Worst-Case Robust MIMO Transmission With Imperfect Channel Knowledge

Jiaheng Wang, *Student Member, IEEE*, and Daniel P. Palomar, *Senior Member, IEEE*

***Abstract*—In this paper, we consider robust transmit strategies, against the imperfectness of the channel state information at the transmitter (CSIT), for multi-input multi-output (MIMO) communication systems. Following a worst-case deterministic model, the actual channel is assumed to be inside an ellipsoid centered at a nominal channel. The objective is to maximize the worst-case received signal-to-noise ratio (SNR), or to minimize the worst-case Chernoff bound of the error probability, thus leading to a maximin problem. Moreover, we also consider the QoS problem, as a complement of the maximin design, which minimizes the transmit power consumption and meanwhile keeps the received SNR above a given threshold for any channel realization in the ellipsoid. It is shown that, for a general class of power constraints, both the maximin and QoS problems can be equivalently transformed into convex problems, or even further into semidefinite programs (SDPs), thus efficiently solvable by the numerical methods. The most interesting result is that the optimal transmit directions, i.e., the eigenvectors of the transmit covariance, are just the right singular vectors of the nominal channel under some mild conditions. This result leads to a channel-diagonalizing structure, as in the cases of perfect CSIT and statistical CSIT with mean or covariance feedback, and reduces the complicated matrix-valued problem to a scalar power allocation problem. Then we provide the closed-form solution to the resulting power allocation problem.**

***Index Terms*—Convex optimization, imperfect CSIT, maximin, MIMO, SDP, worst-case robust designs.**

I. Introduction

MULTI-INPUT multi-output (MIMO) channels, usually arising from the use of multiple transmit and receive antennas, have been well recognized as an effective and practical way to improve the capacity and reliability of wireless communications [1]–[3]. However, the performance of MIMO systems depends, to a substantial extent, on the quality of the channel state information (CSI) available at the transmitter and receiver. In case of no CSI at the transmitter (CSIT), space-time coding techniques [4]–[8] can be used to harvest the diversity gain. When the transmitter knows the channel perfectly, on the other hand, the full benefit of CSI is exploited by precoding techniques. Instead of these two extreme assumptions on CSIT, which space-time coding and precoding are based on, respectively, a practical communication system typically has to confront an intermediate case, i.e., CSIT is available but imperfect. In this paper, we address the problem of finding robust transmit strategies for MIMO systems by taking into account imperfect CSIT.

With the assumption that perfect CSI is available at both ends of a MIMO link, the joint design of linear precoders and equalizers, under a variety of criteria, has been well studied [9]–[15] . As a very important property, it is observed that the optimal linear transceiver often leads to the eigenmode transmission, according to which the channel matrix is diagonalized by the transceiver and the transmission is effectively carried out through a set of parallel subchannels or eigenmodes. In this case, the optimal transmit directions, i.e., the eigenvectors of the transmit covariance matrix, equal to the right singular vectors of the channel matrix. The available transmit power is then allocated over these eigenmodes in a water-filling fashion [1], [3], [10], [12]–[15] . In particular, [13]–[15] showed that, when the objective function belongs to the class of Schur-concave functions, the eigenmode transmission is always optimal, whereas when the objective function is Schur-convex, the eigenmode transmission is still optimal provided a specific rotation is conducted first on the transmitted symbols. Therefore, in the most general case, the channel matrix is always diagonalized, with or without a rotation of the transmitted symbols. This channel-diagonalizing property is of paramount importance as it simplifies a complicated matrix-valued optimization problem to a simple scalar power allocation problem.

CSI is usually estimated at the receiver by using a training sequence, or blind/semi-blind estimation methods. Obtaining CSIT requires either a feedback channel from the receiver to the transmitter, or exploiting the channel reciprocity (e.g., in time division duplexing (TDD) systems). While it is a reasonable approximation to assume perfect CSI at the receiver (CSI), usually CSIT cannot be assumed perfect, due to many factors such as inaccurate channel estimation, quantization of CSIR, erroneous or outdated feedback, and time delays or frequency offsets between the reciprocal channels. Therefore, the imperfectness of CSIT has to be taken into consideration in any practical communication system. There are two classes of models frequently used to characterize imperfect CSI: the stochastic and the deterministic (or worst-case) models. In the stochastic model, the channel is usually modeled as a complex random matrix with normally distributed elements, and the transmitter knows the mean and/or the covariance, i.e., the slowly-varying channel statistics that can be well estimated. The system design is then based on optimizing the average or outage performance. On the

Manuscript received August 11, 2008; accepted January 06, 2009. First published April 21, 2009; current version published July 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Biao Chen. This work was supported by the Hong Kong RGC 618008 research grant.

The authors are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: jhwang@ust.hk; palomar@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2021451

other hand, the deterministic model assumes that the instantaneous channel, though not exactly known, lies in a known set of possible values. The size of the set represents the amount of uncertainty on the channel, i.e., the bigger the set is, the more uncertainty there is. In this case, the robust design aims at optimizing the worst-case performance, which leads to a maximin or minimax formulation, and achieves a guaranteed performance level for any channel realization in the set.

The stochastic model was first introduced into multi-antenna communications for maximizing the ergodic capacity of a multiple-input single-output (MISO) channel with mean or covariance feedback [16], [17], which was later generalized to the MIMO case [18], [19]. Other efforts focused on combining linear precoding with a space-time block code (STBC) by minimizing an upper bound of the pairwise error probability (PEP) or the symbol error rate (SER), with the help of the channel mean [20], [21], or the channel covariance [22], [23], or both [24], [25]. Recently, a linear transceiver design based on a general cost function using the channel mean and covariance was proposed in [26]. Statistical CSIT was also considered for transmitter optimization in a MIMO multi-access channel (MAC) [27], [28]. Interestingly, it turns out that, in most situations [16]–[24], [27], [28] , the eigenmode transmission is still optimal, in the sense that the optimal transmit directions correspond to either the singular vectors of the channel mean in the mean feedback case, or the eigenvectors of the channel covariance matrix in the covariance feedback case. In other words, the channel-diagonalizing property still holds for some particular kinds of statistical CSIT.

The worst-case design based on the deterministic model has a long history in signal processing community [29], [30]. This philosophy has been successfully applied to designing robust beamformers against mismatches in the array response and the covariance matrix [31]–[33], and robustly estimating parameters from a linear model subject to the model uncertainty [34], [35]. As for MIMO channels, both [34] and [36] considered imperfect CSIR and derived the linear equalizers that minimize the worst-case mean square error (MSE), while [37] proposed a robust linear receiver for a MIMO MAC system equipped with an orthogonal STBC (OSTBC) using an approach similar to that in [32]. Regarding robust precoding, [38] proved the optimality of the uniform power allocation for the compound capacity [39] of both a point-to-point MIMO channel and a MIMO MAC, provided the channel belongs to an isotropically unconstrained set. In [40] the authors considered optimizing the compound capacity too, but of a rank-one Ricean MIMO channel. A more practical model, which assumes that the actual channel lies in the neighborhood of a nominal value, was adopted for robust precoder designs in a point-to-point MIMO channel [41]–[43] and a MISO broadcasting channel (BC) [44], [45]. However, [41] omitted the transmit power constraint, an indispensable requirement in practice, to make the problem solvable. Meanwhile, in [42] and [43], the authors simplified their problems by setting the transmit directions equal to the right singular vectors of the nominal channel, but without knowing whether they are optimal or not. So far, it is not clear whether the eigenmode transmission is optimal for the worst-case precoder design. This is precisely one goal of this paper.

In this paper, we consider robust transmit strategies for MIMO communication systems, based on the worst-case optimization, using the deterministic model of imperfect CSIT that is similar to (but more general than) that used in [40], [41], [43], [45]. To be more specific, while assuming perfect CSIR, for the transmitter, we assume that the actual channel lies within an ellipsoid centered at a nominal channel. The design objective, in terms of robustness, is to maximize the worst-case received signal-to-noise ratio (SNR) [42], [43], or to minimize the worst-case Chernoff bound of the PEP [21], [23]–[25] if a STBC is used, thus leading to a maximin problem. Moreover, we also consider the so-called QoS problem formulation, as a complement of the maximin design, whose objective is to minimize the power consumption at the transmitter and meanwhile keep the received SNR above a given threshold for any channel realization in the ellipsoid. Our first main result is that, for a general class of power constraints, both the maximin and QoS problems can be equivalently transformed into convex optimization problems [46] that can be efficiently solved in polynomial time using, e.g., an interior-point method [47]. For some power constraints, the problems simplify further to semidefinite programs (SDPs) [48], [49], a very tractable form of convex optimization.

The imperfect CSIT model considered in this paper can be regarded as the deterministic analogue of the stochastic model of mean feedback, with the nominal channel (the center of the ellipsoid) acting as the counterpart of the channel mean. In light of the optimality of the eigenmode transmission (over the channel mean) in the mean feedback case [20], [24], [27], [28], one may wonder whether it is optimal in the deterministic model as well. As the second main contribution of this paper, we answer affirmatively this question by proving that, for the worst-case design, the optimal transmit directions are the right singular vectors of the nominal channel under some mild conditions. As a special case of our framework, it follows that the transmit directions imposed in [42] (for the spherical uncertainty region) and [43] are actually optimal. Consequently, the complicated matrix-valued problems can be simplified to a scalar power allocation problem without any loss of optimality. Our third main result consists of providing the closed-form solution to the resulting power allocation problem.

The paper is organized as follows. The signal model and the problem formulation are introduced in Section II, where we also discuss the choice of the performance measure, the deterministic model of imperfect CSIT and the power constraints. Section III provides the convex reformulation of the maximin problem with a general power constraint. In Section IV, we prove the optimality of the eigenmode transmission, under some conditions, for the maximin problem. Then the closed-form solution to the resulting power allocation problem is derived in Section V. Section VI addresses the QoS problem, and the numerical results are provided in Section VII. Finally, we conclude with Section VIII.

The following notations are used. Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and standard lower-case letters denote scalars. $\mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$ denote the set of $m \times n$ matrices with real- and complex-valued entries, respectively. $\mathbb{S}_+^n$ denotes the ensemble of all $n \times n$ positive semidefinite matrices. $[\mathbf{X}]_{ij}$ represents the

$(i\text{th}, j\text{th})$ element of matrix $\mathbf{X}$. By $\mathbf{X} \succeq 0$ or $\mathbf{X} \succ 0$, we mean that $\mathbf{X}$ is a Hermitian positive semidefinite or definite matrix, respectively. The operators $(\cdot)^H$, $(\cdot)^{-1}$, $\mathrm{Tr}(\cdot)$ and $|\cdot|$ denote the Hermitian, inverse, trace, and determinant operations, respectively. The maximum eigenvalue of a Hermitian matrix is represented by $\lambda_{\max}(\cdot)$. $\|\cdot\|$ denotes a general norm of a matrix as well as the Euclidean norm of a vector. The Frobenius norm and spectral norm of a matrix are denoted by $\|\cdot\|_F$ and $\|\cdot\|_2$, respectively. $\otimes$ represents the Kronecker product operator and $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex value.

II. Problem Statement

A. Signal Model

We consider a point-to-point communication system equipped with N transmit and M receive antennas. Mathematically, with $\mathbf{H} \in \mathbb{C}^{M\times N}$ being the channel matrix, the system can be represented by a linear model as

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

where $\mathbf{x} \in \mathbb{C}^N$ is the transmitted signal vector, $\mathbf{y} \in \mathbb{C}^M$ is the received signal vector, and $\mathbf{n} \in \mathbb{C}^M$ is a circularly symmetric complex Gaussian noise vector with zero mean and covariance matrix $\sigma_n^2\mathbf{I}$, i.e., $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_n^2\mathbf{I}\right)$.

Let $\mathbf{Q} = E\left[\mathbf{x}\mathbf{x}^H\right]$ be the covariance matrix of the transmitted signal vector, and denote the eigenvalue decomposition (EVD) of $\mathbf{Q}$ by $\mathbf{Q} = \mathbf{U}_q\boldsymbol{\Lambda}_q\mathbf{U}_q^H$, where $\mathbf{U}_q$ is a unitary matrix and $\boldsymbol{\Lambda}_q$ is a diagonal matrix with the diagonal elements $\{p_i\}$, without loss of generality (w.l.o.g.), sorted in decreasing order, i.e., $p_1 \geq \cdots \geq p_N$. Once $\mathbf{Q}$ is known, a precoding matrix $\mathbf{F} \in \mathbb{C}^{N\times N}$ can be obtained by decomposing $\mathbf{Q}$ as $\mathbf{Q} = \mathbf{F}\mathbf{F}^H$. One possible choice is $\mathbf{F} = \mathbf{U}_q\boldsymbol{\Lambda}_q^{1/2}$, where $\mathbf{U}_q$ contains the transmit directions, i.e., the eigenvectors of $\mathbf{Q}$, and p_i has the meaning of the power allocated to the ith eigenmode. Note that other decompositions, such as the Cholesky factorization, are also possible. The transmitted signal vector $\mathbf{x}$ is then obtained through the linear mapping $\mathbf{x} = \mathbf{F}\mathbf{s}$, where $\mathbf{s} \in \mathbb{C}^N$ is a vector of symbols that are zero-mean, unit-power, and uncorrelated, i.e., $E\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{I}$.

The optimal transmit covariance $\mathbf{Q}$ or precoder $\mathbf{F}$, under various criteria, has already been found in [9]–[14] when the channel $\mathbf{H}$ is perfectly known at the transmitter. However, due to many practical issues such as channel estimation errors, feedback errors, quantization and delays, CSIT is usually imperfect and subject to some uncertainty. To model the uncertainty of the channel, we assume that $\mathbf{H}$ belongs to a known set $\mathcal{H}$ of possible states but otherwise unknown, which is usually called the *compound channel* [39]. In this model, the performance measure, denoted by $\Psi(\mathbf{Q}, \mathbf{H})$, is a function of both the transmit covariance matrix and the channel. In this paper, we are interested in the robust design that provides the optimal $\mathbf{Q}$ for the worst channel, hence leading to the following maximin problem:

$$\max_{\mathbf{Q}\in\mathcal{Q}} \min_{\mathbf{H}\in\mathcal{H}} \Psi(\mathbf{Q}, \mathbf{H}) \tag{2}$$

where $\mathcal{Q}$ models the constraints at the transmitter, e.g., power constraints. Such a maximin problem can be interpreted a two-player zero-sum strategic game [38], [50] in which the two players, the transmitter and the channel, are competing against each other. The Nash equilibrium [38], [50] of this game is given by the saddle point $(\mathbf{Q}^\star, \mathbf{H}^\star)$ that satisfies

$$\Psi(\mathbf{Q}, \mathbf{H}^\star) \leq \Psi(\mathbf{Q}^\star, \mathbf{H}^\star) \leq \Psi(\mathbf{Q}^\star, \mathbf{H}) \tag{3}$$

and also serves as the solution to the maximin problem. Nevertheless, existence of a Nash equilibrium depends on the properties of $\Psi(\mathbf{Q}, \mathbf{H})$, $\mathcal{Q}$ and $\mathcal{H}$.

B. Performance Measure

In this paper, we adopt the following performance measure:

$$\Psi(\mathbf{Q}, \mathbf{H}) = \mathrm{Tr}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \tag{4}$$

which corresponds to the following physical interpretations:

1) Received SNR: From the model (1), it is easy to see that the SNR at the receiver is

$$\frac{E\left[\|\mathbf{H}\mathbf{x}\|^2\right]}{E\left[\|\mathbf{n}\|^2\right]} = \frac{1}{M\sigma_n^2}\mathrm{Tr}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \tag{5}$$

which, up to a scaling factor, is equal to $\Psi(\mathbf{Q}, \mathbf{H})$.

2) Chernoff Bound of PEP: When a STBC is used at the transmitter, the system can be represented by

$$\mathbf{Y} = \mathbf{H}\mathbf{F}\mathbf{C} + \mathbf{N} \tag{6}$$

where $\mathbf{C} \in \mathbb{C}^{N\times T}$ is a block codeword of the STBC, $\mathbf{Y} \in \mathbb{C}^{M\times T}$ contains as its columns the received signal vectors of T time slots, and $\mathbf{N} \in \mathbb{C}^{M\times T}$ is the complex white Gaussian noise with power σ_n^2 per time and spatial dimension. It is assumed that the channel remains unchanged during the transmission of a codeword. With a maximum-likelihood (ML) detector at the receiver, the PEP $\mathrm{Prob}(\mathbf{C}_i \to \mathbf{C}_j)$, i.e., the probability that a transmitted codeword $\mathbf{C}_i$ is incorrectly decoded as another codeword $\mathbf{C}_j$, is given by [5], [25]

$$\begin{aligned}\mathrm{Prob}(\mathbf{C}_i \to \mathbf{C}_j) &= Q\left(\sqrt{\frac{\|\mathbf{H}\mathbf{F}(\mathbf{C}_i - \mathbf{C}_j)\|_F^2}{2\sigma_n^2}}\right)\\ &\leq \exp\left(-\frac{\|\mathbf{H}\mathbf{F}(\mathbf{C}_i - \mathbf{C}_j)\|_F^2}{4\sigma_n^2}\right)\\ &= \exp\left(-\frac{1}{4\sigma_n^2}B(\boldsymbol{\Theta}_{i,j})\right)\end{aligned} \tag{7}$$

where $Q(\cdot)$ denotes the standard $\mathbf{Q}$ function [51], $\boldsymbol{\Theta}_{i,j} \triangleq (\mathbf{C}_i - \mathbf{C}_j)(\mathbf{C}_i - \mathbf{C}_j)^H$ is the codeword distance product matrix and $B(\boldsymbol{\Theta}_{i,j}) = \mathrm{Tr}\left(\mathbf{H}\mathbf{F}\boldsymbol{\Theta}_{i,j}\mathbf{F}^H\mathbf{H}^H\right)$. Consequently, minimizing the Chernoff bound of the PEP is equivalent to maximizing $B(\boldsymbol{\Theta}_{i,j})$.

Note that $B(\boldsymbol{\Theta}_{i,j})$ depends on the specific codeword pair $(\mathbf{C}_i, \mathbf{C}_j)$ through $\boldsymbol{\Theta}_{i,j}$. In order to take into account all codewords of the codebook, we can properly choose a $\boldsymbol{\Theta}$ to replace the individual $\boldsymbol{\Theta}_{i,j}$ as in [25], so that a common $B(\boldsymbol{\Theta})$ is maximized. For an orthogonal STBC (OSTBC), each $\boldsymbol{\Theta}_{i,j}$ is a scaled version of the identity matrix, and thus can be expressed as $\theta_{i,j}\mathbf{I}$ with a scalar $\theta_{i,j}$. Then $\boldsymbol{\Theta}$ can be $\theta_{\min}\mathbf{I}$ or $\theta_{\text{avg}}\mathbf{I}$, where $\theta_{\min}$ and θ_{avg} are the minimum and average, respectively, of all $\theta_{i,j}$, $i \neq j$. Therefore, $B(\boldsymbol{\Theta})$ coincides with $\Psi(\mathbf{Q}, \mathbf{H})$. When a non-orthogonal STBC is used, e.g., a quasi-orthogonal STBC (QSTBC) [8], $\boldsymbol{\Theta}_{i,j}$ is not necessarily proportional to the identity matrix. In this case, some relaxation method, as was proposed in [25], can be used to approximate $\boldsymbol{\Theta}_{i,j}$ by a scaled identity matrix, and hence $\boldsymbol{\Theta}$ can be similarly obtained like an OSTBC.

3) Approximation of Mutual Information at Low SNR: If a Gaussian code is used at the transmitter, the mutual information of the MIMO channel in (1) is $\log\det(\mathbf{I} + 1/\sigma_n^2 \mathbf{H}\mathbf{Q}\mathbf{H}^H)$ [1], which can be expanded as [52]

$$\log\det\left(\mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) = \frac{1}{\sigma_n^2}\mathrm{Tr}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) + o\left(\frac{1}{\sigma_n^2}\|\mathbf{H}\mathbf{Q}\mathbf{H}^H\|\right). \quad (8)$$

Therefore, the mutual information can be approximated by a scaled version of $\Psi(\mathbf{Q}, \mathbf{H})$ at low SNR, and in this case the maximin formulation for $\Psi(\mathbf{Q}, \mathbf{H})$ will provide the solution to the compound capacity [39] of the MIMO channel.

4) Approximation of MSE at Low SNR: With perfect CSIR, the optimal linear decoder (or equalizer), under the minimum mean square error (MMSE) criterion, is just the Wiener filter. The resulting MSE is [13]–[15]

$$\mathrm{MSE} = \mathrm{Tr}\left[\left(\mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F}\right)^{-1}\right] \quad (9)$$

which, by using $\mathrm{Tr}[(\mathbf{I} + \mathbf{A})^{-1}] = \mathrm{Tr}(\mathbf{I}) - \mathrm{Tr}(\mathbf{A}) + o(\|\mathbf{A}\|)$, can be expanded as

$$\mathrm{MSE} = N - \frac{1}{\sigma_n^2}\mathrm{Tr}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) + o\left(\frac{1}{\sigma_n^2}\|\mathbf{H}\mathbf{Q}\mathbf{H}^H\|\right). \quad (10)$$

Ignoring the constant term, at low SNR, minimizing the MSE corresponds to maximizing $\Psi(\mathbf{Q}, \mathbf{H})$.

C. Uncertainty Region

To model imperfect CSI deterministically, we assume that the actual channel $\mathbf{H}$ lies in the neighborhood of a nominal channel $\hat{\mathbf{H}}$ that is known to the transmitter. In particular, we consider that $\mathbf{H}$ belongs to the uncertainty region $\mathcal{H} \triangleq \{\mathbf{H} : \|\mathbf{H} - \hat{\mathbf{H}}\| \leq \varepsilon\}$, which is an ellipsoid centered at $\hat{\mathbf{H}}$ with the radius ε (also known to the transmitter). Furthermore, by defining the channel error

$$\boldsymbol{\Delta} \triangleq \mathbf{H} - \hat{\mathbf{H}}, \quad (11)$$

$\mathbf{H} \in \mathcal{H}$ can be equally described by $\boldsymbol{\Delta} \in \mathcal{E} = \{\boldsymbol{\Delta} : \|\boldsymbol{\Delta}\| \leq \varepsilon\}$. In this paper, we consider $\mathcal{E}$ defined by the weighted Frobenius norm $\|\cdot\|_F^{\mathbf{T}}$, i.e.,

$$\mathcal{E} \triangleq \left\{\boldsymbol{\Delta} : \|\boldsymbol{\Delta}\|_F^{\mathbf{T}} \leq \varepsilon\right\} = \left\{\boldsymbol{\Delta} : \mathrm{Tr}\left(\boldsymbol{\Delta}\mathbf{T}\boldsymbol{\Delta}^H\right) \leq \varepsilon^2\right\} \quad (12)$$

where $\mathbf{T}$ is a known positive definite matrix. The usefulness of (12) to model different kinds of channel uncertainty has been justified in [33], [42]. Note that $\mathcal{E}$ includes the deterministic model adopted in [41], [42] (for the spherical uncertainty region) and [43] as a special case of $\mathbf{T} = \mathbf{I}$.

D. Power Constraint

Regarding the power constraints at the transmitter, we will start by considering a very general constraint $\mathbf{Q} \in \mathcal{Q}$, where $\mathcal{Q} \subseteq \mathbb{S}_+^N$ is a nonempty compact convex set. In other words, $\mathbf{Q}$ must be positive semidefinite and within a nonempty compact convex set. This constraint includes all commonly used power constraints as special cases, which will be considered later. Here we list some of them.

1) Sum power constraint:

$$\mathcal{Q}_1 \triangleq \{\mathbf{Q} : \mathbf{Q} \succeq 0,\ \mathrm{Tr}(\mathbf{Q}) \leq P_s\}.$$

2) Maximum power constraint:

$$\mathcal{Q}_2 \triangleq \{\mathbf{Q} : \mathbf{Q} \succeq 0,\ \lambda_{\max}(\mathbf{Q}) \leq P_m\}.$$

3) Per-antenna power constraint:

$$\mathcal{Q}_3 \triangleq \{\mathbf{Q} : \mathbf{Q} \succeq 0, \max_i [\mathbf{Q}]_{ii} \leq P_a\}$$
$$\mathcal{Q}_4 \triangleq \{\mathbf{Q} : \mathbf{Q} \succeq 0, [\mathbf{Q}]_{ii} \leq P_{a,i}, i = 1, \ldots, N\}.$$

It is easily seen that $\mathcal{Q}$ can be any intersection among the above constraint sets. We will first consider the solution to a general set $\mathcal{Q}$, but some constraints, such as the sum and maximum power constraints, may lead to a simple closed-form solution.

III. Convex Reformulation of the Maximin Problem

Using the performance measure of (4) and the definition of the channel error in (11), the maximin problem (2) can be expressed as

$$\max_{\mathbf{Q}\in\mathcal{Q}} \min_{\boldsymbol{\Delta}\in\mathcal{E}} \mathrm{Tr}\left[\left(\hat{\mathbf{H}} + \boldsymbol{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}} + \boldsymbol{\Delta}\right)^H\right]. \quad (13)$$

In this section, we consider the general power constraint $\mathbf{Q} \in \mathcal{Q}$ where $\mathcal{Q} \subseteq \mathbb{S}_+^N$ is a nonempty compact convex set. Since the objective function in (13) is concave (in fact linear) in $\mathbf{Q}$ for fixed $\boldsymbol{\Delta}$, and convex (and quadratic) in $\boldsymbol{\Delta}$ for fixed $\mathbf{Q}$, and the two nonempty compact convex sets $\mathcal{Q}$ and $\mathcal{E}$ are decoupled, there always exists a saddle point [53], which is a solution to the maximin problem (13). Now that we know the existence of a solution, the question is how to characterize and compute it, which is the most difficult part. We will show that this can be achieved by reformulating (13) as a convex problem, which can be globally solved by efficient polynomial-time numerical algorithms, e.g., an interior-point method [47]. Moreover, when some specific power constraints are considered, the resulting convex problem simplifies to an SDP [48], [49].

Proposition 1: Let $\mathcal{Q} \subseteq \mathbb{S}_+^N$ be a nonempty compact set, and $\mathcal{E}$ be defined as in (12). Then, the maximin problem (13) is equivalent to the following problem:

$$\begin{aligned}&\underset{\mathbf{Q},\mu,\mathbf{Z}}{\text{minimize}} \quad \operatorname{Tr}\left[(\mathbf{Z}-\mathbf{Q})\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right]+\varepsilon^2\mu\\ &\text{subject to} \quad \mathbf{Q}\in\mathcal{Q},\ \mu\geq 0\\ &\begin{bmatrix}\mathbf{Z} & \mathbf{Q}\\ \mathbf{Q} & \mathbf{Q}+\mu\mathbf{T}\end{bmatrix}\succeq 0.\end{aligned} \tag{14}$$

Proof: The proof is based on *S-procedure*, a powerful tool for robust quadratic problems.

Lemma 1 (S-Procedure [54]): Let $f_k(\mathbf{x})$, $k=1,2$, be defined as

$$f_k(\mathbf{x})=\mathbf{x}^H\mathbf{A}_k\mathbf{x}+2\text{Re}\left\{\mathbf{b}_k^H\mathbf{x}\right\}+c_k$$

where $\mathbf{A}_k=\mathbf{A}_k^H\in\mathbb{C}^{n\times n}$, $\mathbf{b}_k\in\mathbb{C}^n$ and $c_k\in\mathbb{R}$. Then, the implication $f_1(\mathbf{x})\geq 0\Rightarrow f_2(\mathbf{x})\geq 0$ holds if and only if there exists $\mu\geq 0$ such that

$$\begin{bmatrix}\mathbf{A}_2 & \mathbf{b}_2\\ \mathbf{b}_2^H & c_2\end{bmatrix}-\mu\begin{bmatrix}\mathbf{A}_1 & \mathbf{b}_1\\ \mathbf{b}_1^H & c_1\end{bmatrix}\succeq 0,$$

provided there exists a point $\hat{\mathbf{x}}$ with $f_1(\hat{\mathbf{x}})>0$.

Lemma 2 (Schur's Complement [55]): Let

$$\mathbf{M}=\begin{bmatrix}\mathbf{A} & \mathbf{B}^H\\ \mathbf{B} & \mathbf{C}\end{bmatrix}$$

be a Hermitian matrix. Then, $\mathbf{M}\succeq 0$ if and only if $\mathbf{A}-\mathbf{B}^H\mathbf{C}^{-1}\mathbf{B}\succeq 0$ (assuming $\mathbf{C}\succ 0$), or $\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\succeq 0$ (assuming $\mathbf{A}\succ 0$).[1]

We first transform the maximin problem (13) into

$$\begin{aligned}&\underset{\mathbf{Q}\in\mathcal{Q},t}{\text{maximize}} \quad t\\ &\text{subject to} \quad \operatorname{Tr}\left[\left(\hat{\mathbf{H}}+\boldsymbol{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}}+\boldsymbol{\Delta}\right)^H\right]\geq t,\\ &\qquad\qquad \forall\boldsymbol{\Delta}:\operatorname{Tr}\left(\boldsymbol{\Delta}\mathbf{T}\boldsymbol{\Delta}^H\right)\leq\varepsilon^2.\end{aligned} \tag{15}$$

Let $\boldsymbol{\delta}=\text{vec}(\boldsymbol{\Delta})$ and $\hat{\mathbf{h}}=\text{vec}(\hat{\mathbf{H}})$. It can be easily verified that

$$\operatorname{Tr}\left(\boldsymbol{\Delta}\mathbf{T}\boldsymbol{\Delta}^H\right)=\boldsymbol{\delta}^H\left(\mathbf{T}^T\otimes\mathbf{I}_M\right)\boldsymbol{\delta} \tag{16}$$

$$\begin{aligned}\operatorname{Tr}\left[\left(\hat{\mathbf{H}}+\boldsymbol{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}}+\boldsymbol{\Delta}\right)^H\right]&=\boldsymbol{\delta}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\boldsymbol{\delta}\\ &+2\text{Re}\left\{\hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\boldsymbol{\delta}\right\}+\hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}.\end{aligned} \tag{17}$$

Hence the robust constraint in (15) can be expressed as

$$\begin{aligned}&\boldsymbol{\delta}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\boldsymbol{\delta}+2\text{Re}\left\{\hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\boldsymbol{\delta}\right\}\\ &\quad+\hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}-t\geq 0,\\ &\quad\forall\boldsymbol{\delta}:-\boldsymbol{\delta}^H\left(\mathbf{T}^T\otimes\mathbf{I}_M\right)\boldsymbol{\delta}+\varepsilon^2\geq 0.\end{aligned} \tag{18}$$

According to Lemma 1, (18) holds if and only if there exists $\mu\geq 0$ such that

$$\begin{bmatrix}(\mathbf{Q}+\mu\mathbf{T})^T\otimes\mathbf{I}_M & \left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}\\ \hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right) & \hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}-\varepsilon^2\mu-t\end{bmatrix}\succeq 0. \tag{19}$$

[1]Note that if $\mathbf{C}$ (or $\mathbf{A}$) is non-positive definite, there is a more general version of the Schur's complement involving the pseudoinverse [55].

Based on (19), we will investigate the following two cases: $\mu>0$ and $\mu=0$.

For $\mu>0$, $\mathbf{Q}+\mu\mathbf{T}$ is invertible. Using Lemma 2, (19) amounts to

$$\begin{aligned}&\hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}-\varepsilon^2\mu-t-\hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\\ &\quad\times\left[(\mathbf{Q}+\mu\mathbf{T})^T\otimes\mathbf{I}_M\right]^{-1}\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}\geq 0\end{aligned} \tag{20}$$

which can be simplified to

$$\operatorname{Tr}\left(\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H\right)-\operatorname{Tr}\left[\hat{\mathbf{H}}\mathbf{Q}(\mathbf{Q}+\mu\mathbf{T})^{-1}\mathbf{Q}\hat{\mathbf{H}}^H\right]-\varepsilon^2\mu\geq t. \tag{21}$$

Therefore, (15) is equivalent to the following problem:

$$\underset{\mathbf{Q}\in\mathcal{Q},\mu>0}{\text{maximize}}\ \operatorname{Tr}\left(\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H\right)-\operatorname{Tr}\left[\hat{\mathbf{H}}\mathbf{Q}(\mathbf{Q}+\mu\mathbf{T})^{-1}\mathbf{Q}\hat{\mathbf{H}}^H\right]-\varepsilon^2\mu. \tag{22}$$

Rewrite (22) as a minimization problem

$$\underset{\mathbf{Q}\in\mathcal{Q},\mu>0}{\text{minimize}}\ \operatorname{Tr}\left[\mathbf{Q}(\mathbf{Q}+\mu\mathbf{T})^{-1}\mathbf{Q}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right]-\operatorname{Tr}\left(\mathbf{Q}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)+\varepsilon^2\mu \tag{23}$$

which is equivalent to

$$\begin{aligned}&\underset{\mathbf{Q}\in\mathcal{Q},\mu>0}{\text{minimize}} \quad \operatorname{Tr}\left(\mathbf{Z}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)-\operatorname{Tr}\left(\mathbf{Q}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)+\varepsilon^2\mu\\ &\text{subject to} \quad \mathbf{Q}(\mathbf{Q}+\mu\mathbf{T})^{-1}\mathbf{Q}\preceq\mathbf{Z}.\end{aligned} \tag{24}$$

Using Lemma 2 again, (24) can be equivalently transformed into

$$\begin{aligned}&\underset{\mathbf{Q}\in\mathcal{Q},\mu>0}{\text{minimize}} \quad \operatorname{Tr}\left[(\mathbf{Z}-\mathbf{Q})\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right]+\varepsilon^2\mu\\ &\text{subject to} \quad \begin{bmatrix}\mathbf{Z} & \mathbf{Q}\\ \mathbf{Q} & \mathbf{Q}+\mu\mathbf{T}\end{bmatrix}\succeq 0.\end{aligned} \tag{25}$$

For $\mu=0$, it follows from (19) that

$$\begin{aligned}&\mathbf{Y}=\begin{bmatrix}\mathbf{Q}^T\otimes\mathbf{I}_M & \left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}\\ \hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right) & \hat{\mathbf{h}}^H\left(\mathbf{Q}^T\otimes\mathbf{I}_M\right)\hat{\mathbf{h}}-t\end{bmatrix}\succeq 0\\ &\quad\Rightarrow\left[-\hat{\mathbf{h}}^H\quad 1\right]\mathbf{Y}\begin{bmatrix}-\hat{\mathbf{h}}\\ 1\end{bmatrix}=-t\geq 0\end{aligned} \tag{26}$$

which means the maximum value of the problem (15) is zero (observe that $t=0$ can always be achieved in (15) regardless of $\mathbf{Q}\in\mathcal{Q}$). We can conveniently include this case into (25) by extending $\mu>0$ to $\mu\geq 0$. To prove the equivalence between the extended problem and the original problem, we only need to show that the optimal value of (25) is zero when $\mu=0$. Clearly, when $\mu=0$, a zero objective value can be achieved by setting $\mathbf{Z}=\mathbf{Q}$, which satisfies the constraint of (25). The remaining question is whether $\mathbf{Z}=\mathbf{Q}$ is optimal for (25) with $\mu=0$. Supposing that $\mathbf{Z}=\mathbf{Q}$ is not optimal, there must be some $\hat{\mathbf{Z}}$ and $\hat{\mathbf{Q}}$ resulting in a negative objective value, which in turn implies that there exists some vector $\mathbf{c}$ such that $\mathbf{c}^H\hat{\mathbf{Z}}\mathbf{c}<\mathbf{c}^H\hat{\mathbf{Q}}\mathbf{c}$. On the other hand, however, the constraint in (25) says that

$$\begin{aligned}\begin{bmatrix}\hat{\mathbf{Z}} & \hat{\mathbf{Q}}\\ \hat{\mathbf{Q}} & \hat{\mathbf{Q}}\end{bmatrix}\succeq 0\Rightarrow\left[\mathbf{c}^H\quad -\mathbf{c}^H\right]\begin{bmatrix}\hat{\mathbf{Z}} & \hat{\mathbf{Q}}\\ \hat{\mathbf{Q}} & \hat{\mathbf{Q}}\end{bmatrix}\begin{bmatrix}\mathbf{c}\\ -\mathbf{c}\end{bmatrix}\\ =\mathbf{c}^H\hat{\mathbf{Z}}\mathbf{c}-\mathbf{c}^H\hat{\mathbf{Q}}\mathbf{c}\geq 0\end{aligned} \tag{27}$$

which is in contradiction with $\mathbf{c}^H\hat{\mathbf{Z}}\mathbf{c} < \mathbf{c}^H\hat{\mathbf{Q}}\mathbf{c}$. Therefore, $\mathbf{Z} = \mathbf{Q}$ is optimal for (25) when $\mu = 0$ and the extended problem (25) with $\mu \geq 0$ is equivalent to the original problem. ■

Remark 1: From Propositions 1, if $\mathcal{Q}$ is a nonempty compact convex set, then (14) is a convex problem that can be efficiently solved. As a tractable form of convex optimization, an SDP consists of an affine objective function and constraints defined by linear matrix inequalities (LMIs) [48], [49], [54] . Thus, it is straightforward to see that (14) becomes an SDP if $\mathcal{Q} = \mathcal{Q}_1$ or $\mathcal{Q} = \mathcal{Q}_4$. When $\mathcal{Q} = \mathcal{Q}_2$, one can easily transform (14) into an SDP, since the constraint $\lambda_{\max}(\mathbf{Q}) \leq P_m$ is equivalent to $\mathbf{Q} \preceq P_m\mathbf{I}$. Similarly, when $\mathcal{Q} = \mathcal{Q}_3$, the constraint $\max_i[\mathbf{Q}]_{ii} \leq P_{a,i}$ can be replaced by $[\mathbf{Q}]_{ii} \leq P_{a,i}$, $i = 1, \ldots, N$. In fact, for any combination of $\mathcal{Q}_1$, $\mathcal{Q}_2$, $\mathcal{Q}_3$, and $\mathcal{Q}_4$ (any intersection among them), the problem (14) becomes an SDP.

IV. Optimal Transmit Directions

Although the jointly optimal transmit directions and power allocation can be achieved by decomposing the optimal transmit covariance matrix, which can be found through the method in the previous section, one may wonder whether they can be obtained independently. Even further, one may ask whether there exist optimal channel-diagonalizing transmit directions, just like the perfect CSIT case or some statistical CSIT cases, that can reduce the complicated matrix-valued optimization to a simple power allocation problem. As an important result, in this section, we prove that the optimal transmit directions are just the right singular vectors of the nominal channel matrix, provided some conditions are satisfied.

Before stating our results, it is worth pointing out that the optimum value of the maximin problem (13) is zero if and only if $\mathbf{\Delta} = -\hat{\mathbf{H}}$, which can only happen if $\varepsilon \geq \|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$, i.e., when the channel error is very large. In such a case, there is no guarantee of performance. Hence, to avoid the trivial solution, we assume that $\varepsilon < \|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$, which is reasonable since large channel errors usually leads to unacceptable degradation in performance. Then, we start by giving the following proposition.

Proposition 2: Let $\mathcal{Q} \subseteq \mathbb{S}_+^N$ be a nonempty compact set, and $\mathcal{E}$ be defined as in (12). Then, the maximin problem (13) is equivalent to

$$\underset{\mathbf{Q}\in\mathcal{Q},\mu\geq 0}{\text{maximize}} \;\; \mu\text{Tr}\left[\mathbf{Q}\left(\mathbf{Q}+\mu\mathbf{T}\right)^{-1}\mathbf{T}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right] - \varepsilon^2\mu \qquad (28)$$

where the objective is defined as 0 for $\mu = 0$.

Proof: Note that there are two methods to prove this result, one of which is to exploit the equivalent formulation (22) in the proof of Proposition 1. The other method, which can offer more insights, is provided in Appendix A. ■

Let the EVD of $\mathbf{Q}$ be $\mathbf{Q} = \mathbf{U}_q\mathbf{\Lambda}_q\mathbf{U}_q^H$ with the eigenvalues $p_1 \geq \cdots \geq p_N$, the EVD of $\hat{\mathbf{H}}^H\hat{\mathbf{H}}$ be $\hat{\mathbf{H}}^H\hat{\mathbf{H}} = \mathbf{U}_h\mathbf{\Lambda}_h\mathbf{U}_h^H$ with the eigenvalues $\gamma_1 \geq \cdots \geq \gamma_N$, and the EVD of $\mathbf{T}$ be $\mathbf{T} = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H$ with the eigenvalues $\tau_1 \geq \cdots \geq \tau_N$. Denote by $\mathbf{d}(\mathbf{A})$ and $\boldsymbol{\lambda}(\mathbf{A})$ the vectors consisting of the diagonal elements and the eigenvalues of a square matrix $\mathbf{A}$, respectively. Now, we are ready to state the main results.

Theorem 1: Let $\mathcal{Q} \subseteq \mathbb{S}_+^N$ be a nonempty compact set defined by constraining $\mathbf{Q}$ only through its eigenvalues, and $\mathcal{E}$ be defined as in (12) with $\mathbf{T} = \tau\mathbf{I}$ $(\tau > 0)$. Then, $\mathbf{U}_q = \mathbf{U}_h$ is optimal for the maximin problem (13).

Proof: From Proposition 2, when $\mu = 0$, (13) has a zero objective value and any $\mathbf{Q} \in \mathcal{Q}$ is optimal, so in particular $\mathbf{U}_q = \mathbf{U}_h$ is optimal as well. When $\mu > 0$, the maximin problem (13) with $\mathbf{T} = \tau\mathbf{I}$ is equivalent to

$$\underset{\mathbf{Q}\in\mathcal{Q},\mu>0}{\text{maximize}} \;\; \mu\tau\text{Tr}\left[\mathbf{Q}\left(\mathbf{Q}+\mu\tau\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right] - \varepsilon^2\mu. \qquad (29)$$

Lemma 3 ([56, 9.H.1.g, 9.H.1.h]): Let $\mathbf{A}$ and $\mathbf{B}$ be two $N \times N$ positive semidefinite matrices, with eigenvalues $\alpha_1 \geq \cdots \geq \alpha_N$ and $\beta_1 \geq \cdots \geq \beta_N$, respectively. Then,

$$\sum_{i=1}^{N}\alpha_i\beta_{N-i+1} \leq \text{Tr}\left(\mathbf{AB}\right) \leq \sum_{i=1}^{N}\alpha_i\beta_i.$$

According to Lemma 3, it follows that

$$\begin{aligned}
&\mu\tau\text{Tr}\left[\mathbf{Q}\left(\mathbf{Q}+\mu\tau\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right] \\
&= \mu\tau\text{Tr}\left[\mathbf{\Lambda}_q\left(\mathbf{\Lambda}_q+\mu\tau\mathbf{I}\right)^{-1}\mathbf{U}_q^H\mathbf{U}_h\mathbf{\Lambda}_h\mathbf{U}_h^H\mathbf{U}_q\right] \\
&\leq \mu\tau\sum_{i=1}^{N}\frac{p_i}{\mu\tau+p_i}\gamma_i = \sum_{i=1}^{N}\frac{\mu\tau\gamma_i p_i}{\mu\tau+p_i} \qquad (30)
\end{aligned}$$

where the equality holds when $\mathbf{U}_q = \mathbf{U}_h$. Since the power constraint (i.e., the set $\mathcal{Q}$) does not depend on $\mathbf{U}_q$, we can always choose $\mathbf{U}_q = \mathbf{U}_h$ to maximize the objective without affecting the constraints. ■

Theorem 2: Let $\mathcal{Q} = \{\mathbf{Q} : \mathbf{Q} \succeq 0, f_k(\boldsymbol{\lambda}(\mathbf{Q})) \leq P_k, \forall k\}$ where each $f_k(\mathbf{x})$ is a Schur-convex function, and $\mathcal{E}$ be defined as in (12) with $\mathbf{T} \succ 0$ and $\mathbf{U}_t = \mathbf{U}_h$. Then, $\mathbf{U}_q = \mathbf{U}_h$ is optimal for the maximin problem (13).

Proof: The proof is based on showing that the objective is increased by using $\mathbf{U}_q = \mathbf{U}_h$ and at the same time the power constraint is sill satisfied. Similarly, for $\mu = 0$, any $\mathbf{Q} \in \mathcal{Q}$ is optimal, so is $\mathbf{U}_q = \mathbf{U}_h$. For $\mu > 0$, using Proposition 2 and $\mathbf{U}_t = \mathbf{U}_h$, the maximin problem (13) is equivalent to

$$\underset{\mathbf{Q}\in\mathcal{Q},\mu>0}{\text{maximize}} \;\; \mu\text{Tr}\left[\mathbf{U}_h^H\mathbf{Q}\mathbf{U}_h\left(\mathbf{U}_h^H\mathbf{Q}\mathbf{U}_h+\mu\mathbf{\Lambda}_t\right)^{-1}\mathbf{\Lambda}_t\mathbf{\Lambda}_h\right] - \varepsilon^2\mu. \qquad (31)$$

Introduce $\hat{\mathbf{Q}} = \mathbf{U}_h^H\mathbf{Q}\mathbf{U}_h$. Since the power constraint depends only on the eigenvalues of $\mathbf{Q}$, the constraint $\mathbf{Q} \in \mathcal{Q}$ equals to $\hat{\mathbf{Q}} \in \mathcal{Q}$. Thus, we can consider instead

$$\underset{\hat{\mathbf{Q}}\in\mathcal{Q},\mu>0}{\text{maximize}} \;\; \mu\text{Tr}\left[\hat{\mathbf{Q}}\left(\hat{\mathbf{Q}}+\mu\mathbf{\Lambda}_t\right)^{-1}\mathbf{\Lambda}_t\mathbf{\Lambda}_h\right] - \varepsilon^2\mu. \qquad (32)$$

Lemma 4 ([34], [35], [57]): Let $\mathbf{J}_n \in \mathbb{R}^{N\times N}$ be a diagonal matrix with the diagonal elements being ± 1. There are $L = 2^N$ different such matrices indexed from $n = 1$ to L. Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be an arbitrary matrix, and $\mathbf{D}_{\mathbf{A}}$ is a diagonal matrix such that $[\mathbf{D}_{\mathbf{A}}]_{ii} = [\mathbf{A}]_{ii}$ for $i = 1, \ldots, N$. Then, $\mathbf{D}_{\mathbf{A}} = \frac{1}{L}\sum_{n=1}^{L}\mathbf{J}_n\mathbf{A}\mathbf{J}_n$.

Define $G(\hat{\mathbf{Q}}) \triangleq \mathrm{Tr}\left[\hat{\mathbf{Q}}\left(\hat{\mathbf{Q}}+\mu\mathbf{\Lambda}_t\right)^{-1}\mathbf{\Lambda}_t\mathbf{\Lambda}_h\right]$. It follows that

$$\begin{aligned} &G\left(\mathbf{J}_n\hat{\mathbf{Q}}\mathbf{J}_n\right)\\ &= \mathrm{Tr}\left[\mathbf{J}_n\hat{\mathbf{Q}}\mathbf{J}_n\left(\mathbf{J}_n\hat{\mathbf{Q}}\mathbf{J}_n+\mu\mathbf{\Lambda}_t\right)^{-1}\mathbf{\Lambda}_t\mathbf{\Lambda}_h\right]\\ &= \mathrm{Tr}\left[\hat{\mathbf{Q}}\mathbf{J}_n\mathbf{J}_n^{-1}\left(\hat{\mathbf{Q}}+\mu\mathbf{J}_n^{-1}\mathbf{\Lambda}_t\mathbf{J}_n^{-1}\right)^{-1}\mathbf{J}_n^{-1}\mathbf{\Lambda}_t\mathbf{\Lambda}_h\mathbf{J}_n\right]\\ &= \mathrm{Tr}\left[\hat{\mathbf{Q}}\left(\hat{\mathbf{Q}}+\mu\mathbf{\Lambda}_t\right)^{-1}\mathbf{\Lambda}_t\mathbf{\Lambda}_h\right] = G(\hat{\mathbf{Q}}) \end{aligned} \tag{33}$$

where we use the properties that $\mathbf{J}_n$ is a diagonal matrix, $\mathbf{J}_n^{-1} = \mathbf{J}_n$ and $\mathbf{J}_n^2 = \mathbf{I}$. This means that the function $G(\hat{\mathbf{Q}})$ is invariant under the transformation $\hat{\mathbf{Q}} \longmapsto \mathbf{J}_n\hat{\mathbf{Q}}\mathbf{J}_n$. In addition, it can be verified that the Hessian matrix of $G(\hat{\mathbf{Q}})$ is given by

$$-2\mu\mathbf{X}^T \otimes \mathbf{X}\mathbf{\Lambda}_t^2\mathbf{\Lambda}_h\mathbf{X} \preceq 0 \tag{34}$$

where $\mathbf{X} = (\hat{\mathbf{Q}}+\mu\mathbf{\Lambda}_t)^{-1}$. Thus, $G(\hat{\mathbf{Q}})$ is a concave function. Using all above facts, we have

$$\begin{aligned} G(\hat{\mathbf{Q}}) &= \frac{1}{L}\sum_{n=1}^{L} G(\hat{\mathbf{Q}}) = \frac{1}{L}\sum_{n=1}^{L} G\left(\mathbf{J}_n\hat{\mathbf{Q}}\mathbf{J}_n\right)\\ &\leq G\left(\frac{1}{L}\sum_{n=1}^{L}\mathbf{J}_n\hat{\mathbf{Q}}\mathbf{J}_n\right) = G(\mathbf{D}_{\hat{\mathbf{Q}}}) \end{aligned} \tag{35}$$

where $\mathbf{D}_{\hat{\mathbf{Q}}}$ is a diagonal matrix such that $[\mathbf{D}_{\hat{\mathbf{Q}}}]_{ii} = [\hat{\mathbf{Q}}]_{ii}$ for $i = 1,\ldots,N$. The upper bound is achieved when $\hat{\mathbf{Q}}$ is a diagonal matrix.

For any feasible $\hat{\mathbf{Q}}$, we can always choose $\tilde{\mathbf{Q}}$ such that $\tilde{\mathbf{Q}} = \mathbf{D}_{\hat{\mathbf{Q}}}$. Since $\mathbf{d}(\hat{\mathbf{Q}})$ is majorized by $\boldsymbol{\lambda}(\hat{\mathbf{Q}})$ [56, 9.B.1] and each $f_k(\mathbf{x})$ is a Schur-convex function, it follows that

$$f_k(\boldsymbol{\lambda}(\tilde{\mathbf{Q}})) = f_k(\mathbf{d}(\hat{\mathbf{Q}})) \leq f_k(\boldsymbol{\lambda}(\hat{\mathbf{Q}})). \tag{36}$$

Observing that $\hat{\mathbf{Q}} \succeq 0$ implies $\mathbf{D}_{\hat{\mathbf{Q}}} \succeq 0$, hence $\tilde{\mathbf{Q}}$ is feasible too. On the other hand, from (35), $G(\tilde{\mathbf{Q}}) \geq G(\hat{\mathbf{Q}})$. Therefore, in the optimal solution set, there must exist diagonal matrices. Recall that $\hat{\mathbf{Q}} = \mathbf{U}_h^H\mathbf{U}_q\mathbf{\Lambda}_q\mathbf{U}_q^H\mathbf{U}_h$. Any feasible diagonal matrix $\mathbf{D}_{\hat{\mathbf{Q}}} \in \mathcal{Q}$ can be obtained from a feasible $\mathbf{Q} \in \mathcal{Q}$ by setting $\mathbf{U}_q = \mathbf{U}_h$ and $\mathbf{\Lambda}_q = \mathbf{D}_{\hat{\mathbf{Q}}}$. Consequently, we have proved that, in the optimal solution set, there must exist $\mathbf{Q}^\star$ with $\mathbf{U}_q^\star = \mathbf{U}_h$. ■

Remark 2: Theorem 1 indicates that $\mathbf{U}_q = \mathbf{U}_h$ is optimal with a general power constraint relying only on the eigenvalues of $\mathbf{Q}$, provided the uncertainty region is a sphere defined by the Frobenius norm, which is the most frequently used deterministic model [32]–[35], [37], [41]–[45] . When more restrictions are added to the eigenvalues of $\mathbf{Q}$, Theorem 2 shows that $\mathbf{U}_q = \mathbf{U}_h$ is optimal for an ellipsoid uncertainty region if $\hat{\mathbf{H}}^H\hat{\mathbf{H}}$ and $\mathbf{T}$ can be simultaneously diagonalized. Given that both $\mathrm{Tr}\,(\mathbf{Q})$ and $\lambda_{\max}(\mathbf{Q})$ are Schur-convex functions of the eigenvalues of $\mathbf{Q}$ [15], Theorem 2 is applicable to the two most common constraints $\mathcal{Q}_1$ (sum power constraint) and $\mathcal{Q}_2$ (maximum power constraint) as well as their intersection. Therefore, in most cases, the optimality of the eigenmode transmission (over the nominal channel) still holds for the worst-case design, which complies with the cases of perfect CSIT and statistical CSIT with mean or covariance feedback.

Remark 3: Note that the problems considered in [42] (for the spherical uncertainty region) and [43] are a special case of our framework with $\mathcal{Q} = \mathcal{Q}_1$ and $\mathbf{T} = \mathbf{I}$. However, they assumed $\mathbf{U}_q = \mathbf{U}_h$ without knowing whether this is optimal or not, even if [43] offered some sufficient conditions. In this paper, we prove that the assumption made in [42], [43] is actually optimal in more general cases. Consequently, by using Theorem 1 or 2, the complicated matrix-valued maximin problem (13) can be simplified to a scalar power allocation problem without any loss of optimality.

V. Optimal Power Allocation

In this section, we will solve the power allocation problem when the (nominal) channel is diagonalized as a result of Theorems 1 and 2. To be more specific, we first show that, with a general power constraint, the simplified power allocation problem is a convex problem, and thus can be efficiently solved. Then, we consider the sum power constraint and the maximum power constraint, and derive the closed-form solutions to the resulting power allocation problems, respectively.

When $\mathbf{U}_t = \mathbf{U}_h$ and $\mathcal{Q}$ is a nonempty compact convex set satisfying the condition in Theorem 2, from (31), the maximin problem (13) can be simplified to

$$\underset{\{p_i\}:\mathbf{Q}\in\mathcal{Q},\mu\geq 0}{\text{maximize}} \quad \sum_{i=1}^{r}\frac{\mu\tau_i\gamma_i p_i}{\mu\tau_i+p_i} - \varepsilon^2\mu \tag{37}$$

where $r = \mathrm{rank}(\hat{\mathbf{H}})$. Define

$$\varphi_i(\mu, p_i) \triangleq \frac{\mu\tau_i\gamma_i p_i}{\mu\tau_i+p_i}$$

whose Hessian matrix is given by

$$\begin{aligned} &\frac{-2\tau_i^2\gamma_i}{(\mu\tau_i+p_i)^3}\begin{bmatrix}\mu^2 & -\mu p_i\\ -\mu p_i & p_i^2\end{bmatrix}\\ &\qquad = \frac{-2\tau_i^2\gamma_i}{(\mu\tau_i+p)^3}\begin{bmatrix}\mu\\ -p_i\end{bmatrix}\begin{bmatrix}\mu & -p_i\end{bmatrix} \preceq 0 \end{aligned} \tag{38}$$

implying that $\varphi_i(\mu, p_i)$ is a concave function, so is the objective function of (37). Therefore, (37) is a convex problem, admitting globally optimal solutions that can be efficiently found by, e.g., an interior-point method [47]. Similarly, it is easy to verify that when $\mathbf{T} = \tau\mathbf{I}$ ($\tau > 0$) and $\mathcal{Q}$ is a nonempty compact convex set satisfying the condition in Theorem 1, the resulting power allocation problem is a convex problem as well.

A. Sum Power Constraint $\mathcal{Q} = \mathcal{Q}_1$ and $\mathbf{U}_t = \mathbf{U}_h$

From (37), the corresponding power allocation problem is

$$\begin{aligned} \underset{\{p_i\},\mu}{\text{maximize}} \quad & \sum_{i=1}^{r}\frac{\mu\tau_i\gamma_i p_i}{\mu\tau_i+p_i} - \varepsilon^2\mu\\ \text{subject to} \quad & \sum_{i=1}^{N} p_i = P_s\\ & p_1 \geq \cdots \geq p_N \geq 0\\ & \mu \geq 0 \end{aligned} \tag{39}$$

where we explicitly write the decreasing order of $\{p_i\}$. Noticing that $\varphi_i(\mu, p_i)$ is monotonically increasing in p_i for fixed μ, no power will be allocated to zero γ_i, which means $p_i^\star = 0$ for $i > r$. Hence, the sum power constraint can be reduced to $\sum_{i=1}^r p_i = P_s$. The solution to the above problem is given by the following theorem.

Theorem 3: The solution to the problem (39) is

$$p_i^\star = \begin{cases} \tau_i \left[\sqrt{\frac{\gamma_i}{b_k}}(P_s + c_k\mu^\star) - \mu^\star\right], & \text{for } i = 1, \ldots, k \\ 0, & \text{for } i > k \end{cases} \tag{40}$$

with

$$\mu^\star = \frac{P_s}{c_k}\left(\sqrt{\frac{b_k}{b_k - a_k c_k}} - 1\right) \tag{41}$$

where $a_m \triangleq \sum_{j=1}^m \tau_j\gamma_j - \varepsilon^2$, $b_m \triangleq (\sum_{j=1}^m \tau_j\sqrt{\gamma_j})^2$, $c_m \triangleq \sum_{j=1}^m \tau_j$ and $\beta_m \triangleq \sum_{j=1}^m \tau_j(\sqrt{\gamma_j} - \sqrt{\gamma_m})^2$ for $m = 1, \ldots, r$, and $\beta_{r+1} \triangleq +\infty$, and k is an integer such that $\beta_k < \varepsilon^2 \leq \beta_{k+1}$. The optimum value of (39), i.e., the worst-case SNR, is

$$\mathrm{SNR} = \frac{P_s}{c_k^2}\left(\sqrt{b_k} - \sqrt{b_k - a_k c_k}\right)^2. \tag{42}$$

Proof: See Appendix B. ■

Corollary 1: For $\mathcal{Q} = \mathcal{Q}_1$ and $\mathbf{U}_t = \mathbf{U}_h$, the robust transmitter uses only one eigenmode if

$$\varepsilon \leq \sqrt{\tau_1}(\sqrt{\gamma_1} - \sqrt{\gamma_2}). \tag{43}$$

Corollary 2: As $\varepsilon \longrightarrow \|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$, the solution to the problem (39) becomes

$$p_i^\star = \frac{\tau_i\sqrt{\gamma_i}}{\sum_{j=1}^r \tau_j\sqrt{\gamma_j}} P_s, \; i = 1, \ldots, r. \tag{44}$$

Remark 4: According to Theorem 3, the robust transmitter will use multiple eigenmodes to increase the reliability in the worst-case channel. The larger the error radius ε is, i.e., the more uncertainty, the more eigenmodes will be used. Corollary 1 indicates that beamforming along one direction is robust if ε is very small, or if the difference between the largest two singular values of the nominal channel is very large, which implies a nearly rank-one channel. Interestingly, the similar result on the number of the active eigenmodes was also obtained in [42], [43]. However, In contrast to the semi-closed-form solutions in [42], [43], we offer the fully analytical solution in a more general case. Furthermore, from Corollary 2, as ε approaches $\|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$, the worst-case design tends to allocate the transmit power according to the weighted proportion of a singular value of the nominal channel, instead of a uniform distribution. This has been observed in [42], [43] through numerical simulations, but no proper explanation was given. The fundamental reason is that the deterministic model adopted in this paper is not an isotropically unconstrained set [38], but an ellipsoid with the center, i.e., the nominal channel, away from the origin.

B. Maximum Power Constraint $\mathcal{Q} = \mathcal{Q}_2$ and $\mathbf{U}_t = \mathbf{U}_h$

The corresponding power allocation problem is

$$\begin{aligned} \underset{\{p_i\},\mu}{\text{maximize}} \quad & \sum_{i=1}^r \frac{\mu\tau_i\gamma_i p_i}{\mu\tau_i + p_i} - \varepsilon^2\mu \\ \text{subject to} \quad & P_m \geq p_1 \geq \cdots \geq p_N \geq 0 \\ & \mu \geq 0. \end{aligned} \tag{45}$$

Due to the monotonicity of $\varphi_i(\mu, p_i)$, it is easy to see that the optimal power allocation is $p_i^\star = P_m$ for $i = 1, \ldots, r$ and $p_i^\star = 0$ for $i > r$. That is, a uniform distribution with the maximum power on each nonzero eigenmode.

To obtain the optimal objective value, we need to solve

$$\underset{\mu \geq 0}{\text{maximize}} \; \sum_{i=1}^r \frac{\mu\tau_i\gamma_i P_m}{\mu\tau_i + P_m} - \varepsilon^2\mu \tag{46}$$

which is a convex problem since the objective function is strictly concave in μ. The optimal μ can be found by setting the derivative of the objective function to be zero, which turns to finding the positive root of the following equation:

$$h(\mu) \triangleq \sum_{i=1}^r \frac{\tau_i\gamma_i P_m^2}{(\mu\tau_i + P_m)^2} = \varepsilon^2. \tag{47}$$

Unfortunately, this equation does not admit an analytical root, so we have to resort to some numerical methods. Now that $h(\mu)$ is a monotonically decreasing function, the root of (47) can be found through the bisection method, which requires an initial range, say $[\mu_l, \mu_u]$. The lower boundary can be $\mu_l = 0$ since $h(0) = \sum_{i=1}^r \tau_i\gamma_i > \varepsilon^2$. To find an upper boundary μ_u, it follows that

$$h(\mu_u) = \sum_{i=1}^r \frac{\tau_i\gamma_i P_m^2}{(\mu_u\tau_i + P_m)^2} \leq \frac{\sum_{i=1}^r \tau_i\gamma_i P_m^2}{(\mu_u\tau_r + P_m)^2} = \varepsilon^2 \tag{48}$$

from which we obtain

$$\mu_u = \frac{P_m}{\tau_r}\left(\frac{1}{\varepsilon}\sqrt{\sum\nolimits_{i=1}^r \tau_i\gamma_i} - 1\right). \tag{49}$$

If $\mathbf{T} = \tau\mathbf{I}$ ($\tau > 0$), then a closed-form solution to (46) is available

$$\mu^\star = \frac{P_m}{\tau}\left(\frac{1}{\varepsilon}\sqrt{\tau\sum\nolimits_{i=1}^r \gamma_i} - 1\right) \tag{50}$$

which leads to the optimum value

$$\mathrm{SNR} = P_m\left(\sqrt{\sum\nolimits_{i=1}^r \gamma_i} - \frac{\varepsilon}{\sqrt{\tau}}\right)^2. \tag{51}$$

VI. The QoS Problem

In this section, we consider, as the complement of the maximin problem (2), the so-called QoS problem that minimizes the power consumption at the transmitter while keeping the performance measure above a given threshold for any channel realization $\mathbf{H} \in \mathcal{H}$. Specifically, letting $S(\mathbf{Q})$ represent the power consumption function at the transmitter, the QoS problem is

$$\begin{aligned} \underset{\mathbf{Q}\succeq 0}{\text{minimize}} \quad & S(\mathbf{Q}) \\ \text{subject to} \quad & \Psi(\mathbf{Q},\mathbf{H}) \geq \rho, \ \forall \mathbf{H} \in \mathcal{H} \end{aligned} \tag{52}$$

where $\rho > 0$ is the QoS threshold. With the performance measure (4) and the channel error (11), the QoS problem can be further expressed as

$$\begin{aligned} \underset{\mathbf{Q}\succeq 0}{\text{minimize}} \quad & S(\mathbf{Q}) \\ \text{subject to} \quad & \mathrm{Tr}\left[\left(\hat{\mathbf{H}}+\mathbf{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}}+\mathbf{\Delta}\right)^H\right] \geq \rho, \ \forall \mathbf{\Delta} \in \mathcal{E} \end{aligned} \tag{53}$$

where $\mathcal{E}$ is defined as in (12).

Fundamentally, given the same $\hat{\mathbf{H}}$ and $\mathcal{E}$, the QoS problem (53) and the maximin problem (13) will have the identical optimal solutions if the threshold ρ is set equal to the optimum value of (13), and the power constraint $\mathcal{Q} = \{\mathbf{Q} : \mathbf{Q} \succeq 0, S(\mathbf{Q}) \leq P\}$ is used in (13). Consequently, the QoS problem can be alternatively solved by solving the maximin problem with properly chosen parameters. It seems not necessary to consider the QoS problem at all. However, to efficiently solve (53) through (13), one needs an analytical relation between the optimum value of (13) and the threshold ρ of (53), which is available only when (13) has a closed-form solution. Otherwise, for each specific $\hat{\mathbf{H}}$, this relation has to be evaluated by some numerical method, e.g., the bisection method, through many iterations. The computational complexity will become prohibitive when $\hat{\mathbf{H}}$ is frequently updated at the transmitter. Therefore, it is desirable to find a direct way to solve the QoS problem.

Proposition 3: Let $\mathcal{E}$ be defined as in (12). Then, the QoS problem (53) is equivalent to

$$\begin{aligned} \underset{\mathbf{Q},\mu,\mathbf{Z}}{\text{minimize}} \quad & S(\mathbf{Q}) \\ \text{subject to} \quad & \mathbf{Q} \succeq 0, \ \ \mu \geq 0 \\ & \mathrm{Tr}\left[(\mathbf{Z}-\mathbf{Q})\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right] + \varepsilon^2\mu + \rho \leq 0 \\ & \begin{bmatrix} \mathbf{Z} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{Q}+\mu\mathbf{T} \end{bmatrix} \succeq 0. \end{aligned} \tag{54}$$

Proof: Similar to that of Proposition 1. ■

Remark 5: If $S(\mathbf{Q})$ is a convex function, then the equivalent problem (54) is a convex problem and thus can be efficiently solved. Some commonly used power consumption functions include:

1) Sum power: $S_1(\mathbf{Q}) = \mathrm{Tr}(\mathbf{Q})$.
2) Maximum power per spatial dimension: $S_2(\mathbf{Q}) = \lambda_{\max}(\mathbf{Q})$.
3) Maximum power per antenna: $S_3(\mathbf{Q}) = \max_i [\mathbf{Q}]_{ii}$.

It is easy to see that, when $S(\mathbf{Q})$ is one of $S_1(\mathbf{Q})$, $S_2(\mathbf{Q})$, $S_3(\mathbf{Q})$ or any positive weighted sum of them, the problem (54) is or can be transformed into an SDP.

Theorem 4: Let $S(\mathbf{Q})$ depends only on the eigenvalues of $\mathbf{Q}$, and $\mathcal{E}$ be defined as in (12) with $\mathbf{T} = \tau\mathbf{I}$ ($\tau > 0$). Then, $\mathbf{U}_q = \mathbf{U}_h$ is optimal for the QoS problem (53).

Proof: See Appendix C. ■

Theorem 5: Let $S(\mathbf{Q}) = \sum_k \alpha_k f_k(\boldsymbol{\lambda}(\mathbf{Q}))$ where $f_k(\mathbf{x})$ is a Schur-convex function and $\alpha_k \geq 0$, $\forall k$, and $\mathcal{E}$ be defined as in (12) with $\mathbf{T} \succ 0$ and $\mathbf{U}_t = \mathbf{U}_h$. Then, $\mathbf{U}_q = \mathbf{U}_h$ is optimal for the QoS problem (53).

Proof: See Appendix C. ■

Remark 6: Theorems 4 and 5 are the counterparts of Theorems 1 and 2, respectively. It is not surprising that the QoS problem (13) has the same optimal transmit directions as the maximin problem (53), since they are complementary. Theorem 5 is applicable when $S(\mathbf{Q}) = S_1(\mathbf{Q})$ or $S_2(\mathbf{Q})$ or any non-negatively weighted sum of them. Noticing that $\mathcal{Q}_1 = \{\mathbf{Q} : \mathbf{Q} \succeq 0, S_1(\mathbf{Q}) \leq P_s\}$ and $\mathcal{Q}_2 = \{\mathbf{Q} : \mathbf{Q} \succeq 0, S_2(\mathbf{Q}) \leq P_m\}$, we can solve the QoS problem (53) by utilizing the closed-form solution of the maximin problem (13) in the case of $\mathcal{Q} = \mathcal{Q}_1$ or $\mathcal{Q}_2$. The only thing needed is to adjust P_s or P_m such that the optimum value of (13) equals to the threshold ρ. For the two cases in Sections V-A and V-B, this relation can be easily found:

1) For $S(\mathbf{Q}) = S_1(\mathbf{Q})$ and $\mathbf{U}_t = \mathbf{U}_h$:

$$P_s = \frac{c_k^2 \rho}{\left(\sqrt{b_k} - \sqrt{b_k - a_k c_k}\right)^2} \tag{55}$$

where k satisfies $\beta_k < \varepsilon^2 \leq \beta_{k+1}$.

2) For $S(\mathbf{Q}) = S_2(\mathbf{Q})$ and $\mathbf{T} = \tau\mathbf{I}$ ($\tau > 0$):

$$P_m = \frac{\rho}{\left(\sqrt{\sum_{i=1}^r \gamma_i} - \frac{\varepsilon}{\sqrt{\tau}}\right)^2}. \tag{56}$$

VII. Numerical Results

In this section, the performance of the proposed robust transmit strategy will be investigated through numerical simulations. To be more specific, we will compare the robust approach with the beamforming strategy that transmits only over the maximum eigenvalue of the nominal channel, and with the equal-power transmission that allocates the transmit power equally over all eigenmodes. For simplicity, we consider $\mathbf{T} = \mathbf{I}$, i.e., a spherical uncertainty region. Note that, for the maximum power constraint set $\mathcal{Q}_2$, all transmit strategies are equal in the sense that they will use the maximum power on each eigenmode (see Section V-B). Therefore, we will focus on the sum power constraint set $\mathcal{Q}_1$. Moreover, to take into account different channels, the elements of the nominal channel $\hat{\mathbf{H}}$ are randomly generated according to zero-mean, unit-variance, i.i.d. Gaussian distributions.

The philosophy of robustness in this paper is to guarantee a performance level for any channel realization in the uncertainty region. In other words, we are interested in the worst-case behavior of a precoder. Therefore, we will compare the worst-case performance of the different transmit strategies. In the case of $\mathcal{Q} = \mathcal{Q}_1$ and $\mathbf{T} = \mathbf{I}$, the worst-case channel error and the received SNR, for the robust precoder, are given by (64) (with μ in (41)) and (42), respectively, and the minimum transmit power satisfying a QoS threshold ρ is given by (55). With the same

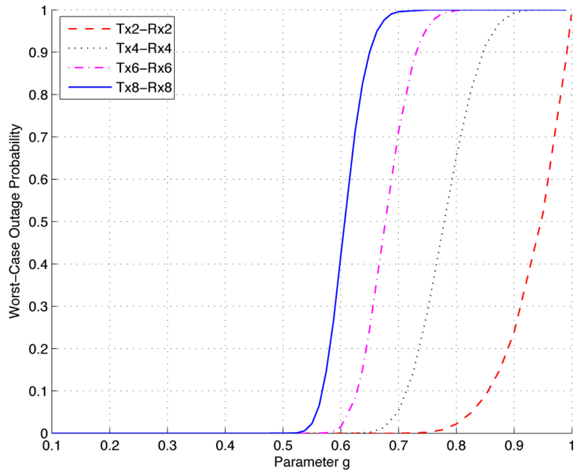

Fig. 1. Worst-case outage probability of beamforming versus the parameter g.

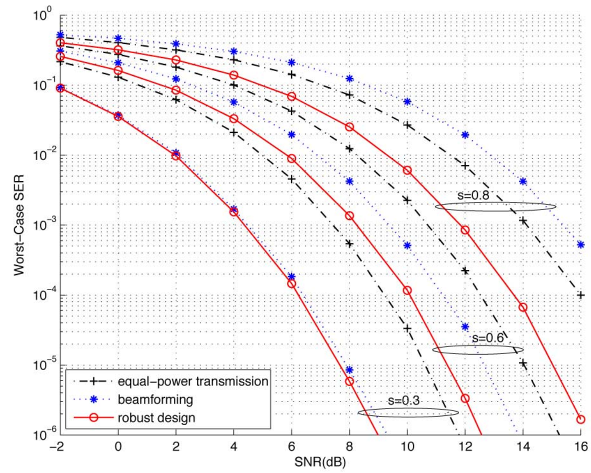

Fig. 2. Worst-case SER versus SNR with different values of s for $M = N = 4$.

configuration, it is not difficult to obtain, for the beamforming strategy, the worst-case error

$$\mathbf{\Delta}_{\rm bf} = -\hat{\mathbf{H}}\mathbf{u}_{h,1}\mathbf{u}_{h,1}^H \frac{\varepsilon}{||\hat{\mathbf{H}}||_2} \tag{57}$$

where $\mathbf{u}_{h,1}$ is the first column of $\mathbf{U}_h$, the worst-case received SNR

$$\mathrm{SNR}_{\rm bf} = P_s \left(||\hat{\mathbf{H}}||_2 - \varepsilon\right)^2 \tag{58}$$

and the minimum transmit power satisfying a QoS threshold ρ

$$P_{s,\rm bf} = \frac{\rho}{\left(||\hat{\mathbf{H}}||_2 - \varepsilon\right)^2}. \tag{59}$$

Similarly, for the equal-power transmission, it can be obtained in the worst-case that

$$\mathbf{\Delta}_{\rm eq} = -\hat{\mathbf{H}}\frac{\varepsilon}{||\hat{\mathbf{H}}||_F} \tag{60}$$

$$\mathrm{SNR}_{\rm eq} = \frac{P_s}{N}\left(||\hat{\mathbf{H}}||_F - \varepsilon\right)^2 \tag{61}$$

$$P_{s,\rm eq} = \frac{N\rho}{\left(||\hat{\mathbf{H}}||_F - \varepsilon\right)^2}. \tag{62}$$

One important thing worth pointing out is that if $\varepsilon \geq ||\hat{\mathbf{H}}||_2$, then from (57) to (59), the worst-case SNR of beamforming is zero. Even if we assume $\varepsilon < ||\hat{\mathbf{H}}||_F$, there is still a probability that beamforming could not guarantee any performance in the uncertainty region because of $||\hat{\mathbf{H}}||_2 \leq ||\hat{\mathbf{H}}||_F$. We call this probability the worst-case outage probability of beamforming. Letting $\varepsilon = g||\hat{\mathbf{H}}||_F$ with $g \in [0, 1)$, then the worst-case outage probability is $\mathrm{Prob}(g||\hat{\mathbf{H}}||_F \geq ||\hat{\mathbf{H}}||_2)$. In Fig. 1, we plot this probability versus g with different numbers of transmit and receive antennas. As expected, the outage probability grows as g increases. Meanwhile, for a fixed g, the more antennas the larger the outage probability is. This is because, as the dimension of the channel increases, $\lambda_{\max}(\hat{\mathbf{H}}^H\hat{\mathbf{H}})$ becomes a smaller fraction of $\mathrm{Tr}(\hat{\mathbf{H}}^H\hat{\mathbf{H}})$. Since setting $\varepsilon = g||\hat{\mathbf{H}}||_F$ may result in an outage for beamforming, we will use $\varepsilon = s||\hat{\mathbf{H}}||_2$ with $s \in [0, 1)$ so that all three transmit strategies can work in their worst-case situations. Nevertheless, it is possible that ε is a small proportion to $||\hat{\mathbf{H}}||_F$ even if s tends to 1.

In Fig. 2, we plot the symbol error rates (SERs) of the three transmit strategies in their worst-case channels versus SNR for different values of s. With four antennas equipped at both ends of the link, i.e., $M = N = 4$, the QPSK modulated symbols are encoded at the transmitter according to a 3/4-rate complex OSTBC introduced in [7], and decoded by a ML detector at the receiver. The worst-case SER is averaged over $\hat{\mathbf{H}}$. As observed from Fig. 2, the robust approach offers the lowest worst-case SER among all transmit strategies, which complies with our design objective. When s is small, i.e., the channel error is small, the performance of beamforming is close to that of the robust approach. This is consistent with Corollary 1 which says that when ε is small the robust strategy coincides with beamforming. On the other hand, as s increases (so does ε), the performance gap between beamforming and the robust approach becomes larger, and eventually beamforming is outperformed by the equal-power transmission. This fact can be more evidently observed in Fig. 3, where the worst-case SER versus s for a given $\mathrm{SNR} = 6$ dB is displayed.

An alternative way to investigate the robustness capability is to compare the minimum transmit power needed to satisfy a QoS threshold. Fig. 4 shows the average minimum transmit power for various QoS thresholds, and Fig. 5 provides the relation between the average minimum transmit power and s for a given QoS threshold $\rho = 6$ dB. The numbers of the transmit and receive antennas are still $M = N = 4$. As can be seen, considerable transmit power is saved by using the robust approach, compared to the other two strategies. Beamforming needs less transmit power than the equal-power transmission when s is small, but more power than the equal-power transmission when s becomes large. Therefore, the results shown in Figs. 4 and 5 conform to those in Figs. 2 and 3.

As mentioned in Section II-B, the maximin problem (13) provides an approximate solution to the *compound capacity* [39], i.e., the maximum mutual information for the worst-case

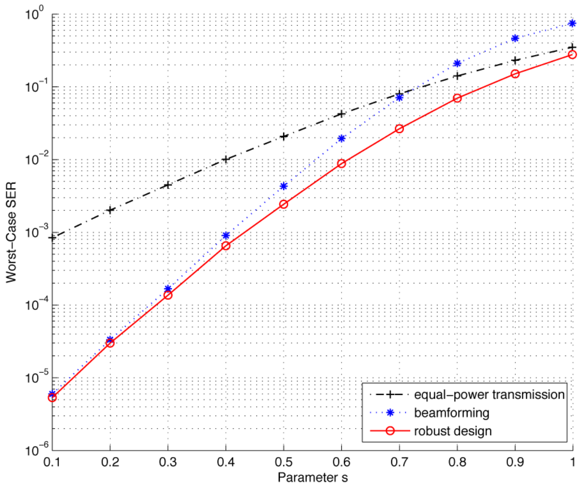

Fig. 3. Worst-case SER versus the parameter s at $\mathrm{SNR} = 6$ dB for $M = N = 4$.

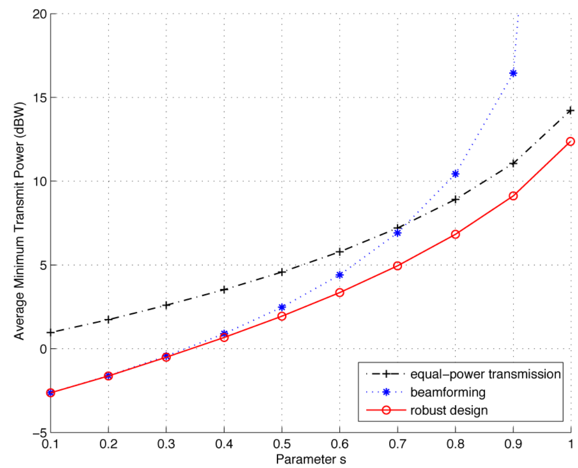

Fig. 5. Average minimum transmit power satisfying a QoS threshold $\rho = 6$ dB versus the parameter s for $M = N = 4$.

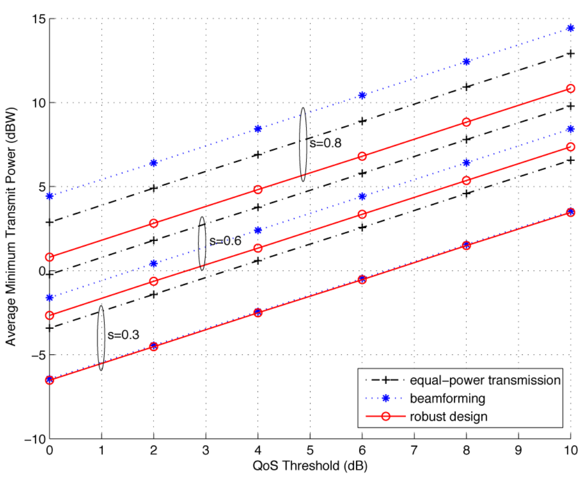

Fig. 4. Average minimum transmit power versus the QoS threshold with different values of s for $M = N = 4$.

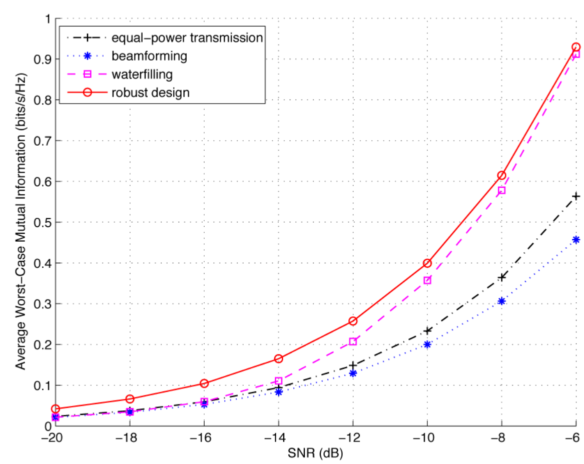

Fig. 6. Average worst-case mutual information versus SNR for $s = 0.7$ and $M = N = 6$.

channel, of the MIMO channel at low SNR. In Fig. 6, the average worst-case mutual information of different transmit strategies is compared under the setting $M = N = 6$ and $s = 0.7$ (which corresponds on average $g = 0.55$). Under the channel uncertainty model (12), the exact worst-case channel errors, in terms of mutual information, for the transmit strategies compared here are unknown so far (and still an open problem). Therefore, we proximately use, for the equal-power transmission, beamforming and robust strategies, the same worst-case channel errors as in Figs. 2–5. The worst-case channel error for the waterfilling strategy, which conducts the waterfilling power allocation over the nominal channel, is the worst one among 100 randomly generated channel errors, whose singular vectors are restricted to be the same as those of the nominal channel. It is shown that, at low SNR, the robust approach provides the maximum worst-case mutual information among all strategies.

The numerical results verify the fact that it is always better to exploit the channel knowledge at the transmitter. When the channel uncertainty is very small, one may use the beamforming strategy, since it is nearly robust in this case. When the channel uncertainty is quite large, although better than beamforming, the equal-power transmission, which assumes no CSIT, is still inferior to the robust approach. For the whole range of uncertainty, especially moderate uncertainty, the robust approach provides the best worst-case performance by using imperfect CSIT through a reasonable deterministic model.

VIII. Conclusion

We have addressed the problem of finding robust transmit strategies for MIMO communication systems with imperfect CSIT. The imperfectness of CSIT was characterized by a deterministic model, which assumes the actual channel within an ellipsoid centered at a nominal channel. Then, we formulated the robust transmitter design as a maximin problem, which maximizes the worst-case received SNR, or minimizes the worst-case Chernoff bound of the PEP for a STBC. We have also considered the QoS problem that minimizes the transmit power while keeping the received SNR above a given QoS threshold. For a general class of power constraints, both the maximin and QoS problems can be reformulated into convex optimization problems, or even further into SDPs, and thus can be efficiently solved by the numerical methods. More importantly, under some mild conditions, the optimal transmit directions, i.e., the eigenvectors of the transmit covariance matrix, are just the right singular vector of the nominal channel. This means that the eigenmode transmission (over the nominal channel) is still optimal for the worst-case design, thus being consistent with the cases of perfect CSIT and statistical CSIT with mean or covariance feedback. Consequently, the complicated matrix-valued maximin and QoS problems can be simplified to the scalar power allocation problems. Finally, we provided the closed-form solutions to the resulting power allocation problems.

Appendix A
Proof of Proposition 2

The proof begins by deriving the dual problem of the inner minimization of the maximin problem (13). Then, we replace the inner minimization by its dual maximization, hence transforming the maximin problem to a maximization problem.

The inner minimization of (13) is

$$\underset{\mathbf{\Delta}\in\mathcal{E}}{\text{minimize}}\ \text{Tr}\left[\left(\hat{\mathbf{H}}+\mathbf{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}}+\mathbf{\Delta}\right)^H\right] \quad (63)$$

where $\mathcal{E}=\{\mathbf{\Delta}:\text{Tr}(\mathbf{\Delta}\mathbf{T}\mathbf{\Delta}^H)\leq\varepsilon^2\}$, and whose Lagrangian is given by

$$l(\mathbf{\Delta},\mu)=\text{Tr}\left[\left(\hat{\mathbf{H}}+\mathbf{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}}+\mathbf{\Delta}\right)^H\right] +\mu\left[\text{Tr}\left(\mathbf{\Delta}\mathbf{T}\mathbf{\Delta}^H\right)-\varepsilon^2\right] \quad (64)$$

with the Lagrange multiplier $\mu\geq 0$. Now that (64) is a convex function of $\mathbf{\Delta}$ for fixed μ, its minimizer can be found, according to the KKT conditions [46], by letting $\partial l\left(\mathbf{\Delta},\mu\right)/\partial\mathbf{\Delta}=0$.

For $\mu>0$, the minimizer of $l\left(\mathbf{\Delta},\mu\right)$ can be easily obtained as

$$\mathbf{\Delta}^\star=-\hat{\mathbf{H}}\mathbf{Q}\left(\mathbf{Q}+\mu\mathbf{T}\right)^{-1}. \quad (65)$$

Substituting (65) back into $l\left(\mathbf{\Delta},\mu\right)$, we have the dual function

$$\begin{aligned} g(\mu) &= l\left(\mathbf{\Delta}^\star,\mu\right)=\text{Tr}\left(\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H\right) \\ &\quad -\text{Tr}\left[\hat{\mathbf{H}}\mathbf{Q}\left(\mathbf{Q}+\mu\mathbf{T}\right)^{-1}\mathbf{Q}\hat{\mathbf{H}}^H\right]-\varepsilon^2\mu \\ &= \mu\text{Tr}\left[\mathbf{Q}\left(\mathbf{Q}+\mu\mathbf{T}\right)^{-1}\mathbf{T}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right]-\varepsilon^2\mu. \end{aligned} \quad (66)$$

Given that (63) is a convex problem with a compact convex feasible set containing nonempty interior (Slater's condition [46] is satisfied), there is no gap between the dual problem and the primal problem. As a result, the inner minimization (63) can be equivalently replaced by its dual maximization, and the maximin problem (13) becomes the maximization problem (28).

For $\mu=0$, the minimizer of $l\left(\mathbf{\Delta},\mu\right)$ is any $\mathbf{\Delta}$ satisfying $\mathbf{\Delta}\mathbf{Q}=-\hat{\mathbf{H}}\mathbf{Q}$ (for example one choice is $\mathbf{\Delta}=-\hat{\mathbf{H}}$), which leads to the dual function $g\left(0\right)=0$ as well as a zero objective value of the maximin problem (13). Thus, we can include the case of $\mu=0$ into (28) by defining the objective of (28) as 0 at $\mu=0$. On the other hand, (13) admits a zero objective value if and only if $\varepsilon\geq\|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$. When $\varepsilon<\|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$ (to avoid the trivial solution), the optimal $\mathbf{Q}$ of (13) can never satisfy $\mathbf{\Delta}\mathbf{Q}=-\hat{\mathbf{H}}\mathbf{Q}$ (neither can $\mathbf{\Delta}$ be equal to $-\hat{\mathbf{H}}$). This means that for small channel errors, i.e., $\varepsilon<\|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$, we can consider $\mu>0$ without loss of any optimality in (28). The proof of Proposition 2 is completed.

Appendix B
Proof of Theorem 3

Lemma 5: Let $\{t_n\}_{n=1}^{N+1}$ be an increasing sequence. Then the piecewise function

$$f\left(t\right)=\begin{cases} f_n\left(t\right), & \text{if } t\in\left[t_n,t_{n+1}\right), n=1,\ldots,N \\ 0, & \text{otherwise}\end{cases}$$

is concave if $f_n\left(t\right)$ is concave in $[t_n,t_{n+1})$ for $n=1,\ldots,N$, and $f\left(t\right)$ is smooth.

Proof of Lemma 5: We only need to prove that the concatenation of two functions $f_1(t)$ and $f_2(t)$ is concave, because the same reasoning can be recursively applied to $f_n(t)$, $n=3,\ldots,N$. Denote the derivative of $f_n(t)$ at t_0 by $f_n'(t_0)$, and let $x\in[t_1,t_2)$ and $y\in[t_2,t_3)$. According to the first-order condition of a concave function [46], it follows that

$$f_1\left(t_2\right)\leq f_1\left(x\right)+f_1'\left(x\right)\left(t_2-x\right) \quad (67)$$

$$f_2\left(y\right)\leq f_2\left(t_2\right)+f_2'\left(t_2\right)\left(y-t_2\right). \quad (68)$$

Note that the smoothness of $f(t)$ implies

$$f_1\left(t_2\right)=f_2\left(t_2\right),\ \lim_{t\longrightarrow t_2}f_1'\left(t\right)=f_1'\left(t_2\right)=f_2'\left(t_2\right). \quad (69)$$

The concavity of each $f_n\left(t\right)$ and the smoothness of $f\left(t\right)$ imply that $f'\left(t\right)$ is a continuous non-increasing function, so we have

$$f_1'\left(x\right)\geq f_1'\left(t_2\right). \quad (70)$$

Adding (67) to (68), and using (69) and (70), we obtain

$$f_2\left(y\right)\leq f_1\left(x\right)+f_1'\left(x\right)\left(y-x\right) \quad (71)$$

or equivalently,

$$f\left(y\right)\leq f\left(x\right)+f'\left(x\right)\left(y-x\right) \quad (72)$$

thus proving that $f\left(t\right)$ is concave. ∎

As a reminder, we write down again some important parameters in Theorem 3

$$a_m\triangleq\sum_{j=1}^m\tau_j\gamma_j-\varepsilon^2,\ b_m\triangleq\left(\sum_{j=1}^m\tau_j\sqrt{\gamma_j}\right)^2, c_m\triangleq\sum_{j=1}^m\tau_j. \quad (73)$$

To solve the power allocation problem (39), we first fix μ and find the optimal $\{p_i\}$ in terms of μ. It is easy to obtain

$$p_i^\star=\begin{cases}\mu\tau_i\left(\sqrt{\frac{\gamma_i}{\eta}}-1\right), & \text{for } i=1,\ldots,m \\ 0, & \text{for } i>m\end{cases} \quad (74)$$

where $\eta\geq 0$ is the Lagrange multiplier, m is an integer such that $\gamma_m>\eta\geq\gamma_{m+1}$ and we define $\gamma_{r+1}\triangleq 0$. It can be assumed w.l.o.g. that γ_i, $i=1,\ldots,r+1$, are distinct, since η will not be within $[\gamma_{m+1},\gamma_m)$ if $\gamma_m=\gamma_{m+1}$. Substituting (74) into the sum power constraint $\sum_{i=1}^{r} p_i = P_s$, the Lagrange multiplier η can be represented by

$$\sqrt{\eta} = \frac{\mu \sum_{j=1}^{m} \tau_j \sqrt{\gamma_j}}{\mu \sum_{j=1}^{m} \tau_j + P_s} = \frac{\mu \sqrt{b_m}}{c_m \mu + P_s}. \quad (75)$$

Replacing η in (74) with (75), we have

$$p_i^\star = \begin{cases} \tau_i \left[\sqrt{\frac{\gamma_i}{b_m}} (P_s + c_m \mu) - \mu \right], & \text{for } i = 1, \dots, m \\ 0, & \text{for } i > m. \end{cases} \quad (76)$$

Now we search the optimal μ. Substituting (74) and (75) into the problem (39), then for a specified m, we obtain the following objective function:

$$\begin{aligned} h_m(\mu) &= \left(\sum_{i=1}^{m} \tau_i \gamma_i - \varepsilon^2 \right) \mu \\ &\quad - \frac{\mu^2}{\mu \sum_{j=1}^{m} \tau_i + P_s} \sum_{i=1}^{m} \sum_{j=1}^{m} \tau_i \tau_j \sqrt{\gamma_i \gamma_j} \\ &= \frac{(a_m c_m - b_m) \mu^2 + a_m P_s \mu}{c_m \mu + P_s} \end{aligned} \quad (77)$$

whose first- and second-order derivatives are given by

$$h_m'(\mu) = \frac{(a_m c_m - b_m) c_m \mu^2 + 2 (a_m c_m - b_m) P_s \mu + a_m P_s^2}{(c_m \mu + P_s)^2} \quad (78)$$

$$h_m''(\mu) = \frac{-2 b_m P_s^2}{(c_m \mu + P_s)^3} < 0 \quad (79)$$

indicating that $h_m(\mu)$ is a strictly concave function. It is important to remember that, once m is specified, μ is indirectly constrained by the assumption $\gamma_m > \eta \geq \gamma_{m+1}$, which leads to

$$\frac{P_s \sqrt{\gamma_{m+1}}}{\sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_{m+1}} \right)} \leq \mu < \frac{P_s \sqrt{\gamma_m}}{\sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_m} \right)}. \quad (80)$$

Define

$$\alpha_m \triangleq \frac{P_s \sqrt{\gamma_m}}{\sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_m} \right)}. \quad (81)$$

It is easily verified that

$$\alpha_{m+1} = \frac{P_s \sqrt{\gamma_{m+1}}}{\sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_{m+1}} \right)} \quad (82)$$

and α_m is decreasing as m increases. Define $\alpha_1 \triangleq +\infty$ and $\alpha_{r+1} \triangleq 0$. Then the region $\mu > 0$ can be divided into r consecutive intervals

$$\alpha_{m+1} \leq \mu < \alpha_m, \quad m = 1, \dots, r. \quad (83)$$

Consequently, taking into account all $m = 1, \dots, r$, the objective is actually a piecewise function as follows:

$$h(\mu) = \begin{cases} h_m(\mu), & \text{if } \mu \in [\alpha_{m+1}, \alpha_m], m = 1, \dots, r \\ 0, & \text{otherwise.} \end{cases} \quad (84)$$

Although complicated, it can be calculated that

$$\begin{aligned} h_m'(\alpha_m) &= c_m \gamma_m - 2 \sqrt{b_m} \sqrt{\gamma_m} + a_m \\ &= \sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_m} \right)^2 - \varepsilon^2 \\ &= \sum_{j=1}^{m-1} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_m} \right)^2 - \varepsilon^2 \end{aligned} \quad (85)$$

$$\begin{aligned} h_{m-1}'(\alpha_m) &= c_{m-1} \gamma_m - 2 \sqrt{b_{m-1}} \sqrt{\gamma_m} + a_{m-1} \\ &= \sum_{j=1}^{m-1} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_m} \right)^2 - \varepsilon^2. \end{aligned} \quad (86)$$

Since $h_m'(\mu)$ is smooth in $[\alpha_{m+1}, \alpha_m)$, and $h_m'(\alpha_m) = h_{m-1}'(\alpha_m)$ for $m = 2, \dots, r$, the piecewise function $h(\mu)$ is smooth too. Hence, according to Lemma 5, $h(\mu)$ is a concave function.

As a result, we have a convex problem

$$\underset{\mu \geq 0}{\text{maximize}} \;\; h(\mu) \quad (87)$$

whose solution can be found by setting $h'(\mu) = 0$ and taking the positive root. Assuming that the root lies in $[\alpha_{k+1}, \alpha_k)$, then in this region $h(\mu) = h_k(\mu)$ and $h'(\mu) = h_k'(\mu)$. Referring to the first-order derivative (78), $h_k'(\mu) = 0$ results in the following equation:

$$(a_k c_k - b_k) c_k \mu^2 + 2 (a_k c_k - b_k) P_s \mu + a_k P_s^2 = 0 \quad (88)$$

which admits only one positive root

$$\mu^\star = \frac{P_s}{c_k} \left(\sqrt{\frac{b_k}{b_k - a_k c_k}} - 1 \right). \quad (89)$$

Substituting (89) back into $h_k(\mu)$, the optimum value is then given by

$$h_k(\mu^\star) = \frac{P_s}{c_k^2} \left(\sqrt{b_k} - \sqrt{b_k - a_k c_k} \right)^2. \quad (90)$$

Note that the assumption $\mu^\star \in [\alpha_{k+1}, \alpha_k)$ leads to

$$\sum_{j=1}^{k} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_k} \right)^2 < \varepsilon^2 \leq \sum_{j=1}^{k} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_{k+1}} \right)^2. \quad (91)$$

Define

$$\beta_m \triangleq \sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_m} \right)^2. \quad (92)$$

It is easily verified that

$$\beta_{m+1} = \sum_{j=1}^{m} \tau_j \left(\sqrt{\gamma_j} - \sqrt{\gamma_{m+1}} \right)^2 \quad (93)$$

and β_m is increasing as m increases. Define $\beta_{r+1} \triangleq +\infty$. Then region $\varepsilon^2 > 0$ can also be divided into r consecutive intervals

$$\beta_m < \varepsilon^2 \leq \beta_{m+1}, m = 1, \dots, r. \quad (94)$$

Therefore, k can be found to be an integer such that $\varepsilon^2 \in (\beta_k, \beta_{k+1}]$. The proof of Theorem 3 is completed.

APPENDIX C
PROOF OF THEOREMS 4 AND 5

The robust QoS constraint in (53) can be written as

$$\min_{\boldsymbol{\Delta}\in\mathcal{E}} \mathrm{Tr}\left[\left(\hat{\mathbf{H}}+\boldsymbol{\Delta}\right)\mathbf{Q}\left(\hat{\mathbf{H}}+\boldsymbol{\Delta}\right)^{H}\right] \geq \rho \tag{95}$$

which enables us to replace the minimization by its dual problem again. Using the similar reasoning in the proof of Proposition 2 (note that here $\mu = 0$ will result in an infeasible case $0 \geq \rho$, assuming a positive QoS threshold), the constraint (95) can be equivalently replaced by

$$\max_{\mu>0} g(\mu) \geq \rho \tag{96}$$

where $g(\mu)$ is given by (66). The constraint (96) says that there must be a $\mu > 0$ such that $g(\mu) \geq \rho$. In other words, the QoS problem (53) is equivalent to

$$\begin{aligned} &\underset{\mathbf{Q}\succeq 0}{\text{minimize}} && S(\mathbf{Q}) \\ &\text{subject to} && g(\mu) \geq \rho,\ \mu > 0. \end{aligned} \tag{97}$$

Starting from (97), we can use the similar methods as in the proofs of Theorems 1 and 2 to show that, by setting $\mathbf{U}_q = \mathbf{U}_h$, the objective is unchanged or decreased while the constraints are still satisfied.

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Payaró, Dr. G. Scutari, and Dr. Y. Huang for the invaluable discussions.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.

[2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

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

[4] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, Jul. 1999.

[7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.

[8] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, pp. 1–4, Jan. 2001.

[9] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, pp. 3221–3231, Dec. 1994.

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

[11] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, pp. 2198–2206, Dec. 2001.

[12] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, pp. 1051–1064, May 2002.

[13] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, pp. 2381–2401, Sep. 2003.

[14] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Processing*, vol. 52, pp. 1179–1197, May 2004.

[15] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Found. Trends Commun. Inf. Theory*, vol. 3, no. 4–5, pp. 331–551, 2006.

[16] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2632–2639, Sep. 2001.

[17] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, pp. 3440–3453, Dec. 2004.

[18] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[19] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 662–671, Mar. 2006.

[20] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, pp. 2599–2613, Oct. 2002.

[21] L. Liu and H. Jafarkhani, "Application of quasi-orthogonal space-time block codes in beamforming," *IEEE Trans. Signal Processing*, vol. 53, pp. 54–63, Jan. 2005.

[22] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1673–1690, Jul. 2003.

[23] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, pp. 239–241, Jun. 2002.

[24] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 48, pp. 611–627, Mar. 2002.

[25] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with nonzero mean in space-time coded systems," *IEEE Trans. Signal Processing*, vol. 54, pp. 2318–2332, Jun. 2006.

[26] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Processing*, vol. 56, pp. 3678–3689, Aug. 2008.

[27] A. Soysal and S. Ulukus, "Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1402–1412, Sep. 2007.

[28] E. A. Jorswieck, A. Sezgin, H. Boche, and E. Costa, "Multiuser MIMO MAC with statistical CSI and MMSE receiver: Feedback strategies and transmitter optimization," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Vancouver, BC, Canada, Jul. 2006.

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

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

[31] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Processing*, vol. 51, pp. 1702–1715, Jul. 2003.

[32] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, pp. 313–324, Feb. 2003.

[33] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Processing*, vol. 53, pp. 1684–1696, May 2005.

[34] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Processing*, vol. 52, pp. 1931–1946, Jul. 2004.

[35] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, pp. 168–181, Jan. 2005.

[36] Y. Guo and B. C. Levy, "Robust MSE equalizer design for MIMO communication systems in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 54, pp. 1840–1852, May 2006.

[37] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multiaccess MIMO systems with imperfect channel state information," *IEEE Trans. Signal Processing*, vol. 53, pp. 3081–3090, Aug. 2005.
[38] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1707–1727, Jul. 2003.
[39] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.
[40] A. Wiesel, Y. C. Eldar, and S. Shamai, "Optimization of the MIMO compound capacity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1094–1101, Mar. 2007.
[41] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known MIMO communications channels," *IEEE Trans. Signal Processing*, vol. 53, pp. 2918–2930, Aug. 2005.
[42] A. Pascual-Iserte, D. P. Palomar, A. I. Pérez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with partial channel state information based on convex optimization," *IEEE Trans. Signal Processing*, vol. 54, pp. 346–360, Jan. 2006.
[43] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, "Robust transmit eigen beamforming based on imperfect channel state information," *IEEE Trans. Signal Processing*, vol. 54, pp. 1596–1609, May 2006.
[44] M. Payaró, A. Pascual-Iserte, and M. A. Lagunas, "Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1390–1401, Sep. 2007.
[45] M. B. Shenouda and T. N. Davidson, "Convex conic formulations of robust downlink precoder designswith quality of service constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 714–724, Dec. 2007.
[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
[47] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, Studies in Applied Mathematics, 1994, vol. 13.
[48] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.
[49] A. Ben-Tal and A. Nemirovski, "Lectures on modern convex optimization: Analysis, algorithms, and engineering applications," *MPS-SIAM Series on Optimization*, 2001.
[50] M. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.
[51] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
[52] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1319–1343, Jun. 2002.
[53] R. T. Rockafellar, *Convex Analysis*, 2nd ed. Princeton, NJ: Princeton Univ. Press, 1970.
[54] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM Studies in Applied Mathematics, 1994.
[55] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.
[56] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[57] M. Payaró, A. Wiesel, J. Yuan, and M. A. Lagunas, "On the capacity of linear vector Gaussian channels with magnitude knowledge and phase uncertainty," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP)*, Toulouse, France, May 2006.

Jiaheng Wang (S'08) received the B.E. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 2001 and 2006, respectively. He is currently working towards the Ph.D. degree in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology.

From 2001 to 2003, he was with the National Mobile Communications Research Lab, Southeast University, Nanjing, China, as a full research assistant. His research interests are mainly on optimization in signal processing, wireless communications and networks.

Daniel P. Palomar (S'99–M'03–SM'08) received the Electrical Engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

Since 2006, he has been an Assistant Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong. He has held several research appointments, namely, at King's College London (KCL), London, U.K.; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza", Rome, Italy; and Princeton University, Princeton, NJ.

Dr. Palomar is an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 special issue on "Convex Optimization for Signal Processing," was a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 special issue on "Game Theory in Communication Systems," as well as the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 special issue on "Optimization of MIMO Transceivers for Realistic Communication Networks." He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM). He received a 2004/06 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002/03 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC); the 2002/2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation; and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.