MULTI-ACCESS FADING CHANNELS:
PART I: POLYMATROIDAL STRUCTURE, OPTIMAL RESOURCE ALLOCATION AND THROUGHPUT CAPACITIES

David N. Tse and Stephen V. Hanly

MULTI-ACCESS FADING CHANNELS:
PART II: DELAY-LIMITED CAPACITIES

Stephen V. Hanly and David N. Tse

Memorandum No. UCB/ERL M96/69

17 November 1996

MULTI-ACCESS FADING CHANNELS:
PART I: POLYMATROIDAL STRUCTURE, OPTIMAL RESOURCE ALLOCATION AND THROUGHPUT CAPACITIES

David N. Tse and Stephen V. Hanly

MULTI-ACCESS FADING CHANNELS:
PART II: DELAY-LIMITED CAPACITIES

Stephen V. Hanly and David N. Tse

Memorandum No. UCB/ERL M96/69

17 November 1996

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Multi-access Fading Channels: Part I: Polymatroidal Structure, Optimal Resource Allocation and Throughput Capacities *

David N. Tse †and Stephen V. Hanly ‡

Abstract

In multi-access 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 multi-access 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 characterizes 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 a multi-user generalization of the water-pouring construction for single user channels, and exploits the underlying *polymatroidal* structure of the capacity region. In part II, we characterize a *delay-limited* capacity region and obtained analogous results.

1 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 dBs. The problem is particularly acute for real-time traffic such as video, since they have a stringent delay requirement. A general strategy to combat these detrimental effects is through the

*Part of this work was done when both authors were with A.T. & T. Bell Laboratories. This paper has been submitted to the IEEE Transactions on Information Theory.

†Dept. of EECS, U.C. Berkeley, dtse@eecs.berkeley.edu

‡Dept. of Electrical Eng., University of Melbourne, s.hanly@ee.mu.oz.au

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) and FDMA (frequency-division). 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 multi-access 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 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 multi-access Gaussian capacity region. Since the number of constraints defining the capacity region is exponential in the number of users, to obtain efficient solutions we need to exploit the special *polymatroidal* structure of the capacity region. Polymatroidal structure has been used successfully in many resource allocation problems to obtain greedy optimization algorithms (see for example [4].) In this paper, we will show that the multi-access 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 [7] 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 [13] have solved the multi-user version of that problem for the special case of symmetric users with equal rate requirements. Our results on computing the entire throughput capacity region of the multi-access fading channel and the associated optimal power allocation can be viewed as generalization of the classic water-filling solution to the multi-user setting. In a related work, Cheng and Verdu [2] obtained an explicit characterization of the capacity region of the two-user time-invariant multi-access Gaussian channel with inter-symbol interference. We will see that this channel is essentially the "frequency-dual" of the multi-access flat fading channel and our techniques for the latter can be readily applied and provide a general solution to the multi-access 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 [11] 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 [10], [18]), where the goal is to maintain a desired signal-to-noise ratio at *all* fading states. However, those works consider only decoding schemes where a user is decoded treating other users as interference, which is sub-optimal 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 [8], [9] established structural results about the multi-user Gaussian capacity region arising directly from its polymatroidal 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 2 we introduce the Gaussian, multi-access, 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 3, 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 multi-access 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 4 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 5, 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 6, we will present greedy power allocation algorithms when additional power constraints are imposed. These results exploit further properties of polymatroids. Finally, in Section 7, we extend our flat fading model to the case of frequency selective fading.

2 The Multi-access Fading Channel

2.1 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 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 non-selective. Frequency-selective fading will be considered in Section 7.

Suppose each source i codes over a block length of T symbols, where T is the delay, using a codebook C_i of size $2^{R_i T}$ (i.e. at rate R_i bits per channel use). Each codeword $\mathbf{x}$ of the ith user satisfies $\|\mathbf{x}\|_2^2 \leq T\bar{P}_i$. Fix a decoding scheme and assume the messages are chosen with equal probability. Let $p_e(T)$ be the probability of the event that any user is decoded incorrectly. The *capacity region* characterizes the fundamental limits of communication in the multi-access scenario:

Definition 2.1 *The rate-tuple* $\mathbf{R} = (R_1, \ldots, R_M)$ *lies inside the capacity region* $\mathcal{C}$ *iff for every positive* ϵ, *there exists a delay* T, *a codebook of block length* T, *and a decoding scheme such that the probability of error* $p_e(T)$ *is less than* ϵ.

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. $H_i(t) = h_i$ for all time t. The characterization of the capacity region of the multi-access memoryless channel with probability transitions $p(y|x_1, \ldots, x_M)$ is well-known (Ahlswede [1], Liao [12]); it is the set of all rate vectors $\mathbf{R}$ satisfying:

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

for some independent input distribution $p(x_1)p(x_2)\ldots p(x_M)$. (In this paper, for any vector $\mathbf{x}$ we use the notation $\mathbf{x}(S)$ to denote $\sum_{i\in S} \mathbf{x}(S)$.) Note that S is any subset of users in $\{1, 2, \ldots, M\}$. 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 multi-access channel, this capacity region reduces to:

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

where $\mathbf{h} = (h_1, \ldots, h_M)$ and $\mathbf{P} = (P_1, \ldots, P_M)$ Note that this region is characterized by $2^M - 1$ constraints, each corresponding to a non-empty 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 $n!$ vertices in the positive quadrant, each achievable by a successive decoding using one of the $n!$ 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 [6], Shamai and Wyner [16]) and is given by:

$$\{(R_1, \ldots, R_M) : \mathbf{R}(S) \le \mathbb{E}_{\mathbf{H}}[\frac{1}{2} \log(1 + \frac{P}{\sigma^2} \sum_{i\in S} H_i)], \forall S \subset \{1, \ldots, M\}\} \tag{3}$$

where $\mathbf{H} = (H_1, \ldots, H_M)$ is a random vector having the stationary distribution of the joint fading process. An intuitive understanding of this result can be obtained by viewing capacities in terms of time averages of mutual information (Gallager [6]), the rate of flow of which can be viewed as a random process depending on the fading levels of the users. Specifically, at time t, the instantaneous 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:

$$\frac{1}{2}\log(1+\frac{P}{\sigma^2}\sum_{i\in S}H_i(t))$$

(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

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

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 multi-access 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.

2.2 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 slower 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.

A *power control policy* $\mathcal{P}:\Re^M\to\Re^M$ is a mapping such that given a joint fading state $\mathbf{h}=(\mathbf{h_1},\ldots\mathbf{h_M})$ for the users, $\mathcal{P}_i(\mathbf{h})$ 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:

$$C_f(\mathcal{P})\equiv\{\mathbf{R}:\mathbf{R}(S)\leq\mathbb{E}_{\mathbf{H}}[\frac{1}{2}\log(1+\frac{1}{\sigma^2}\sum_{i\in S}H_i\mathcal{P}_i(\mathbf{H}))],\forall S\subset\{1,\ldots,M\}\}\qquad(4)$$

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 2.2 *The throughput capacity region for the multi-access fading Gaussian channel when all the transmitters as well as the receiver know the current state of the channel is given by:*

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

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

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

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.

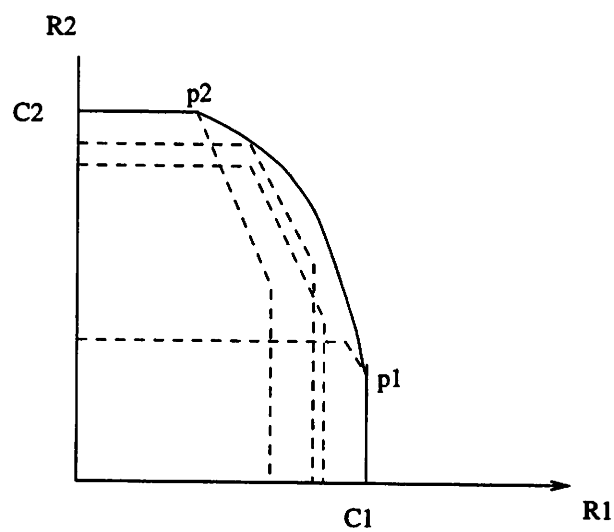

Figure 1: Figure shows a two-user throughput capacity region as a union of capacity regions, each corresponding to a feasible power control $\mathcal{P}$. Note that each of these regions is a pentagon (shown in dashed lines). The boundary surface is the curved part.

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 multi-user, Gaussian channel with a unit power constraint on each user, and in which the fading level for user i is $H_i\mathcal{P}_i(\mathbf{H})$. This channel has capacity region $\mathcal{C}_f(\mathcal{P})$. Consider then any rate $\mathbf{R}$ in the interior of $\mathcal{C}_f(\mathcal{P})$. Given any positive ϵ, we can choose a code length and a codebook (nonrandom) such that the probability of error is less than ϵ. But, as in the proof of Theorem 2.2, 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.

3 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 $\mathcal{C}_f$.

3.1 Polymatroidal Structure

We begin with a few definitions. As before, for a vector $\mathbf{x} \in \Re^M$, we shall use the short-hand notation $\mathbf{x}(S)$ to denote $\sum_{i \in S} x_i$.

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

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

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) $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ (sub modular)

The polyhedron

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

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.*

Polymatroids were introduced by Edmonds [3] where he proved the following key properties. If π is a permutation on the set E, define the vector $\mathbf{v}(\pi) \in \Re^M$ by $v_{\pi(1)}(\pi) = f(\pi(1))$ and $v_{\pi(i)}(\pi) = f(\{\pi(1), \ldots, \pi(i)\}) - f(\{\pi(1), \ldots, \pi(i-1)\})$ for $i = 2, \ldots, M$.

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 λ is a given vector in $\Re^M_+$, then the solution of the optimization problem*

$$\max \lambda \cdot \mathbf{x} \qquad \textit{subject to} \qquad \mathbf{x} \in \mathcal{B}(f) \tag{7}$$

is attained at the point $v(\pi^*)$ where the permutation π^* is given by $\lambda_{\pi^*(1)} \geq \ldots \geq \lambda_{\pi^*(M)}$. 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, 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 $\{\pi(1), \ldots, \pi(k)\}$. 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. Also, after increasing a component of the vector, the algorithm never re-visits 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 λ is a given vector in $\Re^M_+$, then the solution of the optimization problem*

$$\min \lambda \cdot \mathbf{x} \qquad \textit{subject to} \qquad \mathbf{x} \in \mathcal{G}(f) \tag{8}$$

is attained at the point $v(\pi^*)$ where the permutation π^* is given by $\lambda_{\pi^*(1)} \geq \ldots \geq \lambda_{\pi^*(M)}$. 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 multi-access channel with transition matrix $p(y|x_1, \ldots, x_M)$. A similar version of this result was obtained in [9].

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

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

is a polymatroid.

Proof. Let π be a permutation on E and consider the rate vector $R(\pi)$ defined by

$$\begin{aligned} R_{\pi(i)}(\pi) &= I[Y; X_{\pi(i)}|\mathbf{X}(\{\pi(i+1),\dots,\pi(M)\})] \qquad i = 1,\dots,M-1 \\ R_{\pi(M)}(\pi) &= I[Y; X_{\pi(M)}] \end{aligned}$$

These are the capacities achieved by successive decoding in the order given by π, and hence the rate vector $\mathbf{R}(\pi)$ 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 $\mathcal{C}_g(\mathbf{h}, \mathbf{P})$ of a memoryless Gaussian multi-access 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 $\mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h}))$.

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 $\mathbf{h}$, $\mathcal{R}_i(\mathbf{h})$ can be interpreted as the rate allocated to user i while the users are in state $\mathbf{h}$.*

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

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

Furthermore, for any permutation π on E,

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

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

Proof. Define

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

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 $\mathbf{R}(\pi)$, where π ranges over all permutations of E, and

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

But for any π, $\mathbf{R}(\pi) \in \mathcal{E}$, 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.

□

3.2 A Lagrangian characterization of the capacity region

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

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

Definition 3.9 *The boundary surface of $\mathcal{C}(\mathbf{P})$ is the set of those rates such that we cannot increase one component, and remain in $\mathcal{C}(\mathbf{P})$ without decreasing another.*

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 component-wise 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 $\mathcal{C}(\bar{\mathbf{P}})$ and the associated optimal power control policy can be reduced to solving a family of optimization problems over a set of parallel multi-access Gaussian channels.

Lemma 3.10 *A rate vector $\mathbf{R}^*$ lies on the boundary surface of $\mathcal{C}(\bar{\mathbf{P}})$ if and only if there exists a nonnegative $\mu \in \Re^M$ such that $\mathbf{R}^*$ is a solution to the optimization problem:*

$$\max \mu \cdot \mathbf{R} \qquad \textit{subject to } \mathbf{R} \in \mathcal{C}(\bar{\mathbf{P}}). \tag{12}$$

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

$$\max_{(\mathbf{R},\mathbf{P})} \mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P} \qquad \textit{subject to} \qquad \mathbf{R} \in \mathcal{C}_g(\mathbf{h}, \mathbf{P}) \tag{13}$$

and

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

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

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

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

$$\max_{(\mathbf{R},\mathbf{P})} \mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P} \qquad \text{subject to} \qquad \mathbf{R} \in \mathcal{C}_f(\mathcal{P}) \tag{14}$$

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

$$\begin{aligned} R_{\pi(1)} &= \mathbb{E}_{\mathbf{H}}[\frac{1}{2}\log(1 + \frac{H_{\pi(1)}\mathcal{P}_{\pi(1)}(\mathbf{H})}{\sigma^2})] \\ R_{\pi(k)} &= \mathbb{E}_{\mathbf{H}}[\frac{1}{2}\log(1 + \frac{H_{\pi(k)}\mathcal{P}_{\pi(k)}(\mathbf{H})}{\sigma^2 + \sum_{i=1}^{k-1} H_{\pi(i)}\mathcal{P}_{\pi(i)}(\mathbf{H})})] \qquad k = 2, \ldots, M \end{aligned} \tag{15}$$

Hence, the optimization problem (14) is equivalent to

$$\max \mu_1 \mathbb{E}_{\mathbf{H}}[\frac{1}{2}\log(1+\frac{H_{\pi(1)}\mathcal{P}_{\pi(1)}(\mathbf{H})}{\sigma^2})] + \sum_{k=2}^{M} \mu_k \mathbb{E}_{\mathbf{H}}[\frac{1}{2}\log(1+\frac{H_{\pi(k)}\mathcal{P}_{\pi(k)}(\mathbf{H})}{\sigma^2 + \sum_{i=1}^{k-1} H_{\pi(i)}\mathcal{P}_{\pi(i)}(\mathbf{H})})] + \lambda \cdot \mathbb{E}_{\mathbf{H}}[\mathcal{P}(\mathbf{H})] \tag{16}$$

and this is in turn equivalent to

$$\max \mu_1 \frac{1}{2}\log(1 + \frac{H_{\pi(1)}\mathcal{P}_{\pi(1)}(\mathbf{h})}{\sigma^2}) + \sum_{k=2}^{M} \mu_k \frac{1}{2}\log(1 + \frac{H_{\pi(k)}\mathcal{P}_{\pi(k)}(\mathbf{h})}{\sigma^2 + \sum_{i=1}^{k-1} H_{\pi(i)}\mathcal{P}_{\pi(i)}(\mathbf{h})}) + \lambda \cdot \mathcal{P}(\mathbf{h})$$

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

$$\max \mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P} \qquad \text{subject to} \qquad \mathbf{R} \in \mathcal{C}_g(\mathbf{h}, \mathbf{P})$$

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

This completes the proof. $\square$

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

It follows immediately from (15) that the optimal solution will be a successive decoding solution. Lemma 3.8 then shows that the optimal solution $(\mathcal{R}^*(\mathbf{H}), \mathcal{P}^*(\mathbf{H}))$ will be such that $\mathcal{R}^*(\mathbf{h})$ is a corner point of $\mathcal{C}_g(\mathbf{h}, \mathcal{P}^*(\mathbf{h}))$ for every $\mathbf{h}$, with the same ordering π for each $\mathbf{h}$.

3.3 Optimal Power and Rate Allocation

We now consider the problem of determining $(\mathcal{R}^*(\mathbf{h}), \mathcal{P}^*(\mathbf{h}))$ for each fading state $\mathbf{h}$. Note that Lemma 3.10 can be viewed as a multi-access generalization of the Lagrangian formulation for the problem of allocating powers over a set of parallel single-user Gaussian channels ([5]). 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 multi-access setting. Again we make use of the polymatroid structure and the solution will have a greedy flavor.

To make further progress, we now identify further structure in the time-invariant multi-access Gaussian capacity region $\mathcal{C}_g(\mathbf{h}, \mathbf{P})$.

Definition 3.11 *(see [4])*

The rank function f of a polymatroid $\mathcal{B}(f)$ is generalized symmetric *if there exists a vector $\mathbf{y} \in \Re_+^M$ and a non-decreasing concave function g such that $f(S) = g(\mathbf{y}(S))$ for every $S \subset E$.*

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

Lemma 3.12 *Let g be a non-decreasing concave function and for each $\mathbf{y}$, define generalized symmetric rank function $f_{\mathbf{y}}(S) \equiv g(\mathbf{y}(S))$. Fixed a vector $\mathbf{x} \in \Re_+^M$. The set $\{\mathbf{y} : \mathbf{x} \in \mathcal{B}(f_{\mathbf{y}})\}$ is a contra-polymatroid.*

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

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

$$\mathcal{Q}(\mathbf{h}, \mathbf{R}) \equiv \{\mathbf{Q} : \exists \mathbf{P} s.t. Q_i = h_i P_i, \mathbf{R} \in \mathcal{C}_g(\mathbf{h}, \mathbf{P})\}$$

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.

Theorem 3.14 *Consider the problem:*

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

where g is a non-decreasing concave function. Define the utility functions

$$\begin{aligned} u_i(z) &\equiv \mu_i g'(z) - \lambda_i, \qquad i = 1, \ldots, M \\ u^*(z) &\equiv [\max_i u_i(z)]^+ \end{aligned}$$

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

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

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

$$\begin{aligned} x_{\pi(1)} &= g(y_{\pi(1)}) \\ x_{\pi(k)} &= g(\sum_{i=1}^{k} y_{\pi(i)}) - g(\sum_{i=1}^{k-1} y_{\pi(i)}) \end{aligned}$$

for some permutation π. Hence,

$$\begin{aligned} J^* &= \max_{\mathbf{y}} \mu_{\pi(1)} g(y_{\pi(1)}) + \sum_{k=2}^{M} \mu_{\pi(k)} [g(\sum_{i=1}^{k} y_{\pi(i)}) - g(\sum_{i=1}^{k-1} y_{\pi(i)})] - \lambda \cdot \mathbf{y} \\ &= \max_{\mathbf{y}} \sum_{k=1}^{M} \int_{\sum_{i \leq k-1} y_{\pi(i)}}^{\sum_{i \leq k} y_{\pi(i)}} u_{\pi(i)}(z)dz \\ &\leq \int_0^\infty u^*(z)dz \end{aligned}$$

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 $\mathbf{x} = \mathbf{y} = 0$. If $u^*(0) > 0$, then let $0 = z_0 < z_1 < \ldots < z_K$ 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, $k = 0, \ldots K-1$. 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

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

It can be directly verified that

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

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

□

Observe that the solution can be obtained via a greedy algorithm. Starting with $\mathbf{x} = \mathbf{y} = 0$, 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 a solution to the optimization problem (13). The function g is taken to be

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

In terms of the received powers $\mathbf{Q} = (h_1 P_1, \ldots, h_M P_M)$, the optimization problem can be rewritten as:

$$\max \sum_i \mu_i R_i - \sum_i \frac{\lambda_i}{h_i} Q_i \qquad \text{subject to } \mathbf{R}(S) \leq g(\mathbf{Q}(S)) \quad \forall S \subset E$$

The optimal solution is achieved by successive decoding. Any such solution can be represented by a permutation π and set of intervals $[z_i, z_{i+1}]$, $i = 1, \ldots, M$ 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 $\pi(M), \pi(M-1), \ldots, \pi(1)$. 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 noise 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 $\sigma^2 + z$, $z \in [0, \infty)$. See Fig. 3.3 for an example.

The optimal choice is determined by the functions

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

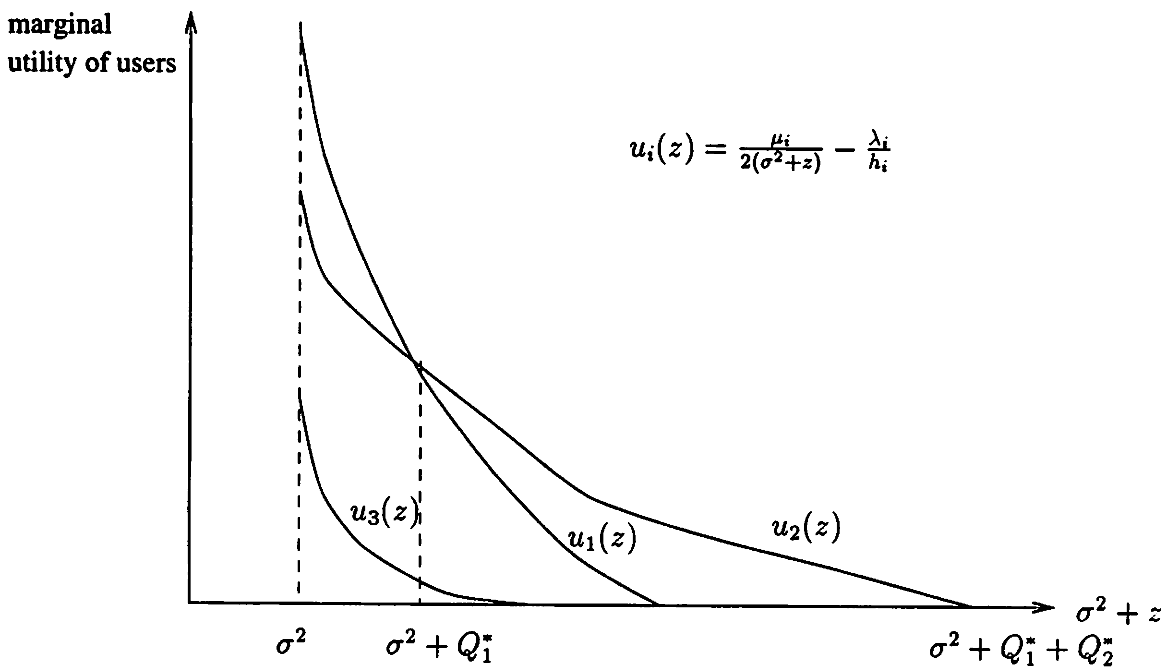

Figure 2: A 3-user example illustrating the greedy power allocation. The x-axis represents the received interference level and y-axis the marginal utility of each user at the interference levels. At each interference level, the user who is selected to transmit is the one with the highest marginal utility. 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.

where $u_i(z) \cdot \delta Q$ can be interpreted as the marginal increase in the value of the objective function due to an amount δQ of power *received* from user i at noise level $\sigma^2 + z$. Thus, the optimal solution is obtained in a greedy fashion by choosing at each noise level z, to transmit the user which will leads to the largest positive marginal increase in the objective function. If no such user can be found, then no user is transmitted at that interference level. 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. In the case when the $\frac{\lambda_i}{h_i}$'s are all distinct (which happens with probability 1), the optimal power and rate allocation is explicitly given by:

$$R_i^*(\mathbf{h}) = \int_{\mathcal{A}_i} \frac{1}{2(\sigma^2 + z)} dz \qquad P_i^*(\mathbf{h}) = |\mathcal{A}_i|$$

where

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

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. Note that the *same* ordering π is used in every fading state, although users may end up using zero power and zero rate in any given fading state.

3.4 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 $\mathcal{C}(\bar{\mathbf{P}})$, 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 μ.

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

Proof. See appendix B. □

It should be noted that the uniqueness result above only holds for positive μ. If some of the rewards μ_i's equal 0, the $\mathbf{R}^*$ which maximizes $\mu \cdot \mathbf{R}$ 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 μ for a parameterization of the boundary surface. We will give a more explicit interpretation of these extreme points in Section 3.5.

For any such positive μ, the above lemma implies that we can define a parameterization $\mathbf{R}^*(\mu)$ which is the unique rate vector on the boundary which maximizes $\mu \cdot \mathbf{R}$. Its value can be obtained using the greedy rate and power allocation solution, with λ 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, $\mathbf{R}^*(\mu)$ has a particularly simple form.

For given μ and λ, let $\mathcal{R}^*(\mathbf{h}, \mu, \lambda)$ and $\mathcal{P}^*(\mathbf{h}, \mu, \lambda)$ 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 due to the greedy allocation procedure, which user to transmit at each interference level z only depends on the values of the marginal utility functions of the user 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,

$$\begin{aligned}
\mathbb{E}_{\mathbf{H}}[\mathcal{R}_i^*(\mathbf{H}, \mu, \lambda)] &= \int_0^\infty \mathbb{E}_{\mathbf{H}}\left[g'(z) I_{\{u_i(z) > u_j(z) \quad \forall j \text{ and } u_i(z) > 0\}}\right] dz \\
&= \int_0^\infty g'(z) \Pr(u_i(z) > u_j(z) \quad \forall j \text{ and } u_i(z) > 0) dz \\
&= \int_0^\infty \frac{1}{2(\sigma^2 + z)} \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k\left(\frac{2\lambda_k h(\sigma^2 + z)}{2\lambda_i(\sigma^2 + z) + (\mu_k - \mu_i)h}\right) f_i(h) dh \right\} dz \\
\mathbb{E}_{\mathbf{H}}[\mathcal{P}_i^*(\mathbf{H}, \mu, \lambda)] &= \int_0^\infty \mathbb{E}_{\mathbf{H}}\left[\frac{1}{h_i} I_{\{u_i(z) > u_j(z) \quad \forall j \text{ and } u_i(z) > 0\}}\right] dz \\
&= \int_0^\infty \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \frac{1}{h} \prod_{k \neq i} F_k\left(\frac{2\lambda_k h(\sigma^2 + z)}{2\lambda_i(\sigma^2 + z) + (\mu_k - \mu_i)h}\right) f_i(h) dh \right\} dz, \quad (17)
\end{aligned}$$

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 $C(\bar{P})$. Note that since $\mathcal{R}^*$ and $\mathcal{P}^*$ are invariant under scalings of the vectors μ and λ, 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 $C(\bar{P})$ is the closure of the parametrically defined surface*

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

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

$$R_i^*(\mu) = \int_0^\infty \frac{1}{2(\sigma^2 + z)} \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \prod_{k \neq i} F_k\left(\frac{2\lambda_k h(\sigma^2 + z)}{2\lambda_i(\sigma^2 + z) + (\mu_k - \mu_i)h}\right) f_i(h) dh \right\} dz \quad (18)$$

where the vector λ is the unique solution of the equations:

$$\int_0^\infty \left\{ \int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^\infty \frac{1}{h} \prod_{k \neq i} F_k\left(\frac{2\lambda_k h(\sigma^2 + z)}{2\lambda_i(\sigma^2 + z) + (\mu_k - \mu_i)h}\right) f_i(h) dh \right\} dz = \bar{P}_i \quad (19)$$

$i = 1, \ldots, M$. 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 2-dimensional integrals. For a given μ, it should therefore be possible to compute λ numerically with low complexity. We shall present an algorithm to do this in Section 4, 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:

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

by reversing the order of integration. Using (19), the constant $\frac{\mu}{2\lambda}$ is shown to satisfy the power constraint

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

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 [7] 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 $\mu_1 = \ldots = \mu_M = 1$, 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 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 $\mathbf{h}$, the user with the maximum value of $u_i(z)$ does not depend on 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:

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

The optimal rates are given by

$$R_i^* = \int_0^\infty \frac{1}{2} \log(1 + \frac{h}{\sigma^2}(\frac{1}{2\lambda_i} - \frac{\sigma^2}{h})^+) \prod_{k \neq i} F_k(\frac{\lambda_k}{\lambda_i} h) f(h)dh, \qquad i = 1, \ldots, M$$

where the constants λ_i's satisfy:

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

This solution was recently obtained by Knopp and Humblet [13].

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 a way to maintain fairer allocation of resources to the users. We consider this issue further in Section 5.

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

3.5 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 $U \equiv \{1, 2, \ldots, M\}$ 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 cancelled, before any user in the complementary subset is decoded, for *every* fading state $\mathbf{h}$. With positive vectors μ and λ, 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 $\mathcal{C}(\mathbf{P})$ are characterized in exactly this way: by a positive (μ, λ) 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 after 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.

4 An Iterative Algorithm for Resource Allocation

In Section 3.2, we provided a Lagrangian characterization of the boundary surface of $\mathcal{C}(\bar{\mathbf{P}})$. In particular, we we characterize a boundary point by a positive rate rewards vector μ, and that associated with this is a unique positive shadow power price vector λ. We now present a simple iterative algorithm to compute λ, for a given μ and average power constraints $\bar{\mathbf{P}}$. In the case when the fadings of the users are independent, this amounts to solving the nonlinear equations (19) for λ 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 reward μ and power constraints $\bar{\mathbf{P}}$ to be given and fixed. Let us define $\mathbf{R}(\lambda)$ and $\mathbf{P}(\lambda)$ to be the rate and average powers under the optimal power control associated with the prices (μ, λ). 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 λ is increased and the other components are held fixed, $P_i(\lambda)$ decreases while $P_j(\lambda)$ 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 $\bar{\mathbf{P}}$, let $\mathbf{R}^*$ be the optimum rate corresponding to the rewards μ, and let λ^* 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 $\lambda(0)$ be an initial arbitrary set of positive power prices. Given the nth iterate $\lambda(n)$, the $n+1$th iterate $\lambda(n+1)$ 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 $\lambda(n)$. (The uniqueness follows from the monotonicity property above.)*

In terms of the equations (19) for the case when the fading is independent, $\lambda_i(n+1)$ is the unique root x of the equation

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

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 $\bar{\mathbf{P}}$, let $\mathbf{R}^*$ be the optimum rate corresponding to the rewards μ^*, and let λ^* be the shadow prices at the point $(\bar{\mathbf{P}}, \mathbf{R}^*)$. Then*

$$\lambda(n) \to \lambda^*, \quad n \uparrow \infty$$

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

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

Lemma 4.4 *(i) For any positive $\lambda(0)$, there exists $\lambda \leq \lambda(0)$ for which $P(\lambda) \geq \bar{P}$.*
(ii) For any positive $\lambda(0)$, there exists $\lambda \geq \lambda(0)$ for which $P(\lambda) \leq \bar{P}$.

Proof. See appendix C. □

Algorithm 4.2 defines a mapping

$$\begin{aligned} T : \Re_+^M &\to \Re_+^M \\ \lambda(n) &\mapsto \lambda(n+1) \end{aligned}$$

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 (19) for given μ. The second follows from Lemma 4.1.

Lemma 4.5 *(i) The vector of shadow prices λ^* corresponding to the point $(\bar{\mathbf{P}}, \mathbf{R}^*)$ is the unique fixed point of T.*

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

The following lemma is also useful.

Lemma 4.6 *(i) If $\lambda(0) \geq T(\lambda(0))$ and we define*

$$\lambda(n) \equiv T^n(\lambda(0)) \quad n = 0, 1, 2, \ldots$$

then $\lambda(n)$ is a decreasing sequence.
(ii) If $\lambda(0) \leq T(\lambda(0))$ then $\lambda(n)$ is an increasing sequence, and $\lambda(n) \uparrow \lambda^*$.
(iii) If $\lambda(0) \geq T(\lambda(0))$ then $\lambda(n) \downarrow \lambda^*$.

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

Proof of Theorem 4.3 Lemma 4.4 guarantees the existence of points $\mathbf{w}(0)$ and $\mathbf{z}(0)$ with the following properties:
(i) $\mathbf{w}(0) \leq \lambda(0) \leq \mathbf{z}(0)$
(ii) $P(\mathbf{w}(0)) \geq \bar{P}$
(iii) $P(\mathbf{z}(0)) \leq \bar{P}$
Now define $\mathbf{w}(n) = T^n(\mathbf{w}(0))$ and $\mathbf{z}(n) = T^n(\mathbf{z}(0))$. It follows from property (ii) and Lemma 4.6(ii) that $\mathbf{w}(n) \uparrow \lambda^*$. Similarly, it follows from property (iii) and Lemma 4.6(iii) that $\mathbf{z}(n) \downarrow \lambda^*$. Finally, it follows from property (i) and the order preserving property of T that $\mathbf{w}(n) \leq \lambda(n) \leq \mathbf{z}(n)$. We conclude that $\lambda(n) \to \lambda^*$. □

Algorithm 4.2 has all the users updating $\lambda(n)$ simultaneously. However, convergence still occurs if users update one at a time, or even asynchronously under certain weak conditions (Mitra [15]). 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 [10], Yates [18]). 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 multi-user decoding, increasing power can *benefit* 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, λ. Nevertheless, monotonicity occurs in λ-space, enabling very similar iterative procedures to be applied.

5 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, let us define the set $\mathcal{D}(\mathbf{R})$ and its boundary surface; it is the "power space equivalent" of the capacity region $\mathcal{C}(\mathbf{P})$:

Definition 5.1 • $\mathcal{D}(\mathbf{R}) \equiv \{\bar{\mathbf{P}} : \mathbf{R} \in \mathcal{C}(\bar{\mathbf{P}})\}$

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

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

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

$$\min \lambda \cdot \mathbf{P} \qquad \textit{subject to } \mathbf{P} \in \mathcal{D}(\mathbf{R}^*) \tag{20}$$

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

$$\max_{(\mathbf{R},\mathbf{P})} \mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P} \qquad \textit{subject to} \qquad \mathbf{R} \in \mathcal{C}_g(\mathbf{h}, \mathbf{P}) \tag{21}$$

and

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

Moreover, for a given λ and $\mathbf{R}^*$, $\bar{\mathbf{P}}$ and μ 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

$$S = \{(\mathbf{R}, \mathbf{P}) : \mathbf{R} \in \mathcal{C}(\mathbf{P})\} = \{(\mathbf{R}, \mathbf{P}) : \mathbf{P} \in \mathcal{D}(\mathbf{R})\}. \tag{22}$$

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

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

Algorithm 5.3 *Let $\mu(0)$ be an initial arbitrary set of positive shadow rewards for rates. Given the nth iterate $\mu(n)$, the $n+1$th iterate $\mu(n+1)$ 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 $\mu(n)$.*

Denote the rate by $\mathbf{R}(\mu(n))$ and the average power by $\mathbf{P}(\mu(n))$ 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 bit rate $\mathbf{R}^*$, let $\bar{\mathbf{P}}$ be the optimum average power corresponding to the prices λ, and let μ^* be the appropriate shadow rewards. Then*

$$\mu(n) \to \mu^*, \quad n \uparrow \infty$$

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

We have seen that given rate rewards μ and power constraints $\bar{\mathbf{P}}$, there exists a unique $\mathbf{R}^*$ which maximizes $\mu \cdot \mathbf{R}$ and unique Lagrangian power prices λ^*. Similarly, given power prices λ and target rates $\mathbf{R}^*$, there exists unique $\bar{\mathbf{P}}$ which minimizes $\lambda \cdot \mathbf{P}$ and unique Lagrangian rewards μ^*. In fact, one can also show that given $(\mathbf{R}^*, \bar{\mathbf{P}})$ on the boundary of S (defined in eqn. (22)), there exist unique μ^*, λ^* such that

$$(\mathbf{R}^*, \bar{\mathbf{P}}) = \text{argmax}_{(\mathbf{R},\mathbf{P}) \in S} \quad \mu^* \cdot \mathbf{R} - \lambda^* \cdot \mathbf{P}$$

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

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 reward earned in the channel is then $\mu^* \cdot \mathbf{R}$. Lemma 3.10 shows that any point $\mathbf{R}^*$ on the interior of the boundary surface of $\mathcal{C}(\bar{\mathbf{P}})$ can be obtained as a maximization of total reward. The lemma shows that at the optimal solution $\mathbf{R}^*$, a set of shadow prices λ^* exist, in the sense that if we change the power constraint by $\Delta\mathbf{P}$, then we change the reward earned by $\lambda^* \cdot \Delta\mathbf{P}$. However, it is clear from Lemma 5.2, that we can interpret λ^* directly as a set of "power prices". To see this, consider problem (20), and interpret $\sum_i \lambda_i P_i$ as the total price of the power vector $\mathbf{P}$. At any solution $\mathbf{P}$, there is an associated shadow reward μ on the rates. Now if we set $\lambda \equiv \lambda^*$, then by the uniqueness of the supporting hyperplane to S at $(\mathbf{R}^*, \bar{\mathbf{P}})$ we must have that $\mu = \mu^*$. 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 (20) are the rate rewards in (12).

We therefore consider the following economic framework for resource allocation. We are given a vector μ^* of bit rate rewards, and a vector λ^* of power prices, and our aim is to find the optimal operating point $(\mathbf{R}^*, \bar{\mathbf{P}})$ such that $\mu^* \cdot \mathbf{R} - \lambda^* \mathbf{P}$ is maximized. Section 3.2 provides a greedy algorithm which attains this optimal operating point.

6 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:

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

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:

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

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

$$\max \mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P} \quad \textit{subject to} \quad \mathbf{R} \in C_g(\mathbf{h}, \mathbf{P}) \quad \textit{and} \quad P_i \leq \hat{P}_i \forall i \tag{23}$$

and

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

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* $\mathbf{R}^*$*.*

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

$$\max \mu \cdot \mathbf{x} - \lambda \cdot \mathbf{y} \quad \text{subject to} \quad \mathbf{x}(S) \leq g(\mathbf{y}(S)) \quad \forall S \subset E, \quad 0 \leq y_i \leq a_i \quad \forall i \tag{24}$$

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 $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_M$. By Lemma 3.2, for any vector $\mathbf{y}$, the maximum value of $\mu \cdot \mathbf{x}$ subject to the polymatroid constraints is given by

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

Hence the optimization problem (24) is equivalent to

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

$$\text{subject to} \quad 0 \leq y_i \leq a_i$$

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

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

and let

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

. We first observe two facts:

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

Fact 2: For $j > i$,

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

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* $\mathbf{y}^{(0)} = 0$. *Set* $k = 0$.

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

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

(where v_j *is in the* i_k*th 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*

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

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

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

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 (24), and the number of steps needed is at most* $2M$.

Proof. See appendix D. □

The optimal power allocation problem with auxiliary constraints (23) can be expressed in terms of the received powers $\mathbf{Q} = (h_1P_1, \ldots, h_MP_M)$:

$$\max \sum_i \mu_i R_i - \sum_i \frac{\lambda_i}{h_i} Q_i \quad \text{subject to } \mathbf{R}(S) \leq g(\mathbf{Q}(S)) \quad \forall S \subset E \quad \text{and} \quad Q_i \leq \frac{\hat{P}_i}{h_i} \forall i$$

where

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

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.

7 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 $\frac{n_0}{2}$. 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 basestation. 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

$$Y(n) = y(\frac{n}{T}), \quad X(n) = x(\frac{n}{T}), \quad H_i(k,n) = \int \frac{\sin(\pi(\frac{k}{T}-\tau))}{\pi(\frac{k}{T}-\tau)} h_i(\tau,\frac{n}{T})d\tau$$

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 multi-access channel with inter-symbol interference (ISI), and a characterization of the capacity region has been obtained by Cheng and Verdu [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

$$\mathcal{F} = \{\mathcal{P} : \int_{\frac{-W}{2}}^{\frac{W}{2}} \mathcal{P}_i(f) df \leq \bar{P}_i \qquad \forall i\}$$

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

$$\bigcup_{\mathcal{P} \in \mathcal{F}} \{\mathbf{R} : \mathbf{R}(S) \leq \int_{\frac{-W}{2}}^{\frac{W}{2}} \log(1 + \frac{\sum_{i \in S} \mathcal{P}_i(f) |\hat{H}_i(f)|^2}{\sigma^2}) \qquad \forall S \subset \{1, \ldots, M\}\} \tag{25}$$

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 multi-user case, which follows almost directly from the results in Section 3. 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 2.2), with the role of the fading state $\mathbf{h}$ now played by frequency f. Using the results of Section 3, 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 multi-access channel with ISI is*

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

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

$$R_i^*(\mu) = \int_0^\infty \frac{1}{(\sigma^2 + z)} \mathrm{m}(A_i(z, \lambda)) dz \tag{26}$$

where

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

and $\mathrm{m}(\cdot)$ *is the Lebesgue measure of a set. The vector* λ *satisfies the equations:*

$$\int_0^\infty \int_{A_i(z,\lambda)} \frac{1}{|\hat{H}_i(f)|^2} df dz = \bar{P}_i \tag{27}$$

$i = 1, \ldots, M.$

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:

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

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 $\sigma^2 + z$, user i transmits if it yields the maximum increase in the objective function $\mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P}$, which is the case if $f \in A_i(z,\lambda)$.

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 [14]. 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 [6]:

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

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 relatively 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 [6]. Consider a typical mobile scenario where the vehicle is moving at 60 km/h and the center frequency of the transmission bandwidth 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 0.018s. Typical delay spread between paths range from 10^{-7} to 1.5×10^{-5} s [17]. 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 2.2, 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:

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

where

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

and $\hat{\mathbf{H}}(f,\omega)=(\hat{H}_i(f,\omega),\ldots,\hat{H}_M(f,\omega))$

Using the techniques of Section 3, 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*

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

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

$$R_i^*(\mu)=\int_0^\infty\frac{1}{2(\sigma^2+z)}\left\{\int_{\frac{-W}{2}}^{\frac{W}{2}}\int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^{\infty}\prod_{k\neq i}F_k\left(f,\frac{1}{\frac{\lambda_i}{\lambda_k h}+\frac{\mu_k-\mu_i}{2\lambda_k(\sigma^2+z)}}\right)f_i(f,h)dhdf\right\}dz\quad(28)$$

where the vector λ satisfies the equations:

$$\int_0^\infty\left\{\int_{\frac{-W}{2}}^{\frac{W}{2}}\int_{\frac{2\lambda_i(\sigma^2+z)}{\mu_i}}^{\infty}\frac{1}{h}\prod_{k\neq i}F_k\left(f,\frac{1}{\frac{\lambda_i}{\lambda_k h}+\frac{\mu_k-\mu_i}{2\lambda_k(\sigma^2+z)}}\right)f_i(f,h)dhdf\right\}dz=\bar{P}_i\qquad(29)$$

$i=1,\ldots,M$.

8 Conclusions

In this paper, we have characterized the throughput capacity region of the multi-access fading channel under optimal power allocation. Just as the solution to the corresponding single-user channel has the water-filling interpretation, our solution can be viewed as a generalization of the water-filling power allocation to the multi-user setting. The solution consists of several steps. First, we use Lagrangian techniques to show that each point on the boundary of the throughput capacity region can be obtained by solving a family of optimization problems over a set of parallel Gaussian multi-access channels, one for each fading state. Second, we exploit the polymatroidal structure of the multi-access Gaussian capacity region to provide a simple greedy solution to each of those optimization problems, despite the fact that there are an exponentially large number of constraints. Third, we show that the Lagrange multipliers associated with the power constraints ("power prices") can be computed by simple iterative procedures. Taken together, these results provide effective algorithms 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.

This problem formulation 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 polymatroidal structure will again help us in characterizing the delay-limited capacity region of the fading channel.

Acknowledgments

This work was initiated when both authors were postdoctoral members of technical staff at the Mathematical Sciences Research Center of A.T. & T. Bell Laboratories. The hospitality of our department head Dr. Debasis Mitra is very much appreciated. Hanly's contribution was completed with the help of an A.R.C. grant, which is gratefully acknowledged.

References

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

[2] Cheng R, Verdu S (1993) Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling *IEEE Trans. on Information Theory, Vol. 39, No. 3*

[3] Edmonds, J. (1970) Submodular functions, matroids and certain polyhedra. In Proc. of Int. Conf. on Combinatorics, Gordon & Beach, 69-87.

[4] Federgruen, A. and H. Groenevelt (1986). The Greedy Procedure for Resource Allocation Problems:Necessary and Sufficient Conditions for Optimality. Operations Research 34, 909-918.

[5] Gallager, R. G. (1968) *Information Theory and Reliable Communication.* Wiley.

[6] Gallager R. (1994) An inequality on the capacity region of multiaccess fading channels. in *Communications and Cryptography - Two sides of one tapestry*, Kluwer, pp. 129-139.

[7] Goldsmith A. (1994) *Design and performance of high-speed communication systems over time-varying radio channels*, Ph. D. thesis, University of California, Berkeley.

[8] Hanly S.V. (1993) Information Capacity of Radio Networks *PhD thesis, University of Cambridge* August.

[9] Hanly S.V., Whiting P.A. (1994) Constraints on capacity in a multi-user channel *International symposium on information theory* Trondheim, Norway.

[10] Hanly S. V. (1995) An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity *IEEE Journal on Selected Areas, special issue on the fundamentals of networking, Vol. 13, No. 7* September.

[11] Hanly S.V. and D. Tse (1995), Multi-Access Fading Channels: Shannon and Delay-Limited Capacities *Proceedings of the 33rd Allerton Conference, Monticello*, IL, Oct.

[12] Liao H. (1972) *Multiple Access Channels* PhD Thesis, Department of Electrical Engineering,University of Hawaii, Honlulu.

[13] R. Knopp and P.A. Humblet (1995) Information capacity and power control in single-cell multiuser communications. *Int. Conf. on Communications*, Seattle, Wash., June.

[14] Medard M. (1995) *The capacity of time varying multiple user channels in wireless communications* PhD Thesis, M.I.T.

[15] Mitra D. (1993) An asynchronous distributed algorithm for power control in cellular radio systems *Fourth winlab workshop on third generation wireless information networks*:249–257

[16] Shamai S, Wyner A.D. (1995) Information Theoretic Considerations for symmetric, cellular multiple access communications channels, Part I *submitted for publication*

[17] Turin, G.L. et al. (1972) A statistical model for urban radio propagation. *IEEE Trans. Vehicular Tech.* vol 21, pp. 1-9, Feb.

[18] Yates R. (1995) A framework for uplink power control in cellular radio systems *IEEE Journal on Selected Areas, special issue on the fundamentals of networking, Vol 13, No. 7*

Appendices

A Proof of Theorem 2.2

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

Conversely, suppose rate $\mathbf{R}$ is achievable. By this we mean that there exists a sequence of codes, indexed by N, with code $\mathcal{C}_N$ of block length N, and with probability of error $\epsilon_N \to 0$. For code $\mathcal{C}_N$, we index the messages of user i by $\{1, 2, \dots, 2^{R_i N}\}$ 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 $\mathbf{X_i}$ for $i = 1, 2, \dots, M$. Note that the codewords can be chosen as a function of the states of the channel.

Let $f(\mathbf{h})$ be the equilibrium probability density of being in fading state $\mathbf{h}$. Without loss of generality, assume that the fading of all users is bounded by 1. For each k, let $I_k = \{0, \frac{1}{k}, \frac{2}{k}, \dots, 1\}^M$ 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 $[1, \dots, N]$ at which times the fading state $\mathbf{H}$ lies in E. Let $Q(N)$ be uniformly distributed on $[1, \dots, N]$. Define

$$V_i^k(E, N) \equiv \mathbb{E}[X_i^2(Q(N)) | Q(N) \in S(E)]$$

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

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

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,

$$\sum_{E \in I_k} V_i^k(E) f(E) \le \bar{P}_i \tag{30}$$

We define $\mathbf{h}(E)$ to be the upper corner of E. Let $\mathbf{H}(n)$ be the fading at time n and define a new value $\hat{\mathbf{H}}(n)$ by $\hat{\mathbf{H}}(n) \equiv \mathbf{h}(E)$ if $\mathbf{H}(n) \in E$. 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 $S \subseteq \{1, 2, \dots, M\}$

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

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

$$\frac{1}{N} I[(X_i)_{i \in S}; Y | (X_i)_{i \in S^c}, \mathbf{H}]$$

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

Taking limits along the convergent subsequence, we obtain

$$
\mathbf{R}(S) \leq \sum_{E\in I_k} f(E)\frac{1}{2}\log(1+\frac{\sum_{i\in S} h_i(E)V_i^k(E)}{\sigma^2}) \tag{31}
$$

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

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

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.

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

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

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:

$$\forall S, \forall \mathbf{h}, \ \sum_{i \in S} \mathcal{R}_i(\mathbf{h}) \leq \frac{1}{2} \log \left(1 + \frac{\sum_{i \in S} \mathcal{P}_i(\mathbf{h}) h_i}{\sigma^2}\right) \tag{32}$$

For any $\mathbf{h}$, consider all subsets S for which there is equality in (32). If there is a user i that is not in any such subset, then $R_i(\mathbf{h})$ can be increased without violating any constraint. But this contradicts the fact that this rate allocation achieves a boundary point of the capacity region $\mathcal{C}(\bar{\mathbf{P}})$. Therefore, every user must be almost surely in a tight constraint, and hence, by the strict concavity of log, $\mathcal{P}^{(1)}(\mathbf{H}) = \mathcal{P}^{(2)}(\mathbf{H})$ almost surely.

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

$$\max_{(\mathbf{R},\mathbf{P})} \mu \cdot \mathbf{R} - \sum_i \frac{\lambda_i}{h_i} \mathbf{P} \tag{33}$$

for every fading state $\mathbf{h}$, for some λ. The only possibility for non-uniqueness of $\mathcal{R}(\mathbf{h})$ 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, $\mathcal{P}_i(\mathbf{h}) = 0$ or $\mathcal{P}_j(\mathbf{h}) = 0$. Together with the fact that the power allocation is unique, we can conclude that there is a also a unique rate allocation.

Now we show that the Lagrangian power prices λ for maximizing $\mu \cdot \mathbf{R}$ subject to average power constraints must also be unique. Without loss of generality, assume that $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_M$. Let $\mathcal{P}$ be the unique optimal power allocation policy; it can be obtained by maximizing $\mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P}$ subject to $\mathbf{R} \in \mathcal{C}_g(\mathbf{h}, \mathbf{P})$ for each fading state, for some choice of λ. We want to show that such a λ must also be unique. We show by induction on k that λ_k must be uniquely specified. Let $\mathbf{h}$ be a fading state for which $\mathcal{P}_1(\mathbf{h}) > 0$; 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.

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

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

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

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

$$\frac{\mu_{k+1}}{\sigma^2 + \sum_{j \geq k+1} \mathcal{P}_j(\mathbf{h})} - \frac{\lambda_{k+1}}{h_{k+1}} = \max_{i \leq k} \frac{\mu_i}{\sigma^2 + \sum_{j \geq k+1} \mathcal{P}_j(\mathbf{h})} - \frac{\lambda_i}{h_i}$$

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

C Proof of Lemma 4.4

(i). Without loss of generality, we assume $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_M$ so that the decoding order is $M, M-1, \ldots, 1$. Let $\epsilon > 0$ be arbitrary. We define a sequence of power prices $\lambda^{(1)}, \lambda^{(2)}, \ldots \lambda^{(M)}$ and from these construct another vector of prices λ. We shall show that λ satisfies the conditions of the lemma. For $m < M$, we take $\lambda^{(m)}$ to be power prices of a fictitious channel in which only users $1, 2, \ldots, m$ 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 $\lambda^{(1)} = 0$, 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 $P_1(\lambda^{(1)})$ is infinite. With $\lambda^{(2)} = (\epsilon_1, 0)$, we have a two user channel, with $\mu = (\mu_1, \mu_2)$ and $\lambda = (\epsilon_1, 0)$. It is clear that here $P_1(\lambda^{(2)}) < \infty$ and $P_2(\lambda^{(2)}) = \infty$. Note that by taking ϵ_1 small we can ensure that $P_1(\lambda^{(2)}) > \bar{P}_1$ and $\epsilon_1 < \epsilon$. This becomes the inductive hypothesis: suppose that with $\lambda^{(m)} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_{m-1}, 0)$ we have $P_i(\lambda^{(m)}) > \bar{P}_i, \;\; \epsilon_i < \epsilon$ for all $i = 1, 2, \ldots, m-1$. Then set $\lambda^{(m+1)} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_m, 0)$, and note that for any ϵ_m this gives a new channel with $m+1$ users. Provided $\epsilon_m > 0$, we must have that

$$P_i(\lambda^{(m+1)}) \geq P_i(\lambda^{(m)}) > \bar{P}_i \quad i = 1, 2, \ldots, m-1.$$

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

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 $i \neq i_k$: (1) if $y_i^{(k)} = a_i$, then $I_i(\mathbf{y}^{(\mathbf{k})}) \geq I_{i_k}(\mathbf{y}^{(\mathbf{k})})$; (2) if $0 < y_i^{(k)} < a_i$, then $I_i(\mathbf{y}^{(\mathbf{k})}) = I_{i_k}(\mathbf{y}^{(\mathbf{k})})$ and $i < i_k$; (3) if $y_i^{(k)} = 0$, then $I_i(\mathbf{y}^{(\mathbf{k})}) \leq I_{i_k}(\mathbf{y}^{(\mathbf{k})})$.

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 $i \neq i_m$, (1) if $y_i^{(m+1)} = a_i$, then by the inductive hypothesis, $I_i(\mathbf{y}^m) \geq I_{i_m}(\mathbf{y}^m)$ and by Fact 1, $I_{i_m}(\mathbf{y}^m) \geq I_{i_m}(\mathbf{y}^{(m+1)})$, so that we have $I_i(\mathbf{y}^{(m+1)}) \geq I_{i_m}(\mathbf{y}^{(m+1)})$; (2) if $0 < y_i^{(m+1)} < a_i$, then by the inductive hypothesis, $I_i(\mathbf{y}^m) = I_{i_m}(\mathbf{y}^m)$ and $i < i_m$, so that together with Fact 2, this implies that $I_i(\mathbf{y}^{(m+1)}) = I_{i_m}(\mathbf{y}^{(m+1)})$; (3) if $y_i^{(m+1)} = 0$, then by the inductive hypothesis and the definition of the algorithm, $I_i(\mathbf{y}^{(m+1)}) \leq I_{i_m}(\mathbf{y}^{(m+1)})$.

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

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

ii) $I_j(\mathbf{y}^{(m+1)}) = I_{i_m}(\mathbf{y}^{(m+1)})$ for some $j > i_m$. In this case, $I_{i_{m+1}}(\mathbf{y}^{(m+1)}) = I_{i_m}(\mathbf{y}^{(m+1)})$ 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 $I_{i_{m+1}}(\mathbf{y}^{(m+1)}) = I_{i_m}(\mathbf{y}^{(m+1)})$ 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 $I_{i_{m+1}}(\mathbf{y}^{(m+1)}) \geq I_i(\mathbf{y}^{(m+1)})$ 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 $\mathbf{y}^*$ satisfies

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

In case iii), $\mathbf{y}^*$ satisfies $y_i^* = a_i$ and $I_i(\mathbf{y}^*) \geq 0$ for all i. Thus, in either case, $\mathbf{y}^*$ 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.

Multi-access Fading Channels: Part II: Delay-Limited Capacities*

Stephen V. Hanly[†] and David N. Tse[‡]

Abstract

In multi-access 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 multi-access 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 have analyzed the *throughput capacity region* which characterizes the long-term achievable rates through the time-varying channel. However, the *delay* experienced depends on how fast the channel varies. In the present paper, part II, we introduce a notion of *delay-limited capacity* which is the maximum rate achievable with delay *independent* of how slow the fading is. We characterize the *delay-limited capacity region* of the multi-access fading channel and the associated optimal resource allocation schemes. We show that successive decoding is optimal, and the optimal decoding order and power allocation can be found explicitly as a function of the fading states; this is a consequence of an underlying polymatroidal structure that we exploit.

1 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 dBs. The problem is particularly acute for real-time traffic such as video, since they have a stringent delay requirement. 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. In part I

*Part of this work was done when both authors were with A.T. & T. Bell Laboratories. This paper has been submitted to the IEEE Transactions on Information Theory.

[†]Dept. of Electrical Eng., University of Melbourne, s.hanly@ee.mu.oz.au

[‡]Dept. of EECS, U.C. Berkeley, dtse@eecs.berkeley.edu

of this paper, we studied the problem of optimal dynamic resource allocation from an information theoretic point of view. We computed the Shannon capacity region of the multi-access fading channel when the transmitters as well as the receiver have access to the channel state, and also characterized the optimal resource allocation schemes.

The Shannon capacity of a channel provides the ultimate limits on the bit rates that are achievable. The capacity itself is not dependent on any delay considerations, and is achievable in an asymptotic sense as delay ("block length") tends to infinity. Thus when we focused on the Shannon capacity of the multi-access fading channel in Part I, we found the Shannon limit over the set of all possible codes. In Part II, we now consider limitations imposed on the possible codes that can be used due to delay constraints.

One way to think about this in the single user case is to identify the Shannon capacity with a "long-term average of mutual information between the user and the receiver" in the channel. That is why we called the Shannon capacity of a fading channel its *throughput capacity* in Part I. On the other hand, there is a notion of "instantaneous mutual information" and this can fluctuate as a function of the fading state. Essentially, in the delay-limited case, we restrict the set of available codes to those for which, through power control, the instantaneous mutual information is kept constant at all times. Without such a restriction the throughput of the channel can be increased but at the expense of having the "instantaneous mutual information" fluctuating with the fading process, leading to delay at the time scale of the channel variations. The situation is more subtle in the multi-user context, but similar ideas go through.

There are many "delay sensitive" applications such as voice and video, for which long delays cannot be tolerated. Unless the fading is fast on the time-scale of tolerable delay, the throughput capacity of Part I is not relevant for these applications. Our delay limited capacity is the appropriate limit for these applications.

The notion of "delay limitedness" is implicit in many works. For example, papers on power control (see Gilhousen *et. al.* [6], Hanly [7], Yates [13]) assume that a desired C/I must be met for every fading state, and this means that the user's mutual information is kept constant in time. The formal notion of delay limited capacity for multi-access channels was defined in Hanly and Tse [8], where we considered the symmetric case with users having the same rate requirements. In the present paper, we focus on characterizing the entire delay-limited capacity region and the associated optimal power control schemes. As in Part I, we shall exploit the convex nature, and underlying polymatroidal structure of this problem. Again, we find that the optimal solution is always successive decoding, and that the optimal power control can be explicitly characterized and has a greedy interpretation.

Part II is organized as follows. In Section 2 we introduce the Gaussian, multi-access, flat fading model and present a coding theorem for the delay limited 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 transmit power based on the channel state. In Section 3, 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 multi-access Gaussian channels, one for each fading state. Given the Lagrange multipliers, which can be interpreted as power prices, the problem is that of finding the optimal "power" allocation as a function of each fading state so as to minimize the total power cost. Here, we exploit the polymatroidal structure of the optimization problem to obtain an explicit solution via a greedy algorithm. In Section 4, we turn to the problem of finding an appropriate set of power prices so that a target delay-limited rate vector can be met within given power constraints. We present an iterative algorithm which, if the target rates are achievable, is guaranteed to converge to the right power prices. Moreover, it also solves a *call admissions* problem by determining if a given set of target rates are indeed achievable.

In the remainder of the paper, we will extend the basic results in several directions. In Section 5, we will present greedy power allocation algorithms when additional power constraints are imposed. These results exploit further properties of polymatroids. In Section 6, we relax the delay limited requirement in two ways. First, we consider a multiple time-scale model, with slow and fast fading, and compute the optimal power control when we are delay limited with respect to the slow fading. Secondly, we consider a frequency selective fading channel, in which rates can be allocated to the different frequencies, but the sum rate over all frequencies must be constant for each fading state. Finally, in Section 7, we explore the implications of these information theoretic results to systems with sub-optimal coding and decoding.

In Hanly and Tse [8], the concept of delay limited capacity is extended to take advantage of statistical multiplexing: it is not always necessary for power control to be used to ensure that "sufficient mutual information is available at every time instant"; this can also be a property of the averaging of the independent fading of a large number of users, even if no power control, or only decentralized power control is employed. In the present paper, however, we allow centralized power control and so do not consider statistical multiplexing of fading.

Our results also provide a link between information theory and the theory of networking. Clearly, the power prices (and in Part I, bit rate rewards) have the potential to be tuned by the network in order to provide control over the radio resources. This is indeed our approach in Section 4, in which a call admission problem is solved by the adaptation of power prices, using an algorithm reminiscent of max-min fair bandwidth allocation algorithms in data networks. In Part I, we employed similar iterative algorithms to control real-time radio resource allocation (see Section 4 in Part I). More generally, there is an economic flavor to our results, as touched on in Section 3, and more directly in Part I, Section 3.4.

2 Delay limited Capacity

As in part I, 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 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 transmit power constraint of $\bar{P}_i$. We shall call $\mathbf{H}(n) = (H_1(n), H_2(n), \ldots, H_M(n))$ the joint fading process.

Suppose each source i codes over a block length of T symbols, where T is the delay, using a codebook $\mathcal{C}_i$ of size $2^{R_i T}$ (i.e. at rate R_i bits per channel use). Each codeword $\mathbf{x}$ of the ith user satisfies $\|\mathbf{x}\|_2^2 \leq T\bar{P}_i$. Fix a decoding scheme and assume the messages are chosen with equal probability. Let $p_e(T)$ be the probability of the event that any user is decoded incorrectly. The following is the definition of the throughput capacity region when both the transmitters and the receiver have access to the channel states. Characterizing this region was our focus in part I.

Definition 2.1 *The rate-tuple $\mathbf{R} = (R_1, \ldots, R_M)$ lies in the interior of the throughput capacity region $C(\bar{\mathbf{P}})$ if and only if for every $\epsilon > 0$, there exists a delay T, codebooks and a decoding scheme such that the probability of error $p_e(T)$ is less than ϵ. Moreover, the codewords can be chosen as a function of the realization of the fading processes.*

The notion of throughput capacity defined above is a natural extension of that for time-invariant Gaussian channels, where rates are achieved with arbitrarily long coding delays. However, there is a subtle but important difference between time-varying and time-invariant Gaussian channels. In the time-invariant Gaussian channel, the delay is needed to average out the Gaussian noise to get small error probabilities, and this is typically quite short. Thus the capacity is not only an upper bound to the achievable performance; it is a useful upper bound in the sense that it is possible to achieve rates close to capacity with acceptable delay, even for real-time traffic. In typical time-varying wireless channels, on the other hand, the fading process is a complex superposition of different effects some of which can be quite slow. Thus the delay required to average out such fading effects may be much longer than the acceptable delay.

To this end, we define a second notion of capacity region for time-varying multi-access channels. Let $\mathcal{H}$ be the set of all possible joint fading states of the users, $\mathcal{Q}$ be a given distribution on $\mathcal{H}$, and $\mathcal{A}(\mathcal{Q})$ be the set of all stationary, ergodic fading processes with stationary distribution $\mathcal{Q}$. We observe from Theorem 2.2 in Part I that the throughput capacity region of the multi-access fading channel depends only on the stationary distribution of the joint fading processes and not on the correlation structure. The following definition of the *delay-limited* capacity region also has this characteristic.

Definition 2.2 *A rate vector $(R_1, \ldots, R_M)$ lies in the interior of the* delay limited *capacity region $C_d(\bar{\mathbf{P}})$, if for every $\epsilon > 0$ there exists a coding delay T such that for every fading process in $\mathcal{A}(\mathcal{Q})$ there exists codebooks and a decoding scheme with $p_e(T) < \epsilon$. Moreover, the codewords can be chosen as a function of the realization of the fading processes.*

Contrast this with Defn. 2.1, where the coding delay can be chosen depending on the specific fading process, the coding delay here has to work *uniformly* for all fading processes with a given stationary distribution. Hence, rates in the delay-limited capacity region can be achieved with delays *independent* of the correlation structure of the fading. Thus the rates in the delay-limited capacity region are essentially those that can be achieved by coding that averages out the white noise but does not average over the fading process. It is an appropriate limit on the performance for traffic with stringent delay requirements and when the fading processes changes relatively slowly (due to users at walking speed for example.) It should also be noted that the throughput capacity region contains the delay-limited capacity region. The notion of delay-limited capacity for multi-access fading channels was first introduced in [8].

In Definition 2.2, we only require that there be a codebook for every realization of every fading process. However, the proof of Theorem 2.3 below shows that we can provide a single codebook of unit power that we scale by the power control policy identified in the theorem. This codebook will work no matter what fading process is chosen (*i.e.* for any correlation structure). By "power control policy", we mean the following:

A *power control policy* $\mathcal{P} : \Re^M \to \Re^M$ is a mapping such that given a joint fading state $\mathbf{h} = (\mathbf{h}_1, \ldots \mathbf{h}_M)$ for the users, $\mathcal{P}_i(\mathbf{h})$ can be interpreted as the transmitter power allocated to user i. Given power control policy $\mathcal{P}$, $\mathbb{E}_{\mathbf{H}}[\mathcal{P}_i(\mathbf{H})]$ is the average power usage for user i.

The following theorem provides a characterization of the delay-limited capacity region for the case when all the transmitters and the receiver know the current state of the channel.

Theorem 2.3 *Assume that the set of possible fading states $\mathcal{H}$ is bounded. The delay-limited capacity region $C_d(\bar{P})$ is given by*

$$C_d(\bar{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\mathbf{h} \in \mathcal{H}} C_g(\mathbf{h}, \mathcal{P}(\mathbf{h})) \tag{2}$$

where $\mathcal{F}$ is the set of all feasible power control policies satisfying the average power constraints, and $C_g(\mathbf{h}, \mathbf{P})$ is the capacity region of the time-invariant Gaussian multi-access channel, given by [1]*:*

$$C_g(\mathbf{h}, \mathbf{P}) = \left\{ \mathbf{R} : \mathbf{R}(S) \leq \frac{1}{2} \log \left(1 + \frac{\sum_{i \in S} h_i^2 P_i}{\sigma^2} \right) \quad \textit{for every } S \subset \{1, \ldots, M\} \right\} \tag{3}$$

[1]Here, as in Part I, for any vector $\mathbf{x}$ and any subset S we use the notation $\mathbf{x}(S)$ to denote $\sum_{i \in S} x_i$.

Proof. See appendix A. □

The intuitive content of the above theorem is that a rate vector $(R_1, \ldots, R_M)$ is achievable in the delay-limited sense if one can choose a feasible power control policy to coordinate the powers of the users such that sufficient mutual information is maintained between the transmitters and the receiver at *all* fading states. Note that this is essentially the information-theoretic version of the objective of standard power control algorithms in which power is allocated to satisfy the signal-to-interference requirements of all the users. Contrast this with the characterization, in Theorem 2.2 of Part I, for the throughput capacity region, where a rate vector $(R_1, \ldots, R_M)$ is achievable as long as there is a feasible power control policy to provide sufficient *long-term* average mutual information, averaged over all fading states. The "instantaneous" mutual information at each fading state, however, fluctuates.

3 Characterization of the Delay-Limited Capacity Region

In this section, we will characterize the optimal power control to achieve points on the boundary of the delay-limited capacity region $C_d(\bar{\mathbf{P}})$. We shall show that successive decoding is always optimal and we shall provide greedy algorithms for obtaining the optimal power control. Using this characterization, we will also provide a necessary and sufficient condition for $\mathbf{R}$ to be inside the capacity region.

3.1 Lagrangian Characterization and Optimal Power Allocation

We first define the boundary surface of $C_d(\bar{\mathbf{P}})$, which is essentially the set of optimal operating points on the capacity region.

Definition 3.1 *The boundary surface of $C_d(\mathbf{P})$ is the set of those rates such that we cannot increase one component, and remain in $C_d(\mathbf{P})$ without decreasing another.*

The following lemma gives a Lagrangian characterization of the capacity region.

Lemma 3.2 *1) A rate vector $\mathbf{R}^*$ lies in $C_d(\bar{\mathbf{P}})$ if and only if there exists a nonnegative $\lambda \in \Re^M$ and a power control policy $\mathcal{P}$ such that for every joint fading state $\mathbf{h}$, $\mathcal{P}(\mathbf{h})$ is a solution to the optimization problem:*

$$\min_{\mathbf{P}} \lambda \cdot \mathbf{P} \qquad \textit{subject to} \qquad \mathbf{R}^* \in C_g(\mathbf{h}, \mathbf{P}) \tag{4}$$

and

$$\mathbb{E}_{\mathbf{H}}[\mathcal{P}_i(\mathbf{H})] \leq \bar{P}_i \qquad i = 1, \ldots, M$$

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 $\mathbf{R}^*$.

2) A rate vector $\mathbf{R}^*$ lies on the boundary surface if and only if there exist λ as above but with all the average power constraints holding with equality.

Analogous to Lemma 3.10 of Part I, this lemma reduces the computation of the optimal power control to a family of optimization problems over a set of parallel time-invariant Gaussian channels. As in the analysis of the throughput capacity region, the vector λ can be interpreted as a set of *power prices* reflecting the power constraints. The important difference is that in this case, we require that the rate vector $\mathbf{R}^*$ be in the Gaussian capacity region $\mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h}))$ for *all* fading states $\mathbf{h}$. This is consistent with the nature of delay-limited capacities.

Proof. Since any rate vector inside the capacity region is dominated by some point on the boundary surface, statement (2) would imply statement (1). Hence, we will focus on proving statement (2).

First note that since the capacity region $\mathcal{C}_d(\bar{\mathbf{P}})$ is convex, a point $\mathbf{R}^*$ is on the boundary surface of the region if and only if it is a solution to the optimization problem:

$$\max_{\mathbf{R}} \mu \cdot \mathbf{R} \qquad \text{subject to } \mathbf{R} \in \mathcal{C}_d(\bar{\mathbf{P}}) \tag{5}$$

for some nonnegative vector μ. Now consider the set

$$S \equiv \{(\mathbf{R}, \mathbf{P}) : \mathbf{R} \in \mathcal{C}_d(\bar{\mathbf{P}})\}$$

By the concavity of the log function, it can readily be verified that S is a convex set. Thus, $\mathbf{R}^*$ solves (5) if and only if there exist nonnegative Lagrange multipliers λ such that $(\mathbf{R}^*, \bar{\mathbf{P}})$ is a solution to the problem:

$$\max_{(\mathbf{R},\mathbf{P})\in S} \mu \cdot \mathbf{R} - \lambda \cdot \mathbf{P}$$

Hence, $\mathbf{R}^*$ is on the boundary surface of $\mathcal{C}_d(\bar{\mathbf{P}})$ if and only if $\bar{\mathbf{P}}$ is a solution to the problem:

$$\min_{\mathbf{P}} \lambda \cdot \mathbf{P} \qquad \text{subject to} \qquad \mathbf{R}^* \in \mathcal{C}_d(\mathbf{P})$$

i.e. if and only if there exists a power control policy $\mathcal{P}^*$ which solves

$$\min_{\mathcal{P}} \lambda \cdot \mathbb{E}_{\mathbf{H}}[\mathcal{P}(\mathbf{H})] \qquad \text{subject to} \qquad \mathbf{R}^* \in \cap_h \mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h}))$$

and

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

We note that this last optimization problem is equivalent to solving (4) for every fading state $\mathbf{h}$. This completes the proof. □

The vector μ can be interpreted as the *rate rewards* and λ as the *power prices*. Thus, a point on the boundary of the capacity region is achieved by maximizing the total revenue for a given rate reward vector μ. Appropriate power prices have to be chosen such that the average power constraints are satisfied.

The computation of the optimal power control is now reduced to solving the optimization problem (4). This is a linear program but one with an exponentially large number of constraints (in M). However, as in part I, we exploit the polymatroid structure of the problem to provide a simple greedy solution to this problem. We first recall the following definition of *contra-polymatroids* and a greedy optimization procedure.

Definition 3.3 *Let $E = \{1, \ldots, M\}$ and $f : 2^E \to \Re_+$ be a set function. The polyhedron*

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

is a contra-polymatroid *if f is a rank function, i.e. 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)

For π a permutation on the set E, define the vector $\mathbf{v}(\pi) \in \Re^M$ by $v_{\pi(1)}(\pi) = f(\pi(1))$ and $v_{\pi(i)}(\pi) = f(\{\pi(1), \ldots, \pi(i)\}) - f(\{\pi(1), \ldots, \pi(i-1)\})$ for $i = 2, \ldots, M$.

Lemma 3.4 *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 λ is a given vector in $\Re_+^M$, then the solution of the optimization problem*

$$\min \lambda \cdot \mathbf{x} \qquad \textit{subject to} \qquad \mathbf{x} \in \mathcal{G}(f) \tag{6}$$

is attained at the point $v(\pi^*)$ where the permutation π^* is given by $\lambda_{\pi^*(1)} \geq \cdots \geq \lambda_{\pi^*(M)}$.

It is straightforward to verify (Corollary 3.13 of Part I) that for a given rate vector $\mathbf{R}^*$ and fading state $\mathbf{h}$, the set of *received powers* that can support $\mathbf{R}^*$,

$$\mathcal{G}(\mathbf{R}^*) \equiv \{\mathbf{Q} : Q_i = h_i \mathcal{P}_i, \mathbf{R}^* \in \mathcal{C}_g(\mathbf{h}, \mathcal{P})\}$$

is a contra-polymatroid with rank function

$$f(S) = \exp(2\mathbf{R}^*(S)) - 1$$

Applying Lemma 3.4, the optimization problem (4) can be readily solved:

$$P^*_{\pi(i)} = \begin{cases} \frac{\sigma^2}{h_{\pi(1)}}[\exp(2R^*_{\pi(1)}) - 1] & \text{if } i = 1 \\ \frac{\sigma^2}{h_{\pi(i)}}[\exp(2\sum_{k=1}^{i} R^*_{\pi(k)}) - \exp(2\sum_{k=1}^{i-1} R^*_{\pi(k)})] & i = 2, \ldots, M \end{cases} \tag{7}$$

where the permutation π satisfy:

$$\frac{\lambda_{\pi(1)}}{h_{\pi(1)}} \geq \ldots \geq \frac{\lambda_{\pi(M)}}{h_{\pi(M)}} \tag{8}$$

This optimal point corresponds to successive decoding in the order given by π, with power allocated to the users such that the target delay-limited rate vector $\mathbf{R}^*$ is achieved. One can think of the successive decoding order π as a way to give priority to different users in the scheduling of resources; a user decoded later in the ordering is given higher priority than a user decoded earlier. This is because users need less transmit power to support their target rates when they are decoded later. The scheduling rule here depends on both the power prices λ and the current fading state. In fact, this rule is analogous to the classic $c-\mu$ rule in scheduling theory (see eg. [12]), as both arise from the polymatroidal structure of the problem. The additional feature in our problem is that the scheduling priority is a dynamic function of the fading state. Another interesting aspect of the solution to the optimization problem (4) is that the solution depends on the power prices λ only via the decoding order. This will simplify our later analysis.

Note that when the power price vector λ is strictly positive, then with probability 1 the ordering is uniquely defined since the fading processes have a continuous stationary distribution. Thus, with probability 1, the solution to the optimization problem (4) is unique. Let us then define $\mathbf{P}(\mathbf{R}^*, \lambda)$ to be the unique average power vector corresponding to the almost surely unique power control policy which solves the optimization problem (4).

In the common case when the fading processes of the users are independent of each other, the average power vector $\mathbf{P}(\mathbf{R}^*, \lambda)$ has a simple form:

$$P_i(\mathbf{R}^*, \lambda) = (\exp(2R_i^*)-1) \int_0^\infty \frac{\sigma^2}{h_i} \prod_{k\neq i} \left\{ P(h_k > \frac{\lambda_k}{\lambda_i} h_i) + P(h_k \leq \frac{\lambda_k}{\lambda_i} h_i) \exp(2R_k^*) \right\} f_i(h_i) dh_i \tag{9}$$

This expression can be obtained by noting that the power allocated to user i depends only on which users have values $\frac{\lambda_k}{h_k}$ greater than that of user i. Note that due to the special structure of the optimal power control policy, the computation of the average power has been reduced from a M-dimensional integral to a 1-dimensional integral.

Combining this with Lemma 3.2, we have the following characterization of the delay-limited capacity region:

Theorem 3.5 *Assume the fading processes of users are independent of each other. Then the rate vector* $\mathbf{R}$ *lies in the delay-limited capacity region* $C_d(\bar{\mathbf{P}})$ *if and only if there exists* $\lambda \in \Re_+^M$ *such that*

$$(\exp(2R_i)-1) \int_0^\infty \frac{\sigma^2}{h_i} \prod_{k\neq i} \left\{ 1 + F_k\left(\frac{\lambda_k}{\lambda_i} h_i\right) (\exp(2R_k) - 1) \right\} f_i(h_i) dh_i \leq \bar{P}_i \qquad i = 1, \ldots, M \tag{10}$$

The power allocation policy that achieves this rate $\mathbf{R}$ *is given by eqn. (7). Moreover,* $\mathbf{R}$ *lies on the boundary surface if and only if there exist* λ *such that (10) holds with equality.*

We can also consider a set $\mathcal{D}_d(\mathbf{R}^*)$: this is the set of average power vectors that can support target delay-limited rates $\mathbf{R}^*$, i.e.

$$\mathcal{D}_d(\mathbf{R}^*) \equiv \{\mathbf{P} : \mathbf{R}^* \in \mathcal{C}_d(\mathbf{P})\}$$

Note that $\mathcal{D}_d(\mathbf{R}^*)$ is the structure in the power space that plays the same role as the capacity region $\mathcal{C}_d(\bar{\mathbf{P}})$ in the rate space. The above results lead to an explicit characterization of the boundary surface of $\mathcal{D}_d(\mathbf{R})$, parameterized by λ.

Theorem 3.6 *Assume the fading processes of users are independent of each other. Then the following equation gives an explicit parameterization of the boundary surface of the region* $\mathcal{D}_d(\mathbf{R}^*)$ *by* $\lambda \in \Re_+^M$.

$$P_i(\lambda) = (\exp(2R_i^*)-1) \int_0^\infty \frac{\sigma^2}{h_i} \prod_{k \neq i} \left\{1 + F_k\left(\frac{\lambda_k}{\lambda_i} h_i\right)(\exp(2R_k^*) - 1)\right\} f_i(h_i)dh_i \qquad i = 1, \ldots, M \tag{11}$$

The above results still leave open important questions: 1) how to check algorithmically if a target rate vector $\mathbf{R}$ is achievable, i.e. in the capacity region $\mathcal{C}_d(\bar{\mathbf{P}})$, and 2) how to find the appropriate power prices λ if $\mathbf{R}$ is indeed in the region. We will return to these questions in Section 4. But first, let us look at some special cases of Theorem 3.5.

3.2 Examples

1)**Single-User Channel:** When $M = 1$, the delay-limited capacity $C_d(\bar{P})$ is given by:

$$C_d(\bar{P}) = \frac{1}{2}\log(1 + \frac{\bar{P}}{\sigma^2 \int_0^\infty \frac{f(h)}{h} dh}) \tag{12}$$

We note that for some fading distribution, the delay-limited capacity may be zero. For example, for Rayleigh fading,

$$f(h) = \frac{1}{a}\exp(-\frac{h}{a})$$

and $\int_0^\infty \frac{f(h)}{h} dh = \infty$, so $C_d = 0$. The problem is that the channel is spending a lot of time close to zero. One approach to deal with this is to allow an event of outage when the channel gets too weak. (This is the approach taken by Ozarow et al [10] and Cheng [2] for situations where there is no power control.) Thus, even for these fading distributions, it is meaningful to consider the notion of delay-limited capacity during the times when the channel is reasonable, and declare an outage otherwise. For many other distributions, such as the log-normal distribution for shadow fading, a non-zero delay-limited capacity is obtained even without the need of allowing outage.

2)**Symmetrical Case [8]:** Consider the case when there are M users, the fading of users are identical and independent, and their power constraints are the same. The symmetric delay-limited capacity C_d is the maximum common rate that can be achieved, and can be obtained by putting $\lambda_i = 1$ for all i in eqns.(10). Simplifying, we find that the capacity satisfies:

$$\frac{\sigma^2}{2}[\exp(2C_d) - 1] \int_0^\infty [1 + F(h)(\exp(2C_d) - 1)]^{M-1} \frac{f(h)}{h} dh = \bar{P}$$

The optimal power control policy has an interesting form. Namely, users are decoded in the order of decreasing channel strengths, with the strongest user decoded first and the weakest user decoded last. Powers are allocated accordingly. If channel strength is determined primarily by the distance to the base-station, then this optimal decoding order results in the smallest possible transmit power for the furthest user to support the desired rate, as he only has to compete with the background noise and not the interference from any other user. This property is particularly appealing in terms of reducing inter-cell interference, as the furthest user will likely cause the most interference in an adjacent cell. Contrast this with the IS-95 CDMA scheme, in which the furthest user has to compete with all other users so that his received power has to be the same as that of the closest user.

3) **Two-User Capacity Region:** When $M = 2$, the boundary of the delay-limited capacity region can be directly calculated by solving the equations (10). Let $\lambda \equiv \frac{\lambda_1}{\lambda_2}$. Then the boundary is the following parametric curve as λ ranges from 0 to ∞:

$$R_1(\lambda) =$$
$$\frac{1}{2}\log\left[1 + \frac{B_2(\lambda)\bar{P}_1 - B_1(\lambda)\bar{P}_2 + A_1A_2 + \sqrt{(B_2(\lambda)\bar{P}_1 - B_1(\lambda)\bar{P}_2 + A_1A_2)^2 + 4A_1A_2B_2(\lambda)\bar{P}_1}}{2A_1B_2(\lambda)}\right]$$
$$R_2(\lambda) =$$
$$\frac{1}{2}\log\left[1 + \frac{B_1(\lambda)\bar{P}_2 - B_2(\lambda)\bar{P}_1 + A_1A_2 + \sqrt{(B_1(\lambda)\bar{P}_2 - B_2(\lambda)\bar{P}_1 + A_1A_2)^2 + 4A_1A_2B_1(\lambda)\bar{P}_2}}{2A_2B_1(\lambda)}\right]$$

where

$$\begin{aligned} A_m &\equiv \int_0^\infty \frac{\sigma^2}{h} f_m(h) dh \qquad m = 1,2 \\ B_1(\lambda) &\equiv \int_0^\infty \frac{\sigma^2}{h} F_2(\frac{h}{\lambda}) f_1(h) dh \\ B_2(\lambda) &\equiv \int_0^\infty \frac{\sigma^2}{h} F_1(\lambda h) f_2(h) dh \end{aligned}$$

The parameter λ can be viewed as a prioritization between the two users. As $\lambda \to 0$, $B_1(\lambda) \to A_1$ and $B_2(\lambda) \to 0$ so $R_2(\lambda) \to \frac{1}{2}\log(1 + \frac{\bar{P}_2}{A_2})$. This is the delay-limited capacity of user 2 when it is given strict priority over user 1 in all fading states (i.e. decoded last), and this is the best rate user 2 can get. Similarly, as $\lambda \to \infty$, $B_1(\lambda) \to 0$ and $B_2(\lambda) \to A_2$ so $R_1(\lambda) \to \frac{1}{2}\log(1+\frac{\bar{P}_1}{A_1})$. This is the delay-limited capacity of user 1 when it is given strict priority in all fading states, and this is the best rate user 1 can get. For λ in between these two extremes, the decoding order of users 1 and 2 changes depending on the fading state. See Fig. 1 for an illustration. Note that in this two-user case, we can parameterize the boundary surface of $\mathcal{C}_d(\bar{\mathbf{P}})$ by $\lambda \in \Re^2_+$. We will comment on whether this can be done in the general M-user case in Appendix B.

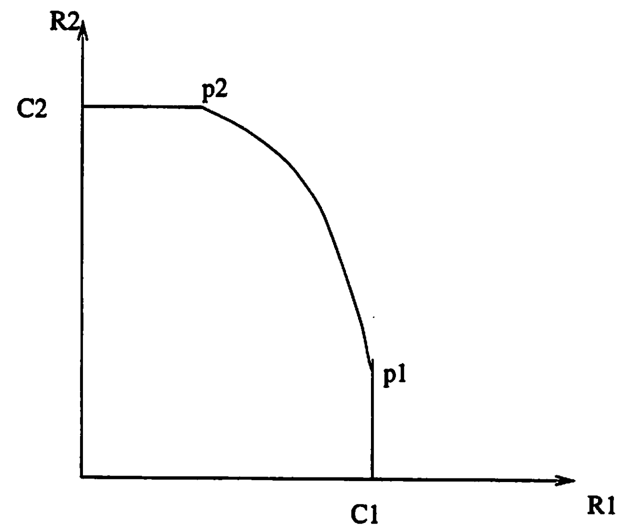

Figure 1: A two-user delay-limited capacity region. The curved part is the boundary surface. The points p_1 and p_2 are the two extreme points of the surface. The point p_1 corresponds to giving absolute priority to user 1, i.e. decoding user 1 after user 2 at every fading state. At this point, user 1 gets rate $C_1 = \frac{1}{2}\log(1+\frac{\bar{P}_1}{A_1})$. And vice versa for point p_2. Note that all other points in the capacity region but not on the curved boundary are dominated by some point on the boundary.

3.3 Extreme points of boundary surface

We now extend the characterization of the interior points of the boundary surface of $\mathcal{D}_d(\mathbf{R})$ to include the extreme points.

Suppose $\mathcal{L}$ is a set of subsets of $E \equiv \{1,2,\ldots,M\}$ 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 $F_2 \subseteq F_1$. This nesting property enables us to define a new decoding rule. Let us use successive decoding, with the ordering determined by λ as before, except now all users in any set $F \in \mathcal{L}$ are decoded before users in F^c, for every fading state $\mathbf{h}$. Thus if $F_1 \subseteq F_2 \ldots \subseteq E$, then $(\lambda_i)_{i\in F_1}$ is used to determine the ordering of users in F_1, and all these users provide interference to users in F_1^c. Inductively, $(\lambda_i)_{i\in F_n\setminus F_{n-1}}$ is used to determine the ordering of users in $F_n\setminus F_{n-1}$, and all the users in F_n provide interference to the users in F_n^c. It is not difficult to show that all extreme points of the boundary surface of $D_d(\mathbf{R}^*)$ are obtained in this way. For the two-user example in Fig. 1, the extreme points are p_1 and p_2.

Let us also extend the notion of $P_i(\mathbf{R}^*, \lambda)$ in the following way.

Definition 3.7 *Given $\mathbf{R}, \lambda \in \Re_+^M$ and $\mathcal{L}$, a set of nested subsets of E, we denote the power vector characterized by $(\mathbf{R}, \lambda, \mathcal{L})$ by $\mathbf{P}(\mathbf{R}, \lambda, \mathcal{L})$.*

Note that $\mathbf{P}(\mathbf{R}, \lambda)$ is not an extreme point of the boundary surface of $\mathcal{D}_d(\mathbf{R})$, but is still representable in this notation:

$$\mathbf{P}(\mathbf{R}, \lambda) \equiv \mathbf{P}(\mathbf{R}, \lambda, \{E\})$$

We shall have use for this extension in Section 4.

3.4 Further remarks concerning the coding theorem

We would like to remark on the decoding schemes to achieve points on the boundary of the delay-limited capacity region. Consider a channel in which the fading state $\mathbf{H}$ is fixed at level $\mathbf{h}$ for all time. It follows immediately from (7) that if users are allocated powers in $\mathcal{P}(\mathbf{h})$ then $\mathbf{R}^*$ is achievable by successive decoding. We conclude that if the fading is sufficiently slow that it does not change during the block length then the optimal solution is to do successive decoding with powers allocated as in (7). This separation of time-scales assumption may be quite reasonable if $\mathbf{H}(n)$ is a slow fading process in relation to the tolerable coding delay (*e.g.* shadow fading). If $\mathbf{H}(n)$ changes during the block length then the optimal power control is still given by (7): it is *as if* successive decoding is being employed as far as power control is concerned, and we shall say that the optimal solution is of "successive decoding type". If we try to do successive decoding, we face the problem that the optimal ordering of the users may change during the blocklength, if the fading changes. This situation does not arise in the non-delay limited case; successive decoding is optimal as shown in Part I. It may be possible to extend successive decoding techniques to deal with fading in the delay limited case (an open problem). In practice, it may be sufficient to update the successive decoding order at the start of each code period, and make an allowance for the fading that occurs within the block length. We would then sacrifice optimality for ease of decoding.

4 An Iterative Algorithm for Resource Allocation

In the previous section, we have characterized the structure of the optimal power allocation and used it for an implicit characterization of the delay-limited capacity region $C_d(\bar{\mathbf{P}})$. The power prices λ play a central role as a mechanism through which resource is allocated to the different users. To achieve a target delay-limited rate vector $\mathbf{R}^*$, we have shown that a simple optimal power control can be obtained, for a *given* power price vector λ. Since the power prices reflect the power constraints on the users, a natural question then is how an appropriate power price vector can be computed for given power constraints. More specifically, we will be concerned with the following problem:

- Is a target delay-limited rate vector $\mathbf{R}^*$ achievable under a given average power constraint $\bar{\mathbf{P}}$? If so, what is an appropriate power price vector?

In the case of independent fading processes, this problem is equivalent to checking if there exists λ such that inequalities (10) can be satisfied. From a networking point of view, a solution to this problem serves the dual purpose of *call admissions* and resource allocation. It determines if a set of users with specified rate requirements is supportable and if so it allocates an appropriate amount of resources via the selection of the power prices.

An equivalent formulation is the optimization problem:

$$\inf_{\lambda>0} \max_{1\leq i\leq M} \frac{P_i(\mathbf{R}^*, \lambda)}{\bar{P}_i} \tag{13}$$

where $P_i(\mathbf{R}^*, \lambda)$ is the average power of the ith user under the optimal power control which minimizes the total power cost $\lambda \cdot \mathbf{P}$ while achieving rates $\mathbf{R}^*$. (In the case of independent fading, $P_i(\mathbf{R}^*, \lambda)$ is given by the explicit expression (9).) By Lemma 3.2, the target rate vector $\mathbf{R}^*$ is achievable with power constraints $\bar{\mathbf{P}}$ if and only if the solution to (13) is no greater than 1. This optimization problem can also be interpreted as finding a solution for *fair* power requirements for the users, weighted by the power constraints of the users.

We will provide an iterative algorithm that solves the problem (13). If the infimum in (13) is achieved at a positive λ^*, the algorithm will converge to it. If this is not the case, then a solution satisfying (13) must be an extreme point of the boundary surface of $\mathcal{D}_d(\mathbf{R}^*)$ (the set of average power vectors that can support $\mathbf{R}^*$.) More generally, we can represent all points on the boundary surface of $\mathcal{D}_d(\mathbf{R}^*)$, including extreme points. by $\mathbf{P}(\mathbf{R}^*, \lambda, \mathcal{L})$, where $\mathcal{L}$ is a set of nested subsets of users giving absolute priority rules that hold irrespective of the fading state. This was discussed in Section 3.3. In general, our algorithm provides the parameters λ^* and $\mathcal{L}^*$, and provably converges to the point $\mathbf{P}^* \equiv \mathbf{P}(\mathbf{R}^*, \lambda^*, \mathcal{L}^*)$ such that $\mathbf{P}^*$ is an optimal solution to (13).

Firstly, it is necessary to develop some notation. Since we assume that $\mathbf{R}^*$ is fixed throughout this section, we shall simplify notation and set

$$\begin{aligned} \mathbf{P}(\lambda) &\equiv \mathbf{P}(\mathbf{R}^*, \lambda) \\ \mathbf{P}(\lambda, \mathcal{L}) &\equiv \mathbf{P}(\mathbf{R}^*, \lambda, \mathcal{L}) \end{aligned}$$

We call $P_i(\lambda)$ the average power of user i at power prices λ, where it is understood that this is the average power to achieve the rate vector $\mathbf{R}^*$ and minimize the total power cost $\lambda \cdot \mathbf{P}$. Also without loss of generality, we can assume that the average power constraint $\bar{P}_i$

is 1 for all users, by appropriate re-scaling of the fading processes. Hence, our problem is

$$\inf_{\lambda>0} \max_i P_i(\lambda)$$

We propose the following iterative algorithm for solving this problem. The basic idea is that at any iteration of the algorithm, we balance the required average powers of all users as much as possible by increasing the power prices of the users with larger average powers. This will result in lowering the required power of such users by giving them higher priority in the decoding order in more of the fading states. However, perfect balancing is not always possible since the required power of a user cannot be lowered beyond giving him highest priority (i.e. last in the decoding order) at *every* fading state.

Start with an arbitrary positive $\lambda^{(0)}$. Let $D^{(0)}$ be the set of users with the largest average powers $P_i(\lambda^{(0)})$ (at power prices $\lambda^{(0)}$) among all users and $U^{(0)}$ be the set of the other users. We now increase the power prices of all users in D by the same factor $c > 1$ and fix the power prices of users in $U^{(0)}$, and call the resulting prices $\lambda(c)$. Let c^* be the smallest value of c such that at power prices $\lambda(c)$, the average power $P_i(\lambda(c))$ of some user i in $D^{(0)}$ equals $P_j(\lambda(c))$ of some user j in $U^{(0)}$. If no such c exists, set $c^* = \infty$. Consider now two cases:

1) c^* **is finite**: Then let $D^{(1)}$ be the set of users with the largest $P_i(\lambda(c))$ among all users and $U^{(1)}$ be the rest. If $U^{(1)}$ is empty, then the algorithm is terminated. Otherwise repeat the iteration with $\lambda(c)$ in place of $\lambda^{(0)}$, $D^{(1)}$ in place of $D^{(0)}$, and $U^{(1)}$ in place of $U^{(0)}$.

2) c^* **is infinite**: In this case, the minimum of the average powers of users in $D^{(0)}$ is greater than the maximum of the average powers of users in $U^{(0)}$ even when absolute priority is given to users in $D^{(0)}$ over users in $U^{(0)}$, i.e. even when users in $U^{(0)}$ are decoded after users in $D^{(0)}$ at *every* fading state. Thus, perfect power balancing is impossible. Then let $L_1 = U^{(0)}$ and users in $D^{(0)}$ from this step on will always be given absolute priority over users in L_1. The power prices of each user i in L_1 will be fixed at $\lambda_i^{(0)}$ and will not be further adjusted in the algorithm. (They determine the power allocation among users in L_1.) The algorithm is now recursively applied to $D^{(0)}$. For users in $D^{(0)}$, use the same power prices as at the start of the previous iteration, (i.e. set $\lambda_i^{(1)} = \lambda_i^{(0)}$ for $i \in D^{(0)}$) and split the set $D^{(0)}$ into a subset $D^{(1)}$ of users with the largest average powers at price $\lambda^{(1)}$ (calculated now assuming users in L_1 are not present since they are decoded earlier at every fading state), and let $U^{(1)}$ be the rest: $D^{(0)} - D^{(1)}$.

After a finite number of iterations of this algorithm, the users will be partitioned into subsets $L_1, L_2, \ldots L_K$ and H, where users in L_i is given absolute priority over users in L_j for $i > j$ and users in H given the highest priority, and such that no further partitioning of H will take place. Let $\mathcal{L}^* = \{L_1, L_1 \cup L_2, \ldots, \cup_{i=1}^{K} L_i, E\}$ be the absolute priority nesting corresponding to this partitioning of the users. We have the following convergence theorem.

Theorem 4.1 *If $\lambda^{(n)}$ is the vector of power prices at iteration n, then*

$$P_i^* \equiv \lim_{n\to\infty} P_i(\lambda^{(n)}, \mathcal{L}^*)$$

exists for all i, and $P^* \equiv \max_i P_i^*$ is the optimal value for the problem (13), i.e.

$$P^* = \inf_{\lambda>0} \max_i P_i(\lambda)$$

Moreover, $P_i^* = P^*$ for every user i in H.

Proof. First, we observe that for any j, the power allocation of the users in the subset L_j do not change after the iteration when the subset L_j is created. To see that, fix a subset L_j, and let $(\lambda_i^*)_{i\in L_j}$ be the power prices of the users in L_j when L_j is created. Let $H_j \equiv L_{j+1} \cup \ldots \cup L_K \cup H$; this is the subset of users which are given higher priority than users in L_j at all fading states. The rest of the users (in $L_1, \ldots, L_{j-1}$) will be given lower priority than users in L_j at all fading states. The optimal power allocation to users in L_j at fading state $\mathbf{h}$ is given by eqn (7):

$$P_{\pi(i)}(\mathbf{h}) = \frac{\sigma^2}{h_{\pi(i)}}\left[\exp\left\{2(\sum_{k\in H_j} R_k^* + \sum_{k=1}^{i} R_{\pi(k)}^*)\right\} - \exp\left\{2(\sum_{k\in H_j} R_k^* + \sum_{k=1}^{i-1} R_{\pi(k)}^*)\right\}\right\}$$

for $i = 1, \ldots, |L_j|$, where π is an ordering of users in the subset L_j satisfying:

$$\frac{\lambda_{\pi(1)}^*}{h_{\pi(1)}} \geq \ldots \geq \frac{\lambda_{\pi(|L_j|)}^*}{h_{\pi(|L_j|)}}$$

The key point is that the power allocation to users in L_j only depends on the power prices of users in L_j, which remain fixed after the iteration when L_j is created, but do not depend on the power prices of the users of higher priority in H_j, which will be changed in future iterations. Thus, the power allocation to users in L_j stay fixed once L_j is created.

Second, we note that for each j, the minimum of the average powers of users in H_j (high priority users) must monotonically increase after the iteration when L_j is formed. This is because as we scale up the power prices of the users with the largest average powers in H_j, the average powers of all the remaining users must monotonically increase. This is a consequence of the fact that the power prices affect the average powers of the users only through the decoding order given by eqn. (8). Similarly, the average powers of users in H_j must all monotonically decrease and so does the maximum.

It can also be seen that for each j, when the partitioning into H_j and L_j occurs, the minimum of the average powers in H_j must be greater than the maximum of the average powers in L_j. Combining this with the two observations above, we conclude at any iteration after L_j is created, the average power of any user in H_j must be greater than that of any user in $L_1, \ldots L_j$. A typical situation is shown in Fig. 2.

In particular, at any iteration after all of $L_1, L_2, \ldots, L_K$ are created, the average powers of any user in H must be greater than the average power of any user in $\cup_{j=1}^K L_j$. Also, the difference between the maximum and the minimum of the average powers of users in H is monotonically decreasing so it must converge to a limit. Since no further partitioning of H occurs, this limit must be zero. Thus, we have proved that for every user i,

$$\lim_{n\to\infty} P_i(\lambda^{(n)}, \mathcal{L}^*) = P_i^*$$

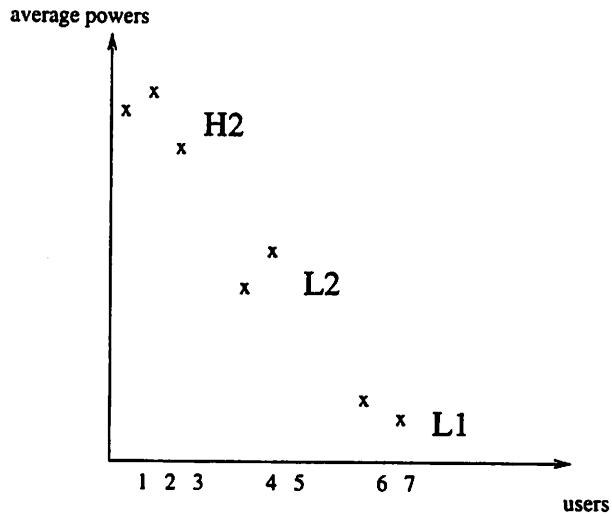

Figure 2: The average powers of the users at the start of an iteration of the resource allocation algorithm. Users are currently partitioned into subsets L_1, L_2 and H_2. Users in H_2 have the largest average powers, and are decoded after users in L_2, which are in turn decoded after users in L_1, at every fading state. The power prices of the users in H_2 will be adjusted in future iterations to further balance their powers; the power prices and average powers of users in L_1 and L_2 will stay fixed.

and

$$P_i^* = \max_{k \in E} P_k^* \qquad \text{for every } i \in H$$

Also, for any power price vector λ,

$$\max_{i \in H} P_i(\lambda) \geq \max_{i \in H} P_i^*.$$

Suppose not. Then there exists λ and n such that

$$P_i(\lambda) < P_i(\lambda^{(n)}, \mathcal{L}^*) \qquad \text{for every } i \in H.$$

This is impossible since under $\mathcal{L}^*$, users in H are already given the highest priority over other users at all fading states and hence $\mathbf{P}(\lambda^{(n)}, \mathcal{L}^*)$ achieves the minimum total average power cost $\sum_{i \in H} \lambda_i^{(n)} P_i$ for users in H. This completes the proof.

□

For the reader who is familiar with flow control problems in virtual circuit networks, this algorithm may be reminiscent of *fair bandwidth allocation* algorithms. Here, the objective is to find a fair average power requirements for the users, weighted by their power constraints. Users in the set H correspond to users whose routes pass through the *bottleneck node*, and have the maximum (weighted) power requirement. In fact, it can be shown that by applying the algorithm recursively to balance the power requirements of users in the subsets $L_1, \ldots L_k$ defined above, one can in fact compute a *min-max fair* solution (see [1] for a max-min fair algorithm for bandwidth allocation.)

5 Auxiliary Constraints on Transmit Power

The constraints on the transmit powers we considered so far are on their long-term *average* value, and under power control, the transmit power will vary depending on the fading state. In practice, one often wants to have some shorter-term constraints on the transmit power as well. These constraints may be due to regulations, or as a way of imposing more stringent 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:

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

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 transmit power of the ith user in every state. We shall now concentrate on the problem of computing the optimal power control subject to these constraints.

We focus on the capacity region:

$$C_d^p(\bar{P}, \hat{P}) \equiv \bigcup_{\mathcal{P} \in \mathcal{F}^p} \bigcap_{\mathbf{h} \in \mathcal{H}} C_g(\mathbf{h}, \mathcal{P}(\mathbf{h}))$$

Lemma 5.1 *A rate vector $\mathbf{R}^*$ lies on the boundary of $C_d^p(\bar{P}, \hat{P})$ if and only if there exists a $\lambda \in \Re^M$ and a power control policy $\mathcal{P}$ such that for every joint fading state $\mathbf{h}$, $\mathcal{P}(\mathbf{h})$ is a solution to the optimization problem:*

$$\min \lambda \cdot \mathbf{P} \qquad \textit{subject to} \qquad \mathbf{R}^* \in C_g(\mathbf{h}, \mathbf{P}) \textit{ and } P_i \leq \hat{P}_i \qquad \forall i \tag{14}$$

and

$$\mathbb{E}_{\mathbf{H}}[\mathcal{P}_i(\mathbf{H})] = \bar{P}_i \qquad i = 1, \ldots, M$$

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 $\mathbf{R}^*$.

The proof of this result is similar to that of Lemma 3.2, and is the analogue of Lemma 5.1 in Part I, and will not be given here.

To solve the optimization problem (14), we first prove a few results about contra-polymatroids.

Definition 5.2 *The rank function f of a contra-polymatroid is said to be* strictly *super-modular if for any subset S, T such that neither is a subset of the other* [2],

$$f(S) + f(T) < f(S \cup T) + f(S \cap T)$$

[2] Clearly, if one is a subset of the other, equality must hold

The following lemma is motivated by a result of Hanly and Whiting [9], which was proved in the context of multi-access capacity regions.

Lemma 5.3 *Let $\mathcal{D}(f)$ be a contra-polymatroid with a strictly supermodular rank function f. Consider any vector $\mathbf{x} \in \mathcal{D}(f)$, and let $S_1, \ldots, S_K$ be the subsets corresponding to the constraints of $\mathcal{D}(f)$ that are tight at $\mathbf{x}$, i.e. these constraints hold with equality at $\mathbf{x}$. Then there exists an ordering π such that*

$$S_{\pi(1)} \subset \ldots \subset S_{\pi(K)}$$

i.e. they are nested.

Proof. Take any two tight constraints corresponding to subsets S_i and S_j. Suppose neither is a subset of the other. Then

$$\begin{aligned} \mathbf{x}(S_i \cup S_j) &= \mathbf{x}(S_i) + \mathbf{x}(S_j) - \mathbf{x}(S_i \cap S_j) \\ &\leq f(S_i) + f(S_j) - f(S_i \cap S_j) \\ &< f(S_i \cup S_j), \end{aligned}$$

a contradiction since $\mathbf{x} \in \mathcal{D}(f)$. Hence, the subsets corresponding to the tight constraints must be nested.

□

Now let a_i's be positive constants, and consider the optimization problem

$$\min \lambda \cdot \mathbf{x} \qquad \text{subject to} \qquad \mathbf{x} \in \mathcal{D}(f) \text{ and } x_i \leq a_i \qquad \forall i \tag{15}$$

where the vector λ satisfies

$$\lambda_1 \geq \ldots \geq \lambda_M \geq 0$$

We will refer to the constraints $x_i \leq a_i$ as *peak* constraints. To motivate the algorithm for solving this problem, we first observe that the algorithm given in Lemma 3.4 (for the same problem but without the peak constraints) can be viewed as a greedy algorithm:

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

With this interpretation, the following greedy algorithm for problem (15) can be viewed as a natural generalization to the case when there are peak constraints:

Algorithm 5.4 • **Initialization**: *Set $x_i^{(0)} = a_i$ for all i. If $\mathbf{x}^{(0)} \notin \mathcal{D}(f)$ then stop. Else set $k = 1$.*

- **Step** k: *Decrease the kth component of $\mathbf{x}$ until a constraint becomes tight. Go to Step $k+1$*

- *Stop after M steps.*

Theorem 5.5 *If $\mathbf{x}^{(0)} \notin \mathcal{D}(f)$, then the optimization problem (15) has an empty feasible region. Otherwise the algorithm 5.4 terminates at an optimal solution to (15).*

Proof.

The first statement follows from the easily verified fact that if $\mathbf{x}, \mathbf{y}$ are two vectors such that $\mathbf{y}_i \leq \mathbf{x}_i \quad \forall i$ and $\mathbf{x} \notin \mathcal{D}(f)$, then $\mathbf{y} \notin \mathcal{D}(f)$.

Now suppose $\mathbf{x}^{(0)} \in \mathcal{D}(f)$ and the algorithm 5.4 terminates at the point $\mathbf{x}^*$. We first show that $\mathbf{x}^*$ is a vertex of the feasible region. At each step k of the algorithm, either the kth component cannot be decreased, in which case the constraint $x_k \leq a_k$ is tight, or it can be decreased until a constraint of $\mathcal{D}(f)$ corresponding to some subset S becomes tight. In any case, at each stage of the algorithm, we are having an additional linearly independent constraint becoming tight. Moreover, since we are always decreasing the components of $\mathbf{x}$, subset constraints that become tight will remain tight. Hence, at termination, there are M linearly independent tight constraints, and $\mathbf{x}$ is a vertex of the feasible region.

Let $S_1, S_2, \ldots, S_K$ be the subset constraints that are tight at $\mathbf{x}^*$. By Lemma 5.3, we can without loss of generality assume that $S_1 \subset \ldots \subset S_K$. Let us now identify the tight peak constraints. Consider the partition of the base set E into $S_1, S_2 - S_1$, $S_3 - S_2, \ldots S_K - S_{K-1}$. Since the tight constraints are all linearly independent, it follows that in each subset $S_j - S_{j-1}$, at most $|S_j - S_{j-1}| - 1$ elements can correspond to tight peak constraints. But since there are $M - K$ tight peak constraints, in fact exactly $|S_j - S_{j-1}| - 1$ elements correspond to peak constraints.

Now, the optimization problem of interest is a linear programming problem. Thus, to verify the optimality of $\mathbf{x}^*$, it suffices to show that that the objective function cannot decrease along any of the M edges of the polyhedron that emanate from $\mathbf{x}^*$. Each edge is obtained by relaxing one of the tight constraints. We consider two cases:

1) Suppose we relax a tight constraint $x_k \leq a_k$, where $k \in S_j - S_{j-1}$ for some j. Let $m \in S_j - S_{j-1}$ be such that the corresponding peak constraint is *not* tight. The edge can be seen to be along the half line:

$$x_k + x_m = x_k^* + x_m^*, x_k \leq x_k^*, x_i = x_i^*, \quad i \neq k, m$$

We first note that $k > m$. For the purpose of contradiction, suppose instead that $k < m$. The point $(x_1^*, \ldots, x_k^* - \epsilon, \ldots, x_m^* + \epsilon, \ldots, x_M^*)$ is in the feasible region, which means that in the kth step of the algorithm, the kth component can be furthered decreased beyond x_k^*. This is a contradiction. Hence, $k > m$. Since the coefficients of the objective function satisfy $\lambda_k \leq \lambda_m$, it follows that the objective function cannot decrease moving along the edge.

2) Suppose we relax a subset constraint corresponding to S_j for some j. If $j < K$, let $k \in S_j - S_{j-1}$ and $m \in S_{j+1} - S_j$ correspond to peak constraints that are not tight at $\mathbf{x}^*$. In this case, the edge can be seen to be along the half line:

$$x_k + x_m = x_k^* + x_m^*, x_k \geq x_k^*, x_i = x_i^*, \quad i \neq k, m$$

Since $\lambda_k < \lambda_m$, it follows that the objective function cannot decrease along this edge. On the other hand, if $j = K$, let $k \in S_K - S_{K-1}$ be the component corresponding to a peak constraint that is not tight. The corresponding edge is along the half line:

$$x_k \geq x_k^*, x_i = x_i^*, \quad i \neq k$$

Clearly, the objective function cannot decrease along this edge.

Hence we conclude that indeed $\mathbf{x}^*$ is an optimal solution.

□

At each step k, algorithm 5.4 has to check when a constraint becomes tight. This is equivalent to the *membership problem*: given a point $\mathbf{x}$, check if $\mathbf{x}$ is in $\mathcal{D}(f)$ or not. For general contra-polymatroids, there is no known efficient combinatorial algorithm to solve this problem (checking every constraint of $\mathcal{D}(f)$ requires complexity exponential in M.) However, for the special case of contra-polymatroids with generalized symmetric rank functions, a very simple test exists. This result is due to Federgruen and Groenevelt [4].

Lemma 5.6 *[4]*

Suppose f is generalized symmetric, i.e. $f(\cdot) = g(\mathbf{y}(\cdot))$ for some convex increasing function g and vector $\mathbf{y}$. Given any $\mathbf{x}$, let σ be a permutation on E such that

$$\frac{x_{\sigma(1)}}{y_{\sigma(1)}} \leq \ldots \leq \frac{x_{\sigma(M)}}{y_{\sigma(M)}}$$

Then $\mathbf{x} \in \mathcal{D}(f)$ if and only if

$$\sum_{i=1}^{m} x_{\sigma(i)} \geq g(\sum_{i=1}^{m} y_{\sigma(i)}) \qquad \forall m = 1, \ldots, M$$

This lemma implies that one only needs to check M constraints to determine if $\mathbf{x}$ is a member of $\mathcal{D}(f)$, instead of $2^M - 1$. Combining this lemma with Algorithm 5.4, we can in fact compute explicitly the value to which the kth component must be decreased to in the kth step of the algorithm. Thus, in the case when $f(\cdot) \equiv g(\mathbf{y}(\cdot))$, the algorithm now becomes:

- **Initialization**: Set $x_i^{(0)} = a_i$ for all i. If $\mathbf{x}^{(0)} \notin \mathcal{D}(f)$ then stop. Else set $k = 1$.
- **Step** k: Let $\sigma^{(k)}$ be a permutation on $\{1, \ldots, k-1, k+1, \ldots, M\}$ such that

$$\frac{x_{\sigma^{(k)}(1)}}{y_{\sigma^{(k)}(1)}} \leq \ldots \leq \frac{x_{\sigma^{(k)}(k-1)}}{y_{\sigma^{(k)}(k-1)}} \leq \frac{x_{\sigma^{(k)}(k+1)}}{y_{\sigma^{(k)}(k+1)}} \leq \ldots \leq \frac{x_{\sigma^{(k)}(M)}}{y_{\sigma^{(k)}(M)}}$$

Then set

$$x_i^{(k)} = \begin{cases} x_i^{(k-1)} & \text{if } i \neq k \\ \max_{j \neq k}[f(S_j \cup \{k\}) - \mathbf{x}(S_j)] & i = k \end{cases}$$

where $S_j \equiv \{\sigma^{(k)}(1), \ldots, \sigma^{(k)}(j)\}$ (noting that the element $\sigma^{(k)}(k)$ does not exist.) Go to step $k+1$.

- Stop after M steps.

Lemma 5.6 implies that at step k of the algorithm, the subset constraints that can become tight are the ones corresponding to the subsets $S_j \cup \{k\}$, for $j = 1, 2, \dots, k-1, k+1, \dots, M$. The value that the kth component should be decreased to is determined by the first of these constraints becoming tight. The complexity of this algorithm is $O(M^2 \log M)$.

By observing that the set of feasible received powers $\mathbf{Q}$ that support a given rate vector $\mathbf{R}^*$ is a contra-polymatroid with generalized symmetric rank function, we can immediately apply the above simplified form of algorithm 5.4 to solve the optimization problem (14). This gives an efficient way to compute the optimal power allocation at a fading state, for given power prices λ. Moreover, the polymatroid theory yields a result of independent interest: an efficient membership test for the Gaussian capacity region. More concretely, given rate vector $\mathbf{R}$ and power constraint $\mathbf{P}$, to check the exponentially large number of constraints:

$$\mathbf{R}(S) \leq \frac{1}{2}\log(1 + \frac{\mathbf{P}(S)}{\sigma^2}) \qquad S \subset E,$$

one needs only to sort $\frac{P_i}{R_i}$'s in ascending order, and check the M nested constraints corresponding to that ordering.

It should be noted that unlike the optimal power control schemes for the previous problems we considered in this paper (parts I and II), the optimal solution for this problem cannot in general be achieved by by successive decoding of the M users. Due to the auxiliary constraints, the optimal solution is not necessarily on a vertex of the capacity region. However, Rimoldi and Urbanke [11] show that each user can be split into at most 2 "virtual users", such that the resulting point *can* be achieved by successive decoding of at most $2M$ virtual users. Their procedure for calculating the power levels that define the splitting is *greedy*, a fact that again arises from the generalized symmetric polymatroidal structure of the Gaussian multi-access capacity region.

6 Multiple Time-Scale Fading and Frequency-Selective Fading Channels

The notions of throughput capacity and delay-limited capacity for fading channels can be viewed as two ends of a spectrum. If we look upon a fading channel as a set of parallel channels, one for each fading state, then the throughput capacity is the maximum total rate that one can achieve by arbitrary allocation of rates and powers over the parallel channels, subject to power constraint. The delay-limited capacity, on the other hand, is the maximum total rate subject to the constraint of a common rate for each of the parallel channels. Thus, one can consider other notions of capacities where the rate allocation policy is not as stringent as in the delay-limited case, but not completely arbitrary as in the throughput capacity. In this section, we will look at two applications of this idea: fading with multiple time-scale dynamics, and frequency-selective fading.

Consider first the situation when the fading processes have two components, one slow and one fast. The slow fading might be due to shadowing, for example, and the fast due to multipath. We assume that the fast fading is sufficiently fast to average out over the tolerable delay, but that we are delay limited with respect to the slow fading. We define a notion of capacity in this context.

For simplicity we take a simple jump Markov model for the slow fading. Let $\mathcal{S}_i$ be the set of slow states for user i. Consider a Markov chain on $\mathcal{S}_i$, with transition probabilities $t(s_i^{(1)}, s_i^{(2)})$ for any $s_i^{(1)}, s_i^{(2)}$ in $\mathcal{S}_i$. Consider also $(p(s_i))_{s_i \in \mathcal{S}_i}$ such that for all s_i, $p(s_i) \in (0,1)$. We then define a Markov process by letting the fading for user i remain in state $s_i^{(1)}$ for a Geometric($p(s_i^{(1)})$) time, and then switching to another state $s_i^{(2)}$ with transition probability $t_i(s_i^{(1)}, s_i^{(2)})$. We assume each user suffers independent slow fading, and so have a Markov process with independent components on the state space $\mathcal{S}$, the cross product of the individual slow state-spaces. Consider a particular stationary, ergodic process $(\mathbf{S}(n))_{n=1}^{\infty}$ of this form, with stationary distribution $\mathcal{Q}$ on $\mathcal{S}$. Let $\mathcal{A}(\mathcal{Q})$ be the set of all such processes with stationary distribution $\mathcal{Q}$.

We also define a set of "conditional fading processes"; for each $j = 1, 2, \ldots$ and each $\mathbf{s} \in \mathcal{S}$, we define independent fading processes $(\mathbf{H}^{(j)}(\mathbf{s}, n))_{n=0}^{\infty}$ such that for all $j, \mathbf{s}$, $(\mathbf{H}^{(j)}(\mathbf{s}, n))_{n=0}^{\infty}$ is a stationary ergodic process on $\mathcal{H}$, where $\mathcal{H}$ is the fading state space. We assume that for all j, $\mathbf{H}^{(j)}(\mathbf{s}, n)$ has stationary distribution $\mathcal{Q}_{\mathbf{s}}$ on $\mathcal{H}$.

Given a slow process $\mathbf{S}(n) \in \mathcal{A}(\mathcal{Q})$, we define an associated fading process $\mathbf{H}(n)$ by

$$\mathbf{H}(n) \equiv \sum_{j=1}^{\infty} I[T_j \leq n < T_{j+1}] \mathbf{H}^{(j)}(\mathbf{S}(n), n - T_j)$$

where $(T_j)_{j=1}^{\infty}$ are the jump times of $\mathbf{S}(n)$. By construction, $(\mathbf{H}(n))_{n=1}^{\infty}$ is stationary and ergodic on $\mathcal{H}$ and, conditional on $\mathbf{S}(n) = \mathbf{s}$, $\mathbf{H}(n)$ has conditional distribution $\mathcal{Q}_{\mathbf{s}}$ on $\mathcal{H}$. Note that we can generate such a fading process by choosing any slow process in $\mathcal{A}(\mathcal{Q})$.

Definition 6.1 *We have a class of fading channels, indexed by the slow processes in $\mathcal{A}(\mathcal{Q})$. We denote the delay limited capacity region with respect to slow fading, and average power constraint $\bar{\mathbf{P}}$ by $C_{ds}(\bar{\mathbf{P}})$. The vector $\mathbf{R} \equiv (R_1, R_2, \ldots, R_M)$ lies in the interior of this region if for every $\epsilon > 0$ there exists a delay T and for for every slow process in $\mathcal{A}(\mathcal{Q})$ there exists a codebook with $2^{R_i T}$ codewords for user i, and a decoding scheme, such that the probability of decoding error is less than ϵ. Each codebook for user i has average power $\bar{\mathbf{P}}_i$ (averaged over the fading).*

From the point of view of a parallel channel decomposition of the fading channel, this corresponds to partitioning the parallel channels into subsets each associated with a slow fading state. In the above definition of delay-limited capacity for multiple time-scale fading channel, one is allowed to do rate allocation among the channels within each subset, but subject to the constraint that the total rate in each subset (slow state) is the same.

Consider the following associated channel: Let $\mathcal{P}(\mathbf{s},\mathbf{h})$ be an arbitrary power control policy, and let the channel have unit power constraints on the users, with the fading for user i being $\mathbf{F}_i(n)\mathcal{P}_i(\mathbf{s},\mathbf{F}(n))$. We assume $\mathbf{s}$ is fixed, and $\mathbf{F}(n)$ has distribution $\mathcal{Q}_s$ on $\mathcal{H}$. We denote the throughput capacity region for this channel by $\mathcal{C}_{ds}(\mathbf{s},\mathcal{P}(\mathbf{s},\cdot))$. Now we return to the channel with both slow and fast fading.

Theorem 6.2 *$\mathcal{C}_{ds}(\bar{P}) = \cup_{\mathcal{P}\in\mathcal{F}} \cap_{\mathbf{s}\in\mathcal{S}} \mathcal{C}_{ds}(\mathbf{s},\mathcal{P}(\mathbf{s},\cdot))$ where $\mathcal{F}$ is the set of power control policies $\mathcal{P}(\mathbf{s},\mathbf{h})$ for which $\mathbb{E}_{S,H}[P(\mathbf{S},\mathbf{H})] = \bar{\mathbf{P}}$.*

The proof of Theorem 6.2 is analogous to that of Theorem 2.3. In particular, it shows that we can choose a single codebook of unit power, and obtain the appropriate codebook for any realization of any slow process in $\mathcal{A}(\mathcal{Q})$ by scaling the elements in this codebook by the appropriate $\mathcal{P}_i(\mathbf{s},\mathbf{h})$.

The dual set $\mathcal{D}_{ds}(\mathbf{R}^*)$ is defined as usual:

$$\mathcal{D}_{ds}(\mathbf{R}^*) \equiv \{\bar{\mathbf{P}} : \mathbf{R}^* \in \mathcal{C}_{ds}(\bar{\mathbf{P}})\}$$

In this section, we shall limit ourselves to the characterization of the extreme points of $\mathcal{D}_{ds}(\mathbf{R}^*)$.

As in Section 3, we characterize any point on the boundary of $\mathcal{D}_{ds}(\mathbf{R}^*)$ by solving the following problem, for every slow state $\mathbf{s}$:

$$\begin{aligned} \min \quad & \sum_i \lambda_i^* \mathbb{E}_{\mathbf{H}|\mathbf{S}=\mathbf{s}}[\mathcal{P}_i(\mathbf{s},\mathbf{H})] s.t. \\ \sum_{i\in\mathcal{L}} R_i^* \quad \leq \quad & \mathbb{E}_{\mathbf{H}|\mathbf{S}=\mathbf{s}}[\frac{1}{2}\log(1+\frac{\sum_{i\in\mathcal{L}}\mathcal{P}_i(\mathbf{s},\mathbf{H})H_i}{\sigma^2})] \end{aligned}$$

Since the delay limited capacity region is convex, there exist Lagrange multipliers $\mu^*(\mathbf{s})$ for which $(\mathbf{R}^*,\mathcal{P}(\mathbf{s},\cdot))$ solves

$$\begin{aligned} \max_{\mathbf{R},\mathcal{P}} \quad & \sum_{i=1}^{M}(\mu_i^*(\mathbf{s})R_i - \lambda_i^*\mathbb{E}_{\mathbf{H}|\mathbf{S}=\mathbf{s}}[\mathcal{P}_i(\mathbf{s},\mathbf{H})]) \\ s.t. \sum_{i\in\mathcal{L}} R_i \quad \leq \quad & \mathbb{E}_{\mathbf{H}|\mathbf{S}=\mathbf{s}}[\frac{1}{2}\log(1+\frac{\sum_{i\in\mathcal{L}}\mathcal{P}_i(\mathbf{s},\mathbf{H})H_i}{\sigma^2})] \end{aligned}$$

which is equivalent to solving

$$\begin{aligned} \max_{\mathcal{R}(\mathbf{s},\mathbf{h}),\mathcal{P}(\mathbf{s},\mathbf{h})} \quad & \sum_{i=1}^{M}(\mu_i^*(\mathbf{s})\mathcal{R}_i(\mathbf{s},\mathbf{h}) - \lambda_i^*\mathcal{P}_i(\mathbf{s},\mathbf{h})) \\ s.t. \sum_{i\in\mathcal{L}} \mathcal{R}_i(\mathbf{s},\mathbf{h}) \quad \leq \quad & \frac{1}{2}\log(1+\frac{\sum_{i\in\mathcal{L}}\mathcal{P}_i(\mathbf{s},\mathbf{h})h_i}{\sigma^2}) \end{aligned} \tag{16}$$

for each $\mathbf{s},\mathbf{h}$. The appropriate $\mu^*(\mathbf{s})$ is determined by the condition

$$\mathbb{E}_{\mathbf{H}|\mathbf{S}=\mathbf{s}}[\mathcal{R}_i(\mathbf{s},\mathbf{H})] = R_i^* \tag{17}$$

A greedy algorithm for solving (16) was presented in Theorem 3.14 in Part I. Moreover, an iterative procedure for computing $\mu^*(\mathbf{s})$ was provided in Algorithm 4.7 of Part I : we start with an arbitrary $\mu(\mathbf{s})$ and update it until (17) holds.

In this section, we have found the minimal cost power control policy to obtain a consistent mutual information vector $\mathbf{R}^*$ over every slow fading state. With this power control, we can obtain any rate strictly below $\mathbf{R}^*$ in a delay limited fashion with respect to slow fading. A very important observation is that to obtain this solution we do not need to know the statistics of the slow fading at all. This is because we have prescribed the delay limited rates $\mathbf{R}^*$ as a constraint, but not the long-term average power consumption. The average power used is obtainable from the solution to the power control problem, but we do not need to know it *a priori*. Moreover, the algorithm that we use to determine $\mu(\mathbf{s})$ does not need to know explicitly the conditional distribution of the fading process given the slow state, but rather it adapts to changes in these statistics.

Another important point is that the solution is of "successive decoding type". If we assume a separation of time-scales to the effect that the code periods are fast relative to the slow fading, then successive decoding is optimal. Given any slow state, we use successive decoding to achieve $\mathbf{R}^*$, as in Section 3; in this case, the decoding order is a function of the slow state $\mathbf{s}$.

The characterization of the extreme points of $\mathcal{C}_{ds}(\bar{\mathbf{P}})$ is slightly more complicated, and we do not attempt it here. Clearly, the calculation of the capacity region requires explicit knowledge of the statistics of the fading, including the slow fading.

Similar reasoning can be applied to the analysis of the delay-limited capacity of frequency-selective fading channels, as defined in Section 6 of Part I. Under an assumption that the product of the delay spread and the Doppler spread is small, one can look upon the frequency-selective fading channel as a time-varying channel where at each fading state, a frequency response is specified for each user, representing the multipath. Thus, it can be viewed as a set of parallel channels, each one jointly specified by the fading state and the frequency. In order to be delay-limited in this channel, each user can allocate rates over the different frequencies but the total rate summed over the frequencies must be the same for each fading state. Thus, the resulting optimization problems are identical to the one studied earlier for multiple time-scale fading processes, and hence the optimal power allocation for given delay-limited rates can be obtained from our theory. This ability of being able to perform dynamic power allocation over different frequencies is an advantage of a wideband system over a narrowband system, especially for delay-sensitive traffic.

7 Power Control for Sub-Optimal Systems

In the previous sections, we have focussed on optimal power control from an information theoretic point of view. We will now demonstrate that the ideas can also be applied, in a straightforward manner, to characterize optimal power control laws for situations when successive decoding is done but non-ideal single-user codes are used so that one is not operating at information theoretic limits.

Consider the multi-access scenario with M users where the mth user has a desired signal-to-interference ratio SIR of α_m. Here, the interference is the sum of the background noise (with power σ^2) and that caused by the users whose signals have not yet been decoded. In general, the SIR requirement of a user depends on the coding scheme, the data rate and the error probability requirement, but we assume that the SIR captures the quality of service requirement of the user. We now ask what is the optimal power control law which maintains the SIR requirements of the users? Focus first on a time-invariant multi-access Gaussian channel where user m has transmit power of P_m and path gain h_m. For a given successive decoding order π, let $\mathcal{F}(\pi,\alpha,\mathbf{h})$ be the set of transmit power vectors $\mathbf{P}=(P_1,\dots,P_M)$ which can support the given SIR vector $\alpha=(\alpha_1,\dots,\alpha_M)$. It is given by:

$$\mathcal{F}(\pi,\alpha,\mathbf{h})=\{\mathbf{P}:\frac{h_{\pi(m)}P_{\pi(m)}}{\sigma^2+\sum_{i<m}h_{\pi(i)}P_{\pi(i)}}\geq\alpha_m \qquad \forall m\}$$

Thus if successive decoding is used, the set of transmit power vectors that can support a given set of SIR requirements α is given by

$$\bigcup_{\pi}\mathcal{F}(\pi,\alpha,\mathbf{h}) \tag{18}$$

Further, if we allow time-sharing between different successive decoding orders, then the set of feasible power vectors is enlarged to the convex hull of (18). Call this polytope $\mathcal{F}(\mathbf{h},\alpha)$.

If we let $R_m\equiv\frac{1}{2}\log(1+\alpha_m)$, i.e. the single-user capacity that can be achieved with a SIR of α_m, then we observe that the set $\mathcal{F}(\mathbf{h},\alpha)$ is the same as

$$\mathcal{G}(\mathbf{h},\mathbf{R})\equiv\{\mathbf{P}:\mathbf{R}\in\mathcal{C}_g(\mathbf{h},\mathbf{P})\}$$

i.e. the set of transmit power vectors such that the rate vector $\mathbf{R}$ is in the multi-access Gaussian capacity region. To see this, note that the only vertex of $\mathcal{F}(\pi,\alpha,\mathbf{h})$ is the power vector in which the SIR's of all users are satisfied with equality. This corresponds to the vertex of $\mathcal{G}(\mathbf{h},\mathbf{R})$ where the successive decoding order is π. Thus, the polytopes $\mathcal{F}(\mathbf{h},\alpha)$ and $\mathcal{G}(\mathbf{h},\mathbf{R})$ have the same set of vertices, and hence must be identical.

With this identification, we can now apply the machinery developed earlier to characterize the optimal power control law to maintain the SIR requirements at all times in a fading channel, subject to transmit power constraints. We allow successive decoding at each fading state, where both the order and the powers can vary with the fading states. Using results in Section 3, we see that the optimal successive decoding order at fading state $\mathbf{h}$ is in increasing $\frac{\lambda_m}{h_m}$, where λ are power prices independent of the fading state, chosen to meet the average power constraints. (Ties can be broken arbitrarily.) For independent fading processes, the boundary of the set of feasible SIR's supportable by given average power constraints $\bar{P}_i$'s consists of vectors α satisfying:

$$\alpha_i\int_0^\infty\frac{\sigma^2}{h_i}\prod_{k\neq i}\left\{1+F_k\left(\frac{\lambda_k}{\lambda_i}h_i\right)\alpha_k\right\}f_i(h_i)dh_i=\bar{P}_i \qquad i=1,\dots,M$$

for some power prices λ. Finally, for a given set of SIR requirements and average transmit power constraints, the algorithm given in Section 4 can be used to determine feasibility and to compute a set of appropriate power prices if feasible.

8 Conclusions

In this paper we have shown that any point on the delay limited capacity region is achievable by solutions of "successive decoding type". Given a set of delay limited bit rates, we have used a Lagrangian characterization of all the possible optimal power vectors to get an explicit parameterization in terms of certain "power prices". Any such optimal solution is obtained by choosing an appropriate set of power prices, and then solving a family of power control problems over a set of parallel time-invariant Gaussian multiple-access channels, one for each fading state. We have exploited the polymatroidal structure of the multi-access Gaussian capacity region to provide a simple greedy solution to each of these power control problems, despite the fact that there are an exponentially large number of constraints. It is also shown that the Lagrange multipliers associated with the power constraints (the power prices) can be computed by simple iterative procedures. We have also addressed the issues of peak power constraints, and extensions of the delay limited concept to multiple time scale fading processes, frequency selective fading and sub-optimal coding schemes.

It is interesting to compare the structure of the optimal schemes for achieving throughput capacities and those for achieving delay-limited capacities. While successive decoding is optimal in both cases, the throughput-optimal schemes maintain the *same* decoding order at all fading states. However, the rates of the users are dynamically adjusted depending on the state, and indeed it is possible that a user may be allocated no rate in some states. For optimal delay-limited schemes, on the other hand, the rates are fixed at all fading states, and the successive decoding order is adjusted to maintain those target rates with the least power cost.

Acknowledgments

This work was initiated when both authors were postdoctoral members of technical staff at the Mathematical Sciences Research Center at A.T. & T. Bell Laboratories. The hospitality of our department head Dr. Debasis Mitra is very much appreciated. Hanly's contribution was completed with the help of an A.R.C. grant, which is gratefully acknowledged.

References

[1] Berteskas, D.P. and R.G. Gallager, *Data Networks*, Prentice Hall, 2nd, ed., 1992.

[2] Cheng R. (1993) On capacity and signal waveform design for slowly fading CDMA channels. *Proc. of 31st Allerton Conf.*, Monticello, Illinois, pp. 11-20.

[3] Edmonds, J. (1970) Submodular functions, matroids and certain polyhedra. In Proc. of Int. Conf. on Combinatorics, Gordon & Beach, 69-87.

[4] Federgruen, A. and H. Groenevelt (1986). The Greedy Procedure for Resource Allocation Problems: Necessary and Sufficient Conditions for Optimality. Operations Research 34, 909-918.

[5] Gallager, R. G. (1985) A perspective on Multiaccess Channels. *IEEE Transactions on Information Theory* IT-31 No.2 :124–142

[6] Gilhousen, K. S., Jacobs, I. M., Padovani, R., Viterbi, A. J ., Weaver, L. A., Wheatley, C. E. (1990) On the capacity of a cellular CDMA system *I.E.E.E. Transactions on Vehicular Technology Vol. 40 No. 2 May 1991.*:303- -312

[7] Hanly S. V. (1995) An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity *IEEE Journal on Selected Areas, special issue on the fundamentals of networking, Vol. 13, No. 7* September.

[8] Hanly,S.V. and D. Tse (1995), Multi-Access Fading Channels: Shannon and Delay-Limited Capacities, *Proceedings of the 33rd Allerton Conference*, Monticello, IL, Oct.

[9] Hanly S.V., Whiting P.A. (1994) Constraints on capacity in a multi-user channel *International symposium on information theory* Trondheim, Norway.

[10] Ozarow L. H., Shamai S., Wyner A. D. (1994) Information-theoretic considerations for cellular mobile radio. *IEEE Trans. on Vehicular Technology*, 43(2):359-378, May.

[11] Rimoldi B, Urbanke R. (1995) A Rate-Splitting Approach to the Gaussian Multiple Access Channel *IEEE Transactions on Information Theory* IT-42 No.2 :364-375.

[12] Shantikumar, J.G. and D.D. Yao (1992), Multiclass queueing systems: Polymatroid structure and optimal scheduling control *Operations Research*, 40(2):293-299.

[13] Yates R. (1995) A framework for uplink power control in cellular radio systems *IEEE Journal on Selected Areas of Communications*, special issue on the fundamentals of networking, Vol 13, No. 7.

Appendices

A Proof of Theorem 2.3

Let $f(\mathbf{h})$ be the equilibrium probability density of being in fading state $\mathbf{h}$. Without loss of generality, assume that the fading of all users is bounded by 1. For each k, let $I_k = \{0, \frac{1}{k}, \frac{2}{k}, \ldots, 1\}^M$ be a partition of the fading state space $[0,1]^M$.

First, suppose that $\mathbf{R}$ is in the interior of $\cap_{\mathbf{h}\in\mathcal{H}} \mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h}))$ for some power control policy $\mathcal{P}$. Let user i generate a random codebook of $2^{R_i T}$ codewords of length T by selecting each symbol at random from a $N(0,1)$ distribution. User i then transmits in time n, the nth symbol of the appropriate codeword, scaled by $\sqrt{\mathcal{P}_i(\mathbf{H}(n))}$. Such a set of codewords then satisfies the power constraint $\bar{\mathbf{P}}$. Given this set of codebooks, let $p(T)$ be the conditional probability of decoding any user incorrectly, using maximum likelihood decoding, under the assumption that the decoder is given the realization $\mathbf{H} = (\mathbf{H}(1), \mathbf{H}(2), \ldots, \mathbf{H}(T))$. For S a subset of $\{1, 2, \ldots, M\}$, let $p(S,T)$ be the conditional probability of decoding any user in S incorrectly, conditional on correctly decoding the users in S^c. The union bound implies

$$p(T) \leq \sum_S p(S,T)$$

As shown in Gallager [5],

$$\begin{aligned} p(S,T) \;\; &\leq \;\; \exp(\rho T \mathbf{R}(S)) \sum_{\mathbf{h}} f(\mathbf{h}) \cdot \\ \sum_{\mathbf{y}} \sum_{(\mathbf{x}_j);j\in S^c} \;\; & Q_j(\mathbf{x}_j|\mathbf{h}) \left[\sum_{(\mathbf{x}_i);i\in S} Q_i(\mathbf{x}_i|\mathbf{h}) p(\mathbf{y}|\mathbf{x},\mathbf{h})^{1/(1+\rho)} \right]^{1+\rho} \end{aligned}$$

for any $\rho > 0$, where $Q_i(\mathbf{x}_i|\mathbf{h})$ is the conditional probability density of $\mathbf{x}_i$ being the codeword of user i, conditional on the fading being $\mathbf{h}$. In our case, we obtain,

$$\begin{aligned} p(S,T) \;\; &\leq \;\; \sum_{\mathbf{h}} f(\mathbf{h}) \cdot \\ & \exp\left(-\rho \left[-T\mathbf{R}(S) + \frac{1}{2} \sum_{n=1}^{T} \log\left(1 + \frac{\sum_{i\in S} \mathcal{P}_i(\mathbf{h}(n))}{\sigma^2(1+\rho)} \right) \right] \right) \end{aligned}$$

By assumption, $\exists\epsilon$ such that

$$\forall \mathbf{h} \in \mathcal{H} \;\; \mathbf{R}(S) \leq \log\left(1 + \frac{\sum_{i\in S}\mathcal{P}_i(\mathbf{h})}{\sigma^2}\right) - \epsilon$$

Thus,

$$p(S,T) \leq \exp(-\rho T(\epsilon - \log(1+\rho))$$

and hence

$$p(T) \le \exp(M \ln 2 - \rho T(\epsilon - \log(1+\rho))) \tag{19}$$

By taking ρ sufficiently small, we have $\epsilon - \log(1+\rho) > 0$ and it follows that $p(T) \to 0$ as $T \uparrow \infty$. Moreover, we have in (19) a bound that decays in T at a rate *independent* of the correlation structure of the fading process. It follows that $\mathbf{R} \in \mathcal{C}_d(\bar{\mathbf{P}})$.

To prove the converse, suppose that $\mathbf{R}$ is an interior point of $\mathcal{C}_d(\bar{\mathbf{P}})$. Recall that we have partitioned the fading state space into cubes $(E_j)_{j=1}^{K^M}$. We consider a sequence of Markov processes defined on $\mathcal{H}$ of the following form. Consider a Markov chain on the "coarse" states E_j with transition probabilities $t(E_j, E_k)$. We use such a chain to define a Markov process on $\mathcal{H}$: conditional on the chain being in coarse state E, we select a fading state for the process by using the stationary distribution conditional on the fading being in E. The process remains in this state for an exponential time $\tau(E) \equiv$ Exponential$(\lambda(E))$ and then selects a new coarse state according to t. We assume that the Markov process has the required stationary distribution on $\mathcal{H}$, by choosing appropriate $(\lambda(E))_{E \in I_k}$. By scaling all $\lambda(E_j)$ by a constant, we can speed up or slow down the rate of fading whilst retaining the required stationary distribution.

For each $T = 1, 2, \ldots$, let $\mathbf{H}^{(T)}$ be such a fading process with the following properties. We assume a random variable $H(0)$ on $\mathcal{H}$ with the stationary distribution of the processes we require. We assume all fading processes start with $H^{(T)}(0) \equiv H(0)$, $T = 1, 2, \ldots$ The initial sojourn time in state $H(0)$ of fading $H^{(T)}$ is given by $\tau_T(H(0))$, where $\tau_T(E_j) \asymp$ Exponential$(r_T \lambda(Ej))$ and independent of $H(0)$ for all j. The constant r_T gives the "rate of fading" for process $H^{(T)}$. Let δ be a fixed, positive constant. By choosing an appropriate decreasing sequence $(r_T)_{T=1}^{\infty}$, $r_T \downarrow 0$, we can ensure that for all j,

$$P(\forall T, \tau_T(E_j) > T) > 1 - \delta \tag{20}$$

Since $R \in \mathcal{C}_d(\bar{\mathbf{P}})$, we can choose for each T and each user i a code of size $2^{R_i T}$ which we label $X^{(T)}(n)$ $n = 1, 2, \ldots, T$ for which the probability of error in channel $\mathbf{H}^{(T)}$ goes to zero with T. Let $p(T)$ be the probability of error for $\mathbf{X}^{(T)}$ under fading $H^{(T)}$. We note that $\mathbf{X}^{(T)}$ may be random; say, with dependence on $H^{(T)}$, although we do not require this. Let $\Omega(E)$ be the subset of the sample space on which $H(0) \in E$ and $\forall T$, $\tau_T(E) > T$. Let Q be uniform on $[0, T]$, and independent of all other variables. Define

$$\begin{aligned}
V(E,T) &\equiv \mathbb{E}[(X^{(T)})^2(Q) | \Omega(E)] \\
W(E,T) &\equiv \mathbb{E}[(X^{(T)})^2(Q) | [H(0) \in E] - \Omega(E)] \\
Z(E,T) &\equiv V(E,T) P(\forall T, \tau_T(E) > T | H(0) \in E] \\
&\quad + W(E,T) P(\exists T : \tau_T(E) \le T | H(0) \in E)
\end{aligned}$$

Then the power constraint is that $\forall T$,

$$\sum_E f(E) Z(E,T) \le \bar{\mathbf{P}}$$

By assumption (20), we have

$$\sum_E f(E)V(E,T) \le \frac{\bar{\mathbf{P}}}{1-\delta}$$

Taking limits along a convergent subsequence, we have that

$$\begin{aligned} V(E,T) &\to V(E) \text{ as } T \uparrow \infty \\ \sum_E f(E)V(E) &\le \frac{\bar{\mathbf{P}}}{1-\delta} \end{aligned}$$

Now let us define a new fading process $\mathbf{H}$ by

$$\mathbf{H}(n) \equiv \sum_E I[\mathbf{H}(0) \in E]\mathbf{h}(E)$$

where $\mathbf{h}(E)$ is the upper corner of the cube E. Note that conditional on $\mathbf{H}(0)$, the fading process is deterministic. Let $q(T|\Omega(E))$ be the conditional probability of error for code $X^{(T)}$ in this new fading channel, conditional on the event $\Omega(E)$. By construction,

$$p(T) \ge P(\Omega(E))q(T|\Omega(E))$$

By assumption, $p(T) \to 0$ as $T \uparrow \infty$, and hence $q(T|\Omega(E)) \to 0$ as $T \uparrow \infty$. But conditional on $\Omega(E)$, we have a a constant fading channel, and a sequence of codes satisfying the power constraint $V(E)$. It follows that for all $E \in I_k$, $\mathbf{R} \in \mathcal{C}_g(\mathbf{h}(E), V(E))$.

Define $\mathcal{F}_{k,\delta}$ to be those power control policies that satisfy the power constraint $\frac{\bar{\mathbf{P}}}{1-\delta}$ and are piecewise constant on each cubic element in I_k. Set

$$\mathcal{P}_{k,\delta}(\mathbf{h}) \equiv \sum_{E \in I_k} V(E)I[\mathbf{h} \in E]$$

and note that $\mathcal{P}_{k,\delta} \in \mathcal{F}_{k,\delta}$. For any power control $\mathcal{P}$, define

$$\bar{\mathcal{C}}_g^{(k)}(\mathbf{h}, \mathcal{P}(\mathbf{h})) \equiv \{\mathbf{R} : \mathbf{R} \in \mathcal{C}_g(\bar{\mathbf{h}}^{(k)}, \mathcal{P}(\mathbf{h}))\}$$

where $\bar{\mathbf{h}}_i^{(k)} = \frac{1}{k}\lceil k\mathbf{h}_i \rceil$ for all $i = 1, 2, \ldots, M$. We have shown that for any $\delta > 0$,

$$\forall \mathbf{h} \in \mathcal{H}, \quad R \in \bar{\mathcal{C}}_g^{(k)}(\mathbf{h}, \mathcal{P}_{k,\delta}(\mathbf{h}))$$

It follows that

$$R \in \cup_{\mathcal{P} \in \mathcal{F}_k} \cap_{\mathbf{h} \in \mathcal{H}} \bar{\mathcal{C}}_g^{(k)}(\mathbf{h}, \mathcal{P}(\mathbf{h})) \tag{21}$$

where $\mathcal{F}_k \equiv \mathcal{F}_{k,0}$. Now by the first part of the proof, we have that

$$\cup_{\mathcal{P} \in \mathcal{F}_k} \cap_{\mathbf{h} \in \mathcal{H}} \mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h})) \subseteq \cup_{\mathcal{P} \in \mathcal{F}} \cap_{\mathbf{h} \in \mathcal{H}} \mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h})) \subseteq \mathcal{C}_d(\bar{\mathbf{P}})$$

We have shown in (21) that

$$\mathcal{C}_d(\bar{\mathbf{P}}) \subseteq \cup_{\mathcal{P} \in \mathcal{F}_k} \cap_{\mathbf{h} \in \mathcal{H}} \bar{\mathcal{C}}_g^{(k)}(\mathbf{h}, \mathcal{P}(\mathbf{h}))$$

But these lower and upper bounds converge as $k \uparrow \infty$, and hence

$$\mathcal{C}_d(\bar{\mathbf{P}}) = \cup_{\mathcal{P} \in \mathcal{F}} \cap_{\mathbf{h} \in \mathcal{H}} \mathcal{C}_g(\mathbf{h}, \mathcal{P}(\mathbf{h}))$$

B Parameterization of Boundary of Capacity Region

One unsatisfactory feature of Section 3 is that we are unable to provide an explicit parameterization of the boundary surface of $\mathcal{C}_d(\bar{\mathbf{P}})$. Theorem 3.5 suggests a parameterization of the boundary surface by $\lambda \in \Re_+^M$, and we discuss this further below.

The following lemma shows that for any $\lambda \in \Re_+^M$ there is at least one $\mathbf{R} \in \Re_+^M$ such that $(\lambda, \mathbf{R})$ solves (10).

Lemma B.1 *Define $x_i = exp(2R_i) - 1$, and the transformation T by*

$$T_i(\mathbf{x}) \equiv \frac{P_i}{\int_0^\infty \frac{\sigma^2}{h} \prod_{k \neq i}(1 + F_k(\frac{\lambda_k h}{\lambda_i})x_k) f_i(h) dh}$$

Then there exists a fixed point of T

Proof. T is continuous, and

$$0 \leq T_i(\mathbf{x}) \leq \frac{P_i}{\int_0^\infty \frac{f_i(h)}{h} dh}$$

and hence T is a mapping from $\prod_i [0, \frac{P_i}{\int_0^\infty \frac{f_i(h)}{h} dh}]$ to itself. By the Brouwer fixed point theorem, there exists a fixed point for T in this set. □

It follows that (10) has a solution in $\mathbf{R}$ for any positive λ. Even if a closed-form parameterization of $\mathcal{C}_d(\bar{\mathbf{P}})$ is not possible, it would be useful to have a computational procedure to find a solution to (10). Consider then the following algorithm, which we might use to try and find such a solution:

$$x \to T^n(x). \tag{22}$$

where x is the starting point of the algorithm, and $T^n(x)$ is the nth iterate. It is easy to show that T^2 satisfies the monotonicity property of Section 4 in Part I. Thus, if T has a unique fixed point then $T^n(x)$ will converge to it from any starting point x. We leave the problem of establishing the uniqueness of the fixed point of T open. It is equivalent to the following conjecture:

Conjecture 1 *1. Given $\bar{\mathbf{P}}$, the mapping $\mu \to \lambda(\bar{\mathbf{P}}, \mu)$ is invertible, implying that we can parameterize the boundary surface of $\mathcal{C}_d(\bar{\mathbf{P}})$ by $\lambda \in \Re_+^M$.*

2. Given $\mathbf{R}^*$, the mapping $\lambda \to \mu(\mathbf{R}^*, \lambda)$ is invertible, implying that we can parameterize the boundary surface of $\mathcal{D}_d(\mathbf{R}^*)$ by $\mu \in \Re_+^M$.

We also conjecture that the analogous results hold in Part I; that is the maps $\lambda(\bar{\mathbf{P}}, \mu)$ and $\mu(\mathbf{R}^*, \lambda)$ are invertible in the throughput capacity case as well.