



# MIT Open Access Articles

## *On Resource Allocation in Fading Multiple Access Channels -- An Efficient Approximate Projection Approach*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<b>Citation</b>	ParandehGheibi, Ali et al. "On Resource Allocation in Fading Multiple-Access Channels&#x2013;An Efficient Approximate Projection Approach." IEEE Transactions on Information Theory 56.9 (2010): 4417-4437. © Copyright 2010 IEEE
<b>As Published</b>	<a href="http://dx.doi.org/10.1109/TIT.2010.2053902">http://dx.doi.org/10.1109/TIT.2010.2053902</a>
<b>Publisher</b>	Institute of Electrical and Electronics Engineers (IEEE)
<b>Version</b>	Final published version
<b>Citable link</b>	<a href="http://hdl.handle.net/1721.1/73045">http://hdl.handle.net/1721.1/73045</a>
<b>Terms of Use</b>	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

# On Resource Allocation in Fading Multiple-Access Channels—An Efficient Approximate Projection Approach

Ali ParandehGheibi, *Student Member, IEEE*, Atilla Eryilmaz, *Member, IEEE*, Asuman Ozdaglar, *Member, IEEE*, and Muriel Médard, *Fellow, IEEE*

**Abstract**—In this paper, we consider the problem of rate and power allocation in a multiple-access channel (MAC). 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 MAC without using queue-length information. First, we address the utility maximization problem in a nonfading channel and present a gradient projection algorithm with approximate projections. 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 present optimal rate and power allocation policies in a fading channel where channel statistics are known. For the case that channel statistics are unknown and the transmission power is fixed, we propose a greedy rate allocation policy and characterize the performance difference of this policy and the optimal policy in terms of channel variations and structure of the utility function. The numerical results 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.

**Index Terms**—Fading channel, multiple access, power control, rate splitting, resource allocation, utility maximization.

## I. INTRODUCTION

**D**YNAMIC allocation of communication resources such as bandwidth or transmission power is a central issue in multiple-access channels (MACs) in view of the time-varying nature of the channel and the interference effects. Most of the

Manuscript received March 28, 2008; revised July 13, 2009. Date of current version August 18, 2010. This work was supported by Stanford University under Contract 18870740-37362-C and by the Defense Advanced Research Projects Agency (DARPA). The material in this paper was presented in part at the 41st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2007. Other parts of this work have been presented at the ITA Workshop, San Diego, CA, 2008 and at WiOpt, Berlin, Germany, 2008.

A. ParandehGheibi, 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 02139-4307 USA (e-mail: parandeh@mit.edu; asuman@mit.edu).

A. Eryilmaz is with the Electrical and Computer Engineering Department, The Ohio State University, Columbus, OH 43210 USA (e-mail: eryilmaz@ecc.osu.edu; medard@mit.edu).

Communicated by S. Ulukus, Associate Editor for Communication Networks.

Color versions of Figures 1–9 in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2010.2053902

existing literature focuses on specific communication schemes such as time-division multiple access (TDMA) [1], code-division multiple access (CDMA) [2], [3], and orthogonal frequency-division multiplexing (OFDM) [4] systems. An exception is the work by Tse and Hanly [5], who 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 code-word length can be arbitrarily long to average over the fading of the channel. Tse and Hanly [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 MAC with perfect CSI. Contrary to the linear case in [5], we consider maximizing a general utility function of transmission rates. Such a general concave utility function allows us to capture different performance metrics such as fairness or delay (cf., [6] and [7]).

Given a utility function, there are different notions of *optimality* for resource allocation policies. Below, we give an overview of three criteria for optimality of a rate allocation policy.

- 1) Long-term optimality: The optimal policy in this case maximizes the utility of the expected achieved rate over the throughput region. This type of metric is interesting when the communication period is significantly large and oscillations in the allocated rate do not matter, e.g., when downloading a large file.

Various works in the literature such as the works by Tse and Hanly [5], Eryilmaz and Srikant [8], and Neely *et al.* [9] consider this notion of optimality.

- 2) Short-term optimality: The optimal policy in this case maximizes the utility function over the instantaneous capacity region at each time slot. This metric is normally employed for delay-sensitive traffic and traffic bursts as well as uncertain environments.

Note that maximizing the expected utility of the allocated rates requires short-term optimality for almost all channel states. Also, we will see later in this paper that short-term optimality criterion coincides with the long-term optimality for linear utility functions.

- 3) Discounted long-term optimality: In this case, the optimal policy maximizes the utility of a discounted average of the allocated rates over the throughput region. This optimality

criterion lies between the two extremes presented above. This criterion allows choosing the discount factor so that we can adapt to the latency requirement of the traffic.

Among several works in the literature addressing this criterion, the works by Agrawal and Subramanian [10] and Stolyar [11] are closely related to our setup. Agrawal and Subramanian [10] develop optimal rate allocation policies under a strict-convexity-type assumption for the capacity region. Stolyar relaxes this assumption in [11] by focusing on a fixed (not time dependent) discount factor, and studying the asymptotic optimality when the discount factor goes to one.

In this paper, we focus on both long-term and short-term optimality criteria of resource allocation policies. Our contributions can be summarized as follows.

We first consider a nonfading MAC 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 vector. Since the capacity region of the MAC 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 [12]. This algorithm also provides the extra information that allows the receiver to decode the message by successive cancellation.

Second, we consider the case where the transmitters do not have the power control feature and channel statistics are not known. We study long-term and short-optimal (greedy) rate allocation policies. We show that the short-term optimal policy, which greedily maximizes the utility function for any given channel state, is suboptimal in the long-term sense for general nonlinear utility functions. However, we can bound the long-term performance difference between the greedy and long-term optimal 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 short-term optimal 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 becomes 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.

Finally, we study jointly optimal rate and power allocation problem in a fading channel 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.

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., [8], [9], [13], and [14]). 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 [8], [9] 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 CSI. 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 [15]). Moreover, the convergence results of queue-length-based policies [8], [9] 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 channel-state-based policies such as the greedy policy seem 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 complexity of the centralized optimization problem of the rate allocation policy [13], [15]–[19]. The effect of these algorithms on the utility achieved is investigated in [15] and [20]. 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.* [21] 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 [22], minimizing the weighted sum of transmission powers [23], and considering quality-of-service

(QoS) constraints [24]. 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 MAC. In Section III, we consider the utility maximization problem in a nonfading channel and present the gradient projection method with approximate projection. In Section IV, we address the optimal rate allocation problem when the transmission powers are fixed and channel statistics are not available. We also present approximate rate allocation policies and study their tracking behavior. In Section V, we generalize the problem to the case where power control is available, and propose jointly optimal rate and power allocation schemes. 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  $\mathbf{x}$ . We write  $\mathbf{x}'$  to denote the transpose of a vector  $\mathbf{x}$ . All vectors in this paper are assumed to be column vectors, and the transpose notation is used for row vectors. We denote the nonnegative orthant by  $\mathbb{R}_+^n$ , i.e.,  $\mathbb{R}_+^n = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq 0\}$ . We use the notation  $\Pr(\cdot)$  for the probability of an event in the Borel  $\sigma$ -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  $\mathbf{x} \in \mathbb{R}^n$ , we have  $\mathcal{P}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{y} \in X} \|\mathbf{x} - \mathbf{y}\|$ , 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 MAC with flat fading effects

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

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 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 CSI. 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:

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

where the  $V_n^i$ 's 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 nonfading case where the channel-state vector is fixed. The capacity region of the Gaussian MAC with no power control is described as follows [25]:

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

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 additive white Gaussian noise (AWGN) channel given by

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

For a MAC 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 [26]

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

where  $\mathbf{H}$  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  $\mathbf{h}$  to the powers allocated to the transmitters  $\boldsymbol{\pi}(\mathbf{h}) = (\pi_1(\mathbf{h}), \dots, \pi_M(\mathbf{h}))$ . Similarly, we can define the rate allocation policy  $\mathcal{R}$  as a function that maps the fading state  $\mathbf{h}$  to the transmission rates  $\mathcal{R}(\mathbf{h})$ . For any given power-control policy  $\boldsymbol{\pi}$ , the capacity region follows from (5) as

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

Tse and Hanly [5] have shown that the throughput capacity of a multiple-access fading channel is given by

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

where  $\mathcal{G} = \{\boldsymbol{\pi} : \mathbb{E}_{\mathbf{H}}[\pi_i(\mathbf{H})] \leq \bar{P}_i, \text{ for all } i\}$  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 GAUSSIAN MAC

In this section, we address the problem of finding the optimal operation rates in a nonfading Gaussian MAC from utility maximization point of view. Without loss of generality, we fix the channel-state vector to unity throughout this section, and denote the capacity region by a simpler notation  $C_g(\mathbf{P})$  instead of  $C_g(\mathbf{P}, \mathbf{1})$ , where  $\mathbf{P} > 0$  denotes the transmission power. A rate vector  $\mathbf{R}^*$  is called *optimal* if it is a solution to the following utility maximization problem for a  $M$ -user channel:

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

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

*Assumption 1:* The following conditions hold.

- The utility function  $u : \mathbb{R}^M \rightarrow \mathbb{R}$  is concave with respect to vector  $\mathbf{R}$ .
- $u(\mathbf{R})$  is monotonically increasing with respect to  $R_i$ , for  $i = 1, \dots, M$ .

*Assumption 2:* There exists a scalar  $B$  such that

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

where  $\partial u(\mathbf{R})$  denotes the subdifferential of  $u$  at  $\mathbf{R}$ , i.e., the set of all subgradients<sup>1</sup> of  $u$  at  $\mathbf{R}$ .

Note that Assumption 2 is standard in the analysis of subgradient methods for nondifferentiable optimization problems [27]. 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 nonnegative 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  $X = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq 0\}$ , where  $A$  is an  $m \times n$  matrix with nonnegative entries. The approximate projection of a vector  $\mathbf{y} \in \mathbb{R}^n$  on  $X$ , denoted by  $\tilde{\mathcal{P}}$ , is given by

$$\tilde{\mathcal{P}}(\mathbf{y}) = \mathcal{P}^+(\mathcal{P}_m(\dots(\mathcal{P}_2(\mathcal{P}_1(\mathbf{y}))))))$$

where  $\mathcal{P}_i$  denotes the exact projection on the half-space  $\{\mathbf{x} \in \mathbb{R}^n | \mathbf{a}_i^T \mathbf{x} \leq b_i\}$ , and  $\mathcal{P}^+$  is projection on the nonnegative orthant  $\mathbb{R}_+^n$ .

An example of approximate projection on a two-user multiple-access capacity region is illustrated in Fig. 1. As shown in the figure, the result of approximate projection is not unique in general, but by definition it terminates in finitely many steps. In order to compute the approximate projection, it is sufficient to successively identify the *violated* constraints and project on their corresponding hyperplanes. In the following, when we

<sup>1</sup>The vector  $\mathbf{g}$  is a subgradient of a concave function  $f : D \rightarrow \mathbb{R}$  at  $\mathbf{x}_0$ , if and only if  $f(\mathbf{x}) - f(\mathbf{x}_0) \leq \mathbf{g}^T(\mathbf{x} - \mathbf{x}_0)$  for all  $\mathbf{x} \in D$ .

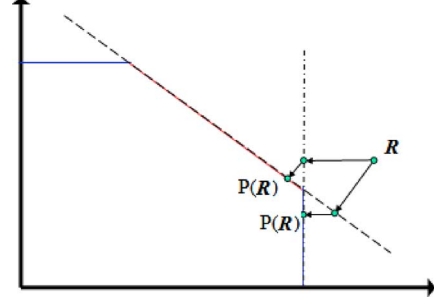


Fig. 1. Approximate projection of  $\mathbf{R}$  on a two-user MAC capacity region.

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 pseudononexpansive as claimed in the following lemma.

*Lemma 1:* The approximate projection  $\tilde{\mathcal{P}}$  on  $C_g(\mathbf{P})$ , the capacity region of the MAC, given by Definition 3 has the following properties.

- For any  $\mathbf{x}$ ,  $\tilde{\mathcal{P}}(\mathbf{x})$  is feasible with respect to  $C_g(\mathbf{P})$ , i.e.,  $\tilde{\mathcal{P}}(\mathbf{x}) \in C_g(\mathbf{P})$ .
- Every  $\mathbf{x} \in C_g(\mathbf{P})$  is a fixed point of  $\tilde{\mathcal{P}}$ , i.e.,  $\tilde{\mathcal{P}}(\mathbf{x}) = \mathbf{x}$ .
- $\tilde{\mathcal{P}}$  is pseudononexpansive, i.e.,

$$\|\tilde{\mathcal{P}}(\mathbf{x}) - \tilde{\mathbf{x}}\| \leq \|\mathbf{x} - \tilde{\mathbf{x}}\|, \quad \text{for all } \tilde{\mathbf{x}} \in C_g(\mathbf{P}). \quad (9)$$

*Proof:* For part (i), note that the constraints defining  $C_g(\mathbf{P})$  are of the form  $A\mathbf{x} \leq \mathbf{b}$  in addition to the nonnegativity constraints, where  $A$  has nonnegative entries. It is straightforward to see that  $\mathcal{P}_i(\mathbf{y})$ , projection of  $\mathbf{y}$  on the half-space  $\mathbf{a}_i^T \mathbf{x} \leq b_i$ , is given by (cf. [28, Sec. II.A.I])

$$\mathcal{P}_i(\mathbf{y}) = \mathbf{y} - \frac{(\mathbf{a}_i^T \mathbf{y} - b_i)^+}{\|\mathbf{a}_i\|} \mathbf{a}_i$$

where  $(z)^+ = \max\{z, 0\}$ . Since  $\mathbf{a}_i$  has only nonnegative entries, no component of  $\mathbf{y}$  is increased after the projection. Hence, the constraint  $i$  will not be violated in the subsequent projections. This shows that given an arbitrary vector  $\mathbf{x}$ , the result of successive projections on the half-spaces corresponding to the constraints  $A\mathbf{x} \leq \mathbf{b}$  is feasible with respect to such constraints, i.e.,  $\mathbf{R} = \mathcal{P}_m(\dots(\mathcal{P}_2(\mathcal{P}_1(\mathbf{x}))))$  satisfies

$$\sum_{i \in S} R_i \leq C \left( \sum_{i \in S} P_i, N_0 \right), \quad \text{for all } S \subseteq \mathcal{M} = \{1, \dots, M\}. \quad (10)$$

Nevertheless,  $\mathbf{R}$  could have negative components. It remains to show that  $\mathbf{R}^+ = \mathcal{P}^+(\mathbf{R}) \in C_g(\mathbf{P})$ . It is clear that

$$\mathbf{R}_i^+ = \begin{cases} R_i, & i \in N^c \\ 0, & i \in N \end{cases}$$

where  $N$  and  $N^c$  are the set of indices of  $\mathbf{R}$  with negative and nonnegative components, respectively. For any  $S \subseteq \mathcal{M}$ , write

$$\begin{aligned} \sum_{i \in S} R_i^+ &= \sum_{i \in S \cap N} R_i^+ + \sum_{i \in S \cap N^c} R_i^+ \\ &= 0 + \sum_{i \in S \cap N^c} R_i \\ &\leq C \left( \sum_{i \in S \cap N^c} P_i, N_0 \right) \leq C \left( \sum_{i \in S} P_i, N_0 \right) \end{aligned} \quad (11)$$

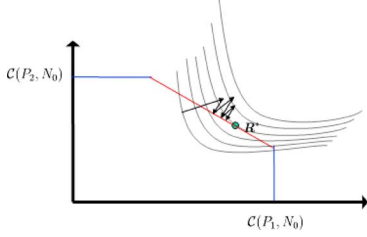


Fig. 2. Gradient projection method with approximate projection on a two-user MAC region.

where the first inequality holds by (10), and the second one is the result of monotonicity of  $C(P, N) = \frac{1}{2} \log(1 + P/N)$  with respect to  $P$ .

Part (ii) is true by definition of  $\tilde{\mathcal{P}}$ , because the set of violated constraints is empty for any feasible point and projection of a feasible point on each half-space gives the same point.

Part (iii) can be verified by successively employing the non-expansiveness property of projection on a closed convex set (see [28, Prop. 2.1.3]). Since  $\tilde{\mathbf{x}}$  is feasible in  $C_g(\mathbf{P})$ , it is a fixed point of  $\mathcal{P}^+$  and  $\mathcal{P}_i$  for all  $i$ . We conclude the claim as follows:

$$\begin{aligned} \|\tilde{\mathcal{P}}(\mathbf{x}) - \tilde{\mathbf{x}}\| &= \|\mathcal{P}^+(\mathcal{P}_m(\dots(\mathcal{P}_2(\mathcal{P}_1(\mathbf{x})))) \\ &\quad - \mathcal{P}^+(\mathcal{P}_m(\dots(\mathcal{P}_2(\mathcal{P}_1(\tilde{\mathbf{x}}))))))\| \\ &\leq \|\mathcal{P}_m(\dots(\mathcal{P}_2(\mathcal{P}_1(\mathbf{x}))) \\ &\quad - \mathcal{P}_m(\dots(\mathcal{P}_2(\mathcal{P}_1(\tilde{\mathbf{x}}))))\| \\ &\quad \vdots \\ &\leq \|\mathcal{P}_1(\mathbf{x}) - \mathcal{P}_1(\tilde{\mathbf{x}})\| \leq \|\mathbf{x} - \tilde{\mathbf{x}}\|. \end{aligned} \quad (12)$$

□

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

$$\mathbf{R}^{k+1} = \tilde{\mathcal{P}}(\mathbf{R}^k + \alpha^k \mathbf{g}^k), \quad \mathbf{g}^k \in \partial u(\mathbf{R}^k) \quad (13)$$

where  $\mathbf{g}^k$  is a subgradient of  $u$  at  $\mathbf{R}^k$ , and  $\alpha^k$  denotes the stepsize. Fig. 2 demonstrates gradient projection iterations for a two-user MAC. The following theorem provides a sufficient condition which can be used to establish convergence of (13) to the optimal solution.

*Theorem 1:* Let Assumptions 1 and 2 hold, and  $\mathbf{R}^*$  be an optimal solution of problem (8). Also, let the sequence  $\{\mathbf{R}^k\}$  be generated by the iteration in (13). If the stepsize  $\alpha^k$  satisfies

$$0 < \alpha^k < \frac{2(u(\mathbf{R}^*) - u(\mathbf{R}^k))}{\|\mathbf{g}^k\|^2} \quad (14)$$

then

$$\|\mathbf{R}^{k+1} - \mathbf{R}^*\| < \|\mathbf{R}^k - \mathbf{R}^*\|. \quad (15)$$

*Proof:* We have

$$\begin{aligned} \|\mathbf{R}^k + \alpha^k \mathbf{g}^k - \mathbf{R}^*\|^2 \\ = \|\mathbf{R}^k - \mathbf{R}^*\|^2 + 2\alpha^k (\mathbf{R}^k - \mathbf{R}^*)' \mathbf{g}^k + (\alpha^k)^2 \|\mathbf{g}^k\|^2. \end{aligned}$$

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

$$(\mathbf{R}^* - \mathbf{R}^k)' \mathbf{g}^k \geq u(\mathbf{R}^*) - u(\mathbf{R}^k). \quad (16)$$

Hence

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

If the stepsize satisfies (14), the above relation yields the following:

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

Now, by applying pseudononexpansiveness of the approximate projection, we have

$$\begin{aligned} \|\mathbf{R}^{k+1} - \mathbf{R}^*\| &= \|\tilde{\mathcal{P}}(\mathbf{R}^k + \alpha^k \mathbf{g}^k) - \mathbf{R}^*\| \\ &\leq \|\mathbf{R}^k + \alpha^k \mathbf{g}^k - \mathbf{R}^*\| < \|\mathbf{R}^k - \mathbf{R}^*\|. \end{aligned}$$

□

*Theorem 2:* Let Assumptions 1 and 2 hold. Also, let the sequence  $\{\mathbf{R}^k\}$  be generated by the iteration in (13). If the stepsize  $\alpha^k$  satisfies (14), then  $\{\mathbf{R}^k\}$  converges to an optimal solution  $\mathbf{R}^*$ .

*Proof:* See [27, Prop. 8.2.7]. □

The convergence analysis for this method can be extended for different stepsize selection rules. For instance, Theorem 1 still holds if we employ the diminishing stepsize (cf., [28, ch. 6]), i.e.,

$$\alpha^k \rightarrow 0, \quad \sum_{k=0}^{\infty} \alpha^k = \infty$$

or more complicated dynamic stepsize selection rules such as the path-based incremental target level algorithm proposed by Brännlund [29]. This stepsize selection rule guarantees convergence to the optimal solution [27], 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 (13) can be performed in polynomial time in  $M$ .

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

$$f(S \cup T) + f(S \cap T) \leq f(S) + f(T), \quad \text{for all } S, T \in 2^{\mathcal{M}}. \quad (17)$$

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

$$f_C(S) = C\left(\sum_{i \in S} P_i, N_0\right), \quad \text{for all } S \subseteq \mathcal{M}. \quad (18)$$

If  $P_i > 0$  for all  $i \in \mathcal{M}$ , then  $f_C(S)$  is submodular. Moreover, the inequality (17) 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 (17). In particular

$$\begin{aligned} & f_C(S) + f_C(T) - f_C(S \cup T) - f_C(S \cap T) \\ &= \frac{1}{2} \log \left[ \frac{(N_0 + \sum_{i \in S} P_i)(N_0 + \sum_{i \in T} P_i)}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\ &= \frac{1}{2} \log \left[ 1 + \frac{\sum_{(i,j) \in (S \setminus T) \times (T \setminus S)} P_i P_j}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\ &\geq 0. \end{aligned} \quad (19)$$

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  $S \subseteq T$ , or  $T \subseteq S$ .  $\square$

*Theorem 3:* For any  $\bar{\mathbf{R}} \in \mathbb{R}_+^M$ , define the *constraint violation* for each constraint of the capacity region (3) as

$$\max \left\{ \sum_{i \in S} \bar{R}_i - C \left( \sum_{i \in S} P_i, N_0 \right), 0 \right\}, \quad S \subseteq \mathcal{M}.$$

Then, the problem of finding the most violated capacity constraint can be written as a submodular function minimization (SFM) problem, which is unconstrained minimization of a submodular function over all  $S \subseteq \mathcal{M}$ .

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

$$f_C(S) - \sum_{i \in S} R_i \geq 0, \quad \text{for all } S \subseteq \mathcal{M}. \quad (20)$$

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

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

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.  $\square$

It was first shown by Grötschel *et al.* [30] that an SFM problem can be solved in polynomial time. There are several fully combinatorial strongly polynomial algorithms in the literature. The best known algorithm for SFM proposed by Orlin [31] 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. This is so since 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 (13), no more than  $M$  constraints can be violated at each iteration. Let us first define the notions of expansion and distance for a polyhedron.

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

$$Q = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}. \quad (21)$$

Define the *expansion* of  $Q$  by  $\delta$ , denoted by  $\mathcal{E}_\delta(Q)$ , as the polyhedron obtained by relaxing all the constraints in (21), i.e.,  $\mathcal{E}_\delta(Q) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b} + \delta \mathbf{1}\}$ , where  $\mathbf{1}$  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 (18). There exists a positive scalar  $\delta$  satisfying

$$\begin{aligned} \delta &\leq \frac{1}{2}(f_C(S) + f_C(T) - f_C(S \cap T) - f_C(S \cup T)), \\ &\quad \text{for all } S, T \in 2^{\mathcal{M}}, \quad S \cap T \neq S, T \end{aligned} \quad (22)$$

and for any such  $\delta$ , the relaxed capacity region  $\mathcal{E}_\delta(C_g(\mathbf{P}))$  of an  $M$ -user MAC violates no more than  $M$  constraints of  $C_g(\mathbf{P})$  defined in (3).

*Proof:* Existence of a positive scalar  $\delta$  satisfying (22) follows directly from Lemma 2, using the fact that neither  $S$  nor  $T$  contains the other one.

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

$$-\sum_{i \in S} R_i < -f_C(S) \quad (23)$$

$$-\sum_{i \in T} R_i < -f_C(T). \quad (24)$$

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

$$\sum_{i \in S \cap T} R_i \leq f_C(S \cap T) + \delta \quad (25)$$

$$\sum_{i \in S \cup T} R_i \leq f_C(S \cup T) + \delta. \quad (26)$$

Note that if  $S \cap T = \emptyset$ , (25) reduces to  $0 \leq \delta$ , which is a valid inequality.

By summing the above inequalities, we conclude

$$\delta > \frac{1}{2}(f_C(S) + f_C(T) - f_C(S \cap T) - f_C(S \cup T)) \quad (27)$$

which is a contradiction.  $\square$

*Theorem 4:* Let Assumptions 1 and 2 hold. Let  $P_1 \leq P_2 \leq \dots \leq P_M$  be the transmission powers.

If the stepsize  $\alpha^k$  in the  $k$ th iteration (13) satisfies

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

for all  $k$  (28)

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

*Proof:* We first show that inequality in (22) holds for the following choice of  $\delta$ :

$$\delta = \frac{1}{4} \log \left[ 1 + \frac{P_1 P_2}{(N_0 + \sum_{i=3}^M P_i)(N_0 + \sum_{i=1}^M P_i)} \right]. \quad (29)$$

In order to verify this, rewrite the right-hand side of (22) as

$$\begin{aligned} & \frac{1}{4} \log \left[ \frac{(N_0 + \sum_{i \in S} P_i)(N_0 + \sum_{i \in T} P_i)}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\ &= \frac{1}{4} \log \left[ 1 + \frac{\sum_{(i,j) \in (S \setminus T) \times (T \setminus S)} P_i P_j}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\ &\geq \frac{1}{4} \log \left[ 1 + \frac{P_1 P_2}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i \in S \cup T} P_i)} \right] \\ &\geq \frac{1}{4} \log \left[ 1 + \frac{P_1 P_2}{(N_0 + \sum_{i \in S \cap T} P_i)(N_0 + \sum_{i=1}^M P_i)} \right] \\ &\geq \frac{1}{4} \log \left[ 1 + \frac{P_1 P_2}{(N_0 + \sum_{i=3}^M P_i)(N_0 + \sum_{i=1}^M P_i)} \right]. \end{aligned}$$

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 nonempty because  $S \cap T \neq S, T$ .

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

$$\begin{aligned} \sum_{i \in S} (R_i^k + \alpha^k g_i^k) &= \sum_{i \in S} R_i^k + \alpha^k \|g^k\| \sum_{i \in S} \frac{g_i^k}{\|g^k\|} \\ &\leq f_C(S) + \frac{\delta}{B\sqrt{M}} B \sum_{i \in S} \frac{g_i^k}{\|g^k\|} \\ &\leq f_C(S) + \delta \end{aligned} \quad (30)$$

where the first inequality follows from Assumption 1(b), Assumption 2, and (28). The second inequality holds because for any unit vector  $\mathbf{d} \in \mathbb{R}^M$ , it is true that

$$\sum_{i \in S} d_i \leq \sum_{i \in S} |d_i| \leq \sqrt{M}. \quad (31)$$

Thus, if  $\alpha^k$  satisfies (28), then  $(\mathbf{R}^k + \alpha^k \mathbf{g}^k) \in \mathcal{E}_\delta(C_g(\mathbf{P}))$ , for some  $\delta$  for which (22) holds. Therefore, by Lemma 3, the number of violated constraints does not exceed  $M$ .  $\square$

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

#### IV. DYNAMIC RATE ALLOCATION IN FADING CHANNELS

In this part, we study the rate allocation problem for a fading channel when transmission powers are fixed to  $\mathbf{P}$ . 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. Throughout this section, we also assume that the channel statistics are not known.<sup>2</sup> The capacity region of the fading MAC for this scenario is a polyhedron given by (5).

We study both long-term and short-term optimal rate allocation policies with respect to a given utility function, which we formally define next. We show that the short-term optimal and long-term optimal policies coincide if the utility function is linear. Moreover, we show that the long-term performance of the short-term policy is close to the long-term optimal policy. The rest of Section IV is dedicated to efficiently computing the short-term optimal policy.

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

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}[\mathcal{R}^*(\mathbf{H})] &= \mathbf{R}^* \in \operatorname{argmax}_{\mathbf{R}} u(\mathbf{R}) \\ &\text{subject to } \mathbf{R} \in C_a(\mathbf{P}). \end{aligned} \quad (32)$$

*Definition 8 (Short-Term Optimal Policy):* A short-term optimal or *greedy* rate allocation policy,<sup>3</sup> denoted by  $\bar{\mathcal{R}}$ , is given by

$$\begin{aligned} \bar{\mathcal{R}}(\mathbf{H}) &= \operatorname{argmax}_{\mathbf{R}} u(\mathbf{R}) \\ &\text{subject to } \mathbf{R} \in C_g(\mathbf{P}, \mathbf{H}) \end{aligned} \quad (33)$$

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  $u(\mathbf{R})$  is assumed to satisfy the following conditions.

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

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

$$\begin{aligned} |u(\mathbf{R}_1) - u(\mathbf{R}_2)| &\leq B(\delta) \|\mathbf{R}_1 - \mathbf{R}_2\|, \\ &\text{for all } \mathbf{R}_i, \|\mathbf{R}_i\|_1 \geq D_\delta, \quad i = 1, 2 \end{aligned}$$

where

$$D_\delta = \inf_{\mathbf{H} \in \mathcal{N}_\delta} \sup_{\mathbf{R} \in C_g(\mathbf{P}, \mathbf{H})} \|\mathbf{R}\|_1. \quad (34)$$

<sup>2</sup>We could also develop optimal rate allocation algorithms for the case where powers are fixed and channel statistics are known. See [32] for more details on this scenario.

<sup>3</sup>We use the terms short-term optimal policy and greedy policy interchangeably throughout this paper.



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

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

Assumption 3(a) is a weakened version of Assumption 2, which imposes a bound on subgradients of the utility function. This assumption only requires a 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  $\alpha$ -fair functions given by

$$f_\alpha(x) = \begin{cases} \frac{x^{1-\alpha}}{1-\alpha}, & \alpha \neq 1 \\ \log(x), & \alpha = 1 \end{cases} \quad (35)$$

which do not satisfy Assumption 2.

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

$$\begin{aligned} \mathbb{E}_{\mathbf{H}} [u(\mathcal{R}^*(\mathbf{H}))] &\leq \mathbb{E}_{\mathbf{H}} [u(\bar{\mathcal{R}}(\mathbf{H}))] \\ &\leq u(\mathbb{E}_{\mathbf{H}} [\bar{\mathcal{R}}(\mathbf{H})]) \\ &\leq u(\mathbb{E}_{\mathbf{H}} [\mathcal{R}^*(\mathbf{H})]) \end{aligned} \quad (36)$$

where the first and third inequalities follow from the feasibility of the long-term 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  $u(\mathbb{E}_{\mathbf{H}} [\mathcal{R}^*(\mathbf{H})]) = \mathbb{E}_{\mathbf{H}} [u(\mathcal{R}^*(\mathbf{H}))]$ , so equality holds throughout in (36) and  $\mathcal{R}(\cdot)$  is indeed long-term optimal as well as being short-term optimal. For nonlinear utility functions, the greedy policy can be strictly suboptimal in the long term.

However, the greedy policy is not arbitrarily worse than the long-term optimal one. In view of (36), we can bound the performance difference  $u(\mathbf{R}^*) - u(\mathbb{E}_{\mathbf{H}} [\bar{\mathcal{R}}(\mathbf{H})])$  by bounding  $|u(\mathbb{E}_{\mathbf{H}} [\mathcal{R}^*(\mathbf{H})]) - u(\mathbb{E}_{\mathbf{H}} [\bar{\mathcal{R}}(\mathbf{H})])|$  or  $|u(\mathbb{E}_{\mathbf{H}} [\mathcal{R}^*(\mathbf{H})]) - \mathbb{E}_{\mathbf{H}} [u(\mathcal{R}^*(\mathbf{H}))]|$  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 both long-term

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 long-term and short-term optimal rate allocation policies as in Definitions 7 and 8, respectively. Then:

(a)  $\bar{\mathcal{R}}(\mathbf{H}) \in \mathcal{F}(C_g(\mathbf{P}, \mathbf{H}))$ , for all  $\mathbf{H}$ ;

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

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

$$\mathbf{R}^* = \mathbb{E}_{\mathbf{H}} [\mathcal{R}^*(\mathbf{H})] \in \mathcal{F}(C_a(\mathbf{P})). \quad (37)$$

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

$$\mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^M \mathcal{R}_i^*(\mathbf{H}) \right] = \mathbb{E}_{\mathbf{H}} \left[ C \left( \sum_{i=1}^M H_i P_i, N_0 \right) \right]. \quad (38)$$

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

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  $\mathbf{H}$  be a random vector with the stationary distribution of the channel-state process, mean  $\bar{\mathbf{H}}$ , and covariance matrix  $\mathbf{K}$ . Then

$$\Pr\left\{d_H(C_g(\mathbf{P}, \mathbf{H}), C_a(\mathbf{P})) > \delta\right\} \leq \frac{\sigma_{\bar{\mathbf{H}}}^2}{\delta^2} \quad (39)$$

where  $\sigma_{\bar{\mathbf{H}}}^2$  is defined as (40) shown at the bottom of the page, in which  $\mathbf{\Gamma}_S$  is given by

$$(\mathbf{\Gamma}_S)_i = \begin{cases} \frac{P_i}{N_0}, & i \in S \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

*Proof:* See Appendix II.  $\square$

The system parameter  $\sigma_{\bar{\mathbf{H}}}^2$  in Lemma 5 is proportional to channel variations, and we expect it to vanish for very small

$$\sigma_{\bar{\mathbf{H}}}^2 \triangleq \frac{1}{4} \sum_{S \subseteq \{1, \dots, M\}} \mathbf{\Gamma}'_S \mathbf{K} \mathbf{\Gamma}_S \left( 1 + \left[ (1 + \mathbf{\Gamma}'_S \bar{\mathbf{H}}) (\sqrt{2 \log(1 + \mathbf{\Gamma}'_S \bar{\mathbf{H}})} - \frac{\sqrt{\mathbf{\Gamma}'_S \mathbf{K} \mathbf{\Gamma}_S}}{2}) \right]^2 \right) \quad (40)$$

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  $u : \mathbb{R}^M \rightarrow \mathbb{R}$  satisfy Assumptions 1 and 3. Also, let  $\mathbf{R}_1^*$  and  $\mathbf{R}_2^*$  be the optimal solutions of maximizing the utility over  $C_a(\mathbf{P})$  and  $C_g(\mathbf{P}, \mathbf{H})$ , respectively. If

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

then, we have

$$\|\mathbf{R}_1^* - \mathbf{R}_2^*\| \leq \delta^{\frac{1}{2}} \left[ \delta^{\frac{1}{2}} + \left( \frac{B(\delta)}{A(\delta)} \right)^{\frac{1}{2}} \right]. \quad (42)$$

*Proof:* See Appendix III.  $\square$

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

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

$$\begin{aligned} & u(\mathbf{R}^*) - u(\mathbb{E}_{\mathbf{H}}[\bar{\mathcal{R}}(\mathbf{H})]) \\ & \leq \frac{\sigma_H^2}{\delta^2} u(\mathbf{R}^*) + \left(1 - \frac{\sigma_H^2}{\delta^2}\right) B(\delta) \left[ \delta^{\frac{1}{2}} + \left( \frac{B(\delta)}{A(\delta)} \right)^{\frac{1}{2}} \right] \delta^{\frac{1}{2}} \end{aligned} \quad (43)$$

where  $\mathbf{R}^* = \mathbb{E}_{\mathbf{H}}[\mathcal{R}^*(\mathbf{H})]$ , 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

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

By Lemma 5, the probability of this event is at least  $1 - \frac{\sigma_H^2}{\delta^2}$ . Using Jensen's inequality as in (36) we can bound the left-hand side of (43) as follows:

$$\begin{aligned} & u(\mathbf{R}^*) - u(\mathbb{E}_{\mathbf{H}}[\bar{\mathcal{R}}(\mathbf{H})]) \\ & \leq u(\mathbf{R}^*) - \mathbb{E}_{\mathbf{H}}[u(\bar{\mathcal{R}}(\mathbf{H}))] \\ & \stackrel{(a)}{\leq} u(\mathbf{R}^*) - \left(1 - \frac{\sigma_H^2}{\delta^2}\right) \mathbb{E}_{\mathbf{H}}[u(\bar{\mathcal{R}}(\mathbf{H}))|\mathcal{V}] \\ & \quad - \Pr(\mathcal{V}^c) \mathbb{E}_{\mathbf{H}}[u(\bar{\mathcal{R}}(\mathbf{H}))|\mathcal{V}^c] \\ & \stackrel{(b)}{\leq} \frac{\sigma_H^2}{\delta^2} u(\mathbf{R}^*) + \left(1 - \frac{\sigma_H^2}{\delta^2}\right) (u(\mathbf{R}^*) - \mathbb{E}_{\mathbf{H}}[u(\bar{\mathcal{R}}(\mathbf{H}))|\mathcal{V}]) \\ & \leq \frac{\sigma_H^2}{\delta^2} u(\mathbf{R}^*) + \left(1 - \frac{\sigma_H^2}{\delta^2}\right) |\mathbb{E}_{\mathbf{H}}[u(\mathbf{R}^*) - u(\bar{\mathcal{R}}(\mathbf{H}))|\mathcal{V}]| \\ & \leq \frac{\sigma_H^2}{\delta^2} u(\mathbf{R}^*) + \left(1 - \frac{\sigma_H^2}{\delta^2}\right) \mathbb{E}_{\mathbf{H}}[|u(\mathbf{R}^*) - u(\bar{\mathcal{R}}(\mathbf{H}))|\mathcal{V}] \end{aligned} \quad (44)$$

where (a) follows from the fact that  $\Pr(\mathcal{V}) \geq 1 - \frac{\sigma_H^2}{\delta^2}$ , and (b) holds by nonnegativity of  $u(\mathbf{R})$ .

On the other hand, by definition of  $D_\delta$  in (34) for any  $\mathbf{H} \in \mathcal{V}$ , write

$$D_\delta = \inf_{\bar{\mathbf{H}} \in \mathcal{V}} \sup_{\mathbf{R} \in C_g(\mathbf{P}, \bar{\mathbf{H}})} \|\mathbf{R}\|_1 \leq \sup_{\mathbf{R} \in C_g(\mathbf{P}, \mathbf{H})} \|\mathbf{R}\|_1 = \|\bar{\mathcal{R}}(\mathbf{H})\|_1$$

where the equality follows from Lemma 4(a) and the fact that the constraint  $\sum_{i=1}^M R_i \leq C(\sum_{i=1}^M H_i P_i)$  is active for any point on the dominant face. The above relation allows us to use Assumption 3(a), which gives

$$|u(\mathbf{R}^*) - u(\bar{\mathcal{R}}(\mathbf{H}))| \leq B(\delta) \|\bar{\mathcal{R}}(\mathbf{H}) - \mathbf{R}^*\|, \quad \text{for all } \mathbf{H} \in \mathcal{V}.$$

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

$$\begin{aligned} & |u(\mathbf{R}^*) - u(\bar{\mathcal{R}}(\mathbf{H}))| \\ & \leq B(\delta) \left( \delta^{\frac{1}{2}} + \left( \frac{B(\delta)}{A(\delta)} \right)^{\frac{1}{2}} \right) \delta^{\frac{1}{2}}, \quad \text{for all } \mathbf{H} \in \mathcal{V} \end{aligned}$$

which implies

$$\begin{aligned} & \mathbb{E}_{\mathbf{H}} \left[ |u(\mathbf{R}^*) - u(\bar{\mathcal{R}}(\mathbf{H}))| \mid \mathcal{V} \right] \\ & \leq B(\delta) \left( \delta^{\frac{1}{2}} + \left( \frac{B(\delta)}{A(\delta)} \right)^{\frac{1}{2}} \right) \delta^{\frac{1}{2}}. \end{aligned} \quad (45)$$

The desired result follows immediately from substituting (45) in (44).  $\square$

Theorem 5 provides a bound parameterized by  $\delta$ . For very small channel variations,  $\sigma_H$  becomes small. Therefore, the parameter  $\delta$  can be picked small enough such that the bound in (43) tends to zero. Fig. 3 illustrates the behavior of the right-hand side of (43) as a function of  $\delta$  for different values of  $\sigma_H$ . For each value of  $\sigma_H$ , the upper bound is minimized for a specific choice of  $\delta$ , which is illustrated by a dot in Fig. 3. As demonstrated in the figure, for smaller channel variations, a smaller gap is achieved and the parameter  $\delta$  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  $u : \mathbb{R}^M \rightarrow \mathbb{R}_+$ . Also, let  $\mathcal{R}^*(\cdot)$  and  $\bar{\mathcal{R}}(\cdot)$  be the long-term and short-term optimal rate allocation policies, defined in Definitions 7 and 8, respectively. Then, for every  $\epsilon \in (0, 1]$

$$u(\mathbf{R}^*) - u(\mathbb{E}_{\mathbf{H}}[\bar{\mathcal{R}}(\mathbf{H})]) \leq \epsilon u(\mathbf{R}^*) + \frac{1}{2}(1 - \epsilon)r(\epsilon)^2 \Omega \quad (46)$$

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

$$\lambda_{\max}(-\nabla^2 u(\boldsymbol{\xi})) \leq \Omega, \quad \text{for all } \boldsymbol{\xi}, \|\boldsymbol{\xi} - \mathbf{R}^*\| \leq r(\epsilon) \quad (47)$$

in which  $\nabla^2$  denotes the Hessian of  $u$ ,  $\lambda_{\max}(Z)$  is the largest eigenvalue of matrix  $Z$ , and  $r(\epsilon)$  is given by

$$r(\epsilon) = \sqrt{M} \frac{\sigma_H}{\sqrt{\epsilon}} + \left[ \sum_{i=1}^M \mathbb{E}_{\mathbf{H}} \left[ \frac{1}{2} \log \left( \frac{(1 + H_i P_i)(1 + \sum_{j \neq i} H_j P_j)}{1 + \sum_{j=1}^M H_j P_j} \right) \right]^2 \right]^{\frac{1}{2}}. \quad (48)$$

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

$$\mathcal{V} = \left\{ \mathbf{H} : d_H(C_g(\mathbf{P}, \mathbf{H}), C_a(\mathbf{P})) \leq \frac{\sigma_H}{\sqrt{\epsilon}} \right\}. \quad (49)$$

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

$$\begin{aligned} \mathbb{E}_{\mathbf{H}} \left[ \frac{1}{2} \log \left( 1 + \frac{H_i P_i}{1 + \sum_{j \neq i} H_j P_j} \right) \right] - \frac{\sigma_H}{\sqrt{\epsilon}} \\ \leq \mathcal{R}_i^*(\mathbf{H}) \leq \mathbb{E}_{\mathbf{H}} \left[ \frac{1}{2} \log(1 + H_i P_i) \right] + \frac{\sigma_H}{\sqrt{\epsilon}}. \end{aligned} \quad (50)$$

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

$$\begin{aligned} |\mathcal{R}_i^*(\mathbf{H}) - R_i^*| \\ \leq \frac{\sigma_H}{\sqrt{\epsilon}} + \mathbb{E}_{\mathbf{H}} \left[ \frac{1}{2} \log \left( \frac{(1 + H_i P_i)(1 + \sum_{j \neq i} H_j P_j)}{1 + \sum_{j=1}^M H_j P_j} \right) \right]. \end{aligned}$$

Therefore

$$\begin{aligned} \|\mathcal{R}^*(\mathbf{H}) - \mathbf{R}^*\| \\ \leq \sqrt{M} \frac{\sigma_H}{\sqrt{\epsilon}} \\ + \left[ \sum_{i=1}^M \mathbb{E}_{\mathbf{H}} \left[ \frac{1}{2} \log \left( \frac{(1 + H_i P_i)(1 + \sum_{j \neq i} H_j P_j)}{1 + \sum_{j=1}^M H_j P_j} \right) \right]^2 \right]^{\frac{1}{2}} \\ = r(\epsilon), \quad \text{for almost all } \mathbf{H} \in \mathcal{V}. \end{aligned} \quad (51)$$

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

$$\begin{aligned} u(\mathbf{R}) &= u(\mathbf{R}^*) + \nabla u(\mathbf{R}^*)'(\mathbf{R} - \mathbf{R}^*) \\ &\quad - \frac{1}{2}(\mathbf{R} - \mathbf{R}^*)'(-\nabla^2 u(\xi))(\mathbf{R} - \mathbf{R}^*) \\ &\geq u(\mathbf{R}^*) + \nabla u(\mathbf{R}^*)'(\mathbf{R} - \mathbf{R}^*) \\ &\quad - \frac{1}{2}\|\mathbf{R} - \mathbf{R}^*\|^2 \lambda_{\max}(-\nabla^2 u(\xi)), \\ &\quad \text{for some } \xi, \|\xi - \mathbf{R}^*\| \leq \|\mathbf{R} - \mathbf{R}^*\|. \end{aligned} \quad (52)$$

In the above relation, let  $\mathbf{R} = \mathcal{R}^*(\mathbf{H})$  for all  $\mathbf{H} \in \mathcal{V}$ . The utility function is concave, so its Hessian is negative definite and we can combine (51) with the above relation to write

$$u(\mathcal{R}^*(\mathbf{H})) \geq u(\mathbf{R}^*) + \nabla u(\mathbf{R}^*)'(\mathcal{R}^*(\mathbf{H}) - \mathbf{R}^*) - \frac{1}{2}r(\epsilon)^2 \Omega, \quad \text{for almost all } \mathbf{H} \in \mathcal{V}. \quad (53)$$

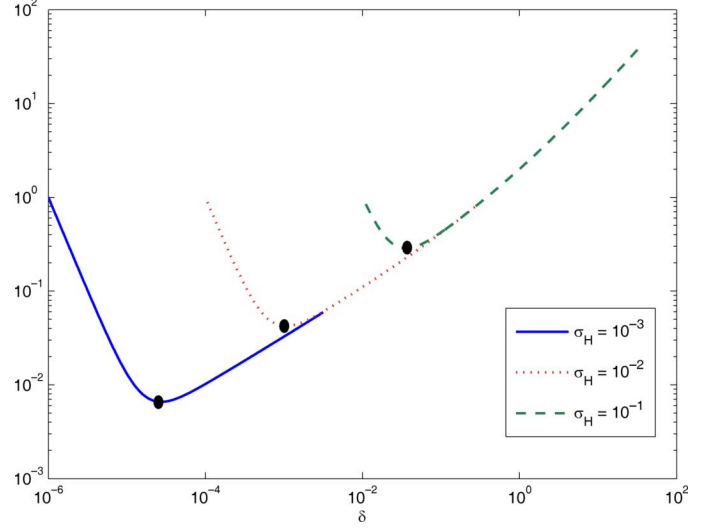


Fig. 3. Parametric upper bound on performance difference between greedy and long-term optimal policies as in the right-hand side of (43) for different channel variations  $\sigma_H$  as a function of  $\delta$ .

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

$$\mathbb{E}_{\mathbf{H}} [u(\mathcal{R}^*(\mathbf{H})) | \mathcal{V}] \geq u(\mathbf{R}^*) - \frac{1}{2}r(\epsilon)^2 \Omega. \quad (54)$$

Hence, we conclude

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

where the first inequality is verified by (36), and the third inequality follows from nonnegativity of the utility function and the inequality in (54).  $\square$

Similarly to Theorem 5, Theorem 6 provides a bound parameterized by  $\epsilon$ . As the utility function tends to have a more linear structure,  $\Omega$  tends to zero. For instance,  $\Omega$  is proportional to  $\alpha$  for a weighted sum  $\alpha$ -fair utility function. Hence, we can choose  $\epsilon$  small such that the right-hand side of (46) goes to zero. The behavior of this upper bound for different values of  $\Omega$  is similar to the one plotted in Fig. 3.

In summary, the performance difference between the greedy (short-term optimal) and the long-term optimal policy is bounded from above by the minimum of the bounds provided by Theorems 5 and 6. Since the greedy policy is short-term optimal and can perform closely to the long-term optimal policy, we focus on developing efficient algorithms to compute the greedy policy.

The greedy policy (cf., Definition 8) requires solving a non-linear program in each time slot. For each channel state, finding even a near-optimal solution of the problem in (33) 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 (13) per time slot.

#### A. Approximate Rate Allocation Policy

In this part, we assume that the CSI is available at each time slot  $n$ , and the computational resources are limited such that a single iteration of the gradient projection method in (13) 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  $\mathbf{R}^\dagger = \operatorname{argmax}_{\mathbf{R} \in C_g(\mathbf{P}, \mathbf{H})} u(\mathbf{R})$ . Then, there exists a positive scalar  $A$  such that

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

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

$$\tilde{\mathcal{R}}(\mathbf{H}(n)) \triangleq \begin{cases} \tilde{\mathcal{R}}(\mathbf{H}(0)), & n = 0 \\ \tilde{\mathbf{R}}_{t(n)}^\tau, & n \geq 1 \end{cases} \quad (55)$$

where

$$\tau = \operatorname{argmax}_{0 \leq j < k-1} u(\tilde{\mathbf{R}}_{t(n)}^j), \quad t(n) = \left\lfloor \frac{n-1}{k} \right\rfloor \quad (56)$$

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

$$\begin{aligned} \tilde{\mathbf{R}}_{t(n)}^0 &= \tilde{P}_{t(n)} \left[ \tilde{\mathcal{R}}(\mathbf{H}(kt(n))) \right] \\ \tilde{\mathbf{R}}_{t(n)}^{j+1} &= \tilde{P}_{t(n)} \left[ \tilde{\mathbf{R}}_{t(n)}^j + \alpha^j \tilde{\mathbf{g}}_{t(n)}^j \right], \quad j = 1, \dots, k-1 \end{aligned} \quad (57)$$

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

For  $k = 1$ , (57) 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 CSI at time slots  $kt, k(t-1), \dots$ . 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  $\{[H_i(n)]_{i=1, \dots, M}\}$  be the fading process that satisfies condition in (2). We have

$$d_H(C_g(\mathbf{P}, \mathbf{H}(n+1)), C_g(\mathbf{P}, \mathbf{H}(n))) \leq W_n \quad (58)$$

where  $\{W_n\}$  are nonnegative independent identically distributed (i.i.d.) 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

$$\begin{aligned} d_H(C_g(\mathbf{P}, \mathbf{H}(n+1)), C_g(\mathbf{P}, \mathbf{H}(n))) &= \max_{S \subseteq \mathcal{M}} \frac{1}{2} \left| \log \left( 1 + \frac{\sum_{i \in S} (H_i(n+1) - H_i(n)) P_i}{1 + \sum_{i \in S} H_i(n) P_i} \right) \right| \\ &\leq \max_{S \subseteq \mathcal{M}} \frac{\sum_{i \in S} |H_i(n+1) - H_i(n)| P_i}{2(1 + \sum_{i \in S} H_i(n) P_i)} \\ &\leq \frac{1}{2} \sum_{i=1}^M |H_i(n+1) - H_i(n)| P_i = \frac{1}{2} \sum_{i=1}^M V_n^i P_i. \end{aligned} \quad (59)$$

Therefore, (58) 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  $\frac{1}{2} \sum_{i=1}^M \hat{v}^i P_i$ .  $\square$

The following useful lemma by Nedić and Bertsekas [33] addresses the convergence rate of the gradient projection method with constant stepsize.

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

$$u\left(\tilde{\mathcal{R}}(\mathbf{H}(n))\right) \geq u\left(\tilde{\mathcal{R}}(\mathbf{H}(kt))\right) - \frac{\alpha B^2 + \epsilon}{2} \quad (60)$$

if  $k$  satisfies

$$k \geq \left\lceil \frac{\|\tilde{\mathbf{R}}_t^0 - \tilde{\mathcal{R}}(\mathbf{H}(kt))\|^2}{\alpha \epsilon} \right\rceil \quad (61)$$

where  $t = t(n) = \lfloor \frac{n-1}{k} \rfloor$ .

*Proof:* See [33, Prop. 2.3].  $\square$

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  $\tilde{\mathcal{R}}$  and  $\tilde{\mathcal{R}}$  be given by Definitions 8 and 9,

respectively. Choose the system parameters  $k$  and  $\alpha$  for the approximate policy in Definition 9 as

$$k = \left\lceil \left( \frac{2B}{Aw'} \right)^{\frac{2}{3}} \right\rceil, \quad \alpha = \left( \frac{16}{B^2A} \right)^{\frac{1}{3}} w'^{\frac{2}{3}}$$

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

$$\|\tilde{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(n))\| \leq 2\theta = 2 \left( \frac{2B}{A} \right)^{\frac{2}{3}} w'^{\frac{4}{3}}. \quad (62)$$

*Proof:* First, we show that

$$\|\tilde{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(kt))\| \leq \theta = \left( \frac{2B}{A} \right)^{\frac{2}{3}} w'^{\frac{4}{3}} \quad (63)$$

where  $t = \lfloor \frac{n-1}{k} \rfloor$ . The proof is by induction on  $t$ . For  $t = 0$ , note that  $\tilde{\mathcal{R}}(\mathbf{H}(n))$  is the result of applying  $k$  steps of gradient projection starting from the optimal solution  $\bar{\mathcal{R}}(\mathbf{H}(0))$ . Hence

$$\tilde{\mathcal{R}}(\mathbf{H}(n)) = \bar{\mathcal{R}}(\mathbf{H}(0)), \quad 0 \leq n \leq k.$$

Thus, the claim is trivially true for  $t = 0$ . Now, suppose that (63) is true for some positive  $t$ . Hence, it also holds for  $n = k(t+1)$  by induction hypothesis, i.e.,

$$\|\tilde{\mathcal{R}}_{t+1}^0 - \bar{\mathcal{R}}(\mathbf{H}(kt))\| \leq \theta. \quad (64)$$

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

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

Thus, by Lemma 6 and the triangle inequality, we have

$$\|\bar{\mathcal{R}}(\mathbf{H}(k(t+1))) - \bar{\mathcal{R}}(\mathbf{H}(kt))\| \leq kw' \leq \theta. \quad (65)$$

Therefore, by another triangle inequality, we conclude from (64) and (65) that the initial point  $\tilde{\mathcal{R}}_{t+1}^0$  for the round  $t+1$  of the iterations is close to the optimal solution  $\bar{\mathcal{R}}(\mathbf{H}(k(t+1)))$ , in particular

$$\|\tilde{\mathcal{R}}_{t+1}^0 - \bar{\mathcal{R}}(\mathbf{H}(k(t+1)))\| \leq 2\theta. \quad (66)$$

Now, we show that for the given value of the stepsize  $\alpha$ , the number of gradient projection steps  $k$  satisfies (61) for  $\epsilon = \alpha B^2$ .

By (66) and plugging the corresponding values of  $\alpha$  and  $\theta$ , we get

$$\begin{aligned} \left\lceil \frac{\|\tilde{\mathcal{R}}_{t+1}^0 - \bar{\mathcal{R}}(\mathbf{H}(k(t+1)))\|^2}{\alpha\epsilon} \right\rceil &\leq \left\lceil \frac{4\theta^2}{\alpha^2 B^2} \right\rceil \\ &= \left\lceil \frac{4 \left( \frac{2B}{A} \right)^{\frac{4}{3}} w'^{\frac{16}{3}}}{\left( \frac{16}{B^2A} \right)^{\frac{2}{3}} w'^{\frac{4}{3}} B^2} \right\rceil \\ &= \left\lceil \left( \frac{2B}{Aw'} \right)^{\frac{2}{3}} \right\rceil = k. \end{aligned}$$

Thus, we can apply Lemma 8 to show

$$\left| u \left( \tilde{\mathcal{R}}(\mathbf{H}(n)) \right) - u \left( \bar{\mathcal{R}}(\mathbf{H}(k(t+1))) \right) \right| \leq \alpha B^2. \quad (67)$$

By Assumption 4, we can write

$$\|\tilde{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(k(t+1)))\| \leq \left( \frac{\alpha B^2}{A} \right)^{\frac{1}{2}} = \theta. \quad (68)$$

Therefore, the proof of (63) is complete by induction.

Similarly to the derivation of (65), by applying Lemmas 6 and 7, we get

$$\|\bar{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(kt))\| \leq kw' \leq \theta \quad (69)$$

and the desired result directly follows from (63) and (69) using the triangle inequality one last time.  $\square$

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 (55) which uses the CSI once in every  $k$  time slots, we present an algorithm which uses the CSI 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 nonnegative random variables as defined in Lemma 7, and  $\gamma$  be a positive constant. Define the sequence  $\{T_i\}$  as

$$\begin{aligned} T_0 &= 0 \\ T_{i+1} &= \min \left\{ t \mid \sum_{n=T_i}^{t-1} W_n \geq \gamma \right\}. \end{aligned} \quad (70)$$

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

$$\hat{\mathcal{R}}(\mathbf{H}(n)) \triangleq \begin{cases} \bar{\mathcal{R}}(\mathbf{H}(0)), & n = 0 \\ \hat{\mathcal{R}}_{t(n)}^\tau, & n \geq 1 \end{cases} \quad (71)$$

where

$$t(n) = \max\{i \mid T_i < n\} \quad (72)$$

$$\tau = \arg \max_{0 \leq j < k-1} u \left( \hat{\mathcal{R}}_{t(n)}^j \right) \quad (73)$$

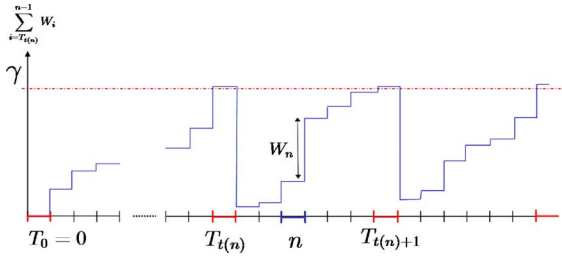


Fig. 4. Improved approximate policy takes  $k$  gradient projection iterations at time  $T_{t(n)}$ , which is the time that the random walk generated by the random variables  $W_n$  reaches the threshold  $\gamma$ .

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

$$\begin{aligned} \hat{\mathbf{R}}_{t(n)}^0 &= \tilde{P}_{t(n)} \left[ \hat{\mathcal{R}}(\mathbf{H}(T_{t(n)})) \right] \\ \hat{\mathbf{R}}_{t(n)}^{j+1} &= \tilde{P}_{t(n)} \left[ \hat{\mathbf{R}}_{t(n)}^j + \alpha^j \hat{\mathbf{g}}_{t(n)}^j \right], \quad j=1, \dots, k-1 \end{aligned} \quad (74)$$

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

Fig. 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 (72), and let  $\bar{w} = \mathbb{E}[W_n]$ . If  $k = \frac{\gamma}{\bar{w}}$ , then we have

$$\lim_{n \rightarrow \infty} \frac{n}{t(n)k} = 1, \quad \text{with probability 1.} \quad (75)$$

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

We can rewrite the limit as follows:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n - N(t(n)\gamma) + N(t(n)\gamma)}{t(n)k} \\ = \lim_{n \rightarrow \infty} \frac{n - N(t(n)\gamma)}{t(n)k} + \bar{w} \frac{N(t(n)\gamma)}{t(n)\gamma}. \end{aligned} \quad (76)$$

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 [34, p. 60]).  $\square$

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  $\hat{\mathcal{R}}$  be given by Definitions 8 and 10, respectively. Also, let  $k = \lfloor \frac{\gamma}{\bar{w}} \rfloor$ , and fix the stepsize to  $\alpha = \frac{A\gamma^2}{B^2}$  in (74), where  $\gamma = c \left(\frac{B}{A}\right)^{\frac{3}{4}} \bar{w}^{\frac{1}{4}}$ , and  $c \geq 1$  is a constant satisfying

$$\left(\frac{B}{A}\right) \frac{(c^2 - 1)^8}{2^8 c^4} = \hat{w}. \quad (77)$$

Then

$$\|\hat{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(n))\| \leq 2\gamma + \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}}. \quad (78)$$

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

$$\|\hat{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(T_t))\| \leq \gamma \quad (79)$$

where  $t$  is defined in (72). The base is trivial. Similar to (64), by induction hypothesis, we have

$$\|\hat{\mathbf{R}}_{t+1}^0 - \bar{\mathcal{R}}(\mathbf{H}(T_t))\| \leq \gamma. \quad (80)$$

By definition of  $T_i$  in (70), we can write

$$d_H \left( C_g \left( \mathbf{P}, \mathbf{H}(T_{t+1}) \right), C_g \left( \mathbf{P}, \mathbf{H}(T_t) \right) \right) \leq \gamma. \quad (81)$$

Thus, by Lemma 6, we have

$$\|\bar{\mathcal{R}}(\mathbf{H}(T_{t+1})) - \bar{\mathcal{R}}(\mathbf{H}(T_t))\| \leq \gamma^{\frac{1}{2}} \left( \gamma^{\frac{1}{2}} + \left(\frac{B}{A}\right)^{\frac{1}{2}} \right). \quad (82)$$

Therefore, by combining (80) and (82) by triangle inequality, we obtain

$$\|\hat{\mathbf{R}}_{t+1}^0 - \bar{\mathcal{R}}(\mathbf{H}(T_{t+1}))\| \leq 2\gamma + \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}}. \quad (83)$$

Using the fact that  $\bar{w} \leq \hat{w} = \left(\frac{B}{A}\right) \frac{(c^2 - 1)^8}{2^8 c^4}$ , we can provide a simpler bound for the right-hand side of (83) as follows:

$$\begin{aligned} \gamma^4 &= c^4 \left(\frac{B}{A}\right)^3 \bar{w} \leq c^4 \left(\frac{B}{A}\right)^3 \cdot \left(\frac{B}{A}\right) \frac{(c^2 - 1)^8}{2^8 c^4} \\ &= \left[ \left(\frac{B}{A}\right)^{\frac{1}{2}} \frac{(c^2 - 1)}{2} \right]^8 \end{aligned} \quad (84)$$

that implies

$$\gamma \leq \frac{(c^2 - 1)}{2} \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}} = \frac{c^2}{2} \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}} - \frac{1}{2} \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}}$$

which gives the following bound on the right-hand side of (83) after rearranging the terms:

$$2\gamma + \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}} \leq c^2 \left(\frac{\gamma B}{A}\right)^{\frac{1}{2}}.$$

Now by plugging the values of  $\alpha$  and  $\gamma$  in terms of system parameters in (61), we can verify that

$$k = \left\lfloor \frac{\gamma}{\bar{w}} \right\rfloor = \left\lfloor \frac{c^4 \frac{\gamma B}{A}}{A \frac{\gamma^2}{B^2} A \gamma^2} \right\rfloor \geq \left\lfloor \frac{\|\hat{\mathbf{R}}_{t+1}^0 - \bar{\mathcal{R}}(\mathbf{H}(T_{t+1}))\|^2}{\alpha \epsilon} \right\rfloor. \quad (85)$$

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

$$\left| u \left( \hat{\mathcal{R}}(\mathbf{H}(n)) \right) - u \left( \bar{\mathcal{R}}(\mathbf{H}(T_{t+1})) \right) \right| \leq \alpha B^2. \quad (86)$$

By exploiting Assumption 4, we have

$$\|\widehat{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(T_{t+1}))\| \leq \left(\frac{\alpha B^2}{A}\right)^{\frac{1}{2}} = \gamma. \quad (87)$$

Therefore, the proof of (79) is complete by induction. Similarly to (82), we have

$$\|\bar{\mathcal{R}}(\mathbf{H}(n)) - \bar{\mathcal{R}}(\mathbf{H}(T_t))\| \leq \gamma^{\frac{1}{2}} \left(\gamma^{\frac{1}{2}} + \left(\frac{B}{A}\right)^{\frac{1}{2}}\right) \quad (88)$$

and (78) follows immediately from (79) and (88) by invoking triangle inequality.  $\square$

Theorems 8 and 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.

In the following, we generalize the results of Section IV to the case of joint rate and power allocation in fading channels.

## V. DYNAMIC RATE AND POWER ALLOCATION IN FADING CHANNEL

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  $\mathcal{R}^*(\mathbf{H}) \in C_g(\boldsymbol{\pi}^*(\mathbf{H}), \mathbf{H})$ , and  $\boldsymbol{\pi}^* \in \mathcal{G}$ . Moreover

$$\mathbb{E}_{\mathbf{H}}[\mathcal{R}^*(\mathbf{H})] = \mathbf{R}^* \in \operatorname{argmax}_{\mathbf{R}} u(\mathbf{R}) \quad \text{subject to } \mathbf{R} \in C(\bar{\mathbf{P}}) \quad (89)$$

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

$$(\mathcal{R}^*(\mathbf{h}), \boldsymbol{\pi}^*(\mathbf{h})) = \operatorname{argmax}_{\mathbf{r}, \mathbf{p}} (\boldsymbol{\mu}'\mathbf{r} - \boldsymbol{\lambda}'\mathbf{p}) \quad \text{subject to } \mathbf{r} \in C_g(\mathbf{p}, \mathbf{h}) \quad (90)$$

where  $\mathbf{h}$  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 (91) shown at the bottom of the

page, 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 (90) can be solved by a simple greedy algorithm (see [5, Lemma 3.2]). It is also shown in [5] that, for positive  $\boldsymbol{\mu}$ , the optimal solution  $\mathbf{R}^*$  to the problem in (89) is *uniquely* obtained. Given the distribution of channel-state process, denoted by  $F_k$  and  $f_k$ , we have (92) shown at the bottom of the page.

The uniqueness of  $\mathbf{R}^*$  follows from the fact that the stationary distribution of the channel-state process has a continuous density [5]. It is worth mentioning that (92) 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  $C(\bar{\mathbf{P}})$  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  $\mathbf{R}^*$ , the optimal solution to (89), is unique. Moreover, by Assumption 1(b), it lies on the boundary of the throughput region. Now suppose that  $\mathbf{R}^*$  is given by some genie. We can choose  $\boldsymbol{\mu}^* = \nabla u(\mathbf{R}^*)$  and  $\tilde{u}(\mathbf{R}) = (\boldsymbol{\mu}^*)'\mathbf{R}$ , as a replacement for the nonlinear utility. By checking the optimality conditions, it can be seen that  $\mathbf{R}^*$  is also the optimal solution of the problem in (89), i.e.,

$$\mathbf{R}^* = \operatorname{argmax}_{\mathbf{R}} (\boldsymbol{\mu}^*)'\mathbf{R} \quad \text{subject to } \mathbf{R} \in C(\bar{\mathbf{P}}). \quad (93)$$

Thus, we can employ the rate and power allocation policies in (90) 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  $\mathbf{R}^*$ . 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 (92). This naturally leads us to the conditional gradient method [28] to compute  $\mathbf{R}^*$ . The  $k$ th iteration of the method is given by

$$\mathbf{R}^{k+1} = \mathbf{R}^k + \alpha^k (\bar{\mathbf{R}}^k - \mathbf{R}^k) \quad (94)$$

where  $\alpha^k$  is the stepsize and  $\bar{\mathbf{R}}^k$  is obtained as

$$\bar{\mathbf{R}}^k \in \operatorname{argmax}_{\mathbf{R} \in C(\bar{\mathbf{P}})} \left( \nabla u(\mathbf{R}^k)'(\mathbf{R} - \mathbf{R}^k) \right) \quad (95)$$

---


$$\int_0^\infty \frac{1}{h} \int_{\frac{2\lambda_i(N_0+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k \left( \frac{2\lambda_k h(N_0+z)}{2\lambda_i(N_0+z) + (\mu_k - \mu_i)h} \right) f_i(h) dh dz = \bar{P}_i \quad (91)$$


---

$$\mathbf{R}_i^*(\boldsymbol{\mu}) = \int_0^\infty \frac{1}{2(N_0+z)} \int_{\frac{2\lambda_i(N_0+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k \left( \frac{2\lambda_k h(N_0+z)}{2\lambda_i(N_0+z) + (\mu_k - \mu_i)h} \right) f_i(h) dh dz. \quad (92)$$

where  $\nabla u(\mathbf{R}^k)$  denotes the gradient vector of  $u(\cdot)$  at  $\mathbf{R}^k$ . 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 subproblem is obtained by (92), in which  $\boldsymbol{\mu}$  is replaced by  $\nabla u(\mathbf{R}^k)$ . By concavity of the utility function and convexity of the capacity region, the iteration (94) will converge to the optimal solution of (89) for appropriate stepsize selection rules such as the Armijo rule or limited maximization rule (cf., [28, pp. 220–222]).

Note that our goal is to determine rate and power allocation policies. Finding  $\mathbf{R}^*$  allows us to determine such policies by the greedy policy in (90) for  $\boldsymbol{\mu}^* = \nabla u(\mathbf{R}^*)$ . It is worth mentioning that all the computations for obtaining  $\mathbf{R}^*$  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.

## 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 approaches. We focus on the case with no power control or knowledge of channel statistics. We also assume that the channel-state processes are generated by independent identical finite-state Markov chains. We consider a weighted  $\alpha$ -fair function as the utility function, i.e.,

$$u(\mathbf{R}) = \sum_i w_i f_\alpha(R_i) \quad (96)$$

where  $f_\alpha(\cdot)$  is given by (35).

We study two different communication scenarios to compare the performance of the greedy policy with the queue-based rate allocation policy by Eryilmaz and Srikant [8]. This policy, parameterized by some parameter  $K$ , uses queue length information to allocate the rates arbitrarily close to the long-term optimal policy by choosing  $K$  large enough. The parameter  $K$  is used to achieve a tradeoff between rate of convergence and suboptimality of the achieved rates. Figs. 5 and 6 illustrate the structure of the transmitters for queue-length-based policy and greedy policy, respectively. As shown in Fig. 5,  $x_i(n)$  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.,

$$\begin{aligned} \boldsymbol{\mu}(n) = \operatorname{argmax} \quad & \sum_{i=1}^M x_i(n) R_i \\ \text{subject to} \quad & \mathbf{R} \in C_g(\mathbf{P}, \mathbf{H}(n)). \end{aligned} \quad (97)$$

The congestion controller proposed in [8] leads to a fair allocation of the rates for a given  $\alpha$ -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:

$$\mathbb{E}[a_i(n) | x_i(n)] = \min \left\{ K \left( \frac{w_i}{x_i(n)} \right)^{\frac{1}{\alpha}}, D \right\} \quad (98)$$

$$\mathbb{E}[a_i^2(n) | x_i(n)] \leq U < \infty, \quad \text{for all } x_i(n) \quad (99)$$

where  $\alpha$ ,  $D$ , and  $U$  are positive constants.

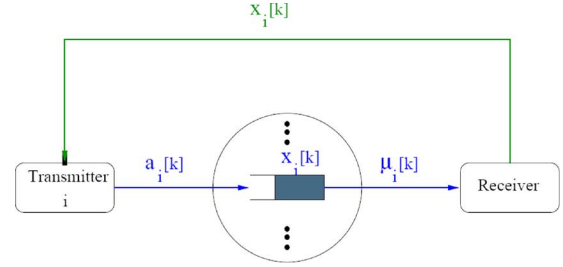


Fig. 5. Structure of the  $i$ th transmitter and the receiver for the queue-length-based policy [8].

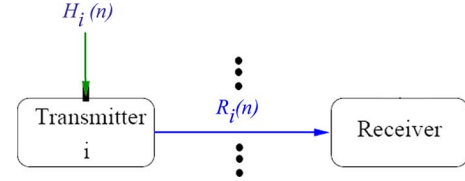


Fig. 6. Structure of the  $i$ th transmitter and the receiver for the presented policies.

In the first scenario, we compare the average achieved rate by the two policies for a communication session with limited duration. In this case, the utility function is given by (96) with  $\alpha = 2$  and  $w_1 = 1.5w_2 = 1.5$ , and the corresponding optimal solution is  $\mathbf{R}^* = (0.60, 0.49)$ . Fig. 7(a) depicts the distance between empirical average rate achieved by the greedy or the queue-length-based policy, and  $\mathbf{R}^*$ , the maximizer of the utility function over the throughput region. Fig. 7(b) demonstrates the performance difference in terms of the value of the utility of average allocated rates. As shown in Fig. 7, the greedy policy outperforms the queue-length-based policy for a communication session with limited duration. The average rate tuples allocated by the greedy and queue-length-based policies are illustrated over the throughput region in Fig. 8. We see that the points allocated by the queue-length-based policy approach the optimal solution  $\mathbf{R}^*$  from the interior of the throughput region, while the greedy policy always allocates rate tuples in a vicinity of the optimal solution. Hence, it achieves better performance within limited number of time slots.

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  $\mathbf{R}^*$ . We have tuned the parameter  $K$  so that the best performance of the queue-length-based policy is achieved within the time frame of the communication session.

Second, we consider a file upload scenario where each user is transmitting a file with fixed finite size to the base station. It is assumed that a file of size  $f_i$  is already stored at transmitter  $i$  at time 0. 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. Fig. 9 demonstrates the ratio of the utilities



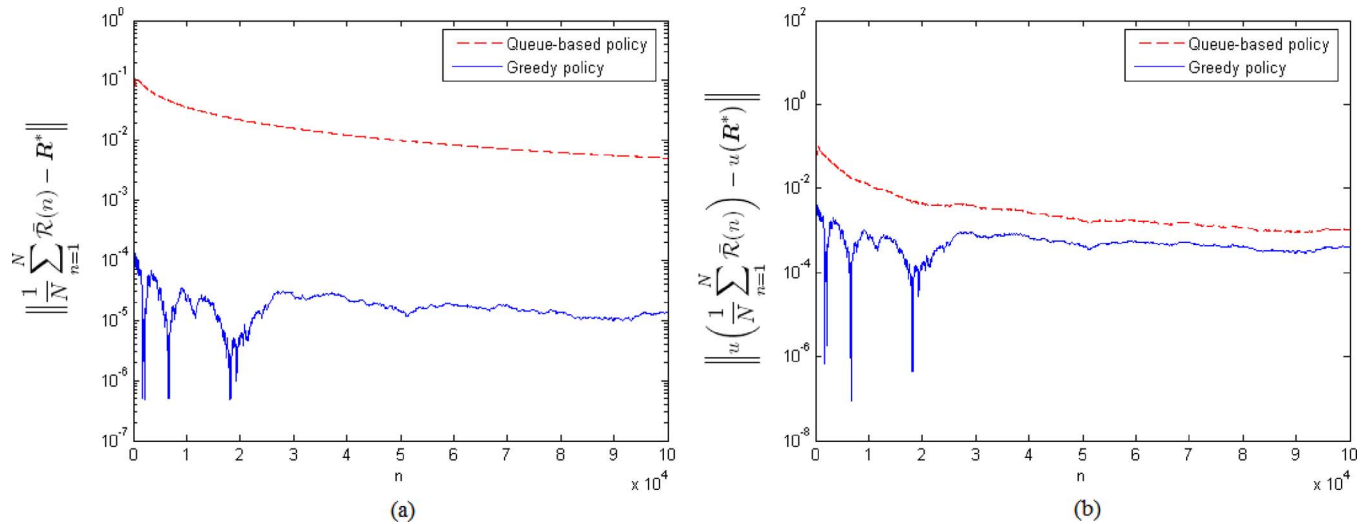


Fig. 7. Performance comparison of greedy and queue-based policies for a communication session with limited duration, for  $\sigma_H = 0.13$ .

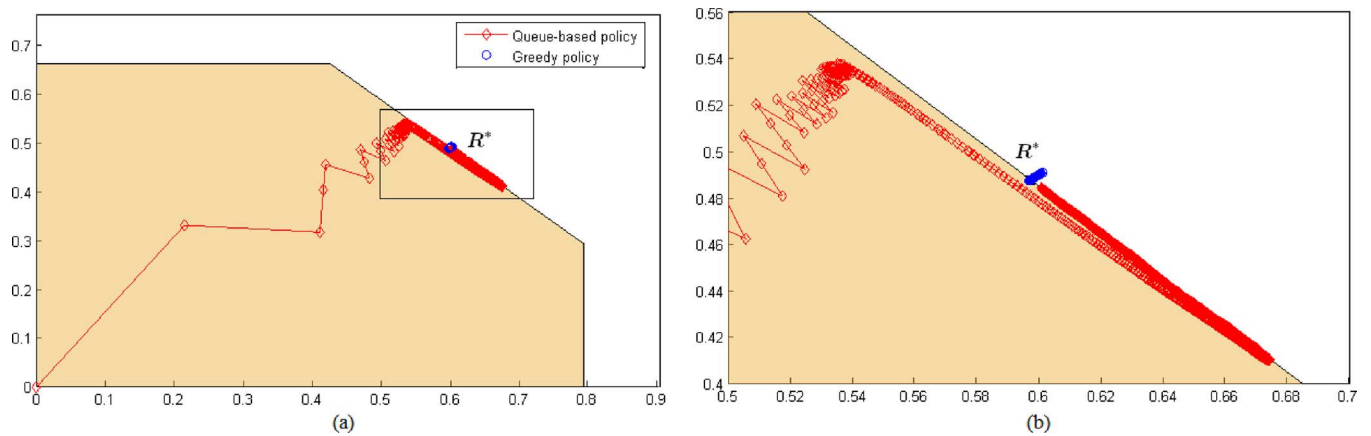


Fig. 8. Average allocated rates by greedy and queue-length-based policies approaching the optimal solution  $\mathbf{R}^*$  of maximizing the utility function over the average capacity region. (b) is a magnification of the box in (a) to demonstrate behavior of the policies in the neighborhood of the optimal solution. The queue-based policy approaches  $\mathbf{R}^*$  from the interior of the region, while the greedy policy approaches  $\mathbf{R}^*$  from the exterior.

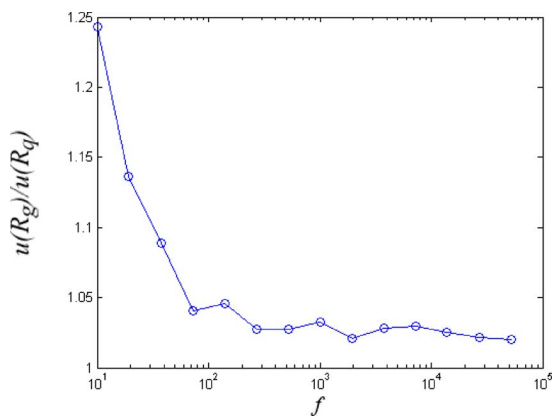


Fig. 9. Performance comparison of greedy and queue-based policies for file upload scenario with respect to file size  $f = f_1 = f_2$ .  $R_g$  and  $R_q$  are expected upload rate of the greedy and the queue-length-based policy, respectively.

of the average upload rates for the greedy and the queue-based policy plotted for different file sizes. We observe that for small

file sizes, the greedy policy achieves a higher utility value compared to the queue-based policy, and this difference decreases by increasing the file size. We can interpret this behavior as follows. For the queue-length-based policy, the transmission queue is initially empty, and almost all of each file is first buffered into the queues with equal rate  $D$  [see (98)]. Then, each queue is emptied by a max-weight scheduler according to (97). Once the files are all buffered in the queues, the queues are emptied with the same rate which is not fair because it does not give any priority to the users based on their utility. In other word, the parameter  $\alpha$  that is supposed to capture different fairness notions does not play any role in this mechanism. For larger file size, the duration for which the entire file is emptied into the queue is negligible compared to the total transmission time. Hence, there is enough time for queues to build up so that the rest of the files are buffered into the queues based on parameters of the utility functions. As a consequence, a higher utility for the average upload rate is achieved. In contrast, the greedy policy always selects the transmission rates by maximizing the utility function instantaneously, which results in close-to-optimal average achieved rates even for small file sizes (see Fig. 8).

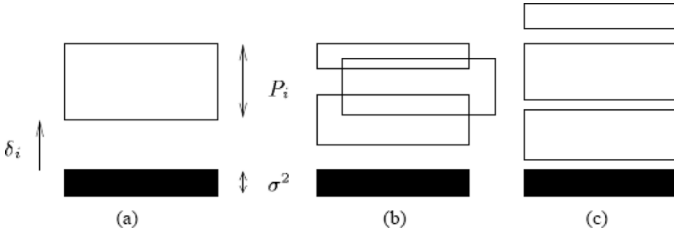


Fig. 10. Graphical representation of messages over MAC [12].

## VII. CONCLUSION

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

We considered several different scenarios for a MAC. 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.

For the case of a fading channel where power control and channel statistics are not available, we propose a greedy rate allocation policy that is short-term optimal but not long-term optimal for nonlinear utility functions. Nevertheless, we showed that its long-term performance in terms of the utility is not arbitrarily worse compared to the long-term 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.

We also 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.

Finally, using computer simulations, we compared the performance of the greedy policy and a queue-length-based policy [8] for a limited period of time. While not relying on any queue-length information, the greedy policy outperformed the queue-length-based policy during the communication session. This suggests that channel-state-based approaches can be more efficient while causing less overhead.

### APPENDIX I

#### ALGORITHM FOR FINDING A VIOLATED CONSTRAINT

In this section, we present an alternative algorithm based on the 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  $(M, \mathbf{P}, \mathbf{R}, N_0)$  is called a *configuration* for an  $M$ -user MAC, where  $\mathbf{R} = (R_1, \dots, R_M)$  is the rate tuple,  $\mathbf{P} = (P_1, \dots, P_M)$  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

$$R_i = C(P_i, N_0 + \delta_i), \quad i = 1, \dots, M. \quad (100)$$

Intuitively, we can think of message  $i$  as rectangles of height  $P_i$ , raised above the noise level by  $\delta_i$ . In fact,  $\delta_i$  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 nonnegative. However, the contrary is not true in general.

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

$$\delta_{i+1} \geq \delta_i + P_i, \quad i = 0, 1, \dots, M-1 \quad (101)$$

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

By the graphical representation described earlier, a configuration is single-user codable if none of the messages are overlapping. Fig. 10(a) gives an example of graphical representing for a message with power  $P_i$  and elevation  $\delta_i$ . Fig. 10(b) and (c) illustrates overlapping and nonoverlapping configurations, respectively.

*Definition 13:* The quadruple  $(m, \mathbf{p}, \mathbf{r}, N_0)$  is a *spinoff* of  $(M, \mathbf{P}, \mathbf{R}, N_0)$  if there exists a surjective mapping  $\phi : \{1, \dots, m\} \rightarrow \{1, \dots, M\}$  such that for all  $i \in \{1, \dots, M\}$ , we have

$$\begin{aligned} P_i &\geq \sum_{j \in \phi^{-1}(i)} p_j \\ R_i &\leq \sum_{j \in \phi^{-1}(i)} r_j \end{aligned}$$

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

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

$$\bar{P} = \sum_{k=1}^d P_{i_k}, \quad \bar{R} = \sum_{k=1}^d R_{i_k}. \quad (102)$$

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

*Proof:* See [12, Th. 1].  $\square$

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 spinoff is trivially equal to the configuration. Otherwise, merge two of the overlapping users into a *hyperuser* 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 spinoff for a given configuration.

It follows directly from the proof of Theorem 10 that this recursive algorithm gives a single-user codable spinoff for an achievable configuration. If the configuration is not achievable, then the algorithm encounters a hyperuser with negative elevation. At this point the algorithm terminates. Suppose that this hyperuser 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

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

where  $S = \{i_1, \dots, i_d\} \subseteq \mathcal{M}$ . Therefore, a hyperuser 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  $M$  hyperusers 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 hyperuser with negative elevation could be found in  $O(1)$  time, or if such a hyperuser does not exist, it takes  $O(M)$  time to find two overlapping hyperusers. In the case that there are no overlapping users and all the elevations are nonnegative, 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.  $\square$

## 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  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  be concave and twice differentiable. Let  $X$  be a random variable with variance  $\sigma_X^2$ . Then

$$f(\mathbb{E}[X]) - \mathbb{E}[f(X)] \leq \sqrt{2M\sigma_X^2} f(\mathbb{E}[X]) - \frac{\sigma_X^2 M}{2} \quad (103)$$

where  $M$  be an upperbound on  $|f''(x)|$ .

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

$$\Pr(|X - \mathbb{E}(X)| > c) \leq \epsilon \quad (104)$$

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

$$\begin{aligned} \mathbb{E}[f(X)] &= \mathbb{E}[f(X) | |X - \mathbb{E}(X)| \leq c] \Pr(|X - \mathbb{E}(X)| \leq c) \\ &\quad + \mathbb{E}[f(X) | |X - \mathbb{E}(X)| > c] \Pr(|X - \mathbb{E}(X)| > c) \\ &\geq (1 - \epsilon) \mathbb{E}[f(X) | |X - \mathbb{E}(X)| \leq c] \\ &\geq \frac{1 - \epsilon}{2} (f(\mathbb{E}[X] + c) + f(\mathbb{E}[X] - c)) \\ &= (1 - \epsilon) f(\mathbb{E}[X]) + \frac{1 - \epsilon}{4} c^2 (f''(\xi_1) + f''(\xi_2)) \quad (105) \end{aligned}$$

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

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

$$f(\mathbb{E}[X]) - \mathbb{E}[f(X)] \leq \frac{1 - \epsilon}{2\epsilon} \sigma_X^2 M + \epsilon f(\mathbb{E}[X]). \quad (106)$$

The right-hand side is minimized for

$$\epsilon^* = \min \left\{ \left( \frac{\sigma_X^2 M}{2f(\mathbb{E}[X])} \right)^{\frac{1}{2}}, 1 \right\}. \quad (107)$$

By substituting  $\epsilon^*$  in (106), the desired result follows immediately.  $\square$

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  $\sigma_X^2$ . Also, let  $Y = \log(1 + X)$ . Then, variance of  $Y$  is upperbounded as

$$\sigma_Y^2 \leq \sigma_X^2 \left( 1 + \left[ (1 + \bar{X}) \left( \sqrt{2 \log(1 + \bar{X})} - \frac{\sigma_X}{2} \right) \right]^2 \right). \quad (108)$$

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

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

where  $\tilde{X}$  is a nonnegative random variable.

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

$$\mathbb{E}[\log(1 + X)] \geq \log(1 + \bar{X}) - \sqrt{2\sigma_X^2 \log(1 + \bar{X})} + \frac{\sigma_X^2}{2}. \quad (110)$$

Hence

$$\begin{aligned} \bar{X} &\geq \hat{X} = \exp \{ \mathbb{E}[\log(1 + X)] \} - 1 \\ &\geq \exp \left\{ \log(1 + \bar{X}) - \sqrt{2\sigma_X^2 \log(1 + \bar{X})} + \frac{\sigma_X^2}{2} \right\} - 1 \\ &\geq \bar{X} - \sigma_X (1 + \bar{X}) \left( \sqrt{2 \log(1 + \bar{X})} - \frac{\sigma_X}{2} \right) \quad (111) \end{aligned}$$

where the first inequality is by (110), and the second relation can be verified after some straightforward manipulation. By combining (109) and (111), the variance of  $Y$  can be bounded as follows:

$$\begin{aligned} \sigma_Y^2 &\leq \mathbb{E}[(X - \hat{X})^2] \\ &\leq \mathbb{E} \left[ \left( X - \bar{X} + \sigma_X(1 + \bar{X}) \left( \sqrt{2 \log(1 + \bar{X})} - \frac{\sigma_X}{2} \right) \right)^2 \right] \\ &= \sigma_X^2 \left( 1 + \left[ (1 + \bar{X}) \left( \sqrt{2 \log(1 + \bar{X})} - \frac{\sigma_X}{2} \right) \right]^2 \right). \end{aligned} \quad (112)$$

□

Now we provide the proof for Lemma 5. Let the random variable  $Y_S$  be defined as follows:

$$Y_S = \frac{1}{2} \log \left( 1 + \sum_{i \in S} \frac{H_i P_i}{N_0} \right), \quad \text{for all } S \subseteq \mathcal{M} = \{1, \dots, M\}. \quad (113)$$

The facet defining constraints of  $C_g(\mathbf{P}, \mathbf{H})$  and  $C_a(\mathbf{P})$  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  $d_H(C_g(\mathbf{P}, \mathbf{H}), C_a(\mathbf{P})) \leq \delta$  if and only if  $|Y_S - \mathbb{E}[Y_S]| \leq \delta$ , for all  $S \subseteq \mathcal{M}$ . Thus, we can write

$$\begin{aligned} &\Pr \left\{ d_H(C_g(\mathbf{P}, \mathbf{H}), C_a(\mathbf{P})) > \delta \right\} \\ &= \Pr \left\{ \max_S |Y_S - \mathbb{E}[Y_S]| > \delta \right\} \\ &\leq \sum_{S \subseteq \mathcal{M}} \Pr \left\{ |Y_S - \mathbb{E}[Y_S]| > \delta \right\} \\ &\leq \frac{1}{\delta^2} \sum_{S \subseteq \mathcal{M}} \sigma_{Y_S}^2 \end{aligned} \quad (114)$$

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.,

$$\sigma_{Y_S}^2 \leq \frac{\sigma_{Z_S}^2}{4} \left( 1 + \left[ (1 + \bar{Z}_S) \left( \sqrt{2 \log(1 + \bar{Z}_S)} - \frac{\sigma_{Z_S}}{2} \right) \right]^2 \right) \quad (115)$$

where

$$\begin{aligned} \bar{Z}_S &= \mathbb{E} \left[ \sum_{i \in S} \frac{H_i P_i}{N_0} \right] = \sum_{i \in S} \Gamma_i \bar{H}_i = \mathbf{\Gamma}'_S \bar{\mathbf{H}}, \\ \sigma_{Z_S}^2 &= \text{var} \left( \sum_{i \in S} \frac{H_i P_i}{N_0} \right) \\ &= \sum_{(i,j) \in S^2} \Gamma_i \Gamma_j \text{cov}(H_i, H_j) = \mathbf{\Gamma}'_S \mathbf{K} \mathbf{\Gamma}_S. \end{aligned}$$

The desired result is concluded by substituting  $\bar{Z}_S$  and  $\sigma_{Z_S}^2$  in (115) and combing the result with (114). □

### 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  $\delta$  contains its expansion by relaxing its constraints by  $\delta$ .

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

$$C_1 = \left\{ \mathbf{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq f(S), \text{ for all } S \subseteq \mathcal{M} \right\} \quad (116)$$

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

*Proof:* By Definition 4, it is straightforward to show that  $C_2$  is also a polymatroid, i.e.,

$$C_2 = \left\{ \mathbf{R} \in \mathbb{R}_+^M : \sum_{i \in S} R_i \leq g(S) = f(S) + \delta, \text{ for all } S \subseteq \mathcal{M} \right\} \quad (117)$$

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  $\mathbf{R} \in \mathbb{R}^M$  be a vertex of  $C_2$ . The polymatroid structure of  $C_2$  implies that  $\mathbf{R}$  is generated by an ordered subset of  $\mathcal{M}$  (see [35, Th. 2.1]). Hence, there is some  $k \in \mathcal{M}$  such that  $R_k = f(\{k\}) + \delta$ . Consider the following construction for  $\mathbf{R}'$ :

$$R'_i = \begin{cases} R_i - \delta, & i = k \\ R_i, & \text{otherwise.} \end{cases} \quad (118)$$

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

$$\sum_{i \in S} R'_i = \sum_{i \in S} R_i - \delta \leq f(S). \quad (119)$$

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 (119), 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  $u(\mathbf{R}_2^*) \geq u(\mathbf{R}_1^*)$ . By Lemma 11, there exists some  $\mathbf{R} \in C_a(\mathbf{P})$  such that  $\|\mathbf{R}_2^* - \mathbf{R}\| \leq \delta$ . Moreover, we can always choose  $\mathbf{R}$  to be on the boundary so that  $\|\mathbf{R}\| \geq D_\delta$ , where  $D_\delta$  is defined in

(34). Therefore, by Assumption 3(a) and the fact that  $u(\mathbf{R}_2^*) \geq u(\mathbf{R}_1^*) \geq u(\mathbf{R})$ , we have

$$u(\mathbf{R}_2^*) - u(\mathbf{R}) = |u(\mathbf{R}_2^*) - u(\mathbf{R})| \leq B \|\mathbf{R}_2^* - \mathbf{R}\| \leq B\delta. \quad (120)$$

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

$$u(\mathbf{R}_1^*) - u(\mathbf{R}) = |u(\mathbf{R}_1^*) - u(\mathbf{R})| \geq A \|\mathbf{R}_1^* - \mathbf{R}\|^2 > B\delta. \quad (121)$$

By subtracting (120) from (121), we obtain  $u(\mathbf{R}_2^*) < u(\mathbf{R}_1^*)$  which is a contradiction. Therefore,  $\|\mathbf{R}_1^* - \mathbf{R}\| \leq (\frac{B}{A}\delta)^{\frac{1}{2}}$ , and the desired result follows immediately by invoking the triangle inequality.  $\square$

## REFERENCES

- [1] X. Wang and G. B. Giannakis, "Energy-efficient resource allocation in time division multiple-access over fading channels," 2005, preprint.
- [2] S. J. Oh, Z. Danlu, and K. M. Wasserman, "Optimal resource allocation in multiservice CDMA networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 811–821, Jul. 2003.
- [3] J. B. Kim and M. L. Honig, "Resource allocation for multiple classes of DS-CDMA traffic," *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 506–519, Mar. 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 *Proc. 41st Annu. Asilomar Conf. Signals Syst. Comput.*, 2007, pp. 265–269.
- [5] D. Tse and S. Hanly, "Multiaccess fading channels part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.
- [6] S. Shenker, "Fundamental design issues for the future internet," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1176–118, Sep. 1995.
- [7] R. Srikant, *Mathematics of Internet Congestion Control*. Cambridge, MA: Birkhauser, 2004.
- [8] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length based scheduling and congestion control," in *Proc. IEEE Infocom*, Miami, FL, Mar. 2005, vol. 3, pp. 1794–1803.
- [9] M. J. Neely, E. Modiano, and C. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *Proc. IEEE Infocom*, Miami, FL, Mar. 2005, pp. 1723–1734.
- [10] V. Subramanian and R. Agrawal, "Optimality of certain channel aware scheduling policies," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Oct. 2002, pp. 1532–1541.
- [11] A. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Oper. Res.*, vol. 53, no. 1, pp. 12–25, 2005.
- [12] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.
- [13] X. Lin and N. Shroff, "The impact of imperfect scheduling on cross-layer rate control in multihop wireless networks," in *Proc. IEEE Infocom*, Miami, FL, Mar. 2005, vol. 3, pp. 1804–1814.
- [14] A. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Syst.*, vol. 50, no. 4, pp. 401–457, 2005.
- [15] A. Eryilmaz, A. Ozdaglar, and E. Modiano, "Polynomial complexity algorithms for full utilization of multi-hop wireless networks," in *Proc. IEEE Infocom*, Anchorage, AL, May 2007, pp. 499–507.
- [16] L. Tassiulas, "Linear complexity algorithms for maximum throughput in radio networks and input queued switches," in *Proc. IEEE Infocom*, 1998, pp. 533–539.
- [17] E. Modiano, D. Shah, and G. Zussman, "Maximizing throughput in wireless networks via gossiping," *ACM SIGMETRICS/IFIP Performance Eval. Rev.*, vol. 34, no. 1, pp. 27–38, 2006.
- [18] L. X. Bui, S. Sanghavi, and R. Srikant, "Distributed link scheduling with constant overhead," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1467–1480, 2009.
- [19] C. Joo, X. Lin, and N. Shroff, "Performance limits of greedy maximal matching in multi-hop wireless networks," in *Proc. IEEE Conf. Decision Control*, New Orleans, LA, Dec. 2007, pp. 1128–1133.
- [20] A. Eryilmaz, A. Ozdaglar, D. Shah, and E. Modiano, "Imperfect randomized algorithms for optimal control of wireless networks," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, Mar. 2008, pp. 922–937.
- [21] S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Optimum power and rate allocation strategies for multiple access fading channels," in *Proc. IEEE Veh. Technol. Conf.*, 2001, vol. 4, pp. 2888–2892.
- [22] E. Yeh and A. Cohen, "Delay optimal rate allocation in multiaccess fading communications," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Oct. 2004, pp. 140–149.
- [23] D. Yu and J. M. Cioffi, "Iterative water-filling for optimal resource allocation in OFDM multiple-access and broadcast channels," in *Proc. IEEE GLOBECOM*, 2006, DOI: 10.1109/GLOCOM.2006.587.
- [24] K. Seong, R. Narasimhan, and J. Cioffi, "Scheduling for fading multiple access channels with heterogeneous QoS constraints," in *Proc. Int. Symp. Inf. Theory*, 2007, pp. 2966–2970.
- [25] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Electr. Eng., Univ. Hawaii, Honolulu, HI, 1972.
- [26] S. Shamai and A. D. Wyner, "Information theoretic considerations for symmetric, cellular, multiple-access fading channels—Part I," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1877–1894, Nov. 1997.
- [27] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Cambridge, MA: Athena Scientific, 2003.
- [28] D. P. Bertsekas, *Nonlinear Programming*. Cambridge, MA: Athena Scientific, 1999.
- [29] U. Brännlund, "On relaxation methods for nonsmooth convex optimization," Ph.D. dissertation, Dept. Math., Royal Inst. Technol., Stockholm, Sweden, 1993.
- [30] M. Grötschel, L. Lovász, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *Combinatorica*, vol. 1, no. 2, pp. 169–197, 1981.
- [31] J. Orlin, "A faster strongly polynomial time algorithm for submodular function minimization," in *Proc. 12th Conf. Integer Programm. Combin. Optim.*, vol. 2007, pp. 240–251.
- [32] A. ParandehGheibi, "Fair resource allocation in multiple access channels," M.Sc. thesis, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, 2008.
- [33] A. Nedić and D. P. Bertsekas, "Stochastic optimization: Algorithms and applications," in *Convergence Rate of Incremental Subgradient Algorithms*, S. Uryasev and P. M. Pardalos, Eds. Norwell, MA: Kluwer, 2000.
- [34] R. Gallager, *Discrete Stochastic Processes*. London, U.K.: Kluwer, 1996.
- [35] R. E. Bixby, W. H. Cunningham, and D. M. Topkis, "The partial order of a polymatroid extreme point," *Math. Oper. Res.*, vol. 10, no. 3, pp. 367–378, 1985.

**Ali ParandehGheibi** (S'08) received the B.Sc. degree (first-class honors) in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2006 and the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 2008, where he is currently working towards the Ph.D. degree in electrical engineering and computer science.

His research interests include network optimization and control, network coding, video streaming, peer-to-peer networks, distributed optimization, gossip algorithms, and social networks.

**Atilla Eryilmaz** (S'00–M'06) received the B.S. degree in electrical and electronics engineering from Bogaziçi University, Istanbul, Turkey, in 1999 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2001 and 2005, respectively.

Between 2005 and 2007, he worked as a Postdoctoral Associate at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge. Currently, he is an Assistant Professor of Electrical and Computer Engineering at The Ohio State University, Columbus. His research interests include communication networks, optimal control of stochastic networks, optimization theory, distributed algorithms, stochastic processes, and network coding.

Dr. Eryilmaz received the 2010 National Science Foundation (NSF) CAREER award.

**Asu Ozdaglar** (S'96–M'02) received the B.S. degree in electrical engineering from the Middle East Technical University, Ankara, Turkey, in 1996 and the S.M. and the Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1998 and 2003, respectively.

Since 2003, she has been a member of the faculty of the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, where she is currently the Class of 1943 Associate Professor. She is also a member of the Laboratory for Information and Decision Systems and the Operations Research Center. Her research interests include optimization theory, with emphasis on nonlinear programming and convex analysis, game theory, with applications in communication, social, and economic networks, and distributed optimization and control. She is the coauthor of the book entitled *Convex Analysis and Optimization* (Cambridge, MA: Athena Scientific, 2003).

Prof. Ozdaglar is the recipient of a Microsoft fellowship, the MIT Graduate Student Council Teaching award, the National Science Foundation (NSF) Career award, and the 2008 Donald P. Eckman award of the American Automatic Control Council. She currently serves as an Associate Editor for the area Optimization Theory, Algorithms and Applications for *The Asia-Pacific Journal of Operational Research* and is a member of the Board of Governors of the Control System Society.

**Muriel Médard** (S'91–M'95–SM'01–F'08) received the B.S. degrees in electrical engineering and computer science and in mathematics, the B.S. degree in humanities, the M.S. degree in electrical engineering, and the Sc.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1989, 1990, 1991, and 1995, respectively.

She is a Professor at the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge. She was previously an Assistant Professor at the Electrical and Computer Engineering Department and a member of the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana. From 1995 to 1998, she was a Staff Member at MIT Lincoln Laboratory in the Optical Communications and the Advanced Networking Groups. Her research interests are in the areas of network coding and reliable communications, particularly for optical and wireless networks.

Prof. Médard serves as an Associate Editor for the Optical Communications and Networking Series of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, as an Associate Editor in Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY, and as a Guest Editor for the joint special issue of the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE/ACM TRANSACTIONS ON NETWORKING on Networking and Information Theory. She has served as a Guest Editor for the IEEE JOURNAL OF LIGHTWAVE TECHNOLOGY and as an Associate Editor for the *OSA Journal of Optical Networking*. She was awarded the IEEE Leon K. Kirchmayer Prize Paper Award 2002 for her paper, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel" (IEEE TRANSACTIONS ON INFORMATION THEORY, vol. 46, no. 3, pp. 935–946, May 2000). She was coawarded the Best Paper Award for G. Weichenberg, V. Chan, M. Médard, "Reliable architectures for networks under stress," (Fourth International Workshop on the Design of Reliable Communication Networks, October 2003, Banff, AB, Canada). She received a National Science Foundation (NSF) Career Award in 2001 and was a cowinner of the 2004 Harold E. Edgerton Faculty Achievement Award, established in 1982 to honor junior faculty members "for distinction in research, teaching and service to the MIT community." She was named a 2007 Gilbreth Lecturer by the National Academy of Engineering.