Multiaccess Fading Channels–Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities

David N. C. Tse, *Member, IEEE*, and Stephen V. Hanly, *Member, IEEE*

***Abstract*—In multiaccess wireless systems, dynamic allocation of resources such as transmit power, bandwidths, and rates is an important means to deal with the time-varying nature of the environment. In this two-part paper, we consider the problem of optimal resource allocation from an information-theoretic point of view. We focus on the multiaccess fading channel with Gaussian noise, and define two notions of capacity depending on whether the traffic is delay-sensitive or not. In part I, we characterize the *throughput capacity region* which contains the long-term achievable rates through the time-varying channel. We show that each point on the boundary of the region can be achieved by successive decoding. Moreover, the optimal rate and power allocations in each fading state can be explicitly obtained in a greedy manner. The solution can be viewed as the generalization of the water-filling construction for single-user channels to multiaccess channels with arbitrary number of users, and exploits the underlying *polymatroid* structure of the capacity region. In part II, we characterize a *delay-limited* capacity region and obtain analogous results.**

***Index Terms*—Fading channels, multiaccess, multiuser water filling, power control, successive cancellation.**

I. Introduction

The mobile wireless environment provides several unique challenges to reliable communication not found in wired networks. One of the most important of these is the time-varying nature of the channel. Due to effects such as multipath fading, shadowing, and path losses, the strength of the channel can fluctuate in the order of tens of decibels. A general strategy to combat these detrimental effects is through the dynamic allocation of resources based on the states of the channels of the users. Such resources may include transmitter power, allocated bandwidth, and bit rates. For example, in the IS-95 CDMA (code-division multiple access) standard, the transmitter powers of the mobiles are controlled such that the received powers at the base station are the same for all mobiles. Thus a user has to be dynamically allocated more power when its reception at the base station is weak. This is to combat the *so-called* near–far problem. Another example is the dynamic channel allocation strategy which aims to adaptively find the best frequencies to transmit at.

Most of the existing work on dynamic resource allocation has been done with respect to *specific* multiple-access schemes, such as CDMA, TDMA (time-division multiple access) and FDMA (frequency-division multiple access). In this paper, we address the problem at a more fundamental level: what are the information theoretically optimal resource allocation schemes and their achievable performance for multiple access? We focus on the single-cell uplink scenario where a set of mobiles communicate to the base station with a single receiver. Our answers are in terms of *capacity regions* of the multiaccess fading channel with Gaussian noise, when both the receiver and the transmitters can track the time-varying channel. To this end, we consider two notions of capacity for the fading channel.

The first is the classic notion of Shannon capacity directly applied to the fading channel. In this definition, the channel statistics are assumed to be fixed, and the codeword length can be chosen arbitrarily long to average over the fading of the channel. Thus to achieve these rates, users will experience delay which depends on how fast the channel varies. We call this the *throughput capacity* as it measures long-term rates, averaged over the fading process.

In contrast, we also define a notion of *delay-limited* capacity for fading channels: these are the rates achievable using codeword lengths which are *independent* of how fast the channel varies. The former notion of capacity is relevant for situations when the delay requirement of the users is much longer than the time scale of the channel fading; it is particularly appropriate for data applications in which delay is not an issue, although it can also be relevant for delay-sensitive traffic if the fading in the channel is sufficiently fast to give tolerable delays. On the other hand, delay-limited capacity is relevant when the delay requirement is shorter than the time scale of channel variations so that one cannot average over the fades and has to maintain the desired rate at all fading states.

We have obtained complete characterizations of these two capacity regions as well as the optimal resource-allocation schemes which attain the points on the boundary of these regions. We compute the boundaries of the capacity regions, and show that every point on the boundary is achievable by

Manuscript received December 19, 1996; revised May 3, 1998. This work was supported in part by the Air Force Office of Scientific Research under Grant F49620-96-1-0199. This work was also supported by the Australian Research Council large grant. Part of this work was performed when both authors were at the Mathematical Research Science Center, AT&T Bell Laboratories. The material in this paper was presented in part at the Information Theory Workshop, Haifa, Israel, June 1996.

D. N. C. Tse is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

S. V. Hanly is with the Department of of Electrical Engineering, University of Melbourne, Melbourne, Australia (e-mail: s.hanly@ee.mu. oz.au).

Publisher Item Identifier S 0018-9448(98)07359-3.

successive decoding, which means that a series of single-user decodings is sufficient to achieve capacity. More precisely, first one user is decoded, treating all other users as noise, then its decoded signal is subtracted from the sum signal, then the next user is decoded and subtracted, and so forth. Thus our solution characterizes the optimal multiple-access schemes, as well as the optimal power allocation. Given the state of the channels, the optimal power allocation can be computed very efficiently and explicitly using *greedy* algorithms.

The optimal power allocations we obtain are solutions to various optimization problems over the multiaccess Gaussian capacity region. Since the number of constraints defining the capacity region is exponential in the number of users, to obtain simple solutions we need to exploit the special *polymatroid* structure of the capacity region. Polymatroid structure has been used successfully in many resource-allocation problems to obtain greedy optimization algorithms (see, for example, [5].) In this paper, we will show that the multiaccess Gaussian capacity region in fact belongs to a special class of *generalized symmetric* polymatroids, and we derive new greedy solutions to various optimization problems for this class of polymatroids.

Goldsmith and Varaiya [8] addressed the problem of computing the throughput capacity of *single- user* fading channels when both the transmitter and the receiver can track the channel. The optimal power allocation is obtained via water-filling over the fading states. Knopp and Humblet [14] have solved the multiuser version of that problem for the special case of symmetric users with equal rate requirements. (A similar result was also presented later in [3].) Our results on computing the entire throughput capacity region of the multiaccess fading channel and the associated optimal power allocation can be viewed as the analog of the classic water-filling solution in the multiuser setting. In a related work, Cheng and Verdú [2] obtained an explicit characterization of the capacity region of the two-user time-invariant multiaccess Gaussian channel with intersymbol interference (ISI). We will see that this channel is essentially the "frequency dual" of the multiaccess flat-fading channel and our techniques for the latter can be readily applied and provide a general solution to the multiaccess ISI channel for an *arbitrary* number of users. Moreover, our results extend to the frequency-selective fading case in a straightforward manner.

The notion of delay-limited capacity was introduced in [12] which obtained results in the symmetric case. The delay-limited power-allocation schemes are similar in flavor to those considered in the CDMA power control literature (see, for example, [11] and [19]), where the goal is to maintain a desired signal-to-noise ratio (SNR) at *all* fading states. However, those works consider only decoding schemes where a user is decoded treating other users as interference, which is suboptimal from an information-theoretic point of view. Our optimal schemes shed some light on the possible improvement by using more complex decoding techniques.

Early work on power control in the Shannon-theoretic context [9], [10] established structural results about the multiuser Gaussian capacity region arising directly from its polymatroid structure. These results provided additional motivation for the present paper.

In Part I of this paper, we will characterize the throughput capacity region and the optimal resource-allocation schemes, while we will relegate the analysis of delay-limited capacities to Part II. Part I is organized as follows. In Section II we introduce the Gaussian, multiaccess, flat-fading model and present a coding theorem for the throughput capacity region when transmitters and receiver can track the channel. This theorem implies that the extra benefit gained from the transmitters tracking the channel is fully realized in the ability to allocate transmitter power based on the channel state. In Section III, we use Lagrangian techniques to show that the optimal power allocation can be obtained by solving a family of optimization problems over a set of parallel time-invariant multiaccess Gaussian channels, one for each fading state. Given the Lagrange multipliers ("power prices") for the average power constraints, the problem is that of finding the optimal "rate" and "power" allocations as a function of each fading state. Here, we exploit the polymatroid structure of the optimization problem to obtain an explicit solution via a greedy algorithm. In Section IV we provide a simple iterative algorithm to compute the power prices for given average power constraints. Together with the greedy power allocation, this yields an efficient algorithm for dynamic resource allocation; moreover, it lends itself naturally to an *adaptive* implementation when the fading statistics are not known. In Section V, we show how the usual economic interpretation of Lagrange multipliers has useful application in radio-resource allocation. In particular, we exploit the symmetry between rate and power to define a power minimization problem, dual to that of maximizing Shannon capacity. In Section VI, we will present greedy power allocation solutions when additional power constraints are imposed. These results exploit further properties of polymatroids. In Section VII, we extend our flat fading model to the case of frequency-selective fading.

Due to the length of the paper, we provide a self-contained summary of the main points of the solution at the end of the paper, in Section VIII.

A word about notation: in this paper we will use boldface letters to denote vector quantities.

II. The Multiaccess Fading Channel

A. Preliminaries

We focus on the uplink scenario where a set of M users communicate to a single receiver. Consider the discrete-time multiple-access Gaussian channel

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

where M is the number of users, $X_i(n)$ and $H_i(n)$ are the transmitted waveform and the fading process of the ith user, respectively, and $Z(n)$ is white Gaussian noise with variance σ^2. We assume that the fading processes for all users are jointly stationary and ergodic, and the stationary distribution has continuous density and is bounded. User i is also subject to an average transmitter power constraint of $\bar{P}_i$. Note that in this basic model, we consider fading effects which are frequency nonselective. Frequency-selective fading will be considered in Section VII.

Consider first the simple situation where the users' locations are fixed and the signal of user i is attenuated by a factor of h_i when received at the base station, i.e., $H_i(n) = h_i$ for all time n. The characterization of the capacity region of the multiaccess memoryless channel with probability transitions $p(y|x_1, \cdots, x_M)$ is well known (Ahlswede [1], Liao [13]); it is the set of all rate vectors $\boldsymbol{R}$ satisfying

$$\boldsymbol{R}(S) \le I[Y; (X_i)_{i\in S} | (X_i)_{i\notin S}] \qquad \forall S \subset \{1, \cdots, M\}$$

for some independent input distribution $p(x_1)p(x_2)\cdots p(x_M)$. (In this paper, for any vector $\boldsymbol{v}$ we use the notation $\boldsymbol{v}(S)$ to denote $\sum_{i\in S} \boldsymbol{v}(i)$.) Note that S is any subset of users in $\{1, 2, \cdots, M\}$. The right-hand side of each of the above inequalities is the mutual information between the output and the inputs of users in S, conditional on the inputs of users not in S. In the case of the Gaussian multiaccess channel, this capacity region reduces to

$$\mathcal{C}_g(\boldsymbol{h}, \boldsymbol{P}) = \left\{ \boldsymbol{R} : \boldsymbol{R}(S) \le \frac{1}{2}\log\left(1 + \frac{\sum_{i\in S} h_i P_i}{\sigma^2}\right) \text{ for every } S \subset \{1, \cdots, M\} \right\} \tag{2}$$

where $\boldsymbol{h} = (h_1, \cdots, h_M)$ and $\boldsymbol{P} = (P_1, \cdots, P_M)$ Note that this region is characterized by $2^M - 1$ constraints, each corresponding to a nonempty subset of users. The right-hand side of each constraint is the joint mutual information per unit time between the subset of the users and the receiver conditional on knowing the transmitted symbols of the other users, under (optimal) independent Gaussian distributed inputs. It can also be interpreted as the maximum sum rate achievable for the given subset of users, with the other users' messages already known at the receiver. Moreover, it is known that the capacity region has precisely $M!$ vertices in the positive quadrant, each achievable by a successive decoding using one of the $M!$ possible orderings.

We now turn to the case of interest where the channels are time-varying due to the motion of the users. When the receiver can perfectly track the channel but the transmitters have no such information, the codewords cannot be chosen as a function of the state of the channel but the decoding can make use of such information. For this scenario, the capacity region is known (Gallager [7], Shamai and Wyner [17]) and is given by

$$\left\{ (R_1, \cdots, R_M) : \boldsymbol{R}(S) \le \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{\sum_{i\in S} H_i P_i}{\sigma^2}\right)\right], \quad \forall S \subset \{1, \cdots, M\} \right\} \tag{3}$$

where $\boldsymbol{H} = (H_1, \cdots, H_M)$ is a random vector having the stationary distribution of the joint fading process. A rigorous proof of this result can be found in [17]. An intuitive understanding of this result can be obtained by viewing capacities in terms of time averages of mutual information (Gallager [7]), the rate of flow of which can be viewed as a random process depending on the fading levels of the users. Specifically, at time n, the rate of flow of joint mutual information between a subset S of users and the receiver, conditional on the other users' messages being known, can be thought of as

$$\frac{1}{2}\log\left(1 + \frac{\sum_{i\in S} H_i(n) P_i}{\sigma^2}\right).$$

(This assumes that the transmitted waveforms are independent Gaussian processes with power P.) Thus the amount of mutual information averaged over a time interval $[0, T]$ is

$$\frac{1}{T}\sum_{n=1}^{T} \frac{1}{2}\log\left(1 + \frac{\sum_{i\in S} H_i(n) P_i}{\sigma^2}\right).$$

As $T \to \infty$, this quantity converges to the right-hand side of the constraint in (3) corresponding to the subset S. This is because of the ergodicity and stationarity of the fading processes.

The multiaccess fading system above is reminiscent of a queuing system with time-varying service rates, corresponding to the instantaneous rates of flow of joint mutual information. In this interpretation, the capacity can be viewed as the throughput of such a queuing system, being the long-term maximum average arrival rates (of mutual information) sustainable by the system. Hence, we will also call this capacity the *throughput capacity* of a fading channel. We will use the terms capacity and throughput capacity interchangeably in this paper, using the latter when we want to emphasize the distinction from other notions of capacity that will be defined in Part II.

B. The Capacity Region Under Dynamic Resource Allocation

We shall now focus on the scenario of interest in this paper, where all the transmitters and the receiver know the current state of the channels of every user. Thus the codewords and the decoding scheme can both depend on the current state of the channels. In practice, this knowledge is obtained from the receiver measuring the channels and feeding back the information to the transmitters. Implicit in this model is the assumption that the channel varies much more slowly than the data rate, so that the tracking of the channel variations can be done accurately and the amount of bits required for feedback is negligible compared to that required for transmitting information. Whereas the transmitters send at constant transmitter power when they do not know the current state of the channel, dynamic power control can be done in response to the changing channels when the transmitters can track the channels. We are interested in characterizing the capacity region in this scenario, with the side-information of the current state of the channel available at both the transmitters and the receiver. Again, we will call this the *throughput capacity* region.

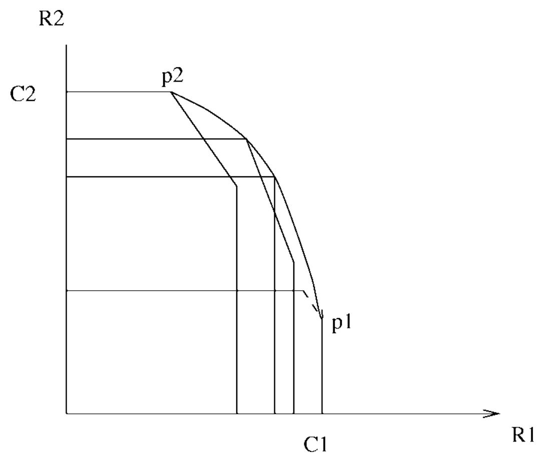

Fig. 1. A two-user throughput capacity region as a union of capacity regions, each corresponding to a feasible power control $\mathcal{P}$. Note that each of these regions is a pentagon. The boundary surface is the curved part.

A *power-control policy* $\mathcal{P}$ is a mapping from the fading state space to $\Re_+^M$. Given a joint fading state $\boldsymbol{h} = (h_1, \cdots, h_M)$ for the users, $\mathcal{P}_i(\boldsymbol{h})$ can be interpreted as the transmitter power allocated to user i. For a given power control policy $\mathcal{P}$, consider the set of rates given by

$$\mathcal{C}_f(\mathcal{P}) \equiv \Bigg\{ \boldsymbol{R} : \boldsymbol{R}(S) \leq \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{1}{\sigma^2}\sum_{i\in S} H_i \mathcal{P}_i(\boldsymbol{H})\right)\right], \quad \forall S \subset \{1, \cdots, M\} \Bigg\} \tag{4}$$

(the subscript f denotes *fading*).

Comparing this with the capacity region (3), one can heuristically think of $\mathcal{C}_f(\mathcal{P})$ as the set of achievable rates when powers are dynamically allocated according to policy $\mathcal{P}$. The following coding theorem substantiates such an interpretation.

Theorem 2.1: The throughput capacity region for the multiaccess fading Gaussian channel when all the transmitters as well as the receiver have side-information of the current state of the channel is given by

$$\mathcal{C}(\bar{\boldsymbol{P}}) \equiv \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_f(\mathcal{P}) \tag{5}$$

where $\mathcal{F}$ is the set of all feasible power control policies satisfying the average power constraint

$$\mathcal{F} \equiv \{\mathcal{P} : \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}_i(\boldsymbol{H})] \leq \bar{P}_i \ \forall i\}.$$

Proof: See Appendix A. □

The above theorem essentially says that the improvement in capacity due to the transmitters having knowledge of the channel state comes solely from the ability to allocate powers according to the channel state. Also, note that since the capacity region is convex, the above characterization implies that time sharing is not required to achieve any point in the capacity region. An example of a two-user capacity region is shown in Fig. 1.

It is worth pointing out that as a result of power control, codewords are random: since the power control depends on the random fading process, so do the codewords themselves. However, consider the multiuser, Gaussian channel with a unit power constraint on each user, and in which the fading level for user i is $H_i\mathcal{P}_i(\boldsymbol{H})$. This channel has capacity region $\mathcal{C}_f(\mathcal{P})$. Consider then any rate $\boldsymbol{R}$ in the interior of $\mathcal{C}_f(\mathcal{P})$. Given any positive ϵ, we can choose a codelength and a codebook (nonrandom) such that the probability of error is less than ϵ. But, as in the proof of Theorem 2.1, we can use this codebook to construct the random codebook for the original fading channel, with the same probability of error. Thus in the original channel, we can use this nonrandom codebook, and scale each symbol by the appropriate power control (dependent on the realization of the fading) to get the random codeword that is transmitted. The receiver can decode since it knows the realization of the fading, and the nonrandom codebooks of the users.

III. Explicit Characterization of the Capacity Region

In this section, we will obtain an explicit characterization of the throughput capacity region (5) as well as the optimal power and rate control policies, and also show that successive decoding is always optimal to get all points on the boundary. We do this by exploiting a special combinatorial structure of the regions $\mathcal{C}_g$ and $\mathcal{C}_f$.

A. Polymatroid Structure

We begin with a few definitions. As before, for a vector $\boldsymbol{v} \in \Re^M$, we shall use the shorthand notation $\boldsymbol{v}(S)$ to denote $\sum_{i\in S} v_i$.

Definition 3.1: Let $E = \{1, \cdots, M\}$ and $f : 2^E \to \Re_+$ be a set function. The polyhedron

$$\mathcal{B}(f) \equiv \{(x_1, \cdots, x_M) : \boldsymbol{x}(S) \leq f(S) \quad \forall S \subset E, x_i \geq 0 \ \forall i\} \tag{6}$$

is a *polymatroid* if the set function f satisfies

1) $f(\emptyset) = 0$ (normalized).
2) $f(S) \leq f(T)$ if $S \subset T$ (nondecreasing).
3) $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ (submodular).

The polyhedron

$$\mathcal{G}(f) \equiv \{(x_1, \cdots, x_M) : \boldsymbol{x}(S) \geq f(S) \ \forall S \subset E\}$$

is a *contra-polymatroid* if f satisfies

1) $f(\emptyset) = 0$ (normalized).
2) $f(S) \leq f(T)$ if $S \subset T$ (nondecreasing).
3) $f(S) + f(T) \leq f(S \cup T) + f(S \cap T)$ (supermodular).

If f satisfies the three properties, f is called a **rank** function in both cases.

Polymatroids were introduced by Edmonds [4] where he proved the following key properties. If π is a permutation on the set E, define the vector $\boldsymbol{v}(\pi) \in \Re^M$ by $v_{\pi(1)}(\pi) = f(\pi(1))$ and

$$v_{\pi(i)}(\pi) = f(\{\pi(1), \cdots, \pi(i)\}) - f(\{\pi(1), \cdots, \pi(i-1)\})$$

for $i = 2, \cdots, M$.

Lemma 3.2: Let $\mathcal{B}(f)$ be a polymatroid. Then $v(\pi)$ is a vertex of $\mathcal{B}(f)$ for every permutation π. Also, any vertex of $\mathcal{B}(f)$ strictly inside the positive orthant must be $v(\pi)$ for some π. Moreover, if $\boldsymbol{\lambda}$ is a given vector in $\Re_+^M$, then a solution of the optimization problem

$$\max \boldsymbol{\lambda} \cdot \boldsymbol{x} \quad \text{subject to } \boldsymbol{x} \in \mathcal{B}(f) \tag{7}$$

is attained at a point $v(\pi^*)$, where the π^* is any permutation such that $\lambda_{\pi^*(1)} \geq \cdots \geq \lambda_{\pi^*(M)}$. Conversely, suppose f is a set function and $\mathcal{B}(f)$ is the polyhedron defined in (6). Then if $v(\pi) \in \mathcal{B}(f)$ for every permutation π, then $\mathcal{B}(f)$ is a polymatroid.

Note that $\mathcal{B}(f)$ is a polyhedron characterized by an exponentially large number of constraints (in M). The above lemma says that the polymatroid structure of $\mathcal{B}(f)$ allows the linear program (7) to be solved efficiently, in fact in time $O(M \log M)$. One can in fact re-interpret the solution of the linear program as that obtained from the following *greedy algorithm*:

- **Initialization**: Set $x_i = 0$ for all i. Set $k = 1$.
- **Step** k: Increase the value of $x_{\pi^*(k)}$ until a constraint becomes tight. Goto Step $k+1$.
- After M steps, the optimal solution is reached.

It can be shown, by the properties of f, that at step k, the constraint that becomes tight is the one that corresponds to the subset $\{\pi(1), \cdots, \pi(k)\}$. Thus this algorithm yields the solution in Lemma 3.2. It is said to be greedy since it is always moving in the direction of steepest ascent of the objective function while staying inside the feasible region. More importantly, after increasing a component of the vector, the algorithm never revisits it again. Thus only M steps are required. We will see that the solutions to all the optimization problems in this paper have this greedy character.

There is an analogous lemma for contra-polymatroids.

Lemma 3.3: Let $\mathcal{G}(f)$ be a contra-polymatroid. Then the points $v(\pi)$ where π is a permutation on E are precisely the vertices of $\mathcal{G}(f)$. Moreover, if $\boldsymbol{\lambda}$ is a given vector in $\Re_+^M$, then a solution of the optimization problem

$$\min \boldsymbol{\lambda} \cdot \boldsymbol{x} \quad \text{subject to } \boldsymbol{x} \in \mathcal{G}(f) \tag{8}$$

is attained at a point $v(\pi^*)$ where π^* is any permutation such that $\lambda_{\pi^*(1)} \geq \cdots \geq \lambda_{\pi^*(M)}$. Conversely, if f is a set function and $v(\pi) \in \mathcal{G}(f)$ for every permutation π, then $\mathcal{G}(f)$ is a contra-polymatroid.

Now consider a discrete memoryless multiaccess channel with transition matrix $p(y|x_1, \cdots, x_M)$. A similar version of the following result was obtained in [10].

Lemma 3.4: For any independent distribution $p(x_1) \cdots p(x_M)$ on the inputs, the polyhedron

$$\{\boldsymbol{R} \in \Re_+^M : \boldsymbol{R}(S) \leq I[Y; \boldsymbol{X}(S) | \boldsymbol{X}(S^c)] \ \forall S \subset E\} \tag{9}$$

is a polymatroid.

Proof: One can directly verify the submodularity of the mutual information function. A shorter proof is as follows. Let π be a permutation on E and consider the rate vector $R(\pi)$ defined by

$$R_{\pi(i)}(\pi) = I[Y; X_{\pi(i)} | \boldsymbol{X}(\{\pi(i+1), \cdots, \pi(M)\})], \quad i = 1, \cdots, M-1$$
$$R_{\pi(M)}(\pi) = I[Y; X_{\pi(M)}].$$

These are the capacities achieved by successive decoding in the order given by π, and hence the rate vector $\boldsymbol{R}(\pi)$ lies in the region (9). Since this is true for every π, by Lemma 3.2, the polyhedron (9) is a polymatroid. □

Corollary 3.5: The capacity region $\mathcal{C}_g(\boldsymbol{h}, \boldsymbol{P})$ of a memoryless Gaussian multiaccess channel is a polymatroid.

Lemma 3.6: Let $\mathcal{P}$ be any power control policy. Then $\mathcal{C}_f(\mathcal{P})$ defined in (4) is a polymatroid.

Proof: By direct verification. □

The following structural result shows that the region $\mathcal{C}_f(\mathcal{P})$ can be written as a weighted sum of the capacity regions of parallel time-invariant Gaussian channels $\mathcal{C}_g(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}))$.

Definition 3.7: A rate allocation policy $\mathcal{R}$ is a mapping from the set of joint fading states to $\Re_+^M$; for each fading state $\boldsymbol{h}$, $\mathcal{R}_i(\boldsymbol{h})$ can be interpreted as the rate allocated to user i while the users are in state $\boldsymbol{h}$.

Lemma 3.8: For any power control policy $\mathcal{P}$

$$\mathcal{C}_f(\mathcal{P}) = \{\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}(\boldsymbol{H})] : \mathcal{R} \text{ is a rate allocation policy s.t. } \forall \boldsymbol{h} \ \mathcal{R}(\boldsymbol{h}) \in \mathcal{C}_g(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}))\}. \tag{10}$$

Furthermore, for any permutation π on E

$$v(\pi) = \mathbb{E}_{\boldsymbol{H}}[v_{\boldsymbol{H}}(\pi)] \tag{11}$$

where $v(\pi)$ is the vertex of $\mathcal{C}_f(\mathcal{P})$ corresponding to the permutation π, and for each state $\boldsymbol{h}$, $v_{\boldsymbol{h}}(\pi)$ is the vertex of $\mathcal{C}_g(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}))$ corresponding to permutation π.

Proof: Define

$$\mathcal{E} \equiv \{\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}(\boldsymbol{H})] : \mathcal{R} \text{ is a rate allocation policy s.t. } \mathcal{R}(\boldsymbol{h}) \in \mathcal{C}_g(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}))\}.$$

By definition, we have that $\mathcal{E} \subseteq \mathcal{C}_f(\mathcal{P})$. But by Lemma 3.6, $\mathcal{C}_f(\mathcal{P})$ is a polymatroid, and hence is the convex hull of successive decoding points $\boldsymbol{R}(\pi)$, where π ranges over all permutations of E, and

$$\sum_{i=1}^{n} R_{\pi_i} = \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{\sum_{i=1}^{n} H_{\pi_i}\mathcal{P}_{\pi_i}(\boldsymbol{H})}{\sigma^2}\right)\right], \quad n = 1, 2, \cdots, M.$$

But for any π, $\boldsymbol{R}(\pi) \in \mathcal{E}$, and hence every extreme point of $\mathcal{C}_f(\mathcal{P})$ lies in $\mathcal{E}$. By the convexity of $\mathcal{E}$, it follows that $\mathcal{E} = \mathcal{C}_f(\mathcal{P})$. This also establishes the second part of the lemma. □

B. A Lagrangian Characterization of the Capacity Region

We shall now make use of the polymatroid structure of $\mathcal{C}_g(\boldsymbol{h}, \boldsymbol{P})$ and $\mathcal{C}_f(\mathcal{P})$ to explicitly characterize the throughput capacity region $\mathcal{C}(\bar{\boldsymbol{P}})$ of the multiaccess fading channel and the optimal power control policies, under an average power constraint $\bar{\boldsymbol{P}}$.

We focus on characterizing the boundary of the region $\mathcal{C}(\bar{\boldsymbol{P}})$, as given in the following definition.

Definition 3.9: The boundary surface of $\mathcal{C}(\bar{\boldsymbol{P}})$ is the set of those rates such that no component can be increased with the other components remaining fixed, while remaining in $\mathcal{C}(\bar{\boldsymbol{P}})$.

For example, the boundary surface of the Gaussian capacity region without fading is simply the points where the constraint for the entire set of users is tight. The points on the boundary surface are in some sense the optimal operating points because any other point in the capacity region is dominated componentwise by some point on the boundary surface. In the two-user example in Fig. 1, the boundary surface is the curved part.

The following lemma shows that the computation of the boundary of the region $\mathcal{C}(\bar{\boldsymbol{P}})$ and the associated optimal power control policy can be reduced to solving a family of optimization problems over a set of parallel multiaccess Gaussian channels.

Lemma 3.10: The boundary surface of $\mathcal{C}(\bar{\boldsymbol{P}})$ is the closure of all points $\boldsymbol{R}^*$ such that $\boldsymbol{R}^*$ is a solution to the optimization problem

$$\max_{\boldsymbol{R}} \mu \cdot \boldsymbol{R} \text{ subject to } \boldsymbol{R} \in \mathcal{C}(\bar{\boldsymbol{P}}) \tag{12}$$

for some positive $\mu \in \Re_+^M$. For a given $\boldsymbol{\mu}$, $\boldsymbol{R}^*$ is a solution to the above problem if and only if there exists a $\boldsymbol{\lambda} \in \Re_+^M$, rate allocation policy $\mathcal{R}(\cdot)$, and power control policy $\mathcal{P}(\cdot)$ such that for every joint fading state $\boldsymbol{h}$, $(\mathcal{R}(\boldsymbol{h}), \mathcal{P}(\boldsymbol{h}))$ is a solution to the optimization problem

$$\max_{(\boldsymbol{r},\boldsymbol{p})} \boldsymbol{\mu} \cdot \boldsymbol{r} - \boldsymbol{\lambda} \cdot \boldsymbol{p} \quad \text{subject to } \boldsymbol{r} \in \mathcal{C}_g(\boldsymbol{h}, \boldsymbol{p}) \tag{13}$$

and

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}_i(\boldsymbol{H})] = R_i^*, \quad \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}_i(\boldsymbol{H})] = \bar{P}_i, \qquad i = 1, \cdots, M$$

where $\bar{P}_i$ is the constraint on the average power of user i.

Proof: The first statement follows from the convexity of the capacity region.

Now consider the set

$$S = \{(\boldsymbol{R}, \boldsymbol{P}) : \boldsymbol{P} \in \Re_+^M, \boldsymbol{R} \in \mathcal{C}(\boldsymbol{P})\}.$$

By the concavity of the log function, it can readily be verified that S is a convex set. Thus there exist Lagrange multipliers $\boldsymbol{\lambda} \in \Re_+^M$ such that $\boldsymbol{R}^*$ is a solution to the optimization problem

$$\max_{(\boldsymbol{R},\boldsymbol{P}) \in S} \boldsymbol{\mu} \cdot \boldsymbol{R} - \boldsymbol{\lambda} \cdot \boldsymbol{P}. \tag{14}$$

Now

$$\mathcal{C}(\boldsymbol{P}) = \bigcup_{\{\mathcal{P}: \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}(\boldsymbol{H})] \le \boldsymbol{P}\}} \mathcal{C}_f(\mathcal{P}) = \bigcup_{\{\mathcal{P}: \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}(\boldsymbol{H})] = \boldsymbol{P}\}} \mathcal{C}_f(\mathcal{P})$$

and hence we can rewrite (14) as an optimization over the set of power control laws

$$\max_{(\boldsymbol{R},\mathcal{P})} \boldsymbol{\mu} \cdot \boldsymbol{R} - \boldsymbol{\lambda} \cdot \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}(\boldsymbol{H})] \quad \text{subject to } \boldsymbol{R} \in \mathcal{C}_f(\mathcal{P}). \tag{15}$$

Let π be the permutation corresponding to a decreasing ordering of the components of the vector $\boldsymbol{\mu}$. By the polymatroid structure of $\mathcal{C}_f(\mathcal{P})$, for any given power control $\mathcal{P}$, $\boldsymbol{\mu} \cdot \boldsymbol{R}$ is maximized at

$$R_{\pi(1)} = \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{H_{\pi(1)}\mathcal{P}_{\pi(1)}(\boldsymbol{H})}{\sigma^2}\right)\right]$$

$$R_{\pi(k)} = \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{H_{\pi(k)}\mathcal{P}_{\pi(k)}(\boldsymbol{H})}{\sigma^2 + \sum_{i=1}^{k-1} H_{\pi(i)}\mathcal{P}_{\pi(i)}(\boldsymbol{H})}\right)\right], \quad k = 2, \cdots, M. \tag{16}$$

Hence, the optimization problem (15) is equivalent to

$$\max_{\mathcal{P}} \mu_1 \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{H_{\pi(1)}\mathcal{P}_{\pi(1)}(\boldsymbol{H})}{\sigma^2}\right)\right] + \sum_{k=2}^{M} \mu_k \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{2}\log\left(1 + \frac{H_{\pi(k)}\mathcal{P}_{\pi(k)}(\boldsymbol{H})}{\sigma^2 + \sum_{i=1}^{k-1} H_{\pi(i)}\mathcal{P}_{\pi(i)}(\boldsymbol{H})}\right)\right] - \boldsymbol{\lambda} \cdot \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}(\boldsymbol{H})] \tag{17}$$

and this is, in turn, equivalent to

$$\max_{\boldsymbol{p}} \mu_1 \frac{1}{2}\log\left(1 + \frac{h_{\pi(1)}p_{\pi(1)}}{\sigma^2}\right) + \sum_{k=2}^{M} \mu_k \frac{1}{2}\log\left(1 + \frac{h_{\pi(k)}p_{\pi(k)}}{\sigma^2 + \sum_{i=1}^{k-1} h_{\pi(i)}p_{\pi(i)}}\right) - \boldsymbol{\lambda} \cdot \boldsymbol{p}$$

for every fading state $\boldsymbol{h}$. But this latter problem is also equivalent to

$$\max_{\boldsymbol{r},\boldsymbol{p}} \boldsymbol{\mu} \cdot \boldsymbol{r} - \boldsymbol{\lambda} \cdot \boldsymbol{p} \quad \text{subject to } \boldsymbol{r} \in \mathcal{C}_g(\boldsymbol{h}, \boldsymbol{p})$$

because of the fact that $\mathcal{C}_g$ is a polymatroid.

This completes the proof. □

One can interpret $\boldsymbol{\mu}$ as a vector of rate rewards, prioritizing the users. A point $\boldsymbol{R}^*$ on the boundary for a given $\boldsymbol{\mu}$ is a rate vector which maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$ over the capacity region $C(\bar{\boldsymbol{P}})$. As $\boldsymbol{\mu}$ varies, we get all points on the boundary of the convex capacity region. The vector $\boldsymbol{\lambda}$ can be interpreted as a set of power prices; for a given $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ is chosen such that the average power constraints are satisfied.

It follows immediately from (16) that an optimal solution will be a successive decoding solution. Lemma 3.8 then shows that the optimal solution $(\mathcal{R}^*(\boldsymbol{H}), \mathcal{P}^*(\boldsymbol{H}))$ will be such that $\mathcal{R}^*(\boldsymbol{h})$ is a corner point of $\mathcal{C}_g(\boldsymbol{h}, \mathcal{P}^*(\boldsymbol{h}))$ for every $\boldsymbol{h}$, with the same ordering π for each $\boldsymbol{h}$. However, *a priori*, the optimizing $\boldsymbol{R}^*$ for a given μ may not be unique. We will see though that the continuity of the stationary density of the fading processes implies that it will be unique.

C. Optimal Power and Rate Allocation

We now consider the problem of determining $(\mathcal{R}^*(\boldsymbol{h}), \mathcal{P}^*(\boldsymbol{h}))$ for each fading state $\boldsymbol{h}$. Note that Lemma 3.10 can be viewed as a multiaccess generalization of the Lagrangian formulation for the problem of allocating powers over a set of parallel single-user Gaussian channels [6]. The solution to the optimization problem in the single user setting is given by the classic water-filling construction. Here we will provide a solution in the multiaccess setting. Again we make use of the polymatroid structure and the solution will have a greedy flavor.

To get some intuition about what the solution may look like, let us first reinterpret the classic water-filling solution for the single-user case. The solution in that setting is to solve, for each fading state h, the optimization problem

$$\max_{r,q}\left[r-\frac{\lambda}{h}q\right] \quad \text{subject to} \quad r\leq\frac{1}{2}\log\left(1+\frac{q}{\sigma^2}\right)$$

where we have formulated the problem in terms of the *received* power q. Equivalently, the problem is

$$\max_{q}\left[\frac{1}{2}\log\left(1+\frac{q}{\sigma^2}\right)-\frac{\lambda}{h}q\right].$$

Here, λ is the Lagrange multiplier (power price) chosen such that the average power constraint P is satisfied. We can write

$$\frac{1}{2}\log\left(1+\frac{q}{\sigma^2}\right)=\int_0^q\frac{1}{2(\sigma^2+z)}\,dz.$$

The integral representation can be given a *rate splitting* interpretation, where the single user can be visualized as being split into many low-rate virtual users, each with received power dz. The total rate is achieved by successive cancellation among these virtual users, with the rate achieved by the virtual user decoded as interference level σ^2+z to be $1/[2(\sigma^2+z)]\cdot dz$.

The optimization problem can be recast in the integral form

$$\max_{g}\int_0^g\left[\frac{1}{2(\sigma^2+z)}-\frac{\lambda}{h}\right]dz.$$

Let us define

$$u(z)\equiv\frac{1}{2(\sigma^2+z)}-\frac{\lambda}{h}$$

and interpret $u(z)\cdot dz$ as the *marginal utility* (rate revenue minus power cost) of adding a virtual user at interference level σ^2+z. The optimal solution can be described by adding more virtual users until the marginal utility of adding any further virtual users is negative. In particular, if $u(0)\leq 0$, then nothing is transmitted at all.

Of course, the resulting optimal received power q^* is the same as that of the water-filling solution, and this seems like a rather convoluted way of presenting the solution. However, the interpretation of the single-user solution suggests a natural conjecture for the optimal solution for the multiuser optimization problem (13). Define the marginal utility function for user i to be

$$u_i(z)\equiv\frac{\mu_i}{2(\sigma^2+z)}-\frac{\lambda_i}{h_i}$$

where $u_i(z)\cdot dz$ can be interpreted as the marginal increase in the value of the overall objective function $\mu\cdot\boldsymbol{r}-\lambda\cdot\boldsymbol{p}$ due to adding a low-rate virtual user of received power dz to user i at interference level σ^2+z. Starting at $z=0$, the optimal solution is obtained by choosing at each interference level σ^2+z, to add a virtual user which will lead to the largest positive marginal increase in the objective function. Here, the choice is whether to add such a virtual user, and if so, to which (physical) user. The interference level σ^2+z is the total received power of all virtual users already added, plus the backgraound noise. The decoding is done by successive cancellation in reverse order of the virtual users added to the algorithm. See Fig. 2 for a three-user example.

We see that the proposed solution has a greedy flavor. To prove that this indeed solves the optimization problem (13), we have to identify further polynomial structure in the time-invariant multiaccess Gaussian capacity region $\mathcal{C}(\boldsymbol{h},\boldsymbol{P})$. Solving this problem in turn leads us to a new result in polynomial theory.

Definition 3.11 (see [5]): The rank function f of a polymatroid $\mathcal{B}(f)$ is *generalized symmetric* if there exists a vector $\boldsymbol{y}\in\Re_+^M$ and a nondecreasing concave function g such that $f(S)=g(\boldsymbol{y}(S))$ for every $S\subset E$.

It can be readily verified that such an f satisfies the three properties of a rank function. We state the following easily proven result.

Lemma 3.12: Let g be a nondecreasing concave function and for each $\boldsymbol{y}$, define the generalized symmetric rank function $f_{\boldsymbol{y}}(S)\equiv g(\boldsymbol{y}(S))$. For all vectors $\boldsymbol{x}\in\Re_+^M$, the set $\{\boldsymbol{y}:\boldsymbol{x}\in\mathcal{B}(f_{\boldsymbol{y}})\}$ is a contra-polymatroid.

Applying this to the capacity region $\mathcal{C}_g(\boldsymbol{h},\boldsymbol{P})$, we get the following "dual" polymatroid structure.

Corollary 3.13: For a given average transmitter power constraint $\boldsymbol{p}$ and fixed $\boldsymbol{h}$, the capacity region $\mathcal{C}_g(\boldsymbol{h},\boldsymbol{p})$ is a polymatroid with generalized symmetric rank function. On the other hand, for a given rate vector $\boldsymbol{r}$, the set of *received powers* that can support $\boldsymbol{r}$

$$\mathcal{Q}(h,\boldsymbol{r})=\{\boldsymbol{q}:\exists\,\boldsymbol{p}\text{ s.t. }q_i=h_ip_i,\boldsymbol{r}\in\mathcal{C}_g(h,\boldsymbol{p})\}$$

is a contra-polymatroid.

We wish to solve (13), and note that by Corollary 3.13, it is sufficient to consider the more general problem stated in Theorem 3.14, in terms of a polymatroid with generalized symmetric rank function. The following is a new result.

Theorem 3.14: Consider the problem

$$\max_{(\boldsymbol{x},\boldsymbol{y})}\boldsymbol{\mu}\cdot\boldsymbol{x}-\boldsymbol{\lambda}\cdot\boldsymbol{y} \quad \text{subject to} \quad \boldsymbol{x}(S)\leq g(\boldsymbol{y}(S)) \quad \forall S\subset E$$

where g is a monotonically increasing concave function. Define the marginal *utility functions*

$$u_i(z)\equiv\mu_i g'(z)-\lambda_i, \qquad i=1,\cdots,M$$
$$u^*(z)\equiv\left[\max_i u_i(z)\right]^+.$$

(Here, $x^+\equiv\max(x,0)$.)

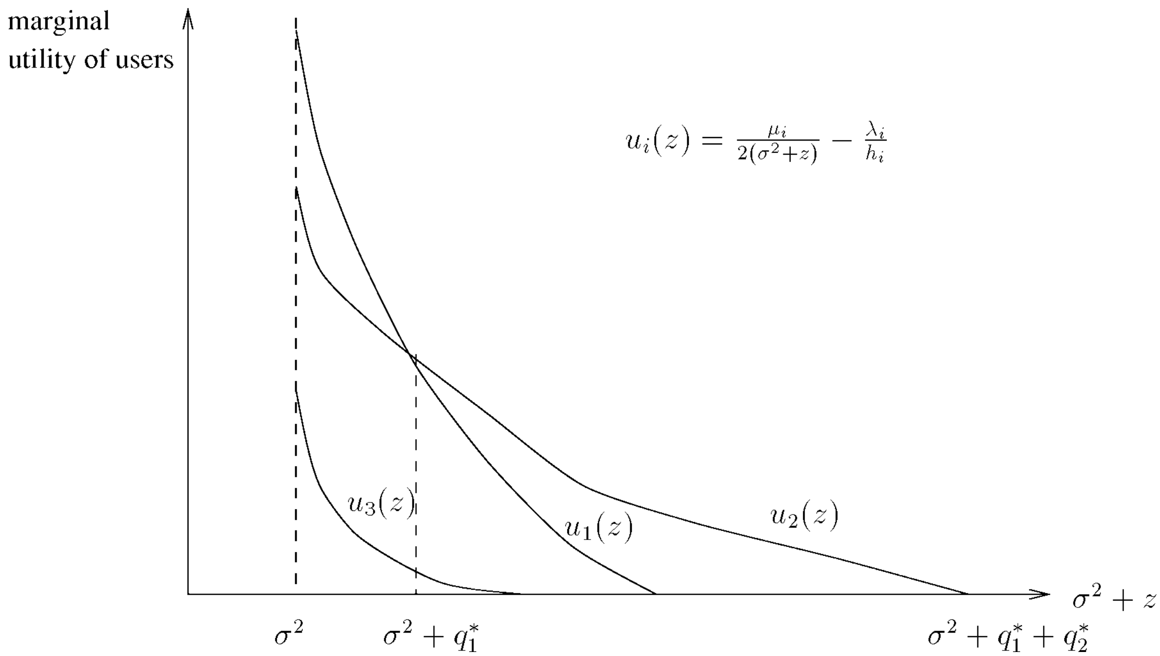

Fig. 2. A three-user example illustrating the greedy power allocation. The x-axis represents the received interference level and y-axis the marginal utility of each user at the interference levels. At each interference level, the (physical) user who is selected to transmit is the one with the highest marginal utility. Here, user 1 gets decoded after user 2, and user 3 gets no power at all. The optimal received powers for user 1 and user 2 are q_1^* and q_2^*, respectively.

Then the solution to the above problem is given by $\int_0^\infty u^*(z)\,dz$ and an optimizing point $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ to achieve this can be found by a greedy algorithm.

Proof: Let J^* be the optimal value for the above problem. For any fixed $\boldsymbol{y}$, the set of feasible $\boldsymbol{x}$ forms a polymatroid, and by Lemma 3.2, the value J^* must be attained at a point satisfying

$$x_{\pi(1)} = g\big(y_{\pi(1)}\big)$$
$$x_{\pi(k)} = g\left(\sum_{i=1}^{k} y_{\pi(i)}\right) - g\left(\sum_{i=1}^{k-1} y_{\pi(i)}\right)$$

for some permutation π. Hence

$$\begin{aligned} J^* &= \max_{\boldsymbol{y}} \mu_{\pi(1)} g\big(y_{\pi(1)}\big) + \sum_{k=2}^{M} \mu_{\pi(k)} \\ &\quad \times \left[g\left(\sum_{i=1}^{k} y_{\pi(i)}\right) - g\left(\sum_{i=1}^{k-1} y_{\pi(i)}\right)\right] - \boldsymbol{\lambda}\cdot\boldsymbol{y} \\ &= \max_{\boldsymbol{y}} \sum_{k=1}^{M} \int_{\sum_{i\le k-1} y_{\pi(i)}}^{\sum_{i\le k} y_{\pi(i)}} u_{\pi(i)}(z)\,dz \\ &\le \int_0^\infty u^*(z)\,dz. \end{aligned}$$

We now show that this upper bound can actually be attained. First, note that by the concavity of g, the function u^* is monotonically decreasing. If $u^*(0) = 0$, then $J^* = 0$ and attained at $\boldsymbol{x} = \boldsymbol{y} = 0$. If $u^*(0) > 0$, then let $0 = z_0 < z_1 < \cdots < z_K$ where z_K is the smallest z for which $u^*(z) = 0$ (if there is no such point, $z_K = \infty$), and such that in the interval $[z_k, z_{k+1}]$, $u^*(z) = u_{i_k}(z)$ for some i_k, $k = 0, \cdots, K-1$. Hence, at z_k, $u_{i_{k-1}}$ intersects u_{i_k}. Now, since g' is monotone, two curves u_i and u_j can intersect at most once. Thus the i_k's are distinct. Pick the point

$$\begin{aligned} y^*_{i_k} &\equiv z_{k+1} - z_k, \qquad k = 0, \cdots K-1 \\ y^*_j &\equiv 0 \quad \text{else} \\ x^*_{i_k} &\equiv g(z_{k+1}) - g(z_k), \qquad k = 0, \cdots K-1 \\ x^*_j &\equiv 0 \quad \text{else.} \end{aligned}$$

It can be directly verified that

$$\boldsymbol{\mu}\cdot\boldsymbol{x}^* - \boldsymbol{\lambda}\cdot\boldsymbol{y}^* = \int_0^\infty u^*(z)\,dz$$

and that $\boldsymbol{x}^*$ is a vertex of the polymatroid with rank function $f(\cdot) \equiv g(\boldsymbol{y}^*(\cdot))$. Thus the upper bound is attained at $(\boldsymbol{x}^*, \boldsymbol{y}^*)$. □

Observe that the solution can be obtained via a greedy algorithm. Starting with $\boldsymbol{x} = \boldsymbol{y} = 0$, the component that gets selected to be increased is the one which leads to the steepest ascent of the objective function. When none of the components leads to an increase in the objective function, the optimal solution is reached. Moreover, the algorithm never revisits a component after finishing increasing it.

Specializing this result to the case of the time-invariant Gaussian channel gives exactly the proposed solution to the optimization problem (13) discussed earlier. The function g is taken to be

$$g(z) \equiv \frac{1}{2}\log\left(1 + \frac{z}{\sigma^2}\right).$$

In terms of the received powers $\boldsymbol{q} = (h_1 p_1, \cdots, h_M p_M)$, the optimization problem can be rewritten as

$$\max \sum_i \mu_i r_i - \sum_i \frac{\lambda_i}{h_i} q_i \quad \text{subject to } \boldsymbol{r}(S) \le g(\boldsymbol{q}(S))\ \forall S \subset E.$$

Moreover, it shows that the optimal solution is achieved by successive decoding among the *actual users*. Any such solution can be represented by a permutation π and a set of intervals $[z, z_{i+1}]$, $i = 1, \cdots, M$ of the real line such that $z_1 = 0, z_{i+1} - z_i$ is the received power of user $\pi(i)$, and users are decoded in the order given by $\pi(M), \pi(M-1), \cdots, \pi(1)$. The value z_i is the total received power of the interfering users when user $\pi(i)$ is decoded. Thus user $\pi(i)$ is decoded at a total interference level of $\sigma^2 + z_i$. One can also think of a solution as the choice of which (if any) user to transmit at every interference level $\sigma^2 + z$, $z \in [0, \infty)$. Refer again to Fig. 2 for an example. Note that in the optimal solution, some users may be allocated zero powers (and hence zero rates), although the priority order (the reverse of the decoding order) of the transmitting users is always in increasing order of the rate rewards μ_i's.

At a given fading state $\boldsymbol{h}$, the optimal rate and power allocated are not unique only when the utility functions of two users coincide, i.e., $\mu_i = \mu_j$ and $\frac{\lambda_i}{h_i} = \frac{\lambda_j}{h_j}$ for some i, j. But since the users have a joint fading distribution with a continuous density, this will only happen on a set of fading states with probability 0. Thus with probability 1, the optimal power and rate allocation is unique and is explicitly given by

$$\mathcal{R}_i^*(\boldsymbol{h}) = \int_{\mathcal{A}_i} \frac{1}{2(\sigma^2 + z)}\, dz \qquad \mathcal{P}_i^*(\boldsymbol{h}) = |\mathcal{A}_i|$$

where

$$\mathcal{A}_i \equiv \{z \in [0, \infty) : u_i(z) > u_j(z)\ \forall j \neq i \text{ and } u_i(z) > 0\}.$$

The proof of Theorem 3.14 illustrates the fact that the optimal point will be a corner point for every fading state, although this also follows directly from Lemmas 3.8 and 3.10.

D. Boundary of the Capacity Region

We now combine the Lagrangian formulation given in Lemma 3.10 and the optimal power and rate allocation solution to give a characterization of the capacity region $\mathcal{C}(\bar{\boldsymbol{P}})$, parameterized by the rate rewards μ. First, we present the following lemma, which allows us to have a well-defined parameterization of the boundary of the capacity region by the rate rewards $\boldsymbol{\mu}$.

Lemma 3.15: Let $\boldsymbol{\mu}$ be a given positive rate reward vector. Then there is a unique $\boldsymbol{R}^*$ on the boundary which maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$, and there is a unique Lagrangian power price $\boldsymbol{\lambda}$ such that the optimal power allocation solving (13) satisfies the average power constraints.

Proof: See Appendix B. □

In the two-user example shown in Fig. 1, this means that every line of negative slope has a unique point of tangency at the curved boundary. In particular, there is no linear part on this boundary. This is in contrast to the (nonfading) Gaussian multiaccess capacity region, where the boundary is a line with slope -1 (in the two-user case). Thus even when $\mu_1 = \mu_2$, the optimal $\boldsymbol{R}^*$ is unique. This is true because when the fading distributions have continuous density, the optimal rate and power allocations are unique with probability 1 even when $\mu_1 = \mu_2$, as explained at the end of Section III-C. Every point on the boundary surface is a corner point of some pentagon $\mathcal{C}_f(\mathcal{P})$, which is the capacity region for a particular power control policy $\mathcal{P}$. The point corresponding to $\mu_1 = \mu_2$ is a corner point of a degenerate pentagon, i.e., a rectangle. Why this is so will be explained later in this subsection.

It should also be noted that the uniqueness result above only holds for positive $\boldsymbol{\mu}$. If some of the rewards μ_i's equal 0, the $\boldsymbol{R}^*$ which maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$ may not be unique. However, it is clear that one can get arbitrarily close to these points (the extreme points of the boundary surface) by letting some of the rewards go to zero. Thus it suffices to focus on the strictly positive reward vectors $\boldsymbol{\mu}$ for a parameterization of the boundary surface. We will give a more explicit interpretation of these extreme points in Section III-E.

For any such positive $\boldsymbol{\mu}$, the above lemma implies that we can define a parameterization $\boldsymbol{R}^*(\mu)$ which is the unique rate vector on the boundary which maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$. Its value can be obtained using the greedy rate and power allocation solution, with $\boldsymbol{\lambda}$ chosen such that the average power constraints are satisfied. In the common case when the fading processes of the users are independent of each other, $\boldsymbol{R}^*(\mu)$ has a particularly simple form.

For the given $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, let $\mathcal{R}^*(\boldsymbol{h}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ and $\mathcal{P}^*(\boldsymbol{h}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ be the optimal solution to the problem (13). Since the stationary distributions of the fading processes have a continuous density, $\Pr(H_i = H_j) = 0$ for all $i \neq j$. We observe that the choice of which user to transmit at each interference level z only depends on the values of the marginal utility functions of the users at z. Thus the average rate and power of each user can be computed first at each interference level z and then integrated over all z. Thus

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}_i^*(\boldsymbol{H}, \boldsymbol{\mu}, \boldsymbol{\lambda})] \\
&\quad = \int_0^\infty \mathbb{E}_{\boldsymbol{H}}\left[g'(z) I_{\{u_i(z) > u_j(z) \quad \forall j \text{ and } u_i(z) > 0\}}\right] dz \\
&\quad = \int_0^\infty g'(z) \Pr(u_i(z) > u_j(z) \quad \forall j \text{ and } u_i(z) > 0)\, dz \\
&\quad = \int_0^\infty \frac{1}{2(\sigma^2 + z)} \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \right. \\
&\qquad\qquad \left. \prod_{k \neq i} F_k\left(\frac{2\lambda_k h(\sigma^2 + z)}{2\lambda_i(\sigma^2 + z) + (\mu_k - \mu_i)h}\right) f_i(h)\, dh \right\} dz \\
&\mathbb{E}_{\boldsymbol{H}}[\mathcal{P}_i^*(\boldsymbol{H}, \boldsymbol{\mu}, \boldsymbol{\lambda})] \\
&\quad = \int_0^\infty \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{h_i} I_{\{u_i(z) > u_j(z) \quad \forall j \text{ and } u_i(z) > 0\}}\right] dz \\
&\quad = \int_0^\infty \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \frac{1}{h} \right. \\
&\qquad\quad \left. \times \prod_{k \neq i} F_k\left(\frac{2\lambda_k h(\sigma^2 + z)}{2\lambda_i(\sigma^2 + z) + (\mu_k - \mu_i)h}\right) f_i(h)\, dh \right\} dz
\end{aligned} \tag{18}$$

where F_i and f_i are the cdf and pdf of the stationary distribution of the fading process for user i, respectively.

Combining this with Lemmas 3.10 and 3.15, we have the following characterization of the throughput capacity region $\mathcal{C}(\bar{P})$. Note that since $\mathcal{R}^*$ and $\mathcal{P}^*$ are invariant under scalings of the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, we can normalize such that $\sum_i \mu_i = 1$.

Theorem 3.16: Assume that the fading processes of users are independent of each other. The boundary of $\mathcal{C}(\bar{P})$ is the closure of the parametrically defined surface

$$\left\{ \boldsymbol{R}^*(\boldsymbol{\mu}) : \boldsymbol{\mu} \in \Re_+^M, \sum_i \mu_i = 1 \right\}$$

where for $i = 1, \cdots, M$

$$R_i^*(\boldsymbol{\mu}) = \int_0^\infty \frac{1}{2(\sigma^2+z)} \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k \left(\frac{2\lambda_k h(\sigma^2+z)}{2\lambda_i(\sigma^2+z)+(\mu_k-\mu_i)h} \right) f_i(h)\, dh \right\} dz \tag{19}$$

where the vector $\boldsymbol{\lambda}$ is the unique solution of the equations

$$\int_0^\infty \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \frac{1}{h} \times \prod_{k \neq i} F_k \left(\frac{2\lambda_k h(\sigma^2+z)}{2\lambda_i(\sigma^2+z)+(\mu_k-\mu_i)h} \right) f_i(h)\, dh \right\} dz = \bar{P}_i \tag{20}$$

$i = 1, \cdots, M$. Moreover, every point can be attained by successive decoding.

Note that due to the special structure of the optimal power control policy, the various expectation terms have been reduced from M-dimensional integrals to two-dimensional integrals. For a given $\boldsymbol{\mu}$, it should therefore be possible to compute $\boldsymbol{\lambda}$ numerically with low complexity. We shall present an algorithm to do this in Section IV, but first let us examine several special cases of Theorem 3.16.

1) **Single-User Channel**: When $M = 1$, the above result reduces to characterizing the capacity of the power-controlled single-user time-varying channel

$$\begin{aligned} R^* &= \int_0^\infty \frac{1}{2(\sigma^2+z)} \left\{ \int_{\frac{2\lambda(\sigma^2+z)}{\mu}}^\infty f(h)\, dh \right\} dz \\ &= \int_0^\infty \frac{1}{2} \log \left(1 + \frac{h}{\sigma^2} \left(\frac{\mu}{2\lambda} - \frac{\sigma^2}{h} \right)^+ \right) f(h)\, dh \end{aligned}$$

by reversing the order of integration. Using (20), the constant $\frac{\mu}{2\lambda}$ is shown to satisfy the power constraint

$$\int_0^\infty \left(\frac{\mu}{2\lambda} - \frac{\sigma^2}{h} \right)^+ f(h)\, dh = \bar{P}.$$

This is just the classic water-filling solution to the problem of power allocation over a set of parallel single-user channels, one for each fading level h. This result was obtained by Goldsmith and Varaiya [8] in the context of the single-user time-varying fading channel. The strategy has the characteristic that more power is used when the channel is good and little or even no power when it is bad.

2) **Maximum Sum-Rate Point**: If we set $\mu_1 = \cdots = \mu_M = 1$, we get the point on the boundary of the capacity region that maximizes the sum of the rates of the individual users. For this case, the marginal utility functions $u_i(z)$'s are given by

$$u_i(z) = \frac{1}{2(\sigma^2+z)} - \frac{\lambda_i}{h_i}.$$

We note that for a given fading state $\boldsymbol{h}$, the marginal utility function of the user with the smallest $\frac{\lambda_i}{h_i}$ dominates all the others for all z. This means that in the optimal strategy, at most one user is allowed to transmit at any given fading state. The optimal power control strategy $\mathcal{P}^*$ can be readily calculated to be

$$\mathcal{P}_i^*(\boldsymbol{h}, \boldsymbol{\lambda}) = \begin{cases} \left(\frac{1}{2\lambda_i} - \frac{\sigma^2}{h_i} \right)^+, & \text{if } h_i > \frac{\lambda_i}{\lambda_j} h_j \text{ for all } j \\ 0, & \text{else.} \end{cases}$$

The optimal rates are given by

$$\begin{aligned} R_i^* = \int_0^\infty & \frac{1}{2} \log \left(1 + \frac{h}{\sigma^2} \left(\frac{1}{2\lambda_i} - \frac{\sigma^2}{h} \right)^+ \right) \\ & \times \prod_{k \neq i} F_k \left(\frac{\lambda_k}{\lambda_i} h \right) f(h)\, dh, \qquad i = 1, \cdots, M \end{aligned}$$

where the constants λ_i's satisfy

$$\int_0^\infty \left(\frac{1}{2\lambda_i} - \frac{\sigma^2}{h} \right)^+ \prod_{k \neq i} F_k \left(\frac{\lambda_k}{\lambda_i} h \right) f(h)\, dh = \bar{P}_i, \qquad i = 1, \cdots, M.$$

This solution was recently obtained by Knopp and Humblet [14]. Note that this power control gives rise to a time-division multiple-access strategy. This explains why in the two-user example of Fig. 1, the point on the boundary corresponding to $\mu_1 = \mu_2$ is in fact the corner point of a rectangular $\mathcal{C}_f(\mathcal{P})$.

3) **Multiple Classes of Users**: While the above strategy maximizes the total throughput of the system, it can be unfair if the fading processes of the users have very different statistics. For example, some of the users may be far away from the base station; they will get lower rates through since their channel is worse that that of the nearby users a lot of the time (there are, of course, still other sources of fluctuations of the channels, such as fading at a faster time scale due to multipaths). One way of remedying this situation is to assign unequal rate rewards to users. Let us consider an example where there are two classes of users. Users in the same class have the same fading statistics and power constraints; the first class can represent users at the cell boundary, while the other class consists of users close to the base station. To maintain equal rates for everyone, we can assign rate rewards μ_1 to all users in class 1, and μ_2 to users in class 2, with $\mu_1 > \mu_2$. By symmetry, the power prices of users in the same class are the same. We observe that at any fading state, the marginal utility function of the user with the best channel within each class dominates those of other users in the same class. Thus the optimal strategy has the form that at each fading state, only the strongest user in each class transmits, and the two

users are decoded by successive cancellation, with the nearby user decoded first. This gives an advantage to the user far away. Adjusting the rate rewards can be thought of as a way to maintain fairer allocation of resources to the users. We consider this issue further in Section V.

Note that in the first two examples, the optimal power control strategy has the special characteristic that the power allocated at each fading state $\boldsymbol{h}$ depends only on $\boldsymbol{h}$ and the Lagrange multipliers. For the general case, the allocated power depends on one additional variable z representing the interference level.

E. Extreme Points of the Boundary Surface

In the previous subsection, we parameterize the boundary of the capacity region by positive reward vectors. By letting some of the rate rewards approach 0, one can get arbitrarily close to the extreme points. We can also give an explicit characterization of the extreme points as follows.

Suppose $\mathcal{L}$ is a set of subsets of $E \equiv \{1, 2, \cdots, M\}$ with the property that all subsets in $\mathcal{L}$ are nested. By this we mean that if $F_1, F_2 \in \mathcal{L}$ then $F_1 \subseteq F_2$ or $F_2 \subseteq F_1$. Then it is possible to insist that all users in a subset in $\mathcal{L}$ are decoded, and canceled, before any user in the complementary subset is decoded, for *every* fading state $\boldsymbol{h}$. With positive vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, we can define the decoding order in each subset, just as before, except that now there is absolute priority given to each subset of users in $\mathcal{L}$ over its complement. The extreme points of the boundary surface of $\mathcal{C}(\boldsymbol{P})$ are characterized in exactly this way: by a positive $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ pair, together with a set of nested subsets of users $\mathcal{L}$.

For example, in the two-user case, as $\frac{\mu_2}{\mu_1} \to 0$, the optimal power allocation and the resulting rate for user 1 approaches that for the single-user fading channel with only user 1 present, i.e., a water-pouring solution. This is the point p_1 in Fig. 1, with user 1 achieving rate C_1. User 2 is always decoded before user 1 in every fading state, and the optimal power control for user 2 is also water-pouring, but regarding the sum of the interference created by user 1 and the background noise as the time-varying noise power. Thus we get to an extreme point of the boundary.

IV. An Iterative Algorithm for Resource Allocation

In Section III-B, we provided a Lagrangian characterization of the boundary surface of $\mathcal{C}(\bar{\boldsymbol{P}})$. In particular, we characterize a boundary point by a positive rate rewards vector $\boldsymbol{\mu}$, and that associated with this is a unique positive shadow power price vector $\boldsymbol{\lambda}$. We now present a simple iterative algorithm to compute $\boldsymbol{\lambda}$ for a given $\boldsymbol{\mu}$ and average power constraints $\bar{\boldsymbol{P}}$. In the case when the fadings of the users are independent, this amounts to solving the nonlinear equations (20) for $\boldsymbol{\lambda}$ in Theorem 3.16. Moreover, the iterative algorithm has a natural adaptive implementation when the exact fading statistics are not known.

Throughout this subsection, we assume a vector of rate rewards $\boldsymbol{\mu}$ and power constraints $\bar{\boldsymbol{P}}$ to be given and fixed. Let us define $\boldsymbol{R}(\boldsymbol{\lambda})$ and $\boldsymbol{P}(\boldsymbol{\lambda})$ to be the rate and average powers under the optimal power control associated with the prices $(\boldsymbol{\mu}, \boldsymbol{\lambda})$. We first present the following monotonicity lemma, which can be verified directly from the greedy power allocation algorithm.

Lemma 4.1: For all i, if the ith component of $\boldsymbol{\lambda}$ is increased and the other components are held fixed, $P_i(\boldsymbol{\lambda})$ decreases while $P_j(\boldsymbol{\lambda})$ increases for $j \neq i$. More generally, for any subset S, if we increase λ_i for all $i \in S$, and hold the remaining λ_j fixed, then average powers of users in S^c will increase.

Given average power $\bar{\boldsymbol{P}}$, let $\boldsymbol{R}^*$ be the optimum rate corresponding to the rewards $\boldsymbol{\mu}$, and let $\boldsymbol{\lambda}^*$ be the shadow power prices. Algorithm 4.2 below generates a sequence $\lambda(n)$ from any starting point $\lambda(0)$ that converges to λ^*.

Algorithm 4.2: Let $\boldsymbol{\lambda}(0)$ be an initial arbitrary set of positive power prices. Given the nth iterate $\boldsymbol{\lambda}(n)$, the $n+1$th iterate $\boldsymbol{\lambda}(n+1)$ is given by the following: for each i, $\lambda_i(n+1)$ is the unique power price for the ith user such that the average power of user i is $\bar{P}_i$ under the optimal power control policy when the power prices of the other users remain fixed at $\boldsymbol{\lambda}(n)$. (The uniqueness follows from the monotonicity property above.)

In terms of (20) for the case when the fading is independent, $\lambda_i(n+1)$ is the unique root x of the equation

$$\int_0^\infty \left\{ \int_{\frac{2x(\sigma^2+z)}{\mu_i^*}}^\infty \frac{1}{h} \times \prod_{k \neq i} F_k\left(\frac{2\lambda_k(n)h(\sigma^2+z)}{2x(\sigma^2+z) + (\mu_k^* - \mu_i^*)h} \right) f_i(h)\, dh \right\} dz = \bar{P}_i$$

which can be solved by binary search if the statistics of the fading are known. Otherwise, one can update the power prices by directly measuring the change in the average power consumption.

Theorem 4.3: Given average power $\bar{\boldsymbol{P}}$, let $\boldsymbol{R}^*$ be the optimum rate corresponding to the rewards $\boldsymbol{\mu}^*$, and let $\boldsymbol{\lambda}^*$ be the shadow prices at the point $(\bar{\boldsymbol{P}}, \boldsymbol{R}^*)$. Then

$$\boldsymbol{\lambda}(n) \to \boldsymbol{\lambda}^*, \qquad n \uparrow \infty$$

and hence $\boldsymbol{R}(\boldsymbol{\lambda}(n)) \to \boldsymbol{R}^*$, and $\boldsymbol{P}(\boldsymbol{\lambda}(n)) \to \bar{\boldsymbol{P}}$.

To prove this theorem, we first consider the following lemma.

Lemma 4.4:

i) For any positive $\boldsymbol{\lambda}(0)$, there exists $\boldsymbol{\lambda} \leq \boldsymbol{\lambda}(0)$ for which $P(\boldsymbol{\lambda}) \geq \bar{P}$.

ii) For any positive $\boldsymbol{\lambda}(0)$, there exists $\boldsymbol{\lambda} \geq \boldsymbol{\lambda}(0)$ for which $P(\boldsymbol{\lambda}) \leq \bar{P}$.

Proof: See Appendix C. □

Algorithm 4.2 defines a mapping

$$T : \Re_+^M \to \Re_+^M$$
$$\lambda(n) \mapsto \lambda(n+1).$$

The following properties of T are useful in the proof of Theorem 4.3. The first follows directly from the uniqueness of the solution of system (20) for given $\boldsymbol{\mu}$. The second follows from Lemma 4.1.

Lemma 4.5:

i) The vector of shadow prices λ^* corresponding to the point $(\bar{\boldsymbol{P}}, \boldsymbol{R}^*)$ is the unique fixed point of T.

ii) The mapping T is order preserving, i.e., $\boldsymbol{\lambda}^{(1)} \leq \boldsymbol{\lambda}^{(2)} \Rightarrow T(\boldsymbol{\lambda}^{(1)}) \leq T(\boldsymbol{\lambda}^{(2)})$.

The following lemma is also useful.

Lemma 4.6:

i) If $\boldsymbol{\lambda}(0) \geq T(\boldsymbol{\lambda}(0))$ and we define

$$\boldsymbol{\lambda}(n) \equiv T^n(\boldsymbol{\lambda}(0)), \qquad n = 0, 1, 2, \cdots$$

then $\boldsymbol{\lambda}(n)$ is a decreasing sequence.

ii) If $\boldsymbol{\lambda}(0) \leq T(\lambda(0))$ then $\boldsymbol{\lambda}(n)$ is an increasing sequence, and $\boldsymbol{\lambda}(n) \uparrow \lambda^*$.

iii) If $\boldsymbol{\lambda}(0) \geq T(\lambda(0))$ then $\boldsymbol{\lambda}(n) \downarrow \lambda^*$.

Proof:

i) The property follows from the order-preserving property of T.

ii) The order-preserving property of T implies that $(\boldsymbol{\lambda}(n))_{n=0}^{\infty}$ is an increasing sequence. However, By Lemma 4.4 ii), there exists a point $\boldsymbol{\lambda}$ for which $\boldsymbol{\lambda}(0) \leq \boldsymbol{\lambda}$ and $P(\boldsymbol{\lambda}) \leq \bar{P}$. By the order preserving property, $\boldsymbol{\lambda}(n) \leq T^n(\boldsymbol{\lambda})\ \forall n$, but since $P(\boldsymbol{\lambda}) \leq \bar{P}$, and part i) holds, it also follows that $T^n(\boldsymbol{\lambda})$ is a decreasing sequence. Hence $(\boldsymbol{\lambda}(n))_{n=1}^{\infty}$ is bounded, and must converge to the unique fixed point $\boldsymbol{\lambda}^*$ of T.

iii) Analogous to ii), but where we use Lemma 4.4 i) to guarantee a *lower* bound to the decreasing sequence $(\boldsymbol{\lambda}(n))_{n=1}^{\infty}$. □

Proof of Theorem 4.3: Lemma 4.4 guarantees the existence of points $\boldsymbol{w}(0)$ and $\boldsymbol{z}(0)$ with the following properties:

i) $\boldsymbol{w}(0) \leq \boldsymbol{\lambda}(0) \leq \boldsymbol{z}(0)$;

ii) $P(\boldsymbol{w}(0)) \geq \bar{P}$;

iii) $P(\boldsymbol{z}(0)) \leq \bar{P}$.

Now define $\boldsymbol{w}(n) = T^n(\boldsymbol{w}(0))$ and $\boldsymbol{z}(n) = T^n(\boldsymbol{z}(0))$. It follows from property ii) and Lemma 4.6 ii) that $\boldsymbol{w}(n) \uparrow \boldsymbol{\lambda}^*$. Similarly, it follows from property iii) and Lemma 4.6 iii) that $\boldsymbol{z}(n) \downarrow \boldsymbol{\lambda}^*$. Finally, it follows from property i) and the order-preserving property of T that $\boldsymbol{w}(n) \leq \boldsymbol{\lambda}(n) \leq \boldsymbol{z}(n)$. We conclude that $\boldsymbol{\lambda}(n) \rightarrow \boldsymbol{\lambda}^*$. □

Algorithm 4.2 has all the users updating $\boldsymbol{\lambda}(n)$ simultaneously. However, convergence still occurs if users update one at a time, or even asynchronously under certain weak conditions (Mitra [16]). An advantage of this is that then users do not need to know the fading statistics. If λ_i is being updated, for example, then binary search can be used to find the new value that achieves $\bar{P}_i$ for user i. This iterative algorithm, together with the greedy power-allocation algorithm described in the last section, yields the following dynamic resource allocation scheme for maximizing the total rate revenue subject to average power constraints: at each fading state, the greedy algorithm computes the optimal rate and power allocation using the current power prices; at a slower time scale, the power prices are adjusted to meet the average power constraints.

The iterative algorithm has the same monotonicity property as other power-control algorithms in the literature (Hanly [11], Yates [19]). In the references quoted, users directly control their access to the "available capacity" by updating their transmit powers. Monotonicity arises from the fact that if a user increases power, this decreases the rates of all other users, causing them to increase power. This occurs because interference from other users is treated as pure noise in these papers. In multiuser decoding, increasing power always benefits other users, so we do not get monotonicity in terms of transmit power alone. Instead, users control access to the "available capacity" through the power prices $\boldsymbol{\lambda}$. Nevertheless, monotonicity occurs in $\boldsymbol{\lambda}$-space, enabling very similar iterative procedures to be applied.

V. An Economic Framework for Resource Allocation

So far, we have formulated the problem of optimal resource allocation in terms of the computation of the capacity region, i.e., given average power constraints, what are the set of achievable rates? This is the standard information-theoretic formulation. However, another question of interest is: what are the average powers needed to support a given set of *target* rates, and the associated optimal resource-allocation schemes? It turns out that there is a complete analogous solution to that problem, and it essentially follows from the symmetry between rate and power.

First, given target rates $\boldsymbol{R}$ let us define the set $\mathcal{D}(\boldsymbol{R})$ and its boundary surface; it is the "power space equivalent" of the capacity region $\mathcal{C}(\boldsymbol{P})$, and contains the set of all average power vectors that can support $\boldsymbol{R}$.

Definition 5.1:

- $\mathcal{D}(\boldsymbol{R}) \equiv \{\bar{\boldsymbol{P}} : \boldsymbol{R} \in \mathcal{C}(\bar{\boldsymbol{P}})\}$.
- The boundary surface of $\mathcal{D}(\boldsymbol{R})$ is the set of those powers such that we cannot decrease one component, and remain in $\mathcal{D}(\boldsymbol{R})$ without increasing another.

Lemma 3.10 provides a Lagrangian characterization of the interior points of the boundary surface of $\mathcal{C}(\bar{\boldsymbol{P}})$. We take any $\boldsymbol{\mu} \in \Re_+^M$ and the lemma shows that this specifies a unique point on the boundary surface of $\mathcal{C}(\bar{\boldsymbol{P}})$. In addition, there is a unique $\boldsymbol{\lambda} \equiv \boldsymbol{\lambda}(\bar{\boldsymbol{P}}, \boldsymbol{\mu})$ associated with this point. We now extend this characterization to the "dual" set $\mathcal{D}(\boldsymbol{R}^*)$:

Lemma 5.2: An average power vector $\bar{\boldsymbol{P}}$ lies in the interior of the boundary surface of $\mathcal{D}(\boldsymbol{R}^*)$ if and only if there exists a positive $\boldsymbol{\lambda} \in \Re_+^M$ such that $\bar{\boldsymbol{P}}$ is a solution to the optimization problem

$$\min \boldsymbol{\lambda} \cdot \boldsymbol{P} \quad \text{subject to} \quad \boldsymbol{P} \in \mathcal{D}(\boldsymbol{R}^*). \tag{21}$$

For a given positive $\boldsymbol{\lambda}$, $\bar{\boldsymbol{P}}$ is a solution to the above problem if and only if there exists a nonnegative $\boldsymbol{\mu} \in \Re^M$, rate allocation policy $\mathcal{R}(\boldsymbol{h})$, and power control policy $\mathcal{P}(\boldsymbol{h})$ such that for every joint fading state $\boldsymbol{h}$, $(\mathcal{R}(\boldsymbol{h}), \mathcal{P}(\boldsymbol{h}))$ is a solution to the optimization problem

$$\max_{(\boldsymbol{r}, \boldsymbol{p})} \boldsymbol{\mu} \cdot \boldsymbol{r} - \boldsymbol{\lambda} \cdot \boldsymbol{p} \quad \text{subject to} \quad \boldsymbol{r} \in \mathcal{C}_g(\boldsymbol{h}, \boldsymbol{p}) \tag{22}$$

and

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}_i(\boldsymbol{H})] = R_i^*, \quad \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}_i(\boldsymbol{H})] = \bar{P}_i, \qquad i = 1, \cdots, M.$$

Moreover, for a given $\boldsymbol{\lambda}$ and $\boldsymbol{R}^*$, $\bar{\boldsymbol{P}}$ and $\boldsymbol{\mu}$ are unique.

Proof: The proof of this lemma is almost identical to that of Lemma 3.10, as both follows from the convexity of the set

$$S = \{(\boldsymbol{R}, \boldsymbol{P}) : \boldsymbol{R} \in \mathcal{C}(\boldsymbol{P})\} = \{(\boldsymbol{R}, \boldsymbol{P}) : \boldsymbol{P} \in \mathcal{D}(\boldsymbol{R})\}. \quad (23)$$

Uniqueness can be proved in a similar manner as in Lemma 3.15. □

Thus each point on the boundary of $\mathcal{D}(\boldsymbol{R}^*)$ is that obtained by minimizing a total cost $\boldsymbol{\lambda} \cdot \boldsymbol{P}$ while supporting the desired rates $\boldsymbol{R}^*$. The greedy algorithm defined in Theorem 3.14 can be used to compute the optimal power and rate allocation, for a given shadow reward $\boldsymbol{\mu}$. To compute $\boldsymbol{\mu}$ for a given $\boldsymbol{\lambda}$ and target rates $\boldsymbol{R}^*$, one can use the following iterative algorithm, entirely analogous to Algorithm 4.2.

Algorithm 5.3: Let $\boldsymbol{\mu}(0)$ be an initial arbitrary set of positive shadow rewards for rates. Given the nth iterate $\boldsymbol{\mu}(n)$, the $n+1$th iterate $\boldsymbol{\mu}(n+1)$ is given by the following: for each i, $\mu_i(n+1)$ is the unique rate reward for the ith user such that the rate of user i is R_i^* under the optimal power control policy when the rate rewards of the other users remain fixed at $\boldsymbol{\mu}(n)$.

Denote the rate by $\boldsymbol{R}(\boldsymbol{\mu}(n))$ and the average power by $\boldsymbol{P}(\boldsymbol{\mu}(n))$ under the optimal power control policy. The proof of the following theorem is entirely analogous to the proof of Theorem 4.3.

Theorem 5.4: Given desired rates $\boldsymbol{R}^*$, let $\bar{\boldsymbol{P}}$ be the optimum average power corresponding to the prices $\boldsymbol{\lambda}$, and let $\boldsymbol{\mu}^*$ be the appropriate shadow rewards. Then

$$\boldsymbol{\mu}(n) \rightarrow \boldsymbol{\mu}^*, \qquad n \uparrow \infty$$

and hence $\boldsymbol{R}(\boldsymbol{\mu}(n)) \rightarrow \boldsymbol{R}^*$, and $\boldsymbol{P}(\boldsymbol{\mu}(n)) \rightarrow \bar{\boldsymbol{P}}$.

We have seen that given rate rewards $\boldsymbol{\mu}$ and power constraints $\bar{\boldsymbol{P}}$, there exists a unique $\boldsymbol{R}^*$ which maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$ and unique Lagrangian power prices $\boldsymbol{\lambda}^*$. Similarly, given power prices $\boldsymbol{\lambda}$ and target rates $\boldsymbol{R}^*$, there exists unique $\bar{\boldsymbol{P}}$ which minimizes $\boldsymbol{\lambda} \cdot \boldsymbol{P}$ and unique Lagrangian rewards $\boldsymbol{\mu}^*$. In fact, one can also show that given $(\boldsymbol{R}^*, \bar{\boldsymbol{P}})$ on the boundary of S (defined in (23)), there exist unique $\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*$ such that

$$(\boldsymbol{R}^*, \bar{\boldsymbol{P}}) = \arg \max_{(\boldsymbol{R}, \boldsymbol{P}) \in S} \boldsymbol{\mu}^* \cdot \boldsymbol{R} - \boldsymbol{\lambda}^* \cdot \boldsymbol{P}$$

i.e., there is a unique supporting hyperplane at $(\boldsymbol{R}^*, \bar{\boldsymbol{P}})$ to S. This fact allows us to give two economic interpretations to $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$.

Let us interpret μ_i^* as the rate reward for user i. That is, user i earns $\mu_i^* R_i$ if it sends with rate R_i. The total revenue earned in the channel is then $\boldsymbol{\mu}^* \cdot \boldsymbol{R}$. Lemma 3.10 shows that any point $\boldsymbol{R}^*$ on the interior of the boundary surface of $\mathcal{C}(\bar{\boldsymbol{P}})$ can be obtained as a maximization of total revenue. The lemma shows that at the optimal solution $\boldsymbol{R}^*$, a set of shadow prices $\boldsymbol{\lambda}^*$ exist, in the sense that if we change the power constraint by $\Delta \boldsymbol{P}$, then we change the revenue earned by $\boldsymbol{\lambda}^* \cdot \Delta \boldsymbol{P}$. However, it is clear from Lemma 5.2 that we can interpret $\boldsymbol{\lambda}^*$ directly as a set of "power prices." To see this, consider problem (21), and interpret $\boldsymbol{\lambda} \cdot \boldsymbol{P}$ as the total price of the power vector $\boldsymbol{P}$. At any solution $\boldsymbol{P}$, there is an associated shadow reward $\boldsymbol{\mu}$ on the rates. Now if we set $\boldsymbol{\lambda} \equiv \boldsymbol{\lambda}^*$, then by the uniqueness of the supporting hyperplane to S at $(\boldsymbol{R}^*, \bar{\boldsymbol{P}})$ we must have that $\boldsymbol{\mu} = \boldsymbol{\mu}^*$. It follows that the shadow prices in the rate maximization problem (12) are the power prices in the "dual" problem, and the shadow rewards in (21) are the rate rewards in (12).

We, therefore, consider the following economic framework for resource allocation. We are given a vector $\boldsymbol{\mu}^*$ of rate rewards, and a vector $\boldsymbol{\lambda}^*$ of power prices, and our aim is to find the optimal operating point $(\boldsymbol{R}^*, \bar{\boldsymbol{P}})$ such that $\boldsymbol{\mu}^* \cdot \boldsymbol{R} - \boldsymbol{\lambda}^* \boldsymbol{P}$ is maximized. Section III-B provides a greedy algorithm which attains this optimal operating point.

VI. Auxiliary Constraints on Transmitted Power

The constraints on the transmitter powers we have considered so far are on their long-term *average* value, and under power control, the transmitter power will vary depending on the fading state. In practice, one often wants to have some shorter term constraints on the transmitter power as well. These constraints may be due to regulations, or as a way of imposing a limit on how much interference a mobile can cause to adjacent cells. To model such auxiliary constraints, we consider the following feasible set of power controls:

$$\mathcal{F}_p \equiv \{\mathcal{P} : \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}_i(\vec{H})] \leq \bar{P}_i \text{ and } \mathcal{P}_i(\boldsymbol{h}) \leq \hat{P}_i \ \forall i \text{ and } \boldsymbol{h} \in \mathcal{H}\}$$

where $\mathcal{H}$ is the set of all possible joint fading states of the users. Thus in addition to the average power constraints, we also have a constraint $\hat{P}_i$ on the transmitter power of the ith user in every state. We will assume that for every i, $\hat{P}_i > \bar{P}_i$. Otherwise, the average power constraint becomes innocuous. We shall now concentrate on the problem of computing the optimal power control subject to these constraints.

We focus on the capacity region

$$\mathcal{C}^p(\bar{P}, \hat{P}) \equiv \bigcup_{\mathcal{P} \in \mathcal{F}^p} \mathcal{C}_f(\mathcal{P})$$

where $C_f(\mathcal{P})$ can be interpreted as the set of achievable rates under power control $\mathcal{P}$.

In parallel to the case when there are only average power constraints, we will characterize this region in terms of the solution to a family of optimization problems over parallel Gaussian channels. The proof of the following lemma is analogous to that of Lemma 3.10.

Lemma 6.1: A rate vector $\boldsymbol{R}^*$ lies on the boundary of $\mathcal{C}^p(\bar{P}, \hat{P})$ if and only if there exist $\boldsymbol{\mu}, \boldsymbol{\lambda} \in \Re^M$, rate-allocation policy $\mathcal{R}(\boldsymbol{h})$ and power-control policy $\mathcal{P}(\boldsymbol{h})$ such that for every joint fading state $\boldsymbol{h}$, $(\mathcal{R}(\boldsymbol{h}), \mathcal{P}(\boldsymbol{h}))$ is a solution to the optimization problem

$$\max \boldsymbol{\mu} \cdot \boldsymbol{r} - \boldsymbol{\lambda} \cdot \boldsymbol{p} \text{ subject to } \boldsymbol{r} \in \mathcal{C}_g(\boldsymbol{h}, \boldsymbol{p}) \text{ and } p_i \leq \hat{P}_i \ \forall i \quad (24)$$

and

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{R}(\boldsymbol{H})] = \boldsymbol{R}^* \qquad \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}(\boldsymbol{H})] = \bar{\boldsymbol{P}}$$

where $\bar{P}_i$ is the constraint on the average power of user i. Moreover, $\mathcal{P}$ is a power-control policy which can achieve the rate vector $\boldsymbol{R}^*$.

Consider the more general optimization problem over polymatroids with generalized symmetric rank function g

$$\max \boldsymbol{\mu}\cdot\boldsymbol{x} - \boldsymbol{\lambda}\cdot\boldsymbol{y} \text{ subject to } \boldsymbol{x}(S) \le g(\boldsymbol{y}(S)) \quad \forall S \subset E, \quad 0 \le y_i \le a_i \quad \forall i \tag{25}$$

where a_i's are given constants. Although there are exponentially large number of constraints, we will exploit the polymatroid structure and given an efficient greedy optimization algorithm.

Without loss of generality, let us assume that $\mu_1 \ge \mu_2 \ge \cdots \ge \mu_M$. By Lemma 3.2, for any vector $\boldsymbol{y}$, the maximum value of $\boldsymbol{\mu}\cdot\boldsymbol{x}$ subject to the polymatroid constraints is given by

$$\sum_{i=1}^{M} \mu_i \left[g\left(\sum_{k=1}^{i} y_k\right) - g\left(\sum_{k=1}^{i-1} y_k\right)\right].$$

Hence the optimization problem (25) is equivalent to

$$\max_{\boldsymbol{y}} \sum_{i=1}^{M-1} (\mu_i - \mu_{i-1}) g\left(\sum_{k=1}^{i} y_k\right) + \mu_M g\left(\sum_{k=1}^{M} y_k\right) - \boldsymbol{\lambda}\cdot\boldsymbol{y}$$

$$\text{subject to } 0 \le y_i \le a_i.$$

We will now demonstrate that the optimal solution can be obtained by a simple combinatorial greedy algorithm with number of steps bounded by $2M$.

Let us define

$$I(\boldsymbol{y}) \equiv \sum_{k=1}^{M-1} (\mu_k - \mu_{k-1}) g\left(\sum_{m=1}^{k} y_m\right) + \mu_M g\left(\sum_{m=1}^{M} y_m\right) - \boldsymbol{\lambda}\cdot\boldsymbol{y}$$

and let

$$I_i(\boldsymbol{y}) \equiv \frac{\partial I}{\partial y_i}(\boldsymbol{y}) = \sum_{k=i}^{M-1} (\mu_k - \mu_{k-1}) g'\left(\sum_{m=1}^{k} y_m\right) + \mu_M g'\left(\sum_{m=1}^{M} y_m\right) - \lambda_i.$$

We first observe two facts.

Fact 1: $I_i(\boldsymbol{y})$ is monotonically decreasing in y_i.

Fact 2: For $j > i$,

$$I_i(\boldsymbol{y}) - I_j(\boldsymbol{y}) = \sum_{k=i}^{j-1} (\mu_k - \mu_{k+1}) g'\left(\sum_{m=1}^{k} y_m\right) + \lambda_j - \lambda_i$$

so that the difference is independent of y_j and decreases monotonically with y_i (by the concavity of g).

Consider now the following algorithm.

Algorithm 6.2:

- **Initialization**: Set $\boldsymbol{y}^{(0)} = 0$. Set $k = 0$.
- **Step** k: Pick an i_k such that $I_{i_k}(\boldsymbol{y}^{(k)}) > 0$, $y_{i_k}^{(k)} < a_{i_k}$, and $I_{i_k}(\boldsymbol{y}^{(k)}) \ge I_j(\boldsymbol{y}^{(k)})$ for all j such that $y_j^{(k)} < a_{i_k}$. If there is no such i_k, then stop. If there is more than one such i_k, pick the largest one. By Fact 2, we know that for each $j > i_k$ there either exists a unique solution $v_j > y_{i_k}^{(k)}$ to the equation

  $$I_j\left(y_1^{(k)}, \cdots, v_j, \cdots, y_M^{(k)}\right) = I_{i_k}\left(y_1^{(k)}, \cdots, v_j, \cdots, y_M^{(k)}\right)$$

  (where v_j is in the i_kth position) or there is no such solution, in which case we set $v_j = \infty$. Also, by Fact 1, let v_0 be the unique solution to the equation

  $$I_{i_k}\left(y_1^{(k)}, \cdots, v_0, \cdots, y_M^{(k)}\right) = 0$$

  if it exists, and let $v_0 = \infty$ otherwise. We now set:

  $$y_i^{(k+1)} = \begin{cases} \min\{a_i, v_0, \min_{j>i} v_j\}, & i = i_k \\ y_i^{(k)}, & i \ne i_k. \end{cases}$$

  Goto step $k+1$.

We note that at each step, we are always increasing the component which leads to the largest rate of positive increase of the objective function and which has not reached the peak constraint. Thus the algorithm is a greedy one.

Theorem 6.3: Algorithm 6.2 terminates at an optimal point for the problem (25), and the number of steps needed is at most $2M$.

Proof: See Appendix D. □

The optimal power-allocation problem with auxiliary constraints (24) can be expressed in terms of the received powers $\boldsymbol{q} = (h_1 p_1, \cdots, h_M p_M)$

$$\max_{(\boldsymbol{r},\boldsymbol{q})} \sum_i \mu_i r_i - \sum_i \frac{\lambda_i}{h_i} q_i \text{ subject to } \boldsymbol{R}(S) \le g(\boldsymbol{q}(S)) \quad \forall S \subset E \text{ and } q_i \le \frac{\hat{P}_i}{h_i} \; \forall i$$

where

$$g(z) \equiv \frac{1}{2}\log\left(1 + \frac{z}{\sigma^2}\right).$$

Thus Algorithm 6.2 can be used to solve this problem. It should also be noted that as in the case without the auxiliary power constraints, successive decoding can be used to achieve an optimizing rate vector.

VII. Frequency-Selective Fading Channels

In the previous sections we have analyzed a flat fading model which is appropriate if the Nyquist sampling period is large compared to the delay spread of the multipaths in the received signal, so that the individual paths are not resolvable in the sampled system. This is typically the case with narrowband transmission. For wideband applications, the multipaths can be resolved, and hence the channel has memory. The appropriate model is the time-varying frequency-selective fading channel. In this section, we will extend some of our previous results to this model.

We start with a continuous-time model. Suppose M users share a total bandwidth W centered around frequency f_0, corrupted by white additive Gaussian noise of spectral density $\frac{\eta_0}{2}$. The average transmitter power of user i is constrained to be less than or equal to $\bar{P}_i$. At time t, the kth path transmitted from the ith user is attenuated by $a_{ik}(t)$ and delayed by $\tau_{ik}(t)$ before being received at the base station. These quantities are time-varying due primarily to the motion of the transmitter but also to the motion of other objects in the system. The baseband representation of the channel is given by

$$y(t) = \sum_{i=1}^{M} \int x_i(t-\tau) h_i(\tau, t)\, d\tau + z(t)$$

where $x_i(\cdot)$ is the transmitted signal of user i, $z(\cdot)$ and $y(\cdot)$ are the complex baseband noise and received signal, respectively, i.e., the actual noise and received signal are $\mathrm{Re}\,[z(t)\exp^{j2\pi f_0 t}]$ and $\mathrm{Re}\,[y(t)\exp^{j2\pi f_0 t}]$, respectively. The time-varying impulse responses h_i's represent the fading effects

$$h_i(\tau, t) = \sum_k \alpha_{ik}(t)\delta(\tau - \tau_{ik}(t))$$

where $\alpha_{ik}(t) = a_{ik}(t)\exp^{j2\pi f_0 \tau_{ik}(t)}$. We assume that there is a bound T_0 on the largest delay of any path, so that $h_i(\tau, t) = 0$ for $\tau < 0$ and $\tau > T_0$. The parameter T_0 is the multipath delay spread.

The fading of the channel stems from both the time variation of the attenuation $a_{ik}(t)$, due to path loss and shadowing effects (slow fading), as well as the constructive and destructive interference between the various paths (fast fading). The latter typically occurs at a much faster time scale than the former.

We now sample the system at a Nyquist rate T and get

$$Y(n) = \sum_{i=1}^{M} \sum_k H_i(k, n) X(n-k) + Z(n)$$

where

$$Y(n) = y\left(\frac{n}{T}\right), \quad X(n) = x\left(\frac{n}{T}\right)$$

$$H_i(k, n) = \int \frac{\sin\left(\pi\left(\frac{k}{T} - \tau\right)\right)}{\pi\left(\frac{k}{T} - \tau\right)} h_i\left(\tau, \frac{n}{T}\right) d\tau.$$

Note that the Nyquist rate T is in general larger than W because the the received signal is spread out due to the time-varying channel.

To begin analyzing the capacity region of this channel, when both the transmitters and the receiver can track the channel, let us first focus on the special case when the channel is time-invariant. In this case, the channel is given by

$$Y(n) = \sum_{i=1}^{M} \sum_k H_i(k) X(n-k) + Z(n).$$

This is the Gaussian multiaccess channel with intersymbol interference (ISI), and a characterization of the capacity region has been obtained by Cheng and Verdú [2]. Let $\hat{H}_i(f)$ be the Fourier transform of the channel. Let $\mathcal{P}$ be a power-allocation policy such that for user i and frequency f, $\mathcal{P}_i(f)$ can be interpreted as the transmitter power that user i allocates at frequency f. Let

$$\mathcal{F} = \left\{ \mathcal{P} : \int_{\frac{-W}{2}}^{\frac{W}{2}} \mathcal{P}_i(f)\, df \le \bar{P}_i \ \forall i \right\}$$

be the set of all feasible power allocation policy. Then the capacity region of the channel is

$$\bigcup_{\mathcal{P} \in \mathcal{F}} \left\{ \boldsymbol{R} : \boldsymbol{R}(S) \le \int_{\frac{-W}{2}}^{\frac{W}{2}} \log\left(1 + \frac{\sum_{i \in S} \mathcal{P}_i(f) |\hat{H}_i(f)|^2}{\sigma^2}\right) \ \forall S \subset \{1, \cdots, M\} \right\} \tag{26}$$

where $\sigma^2 = \eta_0 W$.

In [2], an explicit characterization of the region and the optimal power allocations are obtained for the two-user case. We shall now give the solution in the general multiuser case, which follows almost directly from the results in Section III. The key observation is that the structure of this capacity region is in fact identical to that of the capacity region of the flat fading channel (Theorem 2.1), with the role of the fading state $\boldsymbol{h}$ now played by frequency f. Using the results of Section III, each point on the boundary of the capacity region can be computed via an optimization problem over a set of parallel channels, one for each frequency. In complete analogy to Theorem 3.16, we have the following result.

Theorem 7.1: Assume that for user i and any constant a, the level set $\{f : |\hat{H}_i(f)| = a\}$ has Lebesgue measure 0. Then the boundary of the capacity region of the Gaussian multiaccess channel with ISI is

$$\left\{ \boldsymbol{R}^*(\boldsymbol{\mu}) : \boldsymbol{\mu} \in \Re_+^M, \sum_i \mu_i = 1 \right\}$$

where for $i = 1, \cdots, M$

$$R_i^*(\boldsymbol{\mu}) = \int_0^\infty \frac{1}{(\sigma^2 + z)} m(A_i(z, \boldsymbol{\lambda}))\, dz \tag{27}$$

where

$$A_i(z, \boldsymbol{\lambda}) = \left\{ f \in \left[\frac{-W}{2}, \frac{W}{2}\right] : \frac{\mu_i}{\sigma^2 + z} - \frac{\lambda_i}{|\hat{H}_i(f)|^2} \ge \left[\max_{j \ne i}\left(\frac{\mu_j}{\sigma^2 + z} - \frac{\lambda_j}{|\hat{H}_j(f)|^2}\right)\right]^+ \right\}$$

and $m(\cdot)$ is the Lebesgue measure of a set. The vector $\boldsymbol{\lambda}$ satisfies the equations

$$\int_0^\infty \int_{A_i(z, \boldsymbol{\lambda})} \frac{1}{|\hat{H}_i(f)|^2}\, df\, dz = \bar{P}_i \tag{28}$$

$i = 1, \cdots, M$.

The rate vector on the boundary corresponding to a specific μ can be achieved by successive decoding, with the users decoded in increasing order of μ_i's. The corresponding power allocations to achieve that point are given by

$$\mathcal{P}_i(f) = \frac{1}{|\hat{H}_i(f)|^2} \cdot m(\{z \in [0,\infty) : f \in A_i(z,\boldsymbol{\lambda})\}), \quad f \in \left[\frac{-W}{2}, \frac{W}{2}\right], \quad i = 1, \cdots, M.$$

The interpretation of this power allocation is similar to that in the flat fading case. The variable z represents the received interference caused by users' signals, beyond the background Gaussian noise. At frequency f and received interference level σ^2+z, user i transmits if it yields the maximum increase in the objective function $\boldsymbol{\mu}\cdot\boldsymbol{r}-\boldsymbol{\lambda}\cdot\boldsymbol{p}$, which is the case if $f \in A_i(z,\boldsymbol{\lambda})$.

Next we analyze the general situation when the channel is time-varying. Even for the case when only the receiver can track the channel, there is in general no clean characterization of the capacity region of time-varying frequency-selective fading channels [15]. However, if we make the assumptions that the channel varies very slowly relative to the multipath delay spread and that the time variations are random and ergodic, then the capacity region for that case is given by [7]

$$\left\{\boldsymbol{R} : \boldsymbol{R}(S) \le \mathbb{E}\left[\int_{\frac{-W}{2}}^{\frac{W}{2}} \log\left(1 + \frac{\sum_{i\in S} \bar{P}_i |\hat{H}_i(f,\omega)|^2}{\sigma^2}\right) df\right] \quad \forall S \subset \{1, \cdots, M\}\right\}$$

where $\bar{P}_i$ is the average power constraint of user i. For each realization (time-slot) ω, $\hat{H}_i(\cdot,\omega)$ is the frequency response of user i's channel at fading state ω. The intuition behind this result is that if the time variation is slow relative to the delay spread, the overall channel can be thought of as a set of parallel time-invariant channels. The expectation is taken over all (joint) fading states.

How valid is this assumption in practice? We use here a numerical example in [7]. Consider a typical mobile scenario where the vehicle is moving at 60 km/h and the carrier frequency is 1 GHz. The time constant associated with the fast fading effects due to constructive and destructive interference between paths is of the order of the time taken for the mobile to travel one wavelength at the transmitted frequency. In this example, it is 0.018 s. Typical delay spread between paths range from 10^{-7} to 1.5×10^{-5} seconds [18]. Hence, the time variation due to fast fading is significantly slower than the delay spread. This is even more so when the users are moving at a slower speed. Thus we see that the assumption is quite reasonable for typical wireless situations.

In analogy to Theorem 2.1, it can be shown that the capacity region for this channel when all the transmitters and the receivers can track the channel is given by

$$\bigcup_{\mathcal{P}\in\mathcal{F}} \left\{\boldsymbol{R} : \boldsymbol{R}(S) \le \mathbb{E}\left[\int_{\frac{-W}{2}}^{\frac{W}{2}} \log\left(1 + \frac{\sum_{i\in S} \mathcal{P}_i(f, \hat{\boldsymbol{H}}(f,\omega)) |\hat{H}_i(f,\omega)|^2}{\sigma^2}\right) df\right] \quad \forall S \subset \{1, \cdots, M\}\right\}$$

where

$$\mathcal{F} \equiv \left\{\mathcal{P} : \mathbb{E}\left[\int_{\frac{-W}{2}}^{\frac{W}{2}} \mathcal{P}_i(f, \hat{\boldsymbol{H}}(f,\omega))\, df\right] \le \bar{P}_i \ \forall i\right\}$$

and $\hat{\boldsymbol{H}}(f,\omega) = (\hat{H}_i(f,\omega), \cdots, \hat{H}_M(f,\omega))$

Using the techniques of Section III, each point on the boundary of this capacity region can again be computed via an optimization problem over a set of parallel channels, this time one for each frequency f and fading state ω. This leads to the following generalization of Theorem 3.16 to the frequency selective fading case.

Theorem 7.2: For each frequency f and transmitter i, let the random variable $\hat{H}_i(f,\cdot)$ have continuous cdf $F_i(f,\cdot)$ and density $f_i(f,\cdot)$. Also assume that the fading processes of users are independent of each other. The boundary of the region is the parametrically defined surface

$$\left\{\boldsymbol{R}^*(\boldsymbol{\mu}) : \boldsymbol{\mu} \in \Re_+^M, \sum_i \mu_i = 1\right\}$$

where for $i = 1, \cdots, M$

$$R_i^*(\boldsymbol{\mu}) = \int_0^\infty \frac{1}{2(\sigma^2+z)} \left\{\int_{\frac{-W}{2}}^{\frac{W}{2}} \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \prod_{k\ne i} F_k\left(f, \frac{1}{\frac{\lambda_i}{\lambda_k h} + \frac{\mu_k-\mu_i}{2\lambda_k(\sigma^2+z)}}\right) f_i(f,h)\, dh\, df\right\} dz \tag{29}$$

where the vector $\boldsymbol{\lambda}$ satisfies the equations

$$\int_0^\infty \left\{\int_{\frac{-W}{2}}^{\frac{W}{2}} \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \frac{1}{h} \times \prod_{k\ne i} F_k\left(f, \frac{1}{\frac{\lambda_i}{\lambda_k h} + \frac{\mu_k-\mu_i}{2\lambda_k(\sigma^2+z)}}\right) f_i(f,h)\, dh\, df\right\} dz = \bar{P}_i, \quad i = 1, \cdots, M. \tag{30}$$

VIII. Conclusion

In this paper, we have characterized the throughput capacity region of the multiaccess fading channel with perfect channel state information at the receiver and the transmitters. Just as the solution to the corresponding single-user channel has the water-filling interpretation, our solution can be viewed as the multiuser analog of water filling with an arbitrary number of users. The new mathematical ingredient is provided by the polymatroid structure of the problem, yielding a greedy optimal power allocation. The solution contains various steps, which we summarize in the following.

- For each fading state $\boldsymbol{h} = (h_1, \cdots, h_M)$, define

$$\mathcal{C}_g(\boldsymbol{h}, \boldsymbol{P}) \equiv \left\{\boldsymbol{R} : \sum_{i\in S} R_i \le \frac{1}{2}\log\left(1 + \frac{\sum_{i\in S} h_i P_i}{\sigma^2}\right) \text{ for every } S \subset \{1, \cdots, M\}\right\}$$

where $\boldsymbol{P} = (P_1, \cdots, P_M)$. This can be interpreted as the capacity region of an M-user multiaccess Gaussian channel with noise variance σ^2, transmit powers $\boldsymbol{P}$, and fading state of the users fixed at $\boldsymbol{h}$.

- The fading state $\boldsymbol{h}$ varies according to a stationary and ergodic process. Let $\boldsymbol{H}$ be the random vector with the stationary distribution of the fading state. The capacity region of the fading channel when both the transmitters and the receiver have perfect channel state information is shown to be (Theorem 2.1)
$$\mathcal{C}(\bar{\boldsymbol{P}}) \equiv \bigcup_{\mathcal{P}\in\mathcal{F}} \mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_g(\boldsymbol{H}, \mathcal{P}(\boldsymbol{H}))]$$
where $\bar{P} = (\bar{P}_1, \cdots, \bar{P}_M)$ is the average power constraints of the users and $\mathcal{F}$ is the set of all feasible power allocation policies satisfying the average power constraint
$$\mathcal{F} \equiv \{\mathcal{P} : \mathbb{E}_{\boldsymbol{H}}[\mathcal{P}_i(\boldsymbol{H})] \le \bar{P}_i \ \forall i\}.$$
The region $\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_g(\boldsymbol{H}, \mathcal{P}(\boldsymbol{H}))]$ is the set of all rate vectors which are the weighted average of rate vectors in each of the sets $\mathcal{C}_g(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}))$, weighted according the fading distribution.
- The region $\mathcal{C}(\bar{P})$ is convex. Every point on the boundary of $\mathcal{C}(\bar{P})$ is a solution to an optimization problem
$$\max_{\boldsymbol{R}} \boldsymbol{\mu}\cdot\boldsymbol{R} \text{ subject to } \boldsymbol{R} \in \mathcal{C}(\bar{\boldsymbol{P}}) \tag{31}$$
for some nonnegative $\boldsymbol{\mu}$ (Lemma 3.10).
- The optimal solution to (31) can be solved by decomposition into a set of optimization problem over parallel multiaccess channels, one for each fading state $\boldsymbol{h}$ (Lemma 3.10)
$$\max_{(\boldsymbol{r},\boldsymbol{p})} \boldsymbol{\mu}\cdot\boldsymbol{r} - \boldsymbol{\lambda}\cdot\boldsymbol{p} \text{ subject to } \boldsymbol{r} \in \mathcal{C}_g(\boldsymbol{h}, \boldsymbol{p}) \tag{32}$$
where $\boldsymbol{\lambda}$ is the Lagrangian multiplier for the average power constraints. The optimal solution to (32) gives a power allocation $\mathcal{P}(h)$ and a rate allocation $\mathcal{R}(h)$ in fading state $\boldsymbol{h}$. If $\boldsymbol{\lambda}$ is chosen such that the average power constraints are satisfied (i.e., $\mathbb{E}_{\boldsymbol{H}}[\mathcal{P}(\boldsymbol{H})] = \bar{\boldsymbol{P}}$), then $\boldsymbol{R}^* \equiv \mathbb{E}_{\boldsymbol{H}}[\mathcal{R}(\boldsymbol{H})]$ is an optimal solution to (32).
- The optimization problem (32) has a simple greedy solution, although there are exponentially large number of constraints (in M). (Theorem 3.14). Define
$$g(z) \equiv \frac{1}{2}\log\left(1 + \frac{z}{\sigma^2}\right)$$
$$u_i(z) \equiv \mu_i g'(z) - \frac{\lambda_i}{h_i} = \frac{\mu_i}{2(\sigma^2 + z)} - \frac{\lambda_i}{h_i}$$
$$u^*(z) \equiv \left[\max_i u_i(z)\right]^+.$$
The optimal value for problem (32) is given by
$$\int_0^\infty u^*(z)\, dz.$$
The optimal solution is achieved by *successive decoding* and can be interpreted as follows. Think of $\sigma^2 + z$ as the current "interference level" due to background noise and received powers of users not yet canceled. Start with $z = 0$, and at each z allocate a marginal *received* power δQ to the user i^* with the largest positive $u_i(z)$. Stop when $u_j(z) < 0$ for all j. The marginal increase in rate of user i^* is $g'_{i^*}(z)\cdot\delta q$, decoding at interference level $\sigma^2 + z$. The value $u^*(z)\cdot\delta q$ is therefore the marginal increase in the value of the overall objective function $\boldsymbol{\mu}\cdot\boldsymbol{r} - \boldsymbol{\lambda}\cdot\boldsymbol{p}$ by allocating power δq to the user that will benefit most at the interference level $\sigma^2 + z$. The procedure is thus *greedy*. Integrating over all z gives the optimal rate and power allocation, as well as the successive decoding order to achieve the optimal solution. See Fig. 2 for an example.
- The optimal average powers and rates can be computed explicitly as a function of $\boldsymbol{\lambda}$. This gives a parameterization of the boundary of the capacity region $\mathcal{C}(\bar{\boldsymbol{P}})$ in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ (Theorem 3.16).
- For a given $\boldsymbol{\mu}$, the value of $\boldsymbol{\lambda}$ to meet the average power constraints can be computed by a simple iterative algorithm, which is provably convergent (Theorem 4.3).

Taken together, these results provide simple solutions for computing the throughput capacity region as well as a characterization of the structure of the optimal resource-allocation schemes to achieve the points on the boundary of the region.

The problem formulation considered in this paper suffers from a drawback that delay is not considered; the Shannon capacities are essentially long-term throughput in a time-varying system, and the delay incurred depends on the rate of variations of the fading processes. In the sequel to this paper, we will define a notion of *delay-limited* capacity for the fading channel; these are the rates achievable with delay *independent* of how slow the fading processes are. We will see that polymatroid structure will again help us in characterizing the delay-limited capacity region of the fading channel.

Appendix A
Proof of Theorem 2.1

The proof of this theorem is straightforward other than the technicalities due to the continuous fading distributions.

For any power control policy $\mathcal{P}$, we can reinterpret the channel as a unit transmit power channel with fading $h_i\mathcal{P}_i(\boldsymbol{h})$ for user i. It follows from (3) that all rate vectors in $\mathcal{C}(\mathcal{P})$ are achievable.

Conversely, suppose rate $\boldsymbol{R}$ is achievable. By this we mean that there exists a sequence of codes, indexed by N, with code $\mathcal{C}_N$ of blocklength N, and with probability of error $\epsilon_N \to 0$. For code $\mathcal{C}_N$, we index the messages of user i by $\{1, 2, \cdots, 2^{R_i N}\}$ and user i uses the uniform distribution to select one of these messages, and transmits the corresponding codeword. We denote the resulting random vector by $\boldsymbol{X}_i$ for $i = 1, 2, \cdots, M$. Note that the codewords can be chosen as a function of the states of the channel.

Let $f(\boldsymbol{h})$ be the equilibrium probability density of being in fading state $\boldsymbol{h}$. Without loss of generality, assume that the fading of all users is bounded by 1. For each k, let $I_k = \{0, \frac{1}{k}, \frac{2}{k}, \cdots, 1\}^M$ be a partition of the fading state space $[0,1]^M$. For each cubic element E of partition I_k, let $S(E)$ be that random subset of $[1, \cdots, N]$ at which times the fading state $\boldsymbol{H}$ lies in E. Let $Q(N)$ be uniformly distributed on $[1, \cdots, N]$. Define
$$V_i^k(E, N) \equiv \mathbb{E}\left[X_i^2(Q(N)) | Q(N) \in S(E)\right].$$

Let $f(E)$ be the probability that a random $\boldsymbol{H}$ lies in E. For any message from user i, there is a power constraint on the corresponding codeword. It follows that for each N

$$\sum_{E\in I_k} V_i^k(E,N)f(E) \leq \bar{P}_i.$$

For all cubic elements E such that $f(E)\neq 0$, $V_i^k(E,N)$ are bounded sequences in N. Thus we must have the existence of limiting $V_i^k(E)$ such that there is convergence along a subsequence as $N\to\infty$. Further

$$\sum_{E\in I_k} V_i^k(E)f(E) \leq \bar{P}_i. \tag{33}$$

We define $\boldsymbol{h}(E)$ to be the upper corner of E. Let $\boldsymbol{H}(n)$ be the fading at time n and define a new value $\hat{\boldsymbol{H}}(n)$ by $\hat{\boldsymbol{H}}(n)\equiv \boldsymbol{h}(E)$ if $\boldsymbol{H}(n)\in E$. Define

$$\hat{Y}(n) \equiv \sum_{i=1}^{M} \hat{H}_i(n)X_i(n) + Z(n).$$

By Fano's inequality, we have for any $S\subseteq\{1,2,\cdots,M\}$

$$\boldsymbol{R}(S) \leq \frac{1}{N} I[(\boldsymbol{X}_i)_{i\in S};\boldsymbol{Y}|(\boldsymbol{X}_i)_{i\in S^c},\boldsymbol{H}] + \epsilon_N$$

where $\epsilon_N\to 0$ as $N\to\infty$. But

$$\begin{aligned}
&\frac{1}{N} I[(X_i)_{i\in S};Y|(X_i)_{i\in S^c},\boldsymbol{H}]\\
&= I[(X_i(Q(N)))_{i\in S};Y(Q(N))|(X_i(Q(N)))_{i\in S^c},\\
&\qquad \boldsymbol{H}(Q(N)),Q(N)]\\
&= \sum_{E\in I_k} f(E) I[(X_i(Q(N)))_{i\in S};Y(Q(N))|(X_i(Q(N)))_{i\in S^c},\\
&\qquad \boldsymbol{H}(Q(N)),Q(N),Q(N)\in S(E)]\\
&\leq \sum_{E\in I_k} f(E) I[(X_i(Q(N)))_{i\in S};\hat{Y}(Q(N))|(X_i(Q(N)))_{i\in S^c},\\
&\qquad \hat{H}(Q(N)),Q(N),Q(N)\in S(E)]\\
&\leq \sum_{E\in I_k} f(E)\frac{1}{2}\log\left(1+\frac{\sum_{i\in S} h_i(E)V_i^k(E,N)}{\sigma^2}\right)
\end{aligned}$$

Taking limits along the convergent subsequence, we obtain

$$\boldsymbol{R}(S) \leq \sum_{E\in I_k} f(E)\frac{1}{2}\log\left(1+\frac{\sum_{i\in S} h_i(E)V_i^k(E)}{\sigma^2}\right). \tag{34}$$

Let $\mathcal{F}_k$ be the set of all power controls which are piecewise-constant on the cubic elements of I_k and satisfy the average power constraint. Define

$$\begin{aligned}
\bar{C}_f^{(k)}(\mathcal{P}) \equiv \Bigg\{&\boldsymbol{R} : \boldsymbol{R}(S)\\
&\leq \int_{[0,1]^M} \frac{1}{2}\log\left(1+\frac{1}{\sigma^2}\sum_{i\in S}\frac{1}{k}\lceil kh_i\rceil \mathcal{P}_i(\boldsymbol{h})\right) f(\boldsymbol{h})\,d\boldsymbol{h}\ \forall S\Bigg\}.
\end{aligned}$$

Hence, the above derivation implies that the capacity region $C(\bar{P})$ is bounded by

$$C(\bar{P}) \subset \bigcup_{\mathcal{P}\in\mathcal{F}_k} \bar{C}_f^{(k)}(\mathcal{P}).$$

Combining this with the achievability result, we have for every k the following inner and outer bounds:

$$\bigcup_{\mathcal{P}\in\mathcal{F}_k} C_f(\mathcal{P}) \subset \bigcup_{\mathcal{P}\in\mathcal{F}} C_f(\mathcal{P}) \subset C(\bar{P}) \subset \bigcup_{\mathcal{P}\in\mathcal{F}_k} \bar{C}_f^{(k)}(\mathcal{P}).$$

As $k\to\infty$

$$\bigcup_{\mathcal{P}\in\mathcal{F}_k} C_f(\mathcal{P}) \to \bigcup_{\mathcal{P}\in\mathcal{F}_k} \bar{C}_f^{(k)}(\mathcal{P}).$$

Hence

$$C(\bar{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} C_f(\mathcal{P})$$

and the proof is complete.

Appendix B
Proof of Lemma 3.15

We first claim that there is an almost surely unique rate and power allocation which maximizes $\mu\cdot R$ subject to the average power constraints. (Almost surely with respect to the fading distribution.) Suppose not, and let $(\mathcal{R}^{(j)},\mathcal{P}^{(j)})$, $j=1,2$, be two such rate and power allocations. Define $(\mathcal{R},\mathcal{P})$ by

$$\begin{aligned}
\mathcal{R} &\equiv \frac{1}{2}\left(\mathcal{R}^{(1)}+\mathcal{R}^{(2)}\right)\\
\mathcal{P} &\equiv \frac{1}{2}\left(\mathcal{P}^{(1)}+\mathcal{P}^{(2)}\right).
\end{aligned}$$

Note that this also achieves a point on the boundary of the capacity region. By the concavity of $\log, (\mathcal{R},\mathcal{P})$ is feasible

$$\forall S,\forall\boldsymbol{h}, \sum_{i\in S}\mathcal{R}_i(\boldsymbol{h}) \leq \frac{1}{2}\log\left(1+\frac{\sum_{i\in S}\mathcal{P}_i(\boldsymbol{h})h_i}{\sigma^2}\right). \tag{35}$$

For any $\boldsymbol{h}$, consider all subsets S for which there is equality in (35). If there is a user i that is not in any such subset, then $R_i(\boldsymbol{h})$ can be increased without violating any constraint. But this contradicts the fact that this rate allocation achieves a boundary point of the capacity region $\mathcal{C}(\bar{\boldsymbol{P}})$. Therefore, every user must be almost surely in a tight constraint, and hence, by the strict concavity of $\log, \mathcal{P}^{(1)}(\boldsymbol{H})=\mathcal{P}^{(2)}(\boldsymbol{H})$ almost surely.

Now we consider the issue of uniqueness of rate allocation policy. By Lemma 3.10, any rate-allocation policy $\mathcal{R}(\boldsymbol{h})$ and power-allocation policy $\mathcal{P}(\boldsymbol{h})$ which maximizes $\boldsymbol{\mu}\cdot\boldsymbol{R}$ must solve the optimization problem

$$\max_{(\boldsymbol{r},\boldsymbol{p})} \boldsymbol{\mu}\cdot\boldsymbol{r} - \frac{\lambda_i}{h_i}h_i p_i \tag{36}$$

for every fading state $\boldsymbol{h}$, for some $\boldsymbol{\lambda}$. The only possibility for nonuniqueness of $\mathcal{R}(\boldsymbol{h})$ occurs if $\mu_i=\mu_j$ for some i,j, for then we can reverse the decoding order of i and j without affecting the objective function. However, $\frac{\lambda_i}{h_i}>\frac{\lambda_j}{h_j}$ or *vice versa*, with probability 1, so with probability 1, $\mathcal{P}_i(\boldsymbol{h})=0$ or

$\mathcal{P}_j(\boldsymbol{h}) = 0$. Together with the fact that the power allocation is unique, we can conclude that there is also a unique rate allocation.

Now we show that the Lagrangian power prices $\boldsymbol{\lambda}$ for maximizing $\boldsymbol{\mu} \cdot \boldsymbol{R}$ subject to average power constraints must also be unique. Without loss of generality, assume that $\mu_1 \le \mu_2 \le \cdots \le \mu_M$. Let $\mathcal{P}$ be the unique optimal power-allocation policy; it can be obtained by maximizing $\boldsymbol{\mu} \cdot \boldsymbol{r} - \boldsymbol{\lambda} \cdot \boldsymbol{p}$ subject to $\boldsymbol{r} \in \mathcal{C}_g(\boldsymbol{h}, \boldsymbol{p})$ for each fading state, for some choice of $\boldsymbol{\lambda}$. We want to show that such a $\boldsymbol{\lambda}$ must also be unique. We show by induction on k that λ_k must be uniquely specified. Let $\boldsymbol{h}$ be a fading state for which $\mathcal{P}_1(\boldsymbol{h}) > 0$; in this fading state, user 1 must be decoded first (which means it is last in the priority ordering). Then from the greedy power-allocation algorithm, we see that in this fading state, the total received power must be that value of z such that $u_1(z) = 0$, i.e.,

$$\lambda_1 = \frac{\mu_1 h_1}{2\left(\sigma^2 + \sum_{i=1}^{M} \mathcal{P}_i(\boldsymbol{h})\right)}.$$

Thus λ_1 is uniquely specified. Now assume that $\lambda_1, \cdots, \lambda_k$ are uniquely specified. Let $\boldsymbol{h}$ be a fading state where $\mathcal{P}_{k+1}(\boldsymbol{h}) > 0$. In this fading state, the total received power from users $k+1, k+2, \cdots, M$ must be the value of z such that

$$u_{k+1}(z) = \max_{i \le k} u_i(z)$$

since only users $1, \cdots, k$ can be decoded before user $k+1$. Hence λ_{k+1} must satisfy

$$\frac{\mu_{k+1}}{2\left(\sigma^2 + \sum_{j \ge k+1} \mathcal{P}_j(\boldsymbol{h})\right)} - \frac{\lambda_{k+1}}{h_{k+1}} = \max_{i \le k} \frac{\mu_i}{2\left(\sigma^2 + \sum_{j \ge k+1} \mathcal{P}_j(\boldsymbol{h})\right)} - \frac{\lambda_i}{h_i}.$$

By the induction hypothesis, $\lambda_1, \cdots, \lambda_k$ are uniquely specified and hence so is λ_{k+1}. This completes the proof of uniqueness of the power price vector $\boldsymbol{\lambda}$.

Appendix C
Proof of Lemma 4.4

i) Without loss of generality, we assume

$$\mu_1 \ge \mu_2 \ge \cdots \ge \mu_M$$

so that the decoding order is $M, M-1, \cdots, 1$. Let $\epsilon > 0$ be arbitrary. We define a sequence of power prices $\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}, \cdots \boldsymbol{\lambda}^{(M)}$ and from these construct another vector of prices $\boldsymbol{\lambda}$. We shall show that $\boldsymbol{\lambda}$ satisfies the conditions of the lemma. For $m < M$, we take $\boldsymbol{\lambda}^{(m)}$ to be power prices of a fictitious channel in which only users $1, 2, \cdots, m$ are present. Further, we extend the definition of the channel to allow the price of the power of the user decoded last to be zero, and the power allocated to that user to be infinite. With $\boldsymbol{\lambda}^{(1)} = 0$, we consider a single user channel with $\mu = \mu_1$ and $\lambda = 0$; user 1 occupies the channel alone. With this reward and price, it is clear that $P_1(\boldsymbol{\lambda}^{(1)})$ is infinite. With $\boldsymbol{\lambda}^{(2)} = (\epsilon_1, 0)$, we have a two user channel, with $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\lambda} = (\epsilon_1, 0)$. It is clear that here $P_1(\boldsymbol{\lambda}^{(2)}) < \infty$ and $P_2(\boldsymbol{\lambda}^{(2)}) = \infty$. Note that by taking ϵ_1 small we can ensure that $P_1(\boldsymbol{\lambda}^{(2)}) > \bar{P}_1$ and $\epsilon_1 < \epsilon$. This becomes the inductive hypothesis: suppose that with $\boldsymbol{\lambda}^{(m)} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_{m-1}, 0)$ we have $P_i(\boldsymbol{\lambda}^{(m)}) > \bar{P}_i, \epsilon_i < \epsilon$ for all $i = 1, 2, \cdots, m-1$. Then set $\boldsymbol{\lambda}^{(m+1)} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_m, 0)$, and note that for any ϵ_m this gives a new channel with $m+1$ users. Provided $\epsilon_m > 0$, we must have that

$$P_i(\boldsymbol{\lambda}^{(m+1)}) \ge P_i(\boldsymbol{\lambda}^{(m)}) > \bar{P}_i, \qquad i = 1, 2, \cdots, m-1.$$

By choosing ϵ_m small we can ensure that $P_m(\boldsymbol{\lambda}^{(m+1)}) > \bar{P}_m$ and $\epsilon_m < \epsilon$. Note that $P_{m+1}(\boldsymbol{\lambda}^{(m+1)}) = \infty$. By induction, we terminate with $\boldsymbol{\lambda}^{(M)} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1}, 0)$ for which $P_i(\boldsymbol{\lambda}^{(M)}) > \bar{P}_i$ and $\epsilon_i < \epsilon$, for all $i = 1, 2, \cdots, M-1$, and $P_M(\boldsymbol{\lambda}^{(M)}) = \infty$. Again, by choosing ϵ_M small, and $\boldsymbol{\lambda} \equiv (\epsilon_1, \epsilon_2, \cdots, \epsilon_M)$, we can ensure that both $\bar{P}_M < P_M(\boldsymbol{\lambda}) < \infty$ and $\bar{P}_{M-1} < P_{M-1}(\boldsymbol{\lambda}) < \infty$. This establishes part i) of the lemma.

ii) One can construct such a $\boldsymbol{\lambda}$ in a manner analogous to that in part i).

Appendix D
Proof of Theorem 6.3

We first show by induction on k the following claim.

If i_k is the component to be increased at step k, then for all $i \ne i_k$: 1) if $y_i^{(k)} = a_i$, then $I_i(\boldsymbol{y}^{(k)}) \ge I_{i_k}(\boldsymbol{y}^{(k)})$; 2) if $0 < y_i^{(k)} < a_i$, then $I_i(\boldsymbol{y}^{(k)}) = I_{i_k}(\boldsymbol{y}^{(k)})$ and $i < i_k$; 3) if $y_i^{(k)} = 0$, then $I_i(\boldsymbol{y}^{(k)}) \le I_{i_k}(\boldsymbol{y}^{(k)})$.

For $k = 0$, only case 3) can occur so that the claim is true by definition of i_0. Assume the claim is true at step $k = m$. The i_mth component is updated to $y_{i_m}^{(m+1)}$, and all the other components remain unchanged. For $i \ne i_m$, 1) if $y_i^{(m+1)} = a_i$, then by the inductive hypothesis, $I_i(\boldsymbol{y}^m) \ge I_{i_m}(\boldsymbol{y}^m)$ and by Fact 1, $I_{i_m}(\boldsymbol{y}^m) \ge I_{i_m}(\boldsymbol{y}^{(m+1)})$, so that we have $I_i(\boldsymbol{y}^{(m+1)}) \ge I_{i_m}(\boldsymbol{y}^{(m+1)})$; 2) if $0 < y_i^{(m+1)} < a_i$, then by the inductive hypothesis, $I_i(\boldsymbol{y}^m) = I_{i_m}(\boldsymbol{y}^m)$ and $i < i_m$, so that together with Fact 2, this implies that $I_i(\boldsymbol{y}^{(m+1)}) = I_{i_m}(\boldsymbol{y}^{(m+1)})$; 3) if $y_i^{(m+1)} = 0$, then by the inductive hypothesis and the definition of the algorithm, $I_i(\boldsymbol{y}^{(m+1)}) \le I_{i_m}(\boldsymbol{y}^{(m+1)})$.

Consider now the three possibilities in which the i_mth component can be updated.

i) $I_{i_m}(\boldsymbol{y}^{(m+1)}) = 0$: in this case, the algorithm terminates since by the above, all the other components i either reach the peak constraint (case 1)) or satisfies $I_i(\boldsymbol{y}^{(m+1)}) = 0$ (cases 2) and 3)).

ii) $I_j(\boldsymbol{y}^{(m+1)}) = I_{i_m}(\boldsymbol{y}^{(m+1)})$ for some $j > i_m$. In this case, $I_{i_{m+1}}(\boldsymbol{y}^{(m+1)}) = I_{i_m}(\boldsymbol{y}^{(m+1)})$ for some i_{m+1} such that $y_{i_{m+1}}^{(m+1)} = 0$, and the claim holds for step $m+1$.

iii) $y_{i_m}^{(m+1)} = a_{i_m}$: If there exists an i such that $0 < y_i^{(m+1)} < a_i$, then i_{m+1} will satisfy $I_{i_{m+1}}(\boldsymbol{y}^{(m+1)}) = I_{i_m}(\boldsymbol{y}^{(m+1)})$ and the claim now holds for step $m+1$. If no such i exists but there is an i such that $y_i^{(m+1)} = 0$, then i_{m+1} will be chosen to satisfy $I_{i_{m+1}}(\boldsymbol{y}^{(m+1)}) \geq I_i(\boldsymbol{y}^{(m+1)})$ for all i such that $y_i^{(m+1)} = 0$, and the claim again holds for step $m+1$. Otherwise, the algorithm terminates. Thus in all cases, either the algorithm terminates or the claim holds for step $m+1$. This proves the claim.

We see from above that the algorithm terminates either via case i) or case iii). In case iii), the final point $\boldsymbol{y}^*$ satisfies

$$\begin{aligned} I_i(\boldsymbol{y}^*) &\geq 0, && \text{for } y_i^* = a_i \\ I_i(\boldsymbol{y}^*) &= 0, && \text{for } 0 < y_i^* < a_i \\ I_i(\boldsymbol{y}^*) &\leq 0, && \text{for } y_i^* = 0. \end{aligned}$$

In case iii), $\boldsymbol{y}^*$ satisfies $y_i^* = a_i$ and $I_i(\boldsymbol{y}^*) \geq 0$ for all i. Thus, in either case, $\boldsymbol{y}^*$ satisfies the Kuhn–Tucker conditions and is an optimal point.

We can also see from the above that if a component has already been increased, the only situation when the algorithm returns to that component is in case iii), when another component has reached its peak value. This implies that the event of the algorithm returning to some component that has already been increased can happen at most M times, and hence the algorithm must terminate after at most $2M$ steps.

ACKNOWLEDGMENT

This work was initiated when both authors were Post-Doctoral Members of Technical Staff at the Mathematical Sciences Research Center of AT&T Bell Laboratories. The hospitality of our department head Dr. Debasis Mitra is very much appreciated. We would also like to thank Bixio Rimoldi and Bob Gallager for their careful comments on an earlier draft.

REFERENCES

[1] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd. Int. Symp. Information Theory* (Armenian S.S.R., 1971), pp. 23–52.

[2] R. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. 39, May 1993.

[3] R. Cheng, "Optimal transmit power management on a fading multiaccess channel," in *Proc. IEEE Information Theory Workshop* (Haifa, Israel, June 1996).

[4] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Proc. Calgary Int. Conf. Combinatorial Structures and Applications*. (Calgary, Alta, June 1969), pp. 69–87.

[5] A. Federgruen and H. Groenevelt, "The greedy procedure for resource allocation problems: Necessary and sufficient conditions for optimality," *Oper. Res.*, vol. 34, pp. 909–918, 1986.

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[7] R. Gallager, "An inequality on the capacity region of multiaccess fading channels," in *Communications and Cryptography—Two Sides of One Tapestry*. Boston, MA: Kluwer, 1994, pp. 129–139.

[8] A. Goldsmith and P. Varaiya, "Capacity of fading channel with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1995.

[9] S. V. Hanly, "Information capacity of radio networks," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., Aug. 1993.

[10] S. V. Hanly and P. A. Whiting, "Constraints on capacity in a multi-user channel," presented at the International Symposium on Information Theory, Trondheim, Norway, 1994.

[11] S. V. Hanly, "An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity," *IEEE J. Select. Areas Commun.* (Special Issue on the Fundamentals of Networking), vol. 13, Sept. 1995.

[12] S. V. Hanly and D. Tse, "Multi-access fading channels: Shannon and delay-limited capacities," in *Proc. 33rd Allerton Conf.* (Monticello, IL, Oct. 1995).

[13] H. Liao, "Multiple access channels," Ph.D. dissertation, Dep. Elec. Eng., Univ. Hawaii, Honlulu, 1972.

[14] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," presented at the International Conference on Communications, Seattle, WA, June 1995.

[15] M. Medard, "The capacity of time varying multiple user channels in wireless communications," Ph.D. dissertation, MIT, Cambridge, MA, 1995.

[16] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems," in *4th Winlab Workshop on Third Generation Wireless Information Networks*, 1993, pp. 249–257.

[17] S. Shamai and A. D. Wyner, "Information theoretic considerations for symmetric, cellular, multiple-access fading channels—Part I," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1877–1894, Nov. 1997.

[18] G. L. Turin *et al.*, "A statistical model for urban radio propagation," *IEEE Trans. Veh. Technol.*, vol. 21, pp. 1–9, Feb. 1972.

[19] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.* (Special Issue on the Fundamentals of Networking), vol. 13, no. 7, 1995.