Uniform Power Allocation in MIMO Channels: A Game-Theoretic Approach

Daniel Pérez Palomar, *Student Member, IEEE*, John M. Cioffi, *Fellow, IEEE*, and Miguel Angel Lagunas, *Fellow, IEEE*

***Abstract*—When transmitting over multiple-input–multiple-output (MIMO) channels, there are additional degrees of freedom with respect to single-input–single-output (SISO) channels: the distribution of the available power over the transmit dimensions. If channel state information (CSI) is available, the optimum solution is well known and is based on diagonalizing the channel matrix and then distributing the power over the channel eigenmodes in a "water-filling" fashion. When CSI is not available at the transmitter, but the channel statistics are *a priori* known, an optimal fixed power allocation can be precomputed.**

This paper considers the case in which not even the channel statistics are available, obtaining a robust solution under channel uncertainty by formulating the problem within a game-theoretic framework. The payoff function of the game is the mutual information and the players are the transmitter and a malicious nature. The problem turns out to be the characterization of the capacity of a compound channel which is mathematically formulated as a maximin problem. The uniform power allocation is obtained as a robust solution (under a mild isotropy condition). The loss incurred by the uniform distribution is assessed using the duality gap concept from convex optimization theory. Interestingly, the robustness of the uniform power allocation also holds for the more general case of the multiple-access channel.

***Index Terms*—Capacity compound multiple-input–multiple-output (MIMO) channel, channel uncertainty, game theory, maximin, robust power allocation.**

I. INTRODUCTION

MULTIPLE-input–multiple-output (MIMO) channels arise from the use of multiple dimensions for transmission and reception. Many different scenarios can be modeled as MIMO systems such as wireless communication systems with multiple antennas at both ends of the link (spatial diversity), wireline communications when a bundle of twisted pair copper wires in digital subscriber lines (DSL) is treated as a whole, systems with polarization diversity, or simply when a time-dispersive or frequency-selective channel is properly modeled (e.g., discrete multitone (DMT) and orthogonal frequency-division multiplexing (OFDM)).

Recently, MIMO channels arising from the use of spatial diversity at both the transmitter and the receiver have attracted considerable attention [1], [2]. They have been shown to present a significant increase in capacity over single-input–single-output (SISO) systems because of the constituent parallel subchannels (also termed channel eigenmodes) existing within the MIMO channel.

When channel state information (CSI) is available, the optimal power allocation that achieves capacity is well known [3], [4], [1]. In such a case, capacity is achieved by adapting the transmitted signal to the specific channel realization. To be more specific, the transmit directions need to align with the right singular vectors of the channel. In this way, assuming that a proper rotation is performed at the receiver, the channel matrix is diagonalized and the set of constituent subchannels or eigenmodes is obtained. In addition, the available transmit power has to be optimally allocated over the eigenmodes in a "water-filling" or "water-pouring" fashion [5], [6], [1], [7].

Obtaining a channel estimate at the transmitter requires either a feedback channel or the application of the channel reciprocity property to previous receive channel measurements when the transmit and receive channels are sufficiently correlated such as when the same carrier frequency is used for transmission and reception (provided that the time variation of the channel is not too fast). In many cases, the channel estimate may become significantly inaccurate, mainly due to the time-varying nature of the channel. In fact, in many practical communication systems, the channel is assumed unknown at the transmitter. For those situations, it becomes necessary to utilize transmission techniques (and a transmit power allocation) independent of the actual channel realization.

When CSI at the transmitter (CSIT) is not available, but the channel statistics are *a priori* known, an optimal fixed power allocation (independent of the actual channel realization) can be precomputed. In [2], the capacity of a MIMO channel with no CSIT was obtained assuming a uniform power allocation over the transmit antennas. The choice of the uniform distribution was based on the symmetry of the problem, i.e., the fact that the fading between each transmit–receive pair of antennas was identically distributed and uncorrelated with the fading between any other pair of antennas (spatially uncorrelated channel). The optimality of the uniform power allocation in terms of ergodic capacity for the Gaussian distributed channel with independent and identically distributed (i.i.d.) entries was proved by Telatar

Manuscript received February 21, 2002; revised October 14, 2002. This work was supported in part by the Catalan Government (DURSI) 1999FI 00588, 2001SGR 00268; Samsung Advanced Institute of Technology; and the Spanish Government (CICYT) TIC2000-1025, TIC2001-2356, TIC2002-04594, FIT-070000-2000-649 (MEDEA + A105 UniLAN). The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003.

D. P. Palomar is with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), Campus Nord UPC, 08034 Barcelona, Spain (e-mail: daniel@gps.tsc.upc.es).

J. M. Cioffi is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: cioffi@stanford.edu).

M. A. Lagunas is with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), Campus Nord UPC, 08034 Barcelona, Spain, and with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Edifici Nexus I, 08034 Barcelona, Spain (e-mail: m.a.lagunas@cttc.es).

Communicated by D. N. C. Tse, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.813513

0018-9448/03$17.00 © 2003 IEEE

[1] based on the concavity of the logdet function and on the symmetry of the problem.[1] (Note that if the channel matrix entries are correlated, it is possible to improve upon the uniform power allocation by using some statistical knowledge of the channel, e.g., using a stochastic water-filling solution as proposed in [9].) This result was extended to the multiuser case in [10], [11]. The uniform power distribution has also been shown optimum for some particular cases of interest such as frequency-selective SISO channels [12], [13] and the dual case of flat time-varying SISO channels [14]. Interestingly, the uniform power allocation has also been found optimal in other completely different scenarios such as in noncoherent multiple-antenna channels in the high signal-to-noise ratio (SNR) regime (whenever there are more receive than transmit antennas and for a sufficiently long channel coherence time) [15].

This paper considers the case in which not even the channel statistics are known at the transmitter, obtaining, therefore, a robust power allocation under channel uncertainty. We formulate the problem within a game-theoretic framework [16], [17], in which the payoff function of the game is the mutual information and the players are the transmitter and a malicious nature (different types of games are considered, such as a strategic game, a Stackelberg game, and a mixed-strategy strategic game). Mathematically, this is formulated as a maximin problem that is known to lead to robust solutions [18]. Well-known examples of robust maximin and minimax formulations are universal source coding and universal portfolio [4], [19]. The problem turns out to be the characterization of the capacity of a compound vector Gaussian channel [20], [21]. The uniform power allocation is obtained as the solution of the game in terms of capacity (under the mild condition that the set of channels is isotropically unconstrained, meaning that the transmission "directions" are unconstrained). The results are easily extended to ergodic and outage capacities. The loss in terms of capacity of the robust power allocation with respect to the optimal one (adapted to the specific channel realization) is analyzed using the concept of duality gap arising in convex optimization theory [22], [23].

The robustness of the uniform power allocation from a maximin viewpoint also holds for the more interesting and general case of a multiple-access channel (MAC). In particular, the worst case rate region corresponding to the uniform power distribution is shown to contain the worst case rate region of any other possible power allocation strategy. In other words, the capacity region of the compound vector MAC is achieved when each of the users is using a uniform power allocation.

The paper is structured as follows. Section II introduces the signal model used throughout the paper. The game-theoretic formulation of the problem of robustness is given in Section III. Section IV contains the proof of the optimality of the uniform power allocation in terms of instantaneous capacity along with several illustrative examples and the extension to average and outage statistics of the capacity. Section V deals with the analysis of the loss of performance of the uniform power allocation. The MAC case is considered in Section VI. The final conclusions of the paper are summarized in Section VII.

[1]The proof can be extended to non-Gaussian distributions symmetric with respect to the origin [8].

The following notation is used. Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and light-face italics denote scalars. $\mathbb{R}^{n\times m}$ and $\mathbb{C}^{n\times m}$ denote the set of $n\times m$ matrices with real- and complex-valued entries, respectively. The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote transpose, complex conjugate, and Hermitian operations, respectively. The superscript $(\cdot)^\star$ denotes optimal (do not confuse with the complex conjugate operation). The Frobenius norm of matrix $\boldsymbol{X}$ is represented by $\|\boldsymbol{X}\|_F$ and its trace by $\mathrm{Tr}\,(\boldsymbol{X})$. The ith eigenvalue in decreasing order of matrix $\boldsymbol{X}$ is denoted by $\lambda_i(\boldsymbol{X})$ or $\lambda_{X,\,i}$ (similarly, $\lambda_{\max}(\boldsymbol{X})$ and $\lambda_{X,\,\max}$ denote the maximum eigenvalue). $\boldsymbol{I}_n$ denotes the $n\times n$ identity matrix (the dimension n can be left unspecified whenever it can be inferred from the context). Expressions $p_{\boldsymbol{X}}(\boldsymbol{X})$ and $p_{\boldsymbol{X}}$ denote the probability density function (pdf) of the (possibly matrix-valued) random variable $\boldsymbol{X}$ (the difference between a random variable and a realization of the random variable can always be inferred from the context and are therefore written in the same way). The expectation with respect to $p_{\boldsymbol{X}}(\boldsymbol{X})$ is written as $\mathcal{E}_{p_{\boldsymbol{X}}}[\cdot]$ or simply as $\mathcal{E}[\cdot]$.

II. Signal Model

This paper considers the transmission of a vector signal through a linear MIMO channel immersed in interference and noise. To be more specific, the general vector model used throughout the paper is

$$\boldsymbol{y} = \boldsymbol{H}_c\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{C}^{n_T\times 1}$ is the transmitted vector, $\boldsymbol{H}_c \in \mathbb{C}^{n_R\times n_T}$ is the physical channel matrix that performs a linear transformation on $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{C}^{n_R\times 1}$ is the received signal vector, and $\boldsymbol{n} \in \mathbb{C}^{n_R\times 1}$ is the interference-plus-noise vector with arbitrary covariance matrix $\boldsymbol{R}_n$. (It is assumed without loss of generality that both $\boldsymbol{x}$ and $\boldsymbol{n}$ have zero mean.) As will be argued in Section III, it suffices to consider that $\boldsymbol{n}$ is a proper complex Gaussian random vector [24], i.e., $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0},\, \boldsymbol{R}_n)$. The channel transition probability $p(\boldsymbol{y}\mid\boldsymbol{x})$ is then given by a vector Gaussian distribution parameterized by the channel state $(\boldsymbol{H}_c,\, \boldsymbol{R}_n)$. Recall that the model of (1) has n_T transmit and n_R receive (finite) dimensions. The transmitter is assumed to be constrained in its average power (long-term power constraint [14], [25])

$$\mathcal{E}[\|\boldsymbol{x}\|^2] \le P_T \tag{2}$$

or, equivalently

$$\mathrm{Tr}\,(\boldsymbol{Q}) \le P_T \tag{3}$$

where $\boldsymbol{Q} = \mathcal{E}[\boldsymbol{x}\boldsymbol{x}^H]$ is the covariance matrix of the transmitted vector and P_T is the maximum average transmitted power per transmission. By uniform power allocation we mean $\boldsymbol{Q} = P_T/n_T\, \boldsymbol{I}_{n_T}$ which also implies an independent signaling over the transmit dimensions if a Gaussian code is used. Another interesting constraint is the maximum eigenvalue constraint $\lambda_{\max}(\boldsymbol{Q}) \le \alpha$ (as has been used elsewhere [26]) which will be revisited in Section IV. Note that $\lambda_{\max}(\boldsymbol{Q})$ is an

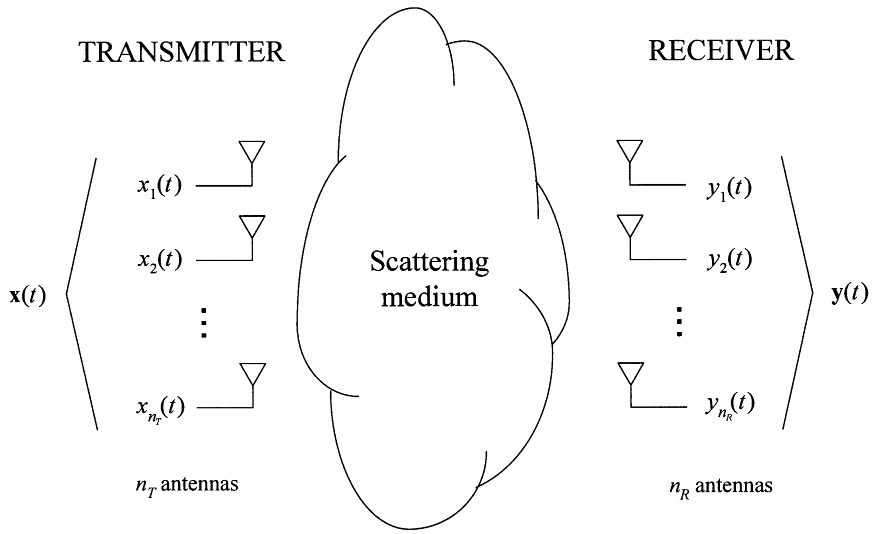

Fig. 1. Example of a MIMO channel arising in wireless communications when multiple antennas are used at both the transmitter and the receiver.

upper bound on the average transmitted power at each transmit dimension $\mathcal{E}[|x_i|^2] \leq \lambda_{\max}(\boldsymbol{Q})$.

For illustration purposes, let us see how the following particular cases fit into the general model of (1): i) a flat channel with multiple antennas at both the transmitter and the receiver (see Fig. 1) fits naturally into the model by letting $[\boldsymbol{H}_c]_{ij}$ represent the fading between the jth transmit antenna and the ith receive one, ii) a frequency-selective SISO channel in time domain can be accommodated by properly choosing $\boldsymbol{H}_c$ as a convolution matrix, and iii) a frequency-selective SISO channel in frequency domain (or multicarrier channel such as an OFDM channel) can also be cast in the general vector model of (1) by choosing $\boldsymbol{H}_c$ diagonal with its ith diagonal element denoting the gain of the ith carrier (of course, the introduction of a cyclic prefix between transmitted blocks is necessary to obtain such a model).

The mutual information between the transmitted and the received signals for a given channel state $\boldsymbol{H}_c$ according to the signal model of (1) (for a given $\boldsymbol{Q}$) is [4], [1]

$$\begin{aligned}\mathcal{I}(\boldsymbol{x};\boldsymbol{y}\mid\boldsymbol{H}_c) &= \mathcal{H}(\boldsymbol{y}\mid\boldsymbol{H}_c) - \mathcal{H}(\boldsymbol{y}\mid\boldsymbol{x},\boldsymbol{H}_c)\\ &\leq \log\det\left(\boldsymbol{I}_{n_R} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \qquad (4)\end{aligned}$$

where $\boldsymbol{H} \triangleq \boldsymbol{R}_n^{-1/2}\boldsymbol{H}_c$ is the whitened channel state. The upper bound is achieved when $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{Q})$ (i.e., a Gaussian code) [4], [1]. (The base of the logarithm will be left unspecified throughout the paper unless otherwise stated.) The channel capacity (assuming $\boldsymbol{H}$ known) over all $\boldsymbol{Q}$ verifying the power constraint of (3) is

$$\mathcal{C}(\boldsymbol{H}) = \max_{\substack{\boldsymbol{Q}:\,\mathrm{Tr}(\boldsymbol{Q})\leq P_T,\\ \boldsymbol{Q}=\boldsymbol{Q}^H\geq\boldsymbol{0}}} \log\det\left(\boldsymbol{I}_{n_R} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \qquad (5)$$

and the capacity-achieving solution is given by a transmit covariance matrix $\boldsymbol{Q}$ that diagonalizes the channel matrix $\boldsymbol{H}$ and distributes the available transmit power among the eigenmodes in a water-filling fashion [3], [4], [1], [6]. We define the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ explicitly as a function of $\boldsymbol{Q}$ and $\boldsymbol{H}$ as in [1]

$$\begin{aligned}\Psi(\boldsymbol{Q},\boldsymbol{H}) &= \log\det\left(\boldsymbol{I}_{n_R} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \qquad (6)\\ &= \log\det\left(\boldsymbol{I}_{n_T} + \boldsymbol{Q}\boldsymbol{H}^H\boldsymbol{H}\right) \qquad (7)\end{aligned}$$

where the determinant identity $\det(\boldsymbol{I}+\boldsymbol{AB}) = \det(\boldsymbol{I}+\boldsymbol{BA})$ has been used.

In wireless communications, the channel may undergo slow and/or fast fading due to shadowing and Doppler effects. Essentially, matrix $\boldsymbol{H}$ is not fixed and changes in time. One possible way to deal with this is by considering the channel as a random variable with a known pdf $p_{\boldsymbol{H}}(\boldsymbol{H})$ which naturally leads to the notions of ergodic capacity and outage capacity [1], [27] (c.f. Section IV-D). In this paper, we are interested in a robust design obtained by including uncertainty about the channel at both the transmitter and the receiver. There is a significant variety of channel models that can be used to model channel uncertainty (see [20] for a great overview of reliable communication under channel uncertainty). If the fading is sufficiently slow (the channel coherence time is much higher than the duration of a transmission[2]), the system can be modeled as a compound channel, where the channel state remains unchanged during the course of a transmission and it is assumed to belong to a set of possible channel states but otherwise unknown [21], [28], [20], [27] (the capacity of the compound vector Gaussian channel was obtained in [29]). For fast fading, however, the compound channel is no longer appropriate and other models, such as a compound finite-state channel (FSC) [20] or an arbitrarily varying channel (AVC) [21], [28], [20] may be necessary. In the AVC, the channel state can arbitrarily change from symbol to symbol during the course of a transmission (see [30] for results on the vector Gaussian AVC). Recall that, in situations where the unknown channel remains unchanged over multiple transmissions, the utilization of a training sequence to estimate the channel at the receiver is particularly attractive. The reader is referred to [20] for a detailed discussion on the applicability of each model.

We consider that the fading is slow enough so that the compound channel model is valid (see [29], for example, where

[2]By "transmission" we mean the transmission of a codeword of block length n, i.e., "n uses of the channel."

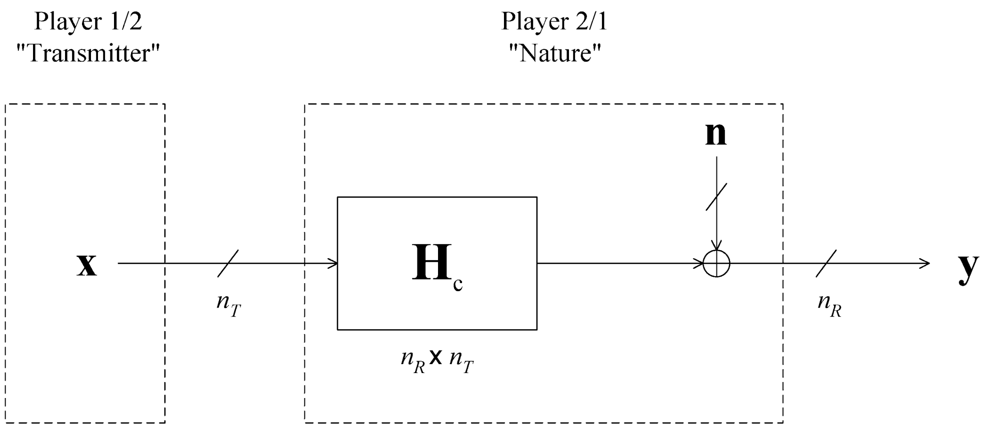

Fig. 2. Communication interpreted as a two-player game.

the compound channel was used to model a wireless MIMO system). In other words, we assume that the transmission duration is sufficiently long so that the information-theoretic coding arguments are valid and sufficiently short so that the channel remains effectively unchanged during a transmission (c.f. [2], [25]). This type of channel is usually referred to as block-fading channel [27], [25].

III. Game-Theoretic Formulation

In this section, the problem of obtaining a robust transmit power allocation when the transmitter does not even know the channel statistics is formulated within the framework of game theory [16], [17]. The idea of robustness implies being able to function in all possible scenarios and, in particular, the worst case scenario. This concept fits naturally into the context of game theory.

We will consider a game in which the payoff function (by which the result of the game is measured) is the mutual information and the players are: the transmitter that selects the best signaling scheme $p(\boldsymbol{x})$ and a malicious nature that chooses the worst communication conditions or channel transition probability $p(\boldsymbol{y} \mid \boldsymbol{x})$. It is interesting to note that the formulation of the communication process explicitly as a game was first proposed more than 40 years ago by Blachman [31] using a mutual information payoff. We constrain our search to Gaussian-distributed signal and noise since it is well known that they constitute a robust solution (a saddle point) to a mutual information game for the memoryless vector channel [32], [33].[3] In this case, $p(\boldsymbol{y} \mid \boldsymbol{x})$ is a vector Gaussian distribution parameterized with the channel state $(\boldsymbol{H}_c, \boldsymbol{R}_n)$ as described in Section II. In the sequel, by "channel" we will simply refer to the whitened channel state $\boldsymbol{H}$ and not to the channel transition probability $p(\boldsymbol{y} \mid \boldsymbol{x})$. The two-player game is illustrated in Fig. 2.

With the previous considerations, the unknowns of the game are the transmit covariance matrix $\boldsymbol{Q}$ and the whitened channel $\boldsymbol{H}$ (which implicitly includes the noise covariance matrix $\boldsymbol{R}_n$ and the original channel $\boldsymbol{H}_c$). The payoff function of the game is then the mutual information given by $\Psi(\boldsymbol{Q}, \boldsymbol{H})$ in (6) or (7). The game would be meaningless and trivial unless we placed restrictions on the players. Therefore, we suppose that the channel $\boldsymbol{H}$ must belong to a set of possible channels $\mathcal{H}$ and, similarly, $\boldsymbol{Q}$ must belong to a set of possible covariance matrices $\mathcal{Q}$. It is important to bear in mind that, for simplicity of notation, we write $\boldsymbol{H} \in \mathcal{H}$ instead of $(\boldsymbol{H}_c, \boldsymbol{R}_n) \in \mathcal{H}_c \times \mathcal{R}_n$ with no loss of generality (one can always define $\mathcal{H}$ as the set of matrices $\boldsymbol{H}$ that can be parameterized as $\boldsymbol{H} = \boldsymbol{R}_n^{-1/2}\boldsymbol{H}_c$ for some $(\boldsymbol{H}_c, \boldsymbol{R}_n) \in \mathcal{H}_c \times \mathcal{R}_n$). The set $\mathcal{Q}$ considered in this paper is defined by the average transmit power constraint of (3)

$$\mathcal{Q} \triangleq \left\{\boldsymbol{Q}\colon \mathrm{Tr}(\boldsymbol{Q}) \leq P_T,\, \boldsymbol{Q} = \boldsymbol{Q}^H \geq \boldsymbol{0}\right\}. \tag{8}$$

We remark that the results of the paper still hold if the eigenvalue constraint $\lambda_{\max}(\boldsymbol{Q}) \leq \alpha$ is utilized instead to define $\mathcal{Q}$. Regarding the set $\mathcal{H}$, since we are interested in finding a robust $\boldsymbol{Q}$ for all possible channels, we would like not to impose any constraint on the allowable set of channels. However, this would be a poor choice because the trivial solution $\boldsymbol{H} = \boldsymbol{0}$ would be obtained. To avoid this effect, we are forced to introduce some artificial constraints (unlike the constraint used to define $\mathcal{Q}$ which is very natural). But this may have the side effect that the solution to the game formulation may depend on the particular constraints chosen. Fortunately, as proved in Section IV, the solution to the game formulation is independent of the particular channel constraints under the mild condition that the constraints guarantee an isotropy property in $\mathcal{H}$ (c.f. Section IV).

As has been previously argued, to take into account the effect of channel uncertainty, we consider that the channel is known to belong to a set of possible channels $\mathcal{H}$ but otherwise unknown. The worst case channel for a given $\boldsymbol{Q}$ is given by the minimizing solution to $\inf_{\boldsymbol{H}\in\mathcal{H}} \Psi(\boldsymbol{Q}, \boldsymbol{H})$. The transmitter will maximize the worst case mutual information over the set $\mathcal{Q}$, yielding the following maximin formulation of the problem:[4]

$$\sup_{\boldsymbol{Q}\in\mathcal{Q}} \inf_{\boldsymbol{H}\in\mathcal{H}} \Psi(\boldsymbol{Q}, \boldsymbol{H}). \tag{9}$$

At this point, it is interesting to recall that a compound channel is precisely a channel that is known to belong to a set of possible channels (unchanged during the course of a transmission) but otherwise unknown [21], [28], [20]. As discussed

[3]For complex-valued signals, the saddle-point property holds for proper complex Gaussian distributions [24].

[4]For the particular sets $\mathcal{Q}$ and $\mathcal{H}$ considered in this paper, the formulation sup-inf reduces to max-min. For the sake of generality, however, we stick to the sup-inf notation throughout the paper.

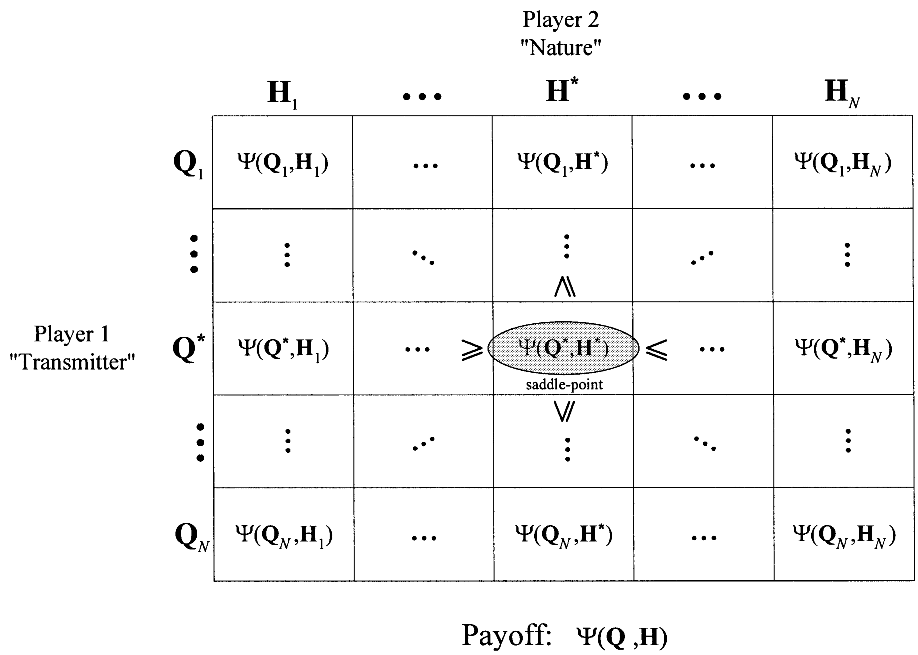

Fig. 3. Two-player zero-sum strategic game in which player 1 (the transmitter) and player 2 (nature) move simultaneously. The optimal power allocation is found as a saddle point (Nash equilibrium). (Note that for illustration purposes the sets $\mathcal{Q}$ and $\mathcal{H}$ have been considered finite.)

in Section II, this type of channel may be useful to model communication under channel uncertainty for sufficiently slow fading. The capacity of the compound channel (the capacity that can be guaranteed for the set of possible channels $\mathcal{H}$) was extensively treated in [21] where an expression similar to (9) was shown to be the capacity of the compound discrete memoryless channel. In [29], the vector Gaussian channel was specifically considered and (9) was indeed shown to be the capacity of the compound vector Gaussian channel when the actual channel state is unknown at both the transmitter and the receiver (under the mild assumption that $\mathcal{H}$ is bounded). Note that knowledge of the channel state at the receiver does not increase the compound channel capacity [21],[5] although the receiver may be simpler to implement with this knowledge. Clearly, the capacity of the compound channel cannot exceed the capacity of any channel in the family. In principle, it may not even be equal to the infimum of the capacities of the individual channels in the family (this is because codes and their decoding sets must be found, not just to give small error probability in the worst channel, but uniformly across the class of channels, which is a more stringent condition) [29], [20].

Alternatively, we can consider the compound channel when the transmitter knows the channel state (as in the previous case, it is indifferent whether the receiver knows the channel state or not [21]). In this case, in principle, a different coding–decoding strategy can be used for each channel realization and the capacity of such a compound channel is given by the following minimax formulation:

$$\inf_{\boldsymbol{H}\in\mathcal{H}} \sup_{\boldsymbol{Q}\in\mathcal{Q}} \Psi\left(\boldsymbol{Q},\, \boldsymbol{H}\right) \tag{10}$$

[5]The intuitive explanation of this effect is that, since the channel state remains fixed for the transmission of the whole codeword, for sufficiently long codes, it can be estimated at the receiver by transmitting, for example, a training sequence with length proportional to $\sqrt{n}$ at no cost of rate as $n \rightarrow \infty$ [21]. In fact, the channel state is not at all required by universal decoders [20].

i.e., the infimum of the capacities of the family of channels $\mathcal{H}$.

From a game-theoretic perspective, the problem can be viewed as a two-player zero-sum (players with diametrically opposed preferences) game, also known as strictly competitive game (the transmitter is the maximizing player and nature is the minimizing player) [16] (see Fig. 2). In the following, we cast the problem in three different types of games: a strategic game both with pure strategies and with mixed strategies and a Stackelberg game.

The simplest formulation (from a game-theoretic standpoint) is that of a strategic game, in which the players select their strategies without knowing the other players' choices, i.e., they "move" simultaneously (see Fig. 3). In such cases, there may exist a set of equilibrium points called Nash equilibria characterized for being robust or locally optimal in the sense that no player wants to deviate from such points. In our case (a two-player zero-sum game), a Nash equilibrium is also termed saddle point $(\boldsymbol{Q}^\star,\, \boldsymbol{H}^\star)$ and it is a simultaneously optimal point for both players (see Fig. 3)

$$\Psi\left(\boldsymbol{Q},\, \boldsymbol{H}^\star\right) \leq \Psi\left(\boldsymbol{Q}^\star,\, \boldsymbol{H}^\star\right) \leq \Psi\left(\boldsymbol{Q}^\star,\, \boldsymbol{H}\right) \tag{11}$$

where $\Psi\left(\boldsymbol{Q}^\star,\, \boldsymbol{H}^\star\right)$ is called the value of the game (whenever it exists) and is equal to the maximin and minimax solutions of (9) and (10) [16], i.e.,

$$\Psi\left(\boldsymbol{Q}^\star,\, \boldsymbol{H}^\star\right) = \sup_{\boldsymbol{Q}\in\mathcal{Q}} \inf_{\boldsymbol{H}\in\mathcal{H}} \Psi\left(\boldsymbol{Q},\, \boldsymbol{H}\right) = \inf_{\boldsymbol{H}\in\mathcal{H}} \sup_{\boldsymbol{Q}\in\mathcal{Q}} \Psi\left(\boldsymbol{Q},\, \boldsymbol{H}\right). \tag{12}$$

Note that one of the major techniques for designing systems that are robust with respect to modeling uncertainties is the minimax approach, in which the goal is the optimization of the worst case performance [34], [18]. Interesting examples of minimax design in information theory are the problem of source coding or data compression when the data distribution is completely unknown and the problem of portfolio investment when nothing is known about the stock market [4], [19]. Both problems can be

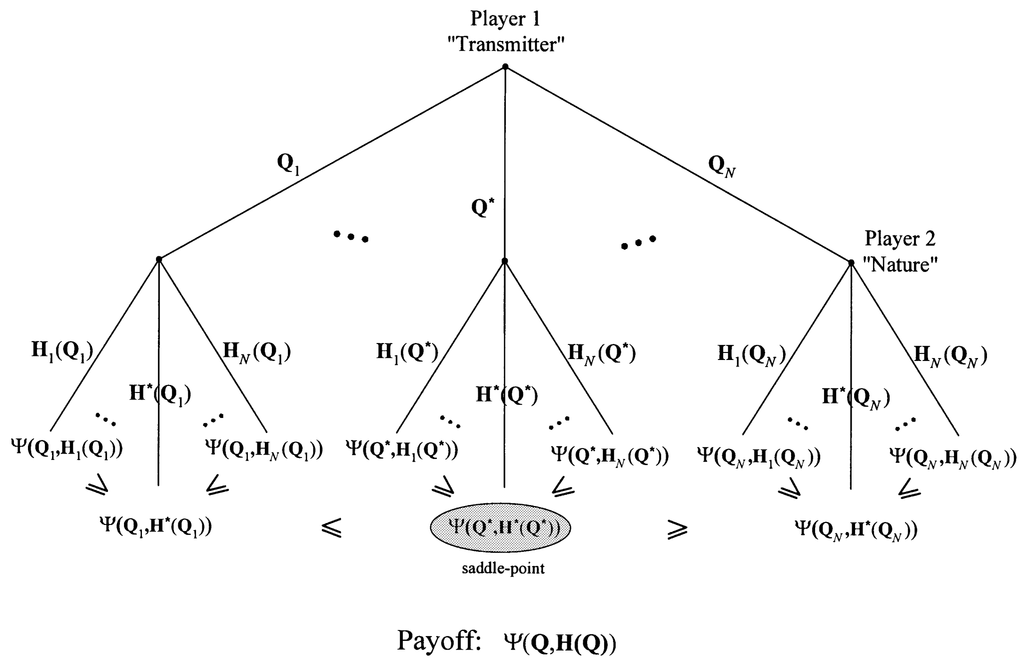

Fig. 4. Two-player zero-sum extensive game in which player 1 (the transmitter) moves first and then player 2 (nature) moves aware of player 1's move, i.e., Stackelberg game. The optimal power allocation is found as a saddle point (subgame perfect equilibrium which is also a Nash equilibrium). (Note that for illustration purposes the sets $\mathcal{Q}$ and $\mathcal{H}$ have been considered finite.)

formulated as a game in which two players compete: the source-encoding scheme versus the data distribution and the portfolio investor versus the market.

In our case, the function $\Psi(\boldsymbol{Q}, \boldsymbol{H})$ may or may not have any saddle point depending on the particular set $\mathcal{H}$ (c.f. Section IV). However, so far, we have only considered pure strategies, i.e., strategies given by a single fixed (deterministic) pair $(\boldsymbol{Q}, \boldsymbol{H})$. The game can be extended to include mixed strategies, i.e., the possibility of choosing a randomization over a set of pure strategies (the randomizations of the different players is independent) [16]. In this case, the payoff is the average of $\Psi(\boldsymbol{Q}, \boldsymbol{H})$ over the mixed strategies $\mathcal{E}_{p_{\boldsymbol{Q}} p_{\boldsymbol{H}}} \Psi(\boldsymbol{Q}, \boldsymbol{H})$ and the saddle point is similarly defined as

$$\mathcal{E}_{p_{\boldsymbol{Q}} p^{\star}_{\boldsymbol{H}}} \Psi(\boldsymbol{Q}, \boldsymbol{H}) \leq \mathcal{E}_{p^{\star}_{\boldsymbol{Q}} p^{\star}_{\boldsymbol{H}}} \Psi(\boldsymbol{Q}, \boldsymbol{H}) \leq \mathcal{E}_{p^{\star}_{\boldsymbol{Q}} p_{\boldsymbol{H}}} \Psi(\boldsymbol{Q}, \boldsymbol{H}). \tag{13}$$

It is well known that a strategic game always has a mixed strategy Nash equilibrium under the assumption that each set of pure strategies is closed, bounded, and convex [16]. In fact, for our specific problem, even if we allow more general sets (which need not be closed, bounded, and convex) such as the set $\mathcal{H}$ defined by $\lambda_{\max}(\boldsymbol{H}^H \boldsymbol{H}) \geq \beta$ (which is nonconvex and unbounded), it can be shown that the problem always has an infinite set of Nash equilibria (c.f. Section IV). One can interpret mixed strategies in different ways. In this problem, perhaps, the most relevant interpretation is to consider the mixed strategy Nash equilibrium as a steady state of an environment in which players act repeatedly, learning other players' mixed strategies (see [16] for other interpretations).

Alternatively, instead of modeling our problem as a strategic game (which, in general, does not have a pure strategy Nash equilibrium), we can formulate it in a more general way as an extensive game[6] in which the selected strategy of a user may depend on the previously selected strategy of another user[7] (as opposed to the previous strategic interpretation in which both players move simultaneously) [16], [17]. For the specific case of a two-player zero-sum game, in the parlance of game theory, such an extensive game is called Stackelberg game [16], [17]. Consider the case in which the transmitter moves first and then nature moves aware of the transmitter's move (see Fig. 4). In such a case, the maximin solution of (9) is always a pure strategy Nash equilibrium. In fact, such a solution is a subgame perfect equilibrium (called in this case Stackelberg equilibrium) which is a more refined definition of equilibrium[8] [16], [17]. In this case, a saddle point $(\boldsymbol{Q}^{\star}, \boldsymbol{H}^{\star}(\boldsymbol{Q}^{\star}))$ is characterized by

$$\Psi(\boldsymbol{Q}, \boldsymbol{H}^{\star}(\boldsymbol{Q})) \leq \Psi(\boldsymbol{Q}^{\star}, \boldsymbol{H}^{\star}(\boldsymbol{Q}^{\star})) \leq \Psi(\boldsymbol{Q}^{\star}, \boldsymbol{H}(\boldsymbol{Q}^{\star})). \tag{14}$$

Similarly, we can also consider the opposite formulation of the Stackelberg game in which nature moves first and then the transmitter moves aware of nature's move with the saddle point given by

$$\Psi(\boldsymbol{Q}(\boldsymbol{H}^{\star}), \boldsymbol{H}^{\star}) \leq \Psi(\boldsymbol{Q}^{\star}(\boldsymbol{H}^{\star}), \boldsymbol{H}^{\star}) \leq \Psi(\boldsymbol{Q}^{\star}(\boldsymbol{H}), \boldsymbol{H}). \tag{15}$$

Note that the saddle points of (14) and (15) are always satisfied by the solutions to problems (9) and (10), respectively.

A significant part of the literature that has modeled communication as a game has dealt with the characterization of the saddle points satisfying (11), i.e., implicitly adopting a

[6]An extensive game is an explicit description of the sequential structure of the decision problems encountered by the players in a strategic situation [16].

[7]By extensive game we always refer to those with perfect information (imperfect information can also be considered) [16].

[8]The solution concept of Nash equilibrium is unsatisfactory in extensive games since it ignores the sequential structure of the decision problem; as a consequence, more refined definitions of equilibrium have been proposed [16].

formulation of the problem as a strategic game. Reference [31] is one of the earliest papers dealing with such a problem using a mutual information payoff. (Note that other payoff functions have also been considered, such as the mean-square error in [35] to deal with communication over a channel with an intelligent jammer.) A two-player zero-sum game was explicitly adopted in [32] obtaining the Gaussian distribution as a saddle point. In [36], m-dimensional strategies were considered in a game-theoretic formulation of communication over channels with block memory, where it was found that memoryless jamming and memoryless coding constitute a saddle point. In [37], a two-player zero-sum game was explicitly formulated for communication in the presence of jamming using a power constraint for both players. In [33], communication under the worst additive noise under a covariance constraint was analyzed (the Gaussian distribution was obtained as a saddle-point solution) with emphasis on covariances satisfying correlation constraints at different lags. The vector Gaussian AVC was considered in [30] obtaining a saddle point given by a water-filling solution for the jammer and for the coder. In [38] , the maximin and minimax problems of (9) and (10) in a multiantenna wireless scenario were solved for a specific set of channels $\mathcal{H}$ defined by $\mathrm{Tr}(\boldsymbol{H}^H\boldsymbol{H}) \geq \beta$, i.e., the two Stackelberg games previously formulated were implicitly considered.

The rest of the paper focuses mainly on finding a robust power allocation when the channel is unknown, i.e., in solving the maximin problem of (9). Such a solution has many interpretations. Under some conditions (obtained in Section IV), it constitutes a saddle point of the strategic game formulation of (11) with the inherent properties of robustness. In any case, if mixed strategies are allowed in the strategic game, the solution to (9) always forms a saddle point defined by (13) (c.f. Section IV). Finally, even if we restrict the game to pure strategies, the solution to (9) always constitutes a saddle point as defined in (14) corresponding to a Stackelberg game. (The opposite minimax problem formulation of (10) is briefly considered in Section IV as well.)

IV. Robust Power Allocation

The main purpose of this section is to solve the maximin formulation of (9) and to characterize the conditions under which the solution forms a saddle point in the strategic formulation of the game (with pure strategies and mixed strategies).

As pointed out in Section III, we have to define some artificial constraint on the channel to avoid the trivial solution. Noting from (6) and (7) that the payoff function $\Psi(\boldsymbol{Q}, \boldsymbol{H})$ depends on $\boldsymbol{H}$ through $\boldsymbol{H}^H\boldsymbol{H}$ (the left singular vectors of $\boldsymbol{H}$ are irrelevant), it is convenient to define $\mathcal{H}$ as

$$\mathcal{H} \triangleq \left\{\boldsymbol{H}\colon \boldsymbol{R}_H \triangleq \boldsymbol{H}^H\boldsymbol{H} \in \mathcal{R}_H\right\}. \tag{16}$$

To define the set $\mathcal{R}_H$ we consider any kind of spectral (eigenvalue) constraint given by

$$\mathcal{R}_H \triangleq \{\boldsymbol{R}_H\colon \{\lambda_i(\boldsymbol{R}_H)\} \in \mathcal{L}_{R_H}\} \tag{17}$$

where $\mathcal{L}_{R_H}$ denotes an arbitrary eigenvalue constraint (in Section IV-A some specific eigenvalue constraints are considered). (Clearly, the set $\mathcal{L}_{R_H}$ cannot contain the all-zero vector that would correspond to $\boldsymbol{H} = \boldsymbol{0}$.) In defining the set $\mathcal{R}_H$ as in (17), we are deliberately leaving the eigenvectors of $\boldsymbol{R}_H$ (equivalently, the right singular vectors of $\boldsymbol{H}$) totally unconstrained. This is so that no preference is given to any signaling direction,[9] i.e., to guarantee the isotropy of $\mathcal{R}_H$ (any direction is possible).

Definition 1: A set of matrices $\mathcal{H}$ is isotropically unconstrained if the right singular vectors of the elements of the set are unconstrained, i.e., if for each $\boldsymbol{H} \in \mathcal{H}$ then $\boldsymbol{HU} \in \mathcal{H}$ for any unitary matrix $\boldsymbol{U}$.

Clearly, the set $\mathcal{H}$ (defined according to (16) and (17)) is isotropically unconstrained. We remark that the results in this paper are valid regardless of the particular eigenvalue constraint chosen to define the set $\mathcal{L}_{R_H}$.

We first obtain two lemmas and then proceed to obtain the uniform power allocation as the maximin solution of (9), i.e., as the capacity-achieving solution of the compound vector Gaussian channel. Note that this could be proved in a shorter way by contradiction, i.e., by showing that, for any given power allocation, we can always find some channel that yields a lower capacity than the minimum capacity corresponding to the uniform power allocation (indeed, this is the technique used in Section VI for the MAC). Nevertheless, we obtain a more complete proof by characterizing the "shape" of the worst channel for any given power allocation and then we give some examples in order to gain insight into the problem.

Lemma 1: Given two positive semidefinite $n \times n$ Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the following holds:

$$\log\det(\boldsymbol{I} + \boldsymbol{AB}) \geq \sum_{i=1}^{n} \log\left(1 + \lambda_{A,\,i}\,\lambda_{B,\,n-i+1}\right) \tag{18}$$

where $\lambda_{A,\,i}$ and $\lambda_{B,\,i}$ denote the eigenvalues in decreasing order $(\lambda_i \geq \lambda_{i+1})$ of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Equality in (18) is achieved for $\boldsymbol{U}_A = \boldsymbol{U}_B\boldsymbol{J}$, where $\boldsymbol{U}_A$ and $\boldsymbol{U}_B$ contain the eigenvectors corresponding to the eigenvalues in decreasing order of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and $\boldsymbol{J}$ is the "backward identity" permutation matrix [39] defined as

$$\boldsymbol{J} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & \ddots & & 1 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 1 & & \ddots & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Proof: See Appendix A. □

Lemma 2: The global optimal solution to the following convex optimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = -\sum_{i=1}^{n} \log\left(1 + x_i\alpha_i\right), \qquad \text{with } 0 \leq \alpha_i \leq \alpha_{i+1} \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i \leq P \\ & x_i \geq x_{i+1} \geq 0, \qquad 1 \leq i \leq n-1 \end{aligned} \tag{19}$$

is given by the uniform solution

$$x_i^\star = P/n, \qquad 1 \leq i \leq n. \tag{20}$$

[9] For a flat multiantenna system, the term "direction" means literally spatial direction.

Proof: From an intuitive viewpoint, we can see that without the constraint $x_i \geq x_{i+1}$, the solution would be a water-filling, which would imply $x_i \leq x_{i+1}$. With the additional constraint, however, the solution will try to water-fill but always verifying the constraint $x_i \geq x_{i+1}$, resulting in $x_i = x_{i+1}$.

This result can be straightforwardly proved in a formal way using majorization theory [40]. First, rewrite the objective function as $f(\boldsymbol{x}) = \sum_{i=1}^{n} g_i(x_i)$ where $g_i(x) = -\log(1 + x\alpha_i)$. Since $g_i'(a) \geq g_{i+1}'(b)$ whenever $a \geq b$, function $f(\boldsymbol{x})$ is Schur-convex [40, Proposition 3.H.2] . Now, from the definition of Schur-convexity [40, Definition 3.A.1] and using the fact that the uniform solution is majorized by any other solution [40, p. 7], it follows that the minimum of $f(\boldsymbol{x})$ is attained by the uniform solution of (20). This result can be alternatively proved using convex optimization theory [22], [23]. □

Before proceeding to the main result, recall that the capacity of the compound vector Gaussian memoryless channel when the channel state is unknown was obtained in [29][10] as

$$\mathcal{C}(\mathcal{H}) = \sup_{\boldsymbol{Q} \in \mathcal{Q}} \inf_{\boldsymbol{H} \in \mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \tag{21}$$

under the mild assumption that $\mathcal{H}$ is bounded (if not, we can simply bound $\mathcal{H}$ by adding the constraint $\lambda_{\max}(\boldsymbol{H}^H\boldsymbol{H}) \leq c$ for a sufficiently large value of c, which can be done without loss of generality based on physical interpretations of the channel $\boldsymbol{H}$). The achievability was proved in [29] by showing the existence of a code (along with the decoding sets). Therefore, in theory, one can always find a code to achieve rates arbitrarily close to capacity and then use a universal decoder that decodes the received word according to the decoding set it belongs to (note that no knowledge of the channel state is required).

Theorem 1: The capacity of the compound vector Gaussian memoryless channel with power constraint P_T, n_T transmit dimensions, and n_R receive dimensions (without knowledge of the channel state) is

$$\mathcal{C}(\mathcal{H}) = \inf_{\boldsymbol{H} \in \mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + P_T/n_T \boldsymbol{H}\boldsymbol{H}^H\right) \tag{22}$$

where the class of channels $\mathcal{H}$ is an isotropically unconstrained set defined by (16) and (17) (unconstrained right-singular vectors). The capacity-achieving solution of (22) is given by a Gaussian code with a uniform power allocation

$$\boldsymbol{Q}^\star = P_T/n_T \boldsymbol{I}_{n_T} \tag{23}$$

which implies an independent signaling over the transmit dimensions.

Proof: Before proceeding further, we give an intuitive explanation of why the uniform power allocation is optimal in (21). Due to the symmetry of the problem, if the transmitter does not use a uniform power distribution, the channel will do an "inverse water-filling," i.e., it will redistribute its singular values so that the highest ones align with the lowest eigenvalues of $\boldsymbol{Q}$ (see Lemma 1). Therefore, maximizing the lowest eigenvalues of $\boldsymbol{Q}$ seems to be appropriate to avoid such a behavior. Indeed, this is achieved by the uniform power allocation.

We use the relation (6) and (7) and the fact that the eigenvectors of $\boldsymbol{R}_H = \boldsymbol{H}^H\boldsymbol{H}$ are unconstrained (see (17)) to simplify the inner minimization of (21) for a given $\boldsymbol{Q}$

$$\begin{aligned}
&\inf_{\boldsymbol{R}_H \in \mathcal{R}_H} \log\det(\boldsymbol{I} + \boldsymbol{Q}\boldsymbol{R}_H) \\
&\quad = \inf_{\{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H}} \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\,\lambda_{R_H,n_T-i+1}\right) \\
&\quad = \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\,\lambda^\star_{R_H,n_T-i+1}(\{\lambda_{Q,i}\})\right)
\end{aligned}$$

where Lemma 1 has been used (the minimizing eigenvectors are chosen according to $\boldsymbol{U}_{R_H} = \boldsymbol{U}_Q\boldsymbol{J}$) and $\{\lambda^\star_{R_H,i}(\{\lambda_{Q,i}\})\}$ denote the minimizing eigenvalues of $\boldsymbol{R}_H$ as a function of $\{\lambda_{Q,i}\}$, which depend on the particular constraint used to define the set $\mathcal{L}_{R_H}$ (in the next subsection, some specific examples of $\mathcal{L}_{R_H}$ are considered).

The outer maximization of (21) can be now written as

$$\begin{aligned}
\max_{\{\lambda_{Q,i}\}} \quad & \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\lambda^\star_{R_H,n_T-i+1}(\{\lambda_{Q,i}\})\right) \\
\text{s.t.} \quad & \sum_i \lambda_{Q,i} \leq P_T \\
& \lambda_{Q,i} \geq \lambda_{Q,i+1} \geq 0 \qquad 1 \leq i \leq n_T - 1
\end{aligned}$$

with solution given by $\lambda^\star_{Q,i} = P_T/n_T\ \forall i$. To show this, we just have to apply Lemma 2

$$\begin{aligned}
&\sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right) \\
&\qquad \leq \sum_{i=1}^{n_T} \log\left(1 + P_T/n_T\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right)
\end{aligned}$$

and then the obvious relation

$$\begin{aligned}
&\inf_{\{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H}} \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\lambda_{R_H,n_T-i+1}\right) \\
&\qquad \leq \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right)
\end{aligned}$$

to finally obtain

$$\begin{aligned}
&\sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\lambda^\star_{R_H,n_T-i+1}(\{\lambda_{Q,i}\})\right) \\
&\qquad \leq \sum_{i=1}^{n_T} \log\left(1 + P_T/n_T\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right).
\end{aligned}$$

Thus, the maximizing solution is given by $\lambda^\star_{Q,i} = P_T/n_T\ \forall i$, i.e., a uniform power allocation $\boldsymbol{Q}^\star = P_T/n_T\boldsymbol{I}_{n_T}$. □

Note that the worst case capacity expression (22) obtained in Theorem 1 can be simplified as

$$\begin{aligned}
&\inf_{\boldsymbol{H} \in \mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + P_T/n_T\boldsymbol{H}\boldsymbol{H}^H\right) \\
&\qquad = \inf_{\{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H}} \sum_{i=1}^{n_T} \log\left(1 + P_T/n_T\lambda_{R_H,i}\right).
\end{aligned} \tag{24}$$

[10]The extension to the complex-valued case is straightforward using the results of [24].

Theorem 1 is basically saying that when the channel state is unknown but known to belong to a set of possible channels $\mathcal{H}$, the optimum solution in the sense of providing the best worst case performance is given by the uniform power allocation of (23). In other words, it is the solution to the problem formulation as a Stackelberg game in which the transmitter moves first as depicted in Fig. 4. Note that if we had used instead the eigenvalue constraint $\lambda_{Q,\max} \leq \alpha$ to define the set $\mathcal{Q}$, it would have immediately followed $\lambda_{Q,i}^{\star} = \alpha \; \forall i$, i.e., a uniform solution as well.

The uniform power allocation and the corresponding minimizing channel always constitute a saddle point of the Stackelberg game as defined in (14). Depending on the specific definition of the set of channels $\mathcal{H}$, they will also form a saddle point of the strategic game as given in (11). The following corollary gives the exact conditions.

Corollary 1: The uniform power allocation

$$\boldsymbol{Q}^{\star} = P_T/n_T \boldsymbol{I}_{n_T}$$

obtained in Theorem 1 and the corresponding minimizing channel form a saddle point of the strategic game given by (11) if and only if the minimizing channel satisfies $\lambda_{R_H,i}^{\star} = \beta \; \forall i$ (in particular, this implies $n_R \geq n_T$).

Proof: Since the right inequality of (11) is satisfied by any solution to (9), it suffices to find the conditions under which the left inequality is satisfied, i.e., $\Psi(\boldsymbol{Q}, \boldsymbol{H}^{\star}) \leq \Psi(P_T/n_T \boldsymbol{I}, \boldsymbol{H}^{\star})$ where $\boldsymbol{H}^{\star} \triangleq \boldsymbol{H}^{\star}(P_T/n_T \boldsymbol{I})$ is the minimizing channel of Theorem 1 corresponding to the uniform power allocation. Recalling that $\Psi(\boldsymbol{Q}, \boldsymbol{H})$ is maximized when the eigenvectors of $\boldsymbol{Q}$ align with the right singular vectors of $\boldsymbol{H}$ and when the eigenvalues of $\boldsymbol{Q}$ water-fill the eigenvalues of $\boldsymbol{H}^H \boldsymbol{H}$ [1], it must be that $\boldsymbol{H}^{\star H} \boldsymbol{H}^{\star}$ is a diagonal matrix and has equal eigenvalues. Thus, it must be that $\boldsymbol{H}^{\star H} \boldsymbol{H}^{\star} = \beta \boldsymbol{I}$ or, equivalently, $\lambda_{R_H,i}^{\star} = \beta \; \forall i$ for some β. □

In the next subsection, specific definitions of $\mathcal{H}$ are considered and Corollary 1 will be invoked to show in which cases the uniform power allocation constitutes a saddle point of the strategic game.

In [33], the existence of a saddle point as defined in (11) was proved for any set of channels $\mathcal{H}$ such that $\boldsymbol{H}_c = \boldsymbol{I}$ and $\boldsymbol{R}_n \in \mathcal{R}_n$ where $\mathcal{R}_n$ is closed, bounded, and convex. With the additional constraint that $\mathcal{R}_n$ be isotropically unconstrained, the existence result of [33] can be combined with Theorem 1 and Corollary 1 to conclude that in such a case, the uniform solution for both the transmitter and the noise always constitute a saddle point of the strategic game as given in (11). We state this in the following corollary for further reference.

Corollary 2: Consider the set of channels $\mathcal{H}$ defined such that $\boldsymbol{H}_c = \boldsymbol{I}$ $(n_T = n_R)$ and $\boldsymbol{R}_n \in \mathcal{R}_n$ where $\mathcal{R}_n$ is closed, bounded, convex, and isotropically unconstrained (i.e., unconstrained eigenvectors). It then follows that the uniform power allocation $\boldsymbol{Q}^{\star} = P_T/n_T \boldsymbol{I}$ and the noise $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}$ always form a saddle point of the strategic game given by (11).

Many papers have obtained a uniform solution for both the transmitter and the noise (or jammer) as mutual information saddle points for the set of noise covariances with power constraint given by $\mathrm{Tr}(\boldsymbol{R}_n) \leq P_n$, e.g., [31], [36], [37] (also [30][11] for the particular case in which the background noise is removed). Corollary 2 generalizes such a result to an arbitrary set of noise covariances $\mathcal{R}_n$ (provided it is closed, bounded, convex, and isotropically unconstrained). Note that a constraint on the channel eigenvalues $\{\lambda_{R_H,i}\}$ can be alternatively expressed (whenever $n_T = n_R$) as a constraint of the form $\boldsymbol{H}_c = \boldsymbol{I}$ and $\boldsymbol{R}_n \in \mathcal{R}_n$ as considered in Corollary 2 since we can write $\lambda_i(\boldsymbol{R}_n) = \lambda_{n_T-i+1}(\boldsymbol{R}_H)^{-1}$.

As mentioned in Section III, even when the strategic game does not have a saddle point or Nash equilibrium, if mixed strategies are allowed the game has then an infinite set of saddle points or Nash equilibria as defined in (13) (see Appendix B). In particular, as proved in Appendix B, the mixed-strategy Nash equilibria are given by a pure strategy for the transmitter $\boldsymbol{Q}^{\star} = P_T/n_T \boldsymbol{I}_{n_T}$ (uniform power allocation) and a mixed strategy for nature that, for example, puts equal probability on each element of the set

$$\left\{ \boldsymbol{H} = \boldsymbol{U}_H \boldsymbol{\Sigma}_H^{\star} \boldsymbol{P} \boldsymbol{V}_H^H : \boldsymbol{P} \in \boldsymbol{\Pi} \right\}$$

where $\boldsymbol{\Sigma}_H^{\star}$ contains in the main diagonal the optimum (worst case) singular values corresponding to $\boldsymbol{Q}^{\star} = P_T/n_T \boldsymbol{I}_{n_T}$ (as in Theorem 1), $\boldsymbol{U}_H$ and $\boldsymbol{V}_H$ are two arbitrary unitary matrices, and $\boldsymbol{\Pi}$ is the set of the $n_T!$ different permutation matrices of size $n_T \times n_T$ (see Appendix B for a proof).

A. Examples of Channel Constraints

In this subsection, to gain further insight into the problem, we analyze in detail some particular constraints to define the set of channels $\mathcal{H}$. In principle, for each of the different constraints, it is possible to directly solve the corresponding maximin problem of (9). Using the result obtained in Theorem 1, however, we already know (provided that the set $\mathcal{H}$ is isotropically unconstrained) that the optimal solution is the uniform power allocation $\boldsymbol{Q}^{\star} = P_T/n_T \boldsymbol{I}_{n_T}$ and that the worst case channel is given by the minimizing solution to (24). It is important to remark that, to find the worst case channel, it is not necessary to solve the minimization for an arbitrary set $\{\lambda_{Q,i}\}$ to obtain $\{\lambda_{R_H,i}^{\star}(\{\lambda_{Q,i}\})\}$; it suffices to consider directly the uniform solution $\lambda_{Q,i} = P_T/n_T \;\; \forall i$ as in (24) and obtain $\{\lambda_{R_H,i}^{\star}(\{P_T/n_T\})\}$, which is a great simplification.

1) General Individual Channel Eigenvalue Constraint $\{\lambda_{R_H,i} \geq \beta_i\}$: Consider a general and individual constraint on each channel eigenvalue $\{\lambda_{R_H,i} \geq \beta_i\}$, where it is assumed that $\beta_i \geq \beta_{i+1} \geq 0$ and that all eigenvalues have a corresponding β_i without loss of generality (if not, one can always set $\beta_i = \beta_{i-1}$ or $\beta_i = 0$ as appropriate). The minimizing channel of (24) is easily obtained by minimizing each of the terms of the right-hand side (RHS) of (24) as

$$\lambda_{R_H,i}^{\star} = \beta_i \qquad 1 \leq i \leq n_T. \tag{25}$$

This solution is, in general, nonuniform and, by Corollary 1, is not a saddle point of the strategic game as given in (11). Note that Corollary 2 cannot be invoked to prove the existence of a

[11] Although [30] deals with the vector Gaussian AVC, the final problem formulation is also given by maximin and minimax mathematical problems.

saddle point (for $\boldsymbol{H}_c = \boldsymbol{I}$) since the constraints expressed in terms of noise eigenvalues

$$\lambda_{n,i} = \lambda^{-1}_{R_H, n_T-i+1} \le \beta^{-1}_{n_T-i+1}$$

in general define a nonconvex and unbounded region for the set $\mathcal{R}_n$.

Consider now a constraint just on the maximum channel eigenvalue $\lambda_{R_H, \max} \ge \beta$. This specific constraint has a special interest since the maximum eigenvalue of $\boldsymbol{R}_H = \boldsymbol{H}^H\boldsymbol{H}$ is an upper bound on the elements of $\boldsymbol{R}_H$ and, in particular, on the received power corresponding to the ith transmit dimension $[\boldsymbol{R}_H]_{ii} = \|\boldsymbol{h}_i\|^2$, where $\boldsymbol{h}_i$ is the ith column of the channel matrix $\boldsymbol{H}$. The minimizing channel is given by

$$\begin{cases} \lambda^\star_{R_H, \max} = \beta \\ \quad \lambda^\star_{R_H, i} = 0, \qquad 2 \le i \le n_T. \end{cases} \tag{26}$$

It is also of interest to consider a constraint just on the minimum channel eigenvalue $\lambda_{R_H, \min} \ge \beta$. The minimizing channel is now

$$\lambda^\star_{R_H, i} = \beta, \qquad 1 \le i \le n_T. \tag{27}$$

For this particular case, the minimizing channel is uniform and then, by Corollary 1, the uniform power allocation forms a saddle point of the strategic game as given in (11). Alternatively, Corollary 2 could have been invoked to show the existence of a saddle point (for $\boldsymbol{H}_c = \boldsymbol{I}$) since the constraints expressed in terms of noise eigenvalues $\lambda_{n, \max} \le \beta^{-1}$ form a closed, bounded, convex, and isotropically unconstrained set $\mathcal{R}_n$.

2) Channel Trace Constraint $\mathrm{Tr}\,(\boldsymbol{R}_H) = \sum_i \lambda_{R_H, i} \ge \beta$*:* The channel trace constraint is probably the most reasonable constraint from a physical standpoint since it represents the total channel energy $\|\boldsymbol{H}\|_F^2 = \mathrm{Tr}\,(\boldsymbol{R}_H)$. In [38], this channel constraint was considered obtaining the same results.

Since the function $f(\boldsymbol{x}) = \sum_{i=1}^n \log(1 + x_i\alpha)$ is Schur-concave ($-f(\boldsymbol{x})$ is Schur-convex) [40, Proposition 3.H.2] and any eigenvalue distribution is majorized by $(\sum_i \lambda_{R_H, i}, 0, \ldots, 0)$ [40, p. 7] (see the proof of Lemma 2 for a similar reasoning), it follows that its minimum value is achieved by

$$\begin{cases} \lambda^\star_{R_H, \max} = \beta \\ \quad \lambda^\star_{R_H, i} = 0, \qquad 2 \le i \le n_T. \end{cases} \tag{28}$$

This solution is clearly nonuniform and, by Corollary 1, does not constitute a saddle point of the strategic game as given in (11). Note that Corollary 2 cannot be invoked either to prove the existence of a saddle point (for $\boldsymbol{H}_c = \boldsymbol{I}$) since the constraint expressed in terms of noise eigenvalues $\sum_i \lambda^{-1}_{n,i} \ge \beta$ defines a nonconvex and unbounded region for the set $\mathcal{R}_n$.

3) Maximum Noise Eigenvalue Constraint $\lambda_{n, \max} \le \sigma^2$*:* This constraint is identical to the minimum channel eigenvalue constraint with solution given by (27).

4) Noise Trace Constraint $\mathrm{Tr}\,(\boldsymbol{R}_n) = \sum_i \lambda_{n,i} \le \sigma^2$*:* This is the constraint considered in most publications since it is a very natural constraint when the noise is interpreted as a jammer constrained in its average transmit power (as is the intended transmitter). See, for example, [31], [36], [37] and also [30] for the particular case in which the background noise is removed.

For this particular constraint, we can directly invoke Corollary 2 to show that the worst case noise is given by

$$\lambda^\star_{n,i} = \sigma^2/n_R, \qquad 1 \le i \le n_R \tag{29}$$

and that the uniform power allocation constitutes a saddle point of the strategic game as given in (11).

5) Banded Noise Covariance Constraint: In [33, Sec. III], a banded noise covariance constraint (a noise with correlation constraints at different lags) was analyzed in detail. Such a constraint is not isotropically unconstrained and, consequently, the results of this paper do not apply. Therefore, we cannot conclude that the uniform power allocation is the maximin solution to the mutual information game. In fact, the saddle-point solution was obtained in [33, Sec. III] to be given by the maximum-entropy extension for the noise and by a water-filling solution for the transmitter which in general is nonuniform.

B. On the Specific Choice of the Channel Constraints

As has been shown in Theorem 1, the uniform power allocation is the solution to the maximin problem of (9). In other words, it is a robust solution under channel uncertainty.

Other aspects and observations of the solution, such as whether it is better to have many antennas or just a few in a multiantenna system, depend on the particular choice of constraints that define the set of channels $\mathcal{H}$ which have to be tailored to each specific application. To illustrate this effect, we now consider some heuristic choices as examples.

Inspired by a communication system with multiple transmit and receive antennas with a unit-energy channel in the sense of expected value $\mathcal{E}[|\boldsymbol{H}_{i,j}|^2] = 1$ (which implies $\mathcal{E}[\mathrm{Tr}(\boldsymbol{H}^H\boldsymbol{H})] = n_T n_R$), we can similarly consider a worst case problem formulation with the trace constraint defined as

$$\mathrm{Tr}(\boldsymbol{H}^H\boldsymbol{H}) \ge \alpha n_T n_R$$

where α is a scaling factor that, for example, guarantees that the constraint is satisfied with a certain probability (if the constraint is not satisfied, an outage event is declared). In this case, using the results of Section IV-A2, the worst case capacity is given by

$$\log(1 + \alpha P_T n_R)$$

from which we can conclude that, while adding transmit antennas does not increase the worst case capacity when the channel state is unknown, adding receive antennas is always beneficial.

Inspired by a set of parallel subchannels, each with unit gain, we can instead define the trace constraint as (assuming $n_T \le n_R$)

$$\mathrm{Tr}(\boldsymbol{H}^H\boldsymbol{H}) \ge \alpha n_T.$$

In this case, the worst case capacity is given by

$$\log\,(1 + \alpha P_T)$$

from which we can conclude that the worst case performance is independent of the number of transmit and receive antennas when the channel state is unknown. However, for this scenario corresponding to a set of parallel subchannels, it may be more

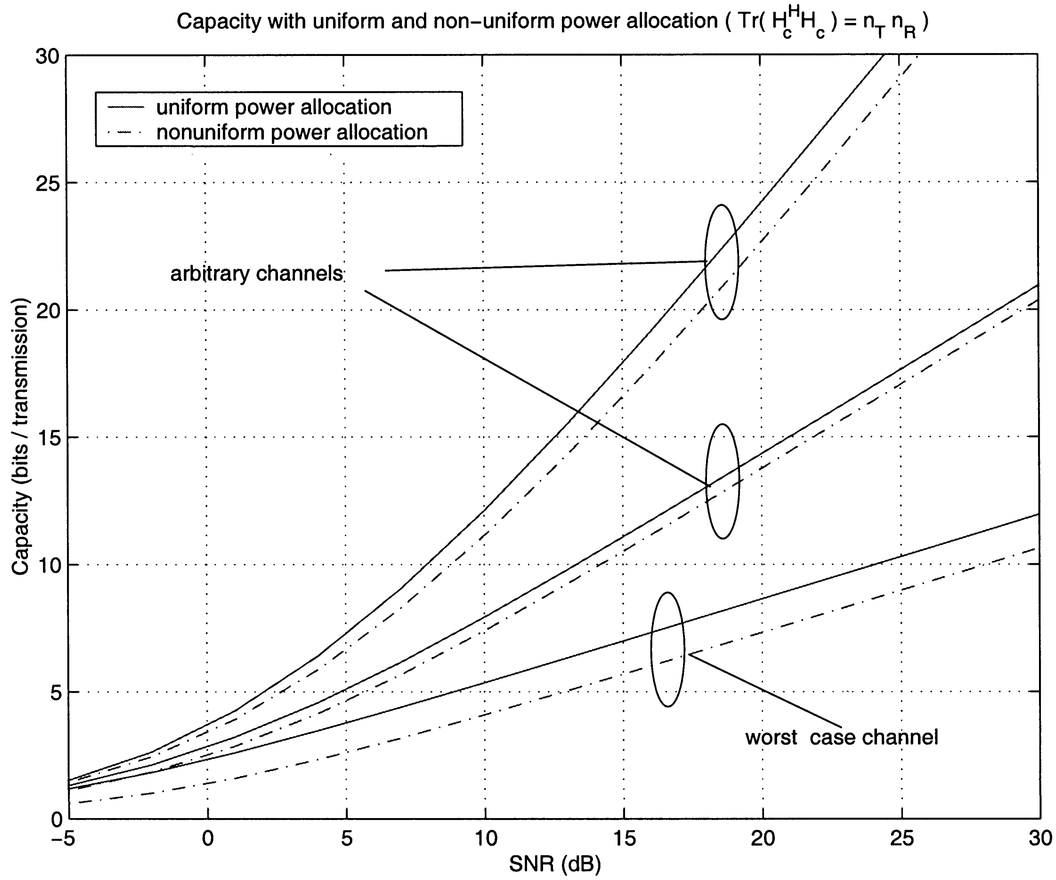

Fig. 5. Capacity of the uniform and a nonuniform (according to the distribution $\boldsymbol{\lambda}_Q = [0.6,\, 0.2,\, 0.1,\, 0.1]^T P_T$) power allocations versus the SNR for two arbitrary channels and for the worst channel of the set defined by $\mathrm{Tr}\left(\boldsymbol{H}_c^H \boldsymbol{H}_c\right) = n_T n_R$ and $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}$.

appropriate to consider the minimum channel eigenvalue constraint (assuming $n_T \leq n_R$)

$$\lambda_{\min}(\boldsymbol{H}^H \boldsymbol{H}) \geq \alpha$$

obtaining a worst case capacity (using the result in Section IV-A1) given by

$$n_T \log\left(1 + \alpha P_T / n_T\right) \underset{n_T \to \infty}{\longrightarrow} \alpha P_T$$

from which it is always beneficial to add transmit and also receive antennas.

A Numerical Example: In Fig. 5, the capacity of the uniform power allocation is compared to that of a nonuniform allocation (simply chosen according to the distribution $\boldsymbol{\lambda}_Q = [0.6,\, 0.2,\, 0.1,\, 0.1]^T P_T$) as a function of the SNR defined as $\mathrm{Tr}\,(\boldsymbol{Q})/\sigma_n^2$, where the noise covariance matrix was fixed to $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}$ and the set of channels $\mathcal{H}_c$ was constrained using the channel trace constraint

$$\mathrm{Tr}(\boldsymbol{H}_c^H \boldsymbol{H}_c) = n_T n_R, \qquad \text{for } n_T = n_R = 4$$

(equivalently, $\mathcal{H}$ is defined by $\mathrm{Tr}(\boldsymbol{H}^H \boldsymbol{H}) = n_T n_R / \sigma_n^2$). The capacities corresponding to two arbitrary channels and to the worst channel adapted to each power distribution are plotted. As expected, the capacity of the uniform distribution is always the best for the worst case channel (note that, in general, for an arbitrary channel, this may or may not be the case).

C. Opposite Problem Formulation: Nature Moves First

For completeness, we now briefly consider the opposite problem formulation, i.e., the minimax problem of (10). A solution to (10) will always be a saddle point as defined in (15) corresponding to the Stackelberg game in which nature moves first and then the transmitter moves aware of natures's move. In some cases, it will also form a saddle point of the strategic game as defined in (11).

It is well known that $\Psi\left(\boldsymbol{Q},\, \boldsymbol{H}\right)$ is maximized when the eigenvectors of $\boldsymbol{Q}$ align with the right singular vectors of $\boldsymbol{H}$ and when the eigenvalues of $\boldsymbol{Q}$ water-fill the eigenvalues of $\boldsymbol{R}_H = \boldsymbol{H}^H \boldsymbol{H}$ [1]. The minimax problem of (10) reduces then to

$$\begin{aligned} \min_{\{\lambda_{R_H,\,i}\}} \quad & \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,\,i}^{\star}\left(\{\lambda_{R_H,\,i}\}\right) \lambda_{R_H,\,i}\right) \\ \text{s.t.} \quad & \{\lambda_{R_H,\,i}\} \in \mathcal{L}_{R_H} \end{aligned} \tag{30}$$

where

$$\lambda_{Q,\,i}^{\star}\left(\{\lambda_{R_H,\,i}\}\right) = (\nu - \lambda_{R_H,\,i}^{-1})^{+}$$

is the water-filling solution, $(x)^{+} \triangleq \max\left(0,\, x\right)$, and ν is the water level chosen to satisfy the power constraint of (8) with equality. Clearly, we can relabel the $\lambda_{R_H,\,i}$'s so that they are in decreasing order without loss of generality and, as a consequence of the water-filling solution, the $\lambda_{Q,\,i}^{\star}$'s will also be in decreasing order.

For the cases considered in Section IV-A in which a saddle point was obtained (minimum channel eigenvalue constraint, maximum noise eigenvalue constraint, noise trace constraint, banded noise covariance constraint), we already know that the same solution is obtained when nature moves first simply by the definition of the saddle point in (11).

For the case of a general individual channel eigenvalue constraint $\{\lambda_{R_H,i} \geq \beta_i\}$, the worst case channel is simply obtained as in (25) (although, in this case, the eigenvalues of $\boldsymbol{Q}$ water-fill those of $\boldsymbol{R}_H$).

The channel trace constraint was considered in [38], where it was found that for low values of the SNR defined as $P_T\beta/\sigma_n^2$ (the noise covariance matrix was assumed fixed and given by $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}$) the worst channel is given by

$$\lambda^{\star}_{R_H,i} = \beta / \min(n_T, n_R), \qquad \forall i$$

and for high values of the SNR, the worst channel is similarly given except that a dominant eigenvalue arises.

D. Extension to Ergodic and Outage Capacities

In addition to analyzing robustness in terms of instantaneous mutual information as given by (6) (which implies a fixed channel state $\boldsymbol{H}$), it is also interesting to consider other statistics of the mutual information such as average and outage values. This implies a random channel state drawn according to some pdf $p_{\boldsymbol{H}}(\boldsymbol{H})$. For communication systems in which the transmission duration is so long as to reveal the long-term ergodic properties of the fading process (assumed to be an ergodic process in time), the ergodic capacity is a useful measure of the achievable bit rate [1], [27]. The mutual information for a given transmit covariance matrix $\boldsymbol{Q}$ is

$$\mathcal{I}^{\text{erg}}(\boldsymbol{Q}) = \mathcal{E}_{p_{\boldsymbol{H}}} \log\det\left(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right).$$

The ergodicity assumption, however, is not necessarily satisfied in practical communication systems operating on fading channels because no significant channel variability may occur during the whole transmission for applications with stringent delay constraints. In these circumstances, the outage capacity defined as the capacity that cannot be supported for only a small outage probability ϵ (also known as ϵ-achievable rate [14], [25]) is an appropriate measure [12], [27], [1]. The mutual information with outage probability ϵ for a given transmit covariance matrix $\boldsymbol{Q}$ is

$$\mathcal{I}_{\epsilon}^{\text{out}}(\boldsymbol{Q}) = \sup_{R}\left\{R\colon \Pr\left\{\log\det\left(\boldsymbol{I}_{n_R} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \leq R\right\} \leq \epsilon\right\}.$$

If the channel pdf were known, then an optimal fixed power allocation (independent of the actual channel realization) could be precomputed to maximize either $\mathcal{I}^{\text{erg}}(\boldsymbol{Q})$ or $\mathcal{I}_{\epsilon}^{\text{out}}(\boldsymbol{Q})$ over the set of $\boldsymbol{Q}$ satisfying the power constraint to obtain the ergodic capacity $\mathcal{C}^{\text{erg}}$ or outage capacity $\mathcal{C}_{\epsilon}^{\text{out}}$, respectively.

The maximin formulation is as in (9), but now the payoff function is given either by $\mathcal{I}^{\text{erg}}(\boldsymbol{Q})$ or $\mathcal{I}_{\epsilon}^{\text{out}}(\boldsymbol{Q})$ instead of (6) and the minimization is over the set of possible channel pdfs $p_{\boldsymbol{H}} \in \mathcal{P}_H$ in which the channel singular vectors are unconstrained (isotropic property), e.g., $\mathcal{E}[\lambda_{\max}(\boldsymbol{H}^H\boldsymbol{H})] \geq \beta$. Without going into details, we justify why the uniform power allocation is also obtained as a robust solution in terms of ergodic and outage capacities. Simply note that $p_{\boldsymbol{H}}$ can always be chosen as a function of the utilized $\boldsymbol{Q}$ to put positive probability only on channel states $\boldsymbol{H}$ with the singular vectors chosen to perform an "inverse water-filling" on $\boldsymbol{Q}$ (c.f. Theorem 1) against which the best solution for the transmitter is a uniform power allocation. Therefore, it is an optimal solution for every choice of $\boldsymbol{H}$ and, hence, for other capacity statistics such as the average and the outage values.

It is interesting to point out that if $p_{\boldsymbol{H}}$ does a randomization over a set of channel states in $\mathcal{H}$ as defined in the previous subsections (this need not be in a general case), the ergodic capacity problem then results in a mixed-strategy formulation of a game in which the pure strategies are defined by $\mathcal{H}$ and, therefore, the previously obtained results on mixed strategy Nash equilibria apply (c.f. Appendix B).

It is important to bear in mind that the optimality of the uniform power allocation in terms of ergodic and outage capacities is in the worst case sense, i.e., when $p_{\boldsymbol{H}}$ is known to belong to a set $\mathcal{P}_H$ but otherwise unknown. Therefore, it cannot be concluded from the obtained results that the uniform power allocation is optimum in terms of outage capacity for the case, for example, of a random $\boldsymbol{H}$ with i.i.d. $\mathcal{CN}(0,1)$ entries which is a well-known open problem as discussed in [1] (where it was conjectured that the uniform power allocation could be the optimal solution, but only over a certain number of transmit dimensions). It is interesting, however, to remark that, by definition, the worst case instantaneous capacity for a set of channels $\mathcal{H}$ as previously considered happens to be the zero-outage capacity (also termed delay-limited capacity [27], [14], [25]) for any $p_{\boldsymbol{H}}$ that puts nonzero probability on each member of $\mathcal{H}$. Unfortunately, this result is not very useful for the case of a random channel $\boldsymbol{H}$ with i.i.d. $\mathcal{CN}(0,1)$ entries, since it has a zero worst case capacity [25].[12]

Note that if one considers that nature can only control the channel eigenvalues but not the eigenvectors, then the optimality of the uniform power allocation need not hold.

V. Cost of Robustness

Robustness is a desirable property that comes with a price. For a given channel state $\boldsymbol{H}$, one can explicitly compute the loss in performance of the robust uniform power distribution with respect to the optimum allocation (obtained with a perfect instantaneous knowledge of the channel state). However, it is also interesting to know the worst case loss of performance for a given class of channels $\mathcal{H}$. In this section, the cost of robustness for a family of channels $\mathcal{H}$ is analyzed using the concept of duality gap arising in convex optimization theory [22], [23] following the approach proposed in [41]. We first review the notion of primal and dual objectives in convex optimization.

A general convex optimization problem (with no equality constraints) is of the form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0, \qquad 1 \leq i \leq m \end{aligned} \tag{31}$$

[12]The worst case capacity studied in this paper is equivalent to the delay-limited capacity considered in [25] under a short-term power constraint (in [25], however, perfect CSI was assumed).

where $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ is the optimization variable, $f_0(\boldsymbol{x})$ is called the primal objective (or objective function), and $f_0(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})$ are convex functions. The Lagrangian associated with the previous problem is defined as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) \tag{32}$$

where λ_i is the Lagrange multiplier or dual variable associated with the ith inequality constraint $f_i(\boldsymbol{x}) \leq 0$. The dual variables always take on nonnegative values.

The dual objective is defined as

$$g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{33}$$

and is a lower bound on the optimal $f_0(\boldsymbol{x})$. For any feasible $(\boldsymbol{x}, \boldsymbol{\lambda})$, we have

$$\begin{aligned} f_0(\boldsymbol{x}) &\geq f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) \\ &\geq \inf_{\boldsymbol{z}} \left(f_0(\boldsymbol{z}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{z}) \right) \\ &= g(\boldsymbol{\lambda}) \end{aligned} \tag{34}$$

and, in particular

$$\inf_{\boldsymbol{x}} f_0(\boldsymbol{x}) \geq \sup_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}) \tag{35}$$

where the maximization is over all nonnegative λ_i's and the minimization is over the original constraint set. The difference between the primal objective $f_0(\boldsymbol{x})$ and the dual objective $g(\boldsymbol{\lambda})$ is called the duality gap

$$\Gamma(\boldsymbol{x}, \boldsymbol{\lambda}) = f_0(\boldsymbol{x}) - g(\boldsymbol{\lambda}) \geq 0. \tag{36}$$

A central result in convex analysis [22] is that for convex optimization problems, under some technical conditions (called constraint qualifications), the duality gap reduces to zero at the optimal $\Gamma(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star) = 0$ or, equivalently, equality in (35) holds. We say then that strong duality holds. From (35), it is clear that the loss of any given $\boldsymbol{x}$ with respect to the optimal $\boldsymbol{x}^\star$ is upper-bounded by the gap function

$$f_0(\boldsymbol{x}) - f_0(\boldsymbol{x}^\star) \leq \Gamma(\boldsymbol{x}, \boldsymbol{\lambda}), \qquad \forall \boldsymbol{\lambda} \tag{37}$$

where equality holds for some $\boldsymbol{\lambda}$ if strong duality holds.

Let us now consider the specific problem at hand. Assuming for the moment a fixed channel state given by $\{\lambda_{R_H,i}\}$, the maximization of the mutual information can be expressed in convex form as (we use in this section logarithms in base 2 and natural logarithms denoted by $\log_2$ and $\ln$, respectively)

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) = -\sum_{i=1}^{n} \log_2(1 + x_i \lambda_{R_H,i}) \\ \text{s.t.} \quad & \sum_i x_i \leq P \\ & x_i \geq 0, \qquad 1 \leq i \leq n \end{aligned} \tag{38}$$

(note that Slater's condition is satisfied and, therefore, strong duality holds) and the Lagrangian is

$$L(\boldsymbol{x}, (\lambda, \boldsymbol{\mu})) = -\sum_{i=1}^{n} \log_2(1 + x_i \lambda_{R_H,i}) + \lambda \left(\sum_i x_i - P \right) - \sum_{i=1}^{n} \mu_i x_i. \tag{39}$$

The dual objective $g(\lambda, \boldsymbol{\mu})$ is obtained by setting $\frac{\partial L}{\partial x_i} = 0$, which gives the water-filling solution

$$x_i + \frac{1}{\lambda_{R_H,i}} = \frac{1}{(\lambda - \mu_i)} \frac{1}{\ln 2}. \tag{40}$$

If we now evaluate $\Gamma(\boldsymbol{x}, (\lambda, \boldsymbol{\mu}))$ at any $\boldsymbol{x}$ and with the Lagrange multipliers chosen so that the water-filling condition (40) is satisfied, then the duality gap is

$$\Gamma(\boldsymbol{x}, \lambda) = -\frac{1}{\ln 2} \sum_{i=1}^{n} \frac{x_i \lambda_{R_H,i}}{1 + x_i \lambda_{R_H,i}} + \lambda P. \tag{41}$$

(Note that a better choice of the Lagrange multipliers to obtain a smaller gap could be made, however, this choice produces a simple closed-form expression.) Using the smallest possible value for λ (such that all the Lagrange multipliers λ and μ_i's are nonnegative)

$$\lambda = \frac{1}{\ln 2} \max_i \left(\frac{\lambda_{R_H,i}}{1 + x_i \lambda_{R_H,i}} \right) \tag{42}$$

and assuming that the power constraint is satisfied with equality $\sum_i x_i = P$, we can write the duality gap as

$$\Gamma(\boldsymbol{x}) = \frac{1}{\ln 2} \sum_{i=1}^{n} x_i \left(\max_j \left(\frac{\lambda_{R_H,j}}{1 + x_j \lambda_{R_H,j}} \right) - \frac{\lambda_{R_H,i}}{1 + x_i \lambda_{R_H,i}} \right). \tag{43}$$

Finally, evaluating the gap for a uniform power allocation $x_i = P/n$, we obtain

$$\Gamma^{\text{uni}}(\boldsymbol{\lambda}_{R_H}) = \frac{1}{\ln 2} \sum_{i=2}^{n} \left(\frac{P/n\, \lambda_{R_H,\max}}{1 + P/n\, \lambda_{R_H,\max}} - \frac{P/n\, \lambda_{R_H,i}}{1 + P/n\, \lambda_{R_H,i}} \right) \tag{44}$$

where we have made explicit the dependence of the gap on the channel eigenvalues $\{\lambda_{R_H,i}\}$ which are assumed in decreasing order. For a channel with equal eigenvalues $\lambda_{R_H,i} = \kappa$, the uniform power allocation is optimum and the gap becomes zero as expected. Note that for $P \rightarrow \infty$ (with positive $\lambda_{R_H,i}$'s) the gap also tends to zero, i.e., for high SNR, the uniform distribution tends to be optimal (this observation was empirically made in [42] and further analyzed in [41]).

Now we can use the closed-form expression in (44) to easily obtain an upper bound on the worst case loss of performance for the class of channels $\mathcal{H}$. For example, if we consider a maximum channel eigenvalue constraint $\lambda_{R_H,\max} \geq \beta$, the gap is

$$\Gamma^{\text{uni}} = \frac{n-1}{\ln 2} \frac{P/n\, \lambda_{R_H,\max}}{1 + P/n\, \lambda_{R_H,\max}} \underset{\lambda_{R_H,\max} \rightarrow \infty}{\longrightarrow} \frac{n-1}{\ln 2}. \tag{45}$$

Note that for a channel trace constraint $\sum_i \lambda_{R_H,i} \geq \beta$, the same gap is obtained. If instead we consider a minimum channel eigenvalue constraint $\lambda_{R_H,\min} \geq \beta$, the gap is

$$\Gamma^{\text{uni}} = \frac{n-1}{\ln 2} \left(\frac{P/n\, \lambda_{R_H,\max}}{1 + P/n\, \lambda_{R_H,\max}} - \frac{P/n\, \beta}{1 + P/n\, \beta} \right) \underset{\lambda_{R_H,\max} \rightarrow \infty}{\longrightarrow} \frac{n-1}{\ln 2} \frac{1}{1 + P/n\, \beta}. \tag{46}$$

In any case, the gap in (44) is always upper-bounded as

$$\Gamma^{\text{uni}}(\boldsymbol{\lambda}_{R_H}) \leq \frac{n-1}{\ln 2} \tag{47}$$

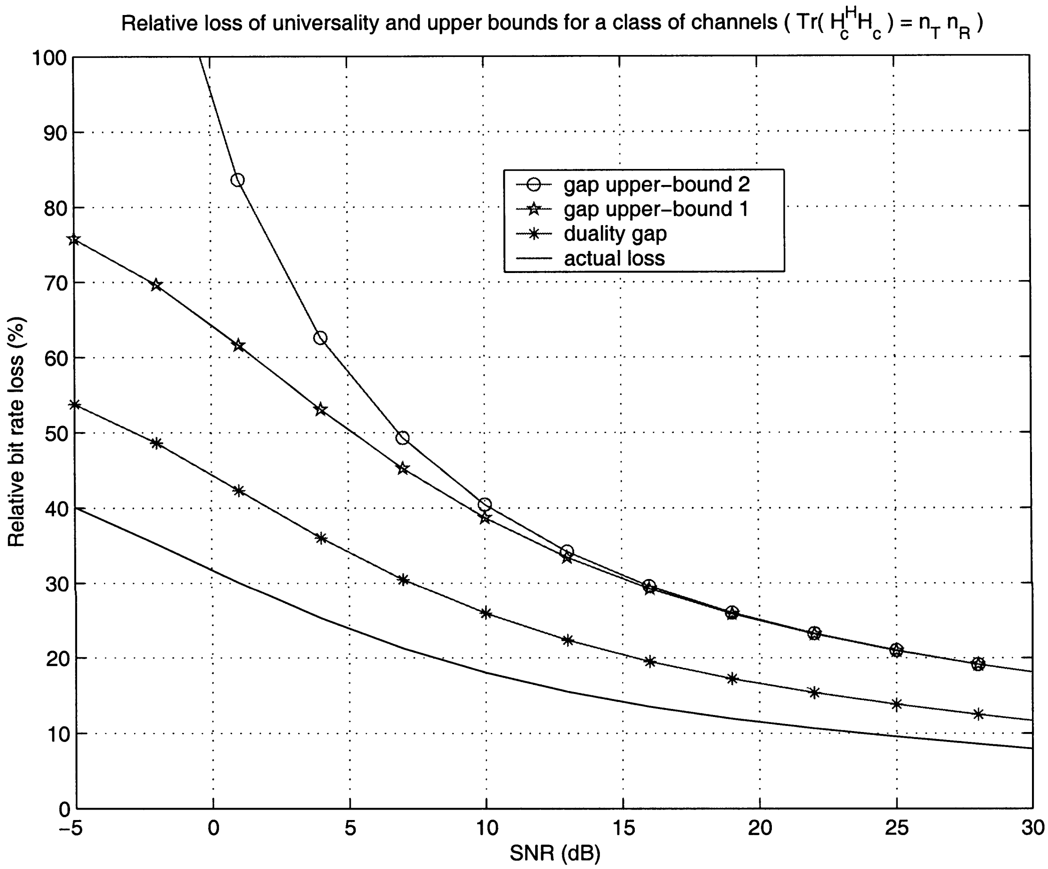

Fig. 6. Relative bit-rate loss and duality gap of the uniform power allocation, along with two upper bounds, versus the SNR for a channel realization ($\boldsymbol{\lambda}_{R_H} =$ (56%, ~44%, 10^{-3}%, 5×10^{-4}%$)^T$) corresponding to the class of channels defined by $\mathrm{Tr}\left(\boldsymbol{H}_c^H \boldsymbol{H}_c\right) = n_T n_R$ and $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}$.

which, in turn, is upper-bounded by $n/\ln 2$ bits per transmission or, equivalently, by 1.4427 bits/transmission/dimension as was found in [41].

Example: As an illustrative example, we consider a channel trace constraint given by $\mathrm{Tr}(\boldsymbol{H}_c^H \boldsymbol{H}_c) = n_T n_R$ for $n_T = n_R = 4$ (the noise covariance matrix was fixed to $\boldsymbol{R}_n = \sigma_n^2 \ \boldsymbol{I}$) and plot in Fig. 6 the actual relative bit-rate loss and duality gap as given in (44) for a channel realization, along with the gap upper bounds of (45) (both the asymptotic and the nonasymptotic versions), as a function of the SNR defined as $\mathrm{Tr}\left(\boldsymbol{Q}\right)/\sigma_n^2$.

VI. Extension to the Multiple-Access Channel

In this section, we extend the previous results on the single-user case and prove the optimality of the uniform power allocation in terms of robustness for the MAC. In particular, we show that all rates inside the capacity region of the compound vector MAC are achieved when each user uses a uniform power allocation. Note that, for the specific case of Gaussian-distributed channel matrices with i.i.d. entries, the optimality of the uniform power allocation in the sense of ergodic capacity was proved in [10], [11] (the proof is the natural extension of that of the single-user case given in [1] based on the concavity of the logdet function).

As in the single-user case of Section IV, we constrain our search to Gaussian-distributed signals and noise, since they constitute a robust solution (a saddle point) to the mutual information game for the memoryless vector MAC (this follows by applying the results of [32], [33] to each of the constraints that define the capacity region).

Consider a scenario with K users, each one transmitting over n_k dimensions with a (possibly different) power constraint $\mathrm{Tr}\left(\boldsymbol{Q}_k\right) \leq P_k$ and with channel $\boldsymbol{H}_k$. The signal model is the natural extension of (1) given by

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{n} \tag{48}$$

where the noise is assumed to be white without loss of generality,[13] i.e., $\boldsymbol{n} \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}\right)$. As in the single-user case of Section IV, we impose some constraints on the set of possible channels $\{\boldsymbol{H}_k\} \in \mathcal{H} \triangleq \mathcal{H}_1 \times \cdots \times \mathcal{H}_K$ to avoid the trivial solution (note that the class of channels seen by each user may be different). We assume that each set $\mathcal{H}_k$ is isotropically unconstrained, i.e., with unconstrained right singular vectors (see Definition 1 in Section IV).

[13] If the noise is not white, the receiver can always prewhiten the received signal and then the signal model in (48) with white noise applies. (It is also possible to explicitly consider a colored noise as was done in Section IV.)

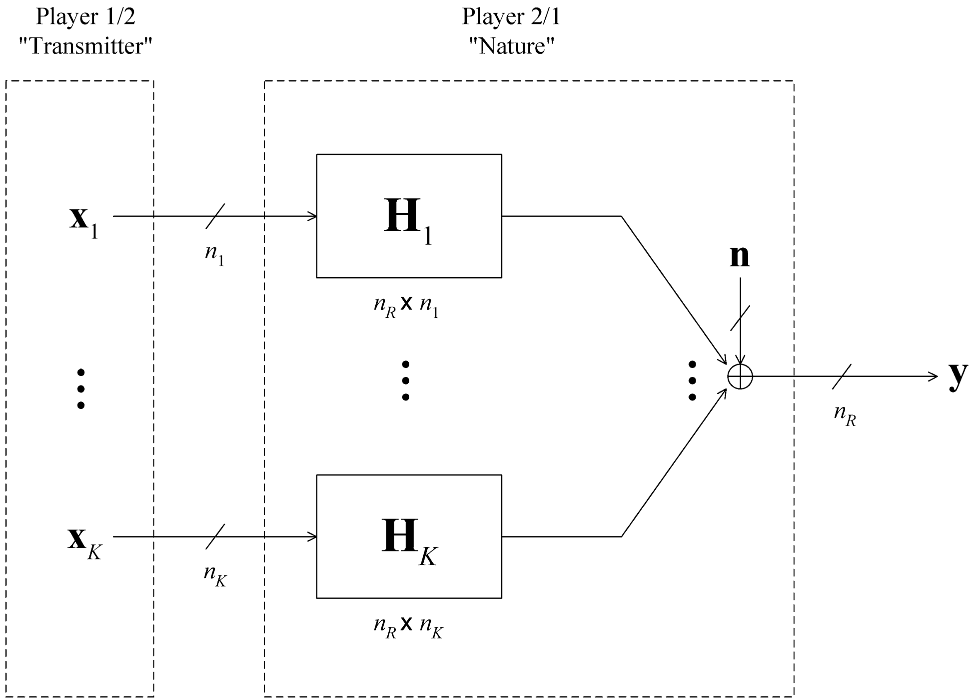

Fig. 7. Communication interpreted as a two-player game for the multiple-access channel.

In the multiple-access channel, we do not deal any more with a single capacity measure but with a capacity region. In particular, the achievable rate region for a given realization of the set of channels $\{\boldsymbol{H}_k\}$ and with a fixed set of covariance matrices $\{\boldsymbol{Q}_k\}$ (Gaussian codes are assumed since they maximize each of the boundaries that define the region) is as in (49) at the bottom of the page [4], [43]. Assuming that the transmit covariance matrices are constrained in their average transmit power, the capacity region is [43], [44]

$$\mathcal{C}\left(\{\boldsymbol{H}_k\}\right) = \bigcup_{\substack{\mathrm{Tr}\,(\boldsymbol{Q}_k) \le P_k, \\ \boldsymbol{Q}_k = \boldsymbol{Q}_k^H \ge 0}} \mathcal{R}\left(\{\boldsymbol{Q}_k\}, \{\boldsymbol{H}_k\}\right). \tag{50}$$

Note that the convex closure operation usually needed [4] is unnecessary in this case because the region is already closed and convex as shown in [43], [44].

From the perspective of robustness under channel uncertainty, we are interested in the worst case capacity region, i.e., in the set of rates that can be achieved regardless of the set of channel states chosen from the set of possible channels $\{\boldsymbol{H}_k\} \in \mathcal{H}$. This can be formulated as a game (see Fig. 7), where one player is the transmitter and the other player, who controls the whole set of channels $\{\boldsymbol{H}_k\}$, is nature. The worst case capacity region is, in fact, the notion of capacity region of the compound MAC [43], [20] (see also [28, p. 288]). Mathematically, the worst case region of the set of achievable rates for a fixed set of transmit covariance matrices $\{\boldsymbol{Q}_k\}$ is expressed as the following intersection:

$$\mathcal{R}\left(\{\boldsymbol{Q}_k\}, \mathcal{H}\right) = \bigcap_{\{\boldsymbol{H}_k\} \in \mathcal{H}} \mathcal{R}\left(\{\boldsymbol{Q}_k\}, \{\boldsymbol{H}_k\}\right) \tag{51}$$

which is closed and convex because it is the intersection of closed and convex sets. Assuming that the transmit covariance matrices are constrained in their average transmit power, the worst case capacity region (capacity region of the compound vector Gaussian MAC) is [43][14]

$$\mathcal{C}\left(\mathcal{H}\right) = \bigcup_{\substack{\mathrm{Tr}\,(\boldsymbol{Q}_k) \le P_k, \\ \boldsymbol{Q}_k = \boldsymbol{Q}_k^H \ge 0}} \mathcal{R}\left(\{\boldsymbol{Q}_k\}, \mathcal{H}\right) \tag{52}$$

which also happens to be closed and convex as shown in Theorem 2. In [45], an expression similar to (52) was obtained as the delay-limited MAC capacity region (although the case of perfect CSIT was considered therein). The worst case capacity region is formally characterized in the following theorem.

[14]As argued in [43], achievability follows easily using randomized codes and the converse is established since reliable communication has to be guaranteed no matter what channel state is in effect. Similarly to the single-user case, the capacity region remains the same if the receiver is uninformed of the channel state [28, p. 293].

$$\mathcal{R}(\{\boldsymbol{Q}_k\}, \{\boldsymbol{H}_k\}) = \left\{ (R_1, \ldots, R_K) : 0 \le \sum_{k \in \mathcal{S}} R_k \le \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k \in \mathcal{S}} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^H\right), \forall \mathcal{S} \subseteq \{1, \ldots, K\} \right\}. \tag{49}$$

Theorem 2: The capacity region of the compound vector Gaussian memoryless MAC composed of K users with power constraints $\{P_k\}$, number of transmit dimensions $\{n_k\}$, and n_R receive dimensions (without knowledge of the channel state) is as in (53) at the bottom of the page, where the class of channels $\mathcal{H}$ is an isotropically unconstrained set (unconstrained right singular vectors). All set of rates within the region of (53) are achieved when each user utilizes a Gaussian code with a uniform power allocation

$$\boldsymbol{Q}_k^\star = P_k/n_k \boldsymbol{I}_{n_k}, \qquad 1 \le k \le K \tag{54}$$

which implies an independent signaling over the transmit dimensions for each user.

Proof: The rate region given by (51) is the intersection of a set of regions each of which is, in turn, defined by the intersection of $2^K - 1$ nontrivial inequalities as in (49). We can, therefore, rewrite the rate region of (51) as the region defined by the more restrictive of each one of the $2^K - 1$ inequalities over the set of possible channels (as was done in [43]) in (55), also at the bottom of the page. Note that the capacity region of the compound vector Gaussian MAC as given by (52) and (55) is the natural counterpart of the capacity of the single-user compound vector Gaussian channel of (21). Similarly, (53) is the natural counterpart of (22).

We have to show now that the inequalities defining the rate region in (55) corresponding to nonuniform power distributions are always more restrictive than for the uniform power distribution, i.e.,

$$\begin{aligned} &\inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right) \\ &\le \inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k \boldsymbol{H}_k\boldsymbol{H}_k^H\right), \\ &\quad \forall \mathcal{S} \subseteq \{1, \ldots, K\} \\ &\quad \forall \boldsymbol{Q}_k\colon \operatorname{Tr}(\boldsymbol{Q}_k) \le P_k,\, \boldsymbol{Q}_k = \boldsymbol{Q}_k^H \ge 0, \quad 1 \le k \le K. \end{aligned} \tag{56}$$

This has the important consequence that the worst case rate region of the uniform power allocation contains the worst case rate region corresponding to any other power allocation strategy, i.e.,

$$\begin{aligned} &\mathcal{R}(\{\boldsymbol{Q}_k\}, \mathcal{H}) \subseteq \mathcal{R}(\{P_k/n_k\boldsymbol{I}_{n_k}\}, \mathcal{H}) \\ &\quad \forall \boldsymbol{Q}_k\colon \operatorname{Tr}(\boldsymbol{Q}_k) \le P_k,\, \boldsymbol{Q}_k = \boldsymbol{Q}_k^H \ge 0, \qquad 1 \le k \le K. \end{aligned}$$

Therefore, the expression of the worst case capacity region in (52) reduces to

$$\mathcal{C}(\mathcal{H}) = \mathcal{R}(\{P_k/n_k\boldsymbol{I}_{n_k}\}, \mathcal{H})$$

which, together with (55), gives the desired result of (53). Now that (52) has been rewritten as (53), it is clear that it is a closed and convex region.

We now focus on proving the inequalities of (56). We first consider a single user in the set $\mathcal{S}$ and show that with a uniform power distribution, the boundary can never decrease. Then, we apply the same idea for the rest of the users in $\mathcal{S}$. Consider the minimization of the boundary with respect to the channel $\boldsymbol{H}_l$ of the lth user in $\mathcal{S}$; for any given set of channels $\{\boldsymbol{H}_k\}_{k\ne l}$, we have

$$\begin{aligned} &\inf_{\boldsymbol{H}_l\in\mathcal{H}_l} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right) \\ &= \inf_{\boldsymbol{H}_l\in\mathcal{H}_l} \log\det\left(\boldsymbol{R}_{n_l} + \boldsymbol{H}_l\boldsymbol{Q}_l\boldsymbol{H}_l^H\right) \\ &= \inf_{\boldsymbol{H}_l\in\mathcal{H}_l} \log\det\left(\boldsymbol{I}_{n_T} + \boldsymbol{Q}_l\boldsymbol{H}_l^H\boldsymbol{R}_{n_l}^{-1}\boldsymbol{H}_l\right) + \log\det(\boldsymbol{R}_{n_l}) \\ &\le \inf_{\boldsymbol{H}_l\in\mathcal{H}_l} \log\det\left(\boldsymbol{I}_{n_T} + P_l/n_l\, \boldsymbol{H}_l^H\boldsymbol{R}_{n_l}^{-1}\boldsymbol{H}_l\right) + \log\det(\boldsymbol{R}_{n_l}) \\ &= \inf_{\boldsymbol{H}_l\in\mathcal{H}_l} \log\det\left(\boldsymbol{I}_{n_R} + P_l/n_l\boldsymbol{H}_l\boldsymbol{H}_l^H + \sum_{\substack{k\in\mathcal{S}\\k\ne l}} \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right) \end{aligned}$$

where

$$\boldsymbol{R}_{n_l} \triangleq \left(\boldsymbol{I}_{n_R} + \sum_{\substack{k\in\mathcal{S}\\k\ne l}} \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right)$$

is the interference-plus-noise covariance matrix seen by the lth user and the inequality comes from invoking Lemmas 1 and 2 as was done when proving Theorem 1 in the single-user case. The previous reasoning can be sequentially applied to each of the users in the set $\mathcal{S}$ to finally obtain

$$\begin{aligned} &\inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right) \\ &\qquad \le \inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k\boldsymbol{H}_k\boldsymbol{H}_k^H\right). \end{aligned}$$

$$\mathcal{C}(\mathcal{H}) = \left\{(R_1, \ldots, R_K) : 0 \le \sum_{k\in\mathcal{S}} R_k \le \inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k\boldsymbol{H}_k\boldsymbol{H}_k^H\right), \forall \mathcal{S} \subseteq \{1, \ldots, K\}\right\} \tag{53}$$

$$\mathcal{R}(\{\boldsymbol{Q}_k\}, \mathcal{H}) = \left\{(R_1, \ldots, R_K)\colon 0 \le \sum_{k\in\mathcal{S}} R_k \le \inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right), \forall \mathcal{S} \subseteq \{1, \ldots, K\}\right\}. \tag{55}$$

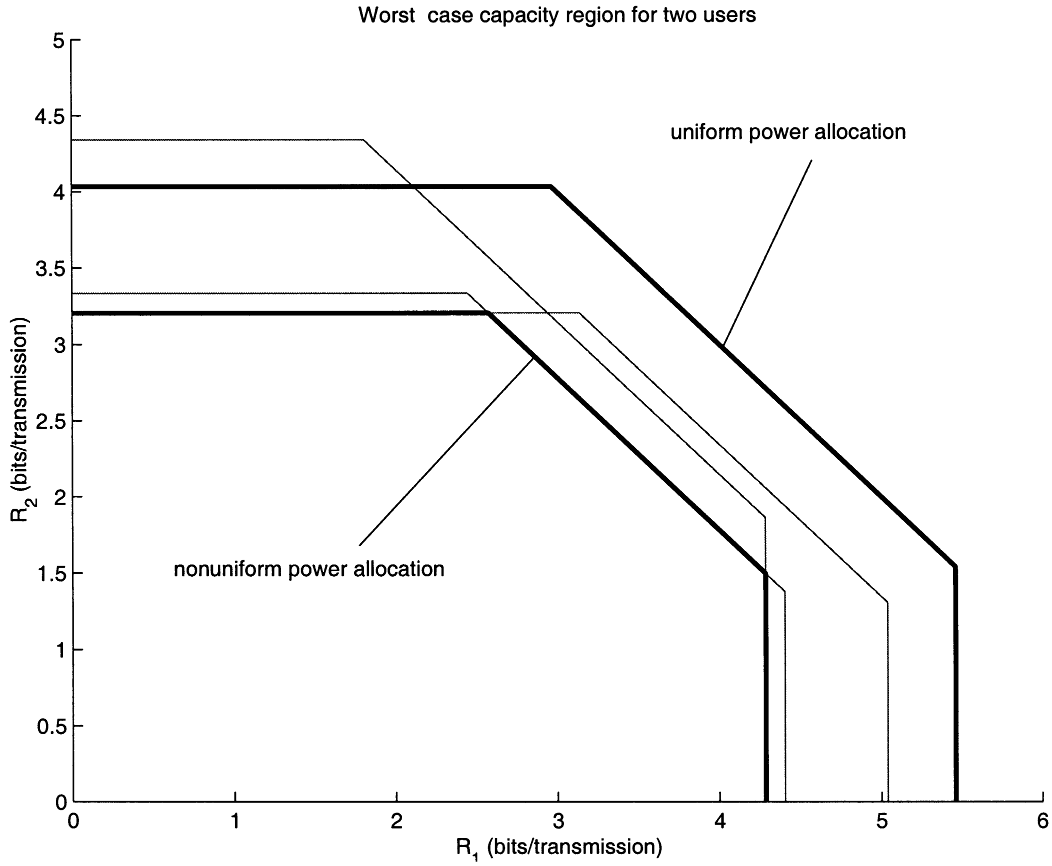

Fig. 8. Worst case capacity region corresponding to the channel eigenvalues $\boldsymbol{\lambda}\left(\boldsymbol{H}_1\boldsymbol{H}_1^H\right) = [9.80,\ 9.24,\ 4.59]^T$ and $\boldsymbol{\lambda}\left(\boldsymbol{H}_2\boldsymbol{H}_2^H\right) = [9.19,\ 5.45,\ 1.29]^T$ when using a uniform and nonuniform (according to $\boldsymbol{\lambda}_{Q_1} = [0.64,\ 0.34,\ 0.02]^T$ and $\boldsymbol{\lambda}_{Q_2} = [0.52,\ 0.40,\ 0.08]^T$) power allocation. The latter is obtained as the intersection of the three capacity regions plotted by thin lines.

Therefore, a nonuniform power allocation always has a lower (or at most equal) worst case boundary for all inequalities defining the capacity region in (56). This concludes the proof. □

It is important to remark that all points inside the worst case capacity region are achieved by the same structure of transmit covariance matrices $\{\boldsymbol{Q}_k\}$, i.e., by a uniform power allocation $\{P_k/n_k\boldsymbol{I}_{n_k}\}$. This is a significant difference with respect to the case with CSIT obtained from (50) in which each point of the region requires, in general, a different structure for the transmit covariance matrices [44].

It is possible to further simplify the expression for each of the boundaries of the worst case capacity region (53) obtained in Theorem 2, provided that the left singular vectors of the class of channels are unconstrained as well (this means unconstrained receive as well as transmit directions). In other words, only the singular values of the channels are constrained and, therefore, $\boldsymbol{H}_k \in \mathcal{H}_k$ if and only if $\{\lambda_i(\boldsymbol{H}_k\boldsymbol{H}_k^H)\} \in \mathcal{L}_{H_k}$ (we similarly define $\mathcal{L}_H \triangleq \mathcal{L}_{H_1} \times \cdots \times \mathcal{L}_{H_K}$). We first state a lemma and then proceed to simplify the boundaries of the worst case capacity region in (53).

Lemma 3: Let $\{\boldsymbol{R}_k\}$ be a set of J $n \times n$ Hermitian matrices. Then, the following inequality is verified:

$$\det\left(\boldsymbol{R}_1 + \cdots + \boldsymbol{R}_J\right) \geq \prod_i \left(\lambda_i\left(\boldsymbol{R}_1\right) + \cdots + \lambda_i\left(\boldsymbol{R}_J\right)\right) \quad (57)$$

where $\lambda_i(\cdot)$ denotes the ith ordered eigenvalue in decreasing order and equality is achieved when all the $\boldsymbol{R}_i$'s have the same eigenvectors with eigenvalues in the same order.

Proof: This result is a generalization of the particular case $J = 2$ considered in [40, Theorem 9.G.3.a] and is proved in Appendix C. □

Since the left singular vectors of the channels $\{\boldsymbol{H}_k\}$ are unconstrained, we can invoke Lemma 3 to obtain

$$\begin{aligned} &\inf_{\{\boldsymbol{H}_k\}\in\mathcal{H}} \log\det\left(\boldsymbol{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k \boldsymbol{H}_k\boldsymbol{H}_k^H\right) \\ &= \inf_{\{\lambda_i(\boldsymbol{H}_k\boldsymbol{H}_k^H)\}\in\mathcal{L}_H} \sum_{i=1}^{n_R} \log\left(1 + \sum_{k\in\mathcal{S}} P_k/n_k \lambda_i\left(\boldsymbol{H}_k\boldsymbol{H}_k^H\right)\right). \quad (58)\end{aligned}$$

This implies that the worst case is obtained by choosing the same left singular vectors for each $\boldsymbol{H}_k$ (the right singular vectors are irrelevant) such that the eigenvalues of $\boldsymbol{H}_k\boldsymbol{H}_k^H$ are ordered in the same way for all k.

Example: In Fig. 8, the worst case capacity region of a two-user system is plotted for a class of 3×3 matrix channels with eigenvalues constrained to be exactly

$$\boldsymbol{\lambda}\left(\boldsymbol{H}_1\boldsymbol{H}_1^H\right) = [9.80,\ 9.24,\ 4.59]^T$$

and

$$\boldsymbol{\lambda}\left(\boldsymbol{H}_2\boldsymbol{H}_2^H\right) = [9.19,\, 5.45,\, 1.29]^T.$$

The three inequalities defining the capacity region corresponding to the uniform power allocation are simultaneously minimized by the same worst case set of channels according to (58). We also plot the worst case capacity region corresponding to a nonuniform power allocation, in particular for

$$\boldsymbol{\lambda}_{Q_1} = [0.64,\, 0.34,\, 0.02]^T$$

and

$$\boldsymbol{\lambda}_{Q_2} = [0.52,\, 0.40,\, 0.08]^T.$$

In this case, however, the three inequalities are not simultaneously minimized by the same choice of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. To obtain the capacity region, therefore, we have to obtain the three capacity regions in which each one of the three inequalities is minimized[15] (plotted in thin lines) and then compute the intersection. This is due to the fact that, in general, there are no channels $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ that simultaneously minimize all inequalities, unlike in the uniform case.

VII. Conclusion

When transmitting a vector signal through a MIMO channel, the power allocation over the transmit dimensions has to be properly chosen. When the instantaneous channel realization is known, the solution is well known and is based on diagonalizing the channel and performing water-filling over the channel eigenmodes. When the channel realization is unknown at the transmitter, but the channel statistics are *a priori* known, the optimal power allocation can, in principle, be precomputed. In this sense, the uniform power allocation has been previously shown in the literature to be optimum in terms of ergodic capacity for some particular cases.

This paper has considered the case in which not even the channel distribution is known at the transmitter. We have formulated the problem within the framework of game theory in which the payoff function of the game is the mutual information and the players are the transmitter and a malicious nature. Mathematically, this has been expressed as a maximin problem, obtaining a robust power allocation under channel uncertainty. This problem characterizes the capacity of the compound vector Gaussian channel. The uniform power allocation has been obtained as a robust solution of the game in terms of capacity for the class of isotropically unconstrained channels (unconstrained "directions"). The loss of capacity when using the uniform power allocation has been analytically bounded, showing that for high SNR the loss is small.

For the more interesting and general case of the MAC, the uniform power allocation for each of the users also constitutes a robust solution. To be more specific, the worst case rate region corresponding to the uniform power distribution is shown to contain the worst case rate region of any other possible power allocation strategy. In other words, the capacity region of the compound vector Gaussian MAC is achieved when each of the users is using a uniform power allocation.

[15] The worst case capacity region corresponding to the nonuniform power allocation in Fig. 8 has been computed by choosing the channels with left singular vectors as dictated by Lemma 3 and by arbitrarily choosing the right singular vectors to diagonalize the transmit covariance matrices and then optimizing over the permutations only. The ultimate worst case capacity region by properly optimizing the right singular vectors may be even smaller.

Appendix A
Proof of Lemma 1

Consider the eigendecompositions

$$\boldsymbol{A} = \boldsymbol{U}_A\boldsymbol{D}_A\boldsymbol{U}_A^H \quad \text{and} \quad \boldsymbol{B} = \boldsymbol{U}_B\boldsymbol{D}_B\boldsymbol{U}_B^H$$

where $\boldsymbol{D}_A = \operatorname{diag}(\{\lambda_{A,i}\})$ and $\boldsymbol{D}_B = \operatorname{diag}(\{\lambda_{B,i}\})$ (we assume eigenvalues in decreasing order). It follows that

$$\det(\boldsymbol{I}+\boldsymbol{A}\boldsymbol{B}) = \det\left(\boldsymbol{I}+\boldsymbol{D}_{\boldsymbol{A}}\tilde{\boldsymbol{U}}^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}\right)$$

where $\tilde{\boldsymbol{U}} = \boldsymbol{U}_B^H\boldsymbol{U}_A$. If $\boldsymbol{A}$ has $n-k$ zero eigenvalues, we can write

$$\boldsymbol{D}_A = \begin{bmatrix}\boldsymbol{D}_{A,1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}, \qquad \boldsymbol{D}_{A,1}\in\mathbb{C}^{k\times k} \quad \text{(nonsingular)}$$

$$\tilde{\boldsymbol{U}} = \begin{bmatrix}\tilde{\boldsymbol{U}}_1 & \tilde{\boldsymbol{U}}_2\end{bmatrix}, \qquad \tilde{\boldsymbol{U}}_1\in\mathbb{C}^{n\times k}$$

and then

$$\begin{aligned}
&\det\left(\boldsymbol{I}+\boldsymbol{D}_{\boldsymbol{A}}\tilde{\boldsymbol{U}}^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}\right)\\
&\quad= \det\left(\boldsymbol{I}+\begin{bmatrix}\boldsymbol{D}_{A,1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{U}}_1^H\\ \tilde{\boldsymbol{U}}_2^H\end{bmatrix}\boldsymbol{D}_B\begin{bmatrix}\tilde{\boldsymbol{U}}_1 & \tilde{\boldsymbol{U}}_2\end{bmatrix}\right)\\
&\quad= \det\left(\boldsymbol{I}+\begin{bmatrix}\boldsymbol{D}_{A,1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{U}}_1^H\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{D}_B\begin{bmatrix}\tilde{\boldsymbol{U}}_1 & \boldsymbol{0}\end{bmatrix}\right)\\
&\quad= \det\left(\boldsymbol{I}+\boldsymbol{D}_{A,1}\tilde{\boldsymbol{U}}_1^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}_1\right)\\
&\quad= \det(\boldsymbol{D}_{A,1})\det\left(\boldsymbol{D}_{A,1}^{-1}+\tilde{\boldsymbol{U}}_1^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}_1\right)\\
&\quad\geq \left(\prod_{i=1}^{k}\lambda_{A,i}\right)\left(\prod_{i=1}^{k}\left(\lambda_{A,k-i+1}^{-1}+\lambda_i\left(\tilde{\boldsymbol{U}}_1^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}_1\right)\right)\right)\\
&\quad\geq \prod_{i=1}^{k}\left(1+\lambda_{A,k-i+1}\,\lambda_{B,n-k+i}\right)\\
&\quad= \prod_{i=1}^{n}\left(1+\lambda_{A,n-i+1}\,\lambda_{B,i}\right)
\end{aligned}$$

where $\lambda_i(\cdot)$ denotes the ith eigenvalue in decreasing order. In the first inequality, we have used the inequality [40, Theorem 9.G.3.a]

$$\det(\boldsymbol{A}+\boldsymbol{B}) \geq \prod_{i=1}^{n}\left(\lambda_i(\boldsymbol{A})+\lambda_i(\boldsymbol{B})\right)$$

with equality verified for $\tilde{\boldsymbol{U}}_1^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}_1$ diagonal, i.e., when $\tilde{\boldsymbol{U}}_1$ is a permutation matrix. In the second inequality, we have used the Poincaré Separation Theorem [46, p. 209]

$$\lambda_{B,i} \geq \lambda_i\left(\tilde{\boldsymbol{U}}_1^H\boldsymbol{D}_B\tilde{\boldsymbol{U}}_1\right) \geq \lambda_{B,i+n-k}, \qquad 1\leq i\leq k$$

with equality verified when $\tilde{\boldsymbol{U}}_1$ (note that $\tilde{\boldsymbol{U}}_1^H\tilde{\boldsymbol{U}}_1 = \boldsymbol{I}_k$) selects the k smallest diagonal elements of $\boldsymbol{D}_B$. Since the logarithm is a monotonic increasing function, taking the logarithm on both sides completes the proof. Equality is achieved for $\tilde{\boldsymbol{U}}$ being a permutation matrix that sorts the diagonal elements of $\boldsymbol{D}_B$ in increasing order, i.e., $\tilde{\boldsymbol{U}} = \boldsymbol{J}$. Note, however, that if $\boldsymbol{A}$ has zero eigenvalues, then $\tilde{\boldsymbol{U}}_2$ can be freely chosen as long as $\tilde{\boldsymbol{U}}$ remains unitary. □

APPENDIX B
MIXED STRATEGY NASH EQUILIBRIA

In this appendix, we characterize the solutions to the mixed-strategy saddle point given by (13).

By the saddle-point property of (13), it must be that

$$\mathcal{E}_{p^\star_{\boldsymbol{H}}}\mathcal{E}_{p^\star_{\boldsymbol{Q}}}\Psi(\boldsymbol{Q}, \boldsymbol{H}) \geq \mathcal{E}_{p^\star_{\boldsymbol{H}}}\Psi(\mathcal{E}_{p^\star_{\boldsymbol{Q}}}\boldsymbol{Q}, \boldsymbol{H}).$$

However, by the concavity of the logdet function [39], it holds that

$$\mathcal{E}_{p_{\boldsymbol{Q}}}\Psi(\boldsymbol{Q}, \boldsymbol{H}) \leq \Psi(\mathcal{E}_{p_{\boldsymbol{Q}}}\boldsymbol{Q}, \boldsymbol{H}).$$

Therefore, it must be the case that

$$\mathcal{E}_{p^\star_{\boldsymbol{Q}}}\Psi(\boldsymbol{Q}, \boldsymbol{H}) = \Psi(\mathcal{E}_{p^\star_{\boldsymbol{Q}}}\boldsymbol{Q}, \boldsymbol{H}), \qquad \forall \boldsymbol{H}: p^\star_{\boldsymbol{H}}(\boldsymbol{H}) > 0$$

which, by the strict concavity of the logdet function [39] and (6), implies that

$$\begin{aligned}&\boldsymbol{H}\boldsymbol{Q}_1\boldsymbol{H}^H = \boldsymbol{H}\boldsymbol{Q}_2\boldsymbol{H}^H\\ &\quad\forall \boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{H}: p^\star_{\boldsymbol{Q}}(\boldsymbol{Q}_1) > 0,\ p^\star_{\boldsymbol{Q}}(\boldsymbol{Q}_2) > 0,\ p^\star_{\boldsymbol{H}}(\boldsymbol{H}) > 0.\end{aligned}$$

Thus, we can conclude that $\boldsymbol{Q}_1 = \boldsymbol{Q}_2$ for $\forall \boldsymbol{Q}_1, \boldsymbol{Q}_2$: $p^\star_{\boldsymbol{Q}}(\boldsymbol{Q}_1) > 0$, $p^\star_{\boldsymbol{Q}}(\boldsymbol{Q}_2) > 0$ (note that if the set of used $\boldsymbol{H}$'s have a common null space, by the nature of the saddle point in (13), all the used $\boldsymbol{Q}$'s will be orthogonal to that subspace). In other words, the optimal mixed strategy $p^\star_{\boldsymbol{Q}}$ reduces to a pure strategy $\boldsymbol{Q}^\star$. We can now invoke Theorem 1: if $\boldsymbol{Q}^\star$ were not the uniform power allocation, the set of optimal $\boldsymbol{H}$'s would align their largest singular values with the smallest eigenvalues of $\boldsymbol{Q}^\star$, and the best solution would then be given by the uniform power allocation $\boldsymbol{Q}^\star = P_T/n_T\boldsymbol{I}$.

The problem now is to find a mixed strategy $p^\star_{\boldsymbol{H}}$ so that the saddle-point conditions are satisfied

$$\mathcal{E}_{p_{\boldsymbol{Q}}p^\star_{\boldsymbol{H}}}\Psi(\boldsymbol{Q},\boldsymbol{H}) \leq \mathcal{E}_{p^\star_{\boldsymbol{H}}}\Psi(P_T/n_T\boldsymbol{I}, \boldsymbol{H}) \leq \mathcal{E}_{p_{\boldsymbol{H}}}\Psi(P_T/n_T\boldsymbol{I}, \boldsymbol{H}).$$

Recall that the mixed strategy $p^\star_{\boldsymbol{H}}$ must satisfy

$$\Psi(P_T/n_T\boldsymbol{I}, \boldsymbol{H}_1) = \Psi(P_T/n_T\boldsymbol{I}, \boldsymbol{H}_2)$$

for $\forall \boldsymbol{H}_1, \boldsymbol{H}_2$: $p^\star_{\boldsymbol{H}}(\boldsymbol{H}_1) > 0$, $p^\star_{\boldsymbol{H}}(\boldsymbol{H}_2) > 0$ [16]. Function $\Psi(P_T/n_T\boldsymbol{I}, \boldsymbol{H})$ only depends on $\boldsymbol{H}$ through its singular values and it is minimized by some optimal set $\{\sigma^\star_{H,i}\}$. Therefore, any $p_{\boldsymbol{H}}$ that puts positive probability on channels $\boldsymbol{H}$'s with singular values given by $\{\sigma^\star_{H,i}\}$ and arbitrary right and left singular vectors satisfies the right inequality of the saddle point. We just have to find the appropriate $p_{\boldsymbol{H}}$ such that the left inequality of the saddle point is also satisfied. An example of such an optimal mixed strategy $p^\star_{\boldsymbol{H}}$ is one that puts equal probability on each element of the set

$$\left\{\boldsymbol{H} = \boldsymbol{U}_H\boldsymbol{\Sigma}^\star_H\boldsymbol{P}\boldsymbol{V}^H_H: \boldsymbol{P} \in \Pi\right\}$$

where $\boldsymbol{\Sigma}^\star_H$ contains in the main diagonal the optimum singular values $\{\sigma^\star_{H,i}\}$, $\boldsymbol{U}_H$ and $\boldsymbol{V}_H$ are two arbitrary unitary matrices, and Π is the set of the $n_T!$ different permutation matrices of size $n_T \times n_T$. To check that the left inequality of the saddle point is verified just note that

$$\begin{aligned}&\mathcal{E}_{p^\star_{\boldsymbol{H}}}\Psi(\boldsymbol{Q}, \boldsymbol{H})\\ &\quad= \frac{1}{n_T!}\sum_{\boldsymbol{P}\in\Pi}\log\det\left(\boldsymbol{I} + \boldsymbol{\Sigma}^{\star H}_H\boldsymbol{\Sigma}^\star_H\boldsymbol{P}\boldsymbol{V}^H_H\boldsymbol{Q}\boldsymbol{V}_H\boldsymbol{P}^H\right)\\ &\quad\leq \log\det\left(\boldsymbol{I} + \boldsymbol{\Sigma}^{\star H}_H\boldsymbol{\Sigma}^\star_H\left(\frac{1}{n_T!}\sum_{\boldsymbol{P}\in\Pi}\left(\boldsymbol{P}\tilde{\boldsymbol{Q}}\boldsymbol{P}^H\right)\right)\right)\\ &\quad\leq \log\det\left(\boldsymbol{I} + \boldsymbol{\Sigma}^{\star H}_H\boldsymbol{\Sigma}^\star_H\mathrm{diag}\left(\frac{1}{n_T!}\sum_{\boldsymbol{P}\in\Pi}\left(\boldsymbol{P}\tilde{\boldsymbol{Q}}\boldsymbol{P}^H\right)\right)\right)\\ &\quad\leq \log\det\left(\boldsymbol{I} + P_T/n_T\,\boldsymbol{\Sigma}^{\star H}_H\boldsymbol{\Sigma}^\star_H\right)\\ &\quad= \mathcal{E}_{p^\star_{\boldsymbol{H}}}\Psi(P_T/n_T\,\boldsymbol{I}, \boldsymbol{H})\end{aligned}$$

where $\tilde{\boldsymbol{Q}} \triangleq \boldsymbol{V}^H_H\boldsymbol{Q}\boldsymbol{V}_H$ and $\mathrm{diag}(\boldsymbol{X})$ denotes a diagonal matrix with the diagonal elements of $\boldsymbol{X}$. The first inequality comes from the concavity of the logdet function, the second from Hadamard's inequality [39], [4], and the third from the fact that the diagonal elements of $\frac{1}{n_T!}\sum_{\boldsymbol{P}\in\Pi}\left(\boldsymbol{P}\tilde{\boldsymbol{Q}}\boldsymbol{P}^H\right)$ equal

$$\frac{1}{n_T}\mathrm{Tr}(\tilde{\boldsymbol{Q}}) = \frac{1}{n_T}\mathrm{Tr}(\boldsymbol{Q}) \leq P_T/n_T.$$

It then follows that $\mathcal{E}_{p_{\boldsymbol{Q}}}\mathcal{E}_{p^\star_{\boldsymbol{H}}}\Psi(\boldsymbol{Q}, \boldsymbol{H}) \leq \mathcal{E}_{p^\star_{\boldsymbol{H}}}\Psi(P_T/n_T\boldsymbol{I}, \boldsymbol{H})$. Thus, we have characterized the uniform power allocation $\boldsymbol{Q}^\star = P_T/n_T\boldsymbol{I}$ as a mixed strategy saddle point of the strategic game.

APPENDIX C
PROOF OF LEMMA 3

In this proof, we make use of the theory of majorization. For definitions and further details, the interested reader is referred to [40].

Using the following consequence of Poincaré separation theorem [46, p. 211]:

$$\max_{\boldsymbol{X}^H\boldsymbol{X}=\boldsymbol{I}_k}\mathrm{Tr}\left(\boldsymbol{X}^H\boldsymbol{A}\boldsymbol{X}\right) = \sum_{i=1}^{k}\lambda_i(\boldsymbol{A})$$

where $\boldsymbol{A}$ is an $n \times n$ Hermitian matrix, $\boldsymbol{X} \in \mathbb{C}^{n\times k}$ with $k \leq n$, and $\lambda_i(\cdot)$ denotes the ith eigenvalue in decreasing order, we obtain

$$\begin{aligned}\sum_{i=1}^{k}\lambda_i\left(\sum_j\boldsymbol{R}_j\right) &= \max_{\boldsymbol{X}^H\boldsymbol{X}=\boldsymbol{I}_k}\mathrm{Tr}\left(\boldsymbol{X}^H\left(\sum_j\boldsymbol{R}_j\right)\boldsymbol{X}\right)\\ &\leq \sum_j\max_{\boldsymbol{X}^H\boldsymbol{X}=\boldsymbol{I}_k}\mathrm{Tr}\left(\boldsymbol{X}^H\boldsymbol{R}_j\boldsymbol{X}\right)\\ &= \sum_{i=1}^{k}\sum_j\lambda_i(\boldsymbol{R}_j).\end{aligned}$$

In addition, for $k = n$, we have

$$\sum_{i=1}^{n} \lambda_i\left(\sum_j \boldsymbol{R}_j\right) = \mathrm{Tr}\left(\sum_j \boldsymbol{R}_j\right) = \sum_j \mathrm{Tr}\,(\boldsymbol{R}_j) = \sum_{i=1}^{n}\sum_j \lambda_i\,(\boldsymbol{R}_j)\,.$$

Therefore, we have proved that the sum of the eigenvalues majorizes the eigenvalues of the sum (see [40] for definitions)

$$\left(\lambda_1\left(\sum_j \boldsymbol{R}_j\right), \ldots, \lambda_n\left(\sum_j \boldsymbol{R}_j\right)\right) \prec \left(\sum_j \lambda_1\,(\boldsymbol{R}_j)\,, \ldots, \sum_j \lambda_n\,(\boldsymbol{R}_j)\right).$$

We can now proceed as in [40, Theorem 9.G.3.a] for $J = 2$. Using [40, Theorem 5.A.2.c], we have

$$\left(\log\lambda_1\left(\sum_j \boldsymbol{R}_j\right), \ldots, \log\lambda_n\left(\sum_j \boldsymbol{R}_j\right)\right) \prec^w \left(\log\sum_j \lambda_1\,(\boldsymbol{R}_j)\,, \ldots, \log\sum_j \lambda_n\,(\boldsymbol{R}_j)\right)$$

or, equivalently

$$\sum_{i=k}^{n} \log\lambda_i\left(\sum_j \boldsymbol{R}_j\right) \geq \sum_{i=k}^{n} \log\sum_j \lambda_i\,(\boldsymbol{R}_j)$$

$$\Longleftrightarrow \log\prod_{i=k}^{n} \lambda_i\left(\sum_j \boldsymbol{R}_j\right) \geq \log\prod_{i=k}^{n}\sum_j \lambda_i\,(\boldsymbol{R}_j)\,.$$

In particular, for $k = 1$

$$\det\,(\boldsymbol{R}_1 + \cdots + \boldsymbol{R}_J) \geq \prod_{i=1}^{n} (\lambda_i\,(\boldsymbol{R}_1) + \cdots + \lambda_i\,(\boldsymbol{R}_J))\,. \quad \square$$

ACKNOWLEDGMENT

The authors would like to thank Wei Yu, Wonjong Rhee, Young-Han Kim, and Jonathan Levin for reviewing the paper and for their helpful comments. They also wish to thank the anonymous reviewers for their comments which helped to improve the paper.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun. (ETT)*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999. (See also a previous version of the paper in AT&T Bell Labs Internal Tech. Memo, June 1995).

[2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

[3] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell Syst. Tech. J.*, vol. 53, no. 5, pp. 745–778, May–June 1974.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[5] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: Transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. Inform. Theory*, vol. 42, pp. 137–160, Jan. 1996.

[6] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[7] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1019–1032, Apr. 1999.

[8] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Tech. Univ. Catalonia (UPC), Barcelona, Spain, May 2003.

[9] D. Shiu and J. M. Kahn, "Power allocation strategies for wireless systems with multiple transmit antennas," Internal report of the University of California, Berkley and Lucent Technologies, July 1998.

[10] I. E. Telatar, "Multi-user multi-antenna Gaussian noise fading channels," private communication, 2001.

[11] W. Rhee and J. M. Cioffi, "Ergodic capacity of multi-antenna Gaussian multiple access channels," in *Proc. 35th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 4–7, 2001.

[12] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[13] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007–2019, Sept. 1999.

[14] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.

[15] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[16] M. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.

[17] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1992.

[18] S. A. Kassam and V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, Mar. 1985.

[19] T. Cover, "Universal portfolios," *Math. Finance*, vol. 1, no. 1, pp. 1–29, Jan. 1991.

[20] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.

[21] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin, Germany: Springer-Verlag, 1978.

[22] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[23] S. Boyd and L. Vandenberghe. (2000) Introduction to Convex Optimization with Engineering Applications. Stanford Univ., Stanford, CA, Course notes. [Online]. Available: http://www.stanford.edu/class/ee364.

[24] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.

[25] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273–1289, May 2001.

[26] G. Ganesan and P. Stoica, "Space-time diversity," in *Signal Processing Advances in Communication*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2000, pp. 79–127.

[27] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2691, Oct. 1998.

[28] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[29] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[30] B. Hughes and P. Narayan, "The capacity of a vector Gaussian arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 34, pp. 995–1003, Sept. 1988.

[31] N. M. Blachman, "Communication as a game," in *Proc. IRE WESCON Conf.*, Aug. 1957, pp. 61–66.

[32] J. M. Borden, D. M. Mason, and R. J. McEliece, "Some information theoretic saddlepoints," *SIAM J. Control and Optimiz.*, vol. 23, no. 1, pp. 129–143, Jan. 1985.

[33] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inform. Theory*, vol. 47, pp. 3072–3081, Nov. 2001.

[34] S. Verdú and V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 328–340, Mar. 1984.

[35] T. Basar, "The Gaussian test channel with an intelligent jammer," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 152–157, Jan. 1983.

[36] W. E. Stark and R. J. McEliece, "On the capacity of channels with block memory," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1273–1289, Mar. 1988.

[37] K. Yanagi, "Optimal mutual information for coders and jammers in mismatched communication channels," *SIAM J. Control and Optimiz.*, vol. 31, no. 1, pp. 41–51, Jan. 1993.

[38] N. Chiurtu, B. Rimoldi, and E. Telatar, "On the capacity of multi-antenna Gaussian channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT 2001)*, Washington, DC, June 24–29, 2001, p. 53.

[39] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.

[40] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[41] W. Yu and J. M. Cioffi, "On constant-power water-filling," in *Proc. IEEE Int. Conf. Communications (ICC 2001)*, vol. 6, Helsinki, Finland, June 11–14, 2001, pp. 1665–1669.

[42] P. S. Chow, "Bandwidth optimized digital transmission techniques for spectrally shaped channels with impulse noise," Ph.D. dissertation, Stanford Univ., Stanford, CA, May 1993.

[43] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 733–751, July 1989.

[44] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," in *IEEE Trans. Inform. Theory*. (See also "Iterative water-filling for Gaussian vector multiple access channels" in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 24–29, 2001, p. 322), submitted for publication.

[45] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels—Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.

[46] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley, 1999.