ASYMPTOTICALLY OPTIMAL WATERFILLING IN VECTOR MULTIPLE ACCESS CHANNELS

by

Pramod Viswanath, David N. C. Tse
And Venkat Anantharam

Memorandum No. UCB/ERL M99/54

6 November 1999

ASYMPTOTICALLY OPTIMAL WATERFILLING IN VECTOR MULTIPLE ACCESS CHANNELS

by

Pramod Viswanath, David N. C. Tse and Venkat Anantharam

Memorandum No. UCB/ERL M99/54

6 November 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Asymptotically Optimal Waterfilling in Vector Multiple Access Channels*

Pramod Viswanath, David N.C. Tse †and Venkat Anantharam ‡

November 6, 1999

Abstract

Dynamic resource allocation is an important means to increase the *sum capacity* of fading multi access channels. In this paper we consider *vector multiaccess channels* (channels where each user has multiple *degrees of freedom*) and study the effect of power allocation as a function of the channel state on the *sum capacity* (or *spectral efficiency*) defined as the maximum sum of rates of users per unit processing gain at which the users can jointly transmit reliably, in an information theoretic sense, assuming random directions of received signal. Direct sequence code division multiple access (DS-CDMA) systems and multiple access systems with multiple antennas at the receiver are two systems that fall under the purview of our model. Our main result is the identification of a simple dynamic power allocation scheme that is optimal in a large system, i.e., with a large number of users and a correspondingly large number of degrees of freedom, for both the ergodic and non-ergodic models. A key feature of this policy is that, for any user, it depends on the instantaneous amplitude of channel state of that user alone and the structure of the policy is "waterfilling". In the context of DS-CDMA and in the special case of no fading, the asymptotically optimal power policy of waterfilling simplifies to constant power allocation over all realizations of signature sequences; this result verifies the conjecture made in [27]. We study the behavior of the asymptotically optimal waterfilling policy in various regimes of number of users per unit degree of freedom and signal to noise ratio (SNR). We also generalize this result to *multiple classes*, i.e., the situation when users in different classes have different average power constraints.

Index Terms: CDMA, Multiple Antenna Systems, Sum Capacity, Spectral Efficiency, Linear MMSE Receivers, Power Control, Waterfilling.

*The authors are with the department of EECS, University of California at Berkeley. Email: {pvi,dtse,ananth}@eecs.berkeley.edu

†Research of this author is supported by an NSF CAREER Award under grant NCR-9734090.

‡Research of the first and third author is supported by NSF under grant IRI 97-12131.

1 Introduction

The focus of this paper is multiaccess vector channels; these are multiaccess channels with multiple *degrees of freedom.* Two common examples of such systems are Direct sequence Code Division Multiple Access (DS-CDMA) and a multiple access channel with multiple antennas at the receiver. The number of degree of freedom in DS-CDMA model is the processing gain and in the antenna model it is the number of received antennas at the receiver. The *signal direction* at the receiver of any user in the CDMA model is its *spreading sequence* (assuming flat fading) and in the antenna model it is the vector of path gains from the user to the different antennas at the receiver. A central problem in this vector multiaccess scheme in fading channels is how to carry out power allocation to increase the *spectral efficiency* of the channel. In this paper, we assume that the signal directions of the users are random (but known at both the transmitter and receiver) and study power allocation policies that aim to maximize the rates at which users can reliably communicate (in an *information theoretic* sense). One fundamental performance measure of a multiaccess channel is *sum capacity* (equivalently spectral efficiency), defined as the maximum sum of rates of users per unit degree of freedom at which the users can transmit reliably. Our focus in this paper will be to identify simple power allocation policies that allow users to communicate at rates (these are long term rates averaged over the fading process) such that the sum of rates is arbitrarily close to the Shannon limit.

Allocation of resources (power, bandwidth, bit rates) in the context of specific multiple access schemes such as TDMA, FDMA and CDMA, with the performance criterion typically being the signal to interference ratio of the users at the receiver, is studied in [4, 7, 9, 28, 22]. In the context of information theoretic power control, existing literature focuses mainly on scalar channels. For the single user scenario [6] identifies waterfilling to be the optimal power allocation as a function of the fading state. This allocation maximizes the rate at which the user can communicate reliably, the rate being averaged over the fading process. In the multiuser scenario [10] studies power allocation strategies of the users as a function of the fading state to maximize the sum of rates at which the users can jointly communicate. It is shown there that the power policy that allows users to achieve sum capacity has the property that only the user with the best channel at any time transmits (if at all) with positive power and the users themselves adopt a waterfilling strategy with respect to their fading states. This paper focuses on multiaccess vector fading channels. Our main results can be summarized as follows:

1. In the DS-CDMA model, we assume that the spreading sequences of the users are random and each user experiences independent flat fading. We consider both *long* and *short* signature sequence models: short signature sequences get repeated every symbol interval while many symbols are transmitted over one duration of a long signature sequence. Our main result is the identification of a simple power allocation policy that is asymptotically optimal (the asymptotic is in the regime of a large number of users and correspondingly large processing gain). This policy is *waterfilling* for each user and depends solely on the amplitude of that user's instantaneous fading amplitude. We show that the waterfilling policy is asymptotically optimal for both the long and

the short signature sequence models.

2. In the multiantenna model, we assume i.i.d. frequency flat fading from the users to each of the antennas at the receiver. Our main result is that the constant power allocation policy (this policy transmits a constant power regardless of the realization of fading amplitudes of the users) is asymptotically optimal (the asymptotic is in the number of users and correspondingly large number of antennas at the receiver).

These results are rather surprising from the point of view of the scalar multiaccess channel result in [10] which shows that the spectral efficiency harnessing the *multiuser diversity* by allocating positive power only to the user with the best channel (if at all) can be substantially more than spectral efficiency obtained by allocating constant powers to the users at all fading states. Our results show that if there are sufficiently many degrees of freedom, the gain in spectral efficiency by having multiuser diversity vanishes.

In other related work on multiaccess vector channels, [16] and [29] study the allocation of signature sequences to achieve sum capacity in non fading channels as a function of the average power constraints of the users. In [8] the authors study the sum capacity of CDMA systems with random long signature sequences in non fading channels. In [27] the authors study the sum capacity of CDMA systems with random long signature sequences for a wide variety of receiver structures: optimal joint detection receivers, linear MMSE receivers, matched filter receivers, and decorrelator receivers. They assume that the users are received at the same power and the channel has no fading. In the special case of constant flat fading in the DS-CDMA long signature sequence model, our main result simplifies to constant power allocation over all realizations of signature sequences and fading states; this verifies the conjecture made in [27]. Recent results on information theoretic power control in non-ergodic scalar fading channels are in [2].

In Section 2 we outline the DS-CDMA fading channel model, formulate the problem and precisely state our main results. In Section 3 we heuristically derive the structure of the optimal power allocation strategy and see that it is waterfilling. This section outlines the key ideas in the identification of asymptotic optimality of the waterfilling strategy and allows the more casual reader to gain insight into our result without entering the technicalities required for the formal proof. In Section 4 we develop the mathematical machinery and some preliminary results required for the proof of our main result. In Section 5 we first give the simpler proof for the no fading case and then give the formal proof in the general case of fading channels. In Section 6 we study various regimes of number of users and SNR and analyze the behavior of the optimal policy in those regimes. We also discuss natural extensions when there are different *classes* of users; users in different classes have different average power constraints. In Section 7 we demonstrate our results by simulating the different power allocation strategies and plot the corresponding sum capacities achieved for flat and Rayleigh fading channels under a wide range of loading of users and SNR. In Section 8, we turn to the multiple antenna model and model the path gains from any user to any antenna by i.i.d complex random variables (in the flat Rayleigh fading model, the amplitude of the path gains are independent Rayleigh random variables). We conclude the paper in Section 9 with some summarizing remarks and suggestions for future work.

2 Model, Problem Formulation and Main Results

2.1 Model

We consider a single cell in a symbol synchronous CDMA system and focus on the up link. There are K users in the system and a single receiver. The processing gain is N and represents the number of degrees of freedom of the multiaccess channel. Throughout this paper we assume that $K = \lfloor \alpha N \rfloor$ where α is a fixed positive number. This assumption simplifies the analysis and notation though only $K/N \to \alpha$ as $N \to \infty$ along with some mild restrictions allows us to derive all the asymptotic results obtained (asymptotic in N) in this paper. Following standard notation (see Section 2.1 of [26]), the baseband received signal in one symbol interval can be expressed as

$$Y(n) = \sum_{i=1}^{K} X_i(n) s_i(n) \mathbf{h}_i(n) + W(n) \, . \tag{1}$$

Here the index n represents time and the received signal Y is regarded as a vector in $\mathbb{C}^N$. Here $s_i(n)$ is the signature sequence of user i regarded as a vector in $\mathbb{R}^N$. We consider both long and short signature sequences (short signature sequences get repeated every symbol interval while many symbols are transmitted over one duration of a long signature sequence). Thus in the long signature sequence model $s_i(n)$ is an independent realization for every time n and in the short signature sequence is fixed for all time. We model the signature sequences as having random i.i.d. entries (the choice and relevance of this model are discussed in [22] and [27]). Here $\mathbf{h}_i$ is the complex fading or path gain from user i to the single base station (receiver). We write the amplitude squared of this complex path gain by $h_i \stackrel{\text{def}}{=} \mathbf{h}_i \mathbf{h}_i^*$. Henceforth, we refer to h_i as the path gain and explicitly say "complex path gain" when referring to $\mathbf{h}_i$. The user symbols are represented by the real valued random variables X_i. W is an additive white complex Gaussian process with variance σ^2. Each user has an average power constraint $\bar{p}$. Our assumptions on the path gains $\mathbf{h}_i$ are conventional (see Section 2 of [21] and [6] for example): We assume that $\left\{ \{\mathbf{h}_i(n)\}_{n \in \mathbb{N}} \right\}_{i=1}^{K}$ is a sequence of independent and identically distributed stationary and ergodic processes; let us denote the (common) stationary distribution of the amplitude squared of the complex fading process by F which is limited to a bounded set $\left[0, \bar{h}\right]$ and has a density. We write $\mathbf{H} = (h_1, \dots, h_K)$ a random vector having the same joint stationary distribution as the fading processes $\{h_i(n)\}_n \, , i = 1 \dots K$.

2.2 Problem Formulation

We first consider short signature sequences. Here the signature sequences, once chosen, are fixed and repeated over every symbol interval. We model the signature sequence of user i as $s_i = \frac{1}{\sqrt{N}} (v_{i1}, \dots, v_{iN})^t$ where $\{v_{il}\}$ is a collection of i.i.d. random variables with zero mean, variance 1, and bounded fourth moment. These random variables are independent of the fading processes $\{h_i(n)\}_{n\in\mathbb{N}}$. Both the random variables $\{v_{ij}\}$ and $\{\mathbf{h}_i\}$ live on the same probability space, say (Ω, μ) and we write $\mathbb{E}[X]$ to mean $\int_\Omega X \, d\mu$ for any X in $L^1(\Omega, \mu)$. We use the notation $\mathbb{E}_{\mathbf{H}}$ to indicate that the integration is carried out only over the fading parameters. Formally, $\mathbb{E}_{\mathbf{H}}[X] \stackrel{\text{def}}{=} \mathbb{E}[X \mid s_1, \ldots, s_K]$.

Conditioned on one sample point or realization of signature sequences, say $s_1, \ldots, s_K$ (we write $S = [s_1, \ldots, s_K]$) the channel model in (1) becomes

$$Y(n) = \sum_{i=1}^{K} X_i(n)\mathbf{h}_i(n)s_i + W(n) . \tag{2}$$

We assume that all the signature sequences (once chosen) are known to both the receiver and all the users. We also assume that the receiver has perfect side information i.e. has perfect knowledge of the fading gains at each channel use. For the situation when the transmitter has no knowledge of the fading gains and the signature sequences are fixed to be $s_1, \ldots, s_K$, the sum capacity of the multiple access channel (MAC) in (2) is:

$$\frac{1}{2N}\mathbb{E}_{\mathbf{H}}\left[\log\det\left(I + \bar{p}\sigma^{-2}\sum_{i=1}^{K} h_i s_i s_i^t\right)\right] . \tag{3}$$

The capacity region for single degree of freedom fading channels with no information of the fading state at the transmitter is in [17] and the intuitive idea behind the proof is in [5] so we omit the proof of (3).

Our interest is in the situation when the transmitter also has perfect knowledge of fading gains. In practice, this knowledge is obtained by the receiver measuring the channels and feeding back the information to the transmitters (users). Implicit in this model is the assumption that the channel varies much more slowly than the data rate, so that the tracking of the channel variations can be done accurately and the number of bits required for feedback is negligible compared to that required for transmitting information. By a power allocation policy, we mean a function from the fading states and signature sequences of the users to the non-negative reals. We let

$$\mathcal{P}_i : (h_1, \ldots, h_K, S) \mapsto \mathbb{R}^+$$

denote a power allocation policy for user i and call the tuple $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_K)$ a power allocation policy. We say that the power allocation policy is *feasible* if for *every* realization of the signature sequences the average power allocated to each user (over the fading process of the users) is no more than $\bar{p}$. Formally, the set of feasible allocations for a fixed realization of signature sequences S is

$$\mathcal{F}_1(S) \stackrel{\text{def}}{=} \{(\mathcal{P}_1, \ldots, \mathcal{P}_K) : \mathbb{E}_{\mathbf{H}}[\mathcal{P}_i(h_1, \ldots, h_K, S)] \leq \bar{p} \; \forall i = 1 \ldots K\} .$$

Now, for every power allocation policy $\mathcal{P} \in \mathcal{F}_1(S)$, define the quantity

$$\mathcal{C}_{sum}(\mathcal{P}, S) \stackrel{\text{def}}{=} \frac{1}{2N}\mathbb{E}_{\mathbf{H}}\left[\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K} h_i s_i s_i^t \mathcal{P}_i(h_1, \ldots, h_K, S)\right)\right] . \tag{4}$$

Comparing with (3), $\mathcal{C}_{sum}(\mathcal{P}, S)$ can be interpreted as the (random, since it depends on the specific realization of the signature sequences) sum capacity of the MAC with powers allocated according to policy $\mathcal{P}$. The following proposition characterizes sum capacity when transmitters also have perfect knowledge of the fading states.

Proposition 2.1 *The sum capacity of the fading Gaussian vector MAC conditioned on a particular realization of the signature sequences (say S) in (2) when both the users and the receiver can perfectly track the fading state and know the signature sequences is*

$$\begin{aligned} \mathcal{C}_{opt}(S) &= \sup_{\mathcal{P}\in\mathcal{F}_1(S)} \mathcal{C}_{sum}(\mathcal{P}, S) \\ &= \sup_{\mathcal{P}\in\mathcal{F}_1(S)} \frac{1}{2N}\mathbb{E}_{\mathbf{H}}\left[\log\det\left(I+\sum_{i=1}^{K}\sigma^{-2}h_i s_i s_i^t \mathcal{P}_i(h_1,\ldots,h_K,S)\right)\right] \end{aligned} \tag{5}$$

A version of the coding theorem in the above proposition appeared as Theorem 2.1 of [21], another version of the above result for single user fading channels is in [6] and we omit the proof. For general S, no closed form solution to the optimization problem in (5) is known. We discuss algorithmic computations that get close to the solution in Section 7.

In the notation of [20], the MAC with short signature sequences in (2) represents a *non-ergodic* channel and the Shannon capacity of the channel is zero; however small the sum rate the users attempt to communicate at there is a non zero probability that the realized signature sequences will render the channel incapable of supporting the rates reliably. Motivated by the approach in [20] and [12] to such channels, we study the tradeoff between the *supportable rate* and *outage probability.* Formally, the supportable sum rate R at an outage probability a is the maximum sum rate at which the users can communicate reliably with sum rate R for all realization of signature sequences but a set $\mathbf{S}$ whose total probability is less than a. In our notation, the supportable rate $R_a^{*(N)}$ is defined as

$$R_a^{*(N)} = \sup\{R:\ \mathbb{P}[\mathcal{C}_{opt}(S) \geq R] \geq 1-a\} \tag{6}$$

For a family of valid power allocations (power allocations for each realization of the] signature sequences), define the quantity

$$R_a(\{\mathcal{P}_S \in \mathcal{F}_1(S)\}_S) = \sup\{R:\ \mathbb{P}[\mathcal{C}_{sum}(\mathcal{P}_S, S) \geq R] \geq 1-a\} \tag{7}$$

and interpreting it as the supportable rate with outage probability at most a when the power allocation policy for the signature sequence realization S is $\mathcal{P}_S$, we have

$$R_a^{*(N)} = \sup_{\{\mathcal{P}_S\in\mathcal{F}_1(S)\}_S} R_a(\{\mathcal{P}_S\}_S) \tag{8}$$

One of the main aims of this paper is to characterize the family of optimal power allocation policies that "achieves" the maximum supportable rate in (8). Our demonstration of a simple power policy (that does not depend on the actual realization of signature sequences and hence the family of power allocations reduce to a single power allocation) that has the supportable rate asymptotically (in N) equal to the optimal $R_a^{*(N)}$ is one of our main results.

We now turn to long signature sequences. Here, many symbols are transmitted over one period of the signature sequence. Thus, the simplifying assumption that the signature sequences are independent copies of identically distributed sequences for *every* channel use is made. Formally, we define $s_i(n) = \frac{1}{\sqrt{N}}\left(v_{i1}(n), \dots v_{iN}(n)\right)^t$ where $\{v_{il}(n)\}$ are i.i.d. random variables with zero mean and variance 1 and finite fourth moment. We retain the assumption that both the receiver and the transmitters (users) have complete side information, namely they have perfect knowledge of the signature sequences and fading gains at all times. As before, power allocation policies are maps from signature sequences and fading states of the users to the non-negative reals. A policy $\mathcal{P} = (\mathcal{P}_1, \dots, \mathcal{P}_K)$ is feasible if for every user i, the average (over signature sequences and fading states of the users) of $\mathcal{P}_i$ is no more than $\bar{p}$. Let the set of *feasible power allocation policies* be denoted by $\mathcal{F}_2^{(N)}$. Formally we have

$$\mathcal{F}_2^{(N)} = \{\mathcal{P} : \mathbb{E}\left[\mathcal{P}_i(h_1, \dots, h_K, S)\right] \leq \bar{p} \text{ and } \mathcal{P}_i(h_1, \dots, h_K, S) \geq 0 \ a.s. \ \forall i = 1 \dots K\} .$$

The Shannon sum capacity of the MAC (recall the channel model in (1)) with perfect side information at both the transmitters and the receiver is given by

$$\bar{C}_{opt}^{(N)} \stackrel{\text{def}}{=} \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \frac{1}{2N} \mathbb{E}\left[\log\det\left(I + \sum_{i=1}^{K} \sigma^{-2} h_i s_i s_i^t \mathcal{P}_i(h_1, \dots, h_K, S)\right)\right] . \tag{9}$$

This result was observed in Section 3 of [27]. For $\mathcal{P} \in \mathcal{F}_2^{(N)}$ defining the quantity

$$\bar{C}_{sum}(\mathcal{P}) \stackrel{\text{def}}{=} \frac{1}{2N} \mathbb{E}\left[\log\det\left(I + \sum_{i=1}^{K} \sigma^{-2} h_i s_i s_i^t \mathcal{P}_i(h_1, \dots, h_K, S)\right)\right] \tag{10}$$

which can be interpreted as the sum capacity of the fading MAC with random long signature sequences when powers are allocated using the policy $\mathcal{P}$, from (9) and (10) it follows that

$$\bar{C}_{opt}^{(N)} = \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \bar{C}_{sum}(\mathcal{P}) . \tag{11}$$

In the case of long signature sequences, we are interested in characterizing power allocation policies that are optimal in the sense of achieving sum capacity equal to $\bar{C}_{opt}$.

2.3 Main Results

The main focus of this paper is in characterizing optimal power allocation policies in two different settings. First, for the long signature sequence model we are interested in the power allocation policy as a function of the realization of signature sequences and fading states subject to an average power constraint that maximizes sum capacity of the MAC in (1). In the second setting, we wish to characterize a family of power allocations as a function of the fading states of the users subject to an average power constraint that maximize the supportable rate at some fixed outage probability. Our main result is the identification of a simple power allocation policy using which both the supportable rate at some outage probability (when the MAC model involves short signature sequences) and sum capacity (when the MAC model involves long signature sequences) are "close" to the optimal values (defined in (8) and (9) respectively). We state this result formally below. Consider the power allocation policies

$$\mathcal{P}_i^{wf} : (h_1, \dots, h_K, S) \mapsto \left(\frac{1}{\lambda} - \frac{1}{\beta_{wf}^* h_i} \right)^+ \quad \forall i = 1 \dots K \tag{12}$$

where we have used the notation $(x)^+$ to indicate $\max\{x, 0\}$. The constant β_{wf}^* is the limiting signal to interference ratio (SIR) of a unit received power user using the linear MMSE estimator in a large system (large processing gain and correspondingly large number of users) with random signature sequences when all other users are following the power allocation policy above in (12). The formal definition and proof of existence of this quantity is in sections 3 and 4. In (12), the constant (Kuhn-Tucker coefficient) λ is a constant that is chosen such that $\mathbb{E}\left[\mathcal{P}_i^{wf}(h_i)\right] = \bar{p}$. Observe that this policy does not depend on the signature sequences of the users and for any user depends only on the fading state of that user at that instant (in the special case when there is no fading this implies that this policy is a static allocation of powers equal to $\bar{p}$ independent of the signature sequences). This power allocation policy is *waterfilling* and generalizes the strategy of [6] for single user fading channels. To see this generalization, recall the optimal power allocation policy for the single user case from [6]:

$$\mathcal{P}(h) = \left(\frac{1}{\lambda} - \frac{\sigma^2}{h} \right)^+ \tag{13}$$

where $\frac{1}{\sigma^2}$ is the SIR seen by a unit received power user in the system (there is only one user in this scenario). Now the generalization is apparent: β_{wf}^* replaces $1/\sigma^2$.

We show that the waterfilling policy of (12) is a "good" power allocation policy for both the long and the short signature sequence models. We also analyze its behavior in various regimes of the number of users per unit processing gain and background noise variance. We enumerate our main results below. We emphasize that these results are true for *any* distribution of the random variables v_{ij} that satisfies the property of zero mean, unit variance and bounded fourth moment and any stationary fading distribution F that has a bounded support and allows a density.

1. Consider the case of long signature sequences. With long signature sequences we show that asymptotically the waterfilling strategy is optimal and identify the gap in sum capacity to be of the order of $N^{-\frac{1}{2}}$ where N is the processing gain of the system. Formally,

$$\bar{C}_{opt}^{(N)} - \bar{C}_{sum}\left(\mathcal{P}^{wf}\right) < O\left(\frac{1}{\sqrt{N}}\right)$$

Note that because of the simplicity of the waterfilling policy, the notation becomes somewhat deceptive: in this equation, $\bar{C}_{sum}\left(\mathcal{P}^{wf}\right)$ does depend on N.

2. Consider now the case of short signature sequences. Our main result in this scenario is

$$R_a\left(\mathcal{P}^{wf}\right) \le R_a^{*(N)} \le \frac{R_a\left(\mathcal{P}^{wf}\right)}{1-a} + o(1)\,.$$

Thus, in a large system the supportable rate using the waterfilling strategy is within a factor $(1-a)$ of the optimal supportable rate. We are interested in very small values of a (typically, a could be 10^{-3} or 10^{-4}) and thus the waterfilling strategy achieves a supportable rate that is close to the optimal rate.

3. For a single user fading channel, it is intuitive (observe the structure of the optimal power allocation policy in (13)) that in high SNR (as $\sigma^2 \to 0$) the loss in sum capacity by using a constant power (equal to $\bar{p}$) allocation policy as compared to the sum capacity by using the optimal waterfilling policy becomes negligible . In the general multiple user scenario we show that the policy (12) at high SNR converges and the limiting policy is the constant power allocation policy for $\alpha \le 1$. Thus the correct extension of the single user high SNR result is that when α (the ratio of users to processing gain) is less than unity, the gain in sum capacity in a large system (large processing gain) over constant power allocation by using an optimal strategy goes to zero at high SNR. On the other hand, there is a strict loss in using constant power allocation when there are more users than the processing gain even in the limit of high SNR. We also give an intuitive explanation of this fact.

4. We have been able to extend our results on the asymptotic optimality of the waterfilling power allocation to the scenario of multiple *classes* in the situation of long signature sequences. Users in different classes have different average power constraints. The asymptotically optimal strategy still has the basic structure of the waterfilling policy (12) but users in different classes have different threshold levels for their waterfilling policies.

3 Heuristic Derivation of the Asymptotically Optimal Power Allocation Strategy

In this section, we first restrict ourselves to long signature sequences channel model and motivate the reason why we can expect asymptotically the waterfilling structure (12) of the optimal power allocation policy. Towards this end, we proceed in the following order: we first review the waterfilling power allocation policy (identified in [6]) for a single user in a (scalar) fading channel. Then, we show the relation of sum capacity to linear minimum mean square error (LMMSE) estimation of users along with successive decoding. We then arrive at a heuristic expression for the optimal power policy in the multiuser scenario.

We begin with the single user, single degree of freedom scenario. Now, the received baseband signal in any channel use is (analogous to (1))

$$Y(n) = \mathrm{h}\,(n)\, X(n) + W(n)$$

where $\{\mathbf{h}(n)\}_{n\in\mathbb{N}}$ is the complex fading process assumed to be stationary and ergodic. As before we denote the amplitude squared process by $\{h_n\}_n$ having a bounded stationary distribution F with a density. W is an additive white complex Gaussian noise process with variance σ^2. We assume that the receiver and the transmitter have perfect channel side information, i.e., the fading gains $\mathbf{h}_n$ are perfectly known to both the transmitter and the receiver. The transmitter has an average power constraint $\bar{p}$. Then, (Theorem 2.1, [6]) the capacity of the channel is

$$\bar{C}_{1user} = \max_{\{\mathcal{P}\geq 0:\mathbb{E}[\mathcal{P}]\leq\bar{p}\}} \frac{1}{2}\mathbb{E}\left[\log\left(1+\frac{h\mathcal{P}(h)}{\sigma^2}\right)\right] \tag{14}$$

and the power allocation that achieves the maximum above is "waterfilling" (refers to the visualization of this scheme):

$$\mathcal{P}^*(h) = \left(\frac{1}{\lambda}-\frac{\sigma^2}{h}\right)^+ \tag{15}$$

where λ is a constant (the Kuhn-Tucker coefficient for the concave function maximization in (14)) that is chosen such that $\mathbb{E}[\mathcal{P}^*(h)] = \bar{p}$. Observe that zero power is transmitted when the fading is below the threshold $h_{wf} \stackrel{\text{def}}{=} \lambda\sigma^2$.

We now turn to the multiuser multiple degrees of freedom scenario. We first restrict our attention to the case when the signature sequences and the fading gains are fixed (to be $s_1,\dots,s_K$ and $h_1,\dots,h_K$ respectively). Let the users have average power constraints $p_1,\dots,p_K$. Then the channel model (1) focusing on one symbol interval is

$$Y = \sum_{i=1}^{K} X_i s_i \mathbf{h}_i + W \tag{16}$$

Sum capacity of this channel was explicitly calculated in (4) as a function of the signature sequences and the user average power constraints as

$$C(p_1,\dots,p_K) = \frac{1}{2N}\log\det\left(I+\sum_{i=1}^{K}\sigma^{-2}s_is_i^tp_ih_i\right). \tag{17}$$

The rate tuples in the capacity region are in general achieved by jointly demodulating the users from the received signal Y (joint typical decoding and joint maximum likelihood estimation are well studied techniques; this classic study is in Chapter 14 of [3]). We focus on the following specific structure of demodulation of the users' symbols from the received signal Y: Fix an ordering of the users. For every symbol interval, following the ordering of the users, users are successively decoded (by estimating the symbols by the LMMSE receiver, and the estimate used to decode that user) and the received signal is stripped off the decoded users. LMMSE) receiver for user i obtains the *optimal* linear estimate (in the sense of minimizing the mean squared error; Chapter 6 of [26] is an excellent reference for this) of the user i symbol X_i from the received vector Y. It was observed in [25] that this scheme allows the users to transmit reliably at a sum rate equal to the sum capacity of the system[1]. We use

[1]In fact, a stronger statement is claimed in [25]: By changing the ordering of the users, this scheme allows the users to transmit reliably at rate tuples corresponding to all the vertices of the capacity region of the channel in (16), by appropriately choosing the ordering of the decoding.

this to interpret an increase in sum capacity by an increase in the power of one user: Let the average power constraint of one user (say user i) be increased by δ. Then the increase in sum capacity (defined in (17)) is

$$C(p_1,\ldots,p_{i-1},p_i+\delta,p_{i+1},\ldots,p_K)-C(p_1,\ldots,p_K)=\frac{1}{2N}\log\left(1+\delta h_i s_i^t\left[\sigma^2 I+\sum_{j=1}^{K}s_j s_j^t p_j h_j\right]^{-1}s_i\right) \tag{18}$$

where we used the matrix inversion lemma $(A+xx^t)^{-1} = A^{-1} - \frac{A^{-1}xx^tA^{-1}}{1+x^tA^{-1}x}$ whenever the terms exist. We can interpret this increase in sum capacity as the rate of a fictitious user (numbered $K+1$) with average power δ, fading gain h_i and signature sequence s_i and is decoded first and then stripped off. The rate achieved by this fictitious user being decoded first is simply

$$R=\frac{1}{2N}\log\left(1+\delta\beta_{K+1}h_i\right)$$

where $\delta\beta_{K+1}h_i$ is the signal to interference ratio (SIR) of the LMMSE estimate of the fictitious user $K+1$. It can be shown that (see Section 2 of [22] or Chapter 6 of [26]) the SIR of the LMMSE estimate of this fictitious $K+1$ user is $\beta_{K+1}\delta h_i$ where

$$\beta_{K+1}=s_i^t\left(\sigma^2 I+\sum_{j=1}^{K}s_j s_j^t p_j h_j\right)^{-1}s_i \tag{19}$$

which is consistent with the expression for the increase in sum capacity in (18).

Recall the expression of the sum capacity $\bar{C}_{opt}^{(N)}$ for the long signature sequence model as an optimization problem in (9):

$$\bar{C}_{opt}^{(N)} \stackrel{\text{def}}{=} \sup_{\mathcal{P}\in\mathcal{F}_2}\frac{1}{2N}\mathbb{E}\left[\log\det\left(I+\sum_{i=1}^{K}\sigma^{-2}h_i s_i s_i^t\mathcal{P}_i\left(h_1,\ldots,h_K,S\right)\right)\right]$$

In Proposition 4.2 we show that this is a concave maximization problem. Thus there exists a Kuhn-Tucker coefficient $\lambda>0$ such that for every realization of $h_1,\ldots,h_K,S$ any positive optimal policy $\mathcal{P}^*$ satisfies the following constraints:

$$\left(\frac{\partial}{\partial p_i}C\left(p_1,\ldots,p_K\right)\right)\left(\mathcal{P}^*\left(h_1,\ldots,h_K,S\right)\right)=\lambda,\ \forall i=1,\ldots,K$$

which can be written as

$$s_i^t\left(\sigma^2 I+\sum_{j=1}^{K}s_j s_j^t h_j\mathcal{P}_j^*\left(h_1,\ldots,h_K,S\right)\right)^{-1}s_i h_i=\lambda,\ \forall i=1\ldots K\,, \tag{20}$$

by using the expression (18) for the increase in the sum capacity by an increase in power of one user in the derivation of (20). For a similar calculation, see Section 3 of [24]. Here λ is the Kuhn-Tucker coefficient (the formal existence and definition is in Proposition 4.4) and is chosen such that the average power constraint of the users is met. Application of the matrix inversion lemma to (20) yields

$$\mathcal{P}_i^*(h_1,\ldots,h_K,S) = \left(\frac{1}{\lambda} - \frac{1}{s_i^t\left(\sigma^2 I + \sum_{j\neq i} s_j s_j^t h_j \mathcal{P}_j^*(h_1,\ldots,h_K,S)\right)^{-1} s_i h_i}\right)^+ \quad \forall i = 1\ldots K \ . \tag{21}$$

Defining

$$\beta_i \stackrel{\text{def}}{=} s_i^t \left(\sigma^2 I + \sum_{j\neq i} s_j s_j^t h_j \mathcal{P}_j^*(h_1,\ldots,h_K,S)\right)^{-1} s_i \tag{22}$$

and observing that $\beta_i h_i \mathcal{P}^*(h_1,\ldots,h_K,S)$ is the (random) SIR of the LMMSE estimate of user i when powers are allocated according to $\mathcal{P}^*$ and substituting in (21), we arrive at the following structure of an optimal power policy:

$$\mathcal{P}_i^* : (h_1,\ldots,h_K,S) \mapsto \left(\frac{1}{\lambda} - \frac{1}{\beta_i h_i}\right)^+ \ . \tag{23}$$

Here β_i is the (random) SIR of the LMMSE estimate of user i when all users are allocating powers optimally. In general we are not aware of a closed form expression for the optimal power allocation in (21). Let us consider the performance of power allocations that have the structure that for any user the policy depends only on the fading gain for that user, i.e., $\mathcal{P}_i$ is of the form $(h_1,\ldots,h_K,S) \mapsto g(h_i)$ for every user i where g is some bounded nonnegative function into the reals. In this situation, Theorem 3.1 of [22] shows that the (random) SIR of any user (say, user 1 to be specific) converges pointwise in a large system. Using our notation we can make this statement precise: β_1 from (22) with $\mathcal{P}_i^*(h_1,\ldots,h_K,S) = g(h_i)$ converges almost surely to $\beta_g^* h_1$ as $N \to \infty$. The positive constant β_g^* depends on α, the background noise variance σ^2, and the function g itself and Theorem 3.1 of [22] identifies β_g^* to be the unique positive solution of a fixed point equation (in general there is no known closed form solution to β_g^*). Thus, in a large system (large N and correspondingly large K) we see that the power allocation

$$\mathcal{P}_i^* : (h_1,\ldots,h_K,S) \stackrel{\text{def}}{=} g_{wf}(h_i) \stackrel{\text{def}}{=} \left(\frac{1}{\lambda} - \frac{1}{\beta_{wf}^* h_i}\right)^+ \tag{24}$$

satisfies the Kuhn-Tucker conditions in (21) asymptotically. Here β_{wf}^* is a positive constant with the following structure: When every user uses a power allocation policy of this form, namely:

$$\mathcal{P}_i : (h_1,\ldots,h_K,S) \mapsto \left(\frac{1}{\lambda} - \frac{1}{\beta h_i}\right)^+$$

for some positive real β (and λ chosen such that the average power, averaged over the fading, is $\bar{p}$), an application of the central result (Theorem 3.1) of [22] shows that the (random) SIR of the LMMSE estimate of any (every) user converges almost surely in a large system to a constant, which we denote by $\tilde{\beta}$. Every choice of β results in a unique asymptotic SIR $\tilde{\beta}$ of the users giving rise to the map $\beta \mapsto \tilde{\beta}$. Since β_{wf}^* denotes the asymptotic SIR of the

LMMSE estimate of any user, it follows that β^*_{wf} must be the fixed point of the map $\beta \mapsto \tilde{\beta}$. Thus, if we assume the existence of the unique fixed point β^*_{wf} and infer (heuristically) that the power policy (24) which asymptotically satisfies the Kuhn-Tucker conditions is close to an optimal power policy, we have heuristically seen the asymptotic structure of an optimal power allocation policy. The nontrivial fact that the map $\beta \mapsto \tilde{\beta}$ has a unique positive fixed point will follow from Lemma 4.10 in Section 4. We also show that there is a simple expression that relates this unique fixed point β^*_{wf} to the corresponding λ and propose a fixed point iteration algorithm to compute the quantities β^*_{wf} and the corresponding λ. In the next section, we develop the mathematical apparatus required to present the formal proof of the asymptotic optimality of the waterfilling power allocation policy we have only heuristically developed in this section.

Thus we have a heuristic derivation of the structure of the asymptotically optimal power allocation policy for the long signature sequence model. Recall the key features of this policy: the policy is independent of the realization of the signature sequences and for each user the policy is waterfilling over the fading process of that user *alone*. We use this structure to show that waterfilling power allocation performs very "close" to the optimal policy even for the short signature sequences model. Towards this end, we make some observations of the limiting sum capacity when using power allocation policies of the type above, i.e., power allocation policies of the form $\mathcal{P}^g_i : (h_1, \ldots, h_K, S) \mapsto g(h_i)$ where g is a non-negative bounded function. We observe that sum capacity with this policy converges *pointwise* in a large system and we make this precise below:

Proposition 3.1

$$\frac{1}{2N} \log\det\left(I + \sum_{i=1}^{K} \sigma^{-2} s_i s_i^t h_i g(h_i)\right) \xrightarrow{a.s., L^1} \bar{C}^g_{sum} \;\; as\ N \to \infty . \tag{25}$$

The proof is found in Appendix A. For the special case when $h_i = 1$ *a.s.* and $g(\cdot) = \bar{p}$ (this is the no fading case with equal received powers for every user), there is a closed form expression for $\bar{C}^g_{sum}$ and (9) of [27] gives the explicit expression (denoting the limiting value as $\bar{C}^s_{sum}$; this is the spectral efficiency with *static* power allocation in the notation of [27]) as:

$$\begin{aligned}\bar{C}^s_{sum}(\bar{p}) &= \frac{\alpha}{2}\log\left(1 + \frac{\bar{p}}{\sigma^2} - \frac{1}{4}\mathcal{F}\left(\frac{\bar{p}}{\sigma^2}, \alpha\right)\right) \\ &+ \frac{1}{2}\log\left(1 + \frac{\bar{p}\alpha}{\sigma^2} - \frac{1}{4}\mathcal{F}\left(\frac{\bar{p}}{\sigma^2}, \alpha\right)\right) \\ &- \frac{\log e}{8\bar{p}}\mathcal{F}\left(\frac{\bar{p}}{\sigma^2}, \alpha\right)\end{aligned} \tag{26}$$

where

$$\mathcal{F}(x, z) \stackrel{\text{def}}{=} \left(\sqrt{x\left(1 + \sqrt{z}\right)^2 + 1} - \sqrt{x\left(1 - \sqrt{z}\right)^2 + 1}\right)^2 .$$

In general, there is no known closed form expression for $\bar{C}^g_{sum}$; however [18] gives some expressions to compute $\bar{C}^g_{sum}$. With the power allocation being $\mathcal{P}^g$ recall the supportable rate $R_a(\mathcal{P}^g)$ at outage probability a defined in (7) as the largest rate such that:

$$\mathbb{P}\left[\mathcal{C}_{sum}(\mathcal{P}^g, S) \geq R\right] \geq 1 - a\,.$$

The reader will observe that we have replaced the family of power allocations in (7) by the single power allocation $\mathcal{P}^g$ since $\mathcal{P}^g$ is independent of the realization of the signature sequences. It follows that

$$\begin{aligned} R_a(\mathcal{P}^g) &\geq \bar{C}^g_{sum} - \frac{2\mathbb{E}\left[\mid \mathcal{C}_{sum}(\mathcal{P}^g, S) - \bar{C}^g_{sum} \mid\right]}{a} \\ &\geq \bar{C}^g_{sum} - o(1) \;\; ; \;\; \text{using Proposition 3.1} \end{aligned} \tag{27}$$

Thus the supportable rate using the power allocation $\mathcal{P}^g$is asymptotically close to the limiting sum capacity with power allocation $\mathcal{P}^g$. Combined with the formal result of the asymptotic optimality of the waterfilling strategy, we use this result in Section 5 to show that the water-filling strategy is also close to the optimal power allocation with short signature sequences.

4 Preliminary Material

In this section we introduce some preliminary results and the mathematical background needed for the formal derivation of our main result: asymptotic optimality of the waterfilling strategy. We begin with the scenario of long signature sequences. Since our main focus is on understanding the optimization problem (9) we begin with some simple observations about its structure its solution.

4.1 Properties of Optimal Power Allocations

The optimization problem in (9) is on an infinite dimensional set (a closed ball in a Banach space) of valid power allocations and it is not clear a priori if the supremum in (9) is actually achieved. In this section, we show that the supremum is actually attained and characterize the set of the optimal power allocations. We proceed via a series of propositions.

1. Our first step is to show that the optimization problem in (9) is well defined. Formally we have the following proposition and the proof is in Appendix B.

   Proposition 4.1 *For every N, $\bar{C}^{(N)}_{opt} \leq \alpha K_c$ where K_c is a constant independent of N and α.*

2. We next show that the function $\bar{C}_{sum}(\mathcal{P})$ is concave. Consider the following proposition.

Proposition 4.2 *For every deterministic $h_1, \ldots, h_K$ and S, the map from the positive orthant in $\mathbb{R}^K$ to the non-negative reals*

$$C : (p_1, \ldots, p_K) \mapsto \frac{1}{2N}\left[\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K} p_i h_i s_i s_i^t\right)\right] \qquad (28)$$

is concave. Furthermore, if $\{h_i s_i s_i^t,\ i = 1 \ldots K\}$ are linearly independent, then C is strictly concave.

The proof is in Appendix C. It then follows that

$$\begin{aligned} \bar{C}_{sum} : \mathcal{F}_2^{(N)} &\to \mathbb{R}_+ \\ (\mathcal{P}_1, \ldots, \mathcal{P}_K)(h_1, \ldots, h_K, S) &\mapsto \frac{1}{2N}\mathbb{E}\left[\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K} h_i s_i s_i^t \mathcal{P}_i(h_1, \ldots, h_K, S)\right)\right] \end{aligned}$$

is also concave.

3. We observe that the power allocation policies that are of interest always meet the average power constraint with equality. Formally, we have the following result:

Proposition 4.3

$$\bar{C}_{opt}^{(N)} = \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)} \cap \{\mathbb{E}[\mathcal{P}_i] = \bar{p},\ i=1,\ldots,K\}} \bar{C}_{sum}(\mathcal{P}) \ .$$

The following (elementary) proof provides an operational interpretation of increasing the average power of one user. Consider $\mathcal{P} \in \mathcal{F}_2^{(N)}$ and $\mathbb{E}[\mathcal{P}_1(h_1, \ldots, h_K, S)] = \bar{p} - \delta$ for some positive δ. Consider the power allocation policy $\hat{\mathcal{P}}_1(h_1, \ldots, h_K, S) = \mathcal{P}_1(h_1, \ldots, h_K, S) + \delta$ and $\hat{\mathcal{P}}_i = \mathcal{P}_i^*$ for $i = 2, \ldots, K$. By definition $\hat{\mathcal{P}} \in \mathcal{F}_2^{(N)}$. Then

$$\begin{aligned} \bar{C}_{sum}\left(\hat{\mathcal{P}}\right) &= \frac{1}{2N}\mathbb{E}\left[\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K} h_i s_i s_i^t \hat{\mathcal{P}}_i(h_1, \ldots, h_K, S)\right)\right] \\ &= \frac{1}{2N}\mathbb{E}\left[\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K} s_i s_i^t h_i \mathcal{P}_i^*(h_1, \ldots, h_K, S) + \sigma^{-2}\delta h_1 s_1 s_1^t\right)\right] \\ &= \bar{C}_{sum}(\mathcal{P}) + \frac{1}{2N}\mathbb{E}\left[\log\left(1 + \frac{\delta\beta_1(\mathcal{P}^*)}{1 + \beta_1(\mathcal{P}) h_1 \mathcal{P}(h_1, \ldots, h_K, S)}\right)\right] \qquad (29) \\ &> \bar{C}_{sum}(\mathcal{P}) \ . \end{aligned}$$

Here $\beta_1(\mathcal{P}) h_1 \mathcal{P}(h_1, \ldots, h_K, S)$ is the (random) SIR of the LMMSE estimate of user 1 when all users are using the power policy $\mathcal{P}$ (an explicit expression for $\beta_1(\mathcal{P})$ is in (39)) and (29) follows from the matrix inversion lemma (as in (18)). Thus, the sum capacity can always be increased by defining a power allocation policy that is pointwise bigger and meets the average power constraint with equality and the proof of the proposition is complete. □

4. The following proposition allows us to use Lagrange multipliers in this maximization of a concave function. N is fixed below.

Proposition 4.4 *There exists (a Kuhn-Tucker coefficient) $\lambda > 0$ such that*

$$\bar{C}_{opt}^{(N)} = \sup_{\mathcal{P}\in\mathcal{F}_0^{(N)}} \left\{ \bar{C}_{sum}(\mathcal{P}) - \frac{\lambda}{2N}\sum_{i=1}^{K} \left(\mathbb{E}\left[\mathcal{P}_i(h_1,\ldots,h_K,S)\right] - \bar{p}\right) \right\} . \tag{30}$$

where

$$\mathcal{F}_0^{(N)} \stackrel{\text{def}}{=} \left\{ \mathcal{P} : \mathcal{P}_i \geq 0 \text{ and } \mathcal{P} \in L^1(h_1,\ldots,h_k,S)\ \forall i = 1\ldots K \right\} . \tag{31}$$

This claim is completely standard for maximization of concave functions in finite dimensions. However $\mathcal{F}_2^{(N)}$ is infinite dimensional and hence this claim needs a formal proof, which is supplied in Appendix D.

5. We now use the previous propositions to show that the supremum in the definition of $\bar{C}_{opt}^{(N)}$ in (9) is actually achieved by a valid power allocation policy. We state this formally in the following proposition and also identify the structure of this optimal power allocation policy. The problem size N is fixed below and the proof is in Appendix E.

Proposition 4.5 *There exists a power allocation policy $\mathcal{P}^* \in \mathcal{F}_2^{(N)}$ such that $\bar{C}_{opt}^{(N)} = \bar{C}_{sum}(\mathcal{P}^*)$. Furthermore, for almost every realization of $h_1,\ldots,h_K$ and S, the optimal power allocation for this realization, denoted by $p_i^* \stackrel{\text{def}}{=} \mathcal{P}_i^*(h_1,\ldots,h_K,S)$, $i = 1\ldots K$, satisfies the equations*

$$p_i^* = \left(\frac{1}{\lambda} - \frac{1}{s_i^t\left(\sigma^2 I + \sum_{j\neq i} s_j s_j^t h_j p_j^*\right)^{-1} s_i h_i} \right)^+ \quad \forall i = 1\ldots K \tag{32}$$

where λ is the same as that given in Proposition 4.4.

6. It is clear from the symmetry in the problem that the optimal power policies $\mathcal{P}_1^*,\ldots,\mathcal{P}_K^*$ are symmetric with respect to the signature sequences and the fading gains. One simple symmetry is given by Proposition 4.3 which allows us to write

$$\mathbb{E}\left[\mathcal{P}_i^*(h_1,\ldots,h_K,S)\right] = \bar{p},\ \forall i = 1\ldots K . \tag{33}$$

Another type of symmetry is in the formal statement below:

Proposition 4.6 *Let $\mathcal{P}^*$ achieve the maximum in (9). Then for every permutation $\sigma \in S_K$ and $\forall i = 1\ldots K$*

$$\mathcal{P}_i^*(h_1,\ldots,h_K,s_1,\ldots,s_K) = \mathcal{P}_{\sigma(i)}^*\left(h_{\sigma(1)},\ldots,h_{\sigma(K)},s_{\sigma(1)},\ldots,s_{\sigma(K)}\right) \ a.s. \tag{34}$$

Proof For every permutation $\sigma \in S_K$ denote

$$\mathcal{P}_i^{\sigma}(h_1,\ldots,h_K,s_1,\ldots,s_K) \stackrel{\text{def}}{=} \mathcal{P}^*_{\sigma(i)}\left(h_{\sigma(1)},\ldots,h_{\sigma(K)},s_{\sigma(1)},\ldots,s_{\sigma(K)}\right).$$

Now,

$$\begin{aligned}\bar{C}_{sum}(\mathcal{P}^{\sigma}) &= \frac{1}{2N}\mathbb{E}\left[\log\det\left(I+\sum_{i=1}^{K} h_i s_i s_i^t \mathcal{P}_i^{\sigma}(h_1,\ldots,h_K,S)\right)\right] \quad (35)\\ &= \bar{C}_{sum}(\mathcal{P}^*) \quad (36)\end{aligned}$$

where (36) follows from the observation that the random variables are permuted (by σ) in (35) and by hypothesis that $h_1,\ldots,h_K$ are exchangeable and $s_1,\ldots,s_K$ independent and identically distributed. Since $\mathcal{P}^*$ is the unique maximizer of (9), we have

$$\mathcal{P}_i^*(h_1,\ldots,h_K,s_1,\ldots,s_K) = \mathcal{P}^*_{\sigma(i)}\left(h_{\sigma(1)},\ldots,h_{\sigma(K)},s_{\sigma(1)},\ldots,s_{\sigma(K)}\right) \quad a.s.$$

which completes the proof of the proposition. □

7. From the structure of the optimal power allocation policy in (32), it follows that the allocations are bounded from above. We need the following technical result that shows that the allocations are *uniformly* bounded from above (uniform in N).

Theorem 4.7 *Let $\mathcal{P}^*$ achieve the maximum in (9). Then*

$$\mathcal{P}_i^*(h_1,\ldots,h_K,S) \le K_p \quad a.s.$$

where K_p is some universal constant (that depends on the fading statistics, α and $\bar{p}$).

This theorem is proved in Appendix F. Using this, sum capacity can be written as

$$\bar{C}_{opt}^{(N)} = \max_{\mathcal{P}\in\mathcal{F}_3^{(N)}} \bar{C}_{sum}(\mathcal{P}) \ . \quad (37)$$

Here

$$\mathcal{F}_3^{(N)} = \left\{\mathcal{P}: \begin{array}{l}\mathcal{P} \text{ satisfies properties (33) and (34)}\\ \mathcal{P}_i(h_1,\ldots,h_K,S)\in[0,K_p] \ a.s.\ \forall i=1\ldots K\end{array}\right\}. \quad (38)$$

4.2 Limiting SIR of LMMSE estimates

In this section we review some recent results about the asymptotic behavior of SIR of the LMMSE estimate in a random spreading environment. Fix a power allocation policy $\mathcal{P} \in \mathcal{F}_3^{(N)}$. Associated with the LMMSE estimate of user i symbol X_i (estimated from the received signal Y) is the performance measure signal to interference ratio (SIR) defined as the ratio of the power of the signal to the power of the interference in the estimate. Recalling (19), we have that the (random) SIR of the LMMSE estimate of user i is $\beta_i(\mathcal{P})\,h_i\mathcal{P}_i(h_1,\ldots,h_K,S)$ where

$$\beta_i(\mathcal{P}) = s_i^t\left[\sigma^2 I + \sum_{j\neq i} h_j s_j s_j^t \mathcal{P}_j(h_1,\ldots,h_K,S)\right]^{-1} s_i \quad \forall i = 1\ldots K \tag{39}$$

The SIR is random since it depends on the particular realizations of the signature sequences and the fading. We further focus our attention on the following class of power allocation policies: $\mathcal{P}$ is independent of the signature sequences and has the structure

$$\mathcal{P}_i(h_1,\ldots,h_K,S) \stackrel{\text{def}}{=} \mathcal{P}_i^g(h_1,\ldots,h_K,S) \mapsto g(h_i)$$

for each $i = 1\ldots K$ where g is a non-negative function bounded by K_g. Denote the corresponding SIRs of the LMMSE estimates (defined in (39)) of the users as $\beta_1(\mathcal{P}^g),\ldots,\beta_K(\mathcal{P}^g)$. Then it is straightforward to see that

$$\text{The random variables } \beta_1(\mathcal{P}^g),\ldots,\beta_K(\mathcal{P}^g) \text{ are identically distributed.} \tag{40}$$

In a large system, the central result of [22] shows that the (random) SIRs converge almost surely to a deterministic constant. Focusing on user 1 alone (without loss of generality), we have the following formal result.

Lemma 4.8 (Theorem 3.1,[22])

.

$$\beta_1(\mathcal{P}^g) \xrightarrow{a.s.} \beta_g^* \quad \text{as } N \to \infty \tag{41}$$

where β_g^* is the unique positive solution to the integral fixed point equation

$$\sigma^2\beta = 1 - \alpha\int_0^{\bar{h}} \frac{\beta h g(h)}{1+\beta h g(h)}\,dF(h). \tag{42}$$

Recall that F is the (same) marginal distribution of the fading gains $h_1,\ldots,h_K$. Convergence of $\beta_1(\mathcal{P}^g)$ in measure first appeared as Theorem 3.1 in [22] and the pointwise convergence (a natural extension) follows as a consequence of the main result in [19] which shows that the empirical distribution of the eigenvalues of the matrix $\sum_{i=1}^K s_i s_i^t h_i$ almost surely converge in distribution to a nonrandom limit.

To get a better feel for this result, consider the special case when there is no fading (we just take $h_i = 1$ *a.s.*) and $g = \bar{p}$ *a.s.* . Let us denote this static power allocation policy by $\bar{\mathcal{P}}$. Then Lemma 4.8 particularizes to

$$\beta_1(\bar{\mathcal{P}}) \xrightarrow{a.s.} \beta^*(\bar{p},\alpha) \tag{43}$$

It is easily verified from (42) that $\beta^*(\bar{p},\alpha)$ is the unique positive solution of the fixed point equation in β

$$\sigma^2\beta = 1 - \alpha\frac{\bar{p}\beta}{1+\bar{p}\beta} \tag{44}$$

and hence $\beta^*(\bar{p},\alpha)$ is the positive root of the quadratic equation (in β)

$$\sigma^2\beta^2\bar{p} + \beta\left(\sigma^2 + \bar{p}(\alpha-1)\right) - 1 = 0 \tag{45}$$

and can be explicitly written out as:

$$\beta^*(\bar{p},\alpha) = \frac{1-\alpha}{2\sigma^2} - \frac{1}{2\bar{p}} + \sqrt{\frac{(1-\alpha)^2}{4\sigma^4} + \frac{1+\alpha}{2\bar{p}\sigma^2} + \frac{1}{4\bar{p}^2}} \tag{46}$$

4.3 Variations around the mean of limiting SIR

For the power allocation policy $\mathcal{P}^g$, we saw in Lemma 4.8 that the SIR of any user converges pointwise. Our first simple observation is that this convergence holds in L^2 also.

$$\begin{aligned}
\beta_1(\mathcal{P}^g) &= s_1^t\left(\sigma^2 I + \sum_{j\neq 1} s_j s_j^t h_j g(h_j)\right)^{-1} s_1; \text{ from (39)} && (47)\\
&\leq \sigma^{-2} s_1^t s_1 \ a.s. && (48)\\
&= \frac{\sigma^{-2}}{N}\sum_{i=1}^{K} v_{1i}^2 \\
\mathbb{E}\left[(\beta_1(\mathcal{P}^g))^2\right] &\leq C_1\sigma^{-4} && (49)
\end{aligned}$$

where C_1 is a constant independent of N. It now follows from (41) and the dominated convergence theorem that

$$\beta_1(\mathcal{P}^g) \xrightarrow{L^2} \beta_g^* . \tag{50}$$

The following result investigates the variation around the mean of the limiting SIR (without loss of generality, focusing only on user 1):

Lemma 4.9

$$\mathbb{E}\left[\left(\beta_1(\mathcal{P}^g) - \beta_g^*\right)^2\right] < \frac{C_2^2}{N} \tag{51}$$

where C_2 is some constant independent of N.

The lemma is proved in Appendix G.

4.4 Existence of β_{wf}^*

In Section 3, we derived heuristically the asymptotic structure of the optimal power allocation policy to be (from (24))

$$\mathcal{P}_i^* : (h_1,\ldots,h_K,S) \mapsto g_{wf}(h_i) \stackrel{\text{def}}{=} \left(\frac{1}{\lambda} - \frac{1}{\beta_{wf}^* h_i}\right)^+ \tag{52}$$

where β^*_{wf} was the limiting SIR of the LMMSE estimate when users adopt the above power allocation policy and λ is a constant chosen such that the average of the power allocation (average of the fading statistics) is equal to $\bar{p}$. We now prove the existence of this quantity β^*_{wf}. From (42), β^*_{wf} is the unique positive solution to the integral fixed point equation:

$$\begin{aligned} \sigma^2\beta^*_{wf} &= 1-\alpha\int_0^{\bar{h}} \frac{\beta^*_{wf}h\left(\frac{1}{\lambda}-\frac{1}{\beta^*_{wf}h}\right)^+ dF(h)}{1+\beta^*_{wf}h\left(\frac{1}{\lambda}-\frac{1}{\beta^*_{wf}h}\right)^+} \\ &= 1-\alpha\int_{\lambda/\beta^*_{wf}}^{\bar{h}}\left(1-\frac{\lambda}{\beta^*_{wf}h}\right)dF(h) \end{aligned} \tag{53}$$

Furthermore, by the average power constraint of $\bar{p}$ on the power allocation in (52), we have another equation relating λ and β^*_{wf}. Denoting the ratio $\frac{\lambda}{\beta^*_{wf}}$ by h_{thr}, the fading threshold level below which no power is transmitted, we see that the average power constraint in our notation yields:

$$h_{thr} = \frac{HM(h_{thr})}{1+\frac{\bar{p}HM(h_{thr})}{1-F(h_{thr})}\beta^*_{wf}} \tag{54}$$

where $HM(h)$ is the "harmonic mean at the level h" defined as

$$HM(h) = (1-F(h))\left(\int_h^{\bar{h}}\frac{dF(\tilde{h})}{\tilde{h}}\right)^{-1}. \tag{55}$$

Observe that $HM(h) > h,\ \forall h \in \left[0,\bar{h}\right]$. As observed by the authors who first derived the fixed point integral equation (42) for the SIR of the LMMSE estimate in [22], in general there is no closed form solution to this fixed point equation. For the special case when there was no fading and all the users were received with the same power, the fixed point equation for the SIR became quadratic (given in (44)) and there is an explicit solution (given by (45)). However, when the power allocation is in this special form we are able to obtain a closed form solution to β^*_{wf}. Continuing from (53), we have

$$\sigma^2\beta^*_{wf} = 1-\alpha\int_{h_{thr}}^{\bar{h}}\left(1-\frac{h_{thr}}{h}\right)dF(h) \tag{56}$$

$$= 1-\alpha(1-F(h_{thr}))+\frac{\alpha h_{thr}(1-F(h_{thr}))}{HM(h_{thr})} \tag{57}$$

$$= 1-\alpha(1-F(h_{thr}))+\alpha\cdot\frac{(1-F(h_{thr}))^2}{1-F(h_{thr})+\beta^*_{wf}\bar{p}HM(h_{thr})} \tag{58}$$

where (56) uses the definition of h_{thr} as $\frac{\lambda}{\beta^*_{wf}}$, (57) follows from our notation of harmonic mean in (55) and we used (54) in (58). Comparing (58) with (44) we see that β^*_{wf} is equal to $\beta^*\left(\frac{\bar{p}HM(h_{thr})}{1-F(h_{thr})}, \alpha(1-F(h_{thr}))\right)$, the SIR of the LMMSE estimate of a unit power user in a large system with all other users received at constant power equal to $\frac{\bar{p}HM(h_{thr})}{1-F(h_{thr})}$ and number of users per unit processing gain equal to $\alpha(1-F(h_{thr}))$. Thus β^*_{wf} has an explicit form as given in (46). Substituting this structure of β^*_{wf} in (54), we see that our claim is verified if we can show the existence of a solution h_{thr} satisfying (54). Denoting

$$\mathcal{K}(h) = \frac{HM(h)}{1 + \frac{\bar{p}HM(h)}{1-F(h)}\ \beta^*\left(\frac{\bar{p}HM(h)}{1-F(h)}, \alpha(1 - F(h))\right)},\ h \in \left(0, \bar{h}\right) \tag{59}$$

we have to show that h_{thr} is the unique positive fixed point of $\mathcal{K}$. The following lemma investigates the fixed points of $\mathcal{K}$ and identifies a convergent fixed point iteration scheme; the proof is found in Appendix H.

Lemma 4.10 *$\mathcal{K}$ has a unique positive fixed point h_{thr}. Furthermore, a fixed point iteration of $\mathcal{K}$ from small enough h converges to h_{thr}.*

5 Proof of Main Result

In this section we formally prove the asymptotic optimality of the waterfilling power allocation strategy heuristically identified earlier. We first focus on the scenario when there is no fading ($h_i = 1$ *a.s.*) and begin with the long signature sequences channel. For this scenario, the authors in [27] conjectured that asymptotically the optimal power allocation policy is to allocate equal powers to all users independent of signature sequences. The waterfilling strategy identified earlier indeed simplifies to the constant power allocation when there is no fading. Our first main result is to show the asymptotic optimality of constant power allocation formally and furthermore identify the loss in sum capacity to be of the order of $\sqrt{N}$. Recall our notation that the policy of static allocation of equal powers is denoted by $\bar{\mathcal{P}}$.

Theorem 5.1 *For the no fading scenario,*

$$\limsup_{N\to\infty} \sqrt{N}\left(\bar{C}^{(N)}_{opt} - \bar{C}_{sum}\left(\bar{\mathcal{P}}\right)\right) < \infty$$

Define the function L (the Lagrangian) as

$$L : \mathcal{P} \mapsto \bar{C}_{sum}(\mathcal{P}) - \left(\frac{1}{2N}\frac{\beta^*}{1+\beta^*\bar{p}}\right)\sum_{i=1}^{K}\mathbb{E}\left[\mathcal{P}_i(S) - \bar{p}\right] .$$

Here $\mathcal{P}$ is any power allocation such that $\mathcal{P}_i \geq 0 a.s.$ and β^* is the positive root of the quadratic in (45). Observe that L is just the sum of $\bar{C}_{sum}$ and a linear functional and hence is also a strictly concave function. Furthermore $L(\mathcal{P}) = \bar{C}_{sum}(\mathcal{P})$ over $\mathcal{F}_3^{(N)}$. Recall our earlier notation (from Section 4) that $\mathcal{P}^*$ is the power policy that maximizes $\bar{C}_{sum}$ over $\mathcal{F}_3^{(N)}$. Fix a realization of signature sequences S. Let (recall earlier notation from (126) and (128))

$$p_i^* \stackrel{\text{def}}{=} \mathcal{P}_i^*(S) \quad \forall i = 1 \ldots K .$$

Then, by the concavity of the map C (Proposition 4.2) we arrive at

$$\begin{aligned} C(p_1^*, \ldots, p_K^*) - C(\bar{p}, \ldots, \bar{p}) &\leq \frac{1}{2N} \sum_{i=1}^{K} s_i^t \left(\sigma^2 I + \sum_{j=1}^{K} s_j s_j^t \bar{p} \right)^{-1} s_i (p_i^* - \bar{p}) & (60) \\ &= \frac{1}{2N} \sum_{i=1}^{K} \frac{\beta_i(\bar{\mathcal{P}})}{1 + \beta_i(\bar{\mathcal{P}})\bar{p}} (p_i^* - \bar{p}) & (61) \end{aligned}$$

where we used (18) to arrive at (60) and (61) is arrived at by using the definition of $\beta_i(\bar{\mathcal{P}})$ from (39) (the quantity $\beta_i(\bar{\mathcal{P}})\bar{p}$ denotes (the (random) SIR of the LMMSE estimate of user i when all the users are transmitting at constant power equal to $\bar{p}$) and the matrix inversion lemma. Integrating (61), we arrive at

$$\begin{aligned} L(\mathcal{P}^*) - L(\bar{\mathcal{P}}) &\leq \frac{1}{2N} \sum_{i=1}^{K} \mathbb{E}\left[\left(\frac{\beta_i(\bar{\mathcal{P}})}{1 + \beta_i(\bar{\mathcal{P}})\bar{p}} - \frac{\beta^*}{1 + \beta^* \bar{p}} \right) (\mathcal{P}_i^*(S) - \bar{p}) \right] & (62) \\ &\leq \frac{K_p}{2N} \sum_{i=1}^{K} \mathbb{E}\left[\left(\frac{\beta_i(\bar{\mathcal{P}})}{1 + \beta_i(\bar{\mathcal{P}})\bar{p}} - \frac{\beta^*}{1 + \beta^* \bar{p}} \right) \right] & (63) \\ &\leq \frac{K K_p}{2N} \mathbb{E}\left[\left(\frac{\beta_1(\bar{\mathcal{P}})}{1 + \beta_1(\bar{\mathcal{P}})\bar{p}} - \frac{\beta^*}{1 + \beta^* \bar{p}} \right) \right] & (64) \\ &\leq \frac{K K_p}{2N} \mathbb{E}\left[\mid \beta_1(\bar{\mathcal{P}}) - \beta^* \mid \right] & (65) \\ &\leq \frac{\alpha K_p C_2}{2\sqrt{N}} & (66) \end{aligned}$$

where (63) follows from Theorem 4.7, (64) from (40), (65) follows from the fact that the map $x \mapsto \frac{x}{1+x}$ is contractive and (66) follows from (43) and Lemma 4.9. Recalling the observation that $L(\mathcal{P}^*) = \bar{C}_{sum}(\mathcal{P}^*)$ and $L(\bar{\mathcal{P}}) = \bar{C}_{sum}(\bar{\mathcal{P}})$, the theorem follows. □

We now focus on the short signature sequences model while retaining the assumption of no fading. The supportable rate at outage probability a with static power allocation

$$\begin{aligned} R_a^{*(N)} &\leq \frac{\mathbb{E}[C_{opt}(S)]}{1-a} & (67) \\ &\leq \frac{\bar{C}_{opt}^{(N)}}{1-a} & (68) \\ &\leq \frac{\bar{C}_{sum}(\bar{\mathcal{P}}) + O\left(\frac{1}{\sqrt{N}}\right)}{1-a} & (69) \end{aligned}$$

$$\leq \frac{\bar{C}^s_{sum} + o(1) + O\left(\frac{1}{\sqrt{N}}\right)}{1-a} \tag{70}$$

$$\leq \frac{R_a\left(\bar{\mathcal{P}}\right) + O\left(\frac{1}{\sqrt{N}}\right) + o(1)}{1-a} \tag{71}$$

where (67) follows from definition of R_a^* in (6) and the Markov inequality, (68) is from the definition of $\bar{C}_{opt}$ in (9) and the fact that the power allocation policy $\mathcal{P}$ defined as (and so as to be measurable in S)

$$\mathcal{P} : (h_1, \dots, h_K, S) \mapsto \mathcal{P}_S(h_1, \dots, h_K) \quad \text{for some } \mathcal{P}_S \in \mathcal{F}_1(S)$$

belongs to $\mathcal{F}_2^{(N)}$ $\forall S$, we used Theorem 5.1 in (69), (70) comes from Proposition 3.1, and finally (71) follows from (27). Thus we have

$$R_a\left(\bar{\mathcal{P}}\right) \leq R_a^{*(N)} \leq \frac{R_a\left(\bar{\mathcal{P}}\right)}{1-a} + o(1). \tag{72}$$

Hence in a large system the static constant power allocation fetches supportable rate which is optimal up to a factor $(1-a)$. Typical values of a that are of interest in this framework are very small and thus the supportable rate with static power allocation is very close to the optimal supportable rate for large N.

We now turn to the general scenario with fading and first consider the long signature sequences model. The proof of the asymptotic optimality of the waterfilling strategy is subtler than in the no fading situation but the essential ideas are contained in the proof of the no fading situation and the heuristic derivation of the waterfilling strategy. Let us denote the waterfilling strategy of (52) by

$$\mathcal{P}_i^{wf} : (h_1, \dots, h_K, S) \mapsto g_{wf}(h_i) \stackrel{\text{def}}{=} \frac{1}{\beta^*_{wf}} \left(\frac{1}{h_{thr}} - \frac{1}{h_i}\right)^+. \tag{73}$$

Recall that $\beta^*_{wf} = \beta^*\left(\frac{\bar{p}HM(h_{thr})}{1-F(h_{thr})}, \alpha\left(1-F\left(h_{thr}\right)\right)\right)$ and the threshold h_{thr} below which no power is transmitted is the unique fixed point of $\mathcal{K}$ in (59). The formal statement of the asymptotic optimality of the waterfilling policy $\mathcal{P}^{wf}$ that also identifies the order of the loss in sum capacity is below:

Theorem 5.2

$$\limsup_{N\to\infty} \sqrt{N}\left(\bar{C}^{(N)}_{opt} - \bar{C}_{sum}\left(\mathcal{P}^{wf}\right)\right) < \infty$$

Proof Define the function L (the Lagrangian) as

$$L : \mathcal{P} \mapsto \bar{C}_{sum}(\mathcal{P}) - \frac{\beta^*_{wf} h_{thr}}{2N} \sum_{i=1}^{K} \mathbb{E}\left[\mathcal{P}_i(h_1, \dots, h_K, S) - \bar{p}\right]$$
$$+ \frac{1}{2N} \sum_{i=1}^{K} \mathbb{E}\left[1_{\{h_i \leq h_{thr}\}} \left(\beta^*_{wf} h_{thr} - h_i \beta_i\left(\mathcal{P}^{wf}\right)\right) \mathcal{P}_i(h_1, \dots, h_K, S)\right] \tag{74}$$

where $\mathcal{P}$ is any power allocation such that $\mathcal{P}_i \geq 0$ *a.s.*. Observe that L is just the sum of $\bar{C}_{sum}$ and a linear functional and hence is also a strictly concave function. Recall our notation from Section 4 of $\mathcal{P}^*$ that maximizes $\bar{C}_{sum}$ over $\mathcal{F}_3^{(N)}$. We proceed by the following steps:

1. We show that $L\left(\mathcal{P}^{wf}\right)$ is close to $L\left(\mathcal{P}^*\right)$ for large enough N. Formally,

$$\mid L\left(\mathcal{P}^*\right) - L\left(\mathcal{P}^{wf}\right) \mid < O\left(\frac{1}{\sqrt{N}}\right) . \tag{75}$$

2. We show that $L\left(\mathcal{P}^*\right) \geq \bar{C}_{sum}\left(\mathcal{P}^*\right)$ for large enough N. Formally,

$$\liminf_{N\to\infty}\left(L\left(\mathcal{P}^*\right) - \bar{C}_{sum}\left(\mathcal{P}^*\right)\right) \geq 0 . \tag{76}$$

Combining the observation that $L\left(\mathcal{P}^{wf}\right) = \bar{C}_{sum}\left(\mathcal{P}^{wf}\right)$ with the two steps above proves the theorem. We first show (75) and then (76).

Analogous to (60), for every realization of fading gains $h_1, \ldots, h_K$ and signature sequences S, we have from the concavity of the map C (Proposition 4.2) that

$$C\left(\mathcal{P}^*\left(h_1, \ldots, h_K, S\right)\right) - C\left(\mathcal{P}^{wf}\left(h_1, \ldots, h_K, S\right)\right) \leq \frac{1}{2N}\sum_{i=1}^{K}\frac{\beta_i\left(\mathcal{P}^{wf}\right) h_i\left(\mathcal{P}_i^*\left(h_1, \ldots, h_K, S\right) - \mathcal{P}_i^{wf}\left(h_i\right)\right)}{1+\beta_i\left(\mathcal{P}^{wf}\right) h_i \mathcal{P}_i^{wf}\left(h_i\right)} . \tag{77}$$

In (77) we have emphasized the fact that $\mathcal{P}_i^{wf}$ is only a function of h_i. Using our notation in (73) and integrating (77), we arrive at

$$L\left(\mathcal{P}^*\right) - L\left(\mathcal{P}^{wf}\right) \leq \frac{1}{2N}\sum_{i=1}^{K}\mathbb{E}\left[1_{\{h_i > h_{thr}\}}\left(\frac{\beta_i\left(\mathcal{P}^{wf}\right) h_i}{\left(1+\beta_i\left(\mathcal{P}^{wf}\right) h_i g_{wf}\left(h_i\right)\right)} - \beta_{wf}^* h_{thr}\right) \cdot \left(\mathcal{P}_i^*\left(h_1, \ldots, h_K, S\right) - g_{wf}\left(h_i\right)\right)\right] . \tag{78}$$

In (78) we used the fact (by definition) that $g_{wf}\left(h_i\right) 1_{\{h_i < h_{thr}\}} = 0$. Continuing from (78),

$$L\left(\mathcal{P}^*\right) - L\left(\mathcal{P}^{wf}\right) \leq \frac{K K_p}{2N}\mathbb{E}\left[1_{\{h_1 > h_{thr}\}} \mid \frac{\beta_1\left(\mathcal{P}^{wf}\right) h_1}{\left(1+\beta_1\left(\mathcal{P}^{wf}\right) h_1 g_{wf}\left(h_1\right)\right)} - \beta_{wf}^* h_{thr} \mid\right] \tag{79}$$

where we used (40) and Theorem 4.7. By definition, β_{wf}^* is equal to $\beta_{g_{wf}}^*$ and thus from Lemma 4.8 we have

$$\beta_1\left(\mathcal{P}^{wf}\right) \xrightarrow{a.s.} \beta_{wf}^* \text{ as } N \to \infty .$$

By definition of g_{wf} we get

$$\frac{1_{\{h_1 > h_{thr}\}}\beta_{wf}^* h_1}{1+\beta_{wf}^* h_1 g_{wf}\left(h_1\right)} = \beta_{wf}^* h_{thr} 1_{\{h_1 > h_{thr}\}} . \tag{80}$$

Using the fact that the map $x \mapsto \frac{x}{1+x}$ is contractive, (80) and (79) yield

$$\begin{aligned} L(\mathcal{P}^*) - L\left(\mathcal{P}^{wf}\right) &\leq \frac{\alpha K_p}{2} \mathbb{E}\left[h_1 1_{\{h_1 > h_{thr}\}} \mid \beta_1\left(\mathcal{P}^{wf}\right) - \beta^*_{wf} \mid\right] \\ &\leq \frac{\alpha C_2 K_p \bar{h}}{2\sqrt{N}} \end{aligned} \tag{81}$$

where we used Lemma 4.9 to arrive at (81). We have thus shown (75).

To show (76), fix $\epsilon > 0$. Using Lemma 4.9, we have from a Chebyshev bound

$$\begin{aligned} \mathbb{P}\left[\frac{\beta_1\left(\mathcal{P}^{wf}\right)}{\beta^*_{wf}} \geq 1 + \epsilon\right] &\leq \frac{\mathbb{E}\left[\mid \beta_1\left(\mathcal{P}^{wf}\right) - \beta^*_{wf} \mid^2\right]}{\beta^{*2}_{wf}\epsilon^2} \\ &\leq \left(\frac{C_2}{\beta^*_{wf}\epsilon\sqrt{N}}\right)^2 \end{aligned} \tag{82}$$

Then, using properties (34) and (33) of $\mathcal{P}^*$ and (40) we have

$$L(\mathcal{P}^*) = \bar{C}_{sum}(\mathcal{P}^*) + \frac{K\beta^*_{wf}}{2N}\mathbb{E}\left[1_{\{h_1 \leq h_{thr}\}}\left(h_{thr} - \frac{h_1\beta_1\left(\mathcal{P}^{wf}\right)}{\beta^*_{wf}}\right)\mathcal{P}^*_1(h_1, \ldots, h_K, S)\right]. \tag{83}$$

Consider the case

$$\liminf_{N\to\infty} \mathbb{E}\left[\mathcal{P}^*_1(h_1, \ldots, h_K, S)\, 1_{\{h_1 \leq h_{thr}\}}\right] = 0\,. \tag{84}$$

Using Theorem 4.7, (84) leads to

$$\begin{aligned} \liminf_{N\to\infty} \mathbb{E}\left[\left(\mathcal{P}^*_1(h_1, \ldots, h_K, S)\right)^2 1_{\{h_1 \leq h_{thr}\}}\right] &\leq K_p \liminf_{N\to\infty} \mathbb{E}\left[\mathcal{P}^*_1(h_1, \ldots, h_K, S)\, 1_{\{h_1 \leq h_{thr}\}}\right] \\ &= 0 \end{aligned} \tag{85}$$

Then it follows from (83) that there exists a subsequence $\{N_{i_n}\}_n$ such that

$$\begin{aligned} L(\mathcal{P}^*) - \bar{C}_{sum}(\mathcal{P}^*) &\geq -\frac{\alpha\bar{h}}{2}\mathbb{E}\left[\beta_1\left(\mathcal{P}^{wf}\right)\mathcal{P}^*_1(h_1, \ldots, h_K, S)\, 1_{\{h_1 \leq h_{thr}\}}\right] \\ &\geq -\frac{\alpha\bar{h}\sqrt{C_1}}{2\sigma^2}\mathbb{E}\left[\left(\mathcal{P}^*_1(h_1, \ldots, h_K, S)\right)^2 1_{\{h_1 \leq h_{thr}\}}\right]^{\frac{1}{2}} \quad (86) \\ \lim_{n\to\infty} L\left(\mathcal{P}^{*(N_{i_n})}\right) - \bar{C}_{sum}\left(\mathcal{P}^{*(N_{i_n})}\right) &\geq 0 \end{aligned}$$

where we used the Cauchy-Schwartz inequality and the bound in (49) in arriving at (86) and we have thus shown (76) (the notation of the superscript N in $\mathcal{P}^{*(N)}$ denotes that $\mathcal{P}^{*(N)} \in \mathcal{F}_3^{(N)}$). Now suppose (84) does not hold and hence we have

$$\liminf_{N\to\infty} \mathbb{E}\left[\mathcal{P}^*_1(h_1, \ldots, h_K, S)\, 1_{\{h_1 \leq h_{thr}\}}\right] > 0\,. \tag{87}$$

We evaluate the integral in (83) over the two disjoint sets $\mathcal{A}_1 \stackrel{\text{def}}{=} \left\{\beta_1\left(\mathcal{P}^{wf}\right) \geq \beta^*_{wf}(1+\epsilon)\right\}$ and $\mathcal{A}_2 \stackrel{\text{def}}{=} \left\{\beta_1\left(\mathcal{P}^{wf}\right) < \beta^*_{wf}(1+\epsilon)\right\}$. As usual, $1_{\mathcal{A}_i}$ denotes the indicator function over the set $\mathcal{A}_i$, $i=1,2$. We have

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{P}^*_1(h_1,\ldots,h_K,S)\left(h_{thr}-\frac{h_1\beta_1\left(\mathcal{P}^{wf}\right)}{\beta^*_{wf}}\right)1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_1}\right] \\
&\qquad\qquad \geq -\frac{K_p h_{thr}}{\beta^*_{wf}}\mathbb{E}\left[\beta_1\left(\mathcal{P}^{wf}\right)1_{\mathcal{A}_1}\right] \qquad (88)\\
&\qquad\qquad \geq -\frac{K_p h_{thr}}{\beta^*_{wf}\sigma^2}\mathbb{E}\left[s_1^t s_1 1_{\mathcal{A}_1}\right] \qquad (89)\\
&\qquad\qquad \geq -\frac{K_p h_{thr}\sqrt{C_1}}{\beta^*_{wf}\sigma^2}\left(\frac{C_2}{\epsilon\beta^*_{wf}\sqrt{N}}\right) \qquad (90)
\end{aligned}$$

where we used Theorem 4.7 in (88), (48) to derive (89), and (49) combined with the Chebyshev bound of (82) in arriving at (90). We also have

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{P}^*_1(h_1\ldots,h_K,S)\left(h_{thr}-\frac{h_1\beta_1\left(\mathcal{P}^{wf}\right)}{\beta^*_{wf}}\right)1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_2}\right]\\
\geq\;& \mathbb{E}\left[\mathcal{P}^*_1(h_1,\ldots,h_K,S)\left(h_{thr}-h_1(1+\epsilon)\right)1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_2}\right]\\
=\;& \mathbb{E}\left[\mathcal{P}^*_1(h_1,\ldots,h_K,S)\left(h_{thr}-h_1\right)1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_2}\right]-\epsilon\mathbb{E}\left[\mathcal{P}^*_1(h_1,\ldots,h_K,S)h_1 1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_2}\right]\\
\geq\;& \mathbb{E}\left[\mathcal{P}^*_1(h_1,\ldots,h_K,S)\left(h_{thr}-h_1\right)1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_2}\right]-\epsilon h_{thr}K_p\left(1-\frac{C_2^2}{\beta^{*2}_{wf}\epsilon^2 N}\right). \qquad (91)
\end{aligned}$$

From (87), we have

$$\liminf_{N\to\infty}\mathbb{E}\left[\mathcal{P}^*_1(h_1,\ldots,h_K,S)\left(h_{thr}-h_1\right)1_{\{h_1\leq h_{thr}\}}1_{\mathcal{A}_2}\right]>0\,. \qquad (92)$$

Letting $\epsilon=\frac{1}{\log N}$ and combining (90), (91), and (92) we have shown (76) that

$$\liminf_{N\to\infty}\left(L\left(\mathcal{P}^*\right)-\bar{C}_{sum}\left(\mathcal{P}^*\right)\right)>0$$

completing the proof. □

The result regarding short signature sequences is completely identical to the argument given in the situation of no fading. Completely analogous to (72) we have

$$R_a\left(\mathcal{P}^{wf}\right)\leq R_a^{*(N)}\leq\frac{R_a\left(\mathcal{P}^{wf}\right)}{1-a}+o(1).$$

6 Optimal Power Allocation and System Parameters

In this section we study the behavior of the waterfilling power allocation strategy in different regimes of the system parameters. In particular we study the effects of the number of users per unit processing gain α and the variance of the background noise σ^2 on the waterfilling strategy. This exercise allows us to comment on the gain in sum capacity with dynamic power allocation over the constant power allocation strategy. We also generalize our results to the situation of multiple classes: users in different classes have different average power constraints.

6.1 Dependence on the number of users per unit processing gain

Recall the waterfilling power allocation strategy defined in (73):

$$\mathcal{P}_i^{wf} : (h_1, \dots, h_K, S) \mapsto g_{wf}(h_i) \stackrel{\text{def}}{=} \frac{1}{\beta_{wf}^*} \left(\frac{1}{h_{thr}} - \frac{1}{h_i} \right)^+ . \qquad (93)$$

Here h_{thr} is the level above which no power is transmitted and β_{wf}^* is the SIR seen by a unit power user in a large system when all the other users are using the power allocation strategy $\mathcal{P}^{wf}$. Following the heuristic derivation of the waterfilling strategy, intuitively one expects that when α is very small there are very few users in a system with a very large processing gain and thus, the users are essentially orthogonal to each other and hence the policy is very similar to the single user waterfilling strategy. In the following result we make this intuitive observation precise:

Proposition 6.1 *Recall $\mathcal{P}^{wf}$, the waterfilling power allocation strategy (93), and the single user waterfilling strategy (15). Then,*

$$h_{thr} \downarrow h_{wf} \text{ and } \beta_{wf}^* \uparrow \sigma^{-2} \quad \text{as} \quad \alpha \downarrow 0 , \qquad (94)$$

$$h_{thr} \uparrow \bar{h} \text{ and } \beta_{wf}^* \downarrow 0 \quad \text{as} \quad \alpha \uparrow \infty . \qquad (95)$$

The proof is found in Appendix I.

6.2 Dependence on SNR, the background noise variance

We begin with the single user situation. It is intuitive that at high SNR (very low background noise variance σ^2), there is so much power available that the waterfilling strategy gains very little over the static power allocation policy, namely equal power allocation over all fading states. This was observed in [6] through simulation studies with Rayleigh and Nakagami fading examples. We make this statement precise and use it to find the structure of the waterfilling strategy at high SNR in the general multiuser scenario. Recall the single user capacity formula from (14):

$$\max_{\{\mathcal{P}\geq 0:\mathbb{E}[\mathcal{P}]\leq\bar{p}\}} \bar{C}(\mathcal{P}) \stackrel{\text{def}}{=} \max_{\{\mathcal{P}\geq 0:\mathbb{E}[\mathcal{P}]\leq\bar{p}\}} \frac{1}{2}\mathbb{E}\left[\log\left(1+\frac{h\mathcal{P}(h)}{\sigma^2}\right)\right]$$

and the optimal power allocation (water filling) from (15) as:

$$\mathcal{P}^*(h) = \left(\frac{\sigma^2}{h_{wf}} - \frac{\sigma^2}{h}\right)^+ . \tag{96}$$

Proposition 6.2 *In high SNR, the optimal power allocation (96) converges to the constant power policy and further more the loss in capacity by using the constant power policy goes to zero. Formally, as $\sigma^2 \to 0$,*

$$\mathcal{P}^* \xrightarrow{a.s.} \bar{p} \tag{97}$$
$$\bar{C}(\mathcal{P}^*) - \bar{C}(\bar{p}) \longrightarrow 0 . \tag{98}$$

The proof is completely elementary. As $\sigma^2 \to 0$, to meet the average power constraint we must have $\frac{\sigma^2}{h_{wf}} \to \bar{p}$. Thus the waterfilling strategy converges to the static power allocation strategy at high SNR showing (97). The gain with waterfilling strategy at any realization of the fading gain h is

$$\begin{aligned}\frac{1}{2}\left(\log\left(1+\frac{h\mathcal{P}^*(h)}{\sigma^2}\right) - \log\left(1+\frac{h\bar{p}}{\sigma^2}\right)\right) &= \frac{1}{2}\left(\log\left(1+\frac{h(\mathcal{P}^*(h)-\bar{p})}{\sigma^2+h\bar{p}}\right)\right) \\ &\leq \frac{1}{2}\log\left(1+\frac{\sigma^2}{h_{wf}\bar{p}}\right)\end{aligned} \tag{99}$$

where we used the definition of $\mathcal{P}^*(h)$ as the single user waterfilling policy in (15). Thus by the dominated convergence theorem and (97) we have shown (98).

We now turn to the multiuser scenario. Based on the single user result above one guesses that when α is very small at high SNR there is not much to gain by using the waterfilling strategy over the static power allocation policy of equal powers at all fading states. The correct extension of this intuition to the multiuser scenario is that when $\alpha \leq 1$ the number of users is less than the degrees of freedom available and each user can essentially null out the other users and we are back in the single user situation. If $\alpha > 1$, this strategy fails and there will be a strict loss with constant power allocation even at high SNR. The precise statement is below and the proof is in Appendix J.

Proposition 6.3 *For every N, at high SNR (i.e., as $\sigma^2 \downarrow 0$),*

1. *For* $\alpha \leq 1$*, we have* $h_{thr} \downarrow 0$ *and* $\beta^*_{wf} \uparrow \infty$*. Furthermore,* $\bar{C}_{sum}\left(\mathcal{P}^{wf}\right) - \bar{C}_{sum}\left(\bar{\mathcal{P}}\right) \to 0$*.*

2. *For* $\alpha > 1$ *we have* $h_{thr} \downarrow h_o > 0$ *and* $\beta^*_{wf} \uparrow \frac{1}{\alpha \bar{p} h_o} < \infty$*. Here* h_o *is the unique positive point of the map*
$$\mathcal{K}_o : h \mapsto \frac{HM(h)\left(\alpha\left(1 - F(h)\right) - 1\right)^+}{\alpha\left(1 - F(h)\right)}.$$
In this case, there is a strict loss in sum capacity by using the equal power allocation scheme.

We would like to give an intuitive explanation as to why this result is a priori feasible: Recall that successive decoding using the LMMSE receiver achieves sum capacity. At high SNR, the LMMSE receiver behaves as a *decorrelator* (Chapter 5 in [26]) and nulls out the multiple access interference. When $\alpha \leq 1$, the entire multiple access interference can be nulled out and thus we are back to the single user channel situation and we have the result that waterfilling makes little difference compared to constant power allocation in this situation. However, when $\alpha > 1$, the multiple access interference is not completely nulled out and the structure of the power strategy of the other users is still relevant. Having provided this intuition, we now dispel another explanation: at first sight it might appear that as N grows large the signature sequences of the users are orthogonal for $\alpha \leq 1$ and are not orthogonal for $\alpha > 1$ and hence provide the intuition for this result. However, as N grows, the users are orthogonal *even* when $\alpha > 1$. In fact, when the random variables v_{ij} are Gaussian, a simple calculation shows that
$$\max_{i \neq j}\left(s_i^t s_j\right)^2 \xrightarrow{a.s.} 0 \quad \text{as } N \to \infty$$
and K grows polynomially in N.

6.3 Multiple Classes

We now turn to a generalization of our model by allowing users to have different average power constraints. In particular, we assume that there are L *classes* of users; users in class l have average power constraint $\bar{p}_l$ for $l = 1 \dots L$. We assume that the number of users of class l is $K_l \stackrel{\text{def}}{=} \lfloor N\alpha_l \rfloor$). For the regime of large N a close observation of the heuristic derivation in Section 3 shows that much of the analysis remains valid in this case also. In particular, when there is no fading, the constant power policy is asymptotically optimal. In the general case of fading, the structure of the optimal power policy based on the asymptotic calculation is still waterfilling (73) but now the Kuhn-Tucker coefficient λ is different for users of different classes and is chosen such that the average power constraints are met: For any user i of class l, the policy is
$$\mathcal{P}_i^* : \left(h_1, \dots, h_{\sum_{l=1}^L K_l}, S\right) \mapsto \left(\frac{1}{\lambda_l} - \frac{1}{\beta^*_{wf} h_i}\right)^+ \tag{100}$$
where β^*_{wf} is the SIR of a unit power user in a large system with users adopting this power strategy and is the solution to the fixed point equation (by an appeal to Lemma 4.8; analogous

to (53)):

$$\sigma^2\beta^*_{wf} = 1 - \sum_{l=1}^{L}\alpha_l \int_{\lambda_l/\beta^*_{wf}}^{\bar{h}} \left(1 - \frac{\lambda_l}{\beta^*_{wf}h}\right) dF(h) \tag{101}$$

Analogous to the continuation in Section 3 for the single class case, we will sketch an argument that ensures the existence of the quantities β^*_{wf} and λ_l and also demonstrates a simple fixed point iteration algorithm that converges to the desired quantities. We will only discuss the major changes from the corresponding steps in Section 3. Denoting $h^{(l)}_{thr} \stackrel{\text{def}}{=} \frac{\lambda_l}{\beta^*_{wf}}$, the level below which no power is transmitted by users of class l, analogous to (54) we have, from the average power constraint on the power policy in (100), that $h^{(l)}_{thr}$ is the solution to the fixed point equation:

$$h^{(l)}_{thr} = \frac{HM\left(h^{(l)}_{thr}\right)}{1 + \frac{\bar{p}_l HM\left(h^{(l)}_{thr}\right)}{1-F\left(h^{(l)}_{thr}\right)}\beta^*_{wf}} . \tag{102}$$

Continuing from (101), analogous to (56), (57) and (58), we have using (102) that

$$\sigma^2\beta^*_{wf} = 1 - \sum_{l=1}^{L}\alpha_l\left(1 - F\left(h^{(l)}_{thr}\right)\right) + \sum_{l=1}^{L}\alpha_l \frac{\left(1 - F\left(h^{(l)}_{thr}\right)\right)^2}{1 - F\left(h^{(l)}_{thr}\right) + \beta^*_{wf}\bar{p}_l HM\left(h^{(l)}_{thr}\right)} . \tag{103}$$

In the single class case we were able to observe that β^*_{wf} was equal to the solution of a quadratic equation (44). The natural extension is the following. Consider a system with processing gain N where K_l users are *received* with the same power p_l (this is equivalent to transmit power p_l but the fading is degenerate, i.e., $h_i(n) = 1$) for $l = 1 \ldots L$. As $N \to \infty$, assuming that $\frac{K_l}{N} \to \alpha_l$ for every class l, it follows from Lemma 4.8 that the asymptotic SIR of a unit (received) power user is a positive constant $\beta^*\left(\{p_l, \alpha_l\}\, l = 1 \ldots L\right)$ that satisfies the fixed point equation (analogous to (44)):

$$\sigma^2\beta = 1 - \sum_{l=1}^{L}\frac{\alpha_l\beta p_l}{1 + \beta p_l} . \tag{104}$$

Comparing (103) with (104) we observe that

$$\beta^*_{wf} = \beta^*\left(\left\{\frac{\bar{p}_l HM\left(h^{(l)}_{thr}\right)}{1 - F\left(h^{(l)}_{thr}\right)}, \alpha_l\left(1 - F\left(h^{(l)}_{thr}\right)\right)\right\}; l = 1 \ldots L\right) . \tag{105}$$

Analogous to the fixed point iteration of the map in (59) for the single class scenario, we define the following maps for each class l:

$$\mathcal{K}_l : (h_1, \ldots, h_L) \mapsto \frac{HM(h_l)}{1 + \frac{\bar{p}_l HM(h_l)}{1-F(h_l)}\beta^*\left(\left\{\frac{\bar{p}_l HM(h_l)}{1-F(h_l)}, \alpha\left(1 - F(h_l)\right)\right\}; l = 1 \ldots L\right)} \tag{106}$$

It follows from (102) and (105) that

$$\mathcal{K}_l\left(h^{(1)}_{thr}, \ldots, h^{(L)}_{thr}\right) = h^{(l)}_{thr} .$$

Analogous to Lemma 4.10, we justify the existence of $h^{(l)}_{thr}$ by the following proposition:

Proposition 6.4 *Consider the fixed point iteration:*

$$\begin{aligned} h_l(0) &\stackrel{\text{def}}{=} 0 \ \forall l = 1 \dots L \\ h_l(n+1) &\stackrel{\text{def}}{=} \mathcal{K}_l(h_1(n), \dots, h_L(n)) \ \forall n \geq 0 \ , \ \forall l = 1 \dots L \ . \end{aligned}$$

Then $\{h_l(n)\}_n$ *is an increasing sequence that converges to* $h_{thr}^{(l)}$ *for each* $l = 1 \dots L$.

Thus $h_{thr}^{(l)}$ exist as the limits of the fixed point iteration above. We omit the proof of this proposition while pointing out the replacement of the key observation (163) in the proof of Lemma 4.10: For every $l = 1 \dots L$,

$$\mathcal{K}_l(h_1, \dots, h_L) \geq h \iff$$
$$\left(\frac{\sigma^2}{\bar{p}_l h} + \alpha_l\right) f(h, h_l) + \sum_{j \neq l} \alpha_j \bar{p}_j HM(h_j) \frac{f(h, h_l)}{\bar{p}_l h + \frac{\bar{p}_j HM(h_j)}{1 - F(h_j)} f(h, h_l)} \geq 1$$

where

$$f(h, h_l) \stackrel{\text{def}}{=} \int_{h_l}^{\bar{h}} \left(1 - \frac{h}{h_o}\right) dF(h_o) \ .$$

This also shows the uniqueness of $h_{thr}^{(l)}$. The formal statement of the optimality of this power allocation solution, analogous to Theorem 5.2 is below and the key ideas of the proof are all contained in the proof of Theorem 5.2.

Theorem 6.5

$$\limsup_{N \to \infty} \sqrt{N} \left(\bar{C}_{opt}^{(N)} - \bar{C}_{sum}\left(\mathcal{P}_l^{wf}; l = 1 \dots L\right)\right) < \infty \quad \textit{and } K_l = \lfloor \alpha_l N \rfloor \ .$$

Extensions of the observations made in Section 6.2 to the multiple class scenario are natural. Constant power allocation (equal to $\bar{p}_l$ for users of class l) to the users incurs no loss in sum capacity as compared to the waterfilling scheme at high SNR if and only if $\sum_{l=1}^{L} \alpha_l \leq 1$.

7 Numerical Examples

In this section we demonstrate the value of our theoretical results by simulating different power control strategies in a Rayleigh fading channel and plotting the corresponding sum capacities achieved for various parameters of loading and SNR. We assumed that the components of the signature sequences are distributed as zero mean, unit variance Gaussian random variables (our theoretical results show that the actual distribution does not matter; so long as it has zero mean, unit variance and bounded fourth moment). In Fig 1, we plot sum capacity with the constant power allocation and also with the optimal power allocation policy (this policy depends on the actual realization of the signature sequences). We observe

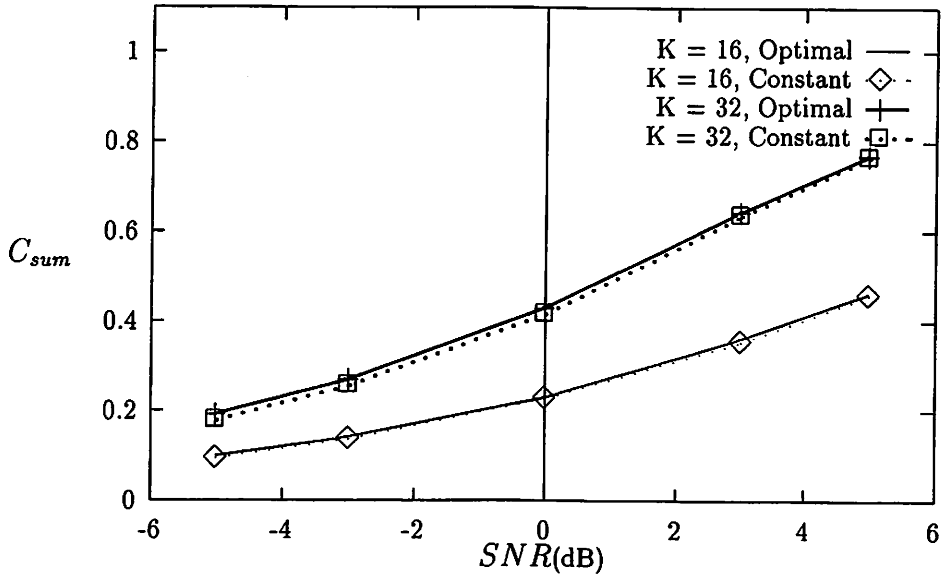

Figure 1: No fading scenario. Sum Capacity is plotted with the optimal allocation and the constant power allocation policies with $N = 32$.

that there is very little difference in sum capacity between these two policies. Thus $N = 32$ is already large enough for the difference to be very small. Assuming Rayleigh fading, Fig 2 and Fig 3 plot sum capacity with three different power allocation policies: the asymptotically optimal waterfilling policy, the optimal power allocation policy (which is a function of the realization of the signature sequences and fading) and the constant power allocation policy, for different values of SNR and number of users equal to $N/2$ and N respectively. The first observation from Fig 2 and Fig 3 is that the sum capacity with the asymptotically optimal policy of waterfilling is already very close to that with the optimal policy even at $N = 32$. Furthermore, from Fig 2 we observe that with the number of users per unit processing gain being small ($\alpha = 0.5$) the difference in sum capacity by using one of these two policies as compared with the constant power allocation policy (constant for all fading levels and realizations of signature sequences) is fairly small. Fig 3 shows that this difference increases when α is increased to 1. This observation is in concord with the observation in [6] that waterfilling gains very little over constant power allocation policy in a single user fading channel for reasonably high SNRs. Proposition 6.3 predicts that the penalty in sum capacity by using the constant power allocation policy grows with the number of users per unit processing gain. We observe this behavior in Fig 4 where we have plotted sum capacity for fixed SNR (5dB) versus the number of users: while there is very little difference in sum capacity between the optimal power allocation and waterfilling policies, the penalty by using constant power allocation policy grows with the number of users.

Even though closed form solutions are not known for the optimal power allocation policy (these depend in general on the instantaneous realizations of the signature sequences and fading gains), we can compute numerically the sum capacity with the optimal power allocation. We used the software `maxdet` available in [31] to arrive at the optimal power allocation;

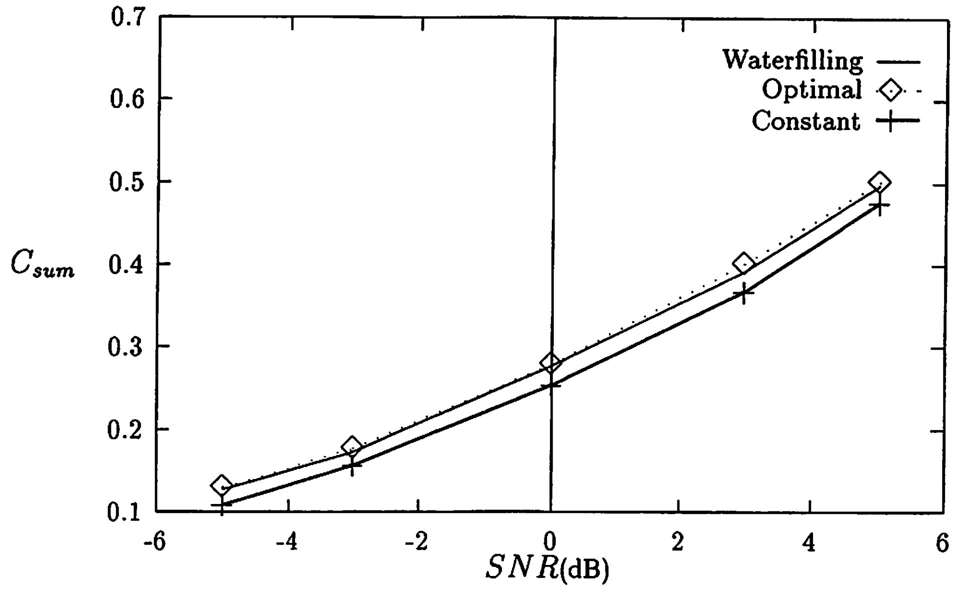

Figure 2: Rayleigh fading scenario with $N = 32$ and $K = 16$. Sum capacity in bits/s/Hz is plotted with the optimal allocation, asymptotically optimal waterfilling allocation and the constant power allocation policies.

the software provides an interior point algorithm to solve the determinant maximization problem:

$$\max_{p_i \geq 0} \quad L(p_1, \ldots, p_K, \lambda_1, \ldots, \lambda_K) \tag{107}$$

$$\text{where} \quad L(p_1, \ldots, p_K, \lambda_1, \ldots, \lambda_K) \stackrel{\text{def}}{=} \log\det\left(I + \sum_{i=1}^{K} s_i s_i^t h_i p_i\right) + \sum_{i=1}^{K} \lambda_i (p_i - \bar{p})$$

We obtained sum capacity at power prices $\lambda_1, \ldots, \lambda_K$ by averaging the scaled (by $1/2N$) maximal value of the optimization problem above (107). Sum capacity is then the smallest value over all power prices (the corresponding prices are known as "equilibrium power prices" or Kuhn-Tucker coefficients; this is from standard Lagrange theory in convex analysis - see Corollary 28.4.1 in [14]). From the proof of Theorem 5.2, we have a good guess for the Kuhn-Tucker coefficients: $\lambda_1 = \cdots = \lambda_K = \beta^*_{wf} h_{thr}$. The *actual* power prices were found by a line search. The solution to the optimization problem (107) with the equilibrium power prices gives the optimal power allocation and thus we arrive numerically at sum capacity with the optimal power allocation policy.

8 Multiple Antenna Systems

We now turn to the multiple antenna model and refer the reader to Section 8 of [22] for a discussion of the standard model and advantages of this diversity scheme. A baseband

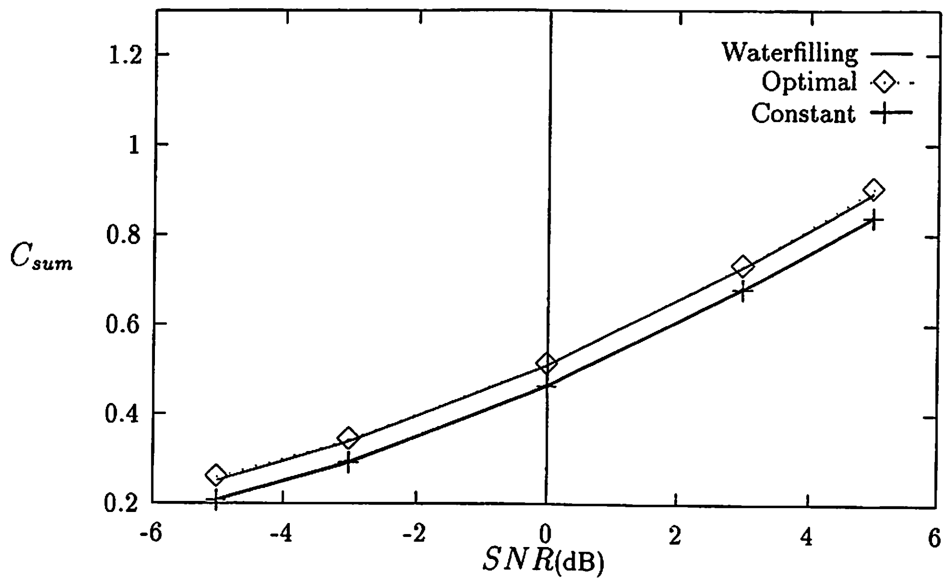

Figure 3: Rayleigh fading scenario with $N = 32$ and $K = 32$. Sum capacity in bits/s/Hz is plotted with the optimal allocation, asymptotically optimal waterfilling allocation and the constant power allocation policies.

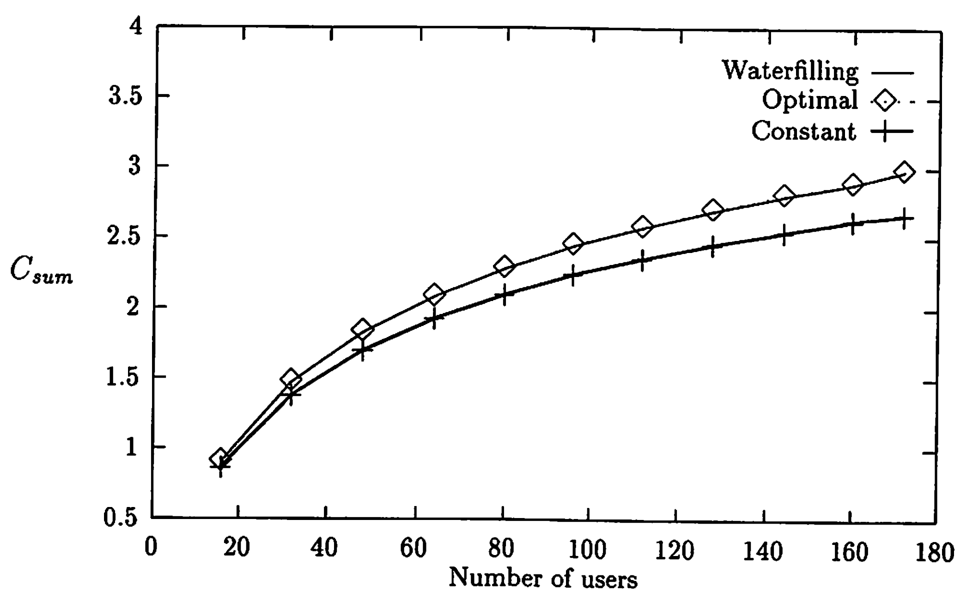

Figure 4: Rayleigh fading scenario with $N = 16$ and SNR = 5dB. Sum capacity is plotted with the optimal allocation, asymptotically optimal waterfilling allocation and the constant power allocation policies versus number of users.

model for a synchronous multiaccess antenna array channel is

$$Y(n) = \sum_{i=1}^{K} X_i(n)\tilde{\mathbf{h}}_i(n) + \mathbf{W}(n) . \tag{108}$$

Here, n denotes the time of channel use, $X_i(n)$ is the transmitted symbol at time n of user i and $Y(n)$ is a N dimensional vector of received symbols at the N antenna elements of the array at the receiver. The vector $\tilde{\mathbf{h}}_i(n)$ represents the fading of the ith user at time n at each of the antenna array elements and the entries are independent and identically distributed complex stationary and ergodic processes. $\mathbf{W}(n)$ is an additive white proper complex Gaussian noise process. Making the assumption that the channel fading of the users can be measured and tracked perfectly (implicit is the notion that the time scale at which the path gains change is much slower than the symbol rate of the system) and that both the transmitter and receiver know the fading state, the multiple antenna MAC model is very similar to that in the long random signature sequence model with no fading (i.e. $h_i = 1 a.s.$ in (1)). The key difference is that the entries of the signature sequences s_i are scaled by $\frac{1}{\sqrt{N}}$ in the CDMA model (1) while the entries of $\tilde{\mathbf{h}}_i$ are not. Thus defining $s_i = \frac{1}{\sqrt{N}}\tilde{\mathbf{h}}_i$ we arrive at the following expression for the sum capacity of the multiple antenna MAC (analogous to (9) and (10)):

$$\bar{C}^A_{opt}(N) \stackrel{\text{def}}{=} \sup_{\mathcal{P}\in\mathcal{F}_2^{(N)}} \bar{C}^A_{sum}(\mathcal{P}) \stackrel{\text{def}}{=} \sup_{\mathcal{P}\in\mathcal{F}_2^{(N)}} \frac{1}{2N}\mathbb{E}\left[\log\det\left(I + N\sum_{i=1}^{K}\sigma^{-2}s_i s_i^t \mathcal{P}_i(S)\right)\right] .$$

Some remarks about this expression are now in order: The power allocation policy $\mathcal{P}$ now depends only on $S = \sqrt{N}\left[\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_K\right]$ and the sum capacity with the power allocation policy $\mathcal{P}$ in nats/s/antenna is written as $\bar{C}^A_{sum}(\mathcal{P})$. The sum capacity of the MAC is (as in (9)) the supremum over all valid power allocation policies. The difference in the expression for sum capacities when compared to that of the CDMA model is that the received power is scaled by N.[2]

We expect the result of Theorem 5.1 to hold in this case also in some appropriate sense, i.e., the power allocation policy of constant power over all fading levels is asymptotically (in the number of antennas at the receiver and corresponding number of users) optimal. We retain our earlier notation of $\bar{\mathcal{P}}$ to denote the static power allocation policy and formalize this result below:

Theorem 8.1

$$\limsup_{N\to\infty}\left(\bar{C}^A_{opt}(N) - \bar{C}^A_{sum}\left(\bar{\mathcal{P}}\right)\right) \leq \frac{\alpha(\alpha-1)^+ \tilde{K}_p}{2\bar{p}}$$

[2]In the underlying physical model, the received power does not arbitrarily increase as N increases. As N becomes too large, either the size of the antenna forces the received power to become constant (since the spacing between the antennas are at least half the wavelength apart) or the distance from the antennas to the users increases (allowing us to keep the same size of the antennas) forcing the received power to become constant. For small values of N, this linear increase in received power is justified. We look at the asymptotic of sum capacity in the regime of large N in this scenario nevertheless since we believe that the asymptotic value of sum capacity is reached even for small values of N. Then linear increase in received power is the relevant model since we are primarily interested in the regime of small number of antennas.

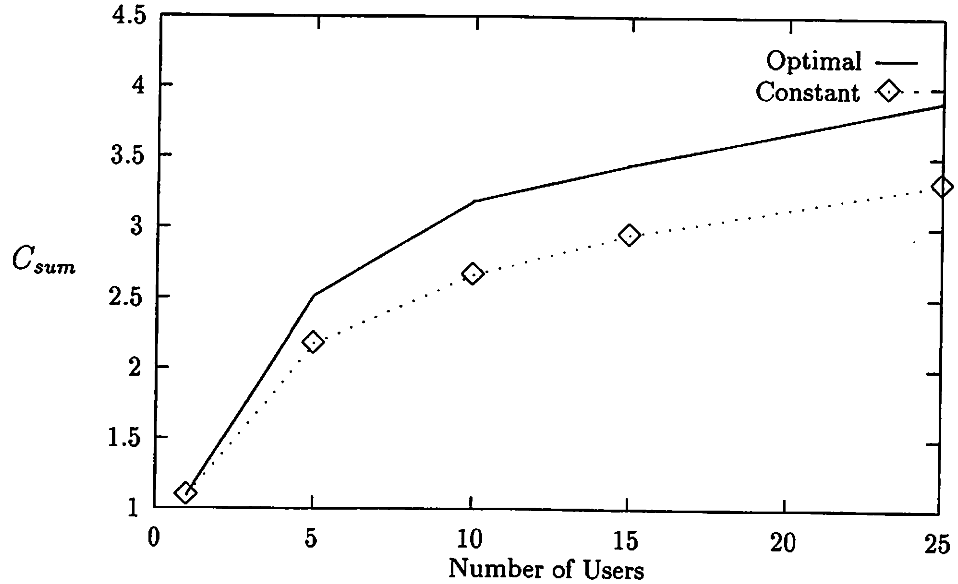

Figure 5: Sum Capacity in bits/s with 1 antenna at the receiver is plotted versus number of users at a fixed SNR level of 5dB with both the optimal and constant power allocations.

Here $\tilde{K}_p$ is some constant independent of N. We relegate the proof of this result to Appendix K. This result is rather surprising from the context of the results for the case of $N = 1$, the single antenna scenario. When there is only one antenna the optimal power allocation policy is to let only the user with the best channel amplitude transmit and for that user to follow the waterfilling power policy [10]. The gain in sum capacity by following this strategy over the suboptimal policy of constant power allocation to the users at all fading levels can be substantial; the larger the number of users, the larger this *multiuser diversity* gain is. Fig 5 plots the sum capacity with both these power policies assuming i.i.d. Rayleigh fading from the users to the single antenna. We can see that with an increasing number of users, the gain in sum capacity is widening. However, when there are a substantial number of antennas, the gain by utilizing multiuser diversity vanishes and the constant power allocation policy performs just as well. In Figure 6, we plot the sum capacity as a function of the SNR of the users when following the optimal policy as well as when following the constant power allocation policy. In practice, a small number of antennas is considered practical at the receiver (to validate our assumption that the paths from any user to each antenna have independent fading, the antennas have to be at least half a wavelength apart and this generally implies a strict restriction on the number of antennas given the size of the receiver). We assume $N = 5$ antennas for our simulations. In this simulation we assumed further that each component of $\tilde{\mathbf{h}}_i(n)$ is i.i.d. complex Gaussian with zero mean and variance 1. We observe that the loss in sum capacity with the constant power allocation policy as compared to the optimal power allocation policy is not huge even when N is very small ($N = 5$ in this simulation example).

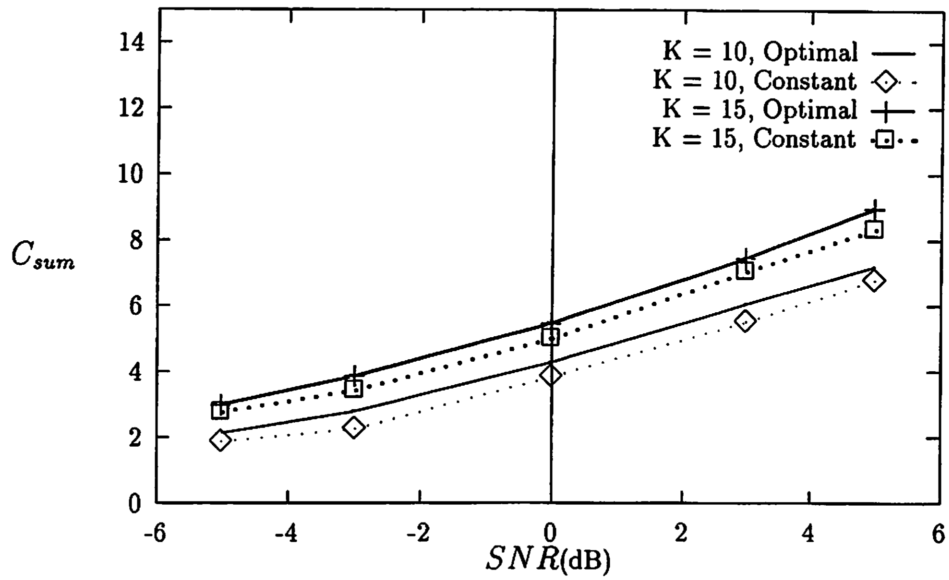

Figure 6: Sum Capacity in bits/s with 5 antennas at the receiver is plotted versus the SNR of the users with both the optimal and constant power allocations.

The antenna model of (108) implicitly assumes that the users are approximately equidistant from the antenna array. The vector of path gains from any user i to the antenna array can more generally be modeled by $\hat{\mathbf{h}}_i\tilde{\mathbf{h}}_i$ where $\tilde{\mathbf{h}}$ is as before and $\hat{\mathbf{h}}_i$ is a scalar complex number that is independent of the vector $\tilde{\mathbf{h}}_i$ and models the (slowly varying, when compared to the components of $\tilde{\mathbf{h}}_i$) component of the overall path gain that captures the distance of user i from the receiver and the shadowing loss of the signal of user i. Then the received signal at the antenna array in symbol interval n is

$$Y(n) = \sum_{i=1}^{K} X_i(n)\hat{\mathbf{h}}_i\tilde{\mathbf{h}}_i(n) + \mathbf{W}(n)\,.$$

This model is now very similar to the DS-CDMA model in (1). The asymptotically optimal power allocation policy in this model, analogous to the results we have for the DS-CDMA case, is *waterfilling*. This policy depends only on the slowly varying component $\hat{\mathbf{h}}_i$ for user i and some constants that depend on the statistics of the stationary fading distributions. The formal result will be analogous to Theorem 5.2 and we do not pursue this here for brevity.

9 Conclusions

Though we have identified the asymptotic structure of the optimal power allocation policy, it remains an important open problem to find a closed form expression for the *exact* optimal solution for every finite system size N and fixed signature sequences and path gains. A closed form expression for the optimal powers as a function of the signature sequences and path gains appears to be unattainable. In fact, exactly this optimization problem (a finite dimensional generalized version of (9)) is considered in [24] where the authors derive interior point algorithms that converge to the optimal allocation; these algorithms have worst-case polynomial (in N, the system size) run time complexity. A software routine that implements the ellipsoidal algorithms for determinant maximization is available in [31]. Nevertheless, the waterfilling power policy identified in this paper is very appealing practically, due to its simplicity and the computation requirements to arrive at it is practically nil. The fading statistics can be estimated and used to adaptively compute the threshold level of the waterfilling strategy using the fixed point iteration outlined in the paper. Though our estimate for the gap in sum capacity between this policy and the optimal policy is of the order of $\sqrt{N}$ for large N, our simulation studies suggest that this gap is very small. even for reasonably small values of N. In the context of multiple antenna systems, we saw this for a very small number of antennas ($N = 5$ in our example). Thus, even if closed form expressions could be found for the optimal power allocation policy, it might not be worthwhile to implement the optimal allocation because of its complexity.

Another natural extension of the problem formulation in this paper is to characterize power policies that maximize any linear functional of the rates at which the users can jointly reliably transmit. This problem was addressed and solved in [21] for multiple access fading channels with a single degree of freedom. The extension of this result to multiple degrees of freedom remains an important open problem.

A Proof of Proposition 3.1

We first recall a special case of the central result of [19] regarding the convergence of the empirical distribution of eigenvalues of a random Hermitian matrices: Let G_N be the empirical distribution function of the eigenvalues of $\sum_{i=1}^{K} s_i s_i^t h_i g(h_i)$ (there are N eigenvalues). Then G_N converges almost surely in distribution to a deterministic distribution G^* where the Stieltjes transform $m(z)$ of G^* satisfies the fixed point equation:

$$m(z) = \frac{1}{-z + \alpha \int \frac{\tau h g(h) dF(h)}{1+\tau m(z)}} \quad \forall z \in \mathbb{C}^+ .$$

Here the Stieltjes transform of a distribution function G is defined as

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda) .$$

It also follows from [19] that the support of G^* is bounded above by $K_g \bar{h} (1 + \sqrt{\alpha})^2$ where K_g denotes the upper bound on the function g. Applying Theorem 1.1 and its corollary of [1] to our case and denoting $K_s = 1 + K_g \bar{h} (1 + \sqrt{\alpha})^2$, we have that

$$\mathbb{P}[G_N(K_s) = 1 \text{ for all large } N] = 1 .$$

Thus we have that

$$\begin{aligned}\frac{1}{2N}\log\det\left(I+\sum_{i=1}^{K}\sigma^{-2}s_is_i^th_ig\left(h_i\right)\right) &= \frac{1}{2}\int\log\left(1+\lambda\sigma^{-2}\right)dG_N\left(\lambda\right)\\ &\xrightarrow{a.s.} \frac{1}{2}\int\log\left(1+\lambda\sigma^{-2}\right)dG^*\left(\lambda\right) \qquad (109)\\ &\stackrel{\text{def}}{=} \bar{C}^g_{sum}\ .\end{aligned}$$

We now show convergence to $\bar{C}^g_{sum}$ in the first moment. We have

$$\begin{aligned}\frac{1}{2N}\log\det\left(I+\sum_{i=1}^{K}\sigma^{-2}s_is_i^th_ig\left(h_i\right)\right) &\le \frac{1}{2N}\sum_{i=1}^{K}\log\left(1+\sigma^{-2}h_ig\left(h_i\right)s_i^ts_i\right) \qquad (110)\\ &\le \frac{1}{2N}\sum_{i=1}^{K}\log\left(1+\sigma^{-2}\bar{h}K_gs_i^ts_i\right) \qquad (111)\\ &\le \frac{\bar{h}K_g}{2N\sigma^2}\sum_{i=1}^{K}s_i^ts_i\end{aligned}$$

where (110) follows from the Hadamard inequality and (111) follows from the fact that $h_i \le \bar{h}$ and g is bounded above by K_g. Since

$$\frac{1}{N}\sum_{i=1}^{K}s_i^ts_i \xrightarrow{a.s.,L^1} \alpha$$

the proposition follows from the pointwise convergence result above (109) and the dominated convergence theorem. □

B Proof of Proposition 4.1

We have

$$\begin{aligned}\bar{C}^{(N)}_{opt} &= \bar{C}_{sum}\left(\mathcal{P}^*\right)\\ &\le \frac{1}{2N}\sum_{i=1}^{K}\mathbb{E}\left[\log\left(1+\sigma^{-2}s_i^ts_ih_i\mathcal{P}_i^*\left(h_1,\dots,h_K,S\right)\right)\right]\\ &= \frac{K}{2N}\mathbb{E}\left[\log\left(1+\sigma^{-2}s_1^ts_1h_1\mathcal{P}_1^*\left(h_1,\dots,h_K,S\right)\right)\right]\\ &\le \frac{K}{2N}\max_{\mathcal{P}\in\mathcal{F}_2^{(N)}}\mathbb{E}\left[\log\left(1+\sigma^{-2}s_1^ts_1h_1\mathcal{P}_1\left(h_1,\dots,h_K,S\right)\right)\right] \qquad (112)\end{aligned}$$

where the derivation of these inequalities is completely straightforward. Using Jensen inequality conditionally on h_1, s_1, we have from (112) that

$$\bar{C}^{(N)}_{opt} \le \frac{\alpha}{2}\sup_{\mathcal{P}\in\mathcal{F}_2}\left\{\mathbb{E}\left[\log\left(1+s_1^ts_1h_1\mathcal{P}\left(h_1s_1^ts_1\right)\right)\right]\right\} \qquad (113)$$

where the set $\mathcal{F}_2$ is defined as

$$\mathcal{F}_2 \stackrel{\text{def}}{=} \left\{\mathcal{P} : h_1 s_1^t s_1 \mapsto \mathbb{R}_+, \mathbb{E}\left[\mathcal{P}\left(h_1 s_1^t s_1\right)\right] \leq \bar{p}\right\} .$$

Now, for every $\mathcal{P} \in \mathcal{F}_2$ we have

$$\begin{aligned} \mathbb{E}\left[\log\left(1 + s_1^t s_1 h_1 \mathcal{P}\left(h_1 s_1^t s_1\right)\right)\right] &\leq \log 2 + \mathbb{E}\left[\log\left(2 h_1 s_1^t s_1 \mathcal{P}\left(h_1 s_1^t s_1\right)\right)\right] \\ &\leq 2\log 2 + \log\left(\mathbb{E}\left[h_1 s_1^t s_1\right]\right) + \log\left(\mathbb{E}\left[\mathcal{P}\left(h_1 s_1^t s_1\right)\right]\right) \\ &\leq 2\log 2 + \log\left(\bar{h}\bar{p}\right) \qquad (114) \end{aligned}$$

where we used Jensen inequality in the derivation of the last but one step. Now combining (114) and (113) we have shown Proposition 4.1 by denoting $K_c = \log 2 + \frac{1}{2}\log\left(\bar{h}\bar{p}\right)$. □

C Proof of Proposition 4.2

This result is fairly well known: [20] shows the concavity of a somewhat modified map but the reference is slightly obscure. The authors in [24] consider a slightly generalized version of the map (28) and mention that the map is known to be concave. We offer the following proof beginning with some notation: For any $x = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$, let $x_{[1]} \geq \cdots \geq x_{[n]}$ denote the components of x in decreasing order, called the *order statistics* of x. Majorization is a partial order on the elements of $\mathbb{R}^n$ and makes precise the vague notion that the components of a vector x are "less spread out" or "more nearly equal" than are the components of a vector y by the statement x is majorized by y.

Definition C.1 *For $x, y \in \mathbb{R}^n$, say that x is majorized by y (or y majorizes x) if*

$$\begin{array}{l} \sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}, \quad k = 1 \ldots n-1 \\ \sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]} \end{array}$$

A simple (trivial, but important) example of majorization between two vectors is the following:

Example C.1 *For every $a \in \mathbb{R}^n$ such that $\sum_{i=1}^n a_i = 1$,*

$$(a_1, \ldots, a_n)^t \quad \textit{majorizes} \quad \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)^t$$

Real functions on $\mathbb{R}^n$ that are monotone nonincreasing in majorization order are called Schur-concave functions. Formally,

Definition C.2 *A real valued function* $\phi : \mathbb{R}^n \to \mathbb{R}$ *is said to be Schur-concave if for all* $x, y \in \mathcal{R}^n$ *such that* y *majorizes* x *we have* $\phi(x) \geq \phi(y)$.

An important class of Schur-concave functions is the following (Theorem 3.C.1 in [11]):

Example C.2 *If* $g : \mathbb{R} \to \mathbb{R}$ *is strictly concave then the symmetric concave function* $\phi(x) = \sum_{i=1}^{n} g(x_i)$ *is Schur-concave.*

Let A and B be two symmetric matrices of dimension $n \times n$. Let λ^A and λ^B denote the vectors of eigenvalues of A and B respectively. The following result (Theorem 9.G.1 in [11]) shows that the eigenvalues of $A + B$ (the components of the vector λ^{A+B}) are less spread out than the sum of the order statistics of the eigenvalues of A and B:

Lemma C.1 *For any two symmetric matrices* A *and* B,

$$\left(\lambda_1^{A+B}, \ldots, \lambda_n^{A+B}\right)^t \text{ is majorized by } \left(\lambda_{[1]}^A + \lambda_{[2]}^B, \ldots, \lambda_{[n]}^A + \lambda_{[n]}^B\right)^t$$

We now use these results to show that the map C is strictly concave. Fix $p^{(1)}$ and $p^{(2)}$ in the positive orthant in $\mathbb{R}^K$. For every $\eta \in (0,1)$ we have

$$\eta \log\det\left(I + \sum_{i=1}^{K} p_i^{(1)} h_i s_i s_i^t\right) + (1-\eta)\log\det\left(I + \sum_{i=1}^{K} p_i^{(2)} h_i s_i s_i^t\right)$$

$$= \eta \sum_{j=1}^{N} \log\left(1 + \lambda_{[j]}^{(1)}\right) + (1-\eta)\sum_{j=1}^{N}\log\left(1 + \lambda_{[j]}^{(2)}\right) \tag{115}$$

$$\leq \sum_{j=1}^{N} \log\left(1 + \eta\lambda_{[j]}^{(1)} + (1-\eta)\lambda_{[j]}^{(2)}\right) \tag{116}$$

$$\leq \sum_{j=1}^{N} \log\left(1 + \lambda_j^{(3)}\right) \tag{117}$$

$$= \log\det\left(I + \sum_{i=1}^{K}\left(\eta p_i^{(1)} + (1-\eta)p_i^{(2)}\right) h_i s_i s_i^t\right) \tag{118}$$

where $\left(\lambda_1^{(t)}, \ldots, \lambda_N^{(t)}\right)$ is the vector of eigenvalues of the matrix $\left(I + \sum_{i=1}^{K} s_i s_i^t h_i p_i^{(t)}\right)$ for $t = 1, 2$. Also, $\left(\lambda_1^{(3)}, \ldots, \lambda_N^{(3)}\right)$ is the vector of eigenvalues of the matrix $\left[I + \sum_{i=1}^{K}\left(\eta p_i^{(1)} + (1-\eta)p_i^{(2)}\right) h_i s_i s_i^t\right]$. Then (115) follows from recalling that $\lambda_{[1]}, \ldots, \lambda_{[N]}$ are the order statistics of the $\lambda_1, \ldots, \lambda_N$. The inequality in (116) follows from concavity of the logarithm function and is strict unless $\lambda_{[j]}^{(1)} = \lambda_{[j]}^{(2)}$ for every $j = 1 \ldots N$. The inequality in (117) follows from Lemma C.1 and from Example C.2. This inequality is strict unless the two matrices $\left\{\left[I + \sum_{i=1}^{K} s_i s_i^t h_i p_i^{(t)}\right]\right\}_{t=1,2}$

have the same eigenvectors (i.e., the two matrices commute) and if the eigenvalue corresponding to an eigenvector for $t = 1$ is $\lambda_{[j]}^{(1)}$ then the eigenvalue for $t = 2$ corresponding to the same eigenvector is $\lambda_{[j]}^{(2)}$ for every $j = 1 \ldots N$. The equality (118) is just definition of the eigenvalues $\lambda^{(3)}$. Thus, we have shown the map in (28) to be concave (in fact, we have shown that the map in (28) is Schur-concave). Furthermore we have shown that the map is strictly concave if and only if $\{h_i s_i s_i^t,\ i = 1 \ldots K\}$ are linearly independent. □

D Proof of Proposition 4.4

We fix N throughout this proof. From Proposition 4.1 we know that $\bar{C}_{opt}^{(N)}$ is finite. Recall the definition of $\mathcal{F}_0^{(N)}$ from (31). Define functions $f_0, \ldots, f_K$ from $\mathcal{F}_0^{(N)}$ to the non-negative reals as follows:

$$\begin{aligned} f_0 : \mathcal{P} &\mapsto \bar{C}_{opt}^{(N)} - \bar{C}_{sum}(\mathcal{P}) \\ f_j : \mathcal{P} &\mapsto \frac{\mathbb{E}(\mathcal{P}_j) - \bar{p}}{2N},\ \forall j = 1 \ldots K . \end{aligned}$$

We first observe that the functions f_j are finite on the domain $\mathcal{F}_0^{(N)}$. Now, by definition of $\bar{C}_{opt}^{(N)}$, the system of equations

$$f_0(\mathcal{P}) < 0, f_1(\mathcal{P}) < 0, \ldots, f_K(\mathcal{P}) < 0$$

has no solution $\mathcal{P} \in \mathcal{F}_0^{(N)}$. Consider the following claim:

$$\exists \lambda_0, \lambda_1, \ldots, \lambda_K \geq 0, \text{ not all zero such that } \sum_{j=0}^{K} \lambda_j f_j(\mathcal{P}) \geq 0,\ \forall \mathcal{P} \in \mathcal{F}_0^{(N)} . \tag{119}$$

Suppose this is true. Our first observation is that $\lambda_0 \neq 0$ since $\sum_{j=1}^{K} \lambda_j f_j(\mathcal{P}) \geq 0$ is impossible for all $\mathcal{P} \in \mathcal{F}_0^{(N)}$. Thus dividing throughout by λ_0, (119) can be rewritten as

$$\bar{C}_{sum}(\mathcal{P}) - \sum_{j=1}^{K} \lambda_j (\mathbb{E}[\mathcal{P}_i] - \bar{p}) \leq \bar{C}_{opt}^{(N)},\ \forall \mathcal{P} \in \mathcal{F}_0^{(N)}$$

and hence

$$\bar{C}_{opt}^{(N)} \geq \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} \bar{C}_{sum}(\mathcal{P}) - \sum_{j=1}^{K} \lambda_j \frac{(\mathbb{E}[\mathcal{P}_j] - \bar{p})}{2N} . \tag{120}$$

By the symmetry among the users, it follows from (120) that for every permutation $\sigma \in S_K$ we have

$$\bar{C}_{opt}^{(N)} \geq \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} \left\{ \bar{C}_{sum}(\mathcal{P}) - \sum_{j=1}^{K} \frac{\lambda_{\sigma(j)}}{2N} (\mathbb{E}[\mathcal{P}_j(h_1, \ldots, h_K, S)] - \bar{p}) \right\} . \tag{121}$$

Observe that the map on the positive orthant of $\mathbb{R}^K$

$$(\mu_1,\ldots,\mu_K)\mapsto \sup_{\mathcal{P}\in\mathcal{F}_0^{(N)}}\left\{\bar{C}_{sum}(\mathcal{P})-\sum_{j=1}^{K}\frac{\mu_j}{2N}\left(\mathbb{E}\left[\mathcal{P}_j(h_1,\ldots,h_K,S)\right]-\bar{p}\right)\right\}$$

is concave. Defining $\lambda=\frac{1}{K}\sum_{i=1}^{K}\lambda_i$ and using in (121) the concavity of the map above we arrive at

$$\bar{C}_{opt}^{(N)}\geq \sup_{\{\mathcal{P}:\mathcal{P}_i\geq 0,\ \forall i=1\ldots K\}}\left\{\bar{C}_{sum}(\mathcal{P})-\frac{\lambda}{2N}\sum_{i=1}^{K}\left(\mathbb{E}\left[\mathcal{P}_i(h_1,\ldots,h_K,S)\right]-\bar{p}\right)\right\}. \tag{122}$$

Now we have for every $\lambda\geq 0$ that

$$\bar{C}_{opt}^{(N)}\leq \bar{C}_{sum}(\mathcal{P})-\frac{\lambda}{2N}\sum_{j=1}^{K}\left(\mathbb{E}\left[\mathcal{P}_j\right]-\bar{p}\right),\ \ \forall\mathcal{P}\in\mathcal{F}_2^{(N)}.$$

Since $\mathcal{F}_2^{(N)}\subset\mathcal{F}_0^{(N)}$ we have for every $\lambda>0$

$$\bar{C}_{opt}^{(N)}\leq \sup_{\mathcal{P}\in\mathcal{F}_0^{(N)}}\left\{\bar{C}_{sum}(\mathcal{P})-\frac{\lambda}{2N}\sum_{j=1}^{K}\left(\mathbb{E}\left[\mathcal{P}_j\right]-\bar{p}\right)\right\} \tag{123}$$

Combining (122) and (123) the proof is complete. It remains now to show (119). To see this, define

$$\begin{aligned}C_1 &= \left\{z=(\eta_0,\ldots,\eta_K):\exists\mathcal{P}\in\mathcal{F}_0^{(N)}\ni f_j(\mathcal{P})<\eta_j,\ \forall j=0\ldots K\right\}\\ C_2 &= \left\{z=(\eta_0,\ldots,\eta_K):\eta_j\leq 0,\ \forall j=0\ldots K\right\}.\end{aligned}$$

It is seen that C_1 is a convex nonempty set in $\mathbb{R}^{K+1}$ and $C_1\cap C_2\neq\phi$. By the separation theorem for convex sets (Theorem 11.3 in [14]) there exist $\lambda_0,\ldots,\lambda_K$, not all zero and real a such that

$$\sum_{j=0}^{K}\lambda_j\eta_j \ \geq\ a,\ \ \forall z\in C_1, \tag{124}$$

$$\sum_{j=0}^{K}\lambda_j\eta_j \ \leq\ a,\ \ \forall z\in C_2. \tag{125}$$

Now (125) implies that $a\geq 0$ and $\lambda_j\geq 0$, $\forall j=0\ldots K$. Fix $\mathcal{P}\in\mathcal{F}_0^{(N)}$. Since $f_j(\mathcal{P})$ is finite for every $j=0\ldots K$ we have for every $\epsilon>0$ that

$$\left(f_0(\mathcal{P})+\epsilon,\ldots,f_K(\mathcal{P})+\epsilon\right)\in C_1$$

and substituting this in (124) we have

$$\sum_{j=0}^{K}\lambda_j\left(f_j(\mathcal{P})+\epsilon\right)\geq 0,\ \forall\epsilon>0,\ \forall\mathcal{P}\in\mathcal{F}_0^{(N)}.$$

Since this is true for every arbitrary $\epsilon>0$, we have shown (119). This completes the proof of Proposition 4.4. □

E Proof of Proposition 4.5

Fix one realization of fading gains $h_1, \ldots, h_K$ and signature sequences S. Since the map C in (28) is concave, any tuple of of powers (denoted by $(p_1^*, \ldots, p_K^*)$) that maximizes

$$C(p_1, \ldots, p_K) - \frac{\lambda}{2N} \sum_{i=1}^{K} (p_i - \bar{p}) \tag{126}$$

in the positive orthant of $\mathbb{R}^K$ has the following structure:

$$\frac{\partial C}{\partial p_i}(p_1^*, \ldots, p_K^*) = \lambda \tag{127}$$

$$p_i^* = \left(\frac{1}{\lambda} - \frac{1}{s_i^t \left(\sigma^2 I + \sum_{j \neq i} s_j s_j^t h_j p_j^* \right)^{-1} s_i h_i} \right)^+ \quad \forall i = 1 \ldots K \; . \tag{128}$$

The derivation of (128) from (127) is completely analogous to that of (21). If the realization $h_1, \ldots, h_K, S$ is such that $\{h_i s_i s_i^t, i = 1 \ldots K\}$ is a linearly independent set then C is strictly concave and the solution $p_1*, \ldots, p_K^*$ in (128) is unique. In general the solution set is a nonempty convex set.

We now construct a power allocation policy that is equal to $p_1^*, \ldots, p_K^*$ at the realization $h_1, \ldots, h_K, S$. If there are no point masses in the distribution F and in the common distribution of v_{ij} then with probability 1 we have $\{h_i s_i s_i^t,\; i = 1 \ldots K\}$ are linearly independent and C is strictly concave. In this case the tuple $(p_1^*, \ldots, p_K^*)$ is uniquely defined almost everywhere (the value depends on the realization of fading gains and signature sequences). In this scenario, we define the power allocation $\mathcal{P}_i^*$ for every user i as:

$$\mathcal{P}_i^*(h_1, \ldots, h_K, S) \stackrel{\text{def}}{=} p_i^* \; . \tag{129}$$

If there are point masses in F and the common distribution of v_{ij} such that there is a positive probability of $\{h_i s_i s_i^t,\; i = 1 \ldots K\}$ being linearly dependent, then on these realizations, the solution set $(p_1^*, \ldots, p_K^*)$ is closed and convex and we select any of these points to be $\mathcal{P}^*$ at that realization of fading gains and signature sequences. Since there is ambiguity in $\mathcal{P}^*$ only on point masses, we still have $\mathcal{P}_i^*$ a measurable function of $h_1, \ldots, h_K, S$ for each $i = 1 \ldots K$. More generally, we can appeal to general measurable selection theorems ([30] is a good review on these results; Theorem 3.1 is relevant to our case) to select a measurable $\mathcal{P}^*$ that satisfies the property (128) at almost every realization of fading gains and sequences. Since for (almost) every realization of fading states and signature sequences $\mathcal{P}^*$ is the maximizer of the map in (126), it follows from Proposition 4.4 that

$$\bar{C}_{opt}^{(N)} = L_\lambda(\mathcal{P}^*) = \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} L_\lambda(\mathcal{P}) \tag{130}$$

where L_λ maps $\mathcal{F}_0^{(N)}$ to the reals as

$$L_\lambda : \mathcal{P} \mapsto \bar{C}_{sum}(\mathcal{P}) - \frac{\lambda}{2N} \sum_{i=1}^{K} \left(\mathbb{E}\left[\mathcal{P}_i(h_1, \ldots, h_K, S)\right] - \bar{p} \right) \; .$$

Furthermore, it follows for any $\mathcal{P} \in \mathcal{F}_2^{(N)}$ that

$$L_\lambda(\mathcal{P}) < \bar{C}_{opt}^{(N)} \text{ for } \mathcal{P} \text{ not satisfying (128) on realizations of positive measure.} \tag{131}$$

Thus if we can show the existence of a power allocation policy $\tilde{\mathcal{P}}^* \in \mathcal{F}_2^{(N)}$ where the supremum of (9) is achieved, the claim of this proposition follows from Proposition 4.3 and (131). We now show the existence of such a power allocation policy.

Fix a realization $h_1, \ldots, h_K, S$ and consider $p_1^*, \ldots, p_K^*$ defined in (128). Since each of the p_i^* is bounded from above (by λ^{-1}) it follows that $\mathcal{P}^* \in \mathcal{F}_0^{(N)}$ and furthermore

$$\mathbb{E}\left[\mathcal{P}_1^*(h_1, \ldots, h_K.S)\right] = \cdots = \mathbb{E}\left[\mathcal{P}_K^*(h_1, \ldots, h_K, S)\right] \geq \bar{p}$$

for each $i = 1 \ldots K$. We used Proposition 4.3 in the observation that $\mathbb{E}[\mathcal{P}_i^*]$ cannot be less than $\bar{p}$ for any $i = 1 \ldots K$. From (130) we conclude that if we can show that $\mathbb{E}[\mathcal{P}_1^*] = \bar{p}$, we have proved the claim of this proposition that

$$\bar{C}_{opt}^{(N)} = L_\lambda(\mathcal{P}^*) = \bar{C}_{sum}(\mathcal{P}^*) \ .$$

Fix $\mu > 0$ and let us denote the (measurably selected) power allocation policy $\mathcal{P}^{*\mu}$ which maximizes L_μ in $\mathcal{F}_0^{(N)}$. In the previous notation $\mathcal{P}^*$ maximizes L_λ. We begin with the following claim for any $0 < a < b$:

$$g : \mu \mapsto \mathbb{E}\left[\mathcal{P}_1^{*\mu}(h_1, \ldots, h_K, S)\right] \quad \text{is continuous on } \mu \in [a, b] \ . \tag{132}$$

Suppose true. Now $g(\mu) < \frac{1}{\mu}$ and thus as $\mu \to \infty$ we arrive at $g(\mu) \to 0$. From Propositions 4.4 and 4.3 we have, for every $p > 0$, that there exists $\mu_p > 0$ such that $g(\mu_p) \geq p$. Using (132), given $\bar{p}$ we have $\tilde{\lambda}$ such that $g\left(\tilde{\lambda}\right) = \bar{p}$. Observe that

$$\begin{aligned}
\bar{C}_{opt}^{(N)} &= \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \bar{C}_{sum}(\mathcal{P}) \leq sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} L_{\tilde{\lambda}} \\
&\leq \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} L_{\tilde{\lambda}}(\mathcal{P}) \leq L_{\tilde{\lambda}}\left(\mathcal{P}^{*\tilde{\lambda}}\right) \\
&= \bar{C}_{sum}\left(\mathcal{P}^{*\tilde{\lambda}}\right)
\end{aligned}$$

where we have used the hypothesis that $g\left(\mathcal{P}^{*\tilde{\lambda}}\right) = \bar{p}$ in the derivation of the last step. Thus

$$\bar{C}_{opt}^{(N)} = \bar{C}_{sum}\left(\mathcal{P}^{*\tilde{\lambda}}\right) \ . \tag{133}$$

We will now show that $\tilde{\lambda}$ must equal λ (proposed by Proposition 4.4) and complete the proof. By the concavity of L_λ, for any $\mathcal{P} \in \mathcal{F}_0^{(N)}$ that does not satisfy (32) on realizations (of fading gains and signature sequences) with positive probability measure, we have $L_\lambda(\mathcal{P}) < L_\lambda(\mathcal{P}^*)$. Using (133) and Proposition 4.4 we arrive at $\lambda = \tilde{\lambda}$. It only remains to show the claim in (132). We only show this for the case when C is strictly concave for almost every realization of $h_1, \ldots, h_K, S$. The extension to the general case when there are realizations of positive measure which lead to non-strict concavity of C is not pursued here. Fix $0 < a < b$ and a realization of $h_1, \ldots, h_K, S$. We first observe that the map

$$G : \mu \mapsto \left(p_1^{*\mu}, \ldots, p_K^{*\mu}\right) \stackrel{\text{def}}{=} \arg \max_{p_i \geq 0, i=1\ldots K} C\left(p_1, \ldots, p_K\right) - \frac{\mu}{2N} \sum_{i=1}^{K} \left(p_i - \bar{p}\right)$$

is continuous for every realization of $h_1, \ldots, h_K, S$ such that C is strictly concave. For such realizations, G is invertible and we have (from (32))

$$s_j^t \left(\sigma^2 I + \sum_{i=1}^{K} h_i p_i^{*\mu} s_i s_i^t\right)^{-1} s_j h_j = \mu, \ \forall j = 1 \ldots K\,. \tag{134}$$

Fix $a \leq \mu \leq b$ and consider $\mu_n \to \mu$ in $[a, b]$ as $n \to \infty$. Observe that the image of $[a, b]$ under G in the positive orthant of $\mathbb{R}^K$ is contained in the box $[0, a^{-1}] \times \cdots [0, a^{-1}]$. Furthermore, the image is closed (using (134)) and thus compact. Now consider the sequence $\{(p_1^{*\mu_n}, \ldots, p_K^{*\mu_n})\}_{n \geq 1}$ in the compact image $G[a, b]$. There exists a subsequence $\{p^{*\mu_{i_n}}\}_{n \geq 1}$ and some $\tilde{\mu} \in [a, b]$ and $p^{*\tilde{\mu}}$ such that $p^{*\mu_{i_n}} \to p^{*\tilde{mu}}$. From the continuity of the inverse of G (using (134)) we arrive at $\mu_{i_n} \to \tilde{\mu}$. By hypothesis, $\mu_n \to \mu$ and thus $\mu = \tilde{\mu}$ allows us to conclude that $G(\mu_n) \to G(\mu)$ showing the continuity of G. Thus for almost every realization, we have shown continuity of G. Fix $\epsilon > 0$ and by Egoroff's theorem (Theorem 3.6.23 in [15]), we have uniform continuity of the map

$$\mu \mapsto \left(p_1^{*\mu}, \ldots, p_K^{*\mu}\right)$$

on a set $\mathcal{E}$ such that $\mathbb{P}\left[(h_1, \ldots, h_K, S) \notin \mathcal{E}\right] < \frac{\epsilon}{2a}$. Hence there exists n_0 such that $\forall n \geq n_0$, we have on $\mathcal{E}$,

$$\mid \mathcal{P}^{*\mu_n} - \mathcal{P}^{*\mu} \mid (h_1, \ldots, h_K, S) < \frac{\epsilon}{2}\,. \tag{135}$$

Then,

$$\begin{aligned} \mid g(\mu_n) - g(\mu) \mid &\leq \mathbb{E}\left[\mid \mathcal{P}^{*\mu_n} - \mathcal{P}^{*\mu} \mid\right] \\ &< \frac{\epsilon}{2} + \mathbb{E}\left[\mid \mathcal{P}^{*\mu_n} - \mathcal{P}^{*\mu} \mid 1_{\{(h_1, \ldots, h_K, S) \in \mathcal{E}\}}\right] \\ &< \epsilon, \ \forall n \geq n_0 \end{aligned}$$

where we used (135) in the last step and the fact for every $a \leq \mu \leq b$ that $\mathcal{P}^{*\mu} < a^{-1}$ in the second step. Since ϵ is arbitrary, we have completed the proof of (132). □

F Proof of Theorem 4.7

Fix the processing gain N and the number of users $K = \lfloor \alpha N \rfloor$. From the argument following Proposition 4.4 and (128) we know that any optimal power allocation has the following structure:

$$\mathcal{P}_i^*\left(h_1, \ldots, h_K, S\right) = \left(\frac{1}{\lambda^{(N)}} - \frac{1}{s_i^t \left(\sigma^2 I + \sum_{j \neq i} s_j s_j^t h_j \mathcal{P}_j^*\left(h_1, \ldots, h_K, S\right)\right)^{-1} s_i h_i}\right)^+ \quad \forall i = 1 \ldots K\,.$$

Here the notation $\lambda^{(N)}$ emphasizes the dependence of (the Kuhn-Tucker coefficient) λ on N. Thus we have $\mathcal{P}_i^* \leq \frac{1}{\lambda^{(N)}}$ *a.s.* and if we can show that

$$\inf_N \left\{\lambda^{(N)} ; N > 0\right\} \stackrel{\text{def}}{=} \frac{1}{K_p} > 0$$

the proof is complete. We now show that $\lambda^{(N)}$ is uniformly lower bounded (uniform in N). Denote (static) power allocations that allocate constant power (say p) for every realization of the fading and signature sequence by $\bar{\mathcal{P}}(p)$. The sum capacity with this static power allocation converges pointwise to a nonzero constant in a large system. Formally,

$$\bar{C}_{sum}\left(\bar{\mathcal{P}}(p)\right) \longrightarrow \bar{C}^s_{sum}(p) > 0 \text{ as } N \to \infty . \tag{136}$$

Using results about eigenvalues of large random matrices, we show a more general version of this result in Proposition 3.1 and $\bar{C}^s_{sum}(p)$ has an explicit expression given in (26). It also follows from this result that $\bar{C}^s_{sum}(p) \to \infty$ as $p \to \infty$. Some simple monotonicity properties of $\bar{C}_{sum}$ and $\bar{C}^s_{sum}$ are as follows:

$$\bar{C}_{sum}\left(\bar{\mathcal{P}}(p_1)\right) > \bar{C}_{sum}\left(\bar{\mathcal{P}}(p_2)\right) \text{ whenever } p_1 > p_2 \text{ for each fixed } N. \tag{137}$$

$$\bar{C}^s_{sum}(p_1) > \bar{C}^s_{sum}(p_2) \text{ whenever } p_1 > p_2 \tag{138}$$

We fix $\bar{\mu}$ such that

$$\bar{C}^s_{sum}\left(\frac{1}{\bar{\mu}}\right) > \alpha\left(K_c + 0.5\right) \tag{139}$$

where K_c is equal to $\log 2 + \frac{1}{2}\log\left(\bar{h}\bar{p}\right)$ defined in the proof of Proposition 4.1. Defining the function on the positive reals

$$g_N(\mu) \stackrel{\text{def}}{=} \sup_{\mathcal{P}\in\mathcal{F}_0^{(N)}} \left\{\bar{C}_{sum}(\mathcal{P}) - \frac{\mu}{2N}\sum_{j=1}^{K}\left(\mathbb{E}\mathcal{P}_j - \bar{p}\right)\right\}$$

we recognize from (123) that $g_N(\mu) \geq \bar{C}^{(N)}_{opt}, \forall \mu \geq 0$. By definition of $\lambda^{(N)}$, from (30) we conclude that

$$g_N\left(\lambda^{(N)}\right) = \min_{\mu\geq 0} g_N(\mu) = \bar{C}^{(N)}_{opt};. \tag{140}$$

Now suppose $\inf_N \lambda^{(N)} = 0$. Then there is a subsequence $\{i_n\}_n$ such that $\lim_{n\to\infty}\lambda^{(i_n)} = 0$ and an integer n_0 such that $\lambda^{(i_n)} < \bar{\mu}$ for all $n > n_0$. By definition, we arrive at

$$\bar{C}^{(i_n)}_{opt} = g_{i_n}\left(\lambda^{(i_n)}\right) \geq \bar{C}_{sum}\left(\bar{\mathcal{P}}\left(\frac{1}{\lambda^{(i_n)}}\right)\right) - \frac{\alpha}{2} . \tag{141}$$

In (141) the power allocation $\bar{\mathcal{P}}\left(\frac{1}{\lambda^{(i_n)}}\right)$ allocates constant power equal to $\frac{1}{\lambda^{(i_n)}}$ for every realization of signature sequences and fading states (recall notation from Section 4.2). Furthermore,

$$\begin{aligned}\liminf_{n\to\infty} \bar{C}_{sum}\left(\bar{\mathcal{P}}\left(\frac{1}{\lambda^{(i_n)}}\right)\right) &\geq \liminf_{n\to\infty} \bar{C}_{sum}\left(\bar{\mathcal{P}}\left(\frac{1}{\bar{\mu}}\right)\right) \\ &= \bar{C}^s_{sum}\left(\frac{1}{\bar{\mu}}\right) \end{aligned} \tag{142}$$

where we used (136) and (137). Combining (141), (142) and (139) we arrive at a contradiction to Proposition 4.1. Thus the Kuhn-Tucker coefficient $\lambda^{(N)}$ is uniformly (in N) lower bounded and denoting the lower bound as K_p^{-1} the proof is complete. □

G Proof of Lemma 4.9

The essential ingredients of the proof are all contained in Lemmas 4.3 and 4.4 of [23] and we only indicate the key points of departure. In particular, a close study of Lemmas 3.2, 4.3 and 4.4 of [23] reveals that the statement made as Lemma 4.9 in this paper is true for the situation when $h_i = 1$ *a.s.* and $g(\cdot) = 1$. Below, we keep consistency with the notation of [23] and point out the main steps in generalizing the results to the general case here. We use $K_p, C_i, i = 2, .., 21$ to denote constants that are independent of N.[3]

Let $\beta_i^{(N)} = s_i^t Z_i^{-1} s_i$ where, (recall notation from (47)) $Z_i = \left(\sigma^2 I + \sum_{l\neq i} s_l s_l^t h_l g\left(h_l\right)\right)$. Let

$$\tilde{\beta}_i^{(N)} = \frac{1}{N}\mathrm{tr} Z_i^{-1} \text{ and } \beta_o^{(N)} = \frac{1}{N}\mathrm{tr} Z^{-1}$$

where $Z = Z_i + s_i s_i^t h_i g\left(h_i\right)$. Let $\bar{\beta}_i^{(N)}$ and $\bar{\beta}_o^{(N)}$ denote $\mathbb{E}\left[\beta_i^{(N)}\right]$ and $\mathbb{E}\left[\beta_o^{(N)}\right]$ respectively. In this notation, we need to prove that

$$\mathbb{E}\left[\left(\beta_1^{(N)} - \beta_g^*\right)^2\right] \leq \frac{C_2^2}{N}. \tag{143}$$

We show (143) by the following sequence of bounds:

$$\begin{aligned}
\mathbb{E}\left[\left(\beta_1^{(N)} - \tilde{\beta}_1^{(N)}\right)^2\right] &\leq \frac{C_3}{N} && (144)\\
\mathrm{Var}\left(\tilde{\beta}_1^{(N)}\right) &\leq \frac{C_4}{N} && (145)\\
\mid \mathbb{E}\left[\tilde{\beta}_1^{(N)}\right] - \beta^* \mid &\leq \frac{C_5}{N} && (146)
\end{aligned}$$

Now, from Lemma 3.2 of [23], it follows that

$$\begin{aligned}
\mathbb{E}\left[\left(\beta_1^{(N)} - \tilde{\beta}_1^{(N)}\right)^2 \mid s_2, \ldots, s_K, \mathbf{H}\right] &\leq \frac{C_6}{N}\left(\lambda_{\max}^2\left(Z_1^{-1}\right)^2\right)\\
&\leq \frac{C_6}{N}\frac{1}{\sigma^4}
\end{aligned}$$

showing (144). To show (145) we closely follow the proof of Lemma 4.3 of [23]. Let $A = Z_1^{-1}$ and $A_j = \left(Z_1 - s_j s_j^t h_j g\left(h_j\right)\right)^{-1}$ for $j \geq 2$. Let $\mathbb{E}_j\left[\cdot\right]$ denote the conditional expectation

[3]The constant K_p used in the proof of this lemma is unrelated to the uniform upper bound K_p on the optimal power policies which was identified in Theorem 4.7; we use it here to keep consistency in our notation with that of [23].

$\mathbb{E}[\cdot \mid s_l, h_l, 2 \le l \le j]$. We denote the received powers as $q_j = h_j g(h_j)$ and $\bar{q} = \mathbb{E}[q_j]$. Using the matrix inversion lemma,

$$\mathbb{E}\mathrm{tr} A - \mathrm{tr} A = \sum_{j=2}^{K} (\mathbb{E}_j - \mathbb{E}_{j-1}) \frac{s_j^t A_j^2 s_j q_j}{1 + s_j^t A_j s_j q_j} \ .$$

Denoting

$$a_j = N^{-1}\mathrm{tr} A_j^2 \ , \ \alpha_j = s_j^t A_j^2 s_j q_j - \bar{q} a_j \ , \ \omega_j = \frac{1}{1 + s_j^t A_j s_j q_j} \ , \ b_N = \frac{1}{1 + \bar{q} N^{-1} \mathbb{E}\mathrm{tr} A_j}$$

$$\zeta_j = s_j^t A_j s_j q_j - \bar{q} N^{-1} \mathbb{E}\mathrm{tr} A_j \ , \ \hat{\zeta}_j = s_j^t A_j s_j q_j - \bar{q} N^{-1} \mathrm{tr} A_j$$

and using some algebra only slightly modified from that in the proof of Lemma 4.3 in [23], we get

$$\begin{aligned} \sum_{j=2}^{K} (\mathbb{E}_j - \mathbb{E}_{j-1}) \frac{s_j^t A_j^2 s_j q_j}{1 + s_j^t A_j s_j q_j} &= b_N \sum_{j=2}^{K} \mathbb{E}_j [\alpha_j] - b_N^2 \sum_{j=2}^{K} \mathbb{E}_j \left[a_j \hat{\zeta}_j \right] \\ & \quad - b_N^2 \sum_{j=2}^{K} (\mathbb{E}_j - \mathbb{E}_{j-1}) \left[\alpha_j \zeta_j - s_j^t A_j^2 s_j q_j \omega_j \zeta_j^2 \right] \\ & \stackrel{\mathrm{def}}{=} W_1 - W_2 - W_3 \ . \end{aligned}$$

Since b_N is uniformly (in N) upper and lower bounded (since h_j is bounded by $\bar{h}$ and g is a bounded function, bounded by K_g), it suffices to estimate $\mathbb{E}[W_1^2/b_N]$ and $\mathbb{E}[W_i^2/b_N^4]$, $i = 2, 3$. We begin with the following key estimates, for $p = 2, 4$:

$$\mathbb{E}[\mid \alpha_j \mid^p] \le C_7 \ . \tag{147}$$

$$\mathbb{E}\left[\mid \hat{\zeta}_j \mid^p\right] \le C_8 \ . \tag{148}$$

$$\mathbb{E}[\mid \zeta_j \mid^p] \le C_9 \ . \tag{149}$$

From Lemma B.1 of [23] (as in Eq (33) of ([23])) we have for $p = 2, 4$

$$\mathbb{E}\left[\left(s_j^t A_j^2 s_j - N^{-1} \mathrm{tr} A_j^2 \right)^p \right] \le C_{10} N^{-\frac{p}{2}} \ . \tag{150}$$

To see (147) with $p = 2$, observe that

$$\begin{aligned} \mathbb{E}\left[\alpha_j^2\right] &= \mathbb{E}\left[\left(s_j^t A_j^2 s_j \right)^2 \right] \mathbb{E}\left[q_j^2\right] - \mathbb{E}\left[\left(N^{-1} \mathrm{tr} A_j^2 \right)^2 \right] \bar{q}^2 \\ &= \mathbb{E}\left[\left(s_j^t A_j^2 s_j - N^{-1} \mathrm{tr} A_j^2 \right)^2 \right] \mathbb{E}\left[q_j^2\right]^2 + \mathbb{E}\left[\left(N^{-1} \mathrm{tr} A_j^2 \right)^2 \right] \left(\mathbb{E}\left[q_j^2\right] - \bar{q}^2 \right) \\ &\le \frac{C_{10} \bar{h}^2 K_g^2}{N} + \frac{\bar{h}^2 K_g^2}{\sigma^8} \end{aligned}$$

where we used (150) and the fact that $\mathrm{tr} A_j^k \le N \sigma^{-2k}$ in the derivation of the last step. This shows (147) for $p = 2$. Also,

$$\mathbb{E}\left[\alpha_j^4\right] = \mathbb{E}\left[\left(s_j^t A_j^2 s_j \right)^4 \right] \mathbb{E}\left[q_j^4\right] - 3\mathbb{E}\left[\left(N^{-1} \mathrm{tr} A_j^2 \right)^4 \right] \bar{q}^4$$

$$
\begin{aligned}
&+6\mathbb{E}\left[\left(s_j^tA_j^2s_j\right)^2\left(N^{-1}\mathrm{tr}A_j^2\right)^2\right]\bar{q}^2\mathbb{E}\left[q_j^2\right]-4\mathbb{E}\left[\left(s_j^tA_j^2s_j\right)^3N^{-1}\mathrm{tr}A_j^2\right]\bar{q}\mathbb{E}\left[q_j^3\right]\\
=\;&\mathbb{E}\left[\left(s_j^tA_j^2s_j-N^{-1}\mathrm{tr}A_j^2\right)^4\right]+3\mathbb{E}\left[\left(N^{-1}\mathrm{tr}A_j^2\right)^4\right]\left(\mathbb{E}\left[q_j^4\right]-\bar{q}^4\right)\\
&-6\mathbb{E}\left[\left(s_j^tA_j^2s_j\right)^2\left(N^{-1}\mathrm{tr}A_j^2\right)^2\right]\left(\mathbb{E}\left[q_j^4\right]-\bar{q}^2\mathbb{E}\left[q_j^2\right]\right)\\
&+4\mathbb{E}\left[\left(s_j^tA_j^2s_j\right)^3N^{-1}\mathrm{tr}A_j^2\right]\left(\mathbb{E}\left[q_j^4\right]-\bar{q}\mathbb{E}\left[q_j^3\right]\right)\\
\le\;&C_{10}N^{-2}+\frac{3\bar{h}^4K_g^4}{\sigma^{16}}+\frac{4\bar{h}^4K_g^4\mathbb{E}\left[\left(s_j^ts_j\right)^3\right]}{\sigma^{16}}
\end{aligned}
$$

where we used (150) in the last step. This shows (147) for $p=4$.

From Eq (34) of [23] we have, for $p=2,4$,

$$
\mathbb{E}\left[\left(s_j^tA_js_j-N^{-1}\mathrm{tr}A_j\right)^p\right]\le C_{11}N^{-\frac{p}{2}}\,. \tag{151}
$$

Observing the similarity in definition of α_j and $\hat{\zeta}_j$ and in equations (150) and (151), we have, completely analogous to the calculation for α_j above,

$$
\begin{aligned}
\mathbb{E}\left[\hat{\zeta}_j^2\right]&\le\frac{C_{11}\bar{h}^2K_g^2}{N}+\frac{\bar{h}^2K_g^2}{\sigma^4}\\
\mathbb{E}\left[\hat{\zeta}_j^4\right]&\le\frac{C_{11}}{N^2}+\frac{3\bar{h}^4K_g^4}{\sigma^8}+\frac{4\bar{h}^4K_g^4\mathbb{E}\left[\left(s_j^ts_j\right)^3\right]}{\sigma^8}
\end{aligned}
$$

thus showing (148).

Now,

$$
\begin{aligned}
\mid\zeta_j-\hat{\zeta}_j\mid^p&=\mid\frac{\bar{q}}{N}\left(\mathrm{tr}A_j-\mathbb{E}\left[\mathrm{tr}A_j\right]\right)\mid^p\\
&\le\mid\frac{\bar{q}}{N}\left(\mathrm{tr}A_j+\mathbb{E}\left[\mathrm{tr}A_j\right]\right)\mid^p\\
&\le\left(\frac{2\bar{q}}{\sigma^2}\right)^p,
\end{aligned}
$$

and an appeal to (148) shows (149).

We now return to the estimates on $W_i, i=1,2,3$. As in the proof of (Lemma 4.5,[23]), we have that $\{\mathbb{E}_j\left[\alpha_j\right]\}$ is a martingale difference sequence and using the Burkholder inequality (Eq (30),[23]), we have, identical to Eq (32) of [23],

$$
\begin{aligned}
\mathbb{E}\left[W_1^2/b_N^2\right]&\le K_p\mathbb{E}\left[\sum_{j=1}^{K}\left(\mathbb{E}_j\left[\alpha_j\right]\right)^2\right]\\
&\le K_p\sum_{j=2}^{K}\mathbb{E}\left[\alpha_j^2\right]\\
&\le\alpha K_pC_7N
\end{aligned}
\tag{152}
$$

where we used (147) in the last step. We have that $\left\{\mathbb{E}_j\left[a_j\hat{\zeta}_j\right]\right\}$ is a martingale difference sequence and using the Burkholder inequality (Eq (30),[23]), we have as in [23]

$$\begin{aligned}
\mathbb{E}\left[W_2^2/b_N^4\right] &= \mathbb{E}\left[\left(\sum_{j=2}^{K}\mathbb{E}_j\left[a_j\hat{\zeta}_j\right]\right)^2\right] \\
&\leq C_{12}\mathbb{E}\left[\sum_{j=1}^{K}\mathbb{E}_{j-1}\left[\left(\mathbb{E}_j\left[a_j\hat{\zeta}_j\right]\right)^2\right]+\left(\mathbb{E}_j\left[a_j\hat{\zeta}_j\right]\right)^2\right] \\
&\leq \frac{2C_{12}}{\sigma^8}\sum_{j=1}^{K}\mathbb{E}\left[\hat{\zeta}_j^2\right] \\
&\leq \frac{2\alpha C_{12}C_8}{\sigma^8}N
\end{aligned} \tag{153}$$

where we used (148) in the last step. The bound involving W_3 is very similar and we have as in [23],

$$\begin{aligned}
\mathbb{E}\left[W_3^2/b_N^4\right] &= \mathbb{E}\left[\left(\sum_{j=2}^{K}\left(\mathbb{E}_j-\mathbb{E}_{j-1}\right)\left[\alpha_j\zeta_j-s_j^tA_j^2s_jq_j\omega_j\zeta_j^2\right]\right)^2\right] \\
&\leq C_{13}\sum_{j=2}^{K}\left(\mathbb{E}\left[\alpha_j^2\zeta_j^2\right]+\mathbb{E}\left[\zeta_j^4\right]\right) \\
&= C_{13}(K-1)\left(\mathbb{E}\left[\alpha_2^2\zeta_2^2\right]+\mathbb{E}\left[\zeta_2^4\right]\right) \\
&\leq C_{13}\alpha\left(\sqrt{C_9C_7}+C_9\right)N\,.
\end{aligned} \tag{154}$$

Thus we conclude that

$$\begin{aligned}
\mathrm{Var}\left(\tilde{\beta}_1^{(N)}\right) &= \frac{1}{N^2}\mathrm{Var}\left(\mathrm{tr}A\right) \\
&= \frac{1}{N^2}\mathbb{E}\left[\left(W_1-W_2-W_3\right)^2\right] \\
&\leq \frac{C_{14}}{N}
\end{aligned}$$

where we used (152), (153) and (154) in the last step. This completes the proof of (145).

To see (146) we closely follow the proof of Lemma 4.4 in [23] and indicate below only the major deviations. Using the fact that h_i is bounded by $\bar{h}$ and that g is a bounded function (bounded above by K_g), following the proof of Lemma 4.4 in [23] we get

$$\mathrm{Var}\left(\beta_i^{(N)}\right)\leq\frac{C_{15}}{N} \quad\text{and}\quad \mid\bar{\beta}_i^{(N)}-\bar{\beta}_o^{(N)}\mid\leq\frac{\bar{h}K_g\alpha}{\sigma^8N}=\frac{C_{16}}{N}\,. \tag{155}$$

Writing $\beta_i^{(N)}=\bar{\beta}_o^{(N)}+\Delta_i^{(N)}$, we obtain as in Eq (41) of [23]

$$\mid\mathbb{E}\left[\Delta_i^{(N)}\right]\mid\leq\frac{C_{17}}{N} \quad\text{and}\quad \mathbb{E}\left[\left(\Delta_i^{(N)}\right)^2\right]\leq\frac{C_{18}}{N}\,. \tag{156}$$

Recalling equation (27) of [22] relating the SIRs attained with $\beta_o^{(N)}$ we have, almost surely, as in Eq (39) of [23],

$$\frac{1}{N}\sum_{i=1}^{K}\frac{\beta_i^{(N)}h_i g(h_i)}{1+\beta_i^{(N)}h_i g(h_i)} = 1-\sigma^2\beta_o^{(N)} \ . \tag{157}$$

In our notation, (157) can be written as

$$\frac{K}{N}-\frac{1}{N}\sum_{i=1}^{K}\frac{1}{1+\left(\bar{\beta}_o^{(N)}+\Delta_i^{(N)}\right)h_i g(h_i)} = 1-\sigma^2\beta_o^{(N)} \ . \tag{158}$$

Writing $\nu_i^{(N)} = \frac{\Delta_i^{(N)}}{1+\bar{\beta}_o^{(N)}h_i g(h_i)}$ we have

$$\begin{aligned}\frac{1}{1+\left(\bar{\beta}_o^{(N)}+\Delta_i^{(N)}\right)h_i g(h_i)} &= \frac{1}{1+\bar{\beta}_o^{(N)}h_i g(h_i)}\cdot\frac{1}{1+\nu_i^{(N)}h_i g(h_i)} \\ &= \frac{1}{1+\bar{\beta}_o^{(N)}h_i g(h_i)}\left(1-\nu_i h_i g(h_i)+\frac{2\left(\nu_i h_i g(h_i)\right)^2}{\left(1+\xi_i\right)^3}\right)\end{aligned} \tag{159}$$

for some $\xi_i \in \left[0,\nu_i^{(N)}\right]\cup\left[\nu_i^{(N)},0\right]$. Now

$$\begin{aligned}\nu_i^{(N)} &= \frac{\beta_i^{(N)}-\bar{\beta}_o^{(N)}}{1+\bar{\beta}_o^{(N)}h_i g(h_i)} \geq \frac{-\bar{\beta}_o^{(N)}}{1+\bar{\beta}_o^{(N)}h_i g(h_i)} \\ &\geq \frac{-1}{\sigma^2+h_i g(h_i)}\end{aligned}$$

and thus $(1+\xi_i)^{-3}\leq C_{19}$. Substituting this upper bound in (159) and integrating (158) we arrive at

$$\left|\frac{K}{N}-\frac{1}{N}\int_0^{\bar{h}}\frac{dF(h)}{1+\bar{\beta}_o^{(N)}hg(h)}-1+\sigma^2\bar{\beta}_o^{(N)}\right| \leq \frac{C_{20}}{N}$$

and

$$\left|\alpha\int_0^{\bar{h}}\frac{\bar{\beta}_o^{(N)}hg(h)\,dF(h)}{1+\bar{\beta}_o^{(N)}hg(h)}-1+\sigma^2\bar{\beta}_o^{(N)}\right| \leq \frac{C_{20}}{N}. \tag{160}$$

Now consider the map

$$f:\ \beta\mapsto\sigma^2\beta-1+\alpha\int_0^{\bar{h}}\frac{\beta hg(h)}{1+\beta hg(h)}dF(h) \ .$$

From (160) and Lemma 4.8 we have

$$\left|f\left(\bar{\beta}_o^{(N)}\right)\right| \leq \frac{C_{20}}{N},\ \ \forall N \text{ and } f\left(\beta_g^*\right)=0 \ . \tag{161}$$

Now,

$$\begin{aligned} f(\beta_1) - f(\beta_2) &= \sigma^2(\beta_1 - \beta_2) + \alpha \int_0^{\bar{h}} \left(\frac{\beta_1 hg(h)}{1+\beta_1 hg(h)} - \frac{\beta_2 hg(h)}{1+\beta_2 hg(h)} \right) dF(h) \\ &= (\beta_1 - \beta_2) \left\{ \sigma^2 + \alpha \int_o^{\bar{h}} \frac{hg(h)}{(1+\beta_1 hg(h))(1+\beta_2 hg(h))} dF(h) \right\} \\ \mid f\left(\bar{\beta}_o^{(N)}\right) - f\left(\beta_g^*\right) \mid &\geq \mid \bar{\beta}_o^{(N)} - \beta_g^* \mid \left\{ \sigma^2 + \alpha \int_o^{\bar{h}} hg(h)\, dF(h) \right\} \\ &\geq C_{21} \mid \bar{\beta}_o^{(N)} - \beta_g^* \mid . \end{aligned} \tag{162}$$

Combining (161) and (162) we have shown (146) completing the proof of Lemma 4.9. □

H Proof of Lemma 4.10

A key observation from the quadratic $\beta^*(\bar{p}, \alpha)$ satisfies (in (45)) is the following:

$$\mathcal{K}\left(\tilde{h}\right) \geq h \Longleftrightarrow \left(\frac{\sigma^2}{\bar{p}h} + \alpha\right)\left(1 - F\left(\tilde{h}\right) - \int_{\tilde{h}}^{\bar{h}} \frac{h}{h_0} dF(h_0)\right) \geq 1 . \tag{163}$$

To see this, define

$$\tilde{p} \stackrel{\text{def}}{=} \frac{\bar{p} HM\left(\tilde{h}\right)}{\left(1 - F\left(\tilde{h}\right)\right)} \text{ and } \tilde{\alpha} \stackrel{\text{def}}{=} \alpha\left(1 - F\left(\tilde{h}\right)\right)$$

Now,

$$\begin{aligned} \mathcal{K}\left(\tilde{h}\right) \geq h &\Longleftrightarrow HM\left(\tilde{h}\right) \geq h + \tilde{p}h\beta^*(\tilde{p}, \tilde{\alpha}) \\ &\Longleftrightarrow \frac{HM\left(\tilde{h}\right) - h}{h\tilde{p}} \geq \beta^*(\tilde{p}, \tilde{\alpha}) && (164) \\ &\Longleftrightarrow \sigma^2 \frac{HM\left(\tilde{h}\right) - h}{h\tilde{p}} + \frac{\tilde{\alpha}\tilde{p}\frac{HM(\tilde{h})-h}{h\tilde{p}}}{1 + \tilde{p}\frac{HM(\tilde{h})-h}{h\tilde{p}}} \geq 1 && (165) \\ &\Longleftrightarrow \sigma^2 \frac{HM\left(\tilde{h}\right) - h}{h\tilde{p}} + \tilde{\alpha}\frac{HM\left(\tilde{h}\right) - h}{HM\left(\tilde{h}\right)} \geq 1 \\ &\Longleftrightarrow \left(\frac{\sigma^2}{\bar{p}h} + \alpha\right)\left(1 - F\left(\tilde{h}\right) - \frac{h\left(1 - F\left(\tilde{h}\right)\right)}{HM\left(\tilde{h}\right)}\right) \geq 1 && (166) \end{aligned}$$

where (165) follows from (164) and (44). Now the claim in (163) follows directly from (166). The following statements now follow from the key observation (163).

$$\frac{h}{\mathcal{K}\left(\tilde{h}\right)} \leq 1 \implies \frac{h}{\mathcal{K}\left(\hat{h}\right)} \leq 1, \;\; \forall h \geq \hat{h} \geq \tilde{h} \in \sup(F) \tag{167}$$

$$\frac{h}{\mathcal{K}\left(\tilde{h}\right)} \geq 1 \implies \frac{h}{\mathcal{K}\left(\hat{h}\right)} \geq 1, \ \forall \hat{h} \geq \tilde{h} \geq h \in \sup(F) \tag{168}$$

$$\mathcal{K}(h) \longrightarrow 0, \quad \text{as} \quad h \longrightarrow \bar{h} \tag{169}$$

If $HM(0) > 0$ we have $\mathcal{K}(0) > 0$, it follows from (169) and by the continuity of F that $\mathcal{K}$ has at least one fixed point. We show that $\mathcal{K}$ has a fixed point h_{thr} by explicit construction of a sequence of points that converges to h_{thr} and in the process uniqueness will follow. Consider the following iteratively defined sequence $\{h(n)\}_{n\in\mathbb{N}}$. Let $h(0) = 0$. and $h(n) = \mathcal{K}(h(n-1))$, $n \geq 1$. We have $h(1) = \mathcal{K}(h(0)) > h(0) = 0$. We show by induction that $h(n) \geq h(n-1)$. Suppose $h(k) \geq h(k-1)$, $\forall k \leq n$. Now, substituting $h = \hat{h} = h(n)$ and $\tilde{h} = h(n-1)$ it follows from (167) that

$$\frac{h(n)}{\mathcal{K}(h(n))} = \frac{h(n)}{h(n+1)} \leq 1$$

This shows that $\{h(n)\}_n$ is an increasing bounded sequence (bounded using (169) and recalling that K is continuous) and hence $h(n) \uparrow h_{thr}$ for some h_{thr} in the support of F and h_{thr} is a fixed point of K. Furthermore, for $h \in \left(h_{thr}, \bar{h}\right)$, it follows from (168) that

$$1 \leq \frac{h_{thr}}{\mathcal{K}(h)} < \frac{h}{\mathcal{K}(h)}$$

and hence $\mathcal{K}(h) < h$ for all $h \in \left(h_{thr}, \bar{h}\right)$. Now suppose $HM(0) = 0$ and thus $\mathcal{K}(0) = 0$. We need to show that for small enough h, we have $\mathcal{K}(h) \geq h$ and thus the fixed point iteration can start from such small enough nonzero h. Substituting $h = \tilde{h}$ in (163) we arrive at $\mathcal{K}\left(\tilde{h}\right) \geq \tilde{h}$ for some $\tilde{h} > 0$ if we show that

$$\int_{\tilde{h}}^{\bar{h}} \left(\frac{1}{\tilde{h}} - \frac{1}{h_0}\right) dF(h_0) \to \infty, \text{ as } h \to 0\,. \tag{170}$$

Observe that the integrand in (170) is the waterfilling power allocation policy in (15) and maximizes the single user capacity in (14). Suppose

$$\sup_{\tilde{h}>0} \int_{\tilde{h}}^{\bar{h}} \left(\frac{1}{\tilde{h}} - \frac{1}{h_0}\right) dF(h_0) < K_d$$

for some constant K_d. Then, we have

$$\begin{aligned} \bar{C}_{1user}\left(\tilde{h}\right) &= \frac{1}{2}\int_{\tilde{h}} \bar{h} \log\left(1 + h_0\sigma^{-2}\left(\frac{1}{\tilde{h}} - \frac{1}{h_0}\right)\right) dF(h_0) \\ &\leq \log 2 + \frac{1}{2}\log\left(\bar{h}K_d\right), \ \forall \tilde{h} > 0 \end{aligned} \tag{171}$$

where we used a technique similar to that used in the proof of Proposition 4.1 to derive the last step. Since $\bar{C}_{1user}$ can be made arbitrarily large by choosing the average power constraint of the power policy $\bar{p}$ arbitrarily large and for every choice of $\bar{p}$ the corresponding single user capacity $\bar{C}_{1user}$ is achieved by the waterfilling policy of the form $\left(\frac{1}{\tilde{h}} - \frac{1}{h}\right)^+$, we have a contradiction to (171). Thus there cannot be a uniform bound K_d and we have shown (170). This shows that h_{thr} is the unique fixed point of $\mathcal{K}$ and a "fixed point iteration" from small enough h converges to h_{thr}. □

I Proof of Proposition 6.1

The proof is quite elementary. We first show (94). Recall the map $\mathcal{K}$ in (59) of which h_{thr} is the unique positive fixed point (Lemma 4.10). Our first observation is that the map $\mathcal{K}$ as a function of α (denoted by $\mathcal{K}_\alpha$) is strictly increasing pointwise with increasing α. Furthermore, for each α the map $\mathcal{K}_\alpha$ is continuous.

Consider the following claim:

$$\mathcal{K}_\alpha(h) \downarrow \mathcal{K}_0(h) \, \forall h \in \left[0, \bar{h}\right] \text{ as } \alpha \downarrow 0 \text{ uniformly in } h \, . \tag{172}$$

where

$$\mathcal{K}_0 : \, h \mapsto \frac{HM(h)}{1 + \frac{\bar{p}\sigma^{-2}HM(h)}{1-F(h)}} \, .$$

To see this claim, let us define

$$\tilde{p} \stackrel{\text{def}}{=} \frac{\bar{p}HM(h)}{\sigma^2(1 - F(h))} \text{ and } \tilde{\alpha} \stackrel{\text{def}}{=} \alpha(1 - F(h))$$

Observe that $\forall h \in \left[0, \bar{h}\right]$,

$$\begin{aligned} \mathcal{K}_\alpha(h) - \mathcal{K}_0(h) &\leq HM(h)\left(\tilde{p} - \sigma^2 \tilde{p}\beta^*\left(\sigma^2\tilde{p}, \tilde{\alpha}\right)\right) \\ &= \frac{HM(h)}{2}\left(\tilde{p}(1+\tilde{\alpha}) + 1 - \sqrt{\tilde{p}^2(1-\tilde{\alpha})^2 + 2\tilde{p}\tilde{\alpha} + 1}\right) \qquad (173) \\ &\leq HM(h)\,\tilde{p}\tilde{\alpha} \\ &\leq \frac{\bar{h}^2\bar{p}}{\sigma^2}\alpha \qquad (174) \end{aligned}$$

where (173) is by definition of $\beta^*(\bar{p}, \alpha)$ in (46). The final upper bound in (174) shows the claim in (172) that K_α converges pointwise uniformly. It follows from (15) and the constraint on the average power to be equal to $\bar{p}$ that h_{wf} is the unique positive solution of the following fixed point equation:

$$h_{wf} = \frac{HM(h_{wf})}{1 + \frac{\bar{p}\sigma^{-2}HM(h_{wf})}{1-F(h_{wf})}} \, . \tag{175}$$

From (175), we see that h_{wf} is the unique positive fixed point of the map $\mathcal{K}_0$. We now claim that the fixed points of the maps $\mathcal{K}_\alpha$ themselves decrease monotonically with decreasing α. Let $h_{thr}^{(\alpha)}$ denote the unique fixed point of the map $\mathcal{K}_\alpha$. Fix $\alpha_2 > \alpha_1$. Define sequences $\{h_i(n)\}_{n\geq 0}$ for $i = 1, 2$ as follows: $h_i(0) = 0$ and $h_i(n) \stackrel{\text{def}}{=} \mathcal{K}_{\alpha_i}(h_i(n-1))$. Then, from Lemma 4.10 it follows that $h_i(n) \uparrow h_{thr}^{(\alpha_i)}$ as $n \uparrow \infty$ for $i = 1, 2$. Thus we have $h_1(n) < h_2(n)$ for every $n > 0$ and we conclude that $h_{thr}^{(\alpha_1)} \leq h_{thr}^{(\alpha_2)}$. Thus $\left\{h_{thr}^{(\alpha)}\right\}_\alpha$ is a decreasing sequence as α is decreasing and converges to, say, h_0. Now $\forall \alpha$,

$$h_{thr}^{(\alpha)} = \mathcal{K}_\alpha\left(h_{thr}^{(\alpha)}\right) \geq \mathcal{K}_0\left(h_{thr}^{(\alpha)}\right) \, .$$

Taking limits as $\alpha \to 0$ and using the continuity of the map $\mathcal{K}_0$ we have

$$h_0 \geq \mathcal{K}_0(h_0) \; . \tag{176}$$

Also, from (174) we have, for every α,

$$\begin{aligned}\mathcal{K}_0\left(h_{thr}^{(\alpha)}\right) &\geq \mathcal{K}_\alpha\left(h_{thr}^{(\alpha)}\right) - \frac{\bar{h}^2\bar{p}\alpha}{\sigma^2} \\ &= h_{thr}^{(\alpha)} - \frac{\bar{h}^2\bar{p}\alpha}{\sigma^2}\end{aligned}$$

and taking limits as $\alpha \to 0$, continuity of $\mathcal{K}_0$ yields

$$\mathcal{K}_0(h_0) \geq h_0 \; . \tag{177}$$

Now (176) and (177) show that $h_0 = \mathcal{K}_0(h_0)$ and thus $h_0 = h_{wf}$, the unique fixed point of $\mathcal{K}_0$. Following the definition of $\beta^*(\bar{p}, \alpha)$ in (45), we have

$$\beta^*\left(\frac{\bar{p}HM\left(h_{thr}^{(\alpha)}\right)}{1-F\left(h_{thr}^{(\alpha)}\right)}, \alpha\left(1-F\left(h_{thr}^{(\alpha)}\right)\right)\right) \longrightarrow \frac{1}{\sigma^2} \quad \text{as } \alpha \to 0 \; . \tag{178}$$

Observing that for every α

$$\mathbb{E}\left[\frac{1}{\beta_{wf}^*}\left(\frac{1}{h_{thr}^{(\alpha)}} - \frac{1}{h}\right)^+\right] = \bar{p} \, ,$$

we have that $h_{thr}^{(\alpha)}$ decreases monotonically with α implies that β_{wf}^* increases monotonically with α. Since we had already observed that the limit of β_{wf}^* is σ^{-2} in (178), we have shown (94). An identical argument now shows (95). □

J Proof of Proposition 6.3

The proof is not too different from that of Proposition 6.1. We observe by definition of $\beta^*(p, \alpha)$ in (45) that, as $\sigma^2 \to 0$,

$$\beta^*(p, \alpha) \to \frac{1}{p(\alpha - 1)^+} \, .$$

This implies that (as in (172)), as $\sigma^2 \to 0$,

$$\mathcal{K}(h) \to \mathcal{K}_o(h) \stackrel{\text{def}}{=} \frac{HM(h)(\alpha(1-F(h))-1)^+}{\alpha(1-F(h))} \, .$$

As in the proof of Proposition 6.1, h_{thr} converges to the fixed point of $\mathcal{K}_o$ as $\sigma^2 \to 0$, denoted by h_o. When $\alpha \leq 1$, we easily identify $h_o = 0$ and when $\alpha > 1$ that $h_o > 0$. The monotonicity arguments follow easily. From the limiting values of h_{thr} and β^*_{wf} we have for each user i that $\mathcal{P}^{wf}_i(h_i) \xrightarrow{a.s.} \bar{p}$ as $\sigma^2 \to 0$ and when $\alpha > 1$ the limiting value of $\mathcal{P}^{wf}$ is different from that of constant power allocation policy and thus there is a strict loss in sum capacity by using the constant power allocation policy as compared to the waterfilling strategy. It still remains to show that when $\alpha \leq 1$ the gain in sum capacity with waterfilling strategy over constant power strategy goes to zero in high SNR. We follow the proof of Proposition 6.2. Fix $\alpha \leq 1$ and processing gain at N and the number of users at $K = \lfloor \alpha N \rfloor$. From the limiting values of h_{thr} and β^*_{wf} we already have that $\mathcal{P}^{wf} \xrightarrow{a.s.} \bar{p}$ as $\sigma^2 \to 0$. We establish a bound akin to (99) and appeal to the dominated convergence theorem concluding the proof:

$$\begin{aligned}
\frac{1}{2N}\left(\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K} s_i s_i^t h_i \mathcal{P}^{wf}_i\right)\right) &\leq \frac{1}{2N}\left(\log\det\left(I + \sigma^{-2}\sum_{i=1}^{K}\frac{s_i s_i^t h_i}{\beta^*_{wf} h_{thr}}\right)\right) && (179)\\
&\leq \frac{1}{2N}\sum_{i=1}^{K}\log\left(1 + \frac{\bar{h} s_i^t s_i}{\sigma^2 \beta^*_{wf} h_{thr}}\right) && (180)\\
&\leq \frac{\bar{h}}{2N\sigma^2\beta^*_{wf} h_{thr}}\sum_{i=1}^{K} s_i^t s_i
\end{aligned}$$

where we used the bound that $\mathcal{P}^{wf}_i(h_i) \leq \frac{1}{\beta^*_{wf} h_{thr}}$ by definition of the waterfilling strategy (73) in (179) and the Hadamard inequality in (180). Analogous to the proof of Proposition 3.1, an application of the dominated convergence theorem completes the proof. □

K Proof of Theorem 8.1

We follow the general approach adopted in the proof of Theorem 5.1. Our first observation is the following replacement of Proposition 3.1:

$$\frac{\bar{C}^A_{sum}(\bar{\mathcal{P}})}{\log N} \xrightarrow{a.s.} C^{\bar{p}}. \tag{181}$$

To see this we use the notation developed in the proof of Proposition 3.1 and write:

$$\begin{aligned}
\frac{\bar{C}^A_{sum}(\bar{\mathcal{P}})}{\log N} &= \frac{1}{2}\int \frac{\log(1 + N\bar{p}\lambda)}{\log N} dG_N(\lambda)\\
&= \frac{1}{2}\log\bar{p} + \frac{1}{2}\int\left(\log\lambda + \frac{\log\left(1 + \frac{1}{N\bar{p}\lambda}\right)}{\log N}\right) dG_N(\lambda)\\
&\xrightarrow{a.s., L^1} \frac{1}{2}\log\bar{p} + \frac{1}{2}\int_{\lambda>0}\log\lambda dG^*(\lambda)
\end{aligned}$$

where in the last step, we used the following arguments: any mass G^* might have at $\lambda = 0$ does no contribute to the integral, Theorem 1.1 and its corollary of [1], and an application of the dominated convergence theorem as in the proof of Proposition 3.1. In this case, G^* has very well known (in the random matrix literature) quarter circle density (Proposition [26]) and thus $C^{\bar{p}}$ can be explicitly expressed as

$$C^{\bar{p}} = \frac{1}{2}\log \bar{p} + \frac{1}{2}\int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \log \lambda \frac{\sqrt{\left(\lambda - (1-\sqrt{\alpha})^2\right)\left((1+\sqrt{\alpha})^2 - \lambda\right)}}{2\pi\alpha\lambda} d\lambda \,.$$

The replacement of Proposition 4.1 is the following:

$$\frac{\bar{C}^{A}_{opt}(N)}{\log N} < \alpha K_c \ , \forall N > 0$$

where K_c is some constant independent of N. The proof of this claim follows very closely that of Proposition 4.1 and we omit it here. Any optimal $\mathcal{P}^*(S)$ has the following structure:

$$\mathcal{P}^*_i(S) = \left(\frac{1}{\lambda} - \frac{1}{s_i^t\left(\frac{\sigma^2}{N} + \sum_{j\neq i} s_j s_j^t \mathcal{P}^*_j(S)\right)^{-1} s_i}\right)^+$$

and Theorem 4.7 is still valid:

$$\mathcal{P}^*_i \leq K_p \ , \forall i = 1 \ldots K, \ \forall N$$

for a possibly different constant K_p than the one used in Theorem 4.7. In this scenario, the SIR of user i with the constant power allocation policy is given by $\beta_i\left(\bar{\mathcal{P}}\right)\bar{p}$ where

$$\beta_i\left(\bar{\mathcal{P}}\right) \stackrel{\text{def}}{=} s_i^t\left(\frac{\sigma^2}{N} + \bar{p}\sum_{j\neq i} s_j s_j^t\right)^{-1} s_i \,.$$

Define the (Lagrangian) function:

$$L : \mathcal{P} \mapsto \bar{C}^{A}_{sum}(\mathcal{P}) - \frac{1}{2N\bar{p}}\sum_{i=1}^{K}\left(\mathbb{E}\left[\mathcal{P}_i(S)\right] - \bar{p}\right) \quad \mathcal{P} \in \mathcal{F}_0^{(N)} \,.$$

As in the proof of Theorem 5.1,

$$\begin{aligned}
L(\mathcal{P}^*) - L\left(\bar{\mathcal{P}}\right) &\leq \frac{\alpha}{2}\mathbb{E}\left[\left(\frac{\beta_1\left(\bar{\mathcal{P}}\right)}{1+\beta_1\left(\bar{\mathcal{P}}\right)\bar{p}} - \frac{1}{\bar{p}}\right)\left(\mathcal{P}^*_i(S) - \bar{p}\right)\right] \\
&\leq \frac{\alpha K_p}{2\bar{p}}\mathbb{E}\left[\frac{1}{1+\beta_1\left(\bar{\mathcal{P}}\right)\bar{p}}\right] \\
&\leq \frac{\alpha K_p}{2\bar{p}^2}\mathbb{E}\left[\frac{1}{\beta_1\left(\bar{\mathcal{P}}\right)}\right] \,. \qquad (182)
\end{aligned}$$

Now

$$\begin{aligned}
\beta_1\left(\bar{\mathcal{P}}\right) &= s_1^t\left(\frac{\sigma^2}{N}+\bar{p}\sum_{j\neq 1}s_js_j^t\right)^{-1}s_1 \\
&\geq s_1^t\left(\frac{\sigma^2}{N_1}+\bar{p}\sum_{j\neq 1}s_js_j^t\right)^{-1}s_1\ ,\ \forall N\geq N_1 \\
\liminf_{N\to\infty}\beta_1\left(\bar{\mathcal{P}}\right) &\geq \beta^*\left(\bar{p},\alpha,\frac{\sigma^2}{N_1}\right),\ \forall N_1\ .
\end{aligned}$$

where $\beta^*\left(\bar{p},\alpha,\sigma^2\right)$ has an explicit expression in (46). Continuing we have

$$\begin{aligned}
\liminf_{N\to\infty}\beta_1\left(\bar{\mathcal{P}}\right) &\geq \sup_{N_1}\beta^*\left(\bar{p},\alpha,\frac{\sigma^2}{N_1}\right) \\
&= \frac{1}{\bar{p}\left(\alpha-1\right)^+}\ .
\end{aligned}$$

Substituting this in (182) shows that

$$\limsup_{N\to\infty}L\left(\mathcal{P}^*\right)-L\left(\bar{\mathcal{P}}\right)\leq\frac{\alpha\left(\alpha-1\right)^+K_p}{2\bar{p}}\ .$$

The observation that $L\left(\mathcal{P}^*\right)=\bar{C}^A_{opt}\left(N\right)$ and $L\left(\bar{\mathcal{P}}\right)=\bar{C}^A_{sum}\left(\bar{\mathcal{P}}\right)$ completes the proof. □

References

[1] Z. D. Bai and J W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices", *Annals of Probability*, 26(1), 1998, pp. 316-345.

[2] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels", *IEEE Transactions on Information Theory*, vol. 45(5), July, 1999, pp. 1468-1489.

[3] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley, 1991.

[4] G. J. Foschini and Z. Miljanic, "Distributed autonomous wireless channel assignment algorithm with power control", *IEEE Trans. on Vehicul. Technol.*, Vol. 44, No.—3, pp.420-429, August 1995.

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

[6] A. Goldsmith and P. Varaiya, "Capacity of fading channel with channel side information", *IEEE Transactions on Information Theory*, Vol. 43, pp. 1986-1992, Nov. 1995.

[7] S. Grandhi, R. Yates and D. Goodman, "Resource allocation for cellular radio systems", *IEEE Trans. on Vehic. Tech.*, 46(3):581-588, August 1997.

[8] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous CDMA channels", *IEEE Trans. on Information Theory*, vol. 44, No. 7, Nov. 1998, pp. 2832-2836.

[9] S. V. Hanly, "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, 1995.

[10] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications", *International Conference on Communications*, Seattle, June 1995.

[11] Marshall, A W. and I. Olkin, *Inequalities: Theory of Majorization and its applications*, Academic Press, 1979.

[12] L. H. Ozarow, S. Shamai and A.D. Wyner, "Information theoretic considerations of cellular mobile radio", *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 359-378, May 1994.

[13] R. Rapajic, M. Honig and G. Woodward, "Multiuser decision-feedback detection: performance bounds and adaptive algorithms", *Proc. 1998 International Symp. on Info. Theory*, August, 1998, p. 34.

[14] T. R. Rockafellar, *Convex Analysis*, Princeton University Press, 1984.

[15] H. L. Royden, *Real Analysis*, Macmillan, 3rd Ed., 1988.

[16] M. Rupf and J. L. Massey, "Optimum sequence multisets for Synchronous code-division multiple-access channels", *IEEE Transactions on Information Theory*, Vol 40, No. 4, pp. 1261-1266, July 1994.

[17] S. Shamai and A. D. Wyner, "Information theoretic considerations for symmetric, cellular, multiple-access fading channels - Part I", *IEEE Transactions on Information Theory*, Vol. 43, No. 6, 1997, pp. 1877-1894.

[18] S. Shamai and S. Verdú, "Capacity of CDMA fading channels", *Information Theory Workshop, Metsovo, Greece*, June 27-July 1, 1999.

[19] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices", *J. Multivariage Anal.,* vol. 54, no. 2, pp. 175-192, 1995.

[20] E. Telatar, "Capacity of multi-antenna Gaussian channels", Bell Labs Technical Memo. 1997.

[21] D. Tse and S. Hanly, "Multi-Access Fading Channels: Part I: Polymatroidal Structure, Optimal Resource Allocation and Throughput Capacities", *IEEE Transactions on Information Theory*, Vol. 44, No. 7, Nov 1998, pp 2796-2815.

[22] D. Tse and S. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity", *IEEE Transactions on Information Theory*, Vol. 45, No. 2, March 1999, pp 641-657.

[23] D. Tse and O. Zeitouni, "Performance of Linear Multiuser Receivers in Random Environments", to appear in *IEEE Transactions on Information Theory.*

[24] L. Vandenberghe, S. Boyd, and S-P. Wu, "Determinant maximization with linear matrix inequality constraints", *SIAM Journal on Matrix Analysis and Applications*, vol. 19, No. 2, April 1998, pp 499-533.

[25] M. K. Varanasi and T.Guess, "Optimum decision feedback multiuser equalization and successive decoding achieves the total capacity of the Gaussian multiple-access channel", Proc.Asilomar conf. on Signals, Systems and Computers, 1997.

[26] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.

[27] S. Verdú and S. Shamai, "Spectral Efficiency with random spreading", *IEEE Transactions on Information Theory*, vol. 45, no. 2, March 1999, pp. 622-640.

[28] P. Viswanath, V. Anantharam and D. Tse, "Optimal Sequences, Power Control and Capacity of Synchronous CDMA Systems with Linear MMSE Multiuser Receivers", *IEEE Transactions on Information Theory*, vol. 45(6), Sept. 1999, pp. 1968-1983.

[29] P. Viswanath and V Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems", *IEEE Transactions on Information Theory*, vol. 45(6), Sept. 1999, pp. 1984-1991.

[30] D. H. Wagner, "Survey of measurable selection theorems: an update", *Proc. Conf. Oberwolfach, 1979*, pp. 176-219, Lecture Notes in Mathematics,794, Springer, 1980.

[31] S.-P. Wu, L. Vandenberghe and S. Boyd, "MAXDET: Software for determinant maximization problems: Users guide, Alpha version", Stanford University, Apr. 1996.