithm gives a single-user codable spin-off for an achievable configuration. If the configuration is not achievable, then the algorithm encounters a hyper-user with negative elevation. At this point the algorithm terminates. Suppose that hyper-user has rate $\bar{R}$ and power $\bar{P}$. Negative elevation is equivalent to the following

$$\bar{R} > C(\bar{P}, N_0).$$

Hence, by Definition 14 we have,

$$\sum_{i \in S} R_i > C(\sum_{i \in S} P_i, N_0).$$

where $S = \{i_1, \ldots, i_d\} \subseteq \mathcal{M}$. Therefore, a hyper-user with negative elevation leads us to a violated constraint in the initial configuration.

Theorem 11: The presented algorithm runs in $O(M^2 \log M)$ time, where M is the number of users.

Proof: The computational complexity of the algorithm can be computed as follows. The algorithm terminates after at most M recursions. At each recursion, all the elevations corresponding to a configuration with at most M hyper-users are computed in $O(M)$ time. It takes $O(M \log M)$ time to sort the elevation in an increasing order. Once the users are sorted by their elevation, a hyper-user with negative elevation could be found in $O(1)$ time, or two if such a hyper-user does not exists it takes $O(M)$ time to find two overlapping hyper-users. In the case that there are no overlapping users and all the elevations are non-negative the input configuration is achievable, and the algorithm terminates with no violated constraint. Hence, computational complexity of each recursion is $O(M) + O(M \log M) + O(M) = O(M \log M)$. Therefore, the algorithm runs in $O(M^2 \log M)$ time. ∎

APPENDIX II
PROOF OF LEMMA 5

First, consider the following lemmas. Lemma 9 bounds Jensen's difference of a random variable for a concave function. The upper bound is characterized in terms of the variance of the random variable.

Lemma 9: Let $f : \mathbb{R} \to \mathbb{R}_+$ be concave and twice differentiable. Let X be a random variable with variance σ_X^2. Then,

$$f(\mathbb{E}[X]) - \mathbb{E}[f(X)] \le \sqrt{2M\sigma_X^2 f(\mathbb{E}[X])} - \frac{\sigma_X^2 M}{2}, \tag{100}$$
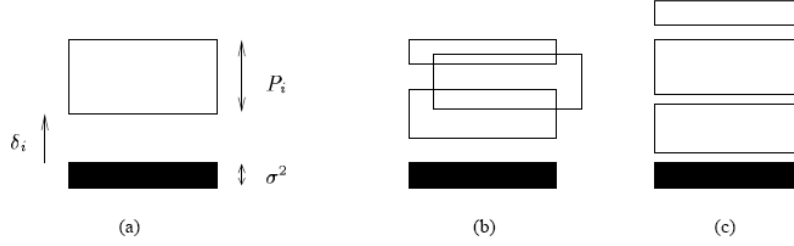
Fig. 10. Graphical representation of messages over multi-access channel [8].

where M be an upper-bound on $|f''(x)|$.

Proof: Pick any $0 < \epsilon \leq 1$. By Chebyshev's inequality we have

$$\mathbf{Pr}\left(|X - \mathbb{E}(X)| > c\right) \leq \epsilon, \tag{101}$$

where $c = \frac{\sigma_X}{\sqrt{\epsilon}}$. Therefore, we have

$$
\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}\left[f(X)\Big||X - \mathbb{E}(X)| \leq c\right]\mathbf{Pr}\left(|X - \mathbb{E}(X)| \leq c\right) \\
&+ \mathbb{E}\left[f(X)\Big||X - \mathbb{E}(X)| > c\right]\mathbf{Pr}\left(|X - \mathbb{E}(X)| > c\right) \\
&\geq (1 - \epsilon)\mathbb{E}\left[f(X)\Big||X - \mathbb{E}(X)| \leq c\right] \\
&\geq \frac{1 - \epsilon}{2}\left(f\left(\mathbb{E}[X] + c\right) + f\left(\mathbb{E}[X] - c\right)\right) \\
&= (1 - \epsilon)f(\mathbb{E}[X]) + \frac{1 - \epsilon}{4}c^2(f''(\xi_1) + f''(\xi_2)), \tag{102}
\end{aligned}
$$

where the first inequality follows from non-negativity of f, and the second and the second inequality follows from concavity of f. The scalars $\xi_1 \in \left[\mathbb{E}[X], \mathbb{E}[X] + c\right]$ and $\xi_2 \in \left[\mathbb{E}[X] - c, \mathbb{E}[X]\right]$ are given by Taylor's theorem.

Given the above relation, for any $\epsilon > 0$ we have

$$f(\mathbb{E}[X]) - \mathbb{E}[f(X)] \leq \frac{1 - \epsilon}{2\epsilon}\sigma_X^2 M + \epsilon f(\mathbb{E}[X]). \tag{103}$$

The right-hand side is minimized for

$$\epsilon^* = \min\left\{\left(\frac{\sigma_X^2 M}{2f(\mathbb{E}[X])}\right)^{\frac{1}{2}}, 1\right\}. \tag{104}$$

By substituting ϵ^* in (103), the desired result follows immediately.

∎

We next provide an upper bound on variance of $Y = \log(1 + X)$ proportional to the variance of X.

Lemma 10: Let $X > 0$ be a random variable with mean $\bar{X}$ and variance σ_X^2, and $Y = \log(1 + X)$ then variance of Y is upper-bounded as

$$\sigma_Y^2 \leq \sigma_X^2\left(1 + \left[(1 + \bar{X})(\sqrt{2\log(1 + \bar{X})} - \frac{\sigma_X}{2})\right]^2\right). \tag{105}$$

Proof: Let $\mathbb{E}(Y) = \log(1 + \hat{X})$ for some $\hat{X} < \bar{X}$. By invoking the mean value theorem, we have

$$
\begin{aligned}
\sigma_Y^2 &= \mathbb{E}\left[\left(\log(1 + X) - \log(1 + \hat{X})\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{1 + \tilde{X}}(X - \hat{X})\right)^2\right] \\
&\leq \mathbb{E}\left[(X - \hat{X})^2\right],
\end{aligned}
\tag{106}
$$

where $\tilde{X}$ is a non-negative random variable.

On the other hand, by employing lemma 9 with $f(x) = \log(1 + x)$, we can write

$$
\mathbb{E}\left[\log(1 + X)\right] \geq \log(1 + \bar{X}) - \sqrt{2\sigma_X^2 \log(1 + \bar{X})} + \frac{\sigma_X^2}{2}.
\tag{107}
$$

Hence,

$$
\begin{aligned}
\bar{X} \geq \hat{X} &= \exp\left\{\mathbb{E}[\log(1 + X)]\right\} - 1 \\
&\geq \exp\left\{\log(1 + \bar{X}) - \sqrt{2\sigma_X^2 \log(1 + \bar{X})} + \frac{\sigma_X^2}{2}\right\} - 1 \\
&\geq \bar{X} - \sigma_X(1 + \bar{X})\left(\sqrt{2\log(1 + \bar{X})} - \frac{\sigma_X}{2}\right),
\end{aligned}
\tag{108}
$$

where the first inequality is by (107), and the second relation can be verified after some straightforward manipulation. By combining (106) and (107) the variance of Y can be bounded as follows

$$
\begin{aligned}
\sigma_Y^2 &\leq \mathbb{E}[(X - \hat{X})^2] \\
&\leq \mathbb{E}\left[\left(X - \bar{X} + \sigma_X(1 + \bar{X})\left(\sqrt{2\log(1 + \bar{X})} - \frac{\sigma_X}{2}\right)\right)^2\right] \\
&= \sigma_X^2\left(1 + \left[(1 + \bar{X})\left(\sqrt{2\log(1 + \bar{X})} - \frac{\sigma_X}{2}\right)\right]^2\right).
\end{aligned}
\tag{109}
$$

∎

Now we provide the proof for Lemma 5. Define random variable Y_S as the following:

$$
Y_S = \frac{1}{2}\log(1 + \sum_{i \in S}\frac{H_i P_i}{N_0}), \quad \text{for all } S \subseteq \mathcal{M} = \{1, \ldots, M\}.
\tag{110}
$$

The facet defining constraints of $C_g(\boldsymbol{P}, \boldsymbol{H})$ and $C_a(\boldsymbol{P})$ are of the form of $\sum_{i \in S} R_i \leq Y_S$ and $\sum_{i \in S} R_i \leq \mathbb{E}[Y_S]$, respectively. Therefore, by Definition 6, we have $d_H\left(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})\right) \leq \delta$ if and only if $|Y_S - \mathbb{E}[Y_S]| \leq \delta$, for all $S \subseteq \mathcal{M}$. Thus, we can write

$$
\begin{aligned}
\mathbf{Pr}\left\{d_H\left(C_g(\boldsymbol{P}, \boldsymbol{H}), C_a(\boldsymbol{P})\right) > \delta\right\} &= \mathbf{Pr}\left\{\max_S |Y_S - \mathbb{E}[Y_S]| > \delta\right\} \\
&\leq \sum_{S \subseteq \mathcal{M}} \mathbf{Pr}\left\{|Y_S - \mathbb{E}[Y_S]| > \delta\right\} \\
&\leq \frac{1}{\delta^2}\sum_{S \subseteq \mathcal{M}} \sigma_{Y_S}^2.
\end{aligned}
\tag{111}
$$

where the first inequality is obtained by union bound, and the second relation is by applying Chebyshev's inequality. On the other hand, $\sigma_{Y_S}^2$ can be bounded from above by employing Lemma 10, i.e.,

$$\sigma_{Y_S}^2 \leq \frac{\sigma_{Z_S}^2}{4}\left(1 + \left[(1 + \bar{Z}_S)(\sqrt{2\log(1 + \bar{Z}_S)} - \frac{\sigma_{Z_S}}{2})\right]^2\right), \tag{112}$$

where

$$\bar{Z}_S = \mathbb{E}\Big[\sum_{i \in S} \frac{H_i P_i}{N_0}\Big] = \sum_{i \in S} \Gamma_i \bar{H}_i = \boldsymbol{\Gamma}'_S \bar{\boldsymbol{H}},$$

$$\sigma_{Z_S}^2 = \mathrm{var}\Big(\sum_{i \in S} \frac{H_i P_i}{N_0}\Big) = \sum_{(i,j) \in S^2} \Gamma_i \Gamma_j \mathrm{cov}(H_i, Hj) = \boldsymbol{\Gamma}'_S K \boldsymbol{\Gamma}_S.$$

The desired result is concluded by substituting $\bar{Z}_S$ and $\sigma_{Z_S}^2$ in (112) and combing the result with (111). ∎

APPENDIX III
PROOF OF LEMMA 6

Let us first state and prove a useful lemma which asserts that Euclidean expansion of a capacity region by δ contains its expansion by relaxing its constraints by δ.

Lemma 11: Let C_1 be a capacity region with *polymatroid* structure, i.e.,

$$C_1 = \Big\{\boldsymbol{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq f(S), \text{ for all } S \subseteq \mathcal{M}\Big\}, \tag{113}$$

where $f(S)$ is a nondecreasing submodular function. Also, let C_2 be an *expansion* of C_1 by δ as defined in Definition 5. Then, for all $\boldsymbol{R} \in C_2$, there exists some $\boldsymbol{R}' \in C_1$ such that $\|\boldsymbol{R} - \boldsymbol{R}'\| \leq \delta$.

Proof: By Definition 15, it is straightforward to show that C_2 is also a polymatroid, i.e.,

$$C_2 = \Big\{\boldsymbol{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq g(S) = f(S) + \delta, \text{ for all } S \subseteq \mathcal{M}\Big\}, \tag{114}$$

where $g(S)$ is a submodular function. By convexity of C_2, we just need to prove the claim for the vertices of C_2. Let $\boldsymbol{R} \in \mathbb{R}^M$ be a vertex of C_2. The polymatroid structure of C_2 implies that $\boldsymbol{R}$ is generated by an ordered subset of $\mathcal{M}$ (see Theorem 2.1 of [32]). Hence, there is some $k \in \mathcal{M}$ such that $R_k = f(\{k\}) + \delta$. Consider the following construction for $\boldsymbol{R}'$:

$$R_i' = \begin{cases} R_i - \delta, & i = k \\ R_i, & \text{otherwise.} \end{cases} \tag{115}$$

By construction, $\boldsymbol{R}'$ is in a δ-neighborhood of $\boldsymbol{R}$. So we just need to show that R' is feasible in C_1. First, let us consider the sets S that contain k. We have

$$\sum_{i \in S} R_i' = \sum_{i \in S} R_i - \delta \leq f(S). \tag{116}$$

Second, consider the case that $k \notin S$.

$$\begin{aligned} \sum_{i \in S} R_i' &= \sum_{i \in S \cup \{k\}} R_i' - R_k + \delta \\ &\leq f(S \cup \{k\}) + \delta - R_k \\ &\leq f(S) + f(\{k\}) + \delta - R_k \\ &= f(S), \end{aligned}$$

where the first inequality comes from (116), and the second inequality is true by submodularity of the function $f(\cdot)$. This completes the proof. ∎

Proof: [of Lemma 6] Without loss of generality assume that $u(\boldsymbol{R}_2^*) \geq u(\boldsymbol{R}_1^*)$. By Lemma 11, there exists some $\boldsymbol{R} \in C_a(\boldsymbol{P})$ such that $\|\boldsymbol{R}_2^* - \boldsymbol{R}\| \leq \delta$. Moreover, we can always choose $\boldsymbol{R}$ to be on the boundary so that $\|\boldsymbol{R}\| \geq D_\delta$, where D_δ is defined in (39). Therefore, by Assumption 3(a) and the fact that $u(\boldsymbol{R}_2^*) \geq u(\boldsymbol{R}_1^*) \geq u(\boldsymbol{R})$, we have

$$u(\boldsymbol{R}_2^*) - u(\boldsymbol{R}) = |u(\boldsymbol{R}_2^*) - u(\boldsymbol{R})| \leq B\|\boldsymbol{R}_2^* - \boldsymbol{R}\| \leq B\delta. \tag{117}$$

Now suppose that $\|\boldsymbol{R}_1^* - \boldsymbol{R}\| > (\frac{B}{A}\delta)^{\frac{1}{2}}$. By Assumption 3(b) we can write

$$u(\boldsymbol{R}_1^*) - u(\boldsymbol{R}) = |u(\boldsymbol{R}_1^*) - u(\boldsymbol{R})| \geq A\|\boldsymbol{R}_1^* - \boldsymbol{R}\|^2 > B\delta. \tag{118}$$

By subtracting (117) from (118) we obtain $u(\boldsymbol{R}_2^*) < u(\boldsymbol{R}_1^*)$ which is a contradiction. Therefore, $\|\boldsymbol{R}_1^* - \boldsymbol{R}\| \leq (\frac{B}{A}\delta)^{\frac{1}{2}}$, and the desired result follows immediately by invoking the triangle inequality. ∎

References

[1] X. Wang and G.B. Giannakis. Energy-efficient resource allocation in time division multiple-access over fading channels. Preprint, 2005.

[2] S.J. Oh, Z. Danlu, and K.M. Wasserman. Optimal resource allocation in multiservice CDMA networks. *IEEE Transactions on Wireless Communications*, 2(4):811–821, 2003.

[3] J.B. Kim and M.L. Honig. Resource allocation for multiple classes of DS-CDMA traffic. *IEEE Transactions on Vehicular Technology*, 49(2):506–519, 2000.

[4] J. Huang, V. Subramanian, R. Berry, and R. Agrawal. Joint scheduling and resource allocation in OFDM systems: Algorithms and performance for the uplink. In *Proceedings of 41st Annual Asilomar Conference on Signals, Systems, and Computers (invited paper)*, 2007.

[5] D. Tse and S. Hanly. Multiaccess fading channels part I: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, 44(7):2796–2815, 1998.

[6] S. Shenker. Fundamental design issues for the future internet. *IEEE Journal on Selected Areas in Communications*, 13:1176–118, 1995.

[7] R. Srikant. *Mathematics of Internet Congestion Control*. Birkhauser, 2004.

[8] B. Rimoldi and R. Urbanke. A rate-splitting approach to the gaussian multiple-access channel. *IEEE Transactions on Information Theory*, 42(2):364–375, 1996.

[9] X. Lin and N. Shroff. The impact of imperfect scheduling on cross-layer rate control in multihop wireless networks. In *Proceedings of IEEE Infocom*, Miami, FL, March 2005.

[10] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length based scheduling and congestion control. In *Proceedings of IEEE Infocom*, volume 3, pages 1794–1803, Miami, FL, March 2005.

[11] M.J. Neely, E. Modiano, and C. Li. Fairness and optimal stochastic control for heterogeneous networks. In *Proceedings of IEEE Infocom*, pages 1723–1734, Miami, FL, March 2005.

[12] A. Stolyar. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems*, 50(4):401–457, 2005.

[13] A. Eryilmaz, A. Ozdaglar, and E. Modiano. Polynomial complexity algorithms for full utilization of multi-hop wireless networks. In *Proceedings of IEEE Infocom*, Anchorage, AL, May 2007.

[14] L. Tassiulas. Linear complexity algorithms for maximum throughput in radio networks and input queued switches. In *Proceedings of IEEE Infocom*, pages 533–539, 1998.

[15] E. Modiano, D. Shah, and G. Zussman. Maximizing throughput in wireless networks via gossiping. In *ACM SIGMETRICS/IFIP Performance*, 2006.

[16] S. Sanghavi, L. Bui, and R. Srikant. Distributed link scheduling with constant overhead, 2007. Technical Report.

[17] C. Joo, X. Lin, and N. Shroff. Performance limits of greedy maximal matching in multi-hop wireless networks. In *Proceedings of IEEE Conference on Decision and Control*, New Orleans LA, December 2007.

[18] A. Eryilmaz, A. Ozdaglar, D. Shah, and E. Modiano. Randomized algorithms for optimal control of wireless networks, 2007. ICCOPT Conference, Hamilton CA.

[19] S. Vishwanath, S.A. Jafar, and A. Goldsmith. Optimum power and rate allocation strategies for multiple access fading channels. In *Proceedings of IEEE VTC*, 2001.

[20] E. Yeh and A. Cohen. Delay optimal rate allocation in multiaccess fading communications. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2004.

[21] D. Yu and J.M. Cioffi. Iterative water-filling for optimal resource allocation in OFDM multiple-access and broadcast channels. In *Proceedings of IEEE GLOBECOM*, 2006.

[22] K. Seong, R. Narasimhan, and J. Cioffi. Scheduling for fading multiple access channels with heterogeneous QoS constraints. In *Proceedings of International Symposium on Information Theory*, 2007.

[23] H. Liao. Multiple access channels. Ph.D. thesis, University of Hawaii, Honolulu, 1972.

[24] S. Shamai and A.D. Wyner. Information theoretic considerations for symmetric, cellular, multiple-access fading channels part I. *IEEE Transactions on Information Theory*, 43(6):1877–1894, 1997.

[25] D.P. Bertsekas, A. Nedić, and A.E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Cambridge, Massachusetts, 2003.

[26] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Cambridge, Massachusetts, 1999.

[27] U. Brännlund. On relaxation methods for nonsmooth convex optimization. Doctoral thesis, Royal Institute of Technology, Stockholm, Sweden, 1993.

[28] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[29] J. Orlin. A faster strongly polynomial time algorithm for submodular function minimization. In *Proceedings of the 12th Conference on Integer Programming and Combinatorial Optimization*, pages 240–251, 2007.

[30] A. Nedić and D.P. Bertsekas. *Convergence Rate of Incremental Subgradient Algorithms*. Stochastic Optimization: Algorithms and Applications (S. Uryasev and P. M. Pardalos, Editors), Kluwer Academic Publishers, 2000.

[31] R. Gallager. *Discrete Stochastic Processes*. Kluwer Academic Publishers, London, United Kingdom, 1996.

[32] R. E. Bixby, W. H. Cunningham, and D. M. Topkis. The partial order of a polymatroid extreme point. *Mathematics of Operations Research*, 10(3):367–378, 1985.