

Asymptotically Optimal Water-Filling in Vector Multiple-Access Channels

Pramod Viswanath, *Member, IEEE*, David N. C. Tse, *Member, IEEE*, and Venkat Anantharam, *Fellow, IEEE*

Abstract—Dynamic resource allocation is an important means to increase the *sum capacity* of fading multiple-access channels (MACs). 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 degree of freedom 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) channels and MACs 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. 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 “water-filling.” In the context of DS-CDMA and in the special case of no fading, the asymptotically optimal power policy of water-filling simplifies to constant power allocation over all realizations of signature sequences; this result verifies the conjecture made in [28]. We study the behavior of the asymptotically optimal water-filling 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—Code-division multiple access (CDMA), linear minimum mean-square error (MMSE) receivers, multiple antenna systems, power control, spectral efficiency, sum capacity, water-filling.

I. INTRODUCTION

THE focus of this paper is vector multiple-access channels (MACs); these are multiple-access channels with multiple *degrees of freedom*. Two common examples of such channels are direct-sequence code-division multiple-access (DS-CDMA) and a MAC with multiple antennas at the receiver. The number of degrees of freedom in the 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 received *spreading sequence* 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 multiple-access fading channel 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 MAC 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 time-division multiple access (TDMA), frequency-division multiple access (FDMA), and code-division multiple access (CDMA), with the performance criterion typically being the signal-to-interference (SIR) ratio of the users at the receiver, is studied in [5], [8], [10], [29], [23]. In the context of information-theoretic power control, existing literature focuses mainly on scalar channels. For the single-user scenario, [7] identifies water-filling 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, [11] 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 water-filling strategy with respect to their fading states. This paper focuses on vector multiple-access 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 frequency 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

Manuscript received November 10, 1999; revised April 19, 2000. The work of P. Viswanath and V. Anantharam was supported by the National Science Foundation under Grant IRI 97-12131. The work of D. N. C. Tse was supported by an NSF CAREER Award under Grant NCR-9734090.

P. Viswanath is with Flarion Technologies, Bedminster, NJ 07921 USA (e-mail: p.viswanath@flarion.com).

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

Communicate by S. Shamai, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(01)00586-7.

large processing gain). This policy is *water-filling* for each user and depends solely on the amplitude of that user's instantaneous fading amplitude. We show that the water-filling policy is asymptotically optimal for both the long and the short signature sequence models.

- 2) In the multiantenna model, we assume independent frequency flat fading from the users to the antenna array at the receiver. The fading from any user to one of the antennas can be considered to have two components: first, a slowly varying component due to distance loss, shadowing effects, and geographical features and second, a fast varying component due to constructive and destructive addition of the various multiple paths at the antenna. Our main result is that a water-filling power policy that, for any user, depends only the slow fading component of the fading channel from that user to the antenna array is asymptotically optimal (the asymptotic is in the number of users and the correspondingly large number of antennas at the receiver).

In the scenario when the slow fading component can be considered constant in the time scale of communication, the water-filling policy simplifies to a constant power allocation policy (this policy transmits a constant power regardless of the realization of fast fading amplitudes of the users). Thus, asymptotically there is no gain in having channel side information at the users. In the multiple-antenna scenario, this result is in sharp contrast from the point of view of the scalar MAC result in [11]. The scalar multiple-access result shows that the gain in spectral efficiency by allowing side information of the channel states to all the users [the optimal way to use this side information is to allocate positive power only to the user with the best channel (if at all)] is significantly higher than the spectral efficiency obtained by not allowing any side information (and thus allocating constant powers to the users at all fading states). From our results, we conclude that if there are sufficiently many degrees of freedom (antennas in this case), the gain in spectral efficiency by having side information at the transmitters vanishes.

In other related work on vector MACs, [16] and [30] study the allocation of signature sequences to achieve sum capacity in nonfading channels as a function of the average power constraints of the users. In [9], the authors study the sum capacity of CDMA systems with random long signature sequences in nonfading channels. In [28], 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 minimum mean-square error (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 [28]. The effect of frequency flat fading on the spectral efficiency of DS-CDMA with random long signature sequences and a wide variety of receiver structures is studied in [19]. Recent results on information—theoretic power control in nonergodic scalar fading channels are in [3]. A comprehen-

sive tutorial of information-theoretic study of fading channels is given in [2].

In Section II, we outline the DS-CDMA fading-channel model, formulate the problem, and precisely state our main results. In Section III, we heuristically derive the structure of the optimal power allocation strategy and see that it is water-filling. This section outlines the key ideas in the identification of asymptotic optimality of the water-filling strategy and allows the more casual reader to gain insight into our result without entering the technicalities required for the formal proof. In Section IV, we develop the mathematical machinery and some preliminary results required for the proof of our main result. In Section V, 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 VI, we study various regimes of number of users and signal-to-noise ratio (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 VII, 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 VIII, we turn to the multiple-antenna model, the other example of a vector MAC. We conclude the paper in Section IX with some summarizing remarks and suggestions for future work.

II. MODEL, PROBLEM FORMULATION, AND MAIN RESULTS

A. Model

We consider a single cell in a symbol synchronous CDMA channel and focus on the uplink. 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 MAC. 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 $K/N \rightarrow \alpha$ as $N \rightarrow \infty$ along with some mild restrictions allows us to derive all the asymptotic results obtained (asymptotic in N) in this paper. Following standard notation (see [27, Sec. 2.1]), the baseband received signal in one symbol interval can be expressed as

$$\mathbf{y}(n) = \sum_{i=1}^K x_i(n) \mathbf{s}_i(n) \tilde{h}_i(n) + \mathbf{w}(n). \quad (1)$$

A word about our notation throughout this paper. We use lower case letters for scalars, bold face lower case letters for vectors (with N components), and upper case for matrices. The superscript tilde on the channel fading states \tilde{h} and \mathbf{h} indicates the quantity is from the complex field. In (1), the index n represents time and the received signal \mathbf{y} is regarded as a vector in \mathbb{C}^N . Here $\mathbf{s}_i(n)$ is the received 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 $\mathbf{s}_i(n)$ is an independent realization for every time n and in the short signature sequence it is fixed for all

time. We model the signature sequences as having random independent and identically distributed (i.i.d.) entries (the choice and relevance of this model are discussed in [23] and [28]). Here, \tilde{h}_i is the complex fading or path gain from user i to the single base station (receiver). The baseband model of (1) represents frequency flat fading which is the appropriate model when the sampling rate is not large enough for the multiple paths to be resolved. We write the amplitude squared of this complex path gain by $h_i \stackrel{\text{def}}{=} \tilde{h}_i \tilde{h}_i^*$. Henceforth, we refer to h_i as the path gain and explicitly say “complex path gain” when referring to \tilde{h}_i . The user symbols are represented by the real-valued random variables x_i . \mathbf{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 \tilde{h}_i are conventional (see [22, Sec. 2] and [7], for example). We assume that $\{\{h_i(n)\}_{n \in \mathbb{N}}\}_{i=1}^K$ is a sequence of i.i.d. stationary and ergodic processes; let us denote the (common) stationary distribution of the amplitude squared of the complex fading process by F which has finite first and second moments.

B. 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 $\mathbf{s}_i = \frac{1}{\sqrt{N}}(v_{i1}, \dots, v_{iN})^t$ where $\{v_{i\ell}\}$ 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 $\{\tilde{h}_i(n)\}_{n \in \mathbb{N}}$. Both the random variables $\{v_{ij}\}$ and $\{\tilde{h}_i\}$ are defined 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 $L^1(\Omega, \mu)$.

Conditioned on one sample point or realization of signature sequences, say $\mathbf{s}_1, \dots, \mathbf{s}_K$ (we write $S = [\mathbf{s}_1, \dots, \mathbf{s}_K]$) the channel model in (1) becomes

$$\mathbf{y}(n) = \sum_{i=1}^K x_i(n) \tilde{h}_i(n) \mathbf{s}_i + \mathbf{w}(n). \quad (2)$$

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 $\mathbf{s}_1, \dots, \mathbf{s}_K$, the sum capacity of the MAC in (2) is

$$\frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \bar{p} \sigma^{-2} \sum_{i=1}^K h_i \mathbf{s}_i \mathbf{s}_i^t \right) \middle| S \right]. \quad (3)$$

The capacity region for single degree of freedom fading channels with no information of the fading state at the transmitter is given in [17] and the intuitive idea behind the proof is given in [6] 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 nonnegative reals. We let

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

denote a power allocation policy for user i and call the tuple $\mathcal{P} = (\mathcal{P}_1, \dots, \mathcal{P}_K)$ 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

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

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

$$\begin{aligned} C_{\text{sum}}(\mathcal{P}, S) & \stackrel{\text{def}}{=} \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sigma^{-2} \sum_{i=1}^K h_i \mathbf{s}_i \mathbf{s}_i^t \mathcal{P}_i(h_1, \dots, h_K, S) \right) \middle| S \right]. \end{aligned} \quad (4)$$

Comparing with (3), $C_{\text{sum}}(\mathcal{P}, S)$ 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

$$\begin{aligned} C_{\text{opt}}(S) &= \sup_{\mathcal{P} \in \mathcal{F}_1(S)} C_{\text{sum}}(\mathcal{P}, S) \\ &= \sup_{\mathcal{P} \in \mathcal{F}_1(S)} \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sum_{i=1}^K \sigma^{-2} h_i \mathbf{s}_i \mathbf{s}_i^t \mathcal{P}_i(h_1, \dots, h_K, S) \right) \middle| S \right]. \end{aligned} \quad (5)$$

A version of the coding theorem in the above proposition appeared as [22, Theorem 2.1], another version of the above result for a single-user fading channels is in [7] 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 VII.

In the notation of [21], the MAC with short signature sequences in (2) represents a *nonergodic* channel and the Shannon capacity of the channel is zero; however small the sum rate the

users attempt to communicate at, there is a nonzero probability that the realized signature sequences will render the channel incapable of supporting the rates reliably. Motivated by the approach in [21] 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 realizations of signature sequences but a set \mathcal{S} whose total probability is less than a . In our notation, the supportable rate $R_a^{*(N)}$ is defined as

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

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

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

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

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

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 reduces 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

$$\mathbf{s}_i(n) = \frac{1}{\sqrt{N}}(v_{i1}(n), \dots, v_{iN}(n))^t$$

where $\{v_{i\ell}(n)\}$ are i.i.d. random variables with zero mean, 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 nonnegative 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

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

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

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

This result was observed in [28, Sec. 3]. For $\mathcal{P} \in \mathcal{F}_2^{(N)}$ defining the quantity

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

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

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

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 $\bar{\mathcal{C}}_{\text{opt}}$.

A word about our notation is in order here. We use the superscript N only when the emphasis on the dependence on the system size is warranted. For example, the superscript N does not appear on $\bar{\mathcal{C}}_{\text{sum}}$ which is a *function* that takes arguments $\mathcal{P} \in \mathcal{F}_2^{(N)}$ and thus the system size N is implicitly contained in the argument \mathcal{P} . On the other hand, we do use the superscript N on $\bar{\mathcal{C}}_{\text{opt}}$ which is a positive number.

C. 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 which is almost optimal for both the short and long signature sequence models. We state this result formally below. Consider the power allocation policies

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

where we have used the notation $(x)^+$ to indicate $\max\{x, 0\}$. The constant β_{wf}^* is the limiting the 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 III and IV. In (12), the constant (Kuhn–Tucker coefficient) λ is chosen such that $\mathbb{E}[\mathcal{P}_i^{\text{wf}}(h_i)] = \bar{p}$. 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 *water-filling* and generalizes the strategy of [7] for single-user fading channels. To see this generalization, recall the optimal power allocation policy for the single-user case from [7]

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

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 water-filling 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 with bounded first and second moments.

- 1) Consider the case of long signature sequences. With long signature sequences we show that asymptotically the water-filling 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

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

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

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

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

Thus, in a large system the supportable rate using the water-filling 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 water-filling 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 $\sigma^2 \rightarrow 0$) 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 water-filling 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 $\alpha \leq 1$. 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) by using an optimal strategy over constant power allocation 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 water-filling 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 water-filling policy (12) but users in different classes have different threshold levels for their water-filling policies.

III. 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 water-filling structure (12) of the optimal power allocation policy. Toward this end, we proceed in the following order: we first review the water-filling power allocation policy (identified in [7]) for a single user in a (scalar) fading channel. Then, we show the relation of sum capacity to linear MMSE (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))

$$y(n) = \tilde{h}(n)x(n) + w(n)$$

where $\{\tilde{h}(n)\}_{n \in \mathbb{N}}$ 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 stationary distribution F with bounded first and second moments. $w(n)$ 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 \tilde{h}_n are perfectly known to both the transmitter and the receiver. The transmitter has an average power constraint \bar{p} . Then, ([7, Theorem 2.1]) the capacity of the channel is

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

and the power allocation that achieves the maximum above is “water-filling” (refers to the visualization of this scheme)

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

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 $h_{\text{wft}} \stackrel{\text{def}}{=} \lambda \sigma^2$.

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 $\mathbf{s}_1, \dots, \mathbf{s}_K$ 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

$$\mathbf{y} = \sum_{i=1}^K x_i \mathbf{s}_i \tilde{h}_i + \mathbf{w}. \quad (16)$$

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

$$C(p_1, \dots, p_K) = \frac{1}{2N} \log \det \left(I + \sum_{i=1}^K \sigma^{-2} \mathbf{s}_i \mathbf{s}_i^t p_i h_i \right). \quad (17)$$

The rate tuples in the capacity region are in general achieved by jointly demodulating the users from the received signal \mathbf{y} . We focus on the following specific structure of demodulation of the users’ symbols from the received signal \mathbf{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 is used to decode that user) and the received signal is stripped off the decoded users. The LMMSE receiver for user i provides the *optimal* linear estimate of the user i symbol x_i from the received vector \mathbf{y} . It was observed in [26] that this scheme allows the users to transmit reliably at a sum rate equal to the sum capacity of the system.¹ We use 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

$$C(p_1, \dots, p_{i-1}, p_i + \delta, p_{i+1}, \dots, p_K) - C(p_1, \dots, p_K) \\ = \frac{1}{2N} \log \left(1 + \delta h_i \mathbf{s}_i^t \left[\sigma^2 I + \sum_{j=1}^K \mathbf{s}_j \mathbf{s}_j^t p_j h_j \right]^{-1} \mathbf{s}_i \right) \quad (18)$$

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 \mathbf{s}_i that is decoded first and then stripped off. It can be shown that (see

¹In fact, a stronger statement is claimed in [26]. 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.

[23, Sec. 2] or [27, Ch. 6]) the SIR of the LMMSE estimate of this fictitious $K+1$ user is $\beta_{K+1} \delta h_i$ where

$$\beta_{K+1} = \mathbf{s}_i^t \left(\sigma^2 I + \sum_{j=1}^K \mathbf{s}_j \mathbf{s}_j^t p_j h_j \right)^{-1} \mathbf{s}_i. \quad (19)$$

Now, the rate achieved by this fictitious user being decoded first is simply

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

where $\delta \beta_{K+1} h_i$ is the SIR of the LMMSE estimate of the fictitious user $K+1$. This expression is consistent with the expression for the increase in sum capacity in (18).

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

$$\bar{C}_{\text{opt}}^{(N)} \stackrel{\text{def}}{=} \sup_{\mathcal{P} \in \mathcal{F}_2} \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sum_{i=1}^K \sigma^{-2} h_i \mathbf{s}_i \mathbf{s}_i^t \mathcal{P}_i(h_1, \dots, h_K, S) \right) \right]. \quad (20)$$

In Proposition 4.2, we show that this is a concave maximization problem and in Proposition 4.5 that the maximum is actually achieved. Thus, there exists a Kuhn–Tucker coefficient $\lambda > 0$ such that a necessary and sufficient condition for the optimality of a power allocation policy \mathcal{P}^* is that for every realization of h_1, \dots, h_K, S the optimal policy \mathcal{P}^* (if positive) satisfies the following constraints:

$$\left(\frac{\partial}{\partial p_i} C(p_1, \dots, p_K) \right) (\mathcal{P}^*(h_1, \dots, h_K, S)) = \lambda \\ \forall i \text{ such that } \mathcal{P}_i^*(h_1, \dots, h_K, S) > 0 \quad (21)$$

$$\left(\frac{\partial}{\partial p_j} C(p_1, \dots, p_K) \right) (\mathcal{P}^*(h_1, \dots, h_K, S)) - \lambda \leq 0 \\ \forall j \text{ such that } \mathcal{P}_j^*(h_1, \dots, h_K, S) = 0. \quad (22)$$

Now (21) can be written as

$$\mathbf{s}_i^t \left(\sigma^2 I + \sum_{j=1}^K \mathbf{s}_j \mathbf{s}_j^t h_j \mathcal{P}_j^*(h_1, \dots, h_K, S) \right)^{-1} \mathbf{s}_i h_i = \lambda \\ \forall i \text{ such that } \mathcal{P}_i^*(h_1, \dots, h_K, S) > 0 \quad (23)$$

by using the expression (18) for the increase in the sum capacity by an increase in power of one user in the derivation of (23). 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 (23) yields

$$\mathcal{P}_i^*(h_1, \dots, h_K, S) \\ = \left(\frac{1}{\lambda} - \frac{1}{\mathbf{s}_i^t \left(\sigma^2 I + \sum_{j \neq i}^K \mathbf{s}_j \mathbf{s}_j^t h_j \mathcal{P}_j^*(h_1, \dots, h_K, S) \right)^{-1} \mathbf{s}_i h_i} \right)^+ \\ \forall i = 1, \dots, K. \quad (24)$$

The special case of $N = 1$ (in this case there is no spreading and $\mathbf{s}_i = 1$) is easily handled. Observe that the choice

$$\mathcal{P}_i^*(h_1, \dots, h_K) = \begin{cases} \left(\frac{1}{\lambda} - \frac{\sigma^2}{h_i}\right)^+, & \text{if } h_i = \max\{h_1, \dots, h_K\} \\ 0, & \text{else} \end{cases} \quad \forall i = 1, \dots, K$$

is a solution to (21) (equivalently, (24)) and (22). Thus, the optimal power allocation policy in the scalar multiple-user case is to allow only the best user to transmit and this user follows the water-filling policy; this was first observed in [11]. Even the case $N = 2$ is quite involved (simple solutions exist only under simplifying assumptions on the entries of the signature sequences). In general, there is no closed-form expression for the optimal power allocation in (24) and the authors in [25] design interior point methods that find approximate solutions to (24). However, in an effort to get more insight into the structure of the optimal power allocation policy we look into the regime of large K and large N . We begin with the following interpretation of the structure of the optimal power policy (24). Defining

$$\beta_i \stackrel{\text{def}}{=} \mathbf{s}_i^t \left(\sigma^2 I + \sum_{j \neq i} \mathbf{s}_j \mathbf{s}_j^t h_j \mathcal{P}_j^*(h_1, \dots, h_K, S) \right)^{-1} \mathbf{s}_i \quad (25)$$

and observing that $\beta_i h_i \mathcal{P}_i^*(h_1, \dots, h_K, S)$ is the (random) SIR of the LMMSE estimate of user i when powers are allocated according to \mathcal{P}^* and substituting in (24), we arrive at the following structure of an optimal power policy:

$$\mathcal{P}_i^*: (h_1, \dots, h_K, S) \mapsto \left(\frac{1}{\lambda} - \frac{1}{\beta_i h_i} \right)^+. \quad (26)$$

Here β_i is the (random) SIR of the LMMSE estimate of user i when all users are allocating powers optimally. 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 $(h_1, \dots, h_K, S) \mapsto g(h_i)$ for every user i where g is some bounded nonnegative function into the reals. In this situation, [23, Theorem 3.1] 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 (25) with $\mathcal{P}_i^*(h_1, \dots, h_K, S) = g(h_i)$ converges almost surely to $\beta_g^* h_1$ as $N \rightarrow \infty$. The positive constant β_g^* depends on α , the background noise variance σ^2 , and the function g itself and [23, Theorem 3.1] 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

$$\mathcal{P}_i^*: (h_1, \dots, h_K, S) \stackrel{\text{def}}{=} g_{\text{wf}}(h_i) \stackrel{\text{def}}{=} \left(\frac{1}{\lambda} - \frac{1}{\beta_{\text{wf}}^* h_i} \right)^+ \quad (27)$$

satisfies the Kuhn–Tucker conditions in (24) asymptotically. Here β_{wf}^* is a positive constant with the following structure. When every user uses a power allocation policy of this form, namely

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

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 [23, Theorem 3.1] 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 (27) 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 IV. 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 water-filling power allocation policy that we have only developed heuristically in this section.

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 water-filling over the fading process of that user *alone*. We use this structure to show that water-filling power allocation performs very “close” to the optimal policy even for the short signature sequences model. Toward 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 $\mathcal{P}_i^g: (h_1, \dots, h_K, S) \mapsto g(h_i)$ where g is a nonnegative 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:

$$\frac{1}{2N} \log \det \left(I + \sum_{i=1}^K \sigma^{-2} \mathbf{s}_i \mathbf{s}_i^t h_i g(h_i) \right) \xrightarrow{a.s., L^1} \bar{C}_{\text{sum}}^g, \quad \text{as } N \rightarrow \infty. \quad (28)$$

Here \bar{C}_{sum}^g is a positive nonrandom quantity and 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}_{sum}^g and [28, eq. 9] gives an explicit expression. In general, there is no known closed-form expression for \bar{C}_{sum}^g ; however, [18] and [19] give some expressions to compute \bar{C}_{sum}^g . 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

$$\mathbb{P}[\mathcal{C}_{\text{sum}}(\mathcal{P}^g, S) \geq R] \geq 1 - a. \quad (29)$$

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

$$\begin{aligned} R_a(\mathcal{P}^g) &\geq \bar{C}_{\text{sum}}^g - \frac{2\mathbb{E}[\|\mathcal{C}_{\text{sum}}(\mathcal{P}^g, S) - \bar{C}_{\text{sum}}^g\|]}{a} \\ &\geq \bar{C}_{\text{sum}}^g - o(1), \quad \text{using Proposition 3.1.} \end{aligned} \quad (30)$$

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 water-filling strategy, we use this result in Section V to show that the water-filling strategy is also close to the optimal power allocation with short signature sequences.

IV. MATHEMATICAL PRELIMINARIES

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 water-filling 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 and its solution.

A. 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 the proof of which is given in Appendix B.

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

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

Proposition 4.3: For every deterministic h_1, \dots, h_K and S , the map from the positive orthant in \mathbb{R}^K to the nonnegative reals

$$C: (p_1, \dots, p_K) \mapsto \frac{1}{2N} \left[\log \det \left(I + \sigma^{-2} \sum_{i=1}^K p_i h_i \mathbf{s}_i \mathbf{s}_i^t \right) \right] \quad (31)$$

is concave. Furthermore, if $\{h_i \mathbf{s}_i \mathbf{s}_i^t, i = 1, \dots, K\}$ are linearly independent, then C is strictly concave.

This result is quite well known. It can be derived from [4, Theorem 16.8.1]. A different proof is available in [32].

- 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:

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

Proof: 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 $\mathbb{E}[\mathcal{P}_1(h_1, \dots, h_K, S)] = \bar{p} - \delta$ for some positive δ . Consider the power allocation policy

$$\hat{\mathcal{P}}_1(h_1, \dots, h_K, S) = \mathcal{P}_1(h_1, \dots, h_K, S) + \delta$$

and

$$\hat{\mathcal{P}}_i = \mathcal{P}_i, \quad \text{for } i = 2, \dots, K.$$

By definition $\hat{\mathcal{P}} \in \mathcal{F}_2^{(N)}$. Then

$$\begin{aligned} & \bar{C}_{\text{sum}}(\hat{\mathcal{P}}) \\ &= \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sigma^{-2} \sum_{i=1}^K h_i \mathbf{s}_i \mathbf{s}_i^t \hat{\mathcal{P}}_i(h_1, \dots, h_K, S) \right) \right] \\ &= \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sigma^{-2} \sum_{i=1}^K \mathbf{s}_i \mathbf{s}_i^t h_i \mathcal{P}_i(h_1, \dots, h_K, S) \right. \right. \\ & \quad \left. \left. + \sigma^{-2} \delta h_1 \mathbf{s}_1 \mathbf{s}_1^t \right) \right] \\ &= \bar{C}_{\text{sum}}(\mathcal{P}) + \frac{1}{2N} \mathbb{E} \\ & \quad \cdot \left[\log \left(1 + \frac{\delta \beta_1(\mathcal{P})}{1 + \beta_1(\mathcal{P}) h_1 \mathcal{P}(h_1, \dots, h_K, S)} \right) \right] \\ &> \bar{C}_{\text{sum}}(\mathcal{P}). \end{aligned} \quad (32)$$

Here $\beta_1(\mathcal{P}) h_1 \mathcal{P}(h_1, \dots, h_K, S)$ 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 given in (42)] and (32) 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. \square

- 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

$$\begin{aligned} \bar{C}_{\text{opt}}^{(N)} &= \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} \\ & \cdot \left\{ \bar{C}_{\text{sum}}(\mathcal{P}) - \frac{\lambda}{2N} \sum_{i=1}^K (\mathbb{E}[\mathcal{P}_i(h_1, \dots, h_K, S)] - \bar{p}) \right\} \end{aligned} \quad (33)$$

where

$$\begin{aligned} \mathcal{F}_0^{(N)} &\stackrel{\text{def}}{=} \{\mathcal{P}: \mathcal{P}_i \geq 0 \\ & \text{and } \mathcal{P} \in L^1(h_1, \dots, h_K, S) \forall i = 1, \dots, K\}. \end{aligned} \quad (34)$$

It follows from Proposition 4.2 that the map $\bar{C}_{\text{sum}}: \mathcal{F}_2^{(N)} \rightarrow \mathbb{R}_+$ defined by (10) is concave. Since $\bar{C}_{\text{opt}}^{(N)}$ defined in (11) is the maximum of \bar{C}_{sum} , a concave function over a convex set $\mathcal{F}_2^{(N)}$, the claim above in Proposition 4.4 would have been completely standard had $\mathcal{F}_2^{(N)}$ been finite-dimensional ([14, Sec. 28] is the classical reference). However, $\mathcal{F}_2^{(N)}$ is infinite-dimensional and hence this claim needs a formal proof, which is supplied in Appendix C.

- 5) We now use the previous propositions to show that the supremum in the definition of $\bar{C}_{\text{opt}}^{(N)}$ 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 D.

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

$$p_i^* = \left(\frac{1}{\lambda} - \frac{1}{\mathbf{s}_i^t \left(\sigma^2 I + \sum_{j \neq i} \mathbf{s}_j \mathbf{s}_j^t h_j p_j^* \right)^{-1} \mathbf{s}_i h_i} \right)^+ \quad \forall i = 1, \dots, K \quad (35)$$

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 $\mathcal{P}_1^*, \dots, \mathcal{P}_K^*$ 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

$$\mathbb{E}[\mathcal{P}_i^*(h_1, \dots, h_K, S)] = \bar{p} \quad \forall i = 1 \dots K. \quad (36)$$

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 π , the power allocation policy defined by $\mathcal{P}_i^\pi(h_1, \dots, h_K, \mathbf{s}_1, \dots, \mathbf{s}_K)$

$$\stackrel{\text{def}}{=} \mathcal{P}_{\pi(i)}^*(h_{\pi(1)}, \dots, h_{\pi(K)}, \mathbf{s}_{\pi(1)}, \dots, \mathbf{s}_{\pi(K)}) \quad \forall i = 1 \dots K$$

also achieves the maximum in (9).

Proof: For every permutation π

$$\begin{aligned} \bar{C}_{\text{sum}}(\mathcal{P}^\pi) &= \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sigma^{-2} \sum_{i=1}^K h_i \mathbf{s}_i \mathbf{s}_i^t \mathcal{P}_i^\pi(h_1, \dots, h_K, S) \right) \right] \\ &= \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + \sigma^{-2} \sum_{i=1}^K h_{\pi(i)} \mathbf{s}_{\pi(i)} \mathbf{s}_{\pi(i)}^t \mathcal{P}_{\pi(i)}^*(h_{\pi(1)}, \dots, h_{\pi(K)}, \mathbf{s}_{\pi(1)}, \dots, \mathbf{s}_{\pi(K)}) \right) \right] \end{aligned} \quad (37)$$

$$= \bar{C}_{\text{sum}}(\mathcal{P}^*) \quad (38)$$

where (38) follows from the observation that the random variables are permuted (by π) in (37) and by the hypothesis that h_1, \dots, h_K are exchangeable and $\mathbf{s}_1, \dots, \mathbf{s}_K$ i.i.d. This completes the proof of the proposition. \square

We conclude by an appeal to the concavity of the map \bar{C}_{sum} in Proposition 4.2 that an optimal power allocation property \mathcal{P}^* has the symmetry property below. For every permutation $\pi \in S$ and $\forall i = 1, \dots, K$,

$$\mathcal{P}_i^*(h_1, \dots, h_K, \mathbf{s}_1, \dots, \mathbf{s}_K) = \mathcal{P}_{\pi(i)}^*(h_{\pi(1)}, \dots, h_{\pi(K)}, \mathbf{s}_{\pi(1)}, \dots, \mathbf{s}_{\pi(K)}). \quad (39)$$

- 7) From the structure of the optimal power allocation policy in (35), 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 $\forall i = 1, \dots, K$

$$\mathcal{P}_i^*(h_1, \dots, h_K, S) \leq K_p \text{ a.s.}$$

where K_p is some universal constant that does not depend on N .

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

$$\bar{C}_{\text{opt}}^{(N)} = \max_{\mathcal{P} \in \mathcal{F}_3^{(N)}} \bar{C}_{\text{sum}}(\mathcal{P}). \quad (40)$$

Here

$$\mathcal{F}_3^{(N)} = \left\{ \mathcal{P}: \begin{array}{l} \mathcal{P} \text{ satisfies properties (36) and (39)} \\ \mathcal{P}_i(h_1, \dots, h_K, S) \in [0, K_p] \text{ a.s. } \forall i = 1 \dots K \end{array} \right\}. \quad (41)$$

B. 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 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 $\beta_i(\mathcal{P}) h_i \mathcal{P}_i(h_1, \dots, h_K, S)$ where

$$\beta_i(\mathcal{P}) = \mathbf{s}_i^t \left[\sigma^2 I + \sum_{j \neq i} h_j \mathbf{s}_j \mathbf{s}_j^t \mathcal{P}_j(h_1, \dots, h_K, S) \right]^{-1} \mathbf{s}_i \quad \forall i = 1 \dots K. \quad (42)$$

The SIR is random since it depends on the particular realizations of the signature sequences and 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

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

for each $i = 1, \dots, K$ where g is a nonnegative function bounded by K_g . Denote the corresponding SIRs of the LMMSE estimates (defined in (42)) of the users as $\beta_1(\mathcal{P}^g), \dots, \beta_K(\mathcal{P}^g)$. Then it is straightforward to see that

The random variables $\beta_1(\mathcal{P}^g), \dots, \beta_K(\mathcal{P}^g)$ are identically distributed. (43)

In a large system, the central result of [23] 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 ([23, Theorem 3.1]):

$$\beta_1(\mathcal{P}^g) \xrightarrow{a.s.} \beta_g^*, \quad \text{as } N \rightarrow \infty \quad (44)$$

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

$$\sigma^2 \beta = 1 - \alpha \int_0^\infty \frac{\beta h g(h)}{1 + \beta h g(h)} dF(h). \quad (45)$$

Recall that F is the (same) marginal distribution of the fading gains h_1, \dots, h_K . Convergence of $\beta_1(\mathcal{P}^g)$ in measure first appeared as [23, Theorem 3.1] and the pointwise convergence (a

natural extension) follows as a consequence of the main result in [20] which shows that the empirical distribution of the eigenvalues of the matrix $\sum_{i=1}^K \mathbf{s}_i \mathbf{s}_i^t h_i$ 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

$$\beta_1(\bar{\mathcal{P}}) \xrightarrow{a.s.} \beta^*(\bar{p}, \alpha). \quad (46)$$

It is easily verified from (45) that $\beta^*(\bar{p}, \alpha)$ is the unique positive solution of the fixed-point equation in β

$$\sigma^2 \beta = 1 - \alpha \frac{\bar{p} \beta}{1 + \bar{p} \beta} \quad (47)$$

and hence $\beta^*(\bar{p}, \alpha)$ is the positive root of the quadratic equation (in β) (this was first observed in [27, eq. (6.62)] for binary random signature sequences)

$$\sigma^2 \beta^2 \bar{p} + \beta(\sigma^2 + \bar{p}(\alpha - 1)) - 1 = 0 \quad (48)$$

and can be explicitly written out as

$$\beta^*(\bar{p}, \alpha) = \frac{1 - \alpha}{2\sigma^2} - \frac{1}{2\bar{p}} + \sqrt{\frac{(1 - \alpha)^2}{4\sigma^4} + \frac{1 + \alpha}{2\bar{p}\sigma^2} + \frac{1}{4\bar{p}^2}}. \quad (49)$$

C. 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 as well

$$\beta_1(\mathcal{P}^g) = \mathbf{s}_1^t \left(\sigma^2 I + \sum_{j \neq 1} h_j \mathbf{s}_j \mathbf{s}_j^t g(h_j) \right)^{-1} \mathbf{s}_1; \quad (50)$$

from (42)

$$\leq \sigma^{-2} \mathbf{s}_1^t \mathbf{s}_1 = \frac{\sigma^{-2}}{N} \sum_{i=1}^K v_{1i}^2 \quad (51)$$

$$\mathbb{E}[(\beta_1(\mathcal{P}^g))^2] \leq C_1 \sigma^{-4} \quad (52)$$

where C_1 is a constant independent of N . It now follows from (44) and the dominated convergence theorem that

$$\beta_1(\mathcal{P}^g) \xrightarrow{L^2} \beta_g^*. \quad (53)$$

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:

$$\mathbb{E}[(\beta_1(\mathcal{P}^g) - \beta_g^*)^2] < \frac{C_2^2}{N} \quad (54)$$

where C_2 is some constant independent of N .

The lemma is proved in Appendix F.

D. Existence of β_{wf}^*

In Section III, we derived heuristically the asymptotic structure of the optimal power allocation policy to be (from (27))

$$\mathcal{P}_i^*: (h_1, \dots, h_K, S) \mapsto g_{\text{wf}}(h_i) \stackrel{\text{def}}{=} \left(\frac{1}{\lambda} - \frac{1}{\beta_{\text{wf}}^* h_i} \right)^+ \quad (55)$$

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 with respect to the fading statistics) is equal to \bar{p} . We now prove

the existence of this quantity β_{wf}^* . From (45), β_{wf}^* is the unique positive solution to the integral fixed-point equation

$$\begin{aligned} \sigma^2 \beta_{\text{wf}}^* &= 1 - \alpha \int_0^\infty \frac{\beta_{\text{wf}}^* h \left(\frac{1}{\lambda} - \frac{1}{\beta_{\text{wf}}^* h} \right)^+ dF(h)}{1 + \beta_{\text{wf}}^* h \left(\frac{1}{\lambda} - \frac{1}{\beta_{\text{wf}}^* h} \right)^+} \\ &= 1 - \alpha \int_{\lambda/\beta_{\text{wf}}^*}^\infty \left(1 - \frac{\lambda}{\beta_{\text{wf}}^* h} \right) dF(h). \end{aligned} \quad (56)$$

Furthermore, by the average power constraint of \bar{p} on the power allocation in (55), we have another equation relating λ and β_{wf}^* . Denoting the ratio $\frac{\lambda}{\beta_{\text{wf}}^*}$ by h_{thr} , the fading threshold level below which no power is transmitted, we see that the average power constraint in our notation yields

$$h_{\text{thr}} = \frac{HM(h_{\text{thr}})}{1 + \frac{\bar{p}HM(h_{\text{thr}})}{1 - F(h_{\text{thr}})}, \beta_{\text{wf}}^*} \quad (57)$$

where $HM(h)$ is the ‘‘harmonic mean of $[h, \infty)$ with respect to the distribution F ’’ defined as

$$HM(h) \stackrel{\text{def}}{=} (1 - F(h)) \left(\int_h^\infty \frac{dF(\tilde{h})}{\tilde{h}} \right)^{-1}. \quad (58)$$

Observe that $HM(h) > h, \forall h \in [0, \infty)$. Continuing from (56), we have

$$\sigma^2 \beta_{\text{wf}}^* = 1 - \alpha \int_{h_{\text{thr}}}^\infty \left(1 - \frac{h_{\text{thr}}}{h} \right) dF(h) \quad (59)$$

$$= 1 - \alpha(1 - F(h_{\text{thr}})) + \frac{\alpha h_{\text{thr}}(1 - F(h_{\text{thr}}))}{HM(h_{\text{thr}})} \quad (60)$$

$$= 1 - \alpha(1 - F(h_{\text{thr}})) + \alpha \frac{(1 - F(h_{\text{thr}}))^2}{1 - F(h_{\text{thr}}) + \beta_{\text{wf}}^* \bar{p} HM(h_{\text{thr}})} \quad (61)$$

where (59) uses the definition of h_{thr} as $\frac{\lambda}{\beta_{\text{wf}}^*}$, (60) follows from our notation of harmonic mean in (58) and we used (57) in (61). Comparing (61) with (47) we see that β_{wf}^* is equal to $\beta^*\left(\frac{\bar{p}HM(h_{\text{thr}})}{1 - F(h_{\text{thr}})}, \alpha(1 - F(h_{\text{thr}}))\right)$, 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 $\frac{\bar{p}HM(h_{\text{thr}})}{1 - F(h_{\text{thr}})}$ and number of users per unit processing gain equal to $\alpha(1 - F(h_{\text{thr}}))$. Thus β_{wf}^* has an explicit form as given in (49). Substituting this structure of β_{wf}^* in (57), we see that our claim is verified if we can show the existence of a solution h_{thr} satisfying (57). Denoting

$$\mathcal{K}(h) = \frac{HM(h)}{1 + \frac{\bar{p}HM(h)}{1 - F(h)} \beta^*\left(\frac{\bar{p}HM(h)}{1 - F(h)}, \alpha(1 - F(h))\right)}, \quad h \in (0, \infty) \quad (62)$$

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 G. From this, we conclude the existence of a unique β_{wf}^* with the properties derived in Section III.

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} .

For completeness, we would like to mention that there is a ‘‘shorter’’ way of seeing the existence of β_{wf}^* from (56) and the average power constraint on the water-filling policy in (55). The

constraint that the average power of the water-filling policy in (55) be \bar{p} is simply

$$\int_{\lambda/\beta_{\text{wf}}^*}^{\infty} \left(\frac{1}{\lambda} - \frac{1}{\beta_{\text{wf}}^* h} \right) dF(h) = \bar{p}. \quad (63)$$

Comparing this with (56) we arrive at

$$\sigma^2 \beta_{\text{wf}}^* = 1 - \alpha \bar{p} \lambda.$$

Combining this with (63), it is straightforward to see that there is a unique solution for λ and β_{wf}^* . However, our “slightly longer” derivation of the existence of β_{wf}^* culminating in the expression $\beta^* \left(\frac{\bar{p} H M(h_{\text{thr}})}{1 - F(h_{\text{thr}})}, \alpha(1 - F(h_{\text{thr}})) \right)$ seems to suggest the following physical interpretation for β_{wf}^* : it is the same SIR as would be seen by a unit power user in a hypothetical system with the same ratio of users to processing gain α and with all the users following the “truncated channel inversion” power policy

$$(h_1, \dots, h_K, S) \mapsto \begin{cases} \frac{\tilde{\lambda}}{h}, & \text{if } h \geq h_{\text{thr}} \\ 0, & \text{else} \end{cases}$$

where h_{thr} is the fixed point of the map \mathcal{K} given by Lemma 4.10 and $\tilde{\lambda}$ is some constant chosen such that the average power constraint is met.

V. PROOF OF MAIN RESULT

In this section we formally prove the asymptotic optimality of the water-filling power allocation strategy heuristically identified earlier in Section III. The proof is quite involved and for pedagogical reasons we first focus on the scenario when there is no fading. In this simple case of no fading, the water-filling power allocation reduces to the simple constant power allocation policy and the proof of this scenario contains several of the key elements of the general proof while being easy to follow.

A. No-Fading Scenario

This is the case when $h_i = 1$ and we begin with the long signature sequences channel model.

1) *Long Signature Sequence Channel Model*: For this scenario, the authors in [28] conjectured that asymptotically the optimal power allocation policy is to allocate equal powers to all users independent of signature sequences. The water-filling 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, to 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, long signature sequence channel model

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

Define the function L (the “Lagrangian”) as

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

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 equation in (48). Observe that L is just the sum of \bar{C}_{sum} and a linear functional and hence is also a concave function (Proposition 4.2). Furthermore, $L(\mathcal{P}) = \bar{C}_{\text{sum}}(\mathcal{P})$ over $\mathcal{F}_3^{(N)}$ [see (41)]. Recall our earlier notation (from Section IV) 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 the statement of Proposition 4.5)

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

Using the concavity of the map C (Proposition 4.2) in the first step (64), we arrive at

$$\begin{aligned} C(p_1^*, \dots, p_K^*) - C(\bar{p}, \dots, \bar{p}) \\ \leq \sum_{i=1}^K \frac{\partial C}{\partial p_i}(\bar{p}, \dots, \bar{p}) \cdot (p_i^* - \bar{p}) \end{aligned} \quad (64)$$

$$= \frac{1}{2N} \sum_{i=1}^K \mathbf{s}_i^t \left(\sigma^2 I + \sum_{j=1}^K \mathbf{s}_j \mathbf{s}_j^t \bar{p} \right)^{-1} \mathbf{s}_i (p_i^* - \bar{p}) \quad (65)$$

$$= \frac{1}{2N} \sum_{i=1}^K \frac{\beta_i(\bar{\mathcal{P}})}{1 + \beta_i(\bar{\mathcal{P}})\bar{p}} (p_i^* - \bar{p}) \quad (66)$$

where we used (18) to arrive at (65) (also see (134) and (135) in the proof of Proposition 4.5). We arrive at (66) by using the definition of $\beta_i(\bar{\mathcal{P}})$ from (42) (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. Averaging both sides of (66), we arrive at

$$\begin{aligned} L(\mathcal{P}^*) - L(\bar{\mathcal{P}}) \\ \leq \frac{1}{2N} \sum_{i=1}^K \mathbb{E} \left[\left(\frac{\beta_i(\bar{\mathcal{P}})}{1 + \beta_i(\bar{\mathcal{P}})\bar{p}} - \frac{\beta^*}{1 + \beta^* \bar{p}} \right) (P_i^*(S) - \bar{p}) \right] \end{aligned} \quad (67)$$

$$\leq \frac{K_p}{2N} \sum_{i=1}^K \mathbb{E} \left[\left(\frac{\beta_i(\bar{\mathcal{P}})}{1 + \beta_i(\bar{\mathcal{P}})\bar{p}} - \frac{\beta^*}{1 + \beta^* \bar{p}} \right) \right] \quad (68)$$

$$= \frac{K K_p}{2N} \mathbb{E} \left[\left(\frac{\beta_1(\bar{\mathcal{P}})}{1 + \beta_1(\bar{\mathcal{P}})\bar{p}} - \frac{\beta^*}{1 + \beta^* \bar{p}} \right) \right] \quad (69)$$

$$\leq \frac{K K_p}{2N} \mathbb{E} [|\beta_1(\bar{\mathcal{P}}) - \beta^*|] \quad (70)$$

$$\leq \frac{\alpha K_p C_2}{2\sqrt{N}} \quad (71)$$

where (68) follows from Theorem 4.7, (69) from (43), (70) follows from the fact that the map $x \mapsto \frac{x}{1+x}$ is contractive, and (71) follows from (46) and Lemma 4.9. Observing that

$$L(\mathcal{P}^*) = \bar{C}_{\text{sum}}(\mathcal{P}^*)$$

and

$$L(\bar{\mathcal{P}}) = \bar{C}_{\text{sum}}(\bar{\mathcal{P}})$$

the theorem follows. \square

2) *Short Signature Sequence Channel Model*: We now turn our attention to the short signature sequences model while retaining the assumption of no fading. Our main result is that the

constant power allocation policy is near optimal in terms of the supportable rate at a given low outage probability value. Recall our definition of $R_a(\mathcal{P}^g)$ in (29) and the notation of $\bar{\mathcal{P}}$ as the constant power allocation policy. The formal result is below and we conclude that in a large system the 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 .

Theorem 5.2: In the no-fading, short signature sequence channel model

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

Proof: The supportable rate at outage probability a with static power allocation satisfies

$$R_a^{*(N)} \leq \frac{\mathbb{E}[C_{\text{opt}}(S)]}{1-a} \quad (73)$$

$$\leq \frac{\bar{C}_{\text{opt}}^{(N)}}{1-a} \quad (74)$$

$$\leq \frac{\bar{C}_{\text{sum}}(\bar{\mathcal{P}}) + O\left(\frac{1}{\sqrt{N}}\right)}{1-a} \quad (75)$$

$$\leq \frac{\bar{C}_{\text{sum}}^{\bar{\mathcal{P}}} + o(1) + O\left(\frac{1}{\sqrt{N}}\right)}{1-a} \quad (76)$$

$$\leq \frac{R_a(\bar{\mathcal{P}}) + O\left(\frac{1}{\sqrt{N}}\right) + o(1)}{1-a} \quad (77)$$

where (73) follows from the definition of R_a^* in (6) and the Markov inequality, (74) 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 (75), (76) comes from Proposition 3.1 with g being the constant function equal to \bar{p} , and, finally, (77) follows from (30). Thus we arrive at (72). \square

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 .

B. General Fading Case

We now turn to the general scenario with flat fading and first consider the long signature sequences model. The proof of the asymptotic optimality of the water-filling 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 water-filling strategy. Let us denote the water-filling strategy of (55) by

$$\mathcal{P}_i^{\text{wf}}: (h_1, \dots, h_K, S) \mapsto g_{\text{wf}}(h_i) \stackrel{\text{def}}{=} \frac{1}{\beta_{\text{wf}}^*} \left(\frac{1}{h_{\text{thr}}} - \frac{1}{h_i} \right)^+ \quad (78)$$

Recall that $\beta_{\text{wf}}^* = \beta^*\left(\frac{\bar{p}HM(h_{\text{thr}})}{1-F(h_{\text{thr}})}, \alpha(1-F(h_{\text{thr}}))\right)$ and the threshold h_{thr} below which no power is transmitted is the unique fixed point of \mathcal{K} in (62). The formal statement of the asymptotic optimality of the water-filling policy \mathcal{P}^{wf} that also identifies the order of the loss in sum capacity is below.

Theorem 5.3:

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

Proof: Define the function L (the ‘‘Lagrangian’’) as

$$\begin{aligned} L: \mathcal{P} \mapsto & \bar{C}_{\text{sum}}(\mathcal{P}) - \frac{\beta_{\text{wf}}^* h_{\text{thr}}}{2N} \sum_{i=1}^K \mathbb{E}[\mathcal{P}_i(h_1, \dots, h_K, S) - \bar{p}] \\ & + \frac{1}{2N} \sum_{i=1}^K \mathbb{E} [1_{\{h_i \leq h_{\text{thr}}\}} (\beta_{\text{wf}}^* h_{\text{thr}} - h_i \beta_i(\mathcal{P}^{\text{wf}})) \\ & \quad \cdot \mathcal{P}_i(h_1, \dots, h_K, S)] \end{aligned} \quad (79)$$

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 IV of \mathcal{P}^* that maximizes \bar{C}_{sum} over $\mathcal{F}_3^{(N)}$. We proceed by the following steps.

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

$$|L(\mathcal{P}^*) - L(\mathcal{P}^{\text{wf}})| < O\left(\frac{1}{\sqrt{N}}\right). \quad (80)$$

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

$$\liminf_{N \rightarrow \infty} (L(\mathcal{P}^*) - \bar{C}_{\text{sum}}(\mathcal{P}^*)) \geq 0. \quad (81)$$

Combining the observation that $L(\mathcal{P}^{\text{wf}}) = \bar{C}_{\text{sum}}(\mathcal{P}^{\text{wf}})$ with the two steps above proves the theorem. We first show (80) and then (81).

Analogous to (64), for every realization of fading gains h_1, \dots, h_K and signature sequences S , we have from the concavity of the map C (Proposition 4.2) that

$$\begin{aligned} & C(\mathcal{P}^*(h_1, \dots, h_K, S)) - C(\mathcal{P}^{\text{wf}}(h_1, \dots, h_K, S)) \\ & \leq \sum_{i=1}^K \frac{\partial C}{\partial p_i}(\mathcal{P}^{\text{wf}}(h_1, \dots, h_K, S)) \\ & \quad \cdot (\mathcal{P}_i^*(h_1, \dots, h_K, S) - \mathcal{P}_i^{\text{wf}}(h_1, \dots, h_K, S)) \quad (82) \\ & = \frac{1}{2N} \sum_{i=1}^K \frac{\beta_i(\mathcal{P}^{\text{wf}}) h_i (\mathcal{P}_i^*(h_1, \dots, h_K, S) - \mathcal{P}_i^{\text{wf}}(h_i))}{1 + \beta_i(\mathcal{P}^{\text{wf}}) h_i \mathcal{P}_i^{\text{wf}}(h_i)}. \end{aligned} \quad (83)$$

In (83) we have emphasized the fact that $\mathcal{P}_i^{\text{wf}}$ is only a function of h_i . Using our notation in (78) and averaging both sides of (83), we arrive at

$$\begin{aligned} L(\mathcal{P}^*) - L(\mathcal{P}^{\text{wf}}) &\leq \frac{1}{2N} \sum_{i=1}^K \mathbb{E} \left[\mathbf{1}_{\{h_i > h_{\text{thr}}\}} \left(\frac{\beta_i(\mathcal{P}^{\text{wf}})h_i}{(1 + \beta_i(\mathcal{P}^{\text{wf}})h_i g_{\text{wf}}(h_i))} - \beta_{\text{wf}}^* h_{\text{thr}} \right) \right. \\ &\quad \left. \cdot (\mathcal{P}_i^*(h_1, \dots, h_K, S) - g_{\text{wf}}(h_i)) \right]. \end{aligned} \quad (84)$$

In (84) we used the fact (by definition) that

$$g_{\text{wf}}(h_i) \mathbf{1}_{\{h_i < h_{\text{thr}}\}} = 0.$$

Continuing from (84)

$$\begin{aligned} L(\mathcal{P}^*) - L(\mathcal{P}^{\text{wf}}) &\leq \frac{KK_p}{2N} \mathbb{E} \left[\mathbf{1}_{\{h_1 > h_{\text{thr}}\}} \left| \frac{\beta_1(\mathcal{P}^{\text{wf}})h_1}{(1 + \beta_1(\mathcal{P}^{\text{wf}})h_1 g_{\text{wf}}(h_1))} - \beta_{\text{wf}}^* h_{\text{thr}} \right| \right] \end{aligned} \quad (85)$$

where we used (43) and Theorem 4.7. By definition, β_{wf}^* is equal to $\beta_{g_{\text{wf}}}^*$ and thus from Lemma 4.8 we have

$$\beta_1(\mathcal{P}^{\text{wf}}) \xrightarrow{\text{a.s.}} \beta_{\text{wf}}^*, \quad \text{as } N \rightarrow \infty.$$

By definition of g_{wf} (recall (78)) we get

$$\frac{\mathbf{1}_{\{h_1 > h_{\text{thr}}\}} \beta_{\text{wf}}^* h_1}{1 + \beta_{\text{wf}}^* h_1 g_{\text{wf}}(h_1)} = \beta_{\text{wf}}^* h_{\text{thr}} \mathbf{1}_{\{h_1 > h_{\text{thr}}\}}. \quad (86)$$

Using the fact that the map $x \mapsto \frac{x}{1+x}$ is contractive, (86) and (85) yield

$$\begin{aligned} L(\mathcal{P}^*) - L(\mathcal{P}^{\text{wf}}) &\leq \frac{\alpha K_p}{2} \mathbb{E} [h_1 \mathbf{1}_{\{h_1 > h_{\text{thr}}\}} |\beta_1(\mathcal{P}^{\text{wf}}) - \beta_{\text{wf}}^*|] \\ &\leq \frac{\alpha C_2 K_p (\mathbb{E}[h_1^2])^{\frac{1}{2}}}{2\sqrt{N}} \end{aligned} \quad (87)$$

where we used Lemma 4.9 and the Cauchy–Schwartz inequality to arrive at (87). We have thus shown (80).

To show (81), fix $\epsilon > 0$. Using Lemma 4.9, we have from a Chebyshev bound

$$\begin{aligned} \mathbb{P} \left[\frac{\beta_1(\mathcal{P}^{\text{wf}})}{\beta_{\text{wf}}^*} \geq 1 + \epsilon \right] &\leq \frac{\mathbb{E} \left[|\beta_1(\mathcal{P}^{\text{wf}}) - \beta_{\text{wf}}^*|^2 \right]}{\beta_{\text{wf}}^{*2} \epsilon^2} \\ &\leq \left(\frac{C_2}{\beta_{\text{wf}}^* \epsilon \sqrt{N}} \right)^2. \end{aligned} \quad (88)$$

Then, using properties (39) and (36) of \mathcal{P}^* and (43) we have

$$\begin{aligned} L(\mathcal{P}^*) &= \bar{C}_{\text{sum}}(\mathcal{P}^*) + \frac{K\beta_{\text{wf}}^*}{2N} \mathbb{E} \left[\mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \left(h_{\text{thr}} - \frac{h_1 \beta_1(\mathcal{P}^{\text{wf}})}{\beta_{\text{wf}}^*} \right) \mathcal{P}_1^*(h_1, \dots, h_K, S) \right]. \end{aligned} \quad (89)$$

Consider the case

$$\liminf_{N \rightarrow \infty} \mathbb{E} [\mathcal{P}_1^*(h_1, \dots, h_K, S) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}] = 0. \quad (90)$$

Using Theorem 4.7, (90) leads to

$$\begin{aligned} \liminf_{N \rightarrow \infty} \mathbb{E} [(\mathcal{P}_1^*(h_1, \dots, h_K, S))^2 \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}] \\ \leq K_p \liminf_{N \rightarrow \infty} \mathbb{E} [\mathcal{P}_1^*(h_1, \dots, h_K, S) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}] \\ = 0. \end{aligned} \quad (91)$$

Then it follows from (89) that there exists a subsequence $\{N_{i_n}\}_n$ such that

$$\begin{aligned} L(\mathcal{P}^*) - \bar{C}_{\text{sum}}(\mathcal{P}^*) &\geq -\frac{\alpha}{2} \mathbb{E} [h_1 \beta_1(\mathcal{P}^{\text{wf}}) \mathcal{P}_1^*(h_1, \dots, h_K, S) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}] \\ &\geq -\frac{\alpha}{2} \mathbb{E} [(\mathcal{P}_1^*(h_1, \dots, h_K, S))^3 \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}]^{\frac{1}{3}} \\ &\quad \cdot \mathbb{E} [(\beta_1(\mathcal{P}^{\text{wf}})h_1)^{\frac{3}{2}}]^{\frac{2}{3}} \end{aligned} \quad (92)$$

$$\geq -\frac{\alpha \sqrt{C_1} \mathbb{E}[h_1^2]^{\frac{1}{2}}}{2\sigma^2} \mathbb{E} [(\mathcal{P}_1^*(h_1, \dots, h_K, S))^3 \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}]^{\frac{1}{3}}. \quad (93)$$

We used Holder inequality to arrive at (92) and Holder inequality again combined with the bound in (52) to arrive at (93). We conclude from (90) and (93) that

$$\liminf_{n \rightarrow \infty} L(\mathcal{P}^{*(N_{i_n})}) - \bar{C}_{\text{sum}}(\mathcal{P}^{*(N_{i_n})}) \geq 0$$

and we have thus shown (81) (the notation of the superscript N in $\mathcal{P}^{*(N)}$ denotes that $\mathcal{P}^{*(N)} \in \mathcal{F}_3^{(N)}$). Now suppose (90) does not hold and hence we have

$$\liminf_{N \rightarrow \infty} \mathbb{E} [\mathcal{P}_1^*(h_1, \dots, h_K, S) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}}] > 0. \quad (94)$$

We evaluate the integral in (89) over the two disjoint sets

$$\mathcal{A}_1 \stackrel{\text{def}}{=} \{\beta_1(\mathcal{P}^{\text{wf}}) \geq \beta_{\text{wf}}^*(1 + \epsilon)\}$$

and

$$\mathcal{A}_2 \stackrel{\text{def}}{=} \{\beta_1(\mathcal{P}^{\text{wf}}) < \beta_{\text{wf}}^*(1 + \epsilon)\}.$$

As usual, $\mathbf{1}_{\mathcal{A}_i}$ denotes the indicator function over the set \mathcal{A}_i , $i = 1, 2$. We have

$$\begin{aligned} \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) \left(h_{\text{thr}} - \frac{h_1 \beta_1(\mathcal{P}^{\text{wf}})}{\beta_{\text{wf}}^*} \right) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_1} \right] \\ \geq -\frac{K_p h_{\text{thr}}}{\beta_{\text{wf}}^*} \mathbb{E} [\beta_1(\mathcal{P}^{\text{wf}}) \mathbf{1}_{\mathcal{A}_1}] \end{aligned} \quad (95)$$

$$\geq -\frac{K_p h_{\text{thr}}}{\beta_{\text{wf}}^* \sigma^2} \mathbb{E} [s_1^t s_1 \mathbf{1}_{\mathcal{A}_1}] \quad (96)$$

$$\geq -\frac{K_p h_{\text{thr}} \sqrt{C_1}}{\beta_{\text{wf}}^* \sigma^2} \left(\frac{C_2}{\epsilon \beta_{\text{wf}}^* \sqrt{N}} \right) \quad (97)$$

where we used Theorem 4.7 in (95), (51) to derive (96), and (52) combined with the Chebyshev bound of (88) in arriving at (97). We also have

$$\begin{aligned}
& \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) \left(h_{\text{thr}} - \frac{h_1 \beta_1(\mathcal{P}^{\text{wf}})}{\beta_{\text{wf}}^*} \right) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_2} \right] \\
& \geq \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) (h_{\text{thr}} - h_1(1 + \epsilon)) \right. \\
& \quad \left. \cdot \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_2} \right] \\
& = \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) (h_{\text{thr}} - h_1) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_2} \right] \\
& \quad - \epsilon \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) h_1 \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_2} \right] \\
& \geq \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) (h_{\text{thr}} - h_1) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_2} \right] \\
& \quad - \epsilon h_{\text{thr}} K_p \left(1 - \frac{C_2^2}{\beta_{\text{wf}}^* \epsilon^2 N} \right). \tag{98}
\end{aligned}$$

From (94), we have

$$\liminf_{N \rightarrow \infty} \mathbb{E} \left[\mathcal{P}_1^*(h_1, \dots, h_K, S) (h_{\text{thr}} - h_1) \mathbf{1}_{\{h_1 \leq h_{\text{thr}}\}} \mathbf{1}_{\mathcal{A}_2} \right] > 0. \tag{99}$$

Letting $\epsilon = \frac{1}{\log N}$ and combining (97)–(99) we have shown (81) that

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

completing the proof. \square

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

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

VI. OPTIMAL POWER ALLOCATION AND SYSTEM PARAMETERS

In this section we study the behavior of the water-filling 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 water-filling 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.

A. Dependence on the Number of Users per Unit Processing Gain

Recall the water-filling power allocation strategy defined in (78)

$$\mathcal{P}_i^{\text{wf}}: (h_1, \dots, h_K, S) \mapsto g_{\text{wf}}(h_i) \stackrel{\text{def}}{=} \frac{1}{\beta_{\text{wf}}^*} \left(\frac{1}{h_{\text{thr}}} - \frac{1}{h_i} \right)^+. \tag{100}$$

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 water-filling 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 water-filling strategy. In the scenario with very large α there are many users with essentially the same signature sequence, and comparing with the Knopp and Humblet strategy [11], we expect that users transmit only when they have very good channels. In the following result we make this intuitive observation precise:

Proposition 6.1: Recall \mathcal{P}^{wf} , the water-filling power allocation strategy (100), and the single-user water-filling strategy (15). Then

$$h_{\text{thr}} \downarrow h_{\text{wf}} \text{ and } \beta_{\text{wf}}^* \uparrow \sigma^{-2} \text{ as } \alpha \downarrow 0 \tag{101}$$

$$h_{\text{thr}} \uparrow \sup\{h: F(h) < 1\} \text{ and } \beta_{\text{wf}}^* \downarrow 0 \text{ as } \alpha \uparrow \infty. \tag{102}$$

The proof is found in Appendix H.

B. Dependence on SNR

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 water-filling strategy gains very little over the static power allocation policy, namely, equal power allocation over all fading states. This was observed in [7] through simulation studies with Rayleigh and Nakagami fading examples. We make this statement precise and use it to find the structure of the water-filling 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

$$\mathcal{P}^*(h) = \left(\frac{\sigma^2}{h_{\text{wf}}} - \frac{\sigma^2}{h} \right)^+. \tag{103}$$

Proposition 6.2: For the single-user channel, at high SNR, the optimal power allocation (103) converges to the constant power policy and furthermore the loss in capacity by using the constant power policy goes to zero. Formally, as $\sigma^2 \rightarrow 0$

$$\mathcal{P}^* \xrightarrow{\text{a.s.}} \bar{p} \tag{104}$$

$$\bar{C}(\mathcal{P}^*) - \bar{C}(\bar{p}) \rightarrow 0. \tag{105}$$

The proof is completely elementary. As $\sigma^2 \rightarrow 0$, to meet the average power constraint we must have $\frac{\sigma^2}{h_{\text{wf}}} \rightarrow \bar{p}$. Thus, the water-filling strategy converges to the static power allocation strategy at high SNR showing (104). The gain with water-filling strategy at any realization of the fading gain h is

$$\begin{aligned}
& \frac{1}{2} \left(\log \left(1 + \frac{h\mathcal{P}^*(h)}{\sigma^2} \right) - \log \left(1 + \frac{h\bar{p}}{\sigma^2} \right) \right) \\
& = \frac{1}{2} \left(\log \left(1 + \frac{h(\mathcal{P}^*(h) - \bar{p})}{\sigma^2 + h\bar{p}} \right) \right) \\
& \leq \frac{1}{2} \log \left(1 + \frac{\sigma^2}{h_{\text{wf}}\bar{p}} \right) \tag{106}
\end{aligned}$$

where we used the definition of $\mathcal{P}^*(h)$ as the single-user water-filling policy in (15). Thus, by the dominated convergence theorem and (104), we have shown (105).

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 water-filling 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 I.

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

1) For $\alpha \leq 1$, we have $h_{\text{thr}} \downarrow 0$ and $\beta_{\text{wf}}^* \uparrow \infty$. Furthermore,

$$\bar{C}_{\text{sum}}(\mathcal{P}^{\text{wf}}) - \bar{C}_{\text{sum}}(\bar{\mathcal{P}}) \rightarrow 0.$$

2) For $\alpha > 1$ we have $h_{\text{thr}} \downarrow h_o > 0$ and $\beta_{\text{wf}}^* \uparrow \frac{1}{\alpha \bar{p} h_o} < \infty$. Here h_o is the unique positive fixed point of the map

$$\mathcal{K}_o: h \mapsto \frac{HM(h)(\alpha(1-F(h)) - 1)^+}{\alpha(1-F(h))}.$$

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* [27, Ch. 5] 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 water-filling 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

$$\max_{i \neq j} (\mathbf{s}_i^t \mathbf{s}_j)^2 \xrightarrow{a.s.} 0$$

as K grows polynomially in N and $N \rightarrow \infty$.

C. 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 III shows that much of the analysis remains valid also in this case. 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 water-filling (78) 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

$$\mathcal{P}_i^*: \left(h_1, \dots, h_{\sum_{i=1}^L K_i}, S \right) \mapsto \left(\frac{1}{\lambda_l} - \frac{1}{\beta_{\text{wf}}^* h_i} \right)^+ \quad (107)$$

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 (56))

$$\sigma^2 \beta_{\text{wf}}^* = 1 - \sum_{l=1}^L \alpha_l \int_{\lambda_l / \beta_{\text{wf}}^*}^{\infty} \left(1 - \frac{\lambda_l}{\beta_{\text{wf}}^* h} \right) dF(h). \quad (108)$$

Analogous to the continuation in Section III 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 III. Denoting $h_{\text{thr}}^{(l)} \stackrel{\text{def}}{=} \frac{\lambda_l}{\beta_{\text{wf}}^*}$, the level below which no power is transmitted by users of class l , analogous to (57) we have, from the average power constraint on the power policy in (107), that $h_{\text{thr}}^{(l)}$ is the solution to the fixed-point equation

$$h_{\text{thr}}^{(l)} = \frac{HM(h_{\text{thr}}^{(l)})}{1 + \frac{\bar{p}_l HM(h_{\text{thr}}^{(l)})}{1-F(h_{\text{thr}}^{(l)})} \beta_{\text{wf}}^*}. \quad (109)$$

Continuing from (108), analogous to (59)–(61), we have using (109) that

$$\begin{aligned} \sigma^2 \beta_{\text{wf}}^* &= 1 - \sum_{l=1}^L \alpha_l \left(1 - F(h_{\text{thr}}^{(l)}) \right) \\ &\quad + \sum_{l=1}^L \alpha_l \frac{\left(1 - F(h_{\text{thr}}^{(l)}) \right)^2}{1 - F(h_{\text{thr}}^{(l)}) + \beta_{\text{wf}}^* \bar{p}_l HM(h_{\text{thr}}^{(l)})}. \end{aligned} \quad (110)$$

In the single class case we were able to observe that β_{wf}^* was equal to the solution of a quadratic equation (47). 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 for $l = 1, \dots, L$. As $N \rightarrow \infty$, assuming that $\frac{K_l}{N} \rightarrow \alpha_l$ for every class l , it follows from Lemma 4.8 that the asymptotic SIR of a unit (received) power user is a positive constant $\beta^* (\{p_l, \alpha_l\}_{l=1, \dots, L})$ that satisfies the fixed-point equation (analogous to (47))

$$\sigma^2 \beta = 1 - \sum_{l=1}^L \frac{\alpha_l \beta p_l}{1 + \beta p_l}. \quad (111)$$

Comparing (110) with (111) we observe that

$$\beta_{\text{wf}}^* = \beta^* \left(\left\{ \frac{\bar{p}_l HM(h_{\text{thr}}^{(l)})}{1 - F(h_{\text{thr}}^{(l)})}, \alpha_l (1 - F(h_{\text{thr}}^{(l)})) \right\}; l=1, \dots, L \right). \quad (112)$$

Analogous to the fixed-point iteration of the map in (62) for the single class scenario, we define the following maps for each class l :

$$\mathcal{K}_l: (h_1, \dots, h_L) \mapsto \frac{HM(h_l)}{1 + \frac{\bar{p}_l HM(h_l)}{1-F(h_l)} \beta^* \left(\left\{ \frac{\bar{p}_l HM(h_l)}{1-F(h_l)}, \alpha(1-F(h_l)) \right\}; l=1, \dots, L \right)}. \quad (113)$$

It follows from (109) and (112) that

$$\mathcal{K}_l \left(h_{\text{thr}}^{(1)}, \dots, h_{\text{thr}}^{(L)} \right) = h_{\text{thr}}^{(l)}.$$

Analogous to Lemma 4.10, we justify the existence of $h_{\text{thr}}^{(l)}$ by the following proposition.

Proposition 6.4: Consider the fixed-point iteration

$$h_l(0) \stackrel{\text{def}}{=} 0 \quad \forall l = 1 \dots L$$

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

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

Thus $\{h_{\text{thr}}^{(l)}\}$ 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 (154) in the proof of Lemma 4.10: For every $l = 1, \dots, L$

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

where

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

This also shows the uniqueness of $h_{\text{thr}}^{(l)}$. The formal statement of the optimality of this power allocation solution, analogous to Theorem 5.3, is given below and the key ideas of the proof are all contained in the proof of Theorem 5.3.

Theorem 6.5:

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

Extensions of the observations made in Section VI-B 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 water-filling scheme at high SNR if and only if $\sum_{l=1}^L \alpha_l \leq 1$.

VII. 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.

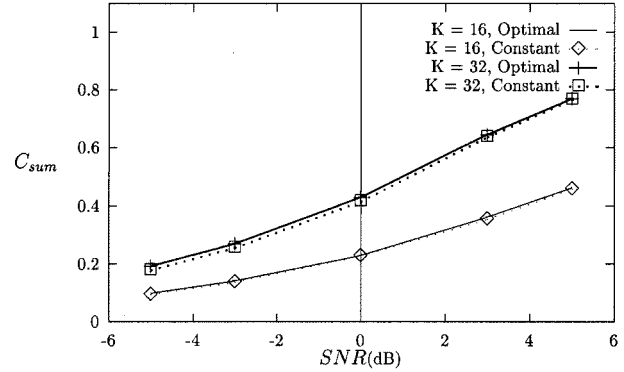


Fig. 1. No-fading scenario. Sum capacity is plotted with the optimal allocation and the constant power allocation policies with $N = 32$.

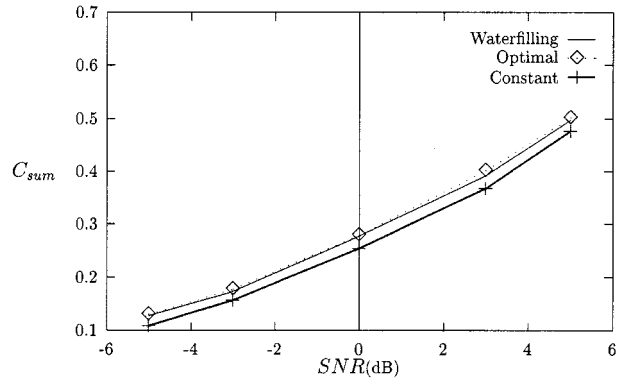


Fig. 2. Rayleigh fading scenario with $N = 32$ and $K = 16$. Sum capacity in bits/s/Hz is plotted with the optimal allocation, asymptotically optimal water-filling allocation, and the constant power allocation policies.

We assumed that the components of the signature sequences are distributed as zero-mean Gaussian random variables with variance $\frac{1}{\sqrt{N}}$ (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 in the scenario of no fading, 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, and was evaluated using the maxdet software, as explained at the end of the section). We observe 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, Figs. 2 and 3 plot sum capacity with three different power allocation policies: the asymptotically optimal water-filling 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 Figs. 2 and 3 is that the sum capacity with the asymptotically optimal policy of water-filling 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

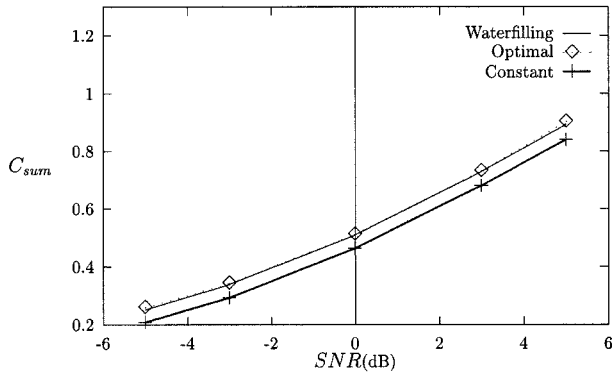


Fig. 3. Rayleigh fading scenario with $N = 32$ and $K = 32$. Sum capacity in bits/s/Hz is plotted with the optimal allocation, asymptotically optimal water-filling allocation, and the constant power allocation policies.

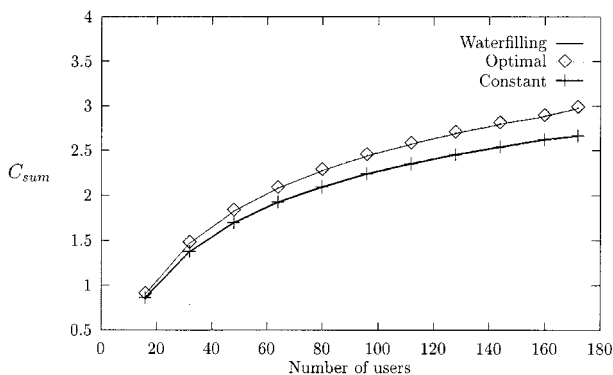


Fig. 4. Rayleigh fading scenario with $N = 16$ and $\text{SNR} = 5$ dB. Sum capacity is plotted with the optimal allocation, asymptotically optimal water-filling allocation, and the constant power allocation policies versus number of users.

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. 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 (5 dB) versus the number of users: while there is very little difference in sum capacity between the optimal power allocation and water-filling 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 [34] to arrive at the optimal power allocation; the software provides an interior point algorithm to solve the determinant maximization problem

$$\max_{p_i \geq 0} L(p_1, \dots, p_K, \lambda_1, \dots, \lambda_K) \quad (114)$$

where $L(p_1, \dots, p_K, \lambda_1, \dots, \lambda_K)$

$$\stackrel{\text{def}}{=} \log \det \left(I + \sum_{i=1}^K \mathbf{s}_i \mathbf{s}_i^H \frac{p_i}{\sigma^2} \right) + \sum_{i=1}^K \lambda_i (p_i - \bar{p}).$$

We obtained sum capacity at power prices $\lambda_1, \dots, \lambda_K$ by averaging the scaled (by $1/2N$) maximal value of the optimization problem above (114). 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 [14, Corollary 28.4.1]). From the proof of Theorem 5.3, we have a good guess for the Kuhn–Tucker coefficients: $\lambda_1 = \dots = \lambda_K = \beta_{\text{wf}}^* h_{\text{thr}}$. The *actual* power prices were found by a line search. The solution to the optimization problem (114) with the equilibrium power prices gives the optimal power allocation and thus we arrive numerically at sum capacity with the optimal power allocation policy.

The spectral efficiency increases with α for fixed SNR and increases with SNR for fixed α . This is clear from the structure of the successive decoding receiver. Every additional user can be decoded first and thus the spectral efficiency always increases with the number of users. However, with other types of receivers [in particular, a bank of linear receivers (the linear MMSE, decorrelator and matched filter)], the spectral efficiency does not increase monotonically with α for a given SNR. The variation of spectral efficiency with loading α for a given SNR is plotted in [19], along with a discussion of the spreading coding tradeoff.

VIII. MULTIPLE-ANTENNA SYSTEMS

In the multiple-antenna model, a baseband model for a synchronous multiple-access antenna array channel is

$$\mathbf{y}(n) = d(N) \sum_{i=1}^K x_i(n) \tilde{\mathbf{h}}_i^s(n) \tilde{\mathbf{h}}_i^f(n) + \mathbf{w}(n). \quad (115)$$

Here n denotes the time of channel use, $x_i(n)$ is the transmitted symbol of user i at time n , and $\mathbf{y}(n)$ is an N -dimensional vector of received symbols at the N antenna elements of the array at the receiver. The vector $\tilde{\mathbf{h}}_i^s(n) \tilde{\mathbf{h}}_i^f(n)$ represents the channel from the i th user to the antenna array at time n . The scalar $\tilde{h}_i^s(n)$ captures the slowly varying component of the fading channel—this depends on the distance from the user to the antenna array and geographical features and thus depends on the user alone. The vector $\tilde{\mathbf{h}}_i^f(n)$ is the fast varying component of the fading channel and is due to the constructive and destructive addition of the various multiple paths at the antenna array. Our assumption is that the antennas are spaced far enough apart (the spacing is at least half a wavelength and depends on the scattering environment) and thus the fast fading components are statistically independent at different antenna elements. We will assume that $\{\tilde{h}_i^s(n)\}_n$ and $\{\tilde{\mathbf{h}}_i^f(n)\}_n$ are independent complex stationary and ergodic processes. As in Section II-A, we denote the amplitude-squared process of $\{\tilde{h}_i^s(n)\}$ by $\{h_i^s(n)\}$ and assume that the stationary distribution (denoted by F) of this stationary, ergodic process has bounded first and second moments. Analogously, we assume that the random variables $\mathbf{h}_i^f(n)$ have

zero mean, unit variance, and bounded fourth moment. Since we have assumed that spacing between the antennas is large enough to allow the fast-fading component of the channel to be independent, we have introduced a penalty term $d(N)$ in the model (115). If the number of antennas is small (say 5 or 10), then $d(N)$ can be taken to be identity. But as the number of antennas increases, $d(N) \rightarrow 0$. In (115), $\mathbf{w}(n)$ is an additive white proper complex Gaussian noise process. As in the DS-CDMA model, we are interested in coherent communication, i.e., the receiver is able to perfectly track the fading channel and we allow feedback of the channel states to the users. Defining $\tilde{\mathbf{s}}_i \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \tilde{\mathbf{h}}_i^f$, by analogy to the DS-CDMA long signature sequence model and results of Section II-B, we arrive at the following expression for the (long-term) sum capacity of the multiple antenna MAC [analogous to (9) and (10)]:

$$\begin{aligned} \bar{C}_{\text{opt}}^{A(N)} &\stackrel{\text{def}}{=} \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \bar{C}_{\text{sum}}^A(\mathcal{P}) \\ &\stackrel{\text{def}}{=} \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \frac{1}{2N} \mathbb{E} \left[\log \det \left(I + d^2(N) N \sum_{i=1}^K \sigma^{-2} h_i^s \tilde{\mathbf{s}}_i \tilde{\mathbf{s}}_i^\dagger \mathcal{P}_i \right. \right. \\ &\quad \left. \left. \cdot (h_1^s, \dots, h_K^s, S) \right) \right]. \quad (116) \end{aligned}$$

Some remarks about this expression are in order: the power allocation policy \mathcal{P} depends on both the slow-fading components h_1^s, \dots, h_K^s and the fast-varying components $S = \frac{1}{\sqrt{N}} [\tilde{\mathbf{h}}_1^f, \dots, \tilde{\mathbf{h}}_K^f]$ and the sum capacity with the power allocation policy \mathcal{P} in nats per second per antenna is written as $\bar{C}_{\text{sum}}^A(\mathcal{P})$. 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 $d^2(N)N$. The quantity $d^2(N)$ captures the physical consequence of the fact that 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 retain the size of the antenna array) forcing the total received power to become constant. One natural assumption on the behavior of $d^2(N)$ is that

$$\frac{1}{N} \leq d^2(N) \leq 1.$$

This says that the total received power with multiple antennas is at least as much as that in the one-antenna case and the increase in received power is no more than linear in the number of antennas. When $d^2(N) = \frac{1}{N}$ the scenario is identical to that of the DS-CDMA model and we have the asymptotic optimality result of Theorem 5.3. When $d^2(N)N \rightarrow \infty$, we are in the regime of high SNR for the DS-CDMA model and can expect the results of Proposition 6.3 to be useful here. We state this result formally below and relegate the proof to Appendix J.

Theorem 8.1: We consider two limiting behaviors of $d^2(N)$.

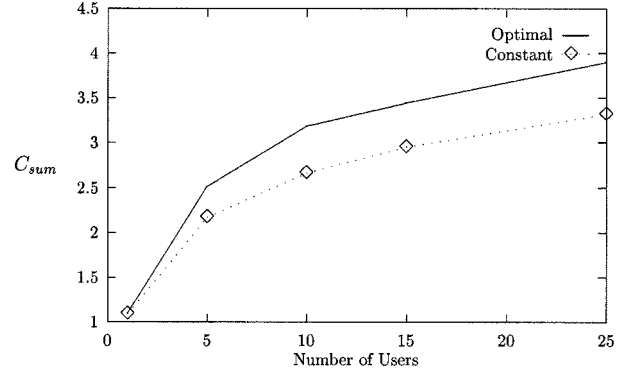


Fig. 5. Sum capacity in bits/s with one antenna at the receiver is plotted versus number of users at a fixed SNR level of 5 dB with both the optimal and constant power allocations.

- 1) Case 1: $d^2(N)N \rightarrow 1$: Define the water-filling policy

$$\mathcal{P}_i^{\text{wf}}: (h_j^s, \tilde{\mathbf{h}}_j^f, \forall j) \mapsto \frac{1}{\beta_{\text{wf}}^*} \left(\frac{1}{h_{\text{thr}}} - \frac{1}{h_i^s} \right)^+$$

where β_{wf}^* and h_{thr} are as defined in Section IV-D. Then

$$\limsup_{N \rightarrow \infty} \sqrt{N} \left(\bar{C}_{\text{opt}}^{A(N)} - \bar{C}_{\text{sum}}^A(\mathcal{P}^{\text{wf}}) \right) < \infty. \quad (117)$$

- 2) Case 2: $d^2(N)N \rightarrow \infty$:

- a) Suppose $\alpha \leq 1$. Then

$$\lim_{N \rightarrow \infty} \left(\bar{C}_{\text{opt}}^{A(N)} - \bar{C}_{\text{sum}}^A(\bar{\mathcal{P}}) \right) = 0. \quad (118)$$

- b) Suppose $\alpha > 1$. Define the water-filling policy

$$\tilde{\mathcal{P}}_i^{\text{wf}}: (h_j^s, \tilde{\mathbf{h}}_j^f, \forall j) \mapsto \alpha \bar{p} h_o \left(\frac{1}{h_o} - \frac{1}{h_i^s} \right)^+$$

where h_o is as defined in Proposition 6.3. Then

$$\lim_{N \rightarrow \infty} \left(\bar{C}_{\text{opt}}^{A(N)} - \bar{C}_{\text{sum}}^A(\tilde{\mathcal{P}}^{\text{wf}}) \right) = 0. \quad (119)$$

Observe from the proof of Proposition 6.3 that h_o decreases to zero as α decreases to 1 and thus the policy $\tilde{\mathcal{P}}^{\text{wf}}$ becomes the constant power allocation policy as α decreases to 1. Theorem 8.1 poses a contrasting picture 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 water-filling power policy [11]. This policy was also seen (in Section III) directly from (35) in Proposition 4.5 that the optimal power policies satisfy. The gain in sum capacity by following this strategy over the sub-optimal policy of constant power allocation to the users at all fading levels can be substantial; the larger the number of users, the larger this gain. 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. This gain stems from two factors: channel state feedback followed by an appropriate power allocation policy and the fact that there are multiple users. However, when there is a substantial number of antennas,

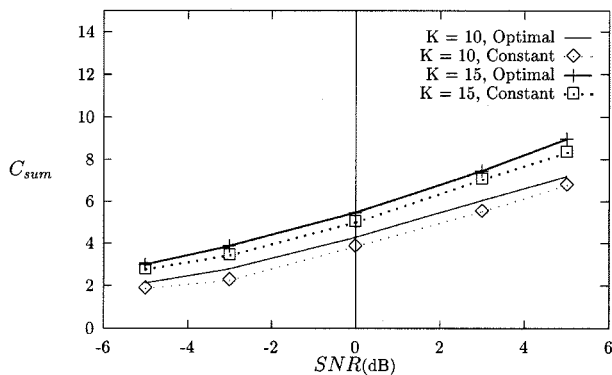


Fig. 6. Sum capacity in bits/s with five antennas at the receiver is plotted versus the SNR of the users with both the optimal and constant power allocations.

the gain that is obtained by feeding back the fast-fading component vanishes, and the water-filling power policy that utilizes only the slow-fading component performs just as well. In Fig. 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 which consider the case when the slow-fading component can be assumed to be constant in the time scale of communication. In this simulation, we assumed further that each component of $\tilde{\mathbf{h}}_i^f(n)$ 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 very minimal even when N is very small ($N = 5$ in this simulation example).

IX. CONCLUSION

The central problem addressed in this paper is the characterization of the optimum power allocation strategy in vector multiple-access frequency flat-fading channels. The common examples of such channels are uplinks of DS-CDMA channels and MACs with multiple antennas at the receiver. For concreteness, the discussion in this section will refer to the multiple-antenna system. The power allocation strategy considered is of the *centralized* type, i.e., each of the transmitters is assumed to be provided with perfect side information (regarding the channel fading states of *all* users). The sum capacity is our fundamental figure of merit to be optimized by appropriately allocating powers to the users as a function of the channel states subject to an average power constraint. While the structure of the optimal power allocation policies is very simple in the case of a single antenna, there is no known closed form to the optimal power allocation policies for finite number of users and multiple antennas. In fact, exactly this optimization problem [a finite-dimensional generalized version

of (9)] is considered in [25] 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 [34].

To get some insight into the nature of the optimal power allocation policy, we have considered in this paper the regime of large number of antennas and large number of users. In this regime, our main result is the identification of an asymptotically optimal simple *water-filling power policy*. This identification is very appealing in practice, due to its simplicity and the computation requirements to implement it are practically nil. Furthermore, the policy depends only on the slow-fading component and thus is robust to channel measurements and delay in feedback to the users. Also, fading statistics can be estimated and used to adaptively compute the threshold level of the water-filling strategy using the fixed-point iteration outlined in the paper. We have also shown that the gap in sum capacity between this water-filling policy and the optimal policy is of the order of \sqrt{N} for large N (uniformly for all ratios of users to antennas, SNR levels, and channel fading distributions). For the usual channel fading statistics such as Rayleigh fading, our simulation studies suggest that this gap is negligible even for very small values of N ($N = 5$ in our example). Another possible regime is to fix the number of antennas (to be N) and let the number of users grow large. In this scenario, it is clear that some of the N users will have almost orthogonal channel gains and the policy that allows only these almost noninterfering users to transmit will be far superior to a policy that lets all users transmit constant power at all times.

In this paper, in the context of the multiple-antenna model we have referred to the number of antennas as the *degrees of freedom*. Another way to have degrees of freedom is to communicate in a wide-band channel. Here the degrees of freedom come in number of frequency bins across which the fading channel can be considered to be independent. In this case, the sum capacity of the channel increases without bound as the bandwidth of the wide-band channel increases. Reference [22, Sec. 7] has considered the nature of optimal power control in scalar multiple-access frequency-selective fading channels and an analogous exercise can be carried out in our vector MAC case. In particular, in the multiple-antenna context and when the slow-fading component can be considered to be constant over the time scale of communication, the policy of constant power at all times equally spread over the entire bandwidth is asymptotically optimal (in the regime of large number of users and antennas).

In this paper, we have considered additive white Gaussian noise as modeling thermal noise and out of cell interference. In scenarios when the out-of-cell interference can be statistically measured, the appropriate model of the additive noise is a colored Gaussian process. The effect of colored noise on the sum capacity and the appropriate power control is studied in [31]. A 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 [22] 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.

APPENDIX A
PROOF OF PROPOSITION 3.1

We first recall a special case of the central result of [20] regarding the convergence of the empirical distribution of eigenvalues of random Hermitian matrices. Let G_N be the empirical distribution function of the eigenvalues of $\sum_{i=1}^K \mathbf{s}_i \mathbf{s}_i^t h_i g(h_i)$ (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

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

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 [20] that the support of G^* is bounded above by some K_s . Applying Theorem 1.1 and its corollary of [1] to our case, we obtain

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

Thus we have that

$$\begin{aligned} & \frac{1}{2N} \log \det \left(I + \sum_{i=1}^K \sigma^{-2} \mathbf{s}_i \mathbf{s}_i^t h_i g(h_i) \right) \\ &= \frac{1}{2} \int \log(1 + \lambda \sigma^{-2}) dG_N(\lambda) \\ & \xrightarrow{a.s.} \frac{1}{2} \int \log(1 + \lambda \sigma^{-2}) dG^*(\lambda) \\ & \stackrel{\text{def}}{=} \bar{C}_{\text{sum}}^g. \end{aligned} \quad (120)$$

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

$$\begin{aligned} & \frac{1}{2N} \log \det \left(I + \sum_{i=1}^K \sigma^{-2} \mathbf{s}_i \mathbf{s}_i^t h_i g(h_i) \right) \\ & \leq \frac{1}{2N} \sum_{i=1}^K \log(1 + \sigma^{-2} h_i g(h_i) \mathbf{s}_i^t \mathbf{s}_i) \quad (121) \\ & \leq \frac{1}{2N} \sum_{i=1}^K \log(1 + \sigma^{-2} K_g \mathbf{s}_i^t \mathbf{s}_i h_i) \\ & \leq \frac{K_g}{2N \sigma^2} \sum_{i=1}^K \mathbf{s}_i^t \mathbf{s}_i h_i \end{aligned} \quad (122)$$

where (121) follows from the Hadamard inequality and (122) follows from the fact that g is bounded above by K_g . Since

$$\frac{1}{N} \sum_{i=1}^K \mathbf{s}_i^t \mathbf{s}_i h_i \xrightarrow{a.s., L^1} \alpha \mathbb{E}[h_1]$$

the proposition follows from the pointwise convergence result above (120) and the dominated convergence theorem. \square

APPENDIX B
PROOF OF PROPOSITION 4.1

We have

$$\begin{aligned} \bar{C}_{\text{opt}}^{(N)} &= \bar{C}_{\text{sum}}(\mathcal{P}^*) \\ &\leq \frac{1}{2N} \sum_{i=1}^K \mathbb{E}[\log(1 + \sigma^{-2} \mathbf{s}_i^t \mathbf{s}_i h_i \mathcal{P}_i^*(h_1, \dots, h_K, S))] \\ &\leq \frac{1}{2N} \sum_{i=1}^K \max_{\mathcal{P}^{(i)} \in \mathcal{F}_2^{(N)}} \mathbb{E} \left[\log \left(1 + \sigma^{-2} \mathbf{s}_1^t \mathbf{s}_1 h_1 \mathcal{P}_i^{(i)} \right. \right. \\ & \quad \left. \left. \cdot (h_1, \dots, h_K, S) \right) \right] \\ &= \frac{K}{2N} \max_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \mathbb{E}[\log(1 + \sigma^{-2} \mathbf{s}_1^t \mathbf{s}_1 h_1 \mathcal{P}_1(h_1, \dots, h_K, S))] \end{aligned} \quad (123)$$

where the derivation of these inequalities is completely straightforward. Using Jensen inequality conditionally on h_1, \mathbf{s}_1 , we have from (123) that

$$\bar{C}_{\text{opt}}^{(N)} \leq \frac{\alpha}{2} \sup_{\mathcal{P} \in \mathcal{F}_2} \{ \mathbb{E}[\log(1 + \sigma^{-2} \mathbf{s}_1^t \mathbf{s}_1 h_1 \mathcal{P}(h_1 \mathbf{s}_1^t \mathbf{s}_1))] \} \quad (124)$$

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

$$\mathcal{F}_2 \stackrel{\text{def}}{=} \{ \mathcal{P}: h_1 \mathbf{s}_1^t \mathbf{s}_1 \mapsto \mathbb{R}_+, \mathbb{E}[\mathcal{P}(h_1 \mathbf{s}_1^t \mathbf{s}_1)] \leq \bar{p} \}.$$

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

$$\begin{aligned} & \mathbb{E}[\log(1 + \sigma^{-2} \mathbf{s}_1^t \mathbf{s}_1 h_1 \mathcal{P}(h_1 \mathbf{s}_1^t \mathbf{s}_1))] \\ & \leq \log 2 + \mathbb{E}[\log(2\sigma^{-2} h_1 \mathbf{s}_1^t \mathbf{s}_1 \mathcal{P}(h_1 \mathbf{s}_1^t \mathbf{s}_1))] \\ & \leq 2 \log 2 + \log(\mathbb{E}[\sigma^{-2} h_1 \mathbf{s}_1^t \mathbf{s}_1]) + \log(\mathbb{E}[\mathcal{P}(h_1 \mathbf{s}_1^t \mathbf{s}_1)]) \\ & \leq 2 \log 2 + \log(\sigma^{-2} \mathbb{E}[h_1] \bar{p}) \end{aligned} \quad (125)$$

where we used Jensen inequality in the derivation of the last but one step. Now combining (125) and (124) we have shown Proposition 4.1 by denoting $K_c = \log 2 + \frac{1}{2} \log(\sigma^{-2} \mathbb{E}[h_1] \bar{p})$. \square

APPENDIX C
PROOF OF PROPOSITION 4.4

We fix N throughout this proof. From Proposition 4.1 we know that $\bar{C}_{\text{opt}}^{(N)}$ is finite. Recall the definition of $\mathcal{F}_0^{(N)}$ from (34). Define functions f_0, \dots, f_K from $\mathcal{F}_0^{(N)}$ to the nonnegative reals as follows:

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

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

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

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

$\exists \lambda_0, \lambda_1, \dots, \lambda_K \geq 0$, not all zero such that

$$\sum_{j=0}^K \lambda_j f_j(\mathcal{P}) \geq 0 \quad \forall \mathcal{P} \in \mathcal{F}_0^{(N)}. \quad (126)$$

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 , (126) can be rewritten as

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

and hence

$$\bar{C}_{\text{opt}}^{(N)} \geq \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} \bar{C}_{\text{sum}}(\mathcal{P}) - \sum_{j=1}^K \lambda_j \frac{(\mathbb{E}[\mathcal{P}_j] - \bar{p})}{2N}. \quad (127)$$

By the symmetry among the users, it follows from (127) that for every permutation π we have

$$\bar{C}_{\text{opt}}^{(N)} \geq \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} \left\{ \bar{C}_{\text{sum}}(\mathcal{P}) - \sum_{j=1}^K \frac{\lambda_{\pi(j)}}{2N} (\mathbb{E}[\mathcal{P}_j(h_1, \dots, h_K, S)] - \bar{p}) \right\}. \quad (128)$$

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

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

is concave (the supremum of linear functionals). Defining $\lambda = \frac{1}{K} \sum_{i=1}^K \lambda_i$ and using in (128) the concavity of the map above we arrive at

$$\bar{C}_{\text{opt}}^{(N)} \geq \sup_{\{\mathcal{P}: \mathcal{P}_i \geq 0 \forall i=1, \dots, K\}} \left\{ \bar{C}_{\text{sum}}(\mathcal{P}) - \frac{\lambda}{2N} \sum_{i=1}^K (\mathbb{E}[\mathcal{P}_i(h_1, \dots, h_K, S)] - \bar{p}) \right\}. \quad (129)$$

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

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

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

$$\bar{C}_{\text{opt}}^{(N)} \leq \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} \left\{ \bar{C}_{\text{sum}}(\mathcal{P}) - \frac{\lambda}{2N} \sum_{j=1}^K (\mathbb{E}[\mathcal{P}_j] - \bar{p}) \right\}. \quad (130)$$

Combining (129) and (130) the proof is complete. It remains now to show (126). To see this, define

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

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

$$\sum_{j=0}^K \lambda_j \eta_j \geq a \quad \forall z \in C_1 \quad (131)$$

$$\sum_{j=0}^K \lambda_j \eta_j \leq a \quad \forall z \in C_2. \quad (132)$$

Now (132) implies that $a \geq 0$ and $\lambda_j \geq 0 \forall j = 0 \dots K$. Fix $\mathcal{P} \in \mathcal{F}_0^{(N)}$. Since $f_j(\mathcal{P})$ is finite for every $j = 0, \dots, K$ we have for every $\epsilon > 0$ that

$$(f_0(\mathcal{P}) + \epsilon, \dots, f_K(\mathcal{P}) + \epsilon) \in C_1$$

and substituting this in (131) we have

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

Since this is true for every arbitrary $\epsilon > 0$, we have shown (126). This completes the proof of Proposition 4.4. \square

APPENDIX D

PROOF OF PROPOSITION 4.5

Fix one realization of fading gains h_1, \dots, h_K and signature sequences S . Since the map C in (31) is concave, any tuple of powers (denoted by (p_1^*, \dots, p_K^*)) that maximizes

$$C(p_1, \dots, p_K) - \frac{\lambda}{2N} \sum_{i=1}^K (p_i - \bar{p}) \quad (133)$$

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

$$\begin{aligned} \frac{\partial C}{\partial p_i}(p_1^*, \dots, p_K^*) &= \lambda \\ p_i^* &= \left(\frac{1}{\lambda} - \frac{1}{\mathbf{s}_i^t \left(\sigma^2 I + \sum_{j \neq i} \mathbf{s}_j \mathbf{s}_j^t h_j p_j^* \right)^{-1} \mathbf{s}_i h_i} \right)^+ \\ &\quad \forall i = 1, \dots, K. \end{aligned} \quad (134)$$

The derivation of (135) from (134) is completely analogous to that of (24). If the realization h_1, \dots, h_K, S is such that $\{h_i \mathbf{s}_i \mathbf{s}_i^t, i = 1, \dots, K\}$ is a linearly independent set then C is strictly concave and the solution p_1^*, \dots, p_K^* in (135) 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^*, \dots, p_K^* at the realization h_1, \dots, h_K, S . If there are no point masses in the distribution F and in the common distribution of v_{ij} then with probability one we have $\{h_i \mathbf{s}_i \mathbf{s}_i^t, i = 1, \dots, K\}$ are linearly independent and C is strictly concave. In this case, the tuple (p_1^*, \dots, 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

$$\mathcal{P}_i^*(h_1, \dots, h_K, S) \stackrel{\text{def}}{=} p_i^*. \quad (136)$$

If there are point masses in F and the common distribution of v_{ij} such that there is a positive probability of $\{h_i \mathbf{s}_i \mathbf{s}_i^t, i = 1, \dots, K\}$ being linearly dependent, then on these realizations, the solution set (p_1^*, \dots, 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, \dots, h_K, S for each $i = 1, \dots, K$. More generally, we can appeal to general measurable selection theorems ([33] is a good review on these results; [33, Theorem 3.1] is relevant to our case) to select a measurable \mathcal{P}^* that satisfies the property (135) 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 (133), it follows from Proposition 4.4 that

$$\bar{C}_{\text{opt}}^{(N)} = L_\lambda(\mathcal{P}^*) = \sup_{\mathcal{P} \in \mathcal{F}_0^{(N)}} L_\lambda(\mathcal{P}) \quad (137)$$

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

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

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

$$L_\lambda(\mathcal{P}) < \bar{C}_{\text{opt}}^{(N)} \text{ for } \mathcal{P} \text{ not satisfying (135) on realizations of positive measure.} \quad (138)$$

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 (138). We now show the existence of such a power allocation policy.

Fix a realization h_1, \dots, h_K, S and consider p_1^*, \dots, p_K^* defined in (135). 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

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

for each $i = 1, \dots, K$. We used Proposition 4.3 in the observation that $\mathbb{E}[\mathcal{P}_i^*]$ cannot be less than \bar{p} for any $i = 1, \dots, K$. From (137) we conclude that if we can show that $\mathbb{E}[\mathcal{P}_1^*] = \bar{p}$, we have proved the claim of this proposition that

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

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 $0 < a < b$:

$$g: \mu \mapsto \mathbb{E}[\mathcal{P}_1^{*\mu}(h_1, \dots, h_K, S)] \text{ is continuous on } \mu \in [a, b]. \quad (139)$$

Suppose that this is true. Now $g(\mu) < \frac{1}{\mu}$ and thus as $\mu \rightarrow \infty$ we arrive at $g(\mu) \rightarrow 0$. 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 (139), given \bar{p} we have $\tilde{\lambda}$ such that $g(\tilde{\lambda}) = \bar{p}$. Observe that

$$\bar{C}_{\text{opt}}^{(N)} = \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} \bar{C}_{\text{sum}}(\mathcal{P}) \leq \sup_{\mathcal{P} \in \mathcal{F}_2^{(N)}} L_{\tilde{\lambda}}$$

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

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

$$\bar{C}_{\text{opt}}^{(N)} = \bar{C}_{\text{sum}}(\mathcal{P}^{*\tilde{\lambda}}). \quad (140)$$

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 (35) on realizations (of fading gains and signature sequences) with positive probability measure, we have $L_\lambda(\mathcal{P}) < L_\lambda(\mathcal{P}^*)$. Using (140) and Proposition 4.4 we arrive at $\lambda = \tilde{\lambda}$. It only remains to show the claim in (139). We only show this for the case when C is strictly concave for almost every realization of h_1, \dots, h_K, S . The extension to the general case when there are realizations of positive measure which lead to nonstrict concavity of C is not pursued here. Fix $0 < a < b$ and a realization of h_1, \dots, h_K, S . We first observe that the map

$$G: \mu \mapsto (p_1^{*\mu}, \dots, p_K^{*\mu})$$

$$\stackrel{\text{def}}{=} \arg \max_{p_i \geq 0, i=1 \dots K} C(p_1, \dots, p_K) - \frac{\mu}{2N} \sum_{i=1}^K (p_i - \bar{p})$$

is continuous for every realization of h_1, \dots, h_K, S such that C is strictly concave. For such realizations, G is invertible and we have [from (35)]

$$\mathbf{s}_j^t \left(\sigma^2 I + \sum_{i=1}^K h_i p_i^{*\mu} \mathbf{s}_i \mathbf{s}_i^t \right)^{-1} \mathbf{s}_j h_j = \mu \quad \forall j = 1 \dots K. \quad (141)$$

Fix $a \leq \mu \leq b$ and consider $\mu_n \rightarrow \mu$ in $[a, b]$ as $n \rightarrow \infty$. Observe that the image of $[a, b]$ under G in the positive orthant of \mathbb{R}^K is contained in the box $[0, a^{-1}] \times \dots \times [0, a^{-1}]$. Furthermore, the image is closed (using (141)) and thus compact. Now consider the sequence $\{(p_1^{*\mu_n}, \dots, p_K^{*\mu_n})\}_{n \geq 1}$ 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 $\mathbf{p}^{*\tilde{\mu}}$ such that $\mathbf{p}^{*\mu_{i_n}} \rightarrow \mathbf{p}^{*\tilde{\mu}}$. From the continuity of the inverse of G (using (141)) we arrive at $\mu_{i_n} \rightarrow \tilde{\mu}$. By hypothesis, $\mu_n \rightarrow \mu$ and thus $\mu = \tilde{\mu}$ allows us to conclude that $G(\mu_n) \rightarrow G(\mu)$ showing the continuity of G . Thus, for almost every realization, we have shown continuity of G . Fix $\epsilon > 0$ and by Egoroff's theorem ([15, Theorem 3.6.23]), we have uniform continuity of the map

$$\mu \mapsto (p_1^{*\mu}, \dots, p_K^{*\mu})$$

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

$$|\mathcal{P}^{*\mu_n} - \mathcal{P}^{*\mu}|(h_1, \dots, h_K, S) < \frac{\epsilon}{2}. \quad (142)$$

Then

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

where we used (142) in the last step and the fact for every $a \leq \mu \leq b$ that $\mathcal{P}^{*a} < a^{-1}$ in the second step. Since ϵ is arbitrary, we have completed the proof of (139). \square

APPENDIX E 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 (135) we know that any optimal power allocation has the following structure:

$$\begin{aligned} \mathcal{P}_i^*(h_1, \dots, h_K, S) \\ = \left(\frac{1}{\lambda^{(N)}} - \frac{1}{\mathbf{s}_i^t (\sigma^2 I + \sum_{j \neq i} \mathbf{s}_j \mathbf{s}_j^t h_j \mathcal{P}_j^*(h_1, \dots, h_K, S))^{-1} \mathbf{s}_i h_i} \right)^+ \\ \forall i = 1, \dots, K. \end{aligned}$$

Here the notation $\lambda^{(N)}$ emphasizes the dependence of (the Kuhn–Tucker coefficient) λ on N . Thus we have

$$\mathcal{P}_i^* \leq \frac{1}{\lambda^{(N)}} \text{ a.s.}$$

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

$$\bar{C}_{\text{sum}}(\bar{\mathcal{P}}(p)) \longrightarrow \bar{C}_{\text{sum}}^s(p) > 0, \quad \text{as } N \rightarrow \infty. \quad (143)$$

Using results about eigenvalues of large random matrices, we show a more general version of this result in Proposition 3.1 and $\bar{C}_{\text{sum}}^s(p)$ has an explicit expression (given in [28]). It also follows from this result that $\bar{C}_{\text{sum}}^s(p) \rightarrow \infty$ as $p \rightarrow \infty$. Some simple monotonicity properties of \bar{C}_{sum} and \bar{C}_{sum}^s are as follows:

$$\bar{C}_{\text{sum}}(\bar{\mathcal{P}}(p_1)) > \bar{C}_{\text{sum}}(\bar{\mathcal{P}}(p_2)) \text{ whenever } p_1 > p_2 \\ \text{for each fixed } N. \quad (144)$$

$$\bar{C}_{\text{sum}}^s(p_1) > \bar{C}_{\text{sum}}^s(p_2) \text{ whenever } p_1 > p_2. \quad (145)$$

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

$$\bar{C}_{\text{sum}}^s\left(\frac{1}{\bar{\mu}}\right) > \alpha(K_c + 0.5) \quad (146)$$

where K_c is equal to $\log 2 + \frac{1}{2} \log(\sigma^{-2} \mathbb{E}[h_1] \bar{p})$ defined in the proof of Proposition 4.1. Defining the function on the positive reals

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

we recognize from (130) that $g_N(\mu) \geq \bar{C}_{\text{opt}}^{(N)}, \forall \mu \geq 0$. By definition of $\lambda^{(N)}$, from (33) we conclude that

$$g_N\left(\lambda^{(N)}\right) = \min_{\mu \geq 0} g_N(\mu) = \bar{C}_{\text{opt}}^{(N)}. \quad (147)$$

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

$$\bar{C}_{\text{opt}}^{(i_n)} = g_{i_n}\left(\lambda^{(i_n)}\right) \geq \bar{C}_{\text{sum}}\left(\bar{\mathcal{P}}\left(\frac{1}{\lambda^{(i_n)}}\right)\right) - \frac{\alpha}{2}. \quad (148)$$

In (148), 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 IV-B]. Furthermore

$$\begin{aligned} \liminf_{n \rightarrow \infty} \bar{C}_{\text{sum}}\left(\bar{\mathcal{P}}\left(\frac{1}{\lambda^{(i_n)}}\right)\right) &\geq \liminf_{n \rightarrow \infty} \bar{C}_{\text{sum}}\left(\bar{\mathcal{P}}\left(\frac{1}{\bar{\mu}}\right)\right) \\ &= \bar{C}_{\text{sum}}^s\left(\frac{1}{\bar{\mu}}\right) \end{aligned} \quad (149)$$

where we used (143) and (144). Combining (148), (149), and (146) 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. \square

APPENDIX F PROOF OF LEMMA 4.9

The essential ingredients of the proof are all contained in [24, Lemmas 4.3 and 4.4] and we only indicate the key points of departure. In particular, a close study of [24, Lemmas 3.2, 4.3 and 4.4] reveals that the statement made as Lemma 4.9 in this paper is true for the situation when $h_i = 1$ and $g(\cdot) = 1$. Below, we hint at a broad outline of the generalization of the results in [24] to the general case here. We use $C_i, i = 2, 3, 4, 5$ to denote constants that are independent of N .

Let $\beta_i^{(N)} = \mathbf{s}_i^t Z_i^{-1} \mathbf{s}_i$ where (recall notation from (50))

$$Z_i = \left(\sigma^2 I + \sum_{l \neq i} s_l \mathbf{s}_l^t h_l g(h_l) \right).$$

Let

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

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

$$\mathbb{E} \left[\left(\beta_1^{(N)} - \beta_g^* \right)^2 \right] \leq \frac{C_2^2}{N}. \quad (150)$$

We show (150) by the following sequence of bounds:

$$\mathbb{E} \left[\left(\beta_1^{(N)} - \tilde{\beta}_1^{(N)} \right)^2 \right] \leq \frac{C_3}{N} \quad (151)$$

$$\text{Var} \left(\tilde{\beta}_1^{(N)} \right) \leq \frac{C_4}{N} \quad (152)$$

$$\left| \mathbb{E} \left[\tilde{\beta}_1^{(N)} \right] - \beta_g^* \right| \leq \frac{C_5}{N}. \quad (153)$$

We omit the sketch of the proofs of (151)–(153) in this paper for brevity. The details of these results can be found in [32]. \square

APPENDIX G
PROOF OF LEMMA 4.10

A key observation from the quadratic equation $\beta^*(\bar{p}, \alpha)$ satisfies (in (48)) is the following:

$$\mathcal{K}(\tilde{h}) \geq h \iff \left(\frac{\sigma^2}{\bar{p}h} + \alpha \right) \left(1 - F(\tilde{h}) - \int_{\tilde{h}}^{\infty} \frac{h}{h_0} dF(h_0) \right) \geq 1. \quad (154)$$

To see this, define

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

Now

$$\begin{aligned} \mathcal{K}(\tilde{h}) \geq h &\iff HM(\tilde{h}) \geq h + \tilde{p}h\beta^*(\tilde{p}, \tilde{\alpha}) \\ &\iff \frac{HM(\tilde{h}) - h}{h\tilde{p}} \geq \beta^*(\tilde{p}, \tilde{\alpha}) \end{aligned} \quad (155)$$

$$\iff \sigma^2 \frac{HM(\tilde{h}) - h}{h\tilde{p}} + \frac{\tilde{\alpha}\tilde{p} \frac{HM(\tilde{h}) - h}{h\tilde{p}}}{1 + \tilde{p} \frac{HM(\tilde{h}) - h}{h\tilde{p}}} \geq 1 \quad (156)$$

$$\begin{aligned} &\iff \sigma^2 \frac{HM(\tilde{h}) - h}{h\tilde{p}} + \tilde{\alpha} \frac{HM(\tilde{h}) - h}{HM(\tilde{h})} \geq 1 \\ &\iff \left(\frac{\sigma^2}{\bar{p}h} + \alpha \right) \left(1 - F(\tilde{h}) - \frac{h(1 - F(\tilde{h}))}{HM(\tilde{h})} \right) \geq 1 \end{aligned} \quad (157)$$

where (156) follows from (155) and (47). Now the claim in (154) follows directly from (157). The following statements now follow from the key observation (154):

$$\frac{h}{\mathcal{K}(\tilde{h})} \leq 1 \implies \frac{h}{\mathcal{K}(\hat{h})} \leq 1 \quad \forall h \geq \hat{h} \geq \tilde{h} \in \text{sup}(F) \quad (158)$$

$$\frac{h}{\mathcal{K}(\tilde{h})} \geq 1 \implies \frac{h}{\mathcal{K}(\hat{h})} \geq 1 \quad \forall \hat{h} \geq \tilde{h} \geq h \in \text{sup}(F) \quad (159)$$

$$\mathcal{K}(h) \longrightarrow 0, \quad \text{as } h \longrightarrow \infty \quad (160)$$

If $HM(0) > 0$ we have $\mathcal{K}(0) > 0$. Then it follows from (160) 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 (158) 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 (160) and recalling that \mathcal{K} is continuous) and

hence $h(n) \uparrow h_{\text{thr}}$ for some h_{thr} in the support of F and h_{thr} is a fixed point of \mathcal{K} . Furthermore, for $h \in (h_{\text{thr}}, \infty)$, it follows from (159) that

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

and hence $\mathcal{K}(h) < h$ for all $h \in (h_{\text{thr}}, \infty)$. 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 (154) we arrive at $\mathcal{K}(\tilde{h}) \geq \tilde{h}$ for some $\tilde{h} > 0$ if we show that

$$\int_{\tilde{h}}^{\infty} \left(\frac{1}{\tilde{h}} - \frac{1}{h_0} \right) dF(h_0) \rightarrow \infty, \quad \text{as } \tilde{h} \rightarrow 0. \quad (161)$$

Observe that the integrand in (161) is the water-filling power allocation policy in (15) and maximizes the single-user capacity in (14). Suppose

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

for some constant K_d . Then, we have

$$\begin{aligned} \bar{C}_{1 \text{ user}}(\tilde{h}) &= \frac{1}{2} \int_{\tilde{h}}^{\infty} \log \left(1 + h_0 \sigma^{-2} \left(\frac{1}{\tilde{h}} - \frac{1}{h_0} \right) \right) dF(h_0) \\ &\leq \log 2 + \frac{1}{2} \log(\sigma^{-2} \mathbb{E}[h_1] K_d) \quad \forall \tilde{h} > 0 \end{aligned} \quad (162)$$

where we used a technique similar to that used in the proof of Proposition 4.1 to derive the last step. Since $\bar{C}_{1 \text{ user}}$ 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 $\bar{C}_{1 \text{ user}}$ is achieved by the water-filling policy of the form $(\frac{1}{\tilde{h}} - \frac{1}{h})^+$, we have a contradiction to (162). Thus, there cannot be a uniform bound K_d and we have shown (161). 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} . \square

APPENDIX H
PROOF OF PROPOSITION 6.1

The proof is quite elementary. We first show (101). Recall the map \mathcal{K} in (62) 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}_α) is strictly increasing pointwise with increasing α . Furthermore, for each α , the map \mathcal{K}_α is continuous.

Consider the following claim:

$$\mathcal{K}_\alpha(h) \downarrow \mathcal{K}_0(h) \text{ uniformly in } h \in [0, \infty) \text{ as } \alpha \downarrow 0 \quad (163)$$

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

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

Observe that $\forall h \in [0, \infty)$

$$\begin{aligned}
& \mathcal{K}_\alpha(h) - \mathcal{K}_0(h) \\
&= \frac{HM(h)\tilde{p}(1-\sigma^2\beta^*(\tilde{p}\sigma^2, \tilde{\alpha}))}{(1+\tilde{p})(1+\sigma^2\tilde{p}\beta^*(\tilde{p}\sigma^2, \tilde{\alpha}))} \\
&\leq \frac{HM(h)\tilde{p}(1-\sigma^2\beta^*(\tilde{p}\sigma^2, \tilde{\alpha}))}{(1+\tilde{p})^2} \\
&= \frac{HM(h)\tilde{p}}{2(1+\tilde{p})^2} \\
&\quad \cdot \left(\tilde{\alpha} + \left(1 + \frac{1}{\tilde{p}}\right) - \sqrt{\left(1 + \frac{1}{\tilde{p}}\right)^2 + \tilde{\alpha} \left(\frac{2}{\tilde{p}} - 2 + \tilde{\alpha}\right)} \right) \quad (164) \\
&= \frac{\sigma^2(1-F(h))}{2\tilde{p}} \left(\frac{\tilde{p}}{1+\tilde{p}} \right)^2 \\
&\quad \cdot \left(\tilde{\alpha} + \left(1 + \frac{1}{\tilde{p}}\right) - \sqrt{\left(1 + \frac{1}{\tilde{p}}\right)^2 + \tilde{\alpha} \left(\frac{2}{\tilde{p}} - 2 + \tilde{\alpha}\right)} \right) \\
&\leq \frac{\sigma^2}{2\tilde{p}} \left(\frac{\tilde{p}}{1+\tilde{p}} \right)^2 \left(\tilde{\alpha} + \frac{\tilde{\alpha} \left(\frac{2}{\tilde{p}} - 2 + \tilde{\alpha}\right)^+}{\sqrt{\left(1 + \frac{1}{\tilde{p}}\right)^2 + \tilde{\alpha} \left(\frac{2}{\tilde{p}} - 2 + \tilde{\alpha}\right)}} \right) \quad (165) \\
&\leq \alpha \frac{\sigma^2}{2\tilde{p}} \left(\frac{\tilde{p}}{1+\tilde{p}} \right)^2 \left(1 + \left(\frac{2}{\tilde{p}} - 2 + \tilde{\alpha}\right)^+ \right) \quad (166)
\end{aligned}$$

where (164) is by definition of $\beta^*(\tilde{p}, \alpha)$ in (49). We used the simple inequality

$$a - \sqrt{a^2 - b} \leq \frac{(b)^+}{\sqrt{a^2 - b}} \quad \forall a > 0 \quad \text{and} \quad a^2 \geq b$$

to arrive at (165). It is straightforward now to verify that

$$K_u \stackrel{\text{def}}{=} \sup_{h \in [0, \infty)} \left(\frac{\tilde{p}}{1+\tilde{p}} \right)^2 \left(1 + \left(\frac{2}{\tilde{p}} - 2 + \tilde{\alpha}\right)^+ \right) < \infty \quad (167)$$

which combined with the final upper bound in (166) shows the claim in (163) that \mathcal{K}_α converges monotonically 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:

$$h_{\text{wf}} = \frac{HM(h_{\text{wf}})}{1 + \frac{\bar{p}\sigma^{-2}HM(h_{\text{wf}})}{1-F(h_{\text{wf}})}}. \quad (168)$$

From (168), 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}_α themselves decrease monotonically with decreasing α . Let $h_{\text{thr}}^{(\alpha)}$ denote the unique fixed point of the map \mathcal{K}_α . 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 $h_i(n) \uparrow h_{\text{thr}}^{(\alpha_i)}$ 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

$h_{\text{thr}}^{(\alpha_1)} \leq h_{\text{thr}}^{(\alpha_2)}$. Thus, $\{h_{\text{thr}}^{(\alpha)}\}_\alpha$ is a decreasing sequence as α is decreasing and converges to, say, h_0 . Now $\forall \alpha$

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

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

$$h_0 \geq \mathcal{K}_0(h_0). \quad (169)$$

Also, from (166) and (167), we have, for every α

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

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

$$\mathcal{K}_0(h_0) \geq h_0. \quad (170)$$

Now (169) and (170) show that $h_0 = \mathcal{K}_0(h_0)$ and thus $h_0 = h_{\text{wf}}$, the unique fixed point of \mathcal{K}_0 .

Following the definition of $\beta^*(\tilde{p}, \alpha)$ in (48), we have

$$\begin{aligned}
\beta^* \left(\frac{\tilde{p}HM \left(h_{\text{thr}}^{(\alpha)} \right)}{1-F \left(h_{\text{thr}}^{(\alpha)} \right)}, \alpha \left(1-F \left(h_{\text{thr}}^{(\alpha)} \right) \right) \right) &\longrightarrow \frac{1}{\sigma^2}, \\
&\text{as } \alpha \rightarrow 0. \quad (171)
\end{aligned}$$

Observing that for every α

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

we have that $h_{\text{thr}}^{(\alpha)}$ 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 (171), we have shown (101). An identical argument now shows (102). \square

APPENDIX I 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 (48) that, as $\sigma^2 \rightarrow 0$

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

This implies that [as in (163)], as $\sigma^2 \rightarrow 0$

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

As in the proof of Proposition 6.1, h_{thr} converges to the fixed point of \mathcal{K}_o as $\sigma^2 \rightarrow 0$, 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 $\mathcal{P}_i^{\text{wf}}(h_i) \xrightarrow{\text{a.s.}} \bar{p}$ as $\sigma^2 \rightarrow 0$ 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 water-filling strategy. It still remains to show that when $\alpha \leq 1$ the gain in sum capacity with water-filling strategy over constant power strategy goes to zero in high SNR. We follow the proof of Proposition 6.2. Fix $\alpha \leq 1$ and N the processing gain. From the limiting values of h_{thr} and β_{wf}^* we already have that $\mathcal{P}_i^{\text{wf}} \xrightarrow{\text{a.s.}} \bar{p}$ as $\sigma^2 \rightarrow 0$ for every $1 \leq i \leq K$. We establish a bound akin to (106) and appeal to the dominated convergence theorem concluding the proof

$$\frac{1}{2N} \left(\log \det \left(I + \sigma^{-2} \sum_{i=1}^K \mathbf{s}_i \mathbf{s}_i^t h_i \mathcal{P}_i^{\text{wf}} \right) \right) \leq \frac{1}{2N} \left(\log \det \left(I + \sigma^{-2} \sum_{i=1}^K \frac{\mathbf{s}_i \mathbf{s}_i^t h_i}{\beta_{\text{wf}}^* h_{\text{thr}}} \right) \right) \quad (172)$$

$$\leq \frac{1}{2N} \sum_{i=1}^K \log \left(1 + \frac{\mathbf{s}_i^t \mathbf{s}_i h_i}{\sigma^2 \beta_{\text{wf}}^* h_{\text{thr}}} \right) \quad (173)$$

$$\leq \frac{1}{2N \sigma^2 \beta_{\text{wf}}^* h_{\text{thr}}} \sum_{i=1}^K \mathbf{s}_i^t \mathbf{s}_i h_i$$

where we used the bound that

$$\mathcal{P}_i^{\text{wf}}(h_i) \leq \frac{1}{\beta_{\text{wf}}^* h_{\text{thr}}}$$

by definition of the water-filling strategy (78) in (172) and the Hadamard inequality [4, sec. 9.6.3] in (173). Analogous to the proof of Proposition 3.1, an application of the dominated convergence theorem completes the proof. \square

APPENDIX J PROOF OF THEOREM 8.1

The first case of $d^2(N)N = 1$ is exactly identical to the DS-CDMA setup and the result is the same as that in Theorem 5.3. The second case is very similar to the DS-CDMA setup at high SNR and thus the result follows the behavior of the water-filling policy in the DS-CDMA model at high SNR (formally analyzed in Proposition 6.3). Below we will sketch a proof of only the situation $\alpha \leq 1$. Any optimal $\mathcal{P}^*(h_1^s, \dots, h_K^s, S)$ has the structure shown in the equation at the bottom of the page and Theorem 4.7 is still valid

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

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 $\beta_i(\bar{\mathcal{P}})\bar{p}$ where

$$\beta_i(\bar{\mathcal{P}}) \stackrel{\text{def}}{=} \tilde{\mathbf{h}}_i^{f \dagger} \left(\frac{\sigma^2}{N} + \bar{p} \sum_{j \neq i} h_j^s \tilde{\mathbf{h}}_j^f \tilde{\mathbf{h}}_j^{f \dagger} \right)^{-1} \tilde{\mathbf{h}}_i^f.$$

Define the (Lagrangian) function

$$L: \mathcal{P} \mapsto \bar{C}_{\text{sum}}^A(\mathcal{P}) - \frac{1}{2N\bar{p}} \sum_{i=1}^K (\mathbb{E}[\mathcal{P}_i(h_1^s, \dots, h_K^s, S)] - \bar{p}), \quad \mathcal{P} \in \mathcal{F}_0^{(N)}.$$

As in the proof of Theorem 5.1

$$\begin{aligned} L(\mathcal{P}^*) - L(\bar{\mathcal{P}}) &\leq \frac{\alpha}{2} \mathbb{E} \left[\left(\frac{\beta_1(\bar{\mathcal{P}})}{1 + \beta_1(\bar{\mathcal{P}})\bar{p}} - \frac{1}{\bar{p}} \right) (\mathcal{P}_1^*(S) - \bar{p}) \right] \\ &\leq \frac{\alpha K_p}{2\bar{p}} \mathbb{E} \left[\frac{1}{1 + \beta_1(\bar{\mathcal{P}})\bar{p}} \right] \\ &\leq \frac{\alpha K_p}{2\bar{p}^2} \mathbb{E} \left[\frac{1}{\beta_1(\bar{\mathcal{P}})} \right]. \end{aligned} \quad (174)$$

Now

$$\begin{aligned} \beta_1(\bar{\mathcal{P}}) &= \tilde{\mathbf{h}}_1^{f \dagger} \left(\frac{\sigma^2}{d^2(N)N} + \bar{p} \sum_{j \neq 1} h_j^s \tilde{\mathbf{h}}_j^f \tilde{\mathbf{h}}_j^{f \dagger} \right)^{-1} \tilde{\mathbf{h}}_1^f \\ &\geq \tilde{\mathbf{h}}_1^{f \dagger} \left(\frac{\sigma^2}{d^2(N_1)N_1} + \bar{p} \sum_{j \neq 1} \tilde{\mathbf{h}}_j^f \tilde{\mathbf{h}}_j^{f \dagger} \right)^{-1} \tilde{\mathbf{h}}_1^f \\ &\quad \forall N \geq N_1 \\ \liminf_{N \rightarrow \infty} \beta_1(\bar{\mathcal{P}}) &\geq \beta_{N_1}^* \quad \forall N_1 \end{aligned} \quad (175)$$

where $\beta_{N_1}^*$ satisfies the fixed-point equation (by an appeal to Lemma 4.8)

$$\frac{\sigma^2}{d^2(N_1)N_1} \beta = 1 - \alpha \int_0^\infty \frac{\beta h g(h)}{1 + \beta h g(h)} dF(h).$$

It follows from this expression for $\beta_{N_1}^*$ and the hypothesis that $d^2(N)N \rightarrow \infty$ and $\alpha \leq 1$ that

$$\beta_{N_1}^* \uparrow \infty \text{ as } N_1 \uparrow \infty. \quad (176)$$

Using (175) and (176) in (174) we arrive at

$$\lim_{N \rightarrow \infty} L(\mathcal{P}^*) - L(\bar{\mathcal{P}}) = 0.$$

The observation that $L(\mathcal{P}^*) = \bar{C}_{\text{opt}}^A(N)$ and $L(\bar{\mathcal{P}}) = \bar{C}_{\text{sum}}^A(\bar{\mathcal{P}})$ completes the proof. \square

$$\mathcal{P}_i^*(h_1, \dots, h_K, S) = \left(\frac{1}{\lambda} - \frac{1}{\tilde{\mathbf{h}}_i^{f \dagger} \left(\frac{\sigma^2}{N} + \sum_{j \neq i} \tilde{\mathbf{h}}_j^f \tilde{\mathbf{h}}_j^{f \dagger} \mathcal{P}_j^*(h_1^s, \dots, h_K^s, S) \right)^{-1} \tilde{\mathbf{h}}_i^f} \right)^+$$

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," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.
- [2] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [3] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.
- [4] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [5] G. J. Foschini and Z. Miljanic, "Distributed autonomous wireless channel assignment algorithm with power control," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 420–429, Aug. 1995.
- [6] 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.
- [7] A. Goldsmith and P. Varaiya, "Capacity of fading channel with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1995.
- [8] S. Grandhi, R. Yates, and D. Goodman, "Resource allocation for cellular radio systems," *IEEE Trans. Vehic. Technol.*, vol. 46, pp. 581–588, Aug. 1997.
- [9] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous CDMA channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2832–2836, Nov. 1998.
- [10] S. V. Hanly, "An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity," *IEEE J. Select. Areas Commun. (Special Issue on the Fundamentals of Networking)*, vol. 13, Sept. 1995.
- [11] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Int. Conf. Communications*, Seattle, WA, June 1995.
- [12] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations of cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
- [13] R. Rapajic, M. Honig, and G. Woodward, "Multiuser decision-feedback detection: Performance bounds and adaptive algorithms," in *Proc. 1998 Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 34.
- [14] T. R. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1984.
- [15] H. L. Royden, *Real Analysis*, 3rd ed. New York: Macmillan, 1988.
- [16] M. Rupf and J. L. Massey, "Optimum sequence multisets for Synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1261–1266, July 1994.
- [17] S. Shamai (Shitz) and A. D. Wyner, "Information theoretic considerations for symmetric, cellular, multiple-access fading channels—Part I," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1877–1894, Nov. 1997.
- [18] S. Shamai (Shitz) and S. Verdú, "Capacity of CDMA fading channels," in *Information Theory Workshop*, Metsovo, Greece, June 27–July 1, 1999.
- [19] —, "Effect of frequency flat fading on spectral efficiency of DS-SS-CDMA with random spreading," *IEEE Trans. Inform. Theory*, submitted for publication.
- [20] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.
- [21] E. Telatar, "Capacity of multi-antenna Gaussian channels," Bell Labs, Tech. Memo., 1997.
- [22] D. Tse and S. Hanly, "Multi-access fading channels—Part I: Polymatroidal structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.
- [23] —, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.
- [24] D. Tse and O. Zeitouni, "Performance of linear multiuser receivers in random environments," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1384–1400, Jan. 2000.
- [25] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. Applic.*, vol. 19, no. 2, pp. 499–533, Apr. 1998.
- [26] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization and successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conf. Signals, Systems and Computers*, 1997.
- [27] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [28] S. Verdú and S. Shamai, "Spectral efficiency with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.
- [29] P. Viswanath, V. Anantharam, and D. Tse, "Optimal sequences, power control and capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1968–1983, Sept. 1999.
- [30] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1984–1991, Sept. 1999.
- [31] —, "Sum capacity of vector multiple access channels with colored noise," Univ. Calif. Berkely/ERL, Memo. M99/47.
- [32] P. Viswanath, "Capacity of vector multiple access channels," Ph.D. dissertation, Univ. Calif. Berkeley, 2000.
- [33] D. H. Wagner, "Survey of measurable selection theorems: An update," in *Proc. Conf. Oberwolfach, 1979 (Lecture Notes in Mathematics no. 794)*. Berlin, Germany: Springer, 1980, pp. 176–219.
- [34] S.-P. Wu, L. Vandenberghe, and S. Boyd, "MAXDET: Software for determinant maximization problems: Users guide, alpha version," Stanford Univ., Apr. 1996.