Statistically Robust Design of Linear MIMO Transceivers

Xi Zhang, *Student Member, IEEE*, Daniel P. Palomar, *Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

***Abstract*—The treatment of channel state information (CSI) is critical in the design of MIMO systems. Accurate CSI at the transmitter is often not possible or may require high feedback rates. Herein, we consider the robust design of linear MIMO transceivers with perfect CSI either at the transmitter or at both sides of the link. The framework considers the design problem where the imperfect CSI consists of the channel mean and covariance matrix or, equivalently, the channel estimate and the estimation error covariance matrix. The robust transceiver design is based on a general cost function of the average MSEs as well as a design with individual MSE based constraints. In particular, a lower bound of the average MSE matrix is explored for the design when only the CSI at the transmitter is imperfect. Under different CSI conditions, the proposed robust transceivers exhibit a similar structure to the transceiver designs for perfect CSI, but with a different equivalent channel and/or noise covariance matrix.**

***Index Terms*—Array signal processing, beamforming, joint transmit-receive equalization, linear precoding, multiple-input-multiple-output (MIMO) channels, robustness.**

I. Introduction

COMMUNICATION links with multiple antennas at both the transmitter and the receiver sides, so-called multiple-input-multiple-output (MIMO) systems, can significantly increase the capacity of band-limited wireless channels, provided that the environment has sufficiently rich scattering. In this respect, it is ideal for future high data rate wireless communications [1].

The design of a MIMO communication system depends on the degree of knowledge of the channel state information (CSI). For a given communication channel, the best spectral efficiency and/or performance is obviously achieved when perfect CSI is available at both sides of the link. Optimal linear transceiver design has been extensively studied in this case (e.g., [2] and [3]). In practical communication systems, imperfect CSI may arise from a variety of sources such as channel estimation errors, quantization of the channel estimate in the feedback channel, outdated channel estimates with respect to the current channel (for time-varying channels), etc. [4]. This effect is of paramount importance in practical implementations. By modeling such imperfections and taking them into account in the transceiver design, a robust high performance link can be achieved.

CSI at the receiver (CSIR) is traditionally acquired via a training sequence that allows the estimation of the channel, or via blind methods that exploit the structure of the transmitted signal or of the channel. CSI at the transmitter (CSIT) can be obtained either by means of feedback from the receiver, or from previous receive measurements if the channel has some reciprocity [4]. Whereas sufficiently accurate CSIR can be assumed in certain cases, CSIT will be far from perfect in most realistic situations. In other cases, both CSIT and CSIR cannot be modeled as perfect. Hence, it is reasonable to assume two different CSI assumptions for design purposes: i) imperfect CSIR with imperfect CSIT, and ii) perfect CSIR with imperfect CSIT.

There are different ways to design a system that is robust to imperfect CSIT. In [4]–[6], worst-case designs are considered. This guarantees a certain system performance for any possible channel sufficiently close to the estimated one. This approach leads to conservative designs, which may translate into a significant increase of the required transmit power. Alternatively, the CSI uncertainty can be modeled statistically. This guarantees a certain system performance averaged over the channel realizations [7], [8]. The latter statistical modeling approach is used in the sequel.

For the statistical or stochastic approach, different types and amount of CSIR and CSIT determine different transceiver designs. Previous work has considered channel mean CSIT or channel covariance CSIT and perfect CSIR. The case of mean CSIT is addressed in [9] for minimizing the average mean square error (MSE), and in [10] [multiple-input-single-output, (MISO)] and [11] (MIMO with ML receiver) for maximizing the mutual information by beamforming. The case when the channel covariance is the only CSIT and the channel mean is assumed to be zero is addressed in [12] (ML receiver) for minimizing an upper bound of the average pairwise error probability (PEP) by eigenbeamforming, and in [10] (MISO) and [11] (MIMO with ML receiver) for maximizing the mutual information by beamforming. When both mean and covariance CSIT are available, a robust design is more involved. The problem of minimizing the average MSE and maximizing the average signal-to-noise ratio (SNR) in MISO channels is considered in [7], and in [8] for minimizing the sum MSE with an equivalent channel based on conditional channel mean and linear transceivers. In [13], the Chernoff bound of the PEP is minimized for a linear transmitter and an ML receiver with space-time block codes.

Herein, we reflect on the robust MIMO linear transceiver design for the cases of i) imperfect CSIR with imperfect CSIT and

Manuscript received July 20, 2006; revised December 18, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Timothy N. Davidson.

X. Zhang and B. Ottersten are with the Signal Processing Lab, ACCESS Linnaeus Center, Royal Institute of Technology (KTH), Stockholm, SE 10044 Sweden (e-mail: xi.zhang@ee.kth.se; bjorn.ottersten@ee.kth.se).

D. P. Palomar is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: danielp@princeton.edu).

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.2008.919384

ii) perfect CSIR with imperfect CSIT. The framework considers the cases of mean and/or covariance feedback CSIT. The design is based on a general cost function of the average MSEs as well as a design with individual constraints based on the MSEs. Under different CSI conditions, the proposed robust transceivers exhibit a very similar structure to the transceiver designs for perfect CSI (with a different equivalent channel matrix and/or noise covariance matrix) and can be unified into the same framework.

The paper is structured as follows. In Section II, a brief description of the concepts of majorization and matrix variate Gaussian distribution is given. Section III describes the MIMO system model, along with a mean/covariance modeling of the CSI. A short review of the transceiver design problem of perfect CSI is included in Section IV. The formulation of the robust transceiver design problem is given in Section V. The robust design with imperfect CSIR and imperfect CSIT is treated in Section VI, whereas the robust design with perfect CSIR and imperfect CSIT is considered in Section VII. The MIMO linear transceiver designs under different CSI conditions are summarized in Section VIII by a unified framework. Numerical examples are included in Section IX to illustrate our theoretical development. Conclusions are drawn in Section X.

II. Preliminaries

The following notation is used in the paper. Uppercase and lowercase boldface denote matrices and vectors respectively. The operators $(\cdot)^*$, $\mathrm{vec}(\cdot)$ and $\mathrm{tr}[\cdot]$ are Hermitian transpose, stacking vectorization operator and trace operator, respectively. The operators $\otimes$ and $\prec$ are Kronecker product and the majorization relation [14], respectively. The operators $\leq$ and $\geq$ for vectors are defined element-wise. The operator $\mathbf{d}[\cdot]$ is the vector consisting of the diagonal elements of the matrix argument, while $\mathrm{diag}[\cdot]$ is the diagonal matrix with the vector argument as diagonal elements, and $[\cdot]_{ij}$ is the (i,j) element of the matrix argument. The set $\mathbb{H}^{n\times n}$ stands for the set of $n\times n$ Hermitian matrices.

The framework developed in the paper for the linear MIMO transceiver design problem is based on majorization theory. For this purpose, we first review some important definitions and properties of Schur-convexity and the matrix variate Gaussian distribution for easy reference.

A. Schur-Convexity

Two special classes of cost functions will be discussed in the sequel: Schur-convex and Schur-concave functions. These functions are defined by the majorization relation, which makes precise the vague notion that components of a vector $\mathbf{x}$ are "less spread out" or "more equal" than the components of another vector $\mathbf{y}$. (see [14] for a complete reference of the subject).

Definition 1: [14, Ch. 1] For any $\mathbf{x}\in\mathbb{R}^{n\times 1}$, let

$$x_{[1]}\geq x_{[2]}\geq\cdots\geq x_{[n]}$$

denote the components of $\mathbf{x}$ in decreasing order.

Definition 2: [14, Ch. 1], Let $\mathbf{x},\mathbf{y}\in\mathbb{R}^{n\times 1}$, the vector $\mathbf{x}$ is majorized by the vector $\mathbf{y}$ if

$$\sum_{i=1}^{k}x_{[i]}\leq\sum_{i=1}^{k}y_{[i]},\quad 1\leq k\leq n-1$$
$$\sum_{i=1}^{n}x_{[i]}=\sum_{i=1}^{n}y_{[i]}$$

and is denoted as $\mathbf{x}\prec\mathbf{y}$.

Definition 3: [14, Ch. 3] A real-valued function g defined on a set $\mathcal{A}\subseteq\mathbb{R}^{n\times 1}$ is said to be Schur-convex on $\mathcal{A}$ if

$$\mathbf{x}\prec\mathbf{y}\quad\text{on}\quad\mathcal{A}\quad\Rightarrow\quad g(\mathbf{x})\leq g(\mathbf{y}).$$

Similarly, function g is said to be Schur-concave on $\mathcal{A}$ if

$$\mathbf{x}\prec\mathbf{y}\quad\text{on}\quad\mathcal{A}\quad\Rightarrow\quad g(\mathbf{x})\geq g(\mathbf{y}).$$

Schur-convex or Schur-concave functions cover many interesting cost functions that are important in wireless communication systems. When there exist multiple data streams, a simple function is usually used to map certain merit functions of these data streams to a single global merit function, which can be easily analyzed and optimized. Here is a brief list of some of these functions that fall into the class of Schur-convex or Schur-concave functions. Their Schur-convexity is defined with regards to the MSEs (cf. [3] for the detailed proof of their Schur-convexity).

Corresponding to Schur-concave functions of the MSEs:
- minimizing the (weighted) sum of the MSEs;
- minimizing the product of the MSEs;
- minimizing the determinant of the MSE matrix;
- maximizing the (weighted) sum of the SNRs;
- maximizing the product of the SNRs.

Corresponding to Schur-convex functions of the MSEs:
- maximizing the harmonic mean of the SNRs;
- minimizing the average BER with equal constellations;
- minimizing the maximum of the MSEs.

B. Matrix Variate Gaussian Distribution

Definition 4: [15] A random matrix $\mathbf{X}\in\mathbb{C}^{m\times n}$ is said to have a matrix variate complex Gaussian distribution with mean $\bar{\mathbf{X}}\in\mathbb{C}^{m\times n}$ and covariance matrix $\mathbf{\Sigma}\otimes\mathbf{\Psi}$ (where $\mathbf{\Sigma}\in\mathbb{C}^{n\times n}$ and $\mathbf{\Psi}\in\mathbb{C}^{m\times m}$ are both positive definite), denoted as $\mathbf{X}\sim\mathcal{CN}(\bar{\mathbf{X}},\mathbf{\Sigma}\otimes\mathbf{\Psi})$, if

$$\mathrm{vec}(\mathbf{X})\sim\mathcal{CN}\left(\mathrm{vec}(\bar{\mathbf{X}}),\mathbf{\Sigma}\otimes\mathbf{\Psi}\right).$$

Lemma 1: [15] Let $\mathbf{X}\sim\mathcal{CN}(\bar{\mathbf{X}},\mathbf{\Sigma}\otimes\mathbf{\Psi})$, then

$$\mathbb{E}\{\mathbf{X}^*\mathbf{M}\mathbf{X}\}=\bar{\mathbf{X}}^*\mathbf{M}\bar{\mathbf{X}}+\mathrm{tr}[\mathbf{\Psi}\mathbf{M}]\mathbf{\Sigma}^T$$
$$\mathbb{E}\{\mathbf{X}\mathbf{M}\mathbf{X}^*\}=\bar{\mathbf{X}}\mathbf{M}\bar{\mathbf{X}}^*+\mathrm{tr}[\mathbf{M}\mathbf{\Sigma}^T]\mathbf{\Psi}.$$

III. System Model

A. Signal Model

Consider a narrow-band MIMO channel with n transmit and m receive antennas, as shown in Fig. 1. Omitting the time index

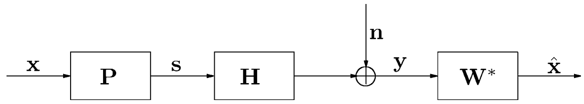

Fig. 1. Scheme of MIMO system model with linear transceiver.

for simplicity, the corresponding discrete-time signal model can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

where the transmitted signal vector is $\mathbf{s} \in \mathbb{C}^{n\times 1}$, the channel matrix is $\mathbf{H} \in \mathbb{C}^{m\times n}$, the received signal vector is $\mathbf{y} \in \mathbb{C}^{m\times 1}$, and the noise vector $\mathbf{n} \in \mathbb{C}^{m\times 1}$ is assumed to be zero-mean circular symmetric complex Gaussian noise with arbitrary covariance matrix, i.e., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_n)$.

We consider the simultaneous transmission of ℓ symbols over the MIMO channel. The design quantities to be optimized are a linear transmitter $\mathbf{P} \in \mathbb{C}^{n\times \ell}$ $(\ell \leq n)$ and a linear receiver $\mathbf{W} \in \mathbb{C}^{m\times \ell}$

$$\mathbf{s} = \mathbf{P}\mathbf{x}$$
$$\hat{\mathbf{x}} = \mathbf{W}^*\mathbf{y}$$

where the data symbol vector, $\mathbf{x} \in \mathbb{C}^{\ell\times 1}$, is zero-mean with unit-energy uncorrelated symbols, i.e., $\mathbb{E}\{\mathbf{x}\mathbf{x}^*\} = \mathbf{I}_\ell$. The total average transmitted power P is defined as

$$P = \mathbb{E}\left\{\|\mathbf{s}\|^2\right\} = \mathrm{tr}[\mathbf{P}\mathbf{P}^*].$$

The SNR is defined as

$$\gamma = \frac{\mathbb{E}\left\{\|\mathbf{s}\|^2\right\}}{\mathbb{E}\left\{\|\mathbf{n}\|^2\right\}} = \frac{P}{\mathrm{tr}[\mathbf{R}_n]}. \tag{1}$$

Although this paper considers arbitrary functions as a measure of the system performance, it will be useful to define the following MSE matrix as a convenient quantity:

$$\begin{aligned}\mathbf{E}(\mathbf{W},\mathbf{P}) &\triangleq \mathbb{E}\left\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^*\right\}\\ &= (\mathbf{W}^*\mathbf{H}\mathbf{P}-\mathbf{I}_\ell)(\mathbf{P}^*\mathbf{H}^*\mathbf{W}-\mathbf{I}_\ell)+\mathbf{W}^*\mathbf{R}_n\mathbf{W}.\end{aligned} \tag{2}$$

Observe that the diagonal elements of $\mathbf{E}$, i.e., $\mathbf{d}[\mathbf{E}]$, contain the MSEs of the established data streams.

B. Modeling of CSI

Because it is either too difficult or too expensive to have perfect CSI at both the transmitter and the receiver side, CSI is usually not an accurate instantaneous channel information. In this work, the channel is described statistically, so we assume that the imperfect CSI consists of the first and second order statistics of the actual channel, i.e., the channel is modeled in the form of a nonzero channel mean $\bar{\mathbf{H}}$ and a channel covariance matrix $\mathbf{R}_H$, or equivalently as a channel estimate and its estimation error covariance. The transceiver $(\mathbf{W}, \mathbf{P})$ optimized for the available CSI will remain constant for a period of time until the CSIR or CSIT is updated.

To model the MIMO channel matrix distribution, we consider the so-called Kronecker model

$$\mathbf{H} = \bar{\mathbf{H}} + \left(\mathbf{R}_H^{Rx}\right)^{1/2}\mathbf{G}\left(\mathbf{R}_H^{Tx}\right)^{T/2} \tag{3}$$

where $\mathbf{G} \in \mathbb{C}^{m\times n}$ has i.i.d. elements distributed as $\mathcal{CN}(0,1)$, and it is the unknown part in the fading estimate. $\mathbf{R}_H^{Rx}$ and $\mathbf{R}_H^{Tx}$ denote the covariance matrices seen from the receiver and transmitter, respectively (cf. [16] and [17]), and that implies the channel covariance matrix $\mathbf{R}_H = \mathbf{R}_H^{Tx} \otimes \mathbf{R}_H^{Rx}$. The resulting distribution of $\mathbf{H}$ is

$$\mathbf{H} \sim \mathcal{CN}\left(\bar{\mathbf{H}}, \mathbf{R}_H^{Tx} \otimes \mathbf{R}_H^{Rx}\right). \tag{4}$$

This statistical model will be simplified in Section VI in order to make the analysis feasible.

IV. Transceiver Design With Perfect CSI: A Review

In a scalar channel, the performance is typically measured in terms of SNR, BER, or MSE. Interestingly, as shown in [3], BER and SNR can both be remapped to functions of MSE. Therefore, it suffices to consider functions of the MSEs of the established data streams (i.e., $\mathbf{d}[\mathbf{E}]$) as a performance measure in a MIMO system. Let $\mathcal{F}_0 : \mathbb{R}^\ell \mapsto \mathbb{R}$ be a cost function increasing in each argument without loss of generality.[1] With perfect CSI, the system optimization problem can be formulated as

$$\begin{aligned}&\underset{\mathbf{W},\mathbf{P}}{\text{minimize}} && \mathcal{F}_0\left(\mathbf{d}[\mathbf{E}]\right)\\ &\text{subject to} && \mathrm{tr}[\mathbf{P}\mathbf{P}^*] \leq P_{\max}\end{aligned} \tag{5}$$

where $P_{\max}$ denotes the maximum power budget at the transmitter. This problem is solved in [2] and [3] (and references therein) for several specific cost functions.

A. Receiver Design

Because the receiver has perfect knowledge of the channel state, it can always optimize $\mathbf{W}$ for each channel realization $\mathbf{H}$. The linear receiver $\mathbf{W}$ that minimizes the MSEs ($\mathbf{d}[\mathbf{E}]$, with $\mathbf{E}$ given in (2)), and therefore the cost function $\mathbf{F}_0$ (as it is increasing in each argument) is the well known Wiener filter

$$\mathbf{W} = (\mathbf{H}\mathbf{P}\mathbf{P}^*\mathbf{H}^* + \mathbf{R}_n)^{-1}\mathbf{H}\mathbf{P} \tag{6}$$

and the resulting MSE matrix is

$$\mathbf{E}(\mathbf{P}) = (\mathbf{I}_\ell + \mathbf{P}^*\mathbf{\Psi}_H\mathbf{P})^{-1} \tag{7}$$

where $\mathbf{\Psi}_H$ is the squared channel matrix defined as

$$\mathbf{\Psi}_H = \mathbf{H}^*\mathbf{R}_n^{-1}\mathbf{H}. \tag{8}$$

B. Transmitter Design

The transmitter design is studied in detail in [3] when both CSIR and CSIT are perfect. In particular, the main results are summarized here for better understanding of the proposed robust design in the following sections.

The nonconvex constrained optimization problem (5), with the receiver $\mathbf{W}$ given in (6), is solved when the transmitter $\mathbf{P}$ has the structure of

$$\mathbf{P} = \mathbf{U}_{\Psi,\ell}\mathbf{\Sigma}_P\mathbf{Q} \tag{9}$$

[1]Note that by increasing we do not mean strictly increasing but just nondecreasing. Also, it would be meaningless to have a function not increasing in the arguments as smaller MSEs are always preferred (cf. [3]).

where the matrix $\mathbf{U}_{\Psi,\ell}$ consists of the ℓ eigenvectors of $\mathbf{\Psi}_H$ corresponding to the ℓ largest eigenvalues in increasing order ($\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\ell$), the diagonal matrix $\mathbf{\Sigma}_P = \mathrm{diag}[\{\sqrt{p_i}\}]$ contains the power allocation $\{p_i\}$ in the diagonal elements, and $\mathbf{Q} \in \mathbb{C}^{\ell \times \ell}$ is a unitary matrix. With the transmitter $\mathbf{P}$ given in (9), the MSE matrix (7) becomes

$$\mathbf{E} = \mathbf{Q}^* \mathrm{diag}\left[\left\{\frac{1}{1+p_i\lambda_i}\right\}\right]\mathbf{Q}. \tag{10}$$

What remains is to choose a proper unitary matrix $\mathbf{Q}$ and the power allocation $\{p_i\}$, both of which depend on the specific function $\mathcal{F}_0$.

The choice of $\mathbf{Q}$ and $\{p_i\}$ can be further simplified in two special cases, when $\mathcal{F}_0$ falls into the class of Schur-concave or Schur-convex functions on $\mathbb{R}^\ell$.[2]

- If $\mathcal{F}_0$ is Schur-concave, then $\mathbf{Q} = \mathbf{I}_\ell$. The optimal transmitter $\mathbf{P}$ actually diagonalizes the equivalent channel covariance matrix $\mathbf{\Psi}_H$. The whole optimization problem is reduced to a power allocation problem, which depends on the specific cost function.
- If $\mathcal{F}_0$ is Schur-convex, then $\mathbf{Q}$ is a unitary matrix such that $\mathbf{E}$ in (10) has identical diagonal elements, e.g., the Hadamard matrix or the unitary discrete Fourier transform (DFT) matrix (cf. [3]). Surprisingly the optimal transmitter $\mathbf{P}$ is independent of the choice of the function $\mathcal{F}_0$. The optimal $\mathbf{P}$ does not fully diagonalize $\mathbf{\Psi}_H$; instead, it produces equal diagonal elements of $\mathbf{E}$. The optimization of the power allocation is equivalent to the minimization of $\mathrm{tr}[\mathbf{E}]$ and can be solved very efficiently.

Although the design of the receiver $\mathbf{W}$ depends on the transmitter $\mathbf{P}$, it is not necessary for the transmitter to feed forward the $\mathbf{P}$ matrix to the receiver because the receiver also has perfect CSI and can calculate $\mathbf{P}$ locally.

C. MSE-Constrained Design

In addition to the design problem (5), which minimizes a cost function of the MSEs with a total transmit power constraint, it is also possible to consider the dual problem that minimizes the total transmit power with a global performance constraint. More general is the design with *individual* MSE constraints:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} \quad & \mathrm{tr}[\mathbf{PP}^*] \\ \text{subject to} \quad & \mathbf{d}[\mathbf{E}] \leq \boldsymbol{\rho} \end{aligned} \tag{11}$$

where $\boldsymbol{\rho}$ is a vector containing the desired MSEs targets for each of the data streams. In this case, the optimal transceiver $(\mathbf{W},\mathbf{P})$ is still given by (6) and (9), but the computation of the unitary matrix $\mathbf{Q}$ is then more involved and depends on the channel realization (cf. [18]).

V. Robust Problem Formulation

In reality, the CSIT and/or the CSIR are prone to perturbations due to practical constraints in wireless systems. Our goal is to design linear transceivers with robustness to such imperfect CSI conditions. The stochastic robust design method will be used in the sequel, following a Bayesian philosophy based on the knowledge of the channel mean and covariance matrix. In contrast to the perfect CSI case [cf. (5)] where an arbitrary increasing function of the *instantaneous* MSEs is considered, the average system performance should be considered. One possible way to formulate the problem is to consider instead an arbitrary increasing function of the *average* MSEs of the data streams, leading to the problem formulation:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} \quad & \mathcal{F}_0\left(\mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right]\right) \\ \text{subject to} \quad & \mathrm{tr}[\mathbf{PP}^*] \leq P_{\max}. \end{aligned} \tag{12}$$

In fact, (12) considers the so-called certainty-equivalent system [19], in which the optimization is only based on the expected value of MSEs. Observe that a statistical robust design does not guarantee any performance target for a particular realization of the random channel, because the statistical variation of the MSEs are ignored in a certainty-equivalent system. If one has to guarantee certain performance or outage for *all* the channel realizations, worst case designs or certainty-equivalent margins [19] should be used.

Another possible way to formulate the robust design problem is to consider the average value of an increasing cost function of the instantaneous MSEs, leading to a different problem formulation:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} \quad & \mathbb{E}\left\{\mathcal{F}_0\left(\mathbf{d}[\mathbf{E}]\right)\right\} \\ \text{subject to} \quad & \mathrm{tr}[\mathbf{PP}^*] \leq P_{\max}. \end{aligned} \tag{13}$$

In this paper, we will focus only on problem (12). Problem (13) is more difficult to deal with because it is generally impossible to find a closed-form expression of $\mathbb{E}\{\mathcal{F}_0\}$ for an arbitrary function. However, the two problems are equivalent when the cost function is linear. When the cost function is nonlinear but differentiable, the two problems can be related via a Taylor expansion, which will be discussed in Section IX.

The MSE-constrained problem (11), in the robust design context, becomes

$$\begin{aligned} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} \quad & \mathrm{tr}[\mathbf{PP}^*] \\ \text{subject to} \quad & \mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right] \leq \boldsymbol{\rho}. \end{aligned} \tag{14}$$

This problem will not be explicitly considered in this paper, but it can be similarly dealt with by combining the approach in this paper with the result in [18].

Two different cases of the robust transceiver design for problem (12) will be addressed in the following sections: i) design with both imperfect CSIR and imperfect CSIT in Section VI, and ii) design with perfect CSIR but imperfect CSIT in Section VII. In both sections, problem (12) is solved in two steps: first a receiver $\mathbf{W}$ is optimized according to different CSIR conditions; then the joint transceiver design is reduced to the optimization of the transmitter $\mathbf{P}$ only. Due to the difference in CSI, the resulting optimal transceivers are different in nature.

VI. Robust Design With Imperfect CSIR and Imperfect CSIT

In this section, the situation when perfect CSI is available neither at transmitter nor at the receiver will be considered. The

[2]Furthermore, the results in the sequel follow verbatim for functions that may not be Schur-concave/convex on $\mathbb{R}^\ell$ but are minimized when the arguments are sorted in decreasing order (or any fixed ordering) and then they become Schur-concave/convex on $\{\mathbf{x} \in \mathbb{R}^\ell | x_1 \geq x_2 \geq \cdots \geq x_\ell\}$.

channel $\mathbf{H}$ is Gaussian distributed as in (4) with CSIR and CSIT in the form of $\bar{\mathbf{H}}$ and $\mathbf{R}_H = \mathbf{R}_H^{Tx} \otimes \mathbf{R}_H^{Rx}$.

Since the CSIR is not perfect, it is not possible to optimize the linear transceiver $(\mathbf{W}, \mathbf{P})$ for each instantaneous channel state $\mathbf{H}$ as in the perfect CSI case in Section IV. Consequently, a fixed transceiver $(\mathbf{W}, \mathbf{P})$ must be calculated for a specific set of $\bar{\mathbf{H}}$ and $\mathbf{R}_H$.

To solve the problem (12), we first give a closed-form expression for the average MSE matrix, $\mathbb{E}\{\mathbf{E}(\mathbf{W}, \mathbf{P})\}$, by invoking Lemma 1. For a given transceiver $(\mathbf{W}, \mathbf{P})$, the MSE matrix (2) averaged over $\mathbf{x}, \mathbf{n}$ and $\mathbf{H}$, is

$$\mathbb{E}\{\mathbf{E}(\mathbf{W}, \mathbf{P})\} = (\mathbf{W}^*\bar{\mathbf{H}}\mathbf{P} - \mathbf{I}_\ell)(\mathbf{P}^*\bar{\mathbf{H}}^*\mathbf{W} - \mathbf{I}_\ell) + \mathbf{W}^*\mathbf{R}_n'\mathbf{W} \tag{15}$$

where the equivalent noise covariance matrix is

$$\mathbf{R}_n' = \mathbf{R}_n + \operatorname{tr}\left[\mathbf{P}\mathbf{P}^*\left(\mathbf{R}_H^{Tx}\right)^T\right]\mathbf{R}_H^{Rx}. \tag{16}$$

A. Receiver Design

The average MSE matrix given in (15) is exactly the same as (2), but with $\bar{\mathbf{H}}$ instead of $\mathbf{H}$ and $\mathbf{R}_n'$ instead of $\mathbf{R}_n$. It follows immediately that the optimal receiver $\mathbf{W}$, which minimizes the diagonal elements of the above average MSE matrix, is still given by (6) but with $\mathbf{H}$ replaced by $\bar{\mathbf{H}}$ and $\mathbf{R}_n$ by $\mathbf{R}_n'$. The average MSE matrix becomes

$$\mathbb{E}\{\mathbf{E}(\mathbf{P})\} = \left(\mathbf{I}_\ell + \mathbf{P}^*\bar{\mathbf{H}}^*\mathbf{R}_n'^{-1}\bar{\mathbf{H}}\mathbf{P}\right)^{-1}. \tag{17}$$

Since $\mathbf{R}_n'$ in (17) depends on $\mathbf{P}$, it follows that $\lim_{\gamma\to\infty} \mathbb{E}\{\mathbf{E}\} \neq \mathbf{0}$, and an MSE floor exists (proof in the next footnote). Such MSE floor can be clearly observed in the numerical examples later in Section IX.

B. Transmitter Design

The optimal design of the transmitter $\mathbf{P}$ is more involved since $\mathbf{P}$ also appears in the equivalent noise covariance $\mathbf{R}_n'$. In order to be able to carry on the analytical computation, we assume that $\mathbf{R}_H^{Tx} = \mathbf{I}_n$. It can be shown then that the optimal solution is always achieved with equality in the power constraint,[3] so that $\mathbf{R}_n'$ becomes

$$\mathbf{R}_n'' = \mathbf{R}_n + P_{\max}\mathbf{R}_H^{Rx}. \tag{18}$$

Now $\mathbf{P}$ no longer appears in the equivalent noise covariance, and the resulting averaged MSE matrix is the same as (7), but with an equivalent covariance matrix $\mathbf{\Psi}_H''$ instead of $\mathbf{\Psi}_H$, which is defined as

$$\mathbf{\Psi}_H'' = \bar{\mathbf{H}}^*\mathbf{R}_n''^{-1}\bar{\mathbf{H}}. \tag{19}$$

So the optimal $\mathbf{P}$ is also given by (9), but with $\mathbf{\Psi}_H''$ instead of $\mathbf{\Psi}_H$. What remains is to choose a proper unitary matrix $\mathbf{Q}$ and the power allocation $\{p_i\}$, which both depend on the specific function $calF_0$.

The robust design can be summarized in the following theorem.

Theorem 1: When both the CSIR and the CSIT are imperfect and in the form of a channel mean $\bar{\mathbf{H}}$ and a channel covariance $\mathbf{R}_H = \mathbf{I}_n \otimes \mathbf{R}_H^{Rx}$, the optimal solution $(\mathbf{W}, \mathbf{P})$ to the robust design problem (12) has the same structure as the perfect CSI one given in (6) and (9), except for replacing the perfect channel $\mathbf{H}$ with $\bar{\mathbf{H}}$, the noise covariance matrix $\mathbf{R}_n$ with the equivalent noise covariance $\mathbf{R}_n''$ in (18), and $\mathbf{\Psi}_H$ with the equivalent covariance matrix $\mathbf{\Psi}_H''$ in (19).

Similar to the case of perfect CSI, the robust design of the receiver $\mathbf{W}$ depends on the transmitter $\mathbf{P}$, but it is not necessary for the transmitter to feed forward the $\mathbf{P}$ matrix to the receiver. The receiver also has the same CSI as the transmitter so it can calculate $\mathbf{P}$ locally.

The MSE-constrained problem (14) is more difficult to solve than (12) in this case, because $\mathbf{R}_n''$ contains $\operatorname{tr}[\mathbf{P}\mathbf{P}^*]$, which is precisely the objective function and cannot be replaced by $P_{\max}$.

Remark: We have discussed the problem when both the transmitter and the receiver have the same imperfect CSI. When CSIR and CSIT are both imperfect but with different imperfectness, i.e., $\bar{\mathbf{H}}$ and/or $\mathbf{R}_H$ are different at the transmitter and receiver, the robust design problem is still an open problem.

VII. Robust Transceiver Design With Perfect CSIR and Imperfect CSIT

It is quite often the case that the channel estimate at the receiver is accurate enough to be modeled as perfect CSIR, but the CSIT is still modeled as imperfect due to feedback-related issues (e.g., feedback errors, quantization, and delay). The robust design problem (12) under such CSI assumptions will be considered here. The channel $\mathbf{H}$ is Gaussian distributed as in (4). The CSIR is perfect in the form of $\mathbf{H}$, while the CSIT is imperfect in the form of $\bar{\mathbf{H}}$ and $\mathbf{R}_H = \mathbf{R}_H^{Tx} \otimes \mathbf{R}_H^{Rx}$.

A. Receiver Design

Exactly as in the design for perfect CSI case, the receiver $\mathbf{W}$ can be designed based on any instantaneous $\mathbf{H}$, which means that the optimal receiver $\mathbf{W}$ is exactly (6). The joint transceiver design problem is, therefore, reduced to the optimization of the linear transmitter $\mathbf{P}$ allowing robust performance with respect to channel imperfectness in the CSIT.

B. Transmitter Design

When $\mathbf{W}$ is given by (6), the instantaneous MSE is given by (7). To simplify the optimization problem (12), a closed-form expression for the averaged MSE matrix $\mathbb{E}\{\mathbf{E}\}$ would be useful just as in Section VI; however, this is very difficult in general. In the following, a robust transmitter design will be proposed based on a tight lower bound of $\mathbb{E}\{\mathbf{E}\}$.

Lemma 2: Given the optimal $\boldsymbol{W}$ in (6) and instantaneous MSE matrix $\mathbf{E}$ defined in (7), the average MSE matrix is lower bounded by $\mathbf{E}'$ as

$$\mathbb{E}\{\mathbf{E}\} \geq \mathbf{E}' = \left(\mathbf{I}_\ell + \mathbf{P}^*\mathbf{\Psi}_H'\mathbf{P}\right)^{-1} \tag{20}$$

[3]Simply rewrite the MSE matrix (17) as $\mathbb{E}\{\mathbf{E}\} = (\mathbf{I}_\ell + \tilde{\mathbf{P}}^*\bar{\mathbf{H}}^*(\mathbf{R}_H^{Rx} + \mathbf{R}_n/\operatorname{tr}[\mathbf{P}\mathbf{P}^*])^{-1}\bar{\mathbf{H}}\tilde{\mathbf{P}})^{-1}$, where $\tilde{\mathbf{P}} = \mathbf{P}/\sqrt{\operatorname{tr}[\mathbf{P}\mathbf{P}^*]}$, and note that for any given $\tilde{\mathbf{P}}$, $\mathbb{E}\{\mathbf{E}\}$ is decreasing in $\operatorname{tr}[\mathbf{P}\mathbf{P}^*]$. Also, when $\gamma = \operatorname{tr}[\mathbf{P}\mathbf{P}^*]/\operatorname{tr}[\mathbf{R}_n] \to \infty$, $\lim_{\gamma\to\infty} \mathbb{E}\{\mathbf{E}\} = (\mathbf{I}_\ell + \tilde{\mathbf{P}}^*\bar{\mathbf{H}}^*(\mathbf{R}_H^{Rx})^{-1}\bar{\mathbf{H}}\tilde{\mathbf{P}})^{-1} \neq \mathbf{0}$

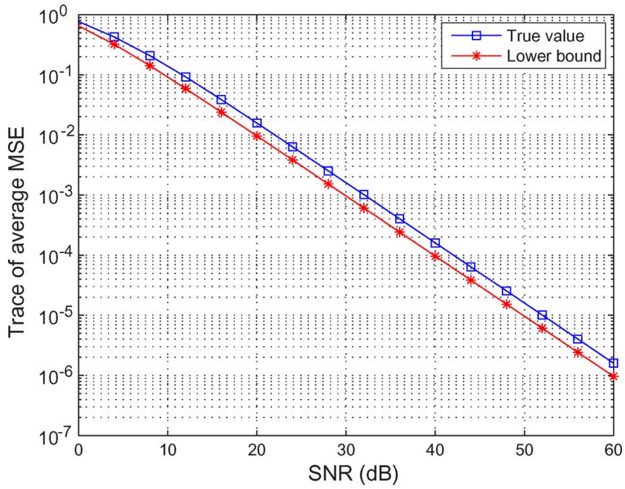

Fig. 2. Trace of average MSE and its lower bound versus SNR ($m = 4, n = 4, \ell = 2$).

where the equivalent channel covariance matrix is

$$\boldsymbol{\Psi}'_H = \bar{\mathbf{H}}^* \mathbf{R}_n^{-1} \bar{\mathbf{H}} + \operatorname{tr}\left[\mathbf{R}_H^{Rx} \mathbf{R}_n^{-1}\right] \left(\mathbf{R}_H^{Tx}\right)^T. \tag{21}$$

Proof: See Appendix A.

Lemma 3: Let $\mathbf{P}^* \boldsymbol{\Psi}'_H \mathbf{P}$ be full rank. The lower bound $\mathbf{E}'$ in (20) is asymptotically tight with the SNR γ in the following sense:

$$\text{i)} \quad \lim_{\gamma\to\infty} \operatorname{tr}\left[\mathbb{E}\{\mathbf{E}\}\right] = \lim_{\gamma\to\infty} \operatorname{tr}[\mathbf{E}'] = 0 \tag{22}$$

$$\text{ii)} \quad 1 \le \lim_{\gamma\to\infty} \frac{\operatorname{tr}\left[\mathbb{E}\{\mathbf{E}\}\right]}{\operatorname{tr}[\mathbf{E}']} \le c_1 \quad \text{for } \ell < m \tag{23}$$

$$\lim_{\gamma\to\infty} \frac{1}{\ln\gamma} \frac{\operatorname{tr}\left[\mathbb{E}\{\mathbf{E}\}\right]}{\operatorname{tr}[\mathbf{E}']} \le c_2 \quad \text{for } \ell = m \tag{24}$$

where $c_1 \ge 1$ and $c_2 > 0$.

Proof: See Appendix A. ∎

Observe that because of (22), there is no MSE floor, as later will be seen in the numerical results in Section IX. This differs from the case when both the CSIT and the CSIR are imperfect, where an MSE floor is always present. The reason is that, even though the transmitter is not well designed for a given realization of the channel, the receiver can be adapted to each realization which can compensate for the mismatch of the transmitter.

The accuracy of the lower bound is illustrated in Fig. 2. In this toy example, true $\mathbb{E}\{\mathbf{E}\}$ is compared with its lower bound $\mathbf{E}'$ by plotting their trace for different SNRs. All the parameters are set to very simple values ($[\mathbf{P}]_{ij} = \delta(i-j)$, $\mathbf{R}_n$ is a scaled identity matrix, $\mathbf{H} \sim \mathcal{CN}(\bar{\mathbf{H}}, \mathbf{I}_n \otimes \mathbf{I}_m)$ with $\bar{\mathbf{H}}$ given in Appendix B, and $P_{\max} = 1$). As shown in Fig. 2, this lower bound is sufficiently close as indicated by (23). If another $\mathbf{P}$ is used, similar results are obtained. Due to the tightness, our robust design in the sequel will be based on this lower bound. Another possibility would be to base the design on some tight upper bound of $\mathbb{E}\{\mathbf{E}\}$, which is not undertaken in this paper.

To solve the optimal transmitter $\mathbf{P}$, first observe that the inequality of Lemma 2 also holds for the diagonal entries element-wise

$$\mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right] \ge \mathbf{d}[\mathbf{E}']$$

so the increasing cost function of the design problem (12) is also lower-bounded as

$$\mathcal{F}_0\left(\mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right]\right) \ge \mathcal{F}_0\left(\mathbf{d}[\mathbf{E}']\right). \tag{25}$$

Replacing the cost function in (12) with the above lower bound, as the expression of $\mathbf{E}'$ has the same format as (7) but with $\boldsymbol{\Psi}'_H$ instead of $\boldsymbol{\Psi}_H$, the optimal transmitter $\mathbf{P}$ based on the tight lower bound (25) is also given by (9) but with $\boldsymbol{\Psi}'_H$ instead of $\boldsymbol{\Psi}_H$. What remains is to choose a proper unitary matrix $\mathbf{Q}$ and the power allocation $\{p_i\}$, which depend on the specific function $\mathcal{F}_0$.

The robust design can be summarized in the theorem below.

Theorem 2: When the CSIR is perfect but the CSIT is imperfect in the form of a channel mean $\bar{\mathbf{H}}$ and a channel covariance $\mathbf{R}_H = \mathbf{R}_H^{Tx} \otimes \mathbf{R}_H^{Rx}$, the optimal solution $(\mathbf{W}, \mathbf{P})$ to the robust design problem (12), with the objective function replaced by the lower bound in (25), has the same structure as the perfect CSI case in (6) and (9), but using the equivalent covariance matrix $\boldsymbol{\Psi}'_H$ in (21) instead of $\boldsymbol{\Psi}_H$.

Notice that even though the receiver has perfect CSI, the design of matrix $\mathbf{W}$ still depends on the transmitter $\mathbf{P}$. If the CSIT is obtained by slow CSI feedback from the receiver, it implies that the receiver also has the same CSIT and can calculate $\mathbf{P}$ locally. If, instead, the CSIT is obtained locally by the transmitter, there are two possible approaches to obtain knowledge of $\mathbf{P}$ at the receiver side: i) the transmitter can simply feed forward the matrix $\mathbf{P}$ to the receiver; ii) separate estimation of the channel statistics $(\bar{\mathbf{H}}, \mathbf{R}_H)$ are performed at the receiver side because the receiver has perfect CSI, and the receiver also calculates $\mathbf{P}$ locally. The second alternative saves bandwidth over feed forward, but it could give different estimates of the channel statistics, which would be a problem.

Remark: As an alternative to the lower bound in (20), it is also possible to expand the average MSE matrix $\mathbb{E}\{\mathbf{E}\}$ by the expectation of a truncated Neumann expansion of $\mathbf{H}^*\mathbf{R}_n^{-1}\mathbf{H}$ as proposed in [20]. When $\mathbf{R}_H^{Rx} = \mathbf{R}_n = \mathbf{I}$, the matrix $\mathbf{H}^*\mathbf{R}_n^{-1}\mathbf{H}$ follows a *noncentral* complex Wishart distribution [21], which is unfortunately very involved in computation as it utilizes Zonal polynomials [22], generalized Laguerre polynomials [23] or it is approximated by a central Wishart distribution [24]. The lower bound approximation presented herein performs better than the approach based on the truncated Neumann expansion in [20] due to the error accumulated from series truncation and moments approximation.

VIII. Unified Framework

Under different CSI assumptions, the linear transceiver design problem has different optimal solutions, but follows a unified framework summarized in Table I.

- The optimal receiver $\mathbf{W}$ is always a Wiener filter, regardless of the CSIT. The CSIR determines the parameters for $\mathbf{W}$: if the CSIR is perfect, $\mathbf{W}$ is calculated using $\mathbf{H}$ and $\mathbf{R}_n$; if the CSIR is imperfect, $\mathbf{W}$ is calculated with $\bar{\mathbf{H}}$ and the equivalent $\mathbf{R}''_n$ in (18).
- The optimal transmitter $\mathbf{P}$ always diagonalizes the equivalent channel covariance matrix $\boldsymbol{\Psi}_H$ (possibly up to a ro-

TABLE I
SUMMARY OF LINEAR TRANSCEIVER DESIGN

	Perfect CSI	Perfect CSIR, imperfect CSIT	Imperfect CSIR/CSIT
Receiver	$\mathbf{W} = (\mathbf{HPP}^*\mathbf{H}^* + \mathbf{R}_n)^{-1}\mathbf{HP}$		$\mathbf{W} = \left(\bar{\mathbf{H}}\mathbf{PP}^*\bar{\mathbf{H}}^* + \mathbf{R}_n''\right)^{-1}\bar{\mathbf{H}}\mathbf{P}$
Transmitter	$\mathbf{P} = \mathbf{U}_{\Psi,\ell}\boldsymbol{\Sigma}_P\mathbf{Q}$		
$\boldsymbol{\Psi}_H$	$\mathbf{H}^*\mathbf{R}_n^{-1}\mathbf{H}$	$\bar{\mathbf{H}}^*\mathbf{R}_n^{-1}\bar{\mathbf{H}} + \mathrm{tr}\left[\mathbf{R}_H^{Rx}\mathbf{R}_n^{-1}\right](\mathbf{R}_H^{Tx})^T$	$\bar{\mathbf{H}}^*\mathbf{R}_n''^{-1}\bar{\mathbf{H}}$

tation) for all CSIT conditions.[4] The equivalent $\boldsymbol{\Psi}_H$ is different for different CSI condition and is determined by both the CSIR and the CSIT: if both the CSIR and the CSIT are perfect, $\boldsymbol{\Psi}_H$ is calculated based on $\mathbf{H}$ and $\mathbf{R}_n$ as in (8); if only the CSIR is perfect, $\boldsymbol{\Psi}_H$ is calculated using $\bar{\mathbf{H}}$ and $\mathbf{R}_n$ as $\boldsymbol{\Psi}'_H$ in (21); if neither the CSIR nor the CSIT is perfect, $\boldsymbol{\Psi}_H$ is calculated with $\bar{\mathbf{H}}$ and $\mathbf{R}''_n$ as $\boldsymbol{\Psi}''_H$ in (19).

IX. NUMERICAL EXAMPLES

We herein present several robust design examples following the development in Sections VI and VII. These originally complicated robust design problems have solutions with clear structure, and are then reduced into simple convex problems with scalar power allocation variables, thus they are easily computed numerically.

The imperfect CSI is in the form of a mean and a covariance as in the Gaussian model (4). The channel mean $\bar{\mathbf{H}}$ is drawn from a complex Gaussian distribution $\bar{\mathbf{H}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_m)$ and kept fixed for all the simulations in this section. The specific value of the channel mean used to generate all the figures is given in Appendix B. Very similar curves are observed with different realizations of $\bar{\mathbf{H}}$. The covariance matrix $\mathbf{R}_H^{Tx}$ is Toeplitz, defined by the correlation coefficient ρ_t $(0 < \rho_t < 1)$ as $[\mathbf{R}_H^{Tx}]_{i,j} = \rho_t^{|i-j|}$. The covariance matrix $\mathbf{R}_H^{Rx}$ is defined similarly by its correlation coefficient ρ_r. The SNR is defined in (1). The Gaussian noise is white both spatially and temporally, so $\mathbf{R}_n$ is a scaled identity matrix. In all the examples, $m = n = 4$, $\ell = 2$, $P_{\max} = 1$, and QPSK is used.[5]

A. Example: Minimizing the Weighted Sum of the Average MSEs

The MSE is the central figure-of-merit of the above set of optimization problems. The nonrobust design of this problem is solved in [2], [3], [25]. The more complicated robust design problem is the minimization of the weighted sum of the average MSEs, which can be solved by our proposed solutions. The problem can be formulated as

$$\begin{array}{ll} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} & \mathbf{w}^T\mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right] \\ \text{subject to} & \mathrm{tr}[\mathbf{PP}^*] \leq P_{\max} \end{array} \tag{26}$$

[4]Observe that the transceiver design for perfect CSIR and imperfect CSIT case is based on a sufficiently close lower bound. Also, the closed-form solution of the transmitter design for imperfect CSIR and CSIT case can only be obtained when $\mathbf{R}_H^{Tx} = \mathbf{I}_n$.

[5]With a larger ℓ close to $\min\{m, n\}$, the average BER or average MSE performance will be more likely dominated by the worst data stream and thus diminish the performance difference between different designs. However, our conclusions will still hold even for larger ℓ. Here $\ell = 2$ is chosen for clear illustrations.

where $\mathbf{w} = [w_1, w_2, \cdots, w_\ell]^T$ is a positive weighting vector with elements in increasing order (or equal) to guarantee that the objective function in (26) is a Schur-concave function (cf. [3]). It follows from the Schur-concavity that $\mathbf{Q} = \mathbf{I}_\ell$.

Imperfect CSIT and Perfect CSIR: Invoking Theorem 2, the optimal receiver $\mathbf{W}$ is given by (6) and optimal transmitter $\mathbf{P}$ is given by (9) but with $\boldsymbol{\Psi}'_H$ given in (21). What remains is to find the optimal power allocation $\{p_i\}$. By using the expressions for $\mathbf{W}$ and $\mathbf{P}$, the problem (26) is simplified to

$$\begin{array}{ll} \underset{\mathbf{P}}{\text{minimize}} & \displaystyle\sum_{i=1}^{\ell} \frac{w_i}{1 + p_i\lambda'_i} \\ \text{subject to} & \displaystyle\sum_{i=1}^{\ell} p_j \leq P_{\max} \\ & p_j \geq 0 \end{array}$$

where $\{\lambda'_1, \lambda'_2, \cdots, \lambda'_\ell\}$ are the ℓ strongest eigenvalues of $\boldsymbol{\Psi}'_H$ in increasing order. This particular scalar optimization problem can be solved efficiently by the water-filling solution

$$p_i = \left(\mu w_i^{1/2}\lambda_i'^{-1/2} - \lambda_i'^{-1}\right)^+ \tag{27}$$

where μ is the water-level satisfying the power constraint with equality.

Imperfect CSIT and CSIR: Invoking Theorem 1, the optimal receiver $\mathbf{W}$ is given by (6), but using $\bar{\mathbf{H}}$ and $\mathbf{R}''_n$. The optimal $\mathbf{P}$ is also given by (9), but using $\boldsymbol{\Psi}''_H$ given in (19). The power allocation $\{p_i\}$ can be solved exactly the same way as in (27), but with λ''_i determined by $\boldsymbol{\Psi}''_H$.

Simulations: Monte Carlo simulations are carried out for a 4 × 4 MIMO system to compare four different linear transceiver designs: i) naive solution that assumes the channel estimate $\bar{\mathbf{H}}$ to be a perfect channel estimate of the instantaneous realization $\mathbf{H}$ and utilizes the solution of Section IV; ii) the robust solution where both CSIT and CSIR are imperfect as in Theorem 1; iii) the robust solution where CSIR is perfect, but CSIT is imperfect as in Theorem 2; iv) the ideal solution assuming exact instantaneous CSIR and CSIT as in Section IV.

The numerical comparisons are shown in Fig. 3. In this example, $\rho_t = 0$, $\rho_r = 0.6$ and the weights $w_i = 1$. The robust design for imperfect CSIT and CSIR performs better than the naive design, but still much worse than the perfect CSI case. An MSE floor is clearly visible, as predicted by (17) in Section VI. The robust design for imperfect CSIT and perfect CSIR performs close to the ideal perfect CSI case and the slope is approximately the same. In this case, no MSE floor can be observed as predicted by (22) in Section VII. This figure also reveals how much the linear transceiver design can gain from having perfect CSIR and/or CSIT: the knowledge of perfect CSIR is clearly

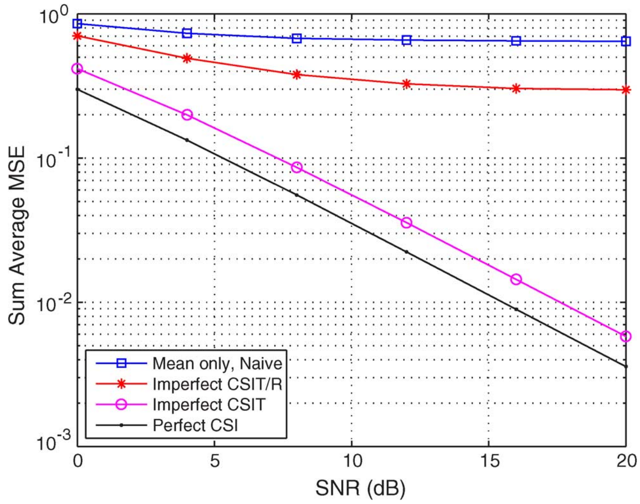

Fig. 3. Sum average MSE versus SNR for different CSI ($m = 4$, $n = 4$, $\ell = 2$, $\rho_r = 0.6$, $\rho_t = 0$).

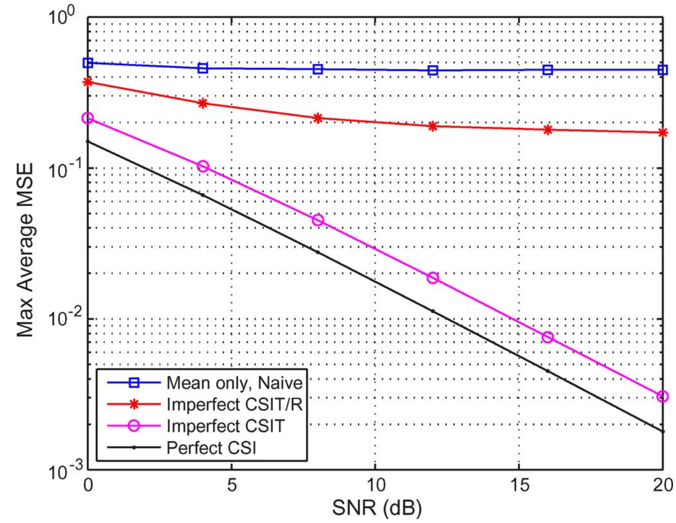

Fig. 4. Max average MSE versus SNR for different CSI ($m = 4$, $n = 4$, $\ell = 2$, $\rho_r = 0.6$, $\rho_t = 0$).

more critical, as it provides significant gain in average MSE. Interestingly, our robust design with imperfect CSIT only suffers a slight loss from the ideal case, although no instantaneous CSIT is available to optimize the transmitter.

B. Example: Minimizing the Maximum of the Average MSEs

In addition to the first example of minimizing the weighted average MSEs, one can also consider minimizing the worst average MSE among the ℓ data streams, because the overall performance is usually dominated by the data stream with the worst average MSE. The problem can be written as

$$\begin{array}{ll} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} & \max_i \mathbf{d}_i\left[\mathbb{E}\{\mathbf{E}\}\right] \\ \text{subject to} & \text{tr}[\mathbf{P}\mathbf{P}^*] \leq P_{\max}. \end{array} \tag{28}$$

The solution to this problem is quite simple because the cost function is a Schur-convex function of the average MSEs (cf. [3]). The unitary matrix $\mathbf{Q}$ is, therefore, a rotation that produces equal diagonal elements of $\mathbb{E}\{\mathbf{E}\}$ and the power allocation can be simply obtained by minimizing $\text{tr}[\mathbb{E}\{\mathbf{E}\}]$, which is a Schur-Concave function and solved as in the first example (26) with equal weights.

Monte Carlo simulations are carried out for a 4×4 MIMO system to compare four different linear transceiver designs with the exactly same setup as previous example. The parameters are $\rho_t = 0$ and $\rho_r = 0.6$. The results are shown in Fig. 4. Similar conclusions can be drawn: the proposed robust designs outperforms the naive design; close-to-ideal performance can be achieved by the proposed robust design even when CSIT is not perfect; perfect CSIR is crucial to obtain acceptable performance in terms of average MSEs.

C. Extensions to Design Problem (13)

All our robust linear transceiver designs are so far based on a cost function $calF_0$ of the average MSEs as in (12). The robust design problem (13) is more difficult to solve, because the expectation operator does not necessarily commute with the function $\mathcal{F}_0$ and can not be applied to the arguments $\mathbf{d}[\mathbf{E}]$. However, these two problems are related: the cost function $\mathbb{E}\{\mathcal{F}_0(\mathbf{d}[\mathbf{E}])\}$ in (13) can be expanded by a Taylor series and then truncated with only the first two terms, which is the cost function in (12) and can be solved by the proposed robust designs, in the hope that the resulting solutions are close, in some sense, to the solutions of the original cost function $\mathbb{E}\{\mathcal{F}_0(\mathbf{d}[\mathbf{E}])\}$.

For a general differentiable function $\mathcal{F}_0$, more specifically, an approximation of $\mathbb{E}\{\mathcal{F}_0(\mathbf{d}[\mathbf{E}])\}$ with bounded error can be obtained as

$$\begin{aligned} \mathbb{E}\left\{\mathcal{F}_0\left(\mathbf{d}[\mathbf{E}]\right)\right\} =& \mathcal{F}_0\left(\mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right]\right) \\ &+\mathbb{E}\left\{\mathbf{d}[\mathbf{E}]-\mathbf{d}[\mathbb{E}\{\mathbf{E}\}]\right\}^T \nabla_{\mathbf{x}}\mathcal{F}_0(\mathbf{x})|_{\mathbf{x}=\mathbf{d}[\mathbb{E}\{\mathbf{E}\}]} \\ &+O\left(\mathbb{E}\left\{\|\mathbf{d}[\mathbf{E}]-\mathbb{E}\left\{\mathbf{d}[\mathbf{E}]\right\}\|^2\right\}\right) \\ =& \mathcal{F}_0\left(\mathbf{d}\left[\mathbb{E}\{\mathbf{E}\}\right]\right) \\ &+O\left(\mathbb{E}\left\{\|\mathbf{d}[\mathbf{E}]-\mathbb{E}\left\{\mathbf{d}[\mathbf{E}]\right\}\|^2\right\}\right). \end{aligned}$$

The error in the approximation depends on both the particular function $\mathcal{F}_0$ and the variance of $\mathbf{E}$.

We illustrate the above idea with the example of minimizing the average BER, which belongs to the design problem (13). The average uncoded BER is a good measure of the uncoded link quality of a communication system. Hence, its minimization can be regarded a good criterion:

$$\begin{array}{ll} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} & \mathbb{E}\left\{\dfrac{1}{\ell}\displaystyle\sum_{i=1}^{\ell}\mathcal{B}\left(d_i(\mathbf{E})\right)\right\} \\ \text{subject to} & \text{tr}[\mathbf{P}\mathbf{P}^*] \leq P_{\max} \end{array} \tag{29}$$

where function $\mathcal{B}(\cdot)$ is a function mapping MSE to BER. We also assume the same M-ary QAM constellation is used for all data streams. For square M-ary QAM constellations, the function $\mathcal{B}$ is convex for the range of MSEs of interest [26] and is given by

$$\mathcal{B}(x) \approx \frac{4}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)\mathcal{Q}\left(\sqrt{\frac{3}{M-1}\left(\frac{1}{x}-1\right)}\right) \tag{30}$$

where $\mathcal{Q}$ is the Q-function defined for the Gaussian distribution.

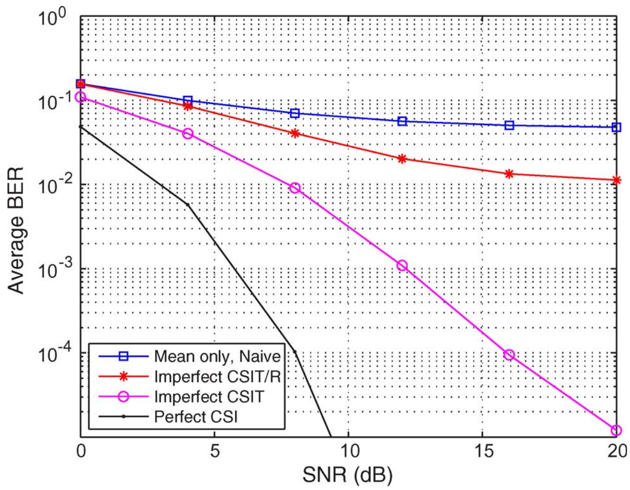

Fig. 5. Average BER versus SNR for different CSI ($m = 4, n = 4, \ell = 2, \rho_r = 0.6, \rho_t = 0$).

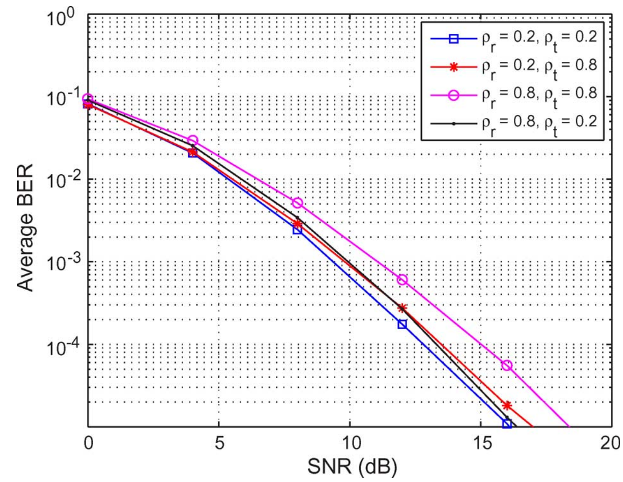

Fig. 6. Average BER versus SNR with different correlations for the robust design of imperfect CSIT and perfect CSIR ($m = 4, n = 4, \ell = 2$).

In order to approximate the above problem to (12), simply take the first two terms in the Taylor expansion for the expectation of the function $\mathcal{B}$.[6] The problem is then approximately reduced to

$$\begin{array}{ll} \underset{\mathbf{W},\mathbf{P}}{\text{minimize}} & \sum_{i=1}^{l} \mathcal{B}\left(d_i\left(\mathbb{E}\{\mathbf{E}\}\right)\right) \\ \text{subject to} & \text{tr}[\mathbf{P}\mathbf{P}^*] \leq P_{\max} \end{array}$$

which indeed belongs to problem (12) and can be solved by our proposed design framework.

First, the cost function is a sum of identical convex functions so it is Schur-convex [14]. The unitary matrix $\mathbf{Q}$ is a rotation that produces equal diagonal elements of $\mathbb{E}\{\mathbf{E}\}$ and the power allocation can be simply obtained by minimizing $\text{tr}[\mathbb{E}\{\mathbf{E}\}]$, which is a Schur-concave function and solved by the first example (26). The optimal receiver $\mathbf{W}$ can be determined according to different CSI conditions as in Theorem 1 or 2.

Monte Carlo simulations are carried out for a 4×4 MIMO system to compare four different linear transceiver designs with the exactly same setup as previous examples. The parameters are $\rho_t = 0$ and $\rho_r = 0.6$. The results are shown in Fig. 5. The proposed robust designs for different CSI work better than the naive design based on channel mean only. Similar to Fig. 3, a perfect CSIR is more important and provides a large gain in SNR. However, unlike Fig. 3, the performance of the robust design with imperfect CSIT and perfect CSIR is now clearly inferior to the perfect CSI design. This is most probably due to the Taylor approximation of $\mathcal{Q}$ function is simple but not very accurate and (30) is not accurate when the interference is not Gaussian.

D. Correlation and K Factor Effect

We examine the impact of channel correlation by varying the correlation coefficients ρ_r and ρ_t. Our robust design for imperfect CSIT and perfect CSIR are tested against four different setups of channel correlations: both transmitter and receiver have high or low correlation, either transmitter or receiver has high correlation. Average BER is used as the cost function and the robust design follows the previous example. The simulation results are shown in Fig. 6. The relative position of the curves suggests that low correlation on both the transmitter and the receiver side results in the best average BER performance, while the high correlated transmitter and receiver combination is the worst. When the SNR is low, the correlation at the receiver is the dominating factor for BER, while the correlation at the transmitter is the dominating factor when the SNR is high. Intuitively, this phenomenon is reasonable as it is less important to utilize more receiver antennas to average out the Gaussian noise effect when the SNR is high.

To study the effect of K factor, the channel model (4) is slightly modified as

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\bar{\mathbf{H}} + \sqrt{\frac{1}{K+1}}\left(\mathbf{R}_H^{Rx}\right)^{1/2}\mathbf{G}\left(\mathbf{R}_H^{Tx}\right)^{T/2}$$

where K determines the relative power of the mean and the covariance of the channel, or equivalently, the channel estimate quality. The channel mean $\bar{\mathbf{H}}$ is normalized such that $\text{tr}[\bar{\mathbf{H}}^*\bar{\mathbf{H}}] = mn$. Therefore a constant average power gain of the channel can be maintained for different correlation coefficients and channel means, i.e.,

$$\mathbb{E}\left\{\text{tr}[\mathbf{H}^*\mathbf{H}]\right\} = mn$$

is satisfied for all K factors.

Again the example of average BER is used to demonstrate the performance, as shown in Fig. 7. In this example, $\rho_t = 0.2$ and $\rho_r = 0.8$. When K is close to 0, which means the channel is close to Rayleigh, the naive design with only channel mean information naturally fails. The proposed robust design benefits from the fact that its equivalent channel covariance matrix incorporates both the channel mean and the channel covariance, for all Ks. When the K factor is extremely high at the right-hand side of the figure, all three methods approach the same performance, because the channel mean dominates over the channel covariance matrix.

[6]This is not an accurate approximation of the expectation of the $\mathcal{Q}$ function. Better approximations could be clearly considered but we limit the scope of the present paper to the considered approximation for its simplicity.

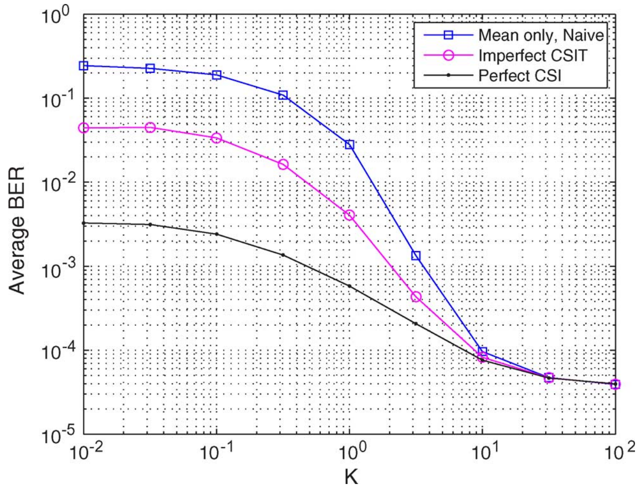

Fig. 7. Average BER versus K factor for the robust design of imperfect CSIT and perfect CSIR ($m=4, n=4, \ell=2, \rho_r=0.8, \rho_t=0.2$, SNR$=10$ dB).

X. Conclusion

We have addressed the robust design problem of linear transceivers for MIMO channels based on a general cost function of the average MSEs, for the cases of i) imperfect CSIR and imperfect CSIT, and ii) perfect CSIR and imperfect CSIT. The proposed framework solves the design problem when the CSI takes the form of a channel mean and a channel covariance matrix. In particular, a lower bound of the average MSE matrix is explored for the design when only the CSIT is imperfect. The resulting designs exhibit a similar structure as in the case of perfect CSI, which can be readily solved by convex optimization approaches in practice. Therefore, linear MIMO transceiver designs with general cost functions of the MSEs or average MSEs can be solved with a unified framework, even for different CSI conditions. By Taylor approximations, it is also possible to extend the proposed designs to the robust design problem based on the expected value of general functions of the *instantaneous* MSEs. Promising gains compared to simple nonrobust designs are confirmed with numerical examples.

Appendix A
Proof of Lemmas 2 and 3

Definition 5: [14, Ch. 16] Let function $g:\mathcal{A}\subseteq\mathbb{C}^{m\times k}\mapsto\mathbb{H}^{n\times n}$. For all random matrices $\mathbf{X}$ taking values in $\mathcal{A}$ and having finite expectation $\mathbb{E}\{\mathbf{X}\}$,

$$g \text{ is matrix}-\text{convex on } \mathcal{A} \quad \Leftrightarrow \quad \mathbb{E}\{g(\mathbf{X})\}\geq g\left(\mathbb{E}\{\mathbf{X}\}\right).$$

This is also the Jensen's inequality for matrix-valued functions.

Lemma 4: [27] On the set of $n\times n$ positive definite Hermitian matrices, the matrix-valued function

$$g(\mathbf{X})=\mathbf{X}^{-1}$$

is matrix-convex, i.e.,

$$\mathbb{E}\{\mathbf{X}^{-1}\}\geq\mathbb{E}\{\mathbf{X}\}^{-1}.$$

We start by proving Lemma 2. As the matrix inverse is strictly matrix-convex on the set of positive definite Hermitian matrices per Lemma 4, the following inequality holds:

$$\begin{aligned}\mathbb{E}\{\mathbf{E}\} &\geq \left(\mathbf{I}_\ell+\mathbf{P}^*\mathbb{E}\left\{\mathbf{H}^*\mathbf{R}_n^{-1}\mathbf{H}\right\}\mathbf{P}\right)^{-1} &(31)\\ &= \left(\mathbf{I}_\ell+\mathbf{P}^*\mathbf{\Psi}_H'\mathbf{P}\right)^{-1} &(32)\end{aligned}$$

where $\mathbf{\Psi}_H'$ is given in (21) and the equality follows from Lemma 1.

The lower bound in (23) of Lemma 3 follows directly from (20). To prove the upper bounds, first denote

$$\begin{aligned}\tilde{\mathbf{R}}_n &= \frac{\mathbf{R}_n}{\mathrm{tr}[\mathbf{R}_n]} &(33)\\ \tilde{\mathbf{\Psi}}_H' &= \mathbb{E}\left\{\mathbf{H}^*\tilde{\mathbf{R}}_n^{-1}\mathbf{H}\right\} &(34)\\ \tilde{\mathbf{P}} &= \frac{\mathbf{P}}{\sqrt{P}} &(35)\end{aligned}$$

so that $\tilde{\mathbf{R}}_n$, $\tilde{\mathbf{\Psi}}_H'$, and $\tilde{\mathbf{P}}$ are now independent of the SNR γ. Let $\{\mu_i\}$ be the eigenvalues of the matrix $\tilde{\mathbf{P}}^*\mathbf{H}^*\tilde{\mathbf{R}}_n^{-1}\mathbf{H}\tilde{\mathbf{P}}$ and $\{\nu_i\}$ be the eigenvalues of the matrix $\tilde{\mathbf{P}}^*\tilde{\mathbf{\Psi}}_H'\tilde{\mathbf{P}}$, both in increasing order. Therefore, the following holds:

$$\begin{aligned}\mathrm{tr}[\mathbf{E}'] &= \mathrm{tr}\left[\left(\mathbf{I}+\gamma\tilde{\mathbf{P}}^*\tilde{\mathbf{\Psi}}_H'\tilde{\mathbf{P}}\right)^{-1}\right]\\ &= \sum_{i=1}^{\ell}(1+\gamma\nu_i)^{-1}\geq(1+\gamma\nu_1)^{-1}. &(36)\end{aligned}$$

As $\tilde{\mathbf{P}}^*\mathbf{\Psi}_H'\tilde{\mathbf{P}}$ is full rank, we have $\nu_1>0$.

The trace of the average MSE matrix can be upper bounded as

$$\begin{aligned}\mathrm{tr}\left[\mathbb{E}\{\mathbf{E}\}\right] &= \mathbb{E}\left\{\mathrm{tr}\left[\left(\mathbf{I}_\ell+\gamma\tilde{\mathbf{P}}^*\mathbf{H}^*\tilde{\mathbf{R}}_n^{-1}\mathbf{H}\tilde{\mathbf{P}}\right)^{-1}\right]\right\}\\ &= \mathbb{E}\left\{\sum_{i=1}^{\ell}\frac{1}{1+\gamma\mu_i}\right\}\\ &\leq \mathbb{E}\left\{\frac{\ell}{1+\gamma\mu_1}\right\}\\ &\leq \mathbb{E}\left\{\frac{\ell}{1+\gamma\mu_1}\Big|\mu_1\leq 1\right\}\\ &\leq c_0\int_0^1\frac{1}{1+\gamma\mu}\mu^{m-\ell}d\mu. &(37)\end{aligned}$$

Because $\tilde{\mathbf{P}}^*\mathbf{H}^*\tilde{\mathbf{R}}_n^{-1}\mathbf{H}\tilde{\mathbf{P}}$ is noncentral quadratic forms [15], c_0 is a positive constant associated with its noncentrality and independent of γ, and (37) is due to the pdf of μ_1.[7]

[7] The pdf of μ_1 takes the form of $f_{\mu_1}(\mu)=C_{\ell,m}\mu^{m-\ell}P_{\ell,m}(\mu)D_\alpha$, where $C_{\ell,m}$ is a constant, $P_{\ell,m}(\mu)$ is a polynomial of degree $(m-\ell)(\ell-1)$, and D_α is a constant obtained by the integral of two related invariant polynomials [28], [29]. Exact values of these parameters are not relevant in this proof. Conditioned on $\mu_1\leq 1$, the pdf can be written as $f_{\mu_1|\mu_1\leq 1}(\mu)=f_{\mu_1}(\mu)/F_{\mu_1}(1)\leq c_0\mu^{m-\ell}$, where $c_0=C_{\ell,m}D_\alpha\max_{0\leq\mu\leq 1}P_{\ell,m}(\mu)/F_{\mu_1}(1)$ and F_{μ_1} is the cumulative distribution function of μ_1.

To prove the upper bound in (23) while $\ell < m$, take the limit as

$$\begin{aligned}\lim_{\gamma\to\infty}\frac{\mathrm{tr}\left[\mathbb{E}\{\mathbf{E}\}\right]}{\mathrm{tr}[\mathbf{E}']} &\le \lim_{\gamma\to\infty}\frac{c_0\int_0^1 \frac{1}{1+\gamma\mu}\mu^{m-\ell}d\mu}{(1+\gamma\nu_1)^{-1}} \quad (38)\\ &= \lim_{\gamma\to\infty}\frac{\frac{c_0}{\gamma}\left[\sum_{k=0}^{m-\ell-1}\frac{(-1)^k}{(m-\ell-k)\gamma^k}+\frac{(-1)^{m-\ell}}{\gamma^{m-\ell}}\ln(1+\gamma)\right]}{(1+\gamma\nu_1)^{-1}} \quad (39)\\ &= c_0\frac{\nu_1}{m-\ell} = c_1. \quad (40)\end{aligned}$$

To prove the upper bound in (24) while $\ell = m$, take the limit as

$$\begin{aligned}\lim_{\gamma\to\infty}\frac{1}{\ln\gamma}\frac{\mathrm{tr}\left[\mathbb{E}\{\mathbf{E}\}\right]}{\mathrm{tr}[\mathbf{E}']} &\le \lim_{\gamma\to\infty}\frac{c_0\int_0^1 \frac{1}{1+\gamma\mu}\mu^{m-\ell}d\mu}{(1+\gamma\nu_1)^{-1}\ln\gamma} \quad (41)\\ &= \lim_{\gamma\to\infty}\frac{c_0\gamma^{-1}\ln(1+\gamma)}{(1+\gamma\nu_1)^{-1}\ln\gamma} \quad (42)\\ &= c_0\nu_1 = c_2 \quad (43)\end{aligned}$$

where both c_1 and c_2 are positive and independent of SNR γ.

To prove (22), because $\nu_i > 0$, the following limit holds:

$$\lim_{\gamma\to\infty}\mathrm{tr}[\mathbf{E}'] = \lim_{\gamma\to\infty}\sum_{i=1}^{\ell}(1+\gamma\nu_i)^{-1} = 0. \quad (44)$$

The limit of the trace of the average MSE matrix is also upper bounded when $\ell < m$

$$\begin{aligned}\lim_{\gamma\to\infty}\mathrm{tr}\left[\mathbb{E}\{\mathbf{E}\}\right] &\le \lim_{\gamma\to\infty} c_0\int_0^1\frac{1}{1+\gamma\mu}\mu^{m-\ell}d\mu\\ &= \lim_{\gamma\to\infty}\frac{c_0}{\gamma}\left[\sum_{k=0}^{m-\ell-1}\frac{(-1)^k}{(m-\ell-k)\gamma^k}\right.\\ &\qquad\left.+\frac{(-1)^{m-\ell}}{\gamma^{m-\ell}}\ln(1+\gamma)\right] = 0 \quad (45)\end{aligned}$$

while when $\ell = m$

$$\begin{aligned}\lim_{\gamma\to\infty}\mathrm{tr}\left[\mathbb{E}\{\mathbf{E}\}\right] &\le \lim_{\gamma\to\infty} c_0\int_0^1\frac{1}{1+\gamma\mu}\mu^{m-\ell}d\mu\\ &= \lim_{\gamma\to\infty}\frac{c_0}{\gamma}\ln(1+\gamma) = 0. \quad (46)\end{aligned}$$

Appendix B
Numerical Simulation Parameter

The channel mean $\bar{\mathbf{H}}$ used in the numerical simulations is

$$\bar{\mathbf{H}} = \begin{bmatrix} .33+.47i & 1.03-.96i & .88-.17i & -.94+.82i\\ .58+.01i & .93-.08i & -.56-.12i & 1.02-.32i\\ .73-.05i & .49-.56i & -.36-.67i & -.39+.72i\\ -.62-1.72i & .51+.95i & 1.00-.88i & -.09-.05i\end{bmatrix}.$$

Acknowledgment

The authors would like to gratefully acknowledge Dr. D. Tuninetti, whose constructive suggestion led to the new tight lower bound of the averaged MSE matrix which simplifies the transceiver design. The authors would also like to extend the acknowledge to the anonymous reviewers and the associate editor for their insightful comments, which helped to improve the manuscript significantly. The first author (X. Zhang) acknowledges Dr. J. Jaldén for the discussion of the proof of Lemma 2.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, 6, no. 3, p. 331, Mar. 1998.

[2] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[3] 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.

[4] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*. Boca Raton, FL: CRC, Aug. 2001.

[5] S. Vorobyov, A. 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.

[6] A. Pascual-Iserte, D. P. Palomar, A. I. Prez-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.

[7] A. P. Iserte, A. Perez-Neira, and M. L. Hernandez, "Exploiting transmission spatial diversity in frequency selective systems with feedback channel," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, pp. IV-85–IV-88, Apr. 2003.

[8] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 1070–1085, Mar. 2005.

[9] E. Jorswieck, A. Sezgin, H. Boche, and E. Costa, "Optimal transmit strategies in MIMO Ricean channels with MMSE receiver," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2004, pp. 3787–3791.

[10] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

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

[12] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, Jun. 2002.

[13] 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.

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

[15] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall/CRC, 2000.

[16] D.-S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 503–513, Mar. 2000.

[17] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wideband MIMO radio channels based on NLOS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.

[18] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[19] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2002.

[20] X. Zhang, D. P. Palomar, and B. Ottersten, "Robust design of linear MIMO transceivers under channel uncertainty," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2006, p. IV.

[21] C. G. Khatri, "Classical statistical analysis based on a cetrain multivariate complex Gaussian distribution," *Annals Math. Statist.*, vol. 36, no. 1, pp. 98–144, Feb. 1965.

[22] A. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, 1964.

[23] J. Roux and E. Raath, "Generalized Laguerre series forms of Wishart distributions," *South African Statist. J.*, vol. 7, pp. 23–34, 1973.

[24] H. Steyn and J. Roux, "Approximations for the noncentral Wishart distribution," *South African Statist. J.*, vol. 6, pp. 165–173, 1972.

[25] H. Sampath, P. Stoica, and A. Paulraj, “Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion,” *IEEE Trans. Commun.*, vol. 49, pp. 2198–2206, Dec. 2001.
[26] J. Proakis and M. Salehi, *Communication Systems Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[27] I. Olkin and J. Pratt, “A multivariate Tchebycheff inequality,” *Ann. Math. Statist.*, vol. 29, no. 1, pp. 226–234, Mar. 1958.
[28] A. W. Davis, “Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory,” *Ann. Inst. Statist. Math.*, vol. 31, pp. 465–485, 1979.
[29] T. Ratnarajah, “Non-central quadratic forms on complex random matrices and applications,” in *Proc. 13th IEEE/SP Workshop Statist. Signal Process.*, Jul. 2005, pp. 555–560.

Xi Zhang (S’03) received the B.Sc. degree in electrical engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2000 and the M.Sc. degree with specialization in wireless systems from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002.

From 2000 to 2001, he was with Microsoft Asia Regional Engineering Center, Shanghai. From April to October 2006, he was with the Information Systems Lab (ISL), Stanford University, CA, as a visiting researcher. At present, he is working toward the Ph.D. degree in telecommunications with the School of Electrical Engineering, KTH. His research interests include transceiver design for MIMO systems and performance analysis and optimization of scheduling algorithms for multiuser MIMO systems.

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

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

Prof. Palomar is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on Game Theory in Communication Systems, and the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on Optimization of MIMO Transceivers for Realistic Communication Networks. He received a 2004–2006 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002–2003 Best Ph.D. Prize in Information Technologies and Communications by the Technical University of Catalonia (UPC); the 2002–2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation; and the 2004 prize for the Best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.

Björn Ottersten (F’04) was born in Stockholm, Sweden, 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, in 1989.

He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, and the Katholieke Universiteit Leuven, Leuven. During 1996–1997, he was Director of Research at ArrayComm Inc., San Jose, CA. In 1991, he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm, and is currently Dean of the School of Electrical Engineering at KTH. From 1992 to 2004, he was Head of the Department for Signals, Sensors, and Systems at KTH. He is also a visiting professor with t the University of Luxembourg. His research interests include wireless communications, stochastic signal processing, sensor array processing, and time-series analysis.

Dr. Ottersten 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*. He has coauthored papers that received an IEEE Signal Processing Society Best Paper Award in 1993, 2001, and 2006.