Optimal Power Allocation over Parallel Gaussian Broadcast Channels

David N.C. Tse*
Dept. of Electrical Engineering and Computer Sciences
University of California at Berkeley
email: dtse@eecs.berkeley.edu

Abstract

We consider the problem of optimal power allocation over a family of parallel Gaussian broadcast channels, each with a different set of noise powers for the users, and obtain a characterization of the optimal solution as well as the resulting capacity region. The solution has a simple greedy structure, just like the corresponding solution to the parallel Gaussian multi-access channel. It is a generalization of the classic water-filling solution for parallel single-user channels. Application of the results to the problem of power control for the downlink wireless fading channel is discussed.

1 Introduction

Many communication channels, such as channels with inter-symbol interference (ISI), fading channels and multi-antenna systems, can be analyzed as a family of parallel Gaussian channels. For example, in the case of ISI channels, each of the parallel channels corresponds to a frequency; in the case of fading channels, each corresponds to a fading state. For single-user parallel Gaussian channels when the transmitter can measure and track the channel, it is well known that capacity can be achieved by an optimal power allocation over the parallel channels. Moreover, the optimal power allocation can be computed via a simple water-filling construction [3].

The concept of decomposition into a family of parallel channels extend to *multi-user* scenarios as well. These channels can be used to model multi-access or broadcast situations

*This author is partially supported by AFOSR under grant F49620-96-1-0199 and by a NSF CAREER Award under grant NCR-9734090. The conference version of this paper has appeared in the International Symposium on Information Theory, Ulm, Germany, 1997.

when each transmitter-receiver pair experiences possibly different ISI or channel fading. For example, in a Gaussian ISI multi-access channel where $H_i(f)$ is the frequency response from the ith transmitter to the receiver, one can view it as a family of parallel multi-access flat Gaussian channel, one for each frequency f, with the path gain from the ith transmitter to the receiver to be $H_i(f)$. It is of interest to obtain the capacity region of such multi-user channels, in analogy to the capacity characterization of single-user parallel Gaussian channels.

Cheng and Verdu have shown that all points in the capacity region of the multi-access ISI channel can be obtained by power allocation over the component parallel channels, and moreover derived the optimal power allocation in the special case of the two-user multi-access channel. A general solution for Gaussian multi-access channels with arbitrary number of users was obtained recently [9, 10], in the context of power control problems for multi-access fading channels. By exploiting the *polymatroid structure* of the multi-access Gaussian capacity region, explicit *greedy* power allocations were obtained to achieve all points on the boundary of the capacity region, yielding an explicit characterization of the region.

In this paper, we will provide an analogous solution for parallel Gaussian broadcast channels. We will obtain, just as for the multi-access channel, explicit greedy power allocations to achieve all point on the boundary of the capacity region, together with an explicit characterization of the region for a given power constraint. Moreover, we will present a simple iterative algorithm to solve the dual problem, that of finding the minimum power required to achieve a given set of target rates. At the end of the paper, we will briefly mention the application of some of these results in the context of power control for the downlink of a wireless fading channel. A more comprehensive study, using some of the results described here, can be found in [8].

After the conference presentation of this work [11], we were informed that a similar optimal power allocation solution was obtained in earlier unpublished work [6]. Our solution and proofs are presented in a simpler form, emphasizing the greedy structure of the optimal solution as well as the similarity to the corresponding solution to the multi-access channel. Moreover, the greedy solution is readily extended to power allocation problems where there are additional power constraints.

2 Parallel Gaussian Broadcast Channels

Consider the M-user Gaussian broadcast channel:

$$Y_i = X + Z_i \qquad i = 1, \ldots, M$$

where the Z_i's are independent zero-mean Gaussian noise and the variance of Z_i is n_i. The transmitter is subjected to a total average power constraint of $\bar{P}$, and we want to send independent information to each of the receivers. This is a degraded broadcast channel and the capacity region is well known [1, 4]. Assuming without loss of generality that $n_1 \leq n_2 \leq \ldots \leq n_M$, the boundary of the capacity region $\mathcal{C}_b(\mathbf{n}, \bar{P})$ is given by

$$\{\mathbf{R} : R_i = \frac{1}{2}\log(1 + \frac{\alpha_i \bar{P}}{n_i + \sum_{j<i} \alpha_j \bar{P}}) \qquad i = 1, \ldots, M, \qquad \text{where } \textstyle\sum_i \alpha_i = 1\} \tag{1}$$

Each point on the boundary corresponds to a choice of α: α_i is the fraction of the total transmit power used for user i's signal. The point is achieved by *superposition coding*, where users' signals are superimposed on each other, together with *interference cancellation*, with the ith user decoding and cancelling the signals intended for the users with noisier channels before decoding its own.

Consider now a family of K parallel broadcast channels, such that in the kth component channel, user i has noise variance $n_i^{(k)}$. The transmitter has a total power constraint of $\bar{P}$. Note that in general, this channel is not degraded, since the orderings of the noise powers of the users are not necessarily the same in each of the component channels. Nevertheless, the capacity region for the case when there are two users and two parallel channels was characterized by El Gamal [5], and the following result is a straightforward generalization to the case when there are M users and K parallel channels.

Theorem 2.1. *The capacity region of the family of parallel broadcast Gaussian channels is given by*

$$\mathcal{C}(\bar{P}) = \bigcup_{\{\mathbf{P}:\sum_{k=1}^{K} P^{(k)} = \bar{P}\}} \sum_{k=1}^{K} \mathcal{C}_b(\mathbf{n}^{(k)}, P^{(k)})$$

(where for two sets A and B, $A + B \equiv \{\mathbf{u} + \mathbf{v} : \mathbf{u} \in A, \mathbf{v} \in B\}$)

Here, $P^{(k)}$ can be interpreted as the total amount of power allocated to the kth component channel. The above theorem says that any achievable rate vector in the overall region is the sum of rate vectors achievable in each of the component broadcast channels, under some power allocation. While the achievability part of the above theorem is obvious, the converse part demonstrates that indeed all optimal capacity-achieving strategies can be viewed as that of power allocation over the component channels.

3 Explicit Characterization of Capacity Region

The above characterization of the capacity region $\mathcal{C}(\bar{P})$ is only implicit, in the sense that it does not give the optimal power and rate allocation (among channels and among users) to achieve each point on the boundary. In this section, we will compute the optimal power and rate allocations, which in turn leads to an explicit characterization of the capacity region $\mathcal{C}(\bar{P})$.

3.1 A Lagrangian Characterization

Before we present the solution, it is instructive to review the corresponding solution for the single-user case. Here, the problem is

$$\max_{\sum_k P^{(k)}=\bar{P}} \sum_k \frac{1}{2}\log\left(1+\frac{P^{(k)}}{n^{(k)}}\right)$$

where $n^{(k)}$ is the noise variance in the kth component channel. The Lagrangian formulation of this convex optimization problem yields:

$$\max_{\{P^{(k)}\}} \sum_k \left[\frac{1}{2}\log\left(1+\frac{P^{(k)}}{n^{(k)}}\right) - \lambda P^{(k)}\right] = \sum_k \max_{P^{(k)}} \left[\frac{1}{2}\log\left(1+\frac{P^{(k)}}{n^{(k)}}\right) - \lambda P^{(k)}\right]$$

with the "power price" λ chosen such that the total power constraint is satisfied. Thus, the overall optimization problem is decomposed into a family of optimization problems, one for each of the component channels. The optimal solution is given by:

$$P^{(k)}* = \left(\frac{1}{2\lambda} - n^{(k)}\right)^+.$$

where we use the notation $x^+ \equiv \max(x, 0)$. This is the classic water-filling solution.

The following lemma shows that the optimal power allocation problem for the broadcast channel can also be decomposed to solving a family of optimization problems, one for each of the parallel channels.

Lemma 3.1. *A rate vector $\mathbf{R}^*$ lies on the boundary surface of $\mathcal{C}(\bar{\mathbf{P}})$ if and only if there exists a nonnegative $\vec{\mu} \in \Re^M$ such that $\mathbf{R}^*$ is a solution to the optimization problem:*

$$\max \vec{\mu} \cdot \mathbf{R} \qquad \textit{subject to } \mathbf{R} \in \mathcal{C}(\bar{\mathbf{P}}). \tag{2}$$

For a given $\vec{\mu}$, a rate vector $\mathbf{R}^*$ solves the above problem if and only if there exist $\lambda \in \Re_+$, rate allocation $\mathbf{R}^{(k)} \in \Re^M$ and power allocation $P^{(k)}$, $k = 1, \ldots K$ such that for every *channel k,* $(\mathbf{R}^{(k)}, P^{(k)})$ *is a solution to the optimization problem:*

$$\max_{(\mathbf{R},P)} \vec{\mu} \cdot \mathbf{R} - \lambda P \qquad \textit{subject to} \qquad \mathbf{R} \in \mathcal{C}(\mathbf{n}^{(k)}, P) \tag{3}$$

and

$$\sum_{k=1}^{k} \mathbf{R}^{(k)} = \mathbf{R}^*, \qquad \sum_{k=1}^{K} P^{(k)} = \bar{P}$$

For a given $\vec{\mu}$ vector, the rate vector $\mathbf{R}^*$ given in the above proposition is the one on the boundary which maximizes $\vec{\mu} \cdot \mathbf{R}$. The vector $\vec{\mu}$ can be interpreted as a set of *rate rewards*, used in prioritizing the users in the resource allocation. As $\vec{\mu}$ is varied, we get all points on the boundary of the convex capacity region; thus, $\vec{\mu}$ can be used to parameterize the boundary of the capacity region. The scalar λ is the Lagrangian multiplier ("power price") chosen such that the total power constraint is satisfied.The vector $\mathbf{R}^{(k)}$ and power $P^{(k)}$ are the optimal rate and power allocated to channel k to achieve $\mathbf{R}^*$.

Proof. The capacity region $\mathcal{C}(\bar{P})$ is convex because we can always perform time-sharing. Hence the first statement follows immediately.

To show the second statement, we first express the capacity region (1) $\mathcal{C}_b(\mathbf{n}, P)$ of an individual broadcast channel in terms of a single inequality, where $\mathbf{n}$ is the vector of noise variances and P is the power constraint.

Without loss of generality, assume that the noise variances n_i's are in increasing order. The rate vector $\mathbf{R}$ on the boundary of the capacity region corresponding to a power allocation $\alpha_1 P, \ldots, \alpha_M P$ is given in (1). Solving the α_i's in terms of the rate vector $\mathbf{R}$ yields an equivalent system of equations:

$$\sum_{i \le m} \alpha_i P = \sum_{i \le m} (n_i - n_{i-1}) \exp\left(2 \sum_{j \ge i} R_j\right) - n_m \qquad m = 1, \ldots, M \tag{4}$$

where $n_0 \equiv 0$. We observe that the right-hand side of the above equation is monotonically increasing in m. Hence, given any rate vector $\mathbf{R}$, provided that eqn. (4) is satisfied for $m = M$, i.e.

$$P = \sum_{i \le M} (n_i - n_{i-1}) \exp\left(2 \sum_{j \ge i} R_j\right) - n_M$$

then there must exist a power allocation $\{\alpha_i\}$ such that eqn. (4) is satisfied for all m, i.e. $\mathbf{R}$ lies on the boundary of the capacity region. Hence, an equivalent characterization of $\mathcal{C}_b(\mathbf{n}, P)$ is $\mathcal{C}_b(\mathbf{n}, P) = \{\mathbf{R} : f(\mathbf{R}) \le P\}$, where

$$f(\mathbf{R}) \equiv \sum_{i=1}^{M} (n_{\pi(i)} - n_{\pi(i-1)}) \exp\left(2 \sum_{j \ge i} R_{\pi(j)}\right) - n_{\pi(M)}$$

and π is a permutation such that $n_{\pi(1)} \leq \ldots \leq n_{\pi(M)}$. We observe that the function f is convex.

Let us now define the set

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

We claim that the set $\mathcal{S}$ is convex. Indeed let $\mathbf{r} \in \mathcal{C}(P)$ and $\mathbf{s} \in \mathcal{C}(Q)$, and $\alpha \in [0,1]$. Let $\mathbf{r} = \sum_k \mathbf{r}^{(k)}$ and $\mathbf{s} = \sum_k \mathbf{s}^{(k)}$, with $\mathbf{r}^{(k)} \in \mathcal{C}_b(\mathbf{n}^{(k)}, P^{(k)})$, $\mathbf{s}^{(k)} \in \mathcal{C}_b(\mathbf{n}^{(k)}, Q^{(k)})$ and $\sum_k P^{(k)} = P, \sum_k Q^{(k)} = Q$. For each channel k, the capacity region is given by:

$$\mathcal{C}_b(\mathbf{n}^{(k)}, P^{(k)}) = \{\mathbf{R} : f_k(\mathbf{R}) = \sum_{i=1}^{M} (n^{(k)}_{\pi(i)} - n^{(k)}_{\pi(i-1)}) \exp\left(2 \sum_{j \geq i} R_{\pi(j)}\right) - n^{(k)}_{\pi(M)} \leq P\}$$

and π is a permutation such that $n^{(k)}_{\pi(1)} \leq \ldots \leq n^{(k)}_{\pi(M)}$. If we define rate vectors $\mathbf{t}^{(k)} \equiv \alpha \mathbf{r}^{(k)} + (1-\alpha)\mathbf{s}^{(k)}$ and power allocation $U^{(k)} = \alpha P^{(k)} + (1-\alpha) Q^{(k)}$, then by the convexity of f_k,

$$f_k(\mathbf{t}^{(k)}) \leq \alpha f_k(\mathbf{r}^{(k)}) + (1-\alpha) f_k(\mathbf{s}^{(k)}) \leq \alpha P^{(k)} + (1-\alpha) Q^{(k)} = U^{(k)}$$

and so $\mathbf{t}^{(k)} \in \mathcal{C}_b(\mathbf{n}^{(k)}, U^{(k)})$. This implies that $\alpha \mathbf{r} + (1-\alpha)\mathbf{s} = \sum_k \mathbf{t}^{(k)}$ is in the capacity region $\mathcal{C}\left(\alpha P + (1-\alpha) Q\right)$. Hence the set $\mathcal{S}$ is convex.

The second statement in the lemma now follows from this fact. By the convexity of $\mathcal{S}$, a rate vector $\mathbf{R}^*$ solves the optimization problem (2) if and only if there exists Lagrange multiplier $\lambda \in \Re_+$ such that $(\mathbf{R}^*, \bar{P})$ is a solution to the optimization problem

$$\max_{(\mathbf{R},P) \in \mathcal{S}} \vec{\mu} \cdot \mathbf{R} - \lambda P \tag{5}$$

By definition of $\mathcal{C}(P)$, $\mathbf{R} \in \mathcal{C}(P)$ if and only there exists $\mathbf{R}^{(k)}$'s and $P^{(k)}$'s such that $\mathbf{R} = \sum_k \mathbf{R}^{(k)}$, $P = \sum_k P^{(k)}$ and $\mathbf{R}^{(k)} \in \mathcal{C}_b(\mathbf{n}^{(k)}, P^{(k)})$ for all component channels k. Hence, the optimization problem (5) now decomposes into a family of independent optimization problems:

$$\max_{\mathbf{R}^{(k)} \in \mathcal{C}_b(\mathbf{n}^{(k)}, P^{(k)})} \vec{\mu} \cdot \mathbf{R}^{(k)} - \lambda P^{(k)} \qquad \forall k.$$

This completes the proof of the lemma. □

3.2 Optimal Power and Rate Allocation

The above lemma implies that to characterize the optimal power and rate allocation, we have to solve the optimization problem (3) for each of the parallel channels. It turns out that there is a simple and explicit greedy solution to the optimization problem (3).

Theorem 3.2. *For given Lagrange multipliers $\vec{\mu}$ and λ and noise variance vector $\mathbf{n}$, consider the optimization problem:*

$$\max_{\mathbf{R},P} \vec{\mu}\cdot\mathbf{R} - \lambda P \qquad s.t. \qquad \mathbf{R} \in \mathcal{C}_b(\mathbf{n}, P) \tag{6}$$

Assume that there are no two users i and j such that $n_i = n_j$ and $\mu_i = \mu_j$.

Define for $i = 1, \ldots, M$ the marginal utility *functions:*

$$u_i(z) \equiv \frac{\mu_i}{2(n_i + z)} - \lambda \tag{7}$$

$$u^*(z) \equiv \left[\max_i u_i(z)\right]^+ \tag{8}$$

and the sets

$$\mathcal{A}_i \equiv \{z \in [0, \infty) : u_i(z) = u^*(z)\}$$

Then the optimal value for the optimization problem (6) is:

$$\int_0^\infty u^*(z)dz$$

and attained at an unique point:

$$R_i^*(\vec{\mu}, \lambda) = \int_{\mathcal{A}_i} \frac{1}{2(n_i + z)}dz \qquad i = 1, \ldots, M \tag{9}$$

$$P^*(\vec{\mu}, \lambda) = \left[\max_i \left(\frac{\mu_i}{\lambda} - n_i\right)\right]^+ \tag{10}$$

To explain the optimal power and rate allocation, let us first re-interpret the classic water-filling solution for the single user case. The solution in that setting is to solve, for each component channel, the optimization problem:

$$\max_{P \geq 0} \left[\frac{1}{2}\log\left(1 + \frac{P}{n}\right) - \lambda P\right],$$

where n is the noise power in the channel. We can write

$$\frac{1}{2}\log\left(1 + \frac{P}{n}\right) = \int_0^P \frac{1}{2(n + z)}dz.$$

This integral representation can be given a *rate splitting* interpretation, where the transmission to a single user can be visualized as being split into many low-rate data streams, each with transmit power dz. The total rate is achieved by successive cancellation among these streams in decreasing order of z, with the rate of the stream decoded at interference level $n + z$ to be $1/[2(n + z)] \cdot dz$.

The optimization problem can be recast in the integral form:

$$\max_{P \geq 0} \int_0^P \left[\frac{1}{2(n+z)} - \lambda \right] dz.$$

Let us define

$$u(z) \equiv \frac{1}{2(n+z)} - \lambda$$

and interpret $u(z) \cdot dz$ as the *marginal utility* (rate revenue minus power cost) of adding a virtual user at interference level $n + z$. The optimal solution can be described by adding more data streams until the marginal utility of adding any further data stream is negative. The power allocated at this point is precisely

$$P^* = \left(\frac{1}{2\lambda} - n \right)^+ .$$

In particular, if $u(0) \leq 0$, then nothing is transmitted at all.

The optimal solution for the multiuser broadcast problem is a natural generalization of this interpretation of the single-user solution. The quantity $u_i(z) \cdot dz$ (u_i defined in eqn. (7)) can be interpreted as the marginal increase in the value of the overall objective function $\vec{\mu} \cdot \mathbf{R} - \lambda P$ due to transmitting a low-rate data stream using power dz to user i at interference level $n_i + z$. Starting at $z = 0$, the optimal solution is obtained in a greedy manner by choosing at each value z, to transmit a data stream of rate $1/[2(n_i + z)] \cdot dz$ to user i which will lead to the largest positive marginal increase in the objective function. Here, the choice is whether to transmit such a data stream, and if so, to which user. The value z can be interpreted as the amount of interference caused by the data streams already allocated. The proof of Theorem 3.2 shows that this interpretation of z is valid in the sense that the data streams added *later* in the procedure will always be transmitted to users with *weaker* channel so that they can always be decoded and subtracted off by the user to which the data stream at level z is transmitted to. When no user can be found such that $u_i(z) > 0$, the procedure terminates. The total rate and power allocated to each user is obtained by aggregating the rates and powers of all the low-rate data streams transmitted to that user, yielding (9) and (10) respectively. An example is shown in Fig. 1.

We now prove that the claimed solution is indeed achievable and optimal.

Proof. Let the optimal solution to (3) be achieved at

$$R^*_{\pi(i)} = \frac{1}{2} \log \left(1 + \frac{\alpha^*_{\pi(i)} P^*}{n_{\pi(i)} + \sum_{j<i} \alpha^*_{\pi(j)}} \right) \qquad i = 1, \ldots, M$$

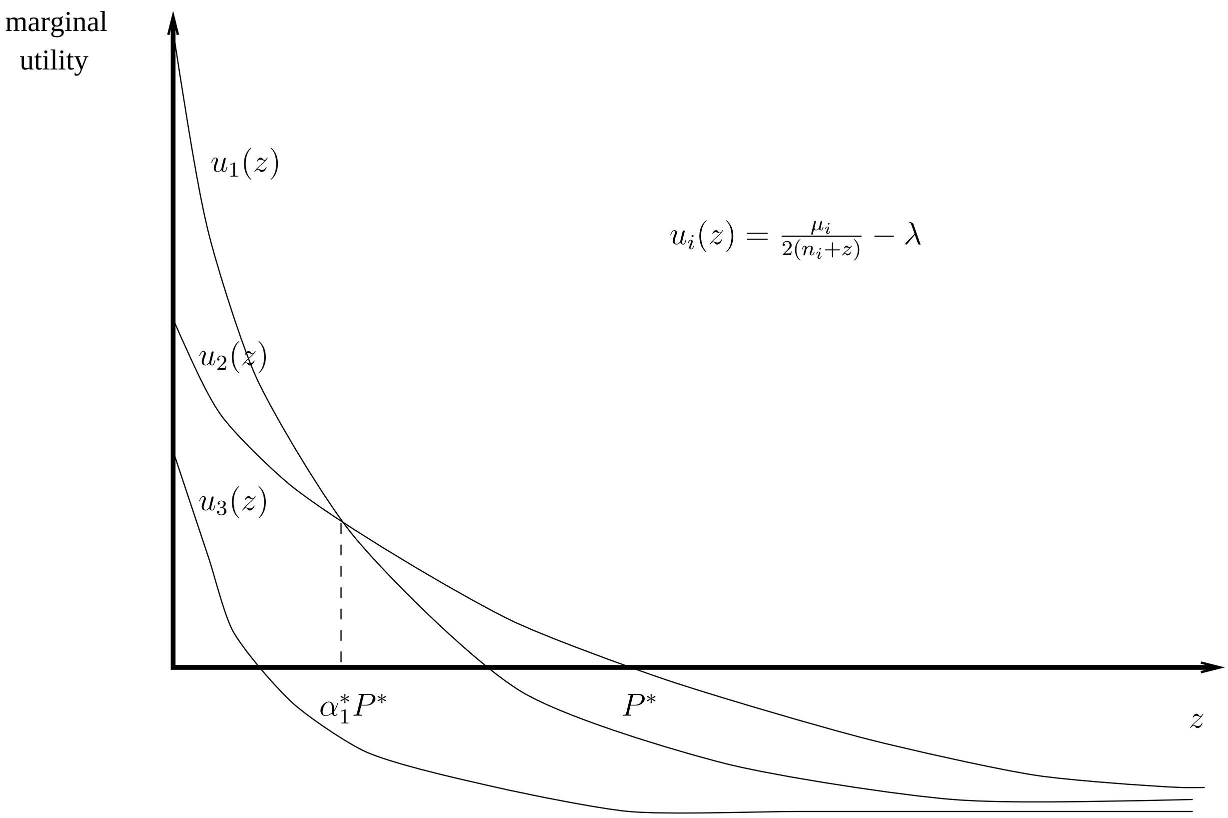

Figure 1: A 3-user example illustrating the greedy power allocation. The x-axis represents the interference level z and y-axis the marginal utility of each user at the interference levels. At each interference level z, the user to which a low-rate data stream is transmitted is the one with the highest marginal utility. Aggregating the data streams yield the optimal rate and power allocation. In this example, the optimal solution is such that user 1 and user 2 are allocated transmit powers $\alpha_1^* P^*$ and $(1 - \alpha_1^*) P^*$ respectively. The solution is achieved by user 2 decoding treating user 1 as noise, while user 1 decodes user 2 signal, strips off its signal, and decodes its own signal in the presence of only the background noise. The proof of the theorem shows that it must be the case that in this example, user 2 has the weaker channel. Note that user 3 gets no power and hence no rate in this component channel.

with power P^*, where π is a permutation of $\{1, \ldots, M\}$ in increasing order of the noise variances and $\sum_{i=1}^M \alpha_i^* = 1$. The optimal value J^* of the problem then satisfies:

$$\begin{aligned}
J^* &= \sum_{i=1}^M \mu_i R_i^* - \lambda P^* \\
&= \frac{1}{2}\sum_{i=1}^M \mu_{\pi(i)} \left[\log\left(n_{\pi(i)} + \sum_{j\le i} \alpha_{\pi(j)} P^*\right) - \log\left(n_{\pi(i)} + \sum_{j\le i-1} \alpha_{\pi(j)} P^*\right)\right] \\
&= \sum_{i=1}^M \int_{\sum_{j\le i-1}\alpha_{\pi(j)}P^*}^{\sum_{j\le i}\alpha_{\pi(j)}P^*} \frac{\mu_{\pi(i)}}{n_{\pi(i)}+z} dz - \lambda P^* \\
&= \sum_{i=1}^M \int_{\sum_{j\le i-1}\alpha_{\pi(j)}P^*}^{\sum_{j\le i}\alpha_{\pi(j)}P^*} u_i(z) dz \\
&\le \int_0^\infty u^*(z) dz
\end{aligned}$$

We also see that if this upper bound is actually attained by some power allocation, then the optimal solution must be unique. Thus our remaining task is to show the achievability of this upper bound. First, note that by the monotonicity of the marginal utility functions u_i's, the function u^* is monotonically decreasing. Also, since $\lim_{z\to\infty} \max_i u_i(z) = -\lambda < 0$, there exists a finite z_0 such that $u^*(z_0) = 0$. Define now the sets

$$\mathcal{A}_i \equiv \{z \in [0, z_0] : u_i(z) = u^*(z)\},$$

which form a partition of $[0, z_0]$. Consider the rate and power allocation:

$$\begin{aligned}
R_i^* &= \int_{\mathcal{A}_i} \frac{1}{2(n_i+z)} dz \qquad i = 1, \ldots, M && (11) \\
\alpha_i^* P^* &= |\mathcal{A}_i| && (12) \\
P^* &= z_0
\end{aligned}$$

It can be seen that

$$\vec{\mu} \cdot \mathbf{R}^* - \lambda P^* = \int_0^\infty u^*(z) dz.$$

So to verify that $(\mathbf{R}^*, P^*)$ is indeed optimal, it suffices to show that $\mathbf{R}^*$ is achievable using total power of P^*, i.e. $\mathbf{R}^* \in \mathcal{C}_b(\mathbf{n}, P^*)$. We will in fact show that the power allocation $\alpha_i^* P^*$ among the users will do it.

Consider any two marginal utility functions $u_i(z)$ and $u_j(z)$ and suppose they intersect at $z = \bar{z}$. Then

$$\frac{\mu_i}{n_i + \bar{z}} = \frac{\mu_j}{n_j + \bar{z}}. \qquad (13)$$

Also,

$$\frac{u_i'(\bar{z})}{u_j'(\bar{z})} = \frac{\mu_i}{\mu_j}\left(\frac{n_j+\bar{z}}{n_i+\bar{z}}\right)^2 = \frac{n_j+\bar{z}}{n_i+\bar{z}} \tag{14}$$

by eqn. (13). We observe that whether this ratio is greater or less than 1 is independent of the intersecting point $\bar{z}$. Since the derivatives of the utility functions are always negative, this implies that the derivative of one function is always greater than the other at all the intersections. Clearly this implies that any two utility functions u_i and u_j can intersect at most once. Since $\mathcal{A}_i$ is the set of all z's where u_i dominates over all other u_j's, this implies that the sets $\mathcal{A}_i$'s must all be contiguous, i.e. single intervals.

Next we investigate how the sets $\mathcal{A}_i$'s are ordered on the real line. Suppose $\mathcal{A}_i$ and $\mathcal{A}_j$ are both non-empty and adjacent to each other, then the point $\bar{z}$ where u_i and u_j intersect is also the point where $\mathcal{A}_i$ and $\mathcal{A}_j$ touch. Suppose also that $\mathcal{A}_i$ is to the left of $\mathcal{A}_j$. This implies that for $z < \bar{z}$, $u_i(z) > u_j(z)$ and for $z > \bar{z}$, $u_i(z) < u_j(z)$, i.e.

$$\frac{u_i'(\bar{z})}{u_j'(\bar{z})} > 1$$

From eqn. (14), this implies that $n_j > n_i$. Thus the sets $\mathcal{A}_i$'s are ordered on the real line in increasing values of the noise variances n_i's, although some of the sets $\mathcal{A}_i$'s can be empty. Let π be the permutation such that $n_{\pi(1)} < \ldots < n_{\pi(M)}$. Using eqn. (12), we can write

$$\mathcal{A}_{\pi(i)} = \left[\sum_{j<i}\alpha^*_{\pi(j)}P^*, \sum_{j\leq i}\alpha^*_{\pi(j)}P^*\right]$$

and using eqn. (11), we can write

$$\begin{aligned} R^*_{\pi(i)} &= \int_{\sum_{j<i}\alpha^*_{\pi(j)}P^*}^{\sum_{j\leq i}\alpha^*_{\pi(j)}P^*} \frac{1}{2(n_i+z)}dz \\ &= \frac{1}{2}\log\left(1+\frac{\alpha^*_{\pi(i)}P^*}{n_i+\sum_{j<i}\alpha^*_{\pi(j)}P^*}\right) \end{aligned}$$

Thus, the rate vector $\mathbf{R}^*$ is achievable by superposition coding and interference cancellation, by allocating power $\alpha_i^* P^*$ to the ith user and decoding it by first canceling off the signals intended for users with noisier channels. The total power used is P^* which is the zero of $u^*(z)$, i.e. the largest of the zeros of the functions $u_i(z)$'s:

$$P^* = \left[\max_i\left(\frac{\mu_i}{\lambda} - n_i\right)\right]^+$$

This concludes the proof. □

We note the assumption that there are no two users with identical rate rewards μ_i's and noise variances n_i's is made without loss of generality. For if there are two such users, we can combine them into a "super-user" with rate and power allocated being the sum of the two individual users. Any optimal solution for the new system would translate into an optimal solution for the original system by any arbitrary split in the power allocated. Naturally, the solution is no longer unique in that case.

It is interesting to observe this optimal power allocation solution has a similar structure to the corresponding solution for parallel Gaussian multi-access channels [10]. In that setting, a component multi-access channel is:

$$Y = \sum_{i=1}^{M} h_i X_i + Z$$

where $\mathbf{h} = (h_1, \ldots, h_M)$ are channel attenuations, and $Z \sim N(0, \sigma^2)$. Each of the parallel channels has a different value of $\mathbf{h}$. User i has a total power constraint of $\bar{P}_i$. Using Lagrangian techniques, it can be shown that for given rate rewards $\vec{\mu}$, the optimal power and rate allocation which maximizes the rate revenue in the overall capacity region solves the following optimization problem for each of the parallel channels:

$$\max_{\mathbf{R},\mathbf{P}} \vec{\mu} \cdot \mathbf{R} - \lambda \cdot \mathbf{P} \qquad \text{s.t.} \qquad \mathbf{R} \in \mathcal{C}_m(\mathbf{h}, , \mathbf{P})$$

where

$$\mathcal{C}_m(\mathbf{h}, \mathbf{P}) = \left\{ \mathbf{R} : \sum_{i \in S} R_i \leq \frac{1}{2} \log \left(1 + \frac{1}{\sigma^2} \sum_{i \in S} h_i P_i \right), \qquad \forall S \subset \{1, \ldots, M\} \right\}$$

is the capacity region of a multi-access channel with transmit power P_i's. Here, λ_i is the Lagrangian multiplier reflecting the total power constraint for user i. If we define the received power of user i to be $Q_i = h_i P_i$ and consider for each user the marginal utility function:

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

then applying exactly the same greedy procedure as for the broadcast channel gives as the optimal rates $\mathbf{R}^*$ and the optimal *received* powers $\mathbf{Q}^*$ allocated for this component channel.

3.3 Boundary of the Capacity Region

Let us now use the optimal rate and power allocation derived above for parallel broadcast channels to compute explicitly the capacity region. For any non-negative $\vec{\mu}$, the uniqueness of the optimal rate and power allocation implies that we can define a parameterization $\mathbf{R}^*(\mu)$ of the boundary, which is the unique rate vector on the boundary which maximizes $\vec{\mu} \cdot \mathbf{R}$. Combining the Lagrangian formulation given in Lemma 3.1 and the optimal power and rate allocation solution in Theorem 3.2, we get:

Theorem 3.3. *Assume that the noise variances of the users are distinct in each of the broadcast channels. Then the boundary is given by:*

$$\{\mathbf{R}^*(\vec{\mu}, \bar{P}) : \sum_i \mu_i = 1\}$$

where

$$R_i^*(\vec{\mu}, \bar{P}) = \int_0^\infty \left\{ \sum_{\{k: u_i^{(k)}(z) = [\max_j u_j^{(k)}(z)]^+\}} \frac{1}{2(n_i^{(k)} + z)} \right\} dz$$

and λ satisfies:

$$\sum_{k=1}^{K} \left[\max_i \left(\frac{\mu_i}{\lambda} - n_i^{(k)} \right) \right]^+ = \bar{P} \tag{15}$$

To compute the optimizing rates $\mathbf{R}^*(\vec{\mu}, \bar{P})$ for given rate rewards $\vec{\mu}$, it is necessary to solve eqn. (15) for the appropriate power price λ. Since the left-hand side of eqn. (15) is monotonically decreasing in λ, this can be done by a simple binary search.

Again, if the noise variances in some of the channels are the same, then there may not be a unique rate vector $\mathbf{R}^*$ maximizing $\vec{\mu} \cdot \mathbf{R}$ for some values of $\vec{\mu}$. These correspond to linear surfaces formed by convex hull of points obtained by giving strict priority to one of the users with identical rate rewards and noise variances.

It is interesting to look at the point on the boundary corresponding to all the rate rewards μ_i's equal, say to 1. The associated rate vector is the one which maximizes the total throughput $\sum_i R_i$. In this case, the marginal utility function of the ith user in the kth component channel is:

$$u_i^{(k)}(z) = \frac{1}{2(n_i^{(k)} + z)} - \lambda.$$

Observe that in a given component channel, the marginal utility functions of all users are parallel: for users with different noise variances, their utility functions do not intersect. We conclude that the optimal power allocation is always allocating all power, if any, to the user with the best reception in each of the component channels. (If more than one user has the best channel, any arbitrary split in the power allocation among these users would be optimal.) Furthermore, if even the noise variance of the best user does not meet a certain threshold, no power is allocated at all. More precisely, the optimal rate and power allocation for the case when the noise variances are distinct in all of the parallel channels is given by:

$$\begin{aligned} R_i^{(k)} &= \begin{cases} \frac{1}{2}\log\left[1+\left(\frac{1}{\lambda n_i^{(k)}}-1\right)^+\right] & \text{if } n_i^{(k)} < n_j^{(k)} \text{ for all } j \neq i \\ 0 & \text{otherwise} \end{cases} \\ P^{(k)} &= \left(\frac{1}{\lambda} - \min_i n_i^{(k)}\right)^+ \end{aligned}$$

Thus the optimal solution has the interesting feature that the information for no more than one user is broadcasted in each of the parallel channels (the user with the best reception), and the power allocation across the channels is a water-filling solution. The optimal power allocation is the same as that of a family of single-user parallel channels with noise variance in each of the channels the same as that of the user with the best reception in the broadcast channel. We note that the corresponding solution for parallel multi-access channels has the same structure [2].

3.4 An Iterative Algorithm for Resource Allocation

We have formulated the problem of optimal power allocation and computation of the resulting capacity region as that of optimizing the total rate revenue $\vec{\mu} \cdot \mathbf{R}$ subject to a total power constraint $\bar{P}$, for arbitrary choice of $\vec{\mu}$. A problem which is of more interest in some applications is in some sense "dual" to this one:

What is the minimum total power required to support a given target rate vector $\tilde{\mathbf{R}}$?

That is:

$$\min P \qquad \text{subject to} \qquad \tilde{\mathbf{R}} \in \mathcal{C}(P). \tag{16}$$

By convexity of the set $\mathcal{S} = \{(\mathbf{R}, P) : \mathbf{R} \in \mathcal{C}(P)\}$, this problem is equivalent to:

$$\min_{(\mathbf{R},P)} P - \vec{\mu} \cdot \mathbf{R} \qquad \text{subject to} \qquad \mathbf{R} \in \mathcal{C}(P) \tag{17}$$

where $\vec{\mu}$ are the Lagrange multipliers chosen such that the target rate vector $\tilde{\mathbf{R}}$ is met. But we have already solved the optimization problem (17): setting $\lambda = 1$ in Theorem 3.2, we get the optimal solution

$$R_i^(\vec{\mu}) = \int_0^\infty \left\{ \sum_{\{k: u_i^{(k)}(z) = [\max_j u_j^{(k)}(z)]^+\}} \frac{1}{2(n_i^{(k)} + z)} \right\} dz \tag{18}$$

$$P(\vec{\mu}) = \sum_{k=1}^{K} \left[\max_i \left(\mu_i - n_i^{(k)} \right) \right]^+ \tag{19}$$

where

$$u_i^{(k)}(z) = \frac{\mu_i}{2(n_i^{(k)} + z)} - 1.$$

Thus, to solve the minimum power problem (16), we have to find a solution $\vec{\mu} = \tilde{\mu}$ to the system of equations $\mathbf{R}(\vec{\mu}) = \tilde{\mathbf{R}}$, and then the minimum power to achieve the rates $\tilde{\mathbf{R}}$ is $P(\tilde{\mu})$. Another view of this is that we are searching for a set of rate rewards $\vec{\mu}$ and power P such that $\tilde{\mathbf{R}}$ maximizes $\vec{\mu} \cdot \mathbf{R}$ subject to $\mathbf{R} \in \mathcal{C}(P)$. Note that we do not require that the solution for $\tilde{\mu}$ is unique; any one solution will do.

We now present a simple iterative algorithm to solve for $\tilde{\mu}$.

Algorithm 3.4. *Start the iteration at $\vec{\mu}(0) = 0$. Given the nth iterate $\vec{\mu}(n)$, the $n+1$th iterate $\vec{\mu}(n+1)$ is given by the following: for each i, $\mu_i(n+1)$ is a rate reward for the ith user such that $R_i(\vec{\mu}) = \tilde{R}_i$, when the rate rewards of the other users remain fixed at $\vec{\mu}(n)$ while the reward for the ith user is adjusted.*

Proposition 3.5. *Each iteration in the above algorithm is well defined, and $\vec{\mu}(n)$ converges to a solution $\tilde{\mu}$ of the system $\mathbf{R}(\vec{\mu}) = \tilde{\mathbf{R}}$.*

The key to the proof of this result is the following monotonicity lemma:

Lemma 3.6. *For all i, if the ith component of $\vec{\mu}$ is increased and the other components are held fixed, the rate $R_i(\lambda)$ remains the same or increases while $R_j(\mu)$ decreases for $j \neq i$.*

This lemma can be easily seen to be true by direct inspection of the expression for $\mathbf{R}(\vec{\mu})$ in eqn. (18).

Proof. (Proposition 3.5) If we fix all components of $\vec{\mu}$ except $\vec{\mu}_i$ and increase $\vec{\mu}_i$ from 0 to ∞, then we see that the marginal utility function $u_i^{(k)}(\cdot)$ increases without bound for every component channel k. This implies that $R_i(\vec{\mu})$ monotonically increases from 0 to ∞. Hence, at each step of the iteration in the algorithm, one can always find for each user a valid $\mu_i(n+1)$ such that its rate equals the target $\tilde{R}_i$, when the rate rewards of other users remain fixed. Moreover, this can be done by a binary search.

To show convergence of the algorithm, it helps to define the mapping representing each iteration:

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

We observe that any fixed point $\vec{\mu}$ of T is a solution of the system $\mathbf{R}(\vec{\mu}) = \tilde{\mathbf{R}}$ that we seek.

It follows from the monotonicity lemma 3.6 that the mapping T is *order preserving*, i.e.

$$\vec{\mu}^{(1)} \leq \vec{\mu}^{(2)} \Rightarrow T(\vec{\mu}^{(1)}) \leq T(\vec{\mu}^{(2)})$$

(where the inequality refers to component-wise.) Starting with $\vec{\mu}(0) = 0$, $\vec{\mu}(1)$ is non-negative. Hence $\vec{\mu}(1) = T(\vec{\mu}(0)) \geq \vec{\mu}(0)$. Applying the monotonic mapping T, we see that $\vec{\mu}(n+1) = T(\vec{\mu}(n)) \geq \vec{\mu}(n)$. If $\vec{\mu}^*$ is a fixed point of T, then since $\vec{\mu}(0) \leq \vec{\mu}^*$, it follows that for all n, $\vec{\mu}(n) = T^n(\vec{\mu}(0)) \leq T^n(\vec{\mu}^*) = \vec{\mu}^*$. Hence, $\{\vec{\mu}(n)\}$ is a monotonically increasing sequence bounded from above, and must converge to a limit. The limit must be a fixed point of T by continuity of T, and hence a solution to the system $\mathbf{R}(\vec{\mu}) = \tilde{\mathbf{R}}$. This completes the proof.

□

3.5 Auxiliary Power Constraints

In some problems, there may also be constraints on the amount of power allocated to *each* of the component broadcast channels in addition to the total power constraint $\bar{P}$. The greedy solution described in Section 3.2 can be extended naturally to handle these types of problem. More concretely, suppose there is a power constraint $\hat{P}$ on the power allocated to each of the channels. Then the associated optimization problem for each of the channels is:

$$\max_{\mathbf{R},P} \vec{\mu} \cdot \mathbf{R} - \lambda P \qquad \text{s.t.} \qquad \mathbf{R} \in \mathcal{C}_b(\mathbf{n}, P) \quad \text{and} \quad P \leq \hat{P} \tag{20}$$

with λ chosen such that the total power constraint is satisfied. By a simple extension of a proof similar to that of Theorem 6, it can be shown that the optimal power and rate allocations can be obtained just as the greedy procedure described, but this time allocating power at most up to the maximum limit of $\hat{P}$. (See Fig. 2.)

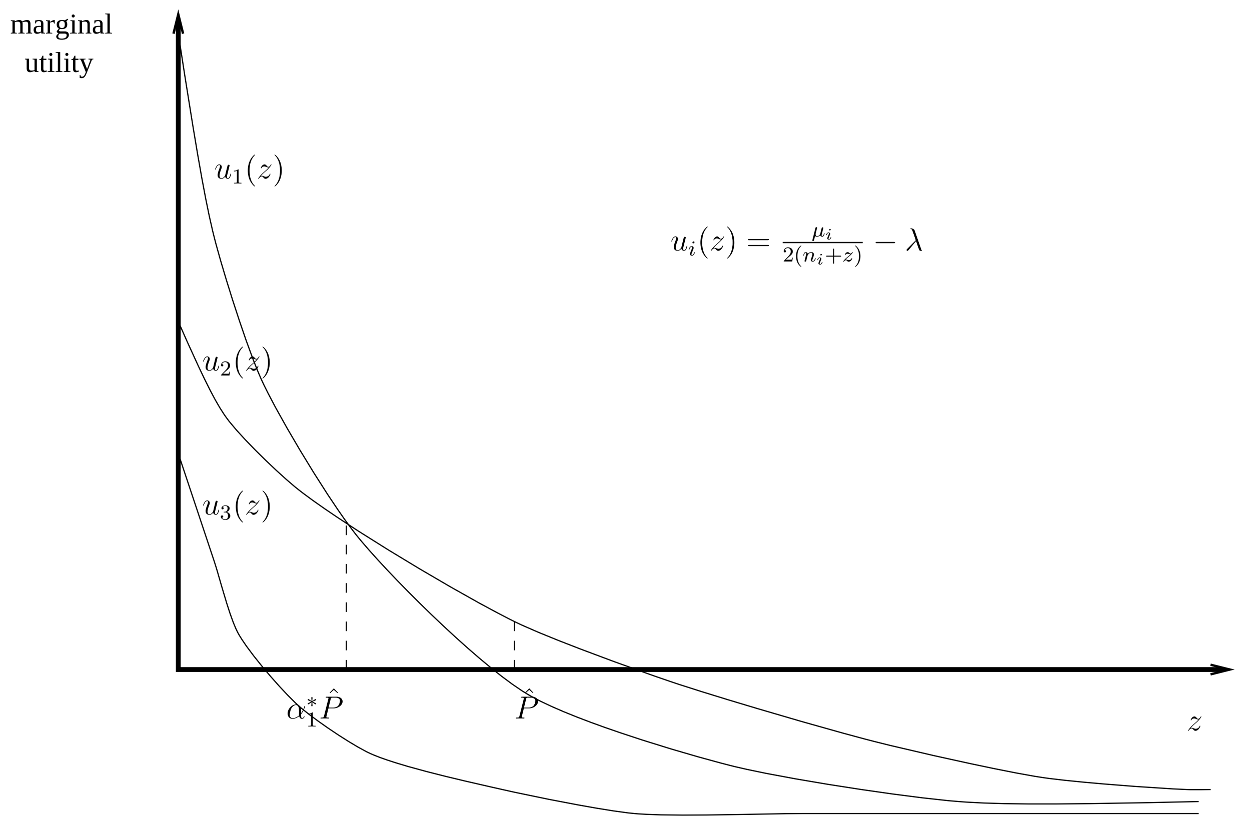

Figure 2: Greedy power allocation under individual power constraint $\hat{P}$ on each of the parallel channel. User 1 gets power $\alpha_1^* \hat{P}$ and user 2 gets power $(1 - \alpha_1^*)\hat{P}$, such that the total power allocated to this parallel channel mets the auxiliary power constraint $\hat{P}$ on each of the parallel channels. As before, user 3 gets no power.

4 Application to a Fading Channel

Consider a discrete-time broadcast fading channel for modeling the downlink of a cell:

$$Y_i(n) = \sqrt{H_i(n)}X(n) + Z_i(n) \qquad i = 1, \ldots, M$$

where $(H_1(n), \ldots, H_M(n))$ is the joint fading process, assumed to be stationary and ergodic with users fading independently. The noise $Z_i(n)$ are i.i.d. Gaussian with zero mean and unit variance. It can be shown that the capacity region when the transmitter can track the fading state of the channel perfectly is given by :

$$\bigcup_{\{\mathcal{P}:\mathbf{E}_{\tilde{\mathbf{H}}}[\mathcal{P}(\mathbf{H})]\leq \bar{P}\}} \mathbf{E}_{\tilde{\mathbf{H}}}[\mathcal{C}(\mathbf{H}, \mathcal{P}(\mathbf{H}))]$$

where $\bar{P}$ is the average total power constraint, $\mathcal{P}$ is a power allocation as a function of the fading state, and $\mathcal{C}(\mathbf{h}, P)$ is the capacity of a broadcast channel with fixed path gains $\mathbf{h}$ and total average power constraint P. The expectation is taken with respect to the stationary distribution of the fading processes. Thus, the fading channel can be viewed as a family of parallel broadcast Gaussian channels, one for each fading state. Using similar arguments as in the last section, we can characterize the optimal power and rate allocation as a function of the fading state $\mathbf{h}$; it is given by Lemma (3), with $n_i^{(k)}$ replaced by $\frac{1}{h_i}$. Also, we can compute the boundary of the region, as in Theorem 3.3, with the sums over the parallel channels replaced by expectations over the stationary distribution of the fading state. The strategy that maximizes the total throughput is one which at any time only broadcasts the information of the user with the strongest reception. This can be thought of as the broadcast "dual" of the optimal strategy proposed by Knopp and Humblet for the multi-access fading channel [7].

If some users have statistically poorer channels than others, the above strategy, although maximizing the total throughput, can lead to unfairness among users. By assigning different rate reward μ_i to users, this unfairness can be compensated for. In fact, for applications in which each user has a target rate R_i to meet, and the goal is to minimize the total power consumption, Algorithm 3.4 can be used to iteratively compute the appropriate rate rewards μ_i's. When applied in real time, this can thought of as a *two-time-scale* adaptive resource allocation procedure. At a slower time-scale, the rate rewards are continuously updated by Algorithm 3.4 to adapt to change in the fading statistics of the users. At a faster time-scale when the rate rewards can be assumed to be fixed, the greedy procedure computes the optimal rate and power allocation for a given fading state.

Under the optimal strategy, the total transmit power will fluctuate depending on the fading state. In some situations, it may be more desirable to have a constraint $\hat{P}$ on the transmit power at *all* fading states, in addition to or in lieu of a long term average power constraint. This is especially relevant on the downlink since, unlike the mobiles in the uplink, the broadcast base-station is usually not battery power limited. Rather, the constraint aims to limit the interference caused in the adjacent cells. A power constraint at all fading states may better reflect that objective. The greedy solution described in Section 3.5 can readily applied to this problem. In the case when there is no average power constraint at all, the power price λ is simply set to be zero. In this case, the strategy that maximizes the total throughput is simply to allocate power $\hat{P}$ to the user with the best channel.

In [8], some of the results described here are used to study the fading channel in greater depth, comparing the performance of the optimal strategy with sub-optimal schemes such as TDMA and FDMA.

Acknowledgment

We would like to thank Professor Andrea Goldsmith for informing us about the work [6] during the preparation of this manuscript.

References

[1] P. Bergmans, "Coding theorem for broadcast channels with degraded components," *IEEE Trans. on Information Theory*, v. 19, n. 2, 197-207, 1973.

[2] R.S. Cheng, S. Verdu "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling", *IEEE Trans. on Information Theory*, Vol. 39, No. 3, pp. 773-786, 1993.

[3] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968.

[4] R.G. Gallager, "Capacity and coding for some broadcast channels", *Probl. Peredachi Inf.*, v.10, n. 3, pp.3-14, 1974.

[5] A.A. El Gamal, "Capacity of the Product and Sum of Two Unmatched Broadcast channels", *Probl. Peredachi Inf.*, v. 16, n. 1, pp.3-23, 1974.

[6] D. Hughes-Hartog, *The capacity of a degraded spectral Gaussian broadcast channel*, Ph.D. thesis, Stanford University, July 1995.

[7] R. Knopp and P.A. Humblet " Information capacity and power control in single-cell multiuser communications", ICC, Seattle, Wash., June 1995.

[8] L. Li and A. Goldsmith , " Capacity and Optimal Resource Allocation for Fading Broadcast Channels", presented at the Allerton Conf. on Communication, Control, and Computing, Sept. 1998.

[9] D.N. Tse and S.V. Hanly, "Capacity Region of the Multi-access Fading Channel Under Dynamic Resource Allocation and Polymatroid Optimization", *Proc. of IEEE Information Theory Workshop*, Haifa, p. 37., 1996.

[10] D.N. Tse and S.V. Hanly, " Multi-Access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities" to appear in the *IEEE Trans. on Information Theory*, Nov., 1998.

[11] D.N. Tse, "Optimal Power Allocation over Parallel Gaussian Broadcast Channels", *Proc. of Int. Symp. on Information Theory*, Ulm Germany, June 1997, p. 27.