Robust MIMO Precoding for Several Classes of Channel Uncertainty

Jiaheng Wang, *Member, IEEE*, Mats Bengtsson, *Senior Member, IEEE*, Björn Ottersten, *Fellow, IEEE*, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—The full potential of multi-input multi-output (MIMO) communication systems relies on exploiting channel state information at the transmitter (CSIT), which is, however, often subject to some uncertainty. In this paper, following the worst-case robust philosophy, we consider a robust MIMO precoding design with deterministic imperfect CSIT, formulated as a maximin problem, to maximize the worst-case received signal-to-noise ratio or minimize the worst-case error probability. Given different types of imperfect CSIT in practice, a unified framework is lacking in the literature to tackle various channel uncertainty. In this paper, we address this open problem by considering several classes of uncertainty sets that include most deterministic imperfect CSIT as special cases. We show that, for general convex uncertainty sets, the robust precoder, as the solution to the maximin problem, can be efficiently computed by solving a single convex optimization problem. Furthermore, when it comes to unitarily-invariant convex uncertainty sets, we prove the optimality of a channel-diagonalizing structure and simplify the complex-matrix problem to a real-vector power allocation problem, which is then analytically solved in a waterfilling manner. Finally, for uncertainty sets defined by a generic matrix norm, called the Schatten norm, we provide a fully closed-form solution to the robust precoding design, based on which the robustness of beamforming and uniform-power transmission is investigated.**

***Index Terms*—Convex uncertainty set, imperfect CSIT, maximin, MIMO, minimax, saddle point, Schatten norm, unitarily-invariant uncertainty set, worst-case robustness.**

Manuscript received July 17, 2012; revised January 5, 2013; accepted March 10, 2013. Date of publication April 12, 2013; date of current version May 20, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Y.-W.Peter Hong. This work was supported in part by the National Basic Research Program of China (973) under 2013CB336600 and 2013CB329204, National Natural Science Foundation of China under 61201174, Natural Science Foundation of Jiangsu, China under BK2012325, and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 228044, and by the Hong Kong RGC 617911 research grant, and by the National Key Special Program of China no. 2012ZX03003005-003.

J. Wang was with the Signal Processing Laboratory, ACCESS Linnaeus Center, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. He is now with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China (e-mail: jhwang@seu.edu.cn).

M. Bengtsson is with ACCESS Linnaeus Center, Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden (e-mail: mats.bengtsson@ee.kth.se).

B. Ottersten is with ACCESS Linnaeus Center, Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden , and also with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg-Kirchberg L-1359, Luxembourg (email: bjorn.ottersten@ee.kth.se; bjorn.ottersten@uni.lu).

D. P. Palomar is with the Hong Kong University of Science and Technology, Hong Kong, and also with the Centre Tecnológic de Telecommunicacions de Catalunya-Hong Kong (CTTC-HK), Hong Kong (e-mail: 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.2013.2258016

I. INTRODUCTION

IT is well known that the performance of multi-input multi-output (MIMO) communication systems depends, to a substantial extent, on the quality of the channel state information (CSI) [1]. The full benefits of MIMO channels are achieved by exploiting CSI at the transmitter (CSIT) and adopting proper precoding techniques [2], [3]. With perfect CSIT, the optimal MIMO precoding has been well studied under various criteria [4]–[6] for either single-user or multi-user communications. In practice, however, CSIT is seldom perfect but subject to some uncertainty due to many practical issues, such as inaccurate channel estimation, quantization of CSI, erroneous or outdated feedback, and time delays or frequency offsets between the reciprocal channels. Therefore, the imperfection of CSIT has to be considered in MIMO precoding designs so that the communication system, on one hand, can fully utilize CSIT, and on the other hand, is robust to various imperfect CSIT.

In the literature, imperfect CSI is modeled by either stochastic or deterministic approaches. The stochastic model assumes that the channel is a random quantity and its instantaneous value is unknown but its statistics, such as the mean and/or the covariance, is known by the transmitter. In this case, the robust design usually aims at optimizing either the long-term average performance [7]–[10] or the outage performance [11]–[14]. On the other hand, the deterministic model, which is more suitable to characterize instantaneous CSI with errors, assumes that the actual channel lies in the neighborhood, often called the uncertainty set or region, of a nominal channel known by the transmitter. The size of this set represents the amount of uncertainty on the channel, i.e., the larger the set is the more uncertainty there is. In this case, a precoding design is said to be robust if it can achieve the best performance in the worst channel within the uncertainty set, or equivalently can guarantee a performance level for any channel in the uncertainty set. Such robust precoding designs can be achieved by optimizing the worst-case performance [15], often leading to a maximin or minimax problem [16]–[32]. Note that, worst-case robustness is related to statistical robustness in some situation. For example, many outage based robust designs are often transformed into deterministic formulations by defining an uncertainty set that has a certain probability [11], [12], [14].

The focus of this paper is on worst-case robust precoding designs based on deterministic imperfect CSIT. As an important branch of robust designs [15], the philosophy of worst-case robustness has been widely used in signal processing [16]–[19]

and communications [20]–[32]. In terms of MIMO communications, [20] studied the compound capacity [33] of a MIMO channel for an isotropically unconstrained uncertainty set, which was recently generalized to a class of norm-based uncertainty sets [21]. The worst-case robust minimum mean square error (MMSE) precoder was proposed in [22]. In [23] and [24], the authors tried to maximize the worst-case received signal-to-noise ratio (SNR) but only focused on a simplified power allocation problem by fixing the transmit directions. Interestingly, it was recently found in [25] and [26] that the transmit directions imposed in [23], [24] are optimal in some situations, which leads to fully analytical robust precoders along with some interesting insights. The worst-case robust precoding was also studied for MIMO multiaccess channels [27], broadcasting channels [28], multi-cell systems [30], [31], and cognitive radio systems [29], [32].

As a consequence of many practical factors that may cause imperfect CSIT, there are various channel uncertainty models commonly used in the literature. For example, the channel error induced by quantizing CSI is generally regarded to fall into a polyhedron [23]. As the most frequently used imperfect CSI model [16]–[29], the channel error is often covered by a sphere or ellipsoid uncertainty set that is usually defined by a matrix or vector norm, such as the Frobenius norm [22]–[25], [27]–[29] or the spectral norm [17], [21], [26], where the shape of the uncertainty set or the coverage of possible imperfection is determined by which norm is used. An uncertainty set can also be defined by other means, for example the Kullback-Leibler divergence [19]. Despite different types of imperfect CSIT that may be encountered in practice, most existing works on worst-case robust designs focused only on one or a few particular uncertainty sets, mainly based on the Frobenius and spectral norms due to their amenability. So far there lacks a unified framework on worst-case robust MIMO precoding that is applicable to various channel uncertainty. The major goal of this paper is to address this open problem.

To be more exact, in this paper we consider a worst-case robust MIMO precoding design, formulated as a maximin problem, to maximize the worst-case received SNR or to minimize the worst-case pairwise error probability (PEP) if a space-time block code (STBC) [34], [35] is used. In contrast with the existing works [16]–[22], [24]–[29] that are based on particular channel uncertainty sets, we try to take into account various channel uncertainty within a unified framework by considering several general classes of uncertainty sets that differ in generality and tractability. We provide robust precoding designs for arbitrary convex uncertainty sets, unitarily-invariant convex uncertainty sets, and the uncertainty sets defined by a generic matrix norm termed the Schatten norm. These general uncertainty sets contain all aforementioned deterministic imperfect CSIT models as special cases. Therefore, the previous works, e.g., [23]–[26], are included as special cases in this unified framework. We show that the formulated worst-case robust MIMO precoding design with the general uncertainty sets can be elegantly handled through convex optimization [36]. The main contributions of this paper are as follows.

We start from the most general case where the uncertainty set is a nonempty compact convex set, which covers all existing uncertainty sets, e.g., [16]–[22], [24]–[29], and may also be used to model more complicated deterministic imperfect CSIT. Note that [23] also considered a convex uncertainty set but only focused on a simplified power allocation problem by imposing some fixed transmit directions without knowing whether they are optimal or not. So far the optimal worst-case robust MIMO precoder for a general convex uncertainty set has still been unknown. In this paper, we solve this open problem by relating the formulated maximin problem to a minimax problem from the dual perspective, and showing that the robust precoder can be efficiently found by solving a single convex optimization problem. As a byproduct, the worst channel for the robust precoder can also be obtained simultaneously. Furthermore, we provide a practical reformulation of the convex problem that can be efficiently handled by general optimization methods as well as software packages.

Then, we consider the case where the convex uncertainty set is unitarily-invariant, which contains many uncertainty sets based on matrix norms, e.g., [16]–[18], [20]–[29]. It is shown that the robust precoder results in a favorable channel-diagonalizing structure, and thus the precoding design can be simplified to a power allocation problem without any loss of optimality. Note that such a desirable structure was only found to be optimal for the uncertainty sets based on some specific matrix norms in [25] and [26]. We further show that, given a particular form of the unitarily-invariant convex set, both the robust precoder and the worst channel can be simultaneously diagonalized, thus fully reducing the complex-matrix robust precoder design to a real-vector convex problem. Moreover, we show that the solution of such a real-vector problem can be analytically obtained via a convenient waterfilling fashion.

Finally, we consider uncertainty sets defined by the Schatten norm [37], which is a special case of unitarily-invariant convex sets, but still general enough to contain most frequently used uncertainty sets based on, e.g., the Frobenius norm [22]–[25], [27]–[29] or the spectral norm [17], [21], [26]. As a generic norm, the Schatten norm includes not only the Frobenius and spectral norms, but also other common matrix norms such as the nuclear norm [38]. In this case, we provide fully closed-form solutions to the robust precoder, based on which we also investigate the robustness of beamforming and equal power transmissions.

The paper is organized as follows. Section II introduces the system model, problem formulation, and various channel uncertainty models. Section III focuses on how to achieve the optimal robust precoder for general convex uncertainty sets in an efficient way. Section IV considers unitarily-invariant convex uncertainty sets and shows the optimality of the eigenmode transmission, while Section V focuses on uncertainty sets defined by the Schatten norm. Numerical results are provided in Section VI and Section VII concludes the paper.

Notation: Uppercase and lowercase boldface denote matrices and vectors, respectively. $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{S}_+$ denote the sets of real numbers, complex numbers, and positive semidefinite matrices, respectively. $\mathbf{I}$ and $\mathbf{1}$ represent the identity matrix and the vector of ones, respectively. $[\mathbf{A}]_{ij}$ denotes the (ith, jth) element of a matrix $\mathbf{A}$, and $\mathbf{e}_i$ denotes the ith column of the identity matrix. By $\mathbf{A} \succeq \mathbf{0}$ or $\mathbf{A} \succ \mathbf{0}$, we mean that $\mathbf{A}$ is a positive semidefinite or definite matrix, respectively. The operators $\geq$ and $\leq$ are defined componentwise for vectors and matrices. The operators $(\cdot)^H$, $(\cdot)^{-1}$, and $\mathrm{Tr}(\cdot)$ denote the Hermitian, inverse, and trace operations, respectively. $\boldsymbol{\sigma}(\mathbf{A})$ and $\boldsymbol{\lambda}(\mathbf{A})$ represent the vectors of singular values and eigenvalues of a matrix $\mathbf{A}$, respectively. The maximum eigenvalue of a Hermitian matrix is denoted by $\lambda_{\max}(\cdot)$. $\|\cdot\|$ denotes a general matrix norm as well as the Euclidean norm of a vector. $\|\cdot\|_*$, $\|\cdot\|_F$, and $\|\cdot\|_2$ denote the nuclear, Frobenius, and spectral norms of a matrix, respectively. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and image parts of a complex value, respectively. We define $(x)_+ \triangleq \max(x,0)$.

II. Problem Statement

A. System Model

Consider a narrowband point-to-point MIMO communication system equipped with N transmit and M receive antennas. Mathematically, the baseband, symbol-sampled system can be represented by a linear model

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

where $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} \in \mathbb{C}^M$ are the transmitted and received signals, respectively, $\mathbf{H} \in \mathbb{C}^{M\times N}$ is the channel matrix, 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., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$. The transmit strategy or precoding is determined by the transmit covariance matrix $\mathbf{Q} = E\{\mathbf{x}\mathbf{x}^H\}$. Indeed, via decomposing $\mathbf{Q} = \mathbf{F}\mathbf{F}^H$, the transmitted symbol vector $\mathbf{s}$, with $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$, can be linearly precoded by $\mathbf{F}$, resulting in $\mathbf{x} = \mathbf{F}\mathbf{s}$. In practice, the transmitter should satisfy the power constraint $\mathbf{Q} \in \mathcal{Q}$ where

$$\mathcal{Q} \triangleq \{\mathbf{Q} : \mathbf{Q} \succeq \mathbf{0},\ \mathrm{Tr}(\mathbf{Q}) \leq P\} \tag{2}$$

and $P \geq 0$ is the budget on the total transmit power.

Under the assumption of perfect CSIT, i.e., the channel $\mathbf{H}$ is perfectly known at the transmitter, the optimal MIMO precoding has been well studied for various criteria [4], [5]. However, due to many practical issues, CSIT is seldom perfect, which thus calls for robust precoding designs that can utilize CSIT and at the same time combat against its imperfection. To model imperfect CSIT, we consider a compound channel model [33] assuming that $\mathbf{H}$ belongs to a known set $\mathcal{H}$, often called an uncertainty set, of possible values but otherwise unknown. In the literature, this imperfect channel model has been widely used in robust designs, and the philosophy behind these robust designs is the so-called worst-case robustness [15], which is achieved by optimizing the system performance for the worst channel in $\mathcal{H}$ [16]–[29].

Specifically, we denote the system performance measure by a utility or payoff function $\Psi(\mathbf{Q},\mathbf{H})$. Then, the worst-case robust transmit strategy is given by the solution to the following maximin problem:

$$\max_{\mathbf{Q}\in\mathcal{Q}} \min_{\mathbf{H}\in\mathcal{H}} \ \Psi(\mathbf{Q},\mathbf{H}) \tag{3}$$

which, namely, offers the best performance for the worst channel within $\mathcal{H}$. As a counterpart of the maximin problem, we also introduce the following minimax problem:

$$\min_{\mathbf{H}\in\mathcal{H}} \max_{\mathbf{Q}\in\mathcal{Q}} \ \Psi(\mathbf{Q},\mathbf{H}) \tag{4}$$

which is, namely, to find the worst channel for the best one of all possible transmit strategies. We will show later that the maximin problem (3) and the minimax problem (4) are closely related.

In this paper, we assume perfect CSI at the receiver (CSIR) and adopt the following payoff or utility function:

$$\Psi(\mathbf{Q},\mathbf{H}) = \mathrm{Tr}(\mathbf{H}\mathbf{Q}\mathbf{H}^H) \tag{5}$$

which is proportional to the received SNR. It can be verified (see Section II in [25]) that maximizing $\Psi(\mathbf{Q},\mathbf{H})$ corresponds to: 1) maximizing the received SNR; 2) minimizing the pairwise error probability (PEP) if a space-time block code (STBC) [34], [35] is used at the transmitter; 3) maximizing a low-SNR approximation of the mutual information; 4) minimizing a low-SNR approximation of the MSE if a linear MMSE equalizer is used at the receiver.

B. Channel Uncertainty Models

In the literature [16]–[29], the uncertainty set $\mathcal{H}$ is often modeled as a neighborhood of a nominal channel $\hat{\mathbf{H}}$ known by the transmitter, where the nominal channel $\hat{\mathbf{H}}$ could be an estimate or feedback of the actual channel $\mathbf{H}$. By defining the channel error $\boldsymbol{\Delta}$ as the difference between the nominal channel and the actual channel as $\boldsymbol{\Delta} \triangleq \hat{\mathbf{H}} - \mathbf{H}$, the uncertainty $\mathbf{H} \in \mathcal{H}$ can be equally described by $\boldsymbol{\Delta} \in \mathcal{E}$ for some set $\mathcal{E}$. Correspondingly, we can rewrite the utility function in (5) based on $\boldsymbol{\Delta}$ as

$$\Psi(\mathbf{Q},\boldsymbol{\Delta}) \triangleq \mathrm{Tr}\left((\hat{\mathbf{H}} - \boldsymbol{\Delta})\mathbf{Q}(\hat{\mathbf{H}} - \boldsymbol{\Delta})^H\right). \tag{6}$$

and thus the maximin and minimax problems (3) and (4) based on $\mathcal{H}$ can be expressed as

$$\max_{\mathbf{Q}\in\mathcal{Q}} \min_{\boldsymbol{\Delta}\in\mathcal{E}} \ \Psi(\mathbf{Q},\boldsymbol{\Delta}) \tag{7}$$

and

$$\min_{\boldsymbol{\Delta}\in\mathcal{E}} \max_{\mathbf{Q}\in\mathcal{Q}} \ \Psi(\mathbf{Q},\boldsymbol{\Delta}) \tag{8}$$

based on $\mathcal{E}$, respectively.

The channel uncertainty set $\mathcal{E}$ provides a convenient way to characterize different types of imperfect CSIT that are caused by different reasons in practice. However, most existing works on worst-case robust MIMO precoding designs, e.g., [16]–[22], [24]–[29], considered only one or several particular choices of the uncertainty set $\mathcal{E}$, mostly focusing on the uncertainty sets based on the Frobenius and spectral norms due to their tractability. So far a unified framework that is applicable to various channel uncertainty sets is still absent. The major goal of this paper is to address this open problem. To be more exact, we consider various kinds of channel uncertainty sets that differ in terms of generality and tractability. Specifically, the uncertainty set $\mathcal{E}$ could be given in any one of the following forms:

1) General Convex Sets: In the most general case, we assume that $\mathcal{E}$ is a nonempty compact convex set, which covers all common uncertainty models as special cases [16]–[19], [21]–[29]. For example, if $\hat{\mathbf{H}}$ results from uniformly quantizing the elements of $\mathbf{H}$ with a step κ, the uncertainty set can be defined as [23]

$$\mathcal{E}_q \triangleq \left\{\mathbf{\Delta} : |\mathrm{Re}\{[\mathbf{\Delta}]_{ij}\}| \le \frac{\kappa}{2},\ |\mathrm{Im}\{[\mathbf{\Delta}]_{ij}\}| \le \frac{\kappa}{2}, \quad \forall i, j\right\}. \tag{9}$$

If the maximin and minimax problems (7) and (8) with a general convex uncertainty set can be solved, so can the special cases. Note that, although [23] also considered a general convex uncertainty set, the authors only focused on a power allocation problem, simplified from (7) by imposing a possibly suboptimal structure on $\mathbf{Q}$. In contrast, we are interested in finding globally optimal robust MIMO precoder in an efficient way. As shown in Section III, this goal can be accomplished by solving a single convex problem.

2) Unitarily-Invariant Convex Sets: In this case, we assume that the uncertainty set $\mathcal{E}$, in addition to being convex, is unitarily-invariant, i.e., $\mathbf{\Delta} \in \mathcal{E}$ implies $\mathbf{U}\mathbf{\Delta}\mathbf{V}^H \in \mathcal{E}$ for arbitrary unitary matrices $\mathbf{U}$ and $\mathbf{V}$, which is still general enough to include most frequently used uncertainty models [16]–[18], [21]–[29]. It is shown in Section IV that the unitarily-invariant condition leads to a favorable channel-diagonalizing structure. We then specify a concrete form of the general unitarily-invariant set as

$$\mathcal{E}_\sigma \triangleq \{\mathbf{\Delta} : f_n(\boldsymbol{\sigma}(\mathbf{\Delta})) \le \varepsilon,\ \forall n\} \tag{10}$$

where each $f_n(\mathbf{x})$ is a symmetric[1] and componentwise nondecreasing function and $f_n(\boldsymbol{\sigma}(\mathbf{\Delta}))$ is convex in $\mathbf{\Delta}$. Note that $\mathcal{E}_\sigma$ contains a number of uncertainty sets defined by matrix norms. We show in Section IV that, for $\mathcal{E} = \mathcal{E}_\sigma$, searching the complex-matrix robust precoder can be simplified to solving a real-vector convex problem. We further show that such a convex problem can be solved in a waterfilling fashion.

3) Uncertainty Sets Based on Matrix Norms: In the literature, the most common way to model the uncertainty of a matrix channel is to use some matrix norm [16]–[18], [21]–[29]. In this paper, we are particularly interested in a generic norm, called the Schatten norm, introduced in Section V. Several well-known examples of the Schatten norm are the nuclear norm $\|\cdot\|_*$ (also known as the trace norm), the Frobenius norm $\|\cdot\|_F$, and the spectral norm $\|\cdot\|_2$ (also known as the 2-norm), based on which we define

$$\mathcal{E}_* \triangleq \{\mathbf{\Delta} : \|\mathbf{\Delta}\|_* \le \varepsilon\} = \left\{\mathbf{\Delta} : \mathrm{Tr}\left((\mathbf{\Delta}^H\mathbf{\Delta})^{\frac{1}{2}}\right) \le \varepsilon\right\} \tag{11}$$

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

$$\mathcal{E}_2 \triangleq \{\mathbf{\Delta} : \|\mathbf{\Delta}\|_2 \le \varepsilon\} = \left\{\mathbf{\Delta} : \lambda_{\max}(\mathbf{\Delta}^H\mathbf{\Delta}) \le \varepsilon^2\right\}. \tag{13}$$

Note that a similar maximin robust design was studied in [23]–[25] for $\mathcal{E}_F$ and in [26] for $\mathcal{E}_2$. However, they are just special cases of the uncertainty set defined by the Schatten norm, which is in turn a subset of $\mathcal{E}_\sigma$ in (10). In this case, we analytically characterize the optimal robust precoder along with some interesting insights.

[1]A function $f(\mathbf{x})$ is symmetric if $f(\mathbf{x}) = f(\mathbf{\Pi}\mathbf{x})$ for any permutation matrix $\mathbf{\Pi}$.

III. Robust Precoder for General Convex Uncertainty Sets

We start from the most general case where the uncertainty set $\mathcal{E}$ is a nonempty compact convex set, and show in this section that the robust MIMO precoder, as the solution to the maximin problem (7), can be efficiently found through convex optimization.

A. Optimal Robust Precoder

Before solving the maximin problem (7), one natural question is whether it admits a solution. Observe that $\Psi(\mathbf{Q}, \mathbf{\Delta})$ is concave (actually linear) in $\mathbf{Q}$ for a fixed $\mathbf{\Delta}$ and convex (and quadratic) in $\mathbf{\Delta}$ for a fixed $\mathbf{Q}$, while the two sets $\mathcal{Q}$ and $\mathcal{E}$ are both nonempty compact convex sets. Therefore, according to [39, Corollary 37.6.2], there always exists a saddle point[2] providing a solution to the maximin problem (7) (and also providing a solution to the minimax problem (8)). Knowing that (7) (as well as (8)) is solvable, now we can focus on the key question on how to find the optimal robust precoder efficiently. The following result provides a positive answer to this question.

Theorem 1: Suppose that $\mathcal{E}$ is a nonempty compact convex set and $\mathcal{Q}$ is defined in (2). Consider the following convex problem:

$$\begin{aligned} &\underset{\mathbf{\Delta}\in\mathcal{E}, t}{\text{minimize}} && Pt \\ &\text{subject to} && (\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta}) \preceq t\mathbf{I} \end{aligned} \tag{14}$$

and let $\mathbf{Z}^\star$ be the optimal Lagrange multiplier associated with the constraint $(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta}) \preceq t\mathbf{I}$. Then, $\mathbf{Z}^\star$ is the optimal solution to the maximin problem (7).

Proof: To show that $\mathbf{Z}^\star$ is a solution to (8), we write the partial Lagrangian of (14) as

$$\begin{aligned} L(\mathbf{\Delta}, t; \mathbf{Z}) &= Pt + \mathrm{Tr}\left(\mathbf{Z}\left((\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta}) - t\mathbf{I}\right)\right) \\ &= (P - \mathrm{Tr}(\mathbf{Z}))\, t + \mathrm{Tr}\left(\mathbf{Z}(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta})\right) \end{aligned} \tag{15}$$

with Lagrange multiplier $\mathbf{Z} \succeq \mathbf{0}$. The dual function is given by

$$G(\mathbf{Z}) = \inf_{\mathbf{\Delta}\in\mathcal{E}, t} L(\mathbf{\Delta}, t; \mathbf{Z}) \tag{16}$$

whose domain is $\mathcal{Z} \triangleq \{\mathbf{Z} : \mathbf{Z} \succeq \mathbf{0}, G(\mathbf{Z}) > -\infty\}$. To guarantee that $G(\mathbf{Z})$ is bounded from below, it follows that $P - \mathrm{Tr}(\mathbf{Z}) = 0$. As a result, we have

$$\mathcal{Z} = \{\mathbf{Z} : \mathbf{Z} \succeq \mathbf{0},\ \mathrm{Tr}(\mathbf{Z}) = P\} \tag{17}$$

$$G(\mathbf{Z}) = \min_{\mathbf{\Delta}\in\mathcal{E}} \mathrm{Tr}\left(\mathbf{Z}(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta})\right) \tag{18}$$

so that the dual problem of (14) is

$$\max_{\mathbf{Z}\in\mathcal{Z}} \min_{\mathbf{\Delta}\in\mathcal{E}} \mathrm{Tr}\left(\mathbf{Z}(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta})\right). \tag{19}$$

[2]A point $(\mathbf{x}^\star, \mathbf{y}^\star)$ is said to be a saddle point of the function $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ with respect to maximizing over $\mathcal{X}$ and minimizing over $\mathcal{Y}$ if $f(\mathbf{x}, \mathbf{y}^\star) \le f(\mathbf{x}^\star, \mathbf{y}^\star) \le f(\mathbf{x}^\star, \mathbf{y}), \forall \mathbf{x} \in \mathcal{X}, \forall \mathbf{y} \in \mathcal{Y}$.

Note that the constraint $\mathrm{Tr}(\mathbf{Z}) = P$ in (19) can be relaxed to $\mathrm{Tr}(\mathbf{Z}) \leq P$, since the optimal $\mathbf{Z}$ is always achieved with equality. Now, comparing (19) with (7), one can find that they are exactly the same with $\mathbf{Z} = \mathbf{Q}$. Therefore, the optimal Lagrange multiplier $\mathbf{Z}^\star$ is also the optimal solution to (7). ■

Given that $(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta})$ is a matrix convex function of $\mathbf{\Delta}$ in the positive semidefinite space $\mathbb{S}_+^N$ [36], $(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta}) \preceq t\mathbf{I}$ is a convex constraint and (14) is indeed a convex problem. Theorem 1 indicates that, for a general convex uncertainty set, the optimal robust MIMO precoder can be efficiently found by solving a single convex optimization problem, i.e., (14), which consists of a differentiable objective function and a compact convex feasible set, thus being solvable by the common gradient-based numerical methods. Note that a similar convex uncertainty set was also considered in [23], which, however, only addressed a simplified power allocation problem by imposing suboptimal transmit directions, thus leading to a suboptimal robust precoder, whereas our solution is globally optimal. The performance gain of our globally optimal robust precoder over the suboptimal one in [23] is shown in Section VI.

Observing that the robust precoder is given by the optimal Lagrange multiplier of (14), one may be curious about what is the physical meaning of the optimal solution of (14). This question is answered by the following result, which establishes a close relation between the maximin and minimax problems (7) and (8).

Proposition 1: The minimax problem (8) is equivalent to the convex problem (14).

Proof: See Appendix A. ■

From Proposition 1, one can see that the convex problem (14) offers not only the solution to the maximin problem (7), but also the solution to the minimax problem (8). Denote by $\mathbf{Q}^\star$ and $\mathbf{\Delta}^\star$ the solutions to the maximin and minimax problems (7) and (8), respectively. According to [40, Proposition 2.6.1], $(\mathbf{Q}^\star, \mathbf{\Delta}^\star)$ is in fact a saddle point of $\Psi(\mathbf{Q}, \mathbf{\Delta})$ over $\mathcal{Q}$ and $\mathcal{E}$. As we have pointed out that the maximin (as well as minimax) problem is solvable due to the existence of a saddle point of $\Psi(\mathbf{Q}, \mathbf{\Delta})$, now Theorem 1 along with Proposition 1 has actually presented an elegant way to compute such a saddle point, i.e., solving the convex problem (14). Meanwhile, the saddle point also implies that $\mathbf{\Delta}^\star$ is the worst channel error for the robust precoder.

B. Practical Reformulation

So far we have theoretically shown that the optimal robust precoder can be achieved by solving (14). However, it should be pointed out that, although (14) is a convex problem, the constraint $(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta}) \preceq t\mathbf{I}$ is given in the form of a matrix convex function defined in $\mathbb{S}_+^N$ [36] but not a linear matrix inequality (LMI), which causes a difficulty in solving (14), because most optimization methods as well as software packages are designed to deal with LMIs but not otherwise arbitrary convex matrix inequalities. Hence, one may wonder: is there any practical method to solve (14) and more importantly to obtain the optimal Lagrange multiplier of (14) conveniently? We provide a positive answer to this question by showing that one can solve, instead of (14), the following equivalent but more tractable problem.

Proposition 2: Let $(\mathbf{\Delta}^\star, t^\star)$ be the optimal solution to the following convex problem:

$$\begin{array}{ll} \underset{\mathbf{\Delta}\in\mathcal{E},t}{\text{minimize}} & Pt \\ \text{subject to} & \begin{bmatrix} t\mathbf{I} & (\hat{\mathbf{H}} - \mathbf{\Delta})^H \\ \hat{\mathbf{H}} - \mathbf{\Delta} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \end{array} \tag{20}$$

and let

$$\mathbf{Y}^\star \triangleq \begin{bmatrix} \mathbf{Y}_{11}^\star & \mathbf{Y}_{12}^\star \\ \mathbf{Y}_{21}^\star & \mathbf{Y}_{22}^\star \end{bmatrix} \in \mathbb{S}_+^{N+M},$$

where $\mathbf{Y}_{11}^\star \in \mathbb{S}_+^N$, $\mathbf{Y}_{22}^\star \in \mathbb{S}_+^M$, and $\mathbf{Y}_{12}^\star = \mathbf{Y}_{21}^{\star H} \in \mathbb{C}^{N\times M}$, be the optimal Lagrange multiplier associated with the constraint

$$\begin{bmatrix} t\mathbf{I} & (\hat{\mathbf{H}} - \mathbf{\Delta})^H \\ \hat{\mathbf{H}} - \mathbf{\Delta} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \tag{21}$$

Then, $(\mathbf{\Delta}^\star, t^\star)$ is also the optimal solution to (14), and $\mathbf{Z}^\star = \mathbf{Y}_{11}^\star$ is the optimal Lagrange multiplier associated with the constraint $(\hat{\mathbf{H}} - \mathbf{\Delta})^H(\hat{\mathbf{H}} - \mathbf{\Delta}) \preceq t\mathbf{I}$ in (14).

Proof: The equivalence between (20) and (14) can be easily established via the Schur complement. The difficulty lies in how to relate the optimal Lagrange multipliers of (14) and (20), which relies on exploring the general optimality conditions of (14) and (20). The detailed proof is given in Appendix B. ■

Proposition 2 combined with Theorem 1 provides a practical and efficient way to find the optimal robust MIMO precoder: one only needs to solve (20) and tailor its optimal Lagrange multiplier. Now that the constraint (21) is an LMI, as a very tractable form in practice, (20) can be efficiently solved by many software packages, e.g., CVX [41] or YALMIP [42]. Such software packages contain numerical methods, e.g., primal-dual interior-point methods [36], that can provide not only the optimal primal variables but also the optimal dual variables, i.e., Lagrange multipliers. As a by-product, the worst channel error for the robust precoder can also be obtained from the primal solution to (20). Summarizing, for a general convex uncertainty set, the robust precoder and the worst channel error can be simultaneously and efficiently computed by solving a standard convex problem.

IV. Robust Precoder for Unitarily-Invariant Convex Uncertainty Sets

In this section, we consider the uncertainty set $\mathcal{E}$ to be a unitarily-invariant convex set, which contains most channel uncertainty models used in practice [16]–[18], [20]–[28]. Since unitarily-invariant convex sets are special cases of general convex sets, the method proposed in the previous section to find the robust precoder in the general convex case is still applicable in special cases. However, we will show that the unitarily-invariant condition leads to a favorable channel-diagonalizing structure of both the robust precoder and worst channel error, and eventually simplifies searching the robust precoder to a waterfilling procedure.

A. General Unitarily-Invariant Sets

The unitary-invariance means that $\mathbf{\Delta} \in \mathcal{E} \Rightarrow \mathbf{U}\mathbf{\Delta}\mathbf{V}^H \in \mathcal{E}$ for any unitary matrices $\mathbf{U} \in \mathbb{C}^{M\times M}$ and $\mathbf{V} \in \mathbb{C}^{N\times N}$, i.e., the set $\mathcal{E}$ is invariant with respect to rotations on the left and the right. Combining the convexity and unitary-invariance of $\mathcal{E}$, we first show in the following that the optimal transmit directions, i.e., the eigenvectors of the optimal transmit covariance matrix, are just the right singular vectors of the nominal channel, thus leading to an eigenmode transmission. In this case, the matrix maximin problem (7) can be simplified into a vector power allocation problem without loss of any optimality. Before stating our result, let us introduce some notations. We denote the eigenvalue decomposition (EVD) of $\mathbf{Q}$ by $\mathbf{Q} = \mathbf{U}_q\mathbf{\Lambda}_q\mathbf{U}_q^H$ with eigenvalues $\mathbf{p} \triangleq \{p_i\}_{i=1}^N$, and the singular-value decomposition (SVD) of $\hat{\mathbf{H}}$ by $\hat{\mathbf{H}} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^H$.

Theorem 2: Suppose that $\mathcal{E}$ is a nonempty compact unitarily-invariant convex set and $\mathcal{Q}$ is defined in (2). Then, there exists a solution $\mathbf{Q}^\star$ to the maximin problem (7) such that $\mathbf{U}_q^\star = \mathbf{V}_h$ and $\mathbf{p}^\star$ is the solution to the following maximin problem:

$$\max_{\mathbf{p}\in\mathcal{P}} \min_{\mathbf{\Delta}\in\mathcal{E}} \sum_{i=1}^N p_i \mathbf{e}_i^H(\mathbf{\Sigma}_h - \mathbf{\Delta})^H(\mathbf{\Sigma}_h - \mathbf{\Delta})\mathbf{e}_i \tag{22}$$

where $\mathcal{P} \triangleq \{\mathbf{p} : \mathbf{p} \geq \mathbf{0}, \mathbf{1}^T\mathbf{p} \leq P\}$ and $\mathbf{e}_i$ is the ith column of the identity matrix.

Proof: The proof is based on showing that by using $\mathbf{U}_q = \mathbf{V}_h$, the objective of (7) reaches an upper bound and the power constraint is still satisfied. The detailed proof is given in Appendix C. ■

It follows from Theorem 2 that the unitary-invariance of the uncertainty set $\mathcal{E}$ is a sufficient condition guaranteeing the optimality of the eigenmode transmission. Among existing works, [23] and [24] imposed the same transmit directions but without knowing whether or when they were optimal, whereas [25] and [26] proved the optimality of the similar channel-diagonalizing structure but only for the uncertainty sets defined by the Frobenius and spectral norms[3] (as special cases of the current proposed framework), i.e., $\mathcal{E}_F$ in (12) and $\mathcal{E}_2$ in (13), respectively. Therefore, our result indicates for the first time that the eigenmode transmission is optimal in terms of worst-case robustness for a general class of uncertainty sets, i.e., unitarily-invariant convex sets.

To find the optimal power allocation, one still needs to solve a maximin problem. Interestingly, similar to the case of general convex uncertainty sets, it was shown in [23] that the solution to the maximin problem (22) can also be obtained by solving a single convex problem as follows:

$$\begin{aligned} \underset{\mathbf{\Delta}\in\mathcal{E},t}{\text{minimize}} \quad & Pt \\ \text{subject to} \quad & \mathbf{e}_i^H(\mathbf{\Sigma}_h - \mathbf{\Delta})^H(\mathbf{\Sigma}_h - \mathbf{\Delta})\mathbf{e}_i \leq t, \\ & i = 1,\ldots,N \end{aligned} \tag{23}$$

where the robust power allocation is given by the optimal Lagrange multipliers associated with the constraints $\mathbf{e}_i^H(\mathbf{\Sigma}_h - \mathbf{\Delta})^H(\mathbf{\Sigma}_h - \mathbf{\Delta})\mathbf{e}_i \leq t$, $i = 1,\ldots,N$ (see [23, Proposition 1]).

[3]Note that the method used in [25] and [26] was based on analytically solving the inner minimization of the maximin problem, which, however, cannot be applied to a general uncertainty set.

So far, we have considered general unitarily-invariant convex uncertainty sets that reduce the maximin problem (7) to the power allocation problem (22) as a result of the diagonalizing structure of the precoder $\mathbf{Q}$. One natural question is whether the channel error $\mathbf{\Delta}$ can also be diagonalized so that the problem (22) or (23) is further simplified to a real-vector problem without involving complex matrices. To answer this question, we examine a particular form of the unitarily-invariant convex uncertainty set based on the singular values of $\mathbf{\Delta}$, in the next section.

B. Unitarily-Invariant Sets Based on Singular Values

In this section, we consider the uncertainty set $\mathcal{E}_\sigma$ defined in (10), as a specific form of the unitarily-invariant convex set. Basically, $\mathcal{E}_\sigma$ is the intersection of the sublevels of $f_n(\boldsymbol{\sigma}(\mathbf{\Delta}))$, $\forall n$, where each $f_n(\mathbf{x})$ is a symmetric and componentwise nondecreasing function and $f_n(\boldsymbol{\sigma}(\mathbf{\Delta}))$ is convex in $\mathbf{\Delta}$. Note that the uncertainty set $\mathcal{E}_\sigma$ is still general enough to include many channel uncertainty models [16]–[18], [21]–[28]. More importantly, in this case, we show that both the robust precoder $\mathbf{Q}$ and the channel error $\mathbf{\Delta}$ admit channel-diagonalizing structures, which simplifies searching the complex-matrix robust precoder to solving a real-vector problem.

Denote the SVD of $\hat{\mathbf{H}}$ by $\hat{\mathbf{H}} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^H$ with singular values $\{\gamma_i\}_{i=1}^N$, where $\gamma_i = 0$ for $i > \min\{M,N\}$, and the SVD of $\mathbf{\Delta}$ by $\mathbf{\Delta} = \mathbf{U}_\delta\mathbf{\Sigma}_\delta\mathbf{V}_\delta^H$ with singular values $\boldsymbol{\delta} = \boldsymbol{\sigma}(\mathbf{\Delta}) \triangleq \{\delta_i\}_{i=1}^N$, where $\delta_i = 0$ for $i > \min\{M,N\}$. Then, we have the following result.

Theorem 3: Suppose that $\mathcal{E} = \mathcal{E}_\sigma$ defined in (10) and $\mathcal{Q}$ is defined in (2). Let $\mathcal{P} \triangleq \{\mathbf{p} : \mathbf{p} \geq \mathbf{0}, \mathbf{1}^T\mathbf{p} \leq P\}$ and $\mathcal{D} \triangleq \{\boldsymbol{\delta} : f_n(\boldsymbol{\delta}) \leq \varepsilon,\ \forall n\}$. Then, the following statements hold.

1) There exists a solution $\mathbf{Q}^\star$ to the maximin problem (7) such that $\mathbf{U}_q^\star = \mathbf{V}_h$ and $\mathbf{p}^\star$ is the solution to the following maximin problem:

$$\max_{\mathbf{p}\in\mathcal{P}} \min_{\boldsymbol{\delta}\in\mathcal{D}} \sum_{i=1}^N (\gamma_i - \delta_i)^2 p_i. \tag{24}$$

2) There exists a solution $\mathbf{\Delta}^\star$ to the minimax problem (8) such that $\mathbf{U}_\delta^\star = \mathbf{U}_h$, $\mathbf{V}_\delta^\star = \mathbf{V}_h$, and $\boldsymbol{\delta}^\star$ is the solution to the following minimax problem:

$$\min_{\boldsymbol{\delta}\in\mathcal{D}} \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^N (\gamma_i - \delta_i)^2 p_i. \tag{25}$$

Proof: See Appendix D. ■

Theorem 3 reveals that, for $\mathcal{E} = \mathcal{E}_\sigma$, both the robust transmit covariance matrix and the worst channel error align with the nominal channel, resulting in a fully channel-diagonalizing structure. In this case, the complex-matrix maximin and minimax problems (7) and (8) can be simplified respectively into the real-vector maximin and minimax problems (24) and (25) without loss of any optimality. Consequently, searching the complex-matrix robust precoder (or worst channel error) reduces to searching its eigenvalues (or singular values), which significantly decreases the computational complexity. Moreover, similar to Theorem 1, the simplified real-vector maximin and minimax problems (24) and (25) are linked by the following convex problem:

$$\begin{aligned}&\underset{\boldsymbol{\delta}\in\mathcal{D},t}{\text{minimize}} \quad Pt\\ &\text{subject to} \quad (\gamma_i-\delta_i)^2\le t,\ i=1,\ldots,N \end{aligned}\tag{26}$$

which is equivalent to the minimax problem (25) and whose the optimal Lagrange multipliers $\boldsymbol{\eta}^\star \triangleq \{\eta_i^\star\}_{i=1}^N$ associated with the constraints $(\gamma_i-\delta_i)^2\le t$, $i=1,\ldots,N$, provide the optimal solution to the maximin problem (24). Therefore, the robust precoder as well as the worst channel error can be obtained by solving (26) eventually.

The unitarily-invariant convex set $\mathcal{E}_\sigma$ covers many channel uncertainty models as special cases, of which two commonly used ones are $\mathcal{E}_F$ in (12) and $\mathcal{E}_2$ in (13), defined by the Frobenius norm [22]–[25], [27], [28] and the spectral norm [17], [21], [26], respectively. Note that these are just two special cases of uncertainty sets defined by the Schatten norm (see Section V). Yet, before going into particular examples of $\mathcal{E}_\sigma$, we will investigate further the optimization problem (26) and show that its solution can be obtained via a waterfilling procedure.

C. Waterfilling Solution for Sets Based on Singular Values

Given the set $\mathcal{D}$ defined in Theorem 3, it is not difficult to see that a solution to the following problem is also a solution to (26):

$$\begin{aligned}&\underset{\boldsymbol{\delta}}{\text{minimize}} \quad \max_{i=1,\ldots,N}(\gamma_i-\delta_i)\\ &\text{subject to} \quad \boldsymbol{\delta}\in\mathcal{D},\ 0\le\delta_i\le\gamma_i,\quad \forall i. \end{aligned}\tag{27}$$

Assume without loss of generality (w.l.o.g.) that $\gamma_1\ge\cdots\ge\gamma_N$. We are particularly interested in characterizing the solution of (27) or equivalently (26) in the following two situations: 1) a coupled constraint set

$$\mathcal{D}_c=\{\boldsymbol{\delta}: f(\boldsymbol{\delta})\le\varepsilon\}\tag{28}$$

where $f:\mathbb{R}^N\to\mathbb{R}$ is a symmetric, strictly increasing, differentiable, and convex function; and 2) a decoupled constraint set

$$\mathcal{D}_d=\{\boldsymbol{\delta}: g(\delta_i)\le\varepsilon,\ i=1,\ldots,N\}\tag{29}$$

where $g:\mathbb{R}\to\mathbb{R}$ is an invertible strictly increasing function. In the former case, one can imagine $\mathcal{D}_c$ as a bottle of water, where the bottle is f, the water is $\boldsymbol{\delta}$, and the water volume is ε. In the latter case, $\mathcal{D}_d$ can be regarded as N bottles of water, where each bottle g is designated to hold the water δ_i.

Let us first focus on the coupled case. Intuitively, to minimize $\max_i(\gamma_i-\delta_i)$, δ_1 should first compensate the difference $\gamma_1-\gamma_2$, then δ_1 and δ_2 together compensate the difference $\gamma_2-\gamma_3$ and so on. As shown in Fig. 1(a), the whole process is like pouring the bottle of water $\mathcal{D}_c$ into the container $\boldsymbol{\gamma}\triangleq\{\gamma_i\}_{i=1}^N$, where the water level μ is given by $\mu=\gamma_1-\delta_1$. This waterfilling procedure is rigorously characterized in Appendix E, based on which we can provide a closed-form solution to (26). To this end, we define $\gamma_{N+1}=0$ and define for $k=1,\ldots,N+1$

$$\begin{aligned}\mathbb{R}^N\ni\boldsymbol{\theta}_k &\triangleq [\underbrace{\gamma_1-\gamma_k,\gamma_2-\gamma_k,\ldots,\gamma_{k-1}-\gamma_k}_{k-1},0,\ldots,0]^T\\ &=\max\{\boldsymbol{\gamma}-\gamma_k\mathbf{1},\ \mathbf{0}\}.\end{aligned}\tag{30}$$

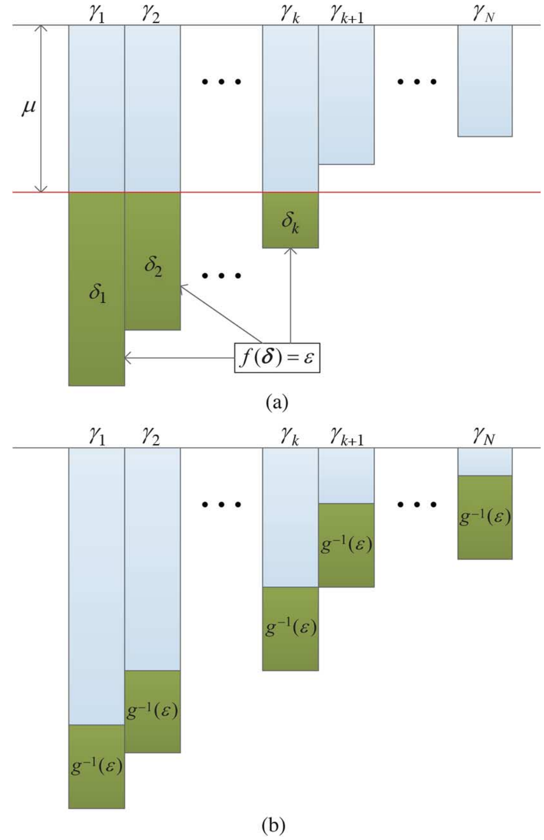


Fig. 1. Solving (27) through a waterfilling procedure. (a) Coupled constraint set $\mathcal{D}_c$. (b) Decoupled constraint set $\mathcal{D}_d$.

Easily observe that $\boldsymbol{\theta}_1=\mathbf{0}$ and $\boldsymbol{\theta}_{N+1}=\boldsymbol{\gamma}$, and that $\boldsymbol{\theta}_k\le\boldsymbol{\theta}_{k+1}$ so $f(\boldsymbol{\theta}_k)\le f(\boldsymbol{\theta}_{k+1})$.

Theorem 4: Suppose that $\mathcal{D}=\mathcal{D}_c$ defined in (28) and $\varepsilon<f(\boldsymbol{\gamma})$.

1) The optimal solution $(\boldsymbol{\delta}^\star,t^\star)$ to (26) is given by $t^\star=\mu^{\star 2}$ and

$$\delta_i^\star=\begin{cases}\gamma_i-\mu^\star, & i=1,\ldots,k\\ 0, & i>k\end{cases}\tag{31}$$

where k is an integer such that

$$f(\boldsymbol{\theta}_k)<\varepsilon\le f(\boldsymbol{\theta}_{k+1})\tag{32}$$

and $\mu^\star\in[\gamma_{k+1},\gamma_k)$ is the root of the equation $f(\boldsymbol{\delta}^\star(\mu^\star))=\varepsilon$.

2) The optimal power allocation $\mathbf{p}^\star$ (or Lagrange multipliers $\boldsymbol{\eta}^\star$ associated with the constraints $(\gamma_i-\delta_i)^2\le t$, $\forall i$, in (26)) is given by

$$p_i^\star=\begin{cases}\frac{P\alpha_i^\star}{\sum_{j\le k}\alpha_j^\star}, & i=1,\ldots,k\\ 0, & i>k\end{cases}\tag{33}$$

where $\alpha_i^\star\triangleq\partial f(\boldsymbol{\delta}^\star)/\partial\delta_i$, $\forall i$.

Proof: See Appendix E. ■

The integer k is the number of active eigenmodes and can be easily determined from (32). Since $f(\boldsymbol{\delta})$ is an increasing function, $f(\boldsymbol{\delta}^\star(\mu^\star))$ is monotonically increasing in $\mu^\star$, meaning that the optimal water level $\mu^\star$ can be efficiently found via the bisection method over $[\gamma_{k+1}, \gamma_k)$. In some situations $\mu^\star$ may be obtained in a closed form (e.g., [25]). The assumption $\varepsilon < f(\boldsymbol{\gamma})$ is to avoid a trivial solution, since if the uncertainty is too large, i.e., $\varepsilon \geq f(\boldsymbol{\gamma})$, the best worst-case performance is zero. Note that similar waterfilling procedures were also investigated in [23] and [24] but only for the uncertainty set defined by the Frobenius norm, i.e., $\mathcal{E}_F$ in (12). It is interesting to see from (33) that the robust power allocation is actually proportional to the partial derivative of the constraint function f of each active eigenmode at the worst channel error, which is discovered for the first time.

Now we consider the decoupled case. Due to the monotonicity and invertibility of g, the decoupled constraint $g(\delta_i) \leq \varepsilon$ can be rewritten as $\delta_i \leq g^{-1}(\varepsilon)$, which determines the amount of water in the bottle i. As shown in Fig. 1(b), the procedure of solving (27) is like pouring the water in each bottle i into each container γ_i independently. Therefore, we obtain the following result.

Theorem 5: Suppose that $\mathcal{D} = \mathcal{D}_d$ defined in (29).

1) The optimal solution $(\boldsymbol{\delta}^\star, t^\star)$ to (26) is given by $t^\star = (\gamma_1 - \delta_1^\star)^2$ and

$$\delta_i^\star = \min\left\{\gamma_i, g^{-1}(\varepsilon)\right\}, \quad i = 1, \ldots, N. \tag{34}$$

2) The optimal power allocation $\mathbf{p}^\star$ (or Lagrange multipliers $\boldsymbol{\eta}^\star$ associated with the constraints $(\gamma_i - \delta_i)^2 \leq t$, $\forall i$, in (26)) is given by

$$p_i^\star = \begin{cases} P, & i = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

Proof: The first part is straightforward. The second part can be obtained by exploring the complimentary condition as shown in Appendix E. ■

Theorem 5 indicates that, in the decoupled case, the robust precoding turns out to be beamforming, one of the simplest transmit strategies using all power on one eigenmode. Because of its simplicity, beamforming is often regarded as a non-robust transmit strategy. However, our result reveals that beamforming is actually robust in some situations (see a further discussion in Section V).

V. Robust Precoder for Uncertainty Sets Based on Matrix Norms

It is very convenient to define the uncertainty set as a sublevel of a norm function, which is always a convex set due to the convexity of an arbitrary norm. Indeed, most existing works on worst-case robust designs, e.g., [16]–[18], [21]–[28], adopted some norm-based uncertainty set. In this section, instead of one particular norm, we consider a generic matrix norm, called the Schatten norm, that covers a range of common matrix norms as special cases.

Definition 1. ([37, Proposition 9.2.3]): Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ with $r = \min\{m, n\}$, and let $p \in [1, \infty]$. Then, the p-Schatten norm $\|\cdot\|_{\sigma p}$ is defined as

$$\|\mathbf{A}\|_{\sigma p} \triangleq \begin{cases} \left(\sum_{i=1}^r \sigma_i^p(\mathbf{A})\right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \sigma_{\max}(\mathbf{A}), & p = \infty. \end{cases} \tag{36}$$

Based on the p-Schatten norm $\|\cdot\|_{\sigma p}$, we define the uncertainty set

$$\mathcal{E}_{\sigma p} \triangleq \{\boldsymbol{\Delta} : \|\boldsymbol{\Delta}\|_{\sigma p} \leq \varepsilon\} \tag{37}$$

where, to avoid a trivial solution, we assume that $\varepsilon < \|\hat{\mathbf{H}}\|_{\sigma p}$. From the definition, it is easy to see that $\mathcal{E}_{\sigma p} \subseteq \mathcal{E}_\sigma$ for $1 \leq p \leq \infty$, and therefore $\mathcal{E}_{\sigma p}$ inherits all favorable properties of $\mathcal{E}_\sigma$ obtained in Section IV. The different p-Schatten norms are related through

$$\|\mathbf{A}\|_{\sigma\infty} \leq \|\mathbf{A}\|_{\sigma q} \leq \|\mathbf{A}\|_{\sigma p} \leq \|\mathbf{A}\|_{\sigma 1} \tag{38}$$

where $p, q \in [1, \infty]$ and $p \leq q$, and therefore we have

$$\mathcal{E}_{\sigma 1} \subseteq \mathcal{E}_{\sigma p} \subseteq \mathcal{E}_{\sigma q} \subseteq \mathcal{E}_{\sigma\infty}. \tag{39}$$

Some well-known examples of the Schatten norm include:

1) The Nuclear Norm (Also Known as the Trace Norm):

$$\|\mathbf{A}\|_* = \operatorname{Tr}\left((\mathbf{A}^H\mathbf{A})^{\frac{1}{2}}\right) = \sigma_1(\mathbf{A}) + \cdots + \sigma_r(\mathbf{A}) = \|\mathbf{A}\|_{\sigma 1} \tag{40}$$

based on which we have defined the uncertainty set $\mathcal{E}_* = \mathcal{E}_{\sigma 1}$ in (11). The nuclear norm can be regarded as a convex approximation of the rank of a matrix, and has been widely used in rank minimization for sparse signal processing [38]. Hence, $\mathcal{E}_*$ approximately describes the uncertainty on the rank of the channel error matrix $\boldsymbol{\Delta}$. Note that $\mathcal{E}_*$ is the smallest one of all $\mathcal{E}_{\sigma p}$.

2) The Frobenius Norm:

$$\begin{aligned}\|\mathbf{A}\|_F &= \left(\operatorname{Tr}(\mathbf{A}^H\mathbf{A})\right)^{\frac{1}{2}} = \left(\sigma_1^2(\mathbf{A}) + \cdots + \sigma_r^2(\mathbf{A})\right)^{\frac{1}{2}} \\ &= \|\mathbf{A}\|_{\sigma 2}\end{aligned} \tag{41}$$

based on which we have defined the uncertainty set $\mathcal{E}_F = \mathcal{E}_{\sigma 2}$ in (12). As the most frequently adopted model in the literature [22]–[25], [27], [28], $\mathcal{E}_F$ represents the uncertainty on the total "power" of all elements of $\boldsymbol{\Delta}$. Meanwhile, from the probabilistic point of view, $\|\boldsymbol{\Delta}\|_F^2 = \|\mathbf{H} - \hat{\mathbf{H}}\|_F^2$ is in fact a closed-form expression of the Kullback-Leibler divergence between the actual and nominal channel models with Gaussian noise [19].

3) The Spectral Norm (Also Known as the 2-Norm):

$$\|\mathbf{A}\|_2 = \lambda_{\max}^{\frac{1}{2}}(\mathbf{A}^H\mathbf{A}) = \sigma_{\max}(\mathbf{A}) = \|\mathbf{A}\|_{\sigma\infty} \tag{42}$$

based on which we have defined the uncertainty set $\mathcal{E}_2 = \mathcal{E}_{\sigma\infty}$ in (13). Intuitively, $\mathcal{E}_2$ models the maximum uncertainty on each eigenmode of the channel [17], [21], [26]. Indeed, we know from (39) that, given the same error radius ε, $\mathcal{E}_{\sigma p} \subseteq \mathcal{E}_{\sigma\infty}$ for $p \in [1, \infty]$. Hence, $\mathcal{E}_{\sigma\infty}$ is the most conservative one among all $\mathcal{E}_{\sigma p}$, modeling the largest channel error.

Since $\mathcal{E}_{\sigma p}$ is a subset of $\mathcal{E}_\sigma$, from Theorem 3, finding the complex-matrix robust precoder reduces to solving the real-vector maximin problem (7), which is in turn equivalent to solving the convex problem (26). Using the uncertainty set $\mathcal{E}_{\sigma p}$, the constraint set $\mathcal{D}$ in (26) is equal to $\mathcal{D}_{\sigma p} \triangleq \{\boldsymbol{\delta} : f_{\sigma p}(\boldsymbol{\delta}) \leq \varepsilon\}$ with

$$f_{\sigma p}(\boldsymbol{\delta}) \triangleq \begin{cases} \left(\sum_{i=1}^N \delta_i^p\right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \max_i\{\delta_i\}, & p = \infty \end{cases} \tag{43}$$

where $\delta_i = 0$ for $i > \min\{M, N\}$. Observe that $\mathcal{D}_{\sigma p}$ can be divided into two categories: 1) the coupled constraint sets $\mathcal{D}_{\sigma p}$ for $p \in [1, \infty)$; and 2) the decoupled constraint set $\mathcal{D}_{\sigma\infty}$. They correspond exactly to the coupled and decoupled cases in Section IV-C, meaning that we can obtain a closed-form solution in a waterfilling fashion as in Section IV-C.

Theorem 6: Suppose that $\mathcal{E} = \mathcal{E}_{\sigma p}$ defined in (37) and $\varepsilon < f_{\sigma p}(\boldsymbol{\gamma})$ for $p \in [1, \infty]$.

1) Let $\delta_i^\star = \gamma_i - \mu^\star$ for $i \le k$ and $\delta_i^\star = 0$ for $i > k$, where k is an integer such that $f_{\sigma p}(\boldsymbol{\theta}_k) < \varepsilon \le f_{\sigma p}(\boldsymbol{\theta}_{k+1})$ with $\boldsymbol{\theta}_k$ defined in (30), and $\mu^\star \in [\gamma_{k+1}, \gamma_k)$ is the root of $f_{\sigma p}(\boldsymbol{\delta}^\star(\mu^\star)) = \varepsilon$. Then, the optimal power allocation for $p \in [1, \infty)$ is given by
$$p_i^\star = \frac{P \delta_i^{\star p-1}}{\sum_{j \le k} \delta_j^{\star p-1}} \tag{44}$$
for $i \le k$ and $p_i^\star = 0$ for $i > k$.
2) Let $\delta_i^\star = \varepsilon, \forall i$. Then, the optimal power allocation for $p = \infty$ is given by $p_1^\star = P$ and $p_i^\star = 0$ for $i > 1$.

Proof: Theorem 6 is the result of applying Theorems 4 and 5 to $\mathcal{D}_{\sigma p}$ and $f_{\sigma p}$. ■

From Theorem 6, we can obtain some interesting insights on worst-case robust MIMO precoding. The following two corollaries concern the optimality of beamforming that transmits data only over one eigenmode, and the optimality of the equal power allocation.

Corollary 1: Suppose that $\mathcal{E} = \mathcal{E}_{\sigma p}$. The robust maximin MIMO precoding is beamforming over the largest eigenmode if either: 1) $p = \infty$; or 2) $p \in [1, \infty)$ and $\varepsilon \le \gamma_1 - \gamma_2$.

Corollary 2: Suppose that $\mathcal{E} = \mathcal{E}_{\sigma p}$. The robust maximin MIMO precoding allocates power equally on the active eigenmodes if either: 1) $p = 1$; or 2) $p \in (1, \infty]$ and $\gamma_1 = \gamma_2 = \cdots = \gamma_k$.

Beamforming is often regarded to be sensitive to imperfect CSIT [4], [5] because of its simplicity. However, our results (and also [26]) reveal that beamforming is actually a robust solution if either the uncertainty set is $\mathcal{E}_{\sigma\infty}$ ($\mathcal{E}_2$), or $\varepsilon \le \gamma_1 - \gamma_2$, i.e., the uncertainty is small or the channel is nearly rank-one. As the most conservative one of $\mathcal{E}_{\sigma p}$, $\mathcal{E}_{\sigma\infty}$ defines the maximum uncertainty on each eigenmode independently, so the robust transmit strategy shall, intuitively, put all power on the strongest eigenmode. Furthermore, one can imagine that, when the channel uncertainty or the size of the channel matrix becomes smaller, the gap between $\mathcal{E}_{\sigma\infty}$ and other uncertainty sets shall become smaller as well. Therefore, we can reasonably infer that beamforming, although might not be optimally robust, is a nearly robust transmit strategy when the channel uncertainty or the channel dimension is small, independently of the shape of the uncertainty set.

The uncertainty set $\mathcal{E}_{\sigma 1}$ ($\mathcal{E}_*$) represents another extreme case of $\mathcal{E}_{\sigma p}$, as it is the smallest one and thus the least conservative one of $\mathcal{E}_{\sigma p}$. Since $\mathcal{E}_{\sigma 1}$ approximately models the uncertainty on the rank of the channel error, the robust transmit strategy may not distinguish between the uncertainty on each eigenmode but treats all active eigenmodes equally, thus leading to an equal power allocation over the active eigenmodes. The number of active eigenmodes k, however, is determined by the total uncertainty, and especially $k \to \operatorname{rank}(\hat{\mathbf{H}})$ as $\varepsilon \to f_{\sigma p}(\boldsymbol{\gamma})$ for $p \in [1, \infty)$. For $\mathcal{E}_{\sigma p}$ with $p \in (1, \infty]$, the equal power allocation is generally not robust unless the channel gains of the active eigenmodes are all equal.

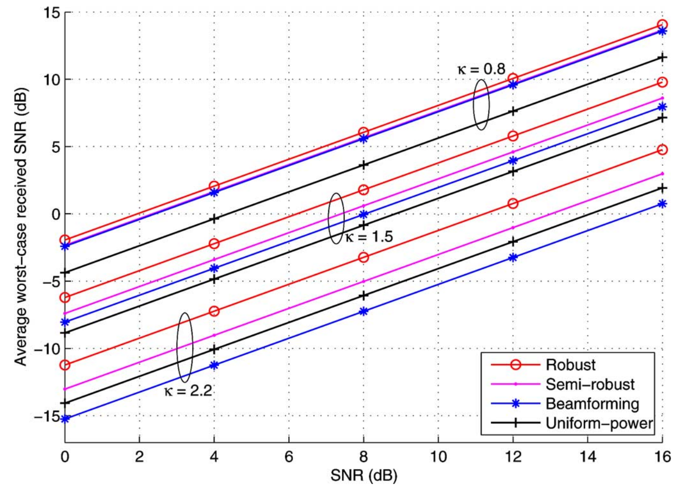


Fig. 2. Average worst-case received SNRs of four precoding strategies versus SNR at $\kappa = 0.8$, 1.5, and 2.2 for $\mathcal{E}_q$ and $M = N = 2$.

VI. Numerical Results

In this section, we demonstrate the effect of the robust MIMO precoding through several numerical examples. According to the philosophy of worst-case robustness, different precoding strategies are compared via their average worst-case performance, where the worst channel error for any given (either non-robust or robust) precoder can be obtained by solving the inner minimization of (7) for a fixed $\mathbf{Q}$ (note that the robust strategy and its worst channel error can be simultaneously obtained by solving (20)). Moreover, to take into account different channels, the worst-case performance is averaged over the nominal channel $\hat{\mathbf{H}}$, whose elements are randomly generated according to zero-mean, unit-variance, i.i.d. Gaussian distributions.

We first consider the uncertainty set $\mathcal{E}_q$ defined in (9), which models the channel uncertainty caused by uniformly quantizing the elements of the actual channel $\mathbf{H}$ with a step κ. The robust precoding, given by the solution of the maximin problem (7), is compared with the beamforming strategy that transmits only over the maximum eigenmode of $\hat{\mathbf{H}}$, the uniform-power strategy that allocates the transmit power equally over all eigenmodes of $\hat{\mathbf{H}}$, and the semi-robust strategy in [23] that provided a robust power allocation but with fixed (suboptimal) transmit directions. Fig. 2 shows the average worst-case received SNRs of the four strategies versus SNR for different quantization steps, and Fig. 3 displays the relation between the average worst-case received SNR and the quantization step. It can be clearly seen that the robust strategy always outperforms the non-robust or semi-robust strategies in terms of worst-case performance, and that the gain becomes larger as the uncertainty increases.

On the other hand, one can also observe from Figs. 2 and 3 that the performance of beamforming, although it is not a robust strategy for $\mathcal{E}_q$, is quite close to that of the robust precoding for small uncertainty. This example verifies our conclusion (in

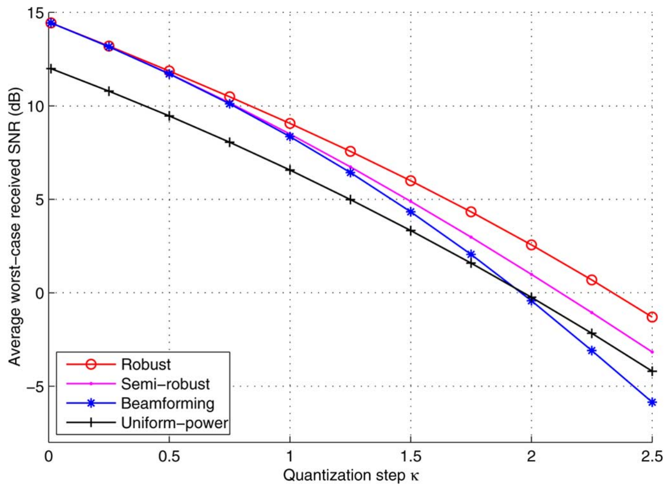


Fig. 3. Average worst-case received SNRs of four precoding strategies versus quantization step κ at $\mathrm{SNR} = 12$ dB for $\mathcal{E}_q$ and $M = N = 2$.

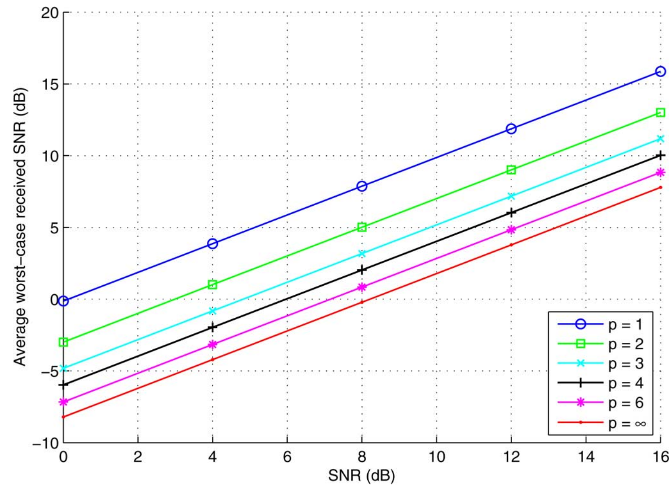


Fig. 4. Average worst-case received SNRs of robust precoding strategies versus SNR at $s = 0.6$ for different $\mathcal{E}_{\sigma p}$ and $M = N = 8$.

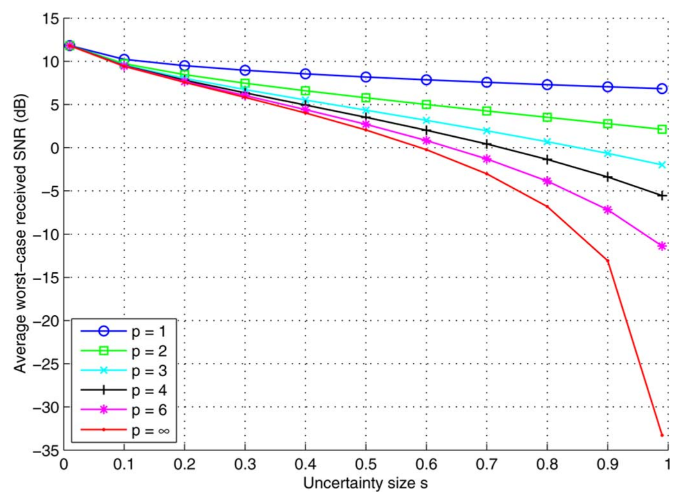


Fig. 5. Average worst-case received SNRs of robust precoding strategies versus uncertainty set size s at $\mathrm{SNR} = 8$ dB for different $\mathcal{E}_{\sigma p}$ and $M = N = 8$.

Section V and also [26]) that beamforming should be nearly robust when the channel uncertainty or the channel dimension is small, but independent of the shape of the uncertainty set. However, as the uncertainty increases, the performance of beamforming becomes worse and worse, and eventually is exceeded by that of the uniform-power strategy.

We then consider the unitarily-invariant uncertainty set $\mathcal{E}_{\sigma p}$ defined in (37) by the Schatten norm, where $p \in [1, \infty]$ determines the shape of $\mathcal{E}_{\sigma p}$, while the size of the uncertainty set is given by the error radius ε. Considering that $\|\cdot\|_{\sigma\infty}$ ($\|\cdot\|_2$) is the smallest one among all Schatten norms, we set a common error radius for all $\mathcal{E}_{\sigma p}$ such that $\varepsilon^2 = s\|\hat{\mathbf{H}}\|_2^2$ with $s \in (0, 1)$, so they can be reasonably compared. As shown in Section V, given the same error radius, the larger the parameter p is, the bigger and thus the more conservative the uncertainty set $\mathcal{E}_{\sigma p}$ is.

Fig. 4 shows the average worst-case received SNRs, achieved by the robust precoding strategies for different $\mathcal{E}_{\sigma p}$, versus SNR, while Fig. 5 displays the relation between the average worst-case received SNR and the uncertainty set size s. From these two figures, one can observe a tradeoff between the conservativeness of the uncertainty model and the system performance, i.e., the more conservative the uncertainty model is, the lower the performance is. Among all $\mathcal{E}_{\sigma p}$ with $p \in [1, \infty]$, $\mathcal{E}_{\sigma\infty}$ is the most conservative set, thus resulting in the lowest worst-case received SNR, whereas $\mathcal{E}_{\sigma 1}$ is the least conservative one, thus leading to the highest performance. In practice, the choice of an uncertainty set depends on the prediction of channel errors--large errors correspond to more conservative uncertainty sets, while small errors correspond to less conservative sets.

In Figs. 6 and 7, we plot the average worst-case symbol error rates (SERs), achieved by the robust precoding strategies for different $\mathcal{E}_{\sigma p}$, versus SNR and the uncertainty set size s, respectively, where we have used an 1/2-rate complex OSTBC [43] and an ML decoder at the receiver. Being consistent with Figs. 4 and 5, the trade off between conservativeness and performance can also be observed in Figs. 6 and 7.

VII. Conclusion

We have considered a robust MIMO precoding design, formulated as a maximin problem, to maximize the worst-case received SNR or minimize the worst-case PEP for an STBC with imperfect CSIT. Various kinds of channel uncertainty models have been taken into account. Specifically, we have considered three classes of general uncertainty models, including convex uncertainty sets, unitarily-invariant convex sets, and uncertainty sets defined by the Schatten norm, respectively, which cover most commonly used uncertainty models as special cases. We have related the formulated maximin problem with a minimax problem from the dual perspective, and shown that the robust MIMO precoder can be efficiently computed by solving a convex optimization problem, or given in an analytical form accompanied by a favorable channel-diagonalizing structure. Based on these results, we have obtained the globally optimal robust MIMO precoder along with the worst channel error, and also investigated the robustness of some common transmit strategies such as beamforming and equal power transmissions.

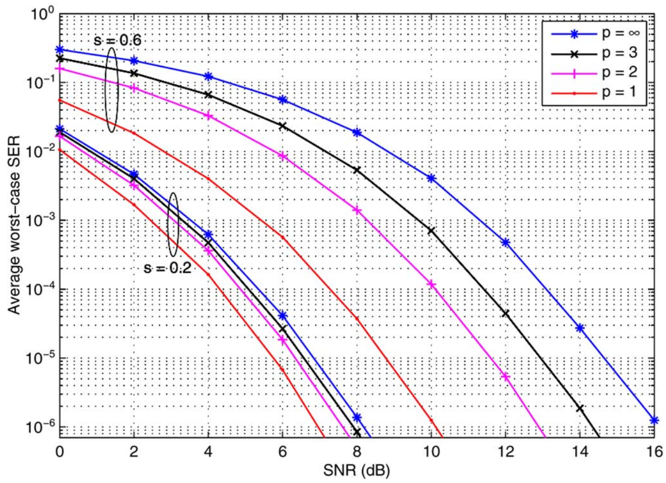


Fig. 6. Average worst-case SERs of robust precoding strategies versus SNR at $s = 0.2$ and 0.6 for different $\mathcal{E}_{\sigma p}$ and $M = N = 4$ with QPSK and 1/2-rate OSTBC.

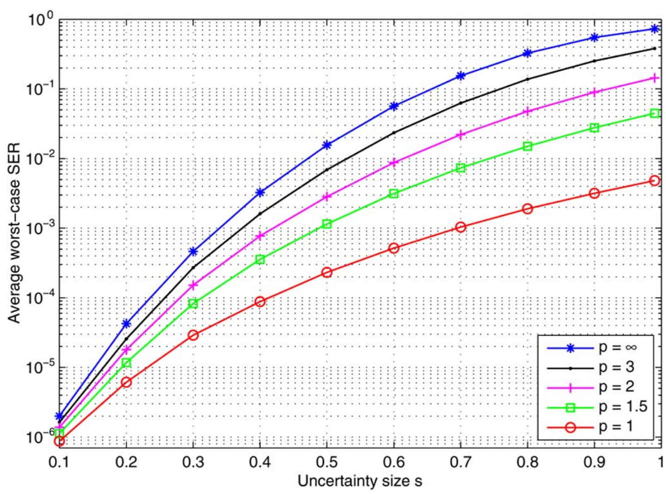


Fig. 7. Average worst-case SERs of robust precoding strategies versus uncertainty set size s at SNR $= 6$ dB for different $\mathcal{E}_{\sigma p}$ and $M = N = 4$ with QPSK and 1/2-rate OSTBC.

APPENDIX

A. Proof of Proposition 1

Lemma 1. ([37, Fact 8.18.18]): 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, $\mathrm{Tr}(\mathbf{AB}) \leq \sum_{i=1}^{N} \alpha_i \beta_i$.

Let $\mathbf{W}(\boldsymbol{\Delta}) \triangleq (\hat{\mathbf{H}} - \boldsymbol{\Delta})^H (\hat{\mathbf{H}} - \boldsymbol{\Delta})$, and denote the EVD of $\mathbf{Q}$ by $\mathbf{Q} = \mathbf{U}_q \boldsymbol{\Lambda}_q \mathbf{U}_q^H$ with eigenvalues $p_1 \geq \cdots \geq p_N$ and the EVD of $\mathbf{W}$ by $\mathbf{W} = \mathbf{U}_w \boldsymbol{\Lambda}_w \mathbf{U}_w^H$ with eigenvalues $w_1 \geq \cdots \geq w_N$. It follows from Lemma 1 that $\mathrm{Tr}(\mathbf{QW}) \leq \sum_{i=1}^{N} p_i w_i$, where the upper bound is achieved if $\mathbf{U}_q = \mathbf{U}_w$. Therefore, the inner maximization of the minimax problem (8) can be simplified to a linear program

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{i=1}^{N} p_i w_i \\ \text{subject to} \quad & \mathbf{p} \geq \mathbf{0},\ \mathbf{1}^T \mathbf{p} \leq P \end{aligned} \tag{45}$$

whose solution is to put all power on the largest w_1, i.e., $p_1 = P$ and $p_i = 0$ for $i \geq 2$. Thus, given any $\boldsymbol{\Delta}$, the inner maximum value is $P w_1 = P \lambda_{\max}(\mathbf{W}(\boldsymbol{\Delta}))$, and the minimax problem (8) can then be expressed as

$$\underset{\boldsymbol{\Delta} \in \mathcal{E}}{\text{minimize}}\ P \lambda_{\max}\left(\mathbf{W}(\boldsymbol{\Delta})\right). \tag{46}$$

Note that (46) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{\Delta} \in \mathcal{E}, t}{\text{minimize}} \quad & Pt \\ \text{subject to} \quad & \mathbf{W}(\boldsymbol{\Delta}) \preceq t\mathbf{I} \end{aligned} \tag{47}$$

which is exactly (14). The proof of Proposition 1 is thus completed.

B. Proof of Proposition 2

Using the Schur complement, it is easy to see that (20) and (14) are equivalent and thus share the same optimal solution or primal variables $(\boldsymbol{\Delta}^\star, t^\star)$. The question is how to relate the optimal Lagrange multipliers or dual variables $\mathbf{Y}^\star$ of (20) and $\mathbf{Z}^\star$ of (14). To this end, we investigate the optimality conditions of (20) and (14), respectively.

According to [40, Proposition 6.2.5], the optimal primal and dual variables of (14) by $(\boldsymbol{\Delta}^\star, t^\star)$ and $\mathbf{Z}^\star$ satisfy the following necessary and sufficient optimality conditions:

$$\boldsymbol{\Delta}^\star \in \mathcal{E}, \quad \mathbf{Z}^\star \succeq \mathbf{0}, \quad \mathrm{Tr}(\mathbf{Z}^\star) = P \tag{48}$$

$$\begin{aligned} (\boldsymbol{\Delta}^\star, t^\star) &\in \arg \min_{\boldsymbol{\Delta} \in \mathcal{E}, t} L(\boldsymbol{\Delta}, t; \mathbf{Z}^\star) \\ &= \mathrm{Tr}\left(\mathbf{Z}^\star (\hat{\mathbf{H}} - \boldsymbol{\Delta})^H (\hat{\mathbf{H}} - \boldsymbol{\Delta})\right) \end{aligned} \tag{49}$$

$$\mathrm{Tr}\left(\mathbf{Z}^\star (t^\star \mathbf{I} - (\hat{\mathbf{H}} - \boldsymbol{\Delta}^\star)^H (\hat{\mathbf{H}} - \boldsymbol{\Delta}^\star))\right) = 0 \tag{50}$$

where $L(\boldsymbol{\Delta}, t; \mathbf{Z})$ is the Lagrangian in (15) over the domain $\mathcal{Z}$ defined in (17). On the other hand, the Lagrangian of (20) is given by

$$\begin{aligned} L(\boldsymbol{\Delta}, t; \mathbf{Y}) &= Pt - \mathrm{Tr}\left(\mathbf{Y} \begin{bmatrix} t\mathbf{I} & (\hat{\mathbf{H}} - \boldsymbol{\Delta})^H \\ \hat{\mathbf{H}} - \boldsymbol{\Delta} & \mathbf{I} \end{bmatrix}\right) \\ &= (P - \mathrm{Tr}(\mathbf{Y}_{11}))\, t \\ &\quad - \mathrm{Tr}\left(\mathbf{Y}_{22} + \mathbf{Y}_{12}(\hat{\mathbf{H}} - \boldsymbol{\Delta}) + \mathbf{Y}_{21}(\hat{\mathbf{H}} - \boldsymbol{\Delta})^H\right) \end{aligned} \tag{51}$$

with dual variable $\mathbf{Y}$. Thus, the dual feasible set is $\mathcal{Y} \triangleq \{\mathbf{Y} \succeq \mathbf{0}, \mathrm{Tr}(\mathbf{Y}_{11}) = P\}$. Similarly, the optimal primal and dual variables of (20) are characterized by the following conditions:

$$\boldsymbol{\Delta}^\star \in \mathcal{E}, \quad \mathbf{Y}^\star \succeq \mathbf{0}, \quad \mathrm{Tr}(\mathbf{Y}_{11}^\star) = P \tag{52}$$

$$\begin{aligned} (\boldsymbol{\Delta}^\star, t^\star) &\in \arg \min_{\boldsymbol{\Delta} \in \mathcal{E}, t} L(\boldsymbol{\Delta}, t; \mathbf{Y}^\star) \\ &= -\mathrm{Tr}\left(\mathbf{Y}_{22}^\star + \mathbf{Y}_{12}^\star(\hat{\mathbf{H}} - \boldsymbol{\Delta}) + \mathbf{Y}_{21}^\star(\hat{\mathbf{H}} - \boldsymbol{\Delta})^H\right) \end{aligned} \tag{53}$$

$$\mathrm{Tr}\left(\mathbf{Y}^\star \begin{bmatrix} t^\star\mathbf{I} & (\hat{\mathbf{H}} - \boldsymbol{\Delta}^\star)^H \\ \hat{\mathbf{H}} - \boldsymbol{\Delta}^\star & \mathbf{I} \end{bmatrix}\right) = 0. \tag{54}$$

Note that, for Hermitian matrices $\mathbf{A}, \mathbf{B} \succeq \mathbf{0}$, $\mathrm{Tr}(\mathbf{AB}) = 0$ if and only if $\mathbf{AB} = \mathbf{0}$. Thus, (54) is equivalent to

$$\begin{bmatrix} \mathbf{Y}_{11}^\star & \mathbf{Y}_{12}^\star \\ \mathbf{Y}_{21}^\star & \mathbf{Y}_{22}^\star \end{bmatrix} \begin{bmatrix} t^\star\mathbf{I} & (\hat{\mathbf{H}} - \boldsymbol{\Delta}^\star)^H \\ \hat{\mathbf{H}} - \boldsymbol{\Delta}^\star & \mathbf{I} \end{bmatrix} = \mathbf{0} \tag{55}$$

which implies $\mathbf{Y}_{21}^\star(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)^H+\mathbf{Y}_{22}^\star=\mathbf{0}$ and $\mathbf{Y}_{21}^\star=\mathbf{Y}_{12}^{\star H}=-(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)\mathbf{Y}_{11}^\star$. Hence, (53) reduces to

$$(\boldsymbol{\Delta}^\star,t^\star)\in\arg\min_{\boldsymbol{\Delta}\in\mathcal{E},t}L(\boldsymbol{\Delta},t;\mathbf{Y}^\star)$$
$$=\operatorname{Tr}\left(\mathbf{Y}_{11}^\star(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)^H(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)\right). \quad (56)$$

Moreover, from (55), we obtain $t^\star\mathbf{Y}_{11}^\star=-\mathbf{Y}_{12}^\star(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)$ and $\mathbf{Y}_{12}^\star=-\mathbf{Y}_{11}^\star(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)^H$, which implies $t^\star\mathbf{Y}_{11}^\star=\mathbf{Y}_{11}^\star(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)^H(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)$ or equivalently

$$\operatorname{Tr}\left(\mathbf{Y}_{11}^\star\left(t^\star\mathbf{I}-(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)^H(\hat{\mathbf{H}}-\boldsymbol{\Delta}^\star)\right)\right)=0. \quad (57)$$

Clearly, if $(\boldsymbol{\Delta}^\star,t^\star,\mathbf{Y}^\star)$ satisfies the conditions (52)–(54), then $(\boldsymbol{\Delta}^\star,t^\star,\mathbf{Z}^\star)$ with $\mathbf{Z}^\star=\mathbf{Y}_{11}^\star$ satisfies the conditions (48)–(50). Therefore, the optimal Lagrange multiplier of (14) is given by $\mathbf{Z}^\star=\mathbf{Y}_{11}^\star$.

C. Proof of Theorem 2

By defining $\tilde{\mathbf{Q}}\triangleq\mathbf{V}_h^H\mathbf{Q}\mathbf{V}_h$ and $\tilde{\boldsymbol{\Delta}}\triangleq\mathbf{U}_h^H\boldsymbol{\Delta}\mathbf{V}_h$ and using the unitarily-invariant property of both $\mathcal{Q}$ and $\mathcal{E}$, the maximin problem (7) can be equivalently expressed as

$$\underset{\tilde{\mathbf{Q}}\in\mathcal{Q}}{\text{maximize}}\ \Psi^\star(\tilde{\mathbf{Q}})\triangleq\min_{\tilde{\boldsymbol{\Delta}}\in\mathcal{E}}\operatorname{Tr}\left((\boldsymbol{\Sigma}_h-\tilde{\boldsymbol{\Delta}})\tilde{\mathbf{Q}}(\boldsymbol{\Sigma}_h-\tilde{\boldsymbol{\Delta}})^H\right). \quad (58)$$

We then use the following tool to show that one optimal $\tilde{\mathbf{Q}}$ is a diagonal matrix.

Lemma 2 ([17]): Let $\mathbf{J}\in\mathbb{R}^{N\times N}$ be a diagonal matrix with the diagonal elements being ±1, and $\mathcal{J}$ denote the set of all $L=2^N$ such matrices. Let $\mathbf{A}\in\mathbb{C}^{N\times N}$ be an arbitrary matrix and $\mathbf{D}_\mathbf{A}$ be a diagonal matrix such that $[\mathbf{D}_\mathbf{A}]_{ii}=[\mathbf{A}]_{ii},\forall i$. Then, $\mathbf{D}_\mathbf{A}=\frac{1}{L}\sum_{\mathbf{J}\in\mathcal{J}}\mathbf{J}\mathbf{A}\mathbf{J}$.

Assume $M\geq N$ for the moment, and define $\mathbf{J}_N\in\mathbb{R}^{N\times N}$ as in Lemma 2 and $\mathbf{J}_M\triangleq\operatorname{diag}(\mathbf{J}_N,\mathbf{J}_{M-N})\in\mathbb{R}^{M\times M}$ where $\mathbf{J}_{M-N}\in\mathbb{R}^{(M-N)\times(M-N)}$ is also defined as in Lemma 2. Using the properties that $\mathbf{J}$ is a diagonal matrix, $\mathbf{J}^2=\mathbf{I}$, and $\boldsymbol{\Sigma}_h\mathbf{J}_N=\mathbf{J}_M\boldsymbol{\Sigma}_h$, and defining $\hat{\boldsymbol{\Delta}}\triangleq\mathbf{J}_M\tilde{\boldsymbol{\Delta}}\mathbf{J}_N$, we have

$$\begin{aligned}
&\Psi^\star(\mathbf{J}_N\tilde{\mathbf{Q}}\mathbf{J}_N)\\
&=\min_{\tilde{\boldsymbol{\Delta}}\in\mathcal{E}}\operatorname{Tr}\left((\boldsymbol{\Sigma}_h\mathbf{J}_N-\tilde{\boldsymbol{\Delta}}\mathbf{J}_N)\tilde{\mathbf{Q}}(\boldsymbol{\Sigma}_h\mathbf{J}_N-\tilde{\boldsymbol{\Delta}}\mathbf{J}_N)^H\right)\\
&=\min_{\tilde{\boldsymbol{\Delta}}\in\mathcal{E}}\operatorname{Tr}\left((\mathbf{J}_M\boldsymbol{\Sigma}_h-\tilde{\boldsymbol{\Delta}}\mathbf{J}_N)\tilde{\mathbf{Q}}(\mathbf{J}_M\boldsymbol{\Sigma}_h-\tilde{\boldsymbol{\Delta}}\mathbf{J}_N)^H\right)\\
&=\min_{\tilde{\boldsymbol{\Delta}}\in\mathcal{E}}\operatorname{Tr}\left((\boldsymbol{\Sigma}_h-\mathbf{J}_M\tilde{\boldsymbol{\Delta}}\mathbf{J}_N)\tilde{\mathbf{Q}}(\boldsymbol{\Sigma}_h-\mathbf{J}_M\tilde{\boldsymbol{\Delta}}\mathbf{J}_N)^H\right)\\
&\overset{(a)}{=}\min_{\hat{\boldsymbol{\Delta}}\in\mathcal{E}}\operatorname{Tr}\left((\boldsymbol{\Sigma}_h-\hat{\boldsymbol{\Delta}})\tilde{\mathbf{Q}}(\boldsymbol{\Sigma}_h-\hat{\boldsymbol{\Delta}})^H\right)\\
&=\Psi^\star(\tilde{\mathbf{Q}})
\end{aligned} \quad (59)$$

where (a) follows from the unitary invariance of $\mathcal{E}$. Recalling that $\Psi^\star(\tilde{\mathbf{Q}})$ is a concave function, it follows from Lemma 2 that

$$\begin{aligned}
\Psi^\star(\tilde{\mathbf{Q}})&=\frac{1}{L}\sum_{\mathbf{J}_N\in\mathcal{J}}\Psi^\star(\mathbf{J}_N\tilde{\mathbf{Q}}\mathbf{J}_N)\\
&\leq\Psi^\star\left(\frac{1}{L}\sum_{\mathbf{J}_N\in\mathcal{J}}\mathbf{J}_N\tilde{\mathbf{Q}}\mathbf{J}_N\right)\\
&=\Psi^\star(\mathbf{D}_{\tilde{\mathbf{Q}}})
\end{aligned} \quad (60)$$

where $\mathbf{D}_{\tilde{\mathbf{Q}}}$ is a diagonal matrix such that $[\mathbf{D}_{\tilde{\mathbf{Q}}}]_{ii}=[\tilde{\mathbf{Q}}]_{ii},\forall i$. This means that, given any feasible $\tilde{\mathbf{Q}}$, we can always achieve a larger (or at least equal) objective value by using $\mathbf{D}_{\tilde{\mathbf{Q}}}$, which is feasible too. Therefore, in the solution set there must exist a diagonal structure, which can always be achieved by setting $\mathbf{U}_q=\mathbf{V}_h$, leading to $\tilde{\mathbf{Q}}=\mathbf{D}_{\tilde{\mathbf{Q}}}=\boldsymbol{\Lambda}_q$. It is then straightforward to rewrite (58) into (22). For $M<N$ the proof is similar.

D. Proof of Theorem 3

We first show that $\mathbf{Q}^\star=\mathbf{V}_h\boldsymbol{\Lambda}_q^\star\mathbf{V}_h^H$ is a solution to the maximin problem (7). It follows from Theorem 2 that $\mathbf{U}_q^\star=\mathbf{V}_h$ and (7) is equivalent to

$$\max_{\mathbf{p}\in\mathcal{P}}\min_{\boldsymbol{\Delta}\in\mathcal{E}}\ g(\boldsymbol{\Delta})\triangleq\operatorname{Tr}\left((\boldsymbol{\Sigma}_h-\boldsymbol{\Delta})\boldsymbol{\Lambda}_q(\boldsymbol{\Sigma}_h-\boldsymbol{\Delta})^H\right). \quad (61)$$

Assuming $M\geq N$, we can then express $\boldsymbol{\Delta}=[\boldsymbol{\Delta}_1^T\ \boldsymbol{\Delta}_2^T]^T$ with $\boldsymbol{\Delta}_1\in\mathbb{C}^{N\times N}$ and $\boldsymbol{\Delta}_2\in\mathbb{C}^{(M-N)\times N}$, and $\boldsymbol{\Sigma}_h=[\boldsymbol{\Lambda}_h\ \mathbf{0}]^T$ and $\boldsymbol{\Sigma}_\delta=[\boldsymbol{\Lambda}_\delta\ \mathbf{0}]^T$ with diagonal matrices $\boldsymbol{\Lambda}_h,\boldsymbol{\Lambda}_\delta\in\mathbb{R}^{N\times N}$. Since[4]

$$\begin{aligned}
\boldsymbol{\sigma}(\boldsymbol{\Delta})&=\boldsymbol{\lambda}^{\frac{1}{2}}(\boldsymbol{\Delta}^H\boldsymbol{\Delta})=\boldsymbol{\lambda}^{\frac{1}{2}}(\boldsymbol{\Delta}_1^H\boldsymbol{\Delta}_1+\boldsymbol{\Delta}_2^H\boldsymbol{\Delta}_2)\\
&\geq\boldsymbol{\lambda}^{\frac{1}{2}}(\boldsymbol{\Delta}_1^H\boldsymbol{\Delta}_1)=\boldsymbol{\sigma}(\boldsymbol{\Delta}_1)
\end{aligned} \quad (62)$$

we have $f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}_1))\leq f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}))$, meaning that if $\boldsymbol{\Delta}\in\mathcal{E}$ then $\tilde{\boldsymbol{\Delta}}=[\boldsymbol{\Delta}_1\ \mathbf{0}]\in\mathcal{E}$. Meanwhile, we have

$$\begin{aligned}
g(\boldsymbol{\Delta})&=\operatorname{Tr}\left((\boldsymbol{\Lambda}_h-\boldsymbol{\Delta}_1)\boldsymbol{\Lambda}_q(\boldsymbol{\Lambda}_h-\boldsymbol{\Delta}_1)^H\right)+\operatorname{Tr}\left(\boldsymbol{\Delta}_2\boldsymbol{\Lambda}_q\boldsymbol{\Delta}_2^H\right)\\
&\geq\operatorname{Tr}\left((\boldsymbol{\Lambda}_h-\boldsymbol{\Delta}_1)\boldsymbol{\Lambda}_q(\boldsymbol{\Lambda}_h-\boldsymbol{\Delta}_1)^H\right)=g(\boldsymbol{\Delta}_1).
\end{aligned} \quad (63)$$

Consequently, to minimize $g(\boldsymbol{\Delta})$, we can set $\boldsymbol{\Delta}_2=\mathbf{0}$ and focus on minimizing $g(\boldsymbol{\Delta}_1)$.

Since $g(\boldsymbol{\Delta}_1)$ and $f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}_1))$ are both convex in $\boldsymbol{\Delta}_1$, by using Lemma 2 and following the same reasoning as in Appendix C, one can obtain

$$g(\mathbf{D}_{\boldsymbol{\Delta}_1})\leq g(\boldsymbol{\Delta}_1)\ \text{and}\ f_n(\boldsymbol{\sigma}(\mathbf{D}_{\boldsymbol{\Delta}_1}))\leq f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}_1)) \quad (64)$$

where $\mathbf{D}_{\boldsymbol{\Delta}_1}$ is a diagonal matrix such that $[\mathbf{D}_{\boldsymbol{\Delta}_1}]_{ii}=[\boldsymbol{\Delta}_1]_{ii},\ \forall i$. The equalities in (64) hold when $\boldsymbol{\Delta}_1$ is a diagonal matrix, which can always be achieved by setting $\mathbf{U}_\delta=\mathbf{U}_h$ and $\mathbf{V}_\delta=\mathbf{V}_h$, resulting in $\boldsymbol{\Delta}_1=\mathbf{D}_{\boldsymbol{\Delta}_1}=\boldsymbol{\Lambda}_\delta$. Therefore, the maximin problem (61) reduces to

$$\max_{\mathbf{p}\in\mathcal{P}}\min_{\boldsymbol{\delta}\in\mathcal{D}}\ \operatorname{Tr}\left((\boldsymbol{\Sigma}_h-\boldsymbol{\Sigma}_\delta)\boldsymbol{\Lambda}_q(\boldsymbol{\Sigma}_h-\boldsymbol{\Sigma}_\delta)^H\right)=\sum_{i=1}^N(\gamma_i-\delta_i)^2p_i \quad (65)$$

which is exactly the maximin problem (24). This equivalence can be similarly proved for $M<N$.

Next, we show that $\boldsymbol{\Delta}^\star=\mathbf{U}_h\boldsymbol{\Sigma}_\delta^\star\mathbf{V}_h^H$ is a solution to the minimax problem (8). Consider the convex problem (20), which is equivalent to (8). Assuming $M\geq N$, we can then express the LMI in (20) As

$$\begin{bmatrix}t\mathbf{I}&(\boldsymbol{\Sigma}_h-\boldsymbol{\Delta})^H\\ \boldsymbol{\Sigma}_h-\boldsymbol{\Delta}&\mathbf{I}\end{bmatrix}=\begin{bmatrix}t\mathbf{I}&(\boldsymbol{\Lambda}_h-\boldsymbol{\Delta}_1)^H&-\boldsymbol{\Delta}_2^H\\ \boldsymbol{\Lambda}_h-\boldsymbol{\Delta}_1&\mathbf{I}&\mathbf{0}\\ -\boldsymbol{\Delta}_2&\mathbf{0}&\mathbf{I}\end{bmatrix}\succeq\mathbf{0}. \quad (66)$$

[4]We assume w.l.o.g. that the elements of $\boldsymbol{\sigma}(\boldsymbol{\Delta})$, $\boldsymbol{\lambda}(\boldsymbol{\Delta}^H\boldsymbol{\Delta})$, $\boldsymbol{\lambda}(\boldsymbol{\Delta}_1^H\boldsymbol{\Delta}_1)$, and $\boldsymbol{\sigma}(\boldsymbol{\Delta}_1)$ are in the same order, and thus the elements of $\boldsymbol{\sigma}(\boldsymbol{\Delta}_1)$ share an arbitrary order as those of $\boldsymbol{\sigma}(\boldsymbol{\Delta})$.

Thus, if $\boldsymbol{\Delta}$ satisfies the LMI, so does $\tilde{\boldsymbol{\Delta}} = [\boldsymbol{\Delta}_1\ \mathbf{0}]$. Moreover, we know that $f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}_1)) \leq f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}))$. Therefore, if $\boldsymbol{\Delta}$ is feasible for (20), so is $\tilde{\boldsymbol{\Delta}}$, implying that (20) can be simplified to

$$\begin{aligned} \underset{\boldsymbol{\Delta}_1, t}{\text{minimize}} \quad & Pt \\ \text{subject to} \quad & f_n\left(\boldsymbol{\sigma}(\boldsymbol{\Delta}_1)\right) \leq \varepsilon, \quad \forall n \\ & \begin{bmatrix} t\mathbf{I} & (\boldsymbol{\Lambda}_h - \boldsymbol{\Delta}_1)^H \\ \boldsymbol{\Lambda}_h - \boldsymbol{\Delta}_1 & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{67}$$

Let $\mathbf{J} \in \mathbb{R}^{N\times N}$ be a diagonal matrix defined in Lemma 2. If $\boldsymbol{\Delta}_1$ satisfies the LMI, then we have

$$\begin{aligned} &\begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{bmatrix} \begin{bmatrix} t\mathbf{I} & (\boldsymbol{\Lambda}_h - \boldsymbol{\Delta}_1)^H \\ \boldsymbol{\Lambda}_h - \boldsymbol{\Delta}_1 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{bmatrix} \\ &= \begin{bmatrix} t\mathbf{I} & (\boldsymbol{\Lambda}_h - \mathbf{J}\boldsymbol{\Delta}_1\mathbf{J})^H \\ \boldsymbol{\Lambda}_h - \mathbf{J}\boldsymbol{\Delta}_1\mathbf{J} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \end{aligned} \tag{68}$$

meaning that $\mathbf{J}\boldsymbol{\Delta}_1\mathbf{J}$ also satisfies the LMI. Since the feasible set defined by an LMI is a convex set, the convex combination $\mathbf{D}_{\boldsymbol{\Delta}_1} = \frac{1}{L}\sum_{\mathbf{J}\in\mathcal{J}} \mathbf{J}\boldsymbol{\Delta}_1\mathbf{J}$ is still in this set and thus satisfies the LMI. Moreover, we know that $f_n(\boldsymbol{\sigma}(\mathbf{D}_{\boldsymbol{\Delta}_1})) \leq f_n(\boldsymbol{\sigma}(\boldsymbol{\Delta}_1))$. Consequently, given any feasible $(\boldsymbol{\Delta}_1, t), (\mathbf{D}_{\boldsymbol{\Delta}_1}, t)$ is feasible too and leads to the same objective value. In the solution set, there must exists a diagonal structure of $\boldsymbol{\Delta}_1$, which can always achieved by setting $\mathbf{U}_\delta = \mathbf{U}_h$ and $\mathbf{V}_\delta = \mathbf{V}_h$, resulting in $\boldsymbol{\Delta}_1 = \mathbf{D}_{\boldsymbol{\Delta}_1} = \boldsymbol{\Lambda}_\delta$.

After proving that $\mathbf{U}_\delta^\star = \mathbf{U}_h$ and $\mathbf{V}_\delta^\star = \mathbf{V}_h$, we can rewrite (67) as

$$\begin{aligned} \underset{\boldsymbol{\delta}, t}{\text{minimize}} \quad & Pt \\ \text{subject to} \quad & f_n(\boldsymbol{\delta}) \leq \varepsilon, \quad \forall n \\ & \begin{bmatrix} t\mathbf{I} & \boldsymbol{\Lambda}_h - \boldsymbol{\Lambda}_\delta \\ \boldsymbol{\Lambda}_h - \boldsymbol{\Lambda}_\delta & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{69}$$

By proper row and column permutations, the LMI in (69) can be expressed as

$$\begin{bmatrix} t & \gamma_i - \delta_i \\ \gamma_i - \delta_i & 1 \end{bmatrix} \succeq \mathbf{0}, \quad i = 1, \ldots, N \tag{70}$$

which, by using the Schur complement, are equivalent to $(\gamma_i - \delta_i)^2 \leq t,\ \forall i$. Consequently, (69) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{\delta}}{\text{minimize}} \quad & P \max_i (\gamma_i - \delta_i)^2 \\ \text{subject to} \quad & f_n(\boldsymbol{\delta}) \leq \varepsilon, \quad \forall n \end{aligned} \tag{71}$$

Since $P\max_i(\gamma_i - \delta_i)^2 = \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^N (\gamma_i - \delta_i)^2 p_i$, (71) is equivalent to the minimax problem (8). The proof is similar for $M < N$.

E. Proof of Theorem 4

To rigorously prove the first part of Theorem 4, we first mathematically characterize the waterfilling solution of (27) by showing the following two properties: 1) Assume that there are k eigenmodes active (i.e., $\delta_i > 0$). Then, $\delta_i > 0$ for $i \leq k$ and $\delta_i = 0$ for $i > k$, i.e., the active eigenmodes correspond to the largest k singular values $\gamma_1, \ldots, \gamma_k$. 2) With k active eigenmodes, we have $\gamma_1 - \delta_1 = \cdots = \gamma_k - \delta_k = \mu$, i.e., the minimum of (27) is achieved at the water level μ. Both properties can be proved via contradictions by investigating w.l.o.g. the simple case of two eigenmodes for $\gamma_1 > \gamma_2$.

For property 1), assume that $\delta_1 = 0$ and $\delta_2 > 0$. It follows that $\gamma_1 - \delta_1 = \gamma_1 > \gamma_2 > \gamma_2 - \delta_2$, so the optimal value of (27) is γ_1. However, by the monotonicity and continuity of f, one can always find an $\alpha > 0$ such that $f(0,0) < f(\alpha, 0) \leq f(0, \delta_2) \leq \varepsilon$. Then, using $\delta_1' = \alpha$ and $\delta_2' = 0$, the objective value of (27) becomes $\max\{\gamma_1 - \alpha, \gamma_2\} < \gamma_1$, which causes a contradiction to the optimal value γ_1. Therefore, if there is one active eigenmode, it must be $\delta_1 > 0$ and $\delta_2 = 0$.

For property 2), we assume that there are two active eigenmodes but $\gamma_1 - \delta_1 > \gamma_2 - \delta_2$. Thus, the optimal value of (27) is $\gamma_1 - \delta_1$. Let $\beta > 0$ such that $\gamma_1 - \delta_1 \geq \gamma_2 - \delta_2 + \beta$. Again by the monotonicity and continuity of f, one can always find an $\alpha > 0$ such that $f(\delta_1, \delta_2 - \beta) < f(\delta_1 + \alpha, \delta_2 - \beta + \alpha) \leq f(\delta_1, \delta_2) \leq \varepsilon$. Then, using $\delta_1' = \delta_1 + \alpha$ and $\delta_2' = \delta_2 - \beta + \alpha$, the objective value of (27) becomes $\max\{\gamma_1 - \delta_1 - \alpha, \gamma_2 - \delta_2 + \beta - \alpha\} = \gamma_1 - \delta_1 - \alpha < \gamma_1 - \delta_1$, which causes a contradiction to the optimal value $\gamma_1 - \delta_1$. Similarly, the assumption of $\gamma_1 - \delta_1 < \gamma_2 - \delta_2$ also leads to a contradiction. Therefore, if there are two active eigenmodes, it must be $\gamma_1 - \delta_1 = \gamma_2 - \delta_2$.

Suppose that there are k active eigenmodes. From the above two properties, the optimal solution of (27) is given by $\delta_i^\star = \gamma_i - \mu^\star$ for $i \leq k$ and $\delta_i^\star = 0$ for $i > k$, where the optimal water level satisfies $\gamma_k > \mu^\star \geq \gamma_{k+1}$, so the water volume ε must satisfy (32). This completes the proof of the first part.

Now we prove the second part of Theorem 4. Note that the optimal $t^\star$ in (26) is given by $t^\star = \mu^{\star 2} = (\gamma_i - \delta_i^\star)^2$ for $i = 1, \ldots, k$. According to the complimentary condition of (26), we have

$$\eta_i^\star \left((\gamma_i - \delta_i^\star)^2 - t^\star\right) = 0, \quad \forall i \tag{72}$$

implying that $\eta_i^\star = 0$ for $i > k$. To find $\eta_i^\star$ for $i \leq k$, we write the Lagrangian of (26)

$$L(\boldsymbol{\delta}, t; \boldsymbol{\eta}, \nu) = Pt + \sum_{i=1}^N \eta_i\left((\gamma_i - \delta_i)^2 - t\right) + \nu(f(\boldsymbol{\delta}) - \varepsilon) \tag{73}$$

from which we have the first-order KKT conditions for $i \leq k$:

$$\begin{aligned} \frac{\partial L(\boldsymbol{\delta}^\star, t^\star; \boldsymbol{\eta}^\star, \nu^\star)}{\partial \delta_i} &= -2\eta_i^\star(\gamma_i - \delta_i^\star) + \nu^\star \frac{\partial f(\boldsymbol{\delta}^\star)}{\partial \delta_i} \\ &= -2\eta_i^\star \mu^\star + \nu^\star \alpha_i^\star = 0 \end{aligned} \tag{74}$$

$$\frac{\partial L(\boldsymbol{\delta}^\star, t^\star; \boldsymbol{\eta}^\star, \nu^\star)}{\partial t} = P - \sum_{i=1}^N \eta_i^\star = P - \sum_{i\leq k} \eta_i^\star = 0. \tag{75}$$

Then, one can easily obtain

$$\nu^\star = \frac{2P\mu^\star}{\sum_{j\leq k}\alpha_j^\star} \qquad \text{and} \qquad \eta_i^\star = \frac{P\alpha_i^\star}{\sum_{j\leq k}\alpha_j^\star}. \tag{76}$$

REFERENCES

[1] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2007.

[2] I. E. Telatar, "Capacity of multi-antenna gaussian channels," AT&T Bell Labs, Internal Tech. Memo, Jun. 1995.

[3] 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.

[4] 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 Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

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

[6] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Max-min fairness linear transceiver design for a multi-user MIMO interference channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–5.

[7] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[8] 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 Process.*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[9] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with nonzero mean in space-time coded systems," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2318–2332, Jun. 2006.

[10] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[11] B. K. Chalise, S. Shahbazpanahi, A. Czylwik, and A. B. Gershman, "Robust downlink beamforming based on outage probability specifications," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3498–3503, Oct. 2007.

[12] M. B. Shenouda and T. N. Davidson, "Probabilistically-constrained approaches to the design of the multiple antenna downlink," in *Proc. 42nd Asilomar Conf.*, Pacific Grove, CA, USA, Oct. 2008, pp. 1120–1124.

[13] K.-Y. Wang, T.-H. Chang, W.-K. Ma, A. M.-C. So, and C.-Y. Chi, "Probabilistic SINR constrained robust transmit beamforming: A Bernstein-type inequality based conservative approach," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3080–3083.

[14] C. R. Rojas, D. Katselis, H. Hjalmarsson, R. Hildebrand, and M. Bengtsson, "Chance constrained input design," in *Proc. 50th IEEE Conf. Decision Control Eur. Control*, Orlando, FL, USA, Dec. 2011, pp. 2957–2962.

[15] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[16] 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 Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[17] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[18] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[19] B. C. Levy and R. Nikoukhah, "Robust least-squares estimation with a relative entropy constraint," *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 89–104, Jan. 2004.

[20] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1707–1727, Jul. 2003.

[21] S. Loyka and C. D. Charalambous, "On the compound capacity of a class of MIMO channels subject to normed uncertainty," *IEEE Trans. Inform. Theory*, vol. 58, no. 4, pp. 2048–2063, Apr. 2012.

[22] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known MIMO communications channels," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2918–2930, Aug. 2005.

[23] 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 Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[24] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, "Robust transmit eigen beamforming based on imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1596–1609, May 2006.

[25] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.

[26] J. Wang and M. Payaró, "On the robustness of transmit beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5933–5938, Nov. 2010.

[27] M. B. Shenouda and T. N. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1015–1024, Aug. 2008.

[28] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.

[29] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.

[30] A. Tajer, N. Prasad, and X. Wang, "Robust linear precoder design for multi-cell downlink transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 235–251, Jan. 2011.

[31] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, May 2012.

[32] J. Wang, G. Scutari, and D. P. Palomar, "Robust MIMO cognitive radio via game theory," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1183–1201, Mar. 2011.

[33] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[34] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[35] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, Formulas*, 2nd ed. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[38] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

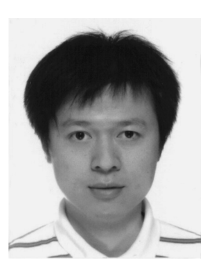
[39] R. T. Rockafellar, *Convex Analysis*, 2nd ed. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[40] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Scientific, 2003.

[41] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, version 1.21, Feb. 2011 [Online]. Available: http://cvxr.com/cvx

[42] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004 [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[43] 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.



Jiaheng Wang (S'08–M'10) received the Ph.D. degree in electrical engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2010. He received the B.E. and M.S. degrees from the Southeast University, Nanjing, China, in 2001 and 2006, respectively.

From 2010 to 2011, he was with the Signal Processing Laboratory, ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently an Associate Professor at the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests mainly include optimization in signal processing, wireless communications and networks.

Mats Bengtsson (M'00–SM'06) received the M.S. degree in computer science from Linköping University, Linköping, Sweden, in 1991 and the Tech. Lic. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1997 and 2000, respectively.

From 1991 to 1995, he was with Ericsson Telecom AB Karlstad. He is currently an Associate Professor at the Signal Processing Laboratory, School of Electrical Engineering, KTH. His research interests include statistical signal processing and its applications to communications, multi-antenna processing, cooperative communication, radio resource management, and propagation channel modelling.

Dr. Bengtsson served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2007 to 2009 and was a member of the IEEE SPCOM Technical Committee from 2007 to 2012.

Björn Ottersten (F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1989.

He has held research positions at the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, and the University of Luxembourg. In 1996 and 1997, he was Director of Research at ArrayComm Inc, a start-up based on his patented technology and located in San Jose, CA, USA. In 1991, he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm. From 1992 to 2004 he was head of the department for Signals, Sensors, and Systems at KTH and from 2004 to 2008 he was Dean of the School of Electrical Engineering at KTH. Currently, Dr. Ottersten is Director for the Interdisciplinary Centre for Security, Reliability and Trust at the University of Luxembourg. As Digital Champion of Luxembourg, he acts as an adviser to European Commissioner Neelie Kroes. His research interests include security and trust, reliable wireless communications, and statistical signal processing.

Dr. Ottersten has served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the editorial board of IEEE Signal Processing Magazine. He is currently editor in chief of the EURASIP Signal Processing Journal and a member of the editorial board of the EURASIP Journal of Applied Signal Processing and Foundations and Trends in Signal Processing. Dr. Ottersten is a Fellow of the IEEE and EURASIP. In 2011 he received the IEEE Signal Processing Society Technical Achievement Award. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, and 2006 and three IEEE conference papers receiving Best Paper Awards. He is a first recipient of the European Research Council advanced research grant.

Daniel P. Palomar (S'99–M'03–SM'08–F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He is an Associate Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He previously held several research appointments at King's College London, London, UK, Technical University of Catalonia (UPC), Barcelona, Spain, Stanford University, Stanford, CA, USA, Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain, Royal Institute of Technology (KTH), Stockholm, Sweden, University of Rome "La Sapienza", Rome, Italy, and Princeton University, Princeton, NJ, USA. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems and communication systems.

Dr. Palmoar serves as an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY, and has been 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," the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on "Game Theory in Communication Systems," and 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). Dr. Palomar is an IEEE Fellow and a Fellow of the Institute for Advance Study (IAS) at HKUST since 2013. He is the recipient of the 2004/2006 Fulbright Research Fellowship, the 2004 Young Author Best Paper Award from the IEEE Signal Processing Society, the 2002/2003 Best Ph.D. Prize in Information Technologies and Communications from the Technical University of Catalonia (UPC), the 2002/2003 Rosina Ribalta First Prize for the Best Doctoral Thesis in Information Technologies and Communications from the Epson Foundation, and the 2004 prize for the Best Doctoral Thesis in Advanced Mobile Communications from the Vodafone Foundation and COIT.