Maximum Mutual Information Design for MIMO Systems With Imperfect Channel Knowledge

Minhua Ding, *Member, IEEE*, and Steven D. Blostein, *Senior Member, IEEE*

***Abstract*—New results on maximum mutual information design for multiple-input multiple-output (MIMO) systems are presented, assuming that both transmitter and receiver know only an estimate of the channel state as well as the transmit and receive correlation. Since an exact capacity expression is difficult to obtain for this case, a tight lower-bound on the mutual information between the input and the output of a MIMO channel has been previously formulated as a design criterion. However, in the previous literature, there has been no analytical expression of the optimum transmit covariance matrix for this lower-bound. Here it is shown that for the general case with channel correlation at both ends, there exists a unique and globally optimum transmit covariance matrix whose explicit expression can be conveniently determined. For the special case with transmit correlation only, the closed-form optimum transmit covariance matrix is presented. Interestingly, the optimal transmitters for the maximum mutual information design and the minimum total mean-square error design share the same structure, as they do in the case with perfect channel state information. Simulation results are provided to demonstrate the effects of channel estimation errors and channel correlation on the mutual information.**

***Index Terms*—Channel state information (CSI), mean-square error (MSE), multiple-input multiple-output (MIMO), mutual information, optimization.**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) systems are known to be capable of providing high data rates without increasing bandwidth in rich scattering wireless fading channels [1]. However, the capacity of a MIMO channel depends on the availability of channel state information (CSI) at both ends. Correspondingly, different transmit strategies should be used with different types of CSI. The case when the fading channel is perfectly known to both ends has been studied in [1]–[3] and [20]. More recently, optimal transmit strategies are obtained for the case when the CSI at the receiver (CSIR) is perfect and the CSI at the transmitter (CSIT) is the channel mean or covariance information [4]. The noncoherent case with no *instantaneous* CSIT or CSIR has been studied in [5] and [6]. A comprehensive overview of the capacity results of MIMO systems can be found in [7]. In the above, either a perfect coherent system (perfect CSIR) or a noncoherent system (no instantaneous CSIR) has been assumed.

In [8], a different MIMO channel scenario is considered, where the CSIR is obtained through channel estimation and contains estimation errors. The CSIT is assumed to be obtained from the receiver via a lossless feedback link and is the same as the CSIR. The CSI at both ends consists of the channel estimate and channel correlation information. Under this assumption of CSI, an exact capacity expression is hardly tractable. Instead, tight upper- and lower-bounds on capacity have been proposed for system design [8], which are generalizations from those for a single-input single-output (SISO) channel [9]. The case when the CSI at both ends consists of channel estimates and channel correlation has been studied in [10] and [11], where the upper- and lower-bounds are shown to be close and thus are both tight. In particular, the lower-bound on the ergodic capacity has been formulated and used as the design criterion [10], [11]. Unfortunately, so far, no expression for the optimum transmit covariance matrix has been obtained for the capacity lower-bound with channel mean (i.e., channel estimate) and channel correlation information at both ends. In this paper, we attempt to solve this problem.

Our main contributions are listed as follows:

- We show that a globally optimum transmit covariance matrix exists for the capacity lower-bound. We also present its expression, which clarifies the transmitter structure and can be conveniently determined.
- The methodology employed to determine the optimum transmit covariance matrix enables us to determine the relationship between the maximum mutual information design and minimum total mean-square error (MSE) design with imperfect CSI.
- In [11], due to the absence of the optimum covariance matrix for the capacity lower-bound, the effects of the same amount of transmit and receive correlation are found to be different (see [11, Figs. 4 and 5]). Based on the optimum transmit covariance matrix obtained here, we reassess the effects of transmit and receive correlation, and observe different results from those in [11].

The differences between our work and that in the literature can be highlighted as follows:

- In [12, Sec. VI], linear MIMO transceiver designs with imperfect CSI at both ends have been considered. The authors have obtained results for the case with receive correlation only, which is mathematically equivalent to the perfect CSI

Manuscript received November 04, 2008; revised June 26, 2009. Date of current version September 15, 2010. This work was performed while M. Ding was with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada, and was supported under the Natural Sciences and Engineering Research Council of Canada Discovery Grant 41731.

M. Ding is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: minhua.ding@ieee.org).

S. D. Blostein is with the Department of Electrical and Computer Engineering, Queen's University at Kingston, Kingston, ON K7L 3N6 Canada (e-mail: steven.blostein@queensu.ca).

Communicated by G. Taricco, Associate Editor for Communications.

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/TIT.2010.2059870

0018-9448/$26.00 © 2010 IEEE

case. However, it is the presence of transmit correlation that requires a nontrivial new problem to be solved. In this paper, we consider the case with transmit correlation alone as well as the general case with both transmit and receive correlation.

- In the special case with only transmit correlation, a numerical search method has been proposed in [10, Sec. IV-B]. For this case, we provide a unique closed-form optimum solution.
- Unlike [10, Sec. IV-B] or [12, Sec. VI], our method is to formulate an equivalent problem first, and then apply the general nonlinear programming method based on the associated Karush–Kuhn–Tucker (KKT) optimality conditions.

A. Notation

Upper (lower) case boldface letters are for matrices (vectors); $\mathbb{E}\{\cdot\}$ denotes statistical expectation and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix; $|\mathbf{A}|$ stands for the determinant of matrix $\mathbf{A}$, whereas $|a|$ denotes the magnitude of scalar a; $(\cdot)^*$ and $(\cdot)^H$ denote the complex conjugate and complex conjugate transpose (Hermitian), respectively; $(b)_+ = \max(b,0)$; $\mathbf{I}_a$ is the $a \times a$ identity matrix; $\mathcal{N}_c(\cdot,\cdot)$ represents the circularly symmetric complex Gaussian distribution; $\mathbf{A} \succeq 0$ means that $\mathbf{A}$ is positive semidefinite; $\|\mathbf{A}\|_F$ is the Frobenius norm of $\mathbf{A}$ defined as $[\mathrm{tr}(\mathbf{A}\mathbf{A}^H)]^{\frac{1}{2}}$.

II. System Model and Problem Formulation

A. System Model

Consider a single-user MIMO communication system. The flat-fading MIMO channel, with n_T antennas at the transmitter and n_R antennas at the receiver, is represented by the $n_R \times n_T$ matrix $\mathbf{H}$. The system is described by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{x}$ is the $n_T \times 1$ zero-mean data vector (channel input) with its covariance matrix given by $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$, and $\mathbf{y}$ is the $n_R \times 1$ received signal vector (channel output). The $n_R \times 1$ Gaussian noise vector $\mathbf{n}$, distributed according to $\mathcal{N}_c(0, \sigma_n^2 \mathbf{I}_{n_R})$, is assumed to be zero-mean, spatially and temporally white, and independent of both data and channel fades. The channel model used here is given by $\mathbf{H} = \mathbf{R}_R^{\frac{1}{2}} \mathbf{H}_w \mathbf{R}_T^{\frac{1}{2}}$ [13], where $\mathbf{R}_T$ and $\mathbf{R}_R$ represent normalized transmit and receive correlation matrices with unit diagonal entries, respectively. The entries of $\mathbf{H}_w$ are independent and identically-distributed (i.i.d.) $\mathcal{N}_c(0,1)$.

In this paper, the imperfect channel state information is modeled in the same way as in [11].[1] Specifically, the CSIR is described by

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E}, \quad \hat{\mathbf{H}} = \mathbf{R}_R^{\frac{1}{2}} \hat{\mathbf{H}}_w \mathbf{R}_T^{\frac{1}{2}}, \quad \mathbf{E} = \mathbf{R}_R^{\frac{1}{2}} \mathbf{E}_w \mathbf{R}_T^{\frac{1}{2}} \tag{1}$$

where $\hat{\mathbf{H}}$ is the estimate of $\mathbf{H}$, and $\mathbf{E}$ is the overall channel estimation error matrix. Spatially white matrices $\hat{\mathbf{H}}_w$ and $\mathbf{E}_w$ are uncorrelated with i.i.d. entries distributed according to $\mathcal{N}_c(0, 1-\sigma_E^2)$ and $\mathcal{N}_c(0, \sigma_E^2)$, respectively. Note that σ_E^2 is the variance of channel estimation error. As in [11], [12, Sec. VI], a lossless feedback link is assumed, i.e., CSIT is the same as CSIR.[2] Thus, $\hat{\mathbf{H}}$, $\mathbf{R}_R$, $\mathbf{R}_T$, σ_E^2 and σ_n^2 represent the CSI known to both ends.

Under the above channel uncertainty model, the channel output can be written as $\mathbf{y} = \hat{\mathbf{H}}\mathbf{x} + \mathbf{E}\mathbf{x} + \mathbf{n}$. The total noise is given by $\mathbf{n}_{\mathrm{total}} = \mathbf{E}\mathbf{x} + \mathbf{n}$, with its mean being the zero vector and its covariance matrix given by

$$\begin{aligned}
\mathbf{R}_{n_{\mathrm{total}}} &= \mathbb{E}\left[(\mathbf{E}\mathbf{x}+\mathbf{n})(\mathbf{E}\mathbf{x}+\mathbf{n})^H\right] \\
&= \mathbb{E}\left[\mathbf{R}_R^{\frac{1}{2}} \mathbf{E}_w \mathbf{R}_T^{\frac{1}{2}} (\mathbf{x}\mathbf{x}^H) (\mathbf{R}_T^{\frac{1}{2}})^H \mathbf{E}_w^H (\mathbf{R}_R^{\frac{1}{2}})^H\right] + \sigma_n^2 \mathbf{I}_{n_R} \\
&= \sigma_E^2 \mathrm{tr}(\mathbf{R}_T \mathbf{Q}) \mathbf{R}_R + \sigma_n^2 \mathbf{I}_{n_R}
\end{aligned} \tag{2}$$

where the expectation is with respect to the distributions of $\mathbf{x}$, $\mathbf{n}$ and $\mathbf{E}_w$, and we have used the result

$$\mathbb{E}\left\{\mathbf{E}_w \mathbf{A} \mathbf{E}_w^H\right\} = \sigma_E^2 \mathrm{tr}(\mathbf{A})\, \mathbf{I}_{n_R}$$

if the entries of matrix $\mathbf{E}_w$ are i.i.d. $\mathcal{N}_c(0, \sigma_E^2)$. Note that $\mathbf{n}_{\mathrm{total}}$ is not Gaussian and an exact capacity expression is hard to obtain. As a result, tight upper- and lower-bounds on capacity have been proposed for system design purposes. In particular, assuming a Gaussian input distribution, though it does not necessarily achieve capacity with the above assumed CSI, the mutual information between $\mathbf{y}$ and $\mathbf{x}$ given $\hat{\mathbf{H}}$ can be bounded as [8]–[11]

$$\underline{I}_{\mathrm{low}} \le I(\mathbf{x}, \mathbf{y} | \hat{\mathbf{H}}) \le \underline{I}_{\mathrm{up}}$$

where

$$\underline{I}_{\mathrm{low}} = \log_2 \left| \mathbf{I}_{n_R} + \hat{\mathbf{H}} \mathbf{Q} \hat{\mathbf{H}}^H \mathbf{R}_{n_{\mathrm{total}}}^{-1} \right| \tag{3}$$

$$\begin{aligned}
\underline{I}_{\mathrm{up}} = {} & \underline{I}_{\mathrm{low}} + \log_2 \left|\mathbf{R}_{n_{\mathrm{total}}}\right| \\
& - \mathbb{E}\left\{\log_2 \left| \sigma_E^2 \left(\mathbf{x}^H \mathbf{R}_T \mathbf{x}\right) \mathbf{R}_R + \sigma_n^2 \mathbf{I}_{n_R} \right|\right\}.
\end{aligned} \tag{4}$$

The expectation in (4) is taken over the distribution of $\mathbf{x}$. $\underline{I}_{\mathrm{low}}$ and $\underline{I}_{\mathrm{up}}$ denote the lower- and upper-bounds on the actual maximum achievable mutual information, respectively.

B. Problem Formulation

As in [8], [10], and [11], below we adopt the capacity lower-bound as the design criterion. To obtain the highest data rate from using the capacity lower-bound, i.e., to get the best out of the worst case [24], we need to solve the following problem [10], [11]

$$I_{\mathrm{low}} = \max_{\mathbf{Q}\succeq 0,\ \mathrm{tr}\{\mathbf{Q}\} \le P_T} \log_2 \left| \mathbf{I}_{n_R} + \hat{\mathbf{H}} \mathbf{Q} \hat{\mathbf{H}}^H \mathbf{R}_{n_{\mathrm{total}}}^{-1} \right|. \tag{5}$$

The lower-bound on the ergodic capacity is then [8], [10]

$$C_{\mathrm{low}} = \mathbb{E}\left[I_{\mathrm{low}}\right] \tag{6}$$

where the expectation is taken with respect to the fading channel distribution.

Note that the short-term power constraint is imposed on the spatial domain. No temporal power allocation is considered. Since the power constraint is imposed across antennas at each

[1]We employ the current model to facilitate the comparison between our results and those in the literature. However, the model of CSI here and the one described by [12, Eq. (3)] or [16, Eqs. (3)–(5)] lead to mathematically equivalent problem formulations.

[2]While this assumption is still far from practical, it is more realistic than the one which assumes the same perfect CSI at both ends [1]–[3], [20]. A design which explicitly models and accounts for errors in the feedback, though of engineering interest, is beyond the scope of this paper.

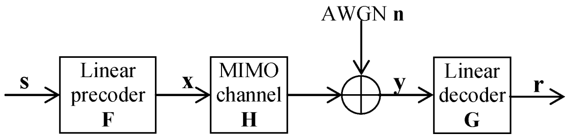

Fig. 1. Virtual auxiliary system model.

fading state (or channel realization), we then perform the maximization for each fading state, i.e., for a given $\mathbf{H}$ (or, equivalently, $\hat{\mathbf{H}}$ here). If the mutual information is maximized for each channel state, the expectation of mutual information over fading distribution is maximized. In a word, in general [26]

$$\max_{\mathbf{Q}\succeq 0,\ \mathrm{tr}(\mathbf{Q})\leq P_T} \mathbb{E}\left[I_{\mathrm{low}}\right] \leq \mathbb{E}\left[\max_{\mathbf{Q}\succeq 0,\ \mathrm{tr}(\mathbf{Q})\leq P_T} I_{\mathrm{low}}\right]$$

where the right-hand side is indeed *the* maximum which one can possibly achieve with a short-term power constraint and with the CSI specified in this paper.

The problem in (5) is considered to be a maximum mutual information design problem with imperfect channel knowledge. This explains the title of this paper.

III. Determining the Optimum Transmit Covariance Matrix for the Capacity Lower-Bound

A. Methodology

Our approach to obtaining the optimum $\mathbf{Q}$ relies on solving an equivalent problem of (5). Toward this end, we introduce a virtual auxiliary precoder-decoder pair $(\mathbf{F}, \mathbf{G})$ in our system model (see Fig. 1), where $\mathbf{F}$ and $\mathbf{G}$ are $n_T \times r_g$ and $r_g \times n_R$ matrices, respectively, and $r_g = \mathrm{rank}(\hat{\mathbf{H}})$. The choice of the sizes of $\mathbf{F}$ and $\mathbf{G}$ is explained in Part 3) of Appendix B. Let $\mathbf{s}$ be a zero-mean $r_g \times 1$ data vector whose entries are i.i.d. with unit variance. Then $\mathbf{x} = \mathbf{F}\mathbf{s}$, and $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H) = \mathbf{F}\mathbf{F}^H$. From Fig. 1,

$$\mathbf{y} = \mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{n} = \hat{\mathbf{H}}\mathbf{F}\mathbf{s} + \mathbf{E}\mathbf{F}\mathbf{s} + \mathbf{n}.$$

The received vector after the decoder is given by $\mathbf{r} = \mathbf{G}\mathbf{y}$. Define the MSE matrix $\mathrm{MSE}(\mathbf{F}, \mathbf{G})$ as

$$\begin{aligned}\mathrm{MSE}(\mathbf{F},\mathbf{G}) &\stackrel{\mathrm{def}}{=} \mathbb{E}\left[(\mathbf{r}-\mathbf{s})(\mathbf{r}-\mathbf{s})^H\right]\\ &= \mathbb{E}\left\{\left[\mathbf{G}(\hat{\mathbf{H}}+\mathbf{E})\mathbf{F}-\mathbf{I}_{r_g}\right]\mathbf{s}\,\mathbf{s}^H \right.\\ &\quad\left.\times\left[\mathbf{G}(\hat{\mathbf{H}}+\mathbf{E})\mathbf{F}-\mathbf{I}_{r_g}\right]^H\right\} + \sigma_n^2\mathbf{G}\mathbf{G}^H\\ &= \mathbf{G}\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H - \mathbf{G}\hat{\mathbf{H}}\mathbf{F} - \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H + \mathbf{I}_{r_g}\\ &\quad + \mathbf{G}\underbrace{\left[\sigma_E^2\mathrm{tr}\left(\mathbf{R}_T\mathbf{F}\mathbf{F}^H\right)\mathbf{R}_R + \sigma_n^2\mathbf{I}_{n_R}\right]}_{\mathbf{R}_{n_{\mathrm{total}}}}\mathbf{G}^H. \end{aligned} \quad (7)$$

In the above, the expectation is with respect to $\mathbf{s}$, $\mathbf{n}$ and $\mathbf{E}_w$.

Lemma 1: The problem in (5) is equivalent to

$$\min_{\mathbf{F},\,\mathbf{G},\,\mathrm{tr}\{\mathbf{F}\mathbf{F}^H\}\leq P_T} \ln\left|\mathrm{MSE}(\mathbf{F},\mathbf{G})\right| \quad (8)$$

where $\ln$ denotes the natural logarithm. Denote the optimum solution for (8) as $(\mathbf{F}_{g_{\mathrm{opt}}}, \mathbf{G}_{g_{\mathrm{opt}}})$. Then $\mathbf{Q}_{g_{\mathrm{opt}}}$ for (5) is related to the optimum solution for (8) by $\mathbf{Q}_{g_{\mathrm{opt}}} = \mathbf{F}_{g_{\mathrm{opt}}}\mathbf{F}_{g_{\mathrm{opt}}}^H$. Furthermore, a global maximum exists for (5) and a global minimum exists for (8).

Proof: See Appendix A.

Remark 1: Based on **Lemma 1**, in order to solve (5), we attempt to solve (8) instead. A similar line of thinking can be found in [2], [14] for the perfect CSI case. Nevertheless, this idea has not been exploited to solve problems associated with imperfect CSI. With imperfect CSI, while the minimum total MSE design in [16] minimizes the trace of the MSE matrix, the maximum mutual information design minimizes its log determinant, which is the capacity lower-bound achievable with a Gaussian input distribution. This is analogous to their relation in the perfect CSI case [2].

Upon reaching this point, we take an approach similar to that in [16] to solve (8), a nonconvex problem.[3] Since the objective and constraint functions of (8) are continuously differentiable with respect to $(\mathbf{F}^*, \mathbf{G}^*)$ (or, equivalently, $(\mathbf{F}, \mathbf{G})$), and the feasible points of (8) satisfy the regularity condition [19, pp. 309–310], the global minimum, which exists per **Lemma 1**, should satisfy the first-order KKT necessary conditions associated with (8) [19, p. 310, Prop. 3.3.1]. Our method is to obtain all the solutions satisfying the KKT necessary conditions, compare the object values induced by them, and then single out the optimum $(\mathbf{F}_{g_{\mathrm{opt}}}, \mathbf{G}_{g_{\mathrm{opt}}})$ pairs from them.[4] Note that if $(\mathbf{F}_{g_{\mathrm{opt}}}, \mathbf{G}_{g_{\mathrm{opt}}})$ is optimum, so is $(\mathbf{F}_{g_{\mathrm{opt}}}\mathbf{U}, \mathbf{U}^H\mathbf{G}_{g_{\mathrm{opt}}})$, where $\mathbf{U}$ is an arbitrary $r_g \times r_g$ unitary matrix. Nevertheless, the optimum transmit covariance matrix $\mathbf{Q}_{g_{\mathrm{opt}}} = \mathbf{F}_{g_{\mathrm{opt}}}\mathbf{F}_{g_{\mathrm{opt}}}^H$ for this group of $(\mathbf{F}_{g_{\mathrm{opt}}}, \mathbf{G}_{g_{\mathrm{opt}}})$ pairs is unique. Below we will refer to $(\mathbf{F}_{g_{\mathrm{opt}}}, \mathbf{G}_{g_{\mathrm{opt}}})$ as an optimum solution for (8) up to a unitary transform.

B. General Results: $\mathbf{R}_T \neq \mathbf{I}_{n_T}$ and $\mathbf{R}_R \neq \mathbf{I}_{n_R}$

Theorem 1: With the MSE matrix given by (7), the optimal structures of the precoder and decoder for (8) (up to a unitary transform) are given by

$$\mathbf{F}_{g_{\mathrm{opt}}} = \left[\sigma_E^2\alpha_g\mathbf{R}_T + \mu_g\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\mathbf{V}_g\mathbf{\Phi}_{Fgo} \quad (9)$$

$$\begin{aligned}\mathbf{G}_{g_{\mathrm{opt}}} =& \mathbf{\Phi}_{Ggo}\mathbf{V}_g^H\left[\sigma_E^2\alpha_g\mathbf{R}_T + \mu_g\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\\ &\times\hat{\mathbf{H}}^H\left[\sigma_E^2\mathrm{tr}(\mathbf{R}_T\mathbf{F}_{g_{\mathrm{opt}}}\mathbf{F}_{g_{\mathrm{opt}}}^H)\mathbf{R}_R + \sigma_n^2\mathbf{I}_{n_R}\right]^{-1}\end{aligned} \quad (10)$$

respectively, where

$$\mathbf{\Phi}_{Fgo} = \left[\mathbf{I}_{r_g} - \mathbf{\Lambda}_g^{-1}\right]_+^{\frac{1}{2}} \quad (11)$$

$$\mathbf{\Phi}_{Ggo} = \left[\mathbf{I}_{r_g} - \mathbf{\Lambda}_g^{-1}\right]_+^{\frac{1}{2}}\mathbf{\Lambda}_g^{-1} \quad (12)$$

[3]An easy way to check this is to consider the special case when all matrices are one-by-one (scalars).

[4]A simple example using this method can be found in [19, p. 11].

and

$$\alpha_g = \mathrm{tr}\left\{\mathbf{G}_{g_{\rm opt}}\mathbf{R}_R\left[\sigma_E^2\mathrm{tr}(\mathbf{R}_T\mathbf{F}_{g_{\rm opt}}\mathbf{F}_{g_{\rm opt}}^H)\mathbf{R}_R+\sigma_n^2\mathbf{I}_{n_R}\right]^{-1} \times\hat{\mathbf{H}}\mathbf{F}_{g_{\rm opt}}\right\}, \quad (13)$$

$$\mu_g = \frac{\sigma_n^2}{P_T}\mathrm{tr}\left\{\mathbf{G}_{g_{\rm opt}}\left[\sigma_E^2\mathrm{tr}(\mathbf{R}_T\mathbf{F}_{g_{\rm opt}}\mathbf{F}_{g_{\rm opt}}^H)\mathbf{R}_R+\sigma_n^2\mathbf{I}_{n_R}\right]^{-1} \times\hat{\mathbf{H}}\mathbf{F}_{g_{\rm opt}}\right\}. \quad (14)$$

Matrices $\mathbf{V}_g$ and $\mathbf{\Lambda}_g$ are defined by the following eigenvalue decomposition

$$\left[\sigma_E^2\alpha_g\mathbf{R}_T+\mu_g\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\hat{\mathbf{H}}^H \times\left[\sigma_E^2\mathrm{tr}\left(\mathbf{R}_T\mathbf{F}\mathbf{F}^H\right)\mathbf{R}_R+\sigma_n^2\mathbf{I}_{n_R}\right]^{-1} \times\hat{\mathbf{H}}\left[\sigma_E^2\alpha_g\mathbf{R}_T+\mu_g\mathbf{I}_{n_T}\right]^{-\frac{1}{2}} = \left[\mathbf{V}_g\tilde{\mathbf{V}}_g\right]\begin{pmatrix}\mathbf{\Lambda}_g & 0\\ 0 & \tilde{\mathbf{\Lambda}}_g\end{pmatrix}\left[\mathbf{V}_g\tilde{\mathbf{V}}_g\right]^H \quad (15)$$

(in which $\mathbf{F}$ is replaced by $\mathbf{F}_{g_{\rm opt}}$), and $r_g = \mathrm{rank}(\hat{\mathbf{H}}) = \mathrm{rank}(\mathbf{\Lambda}_g)$. The $n_T \times (n_T - r_g)$ matrix $\tilde{\mathbf{V}}_g$ consists of basis vectors for the null space of (15). The entries of the diagonal matrix $\tilde{\mathbf{\Lambda}}_g$ are all zero. The $n_T \times r_g$ matrix $\mathbf{V}_g$ is composed of the eigenvectors corresponding to the nonzero eigenvalues. Without loss of generality, the entries of the diagonal matrix $\mathbf{\Lambda}_g$ are arranged in nonincreasing order. By inserting (9) and (10) into (13) and (14), two equations can be obtained with μ_g and α_g being the only two unknowns, which can be determined numerically.

Proof: See Appendix B.

Theorem 1 clearly describes the structures of the optimal precoder and decoder. However, μ_g and α_g need to be determined numerically. Inspired by [15], we provide a KKT-conditions-based iterative algorithm to determine the optimum solution as given in Table I. This algorithm has the property of reducing the value of the objective function at each iteration. Based on a large number of simulations, starting from a nonzero matrix $\mathbf{F}_0$, the algorithm in Table I yields the optimum $(\mathbf{F}_{g_{\rm opt}}, \mathbf{G}_{g_{\rm opt}})$ up to a unitary transform. Unfortunately, a rigorous convergence analysis of this iterative algorithm has not been obtained to date, but represents a subject for future investigation.

The following corollary follows immediately from **Lemma 1** and **Theorem 1**.

Corollary 1: The unique optimum covariance matrix for (5) is given by $\mathbf{Q}_{g_{\rm opt}} = \mathbf{F}_{g_{\rm opt}}\mathbf{F}_{g_{\rm opt}}^H$, where $\mathbf{F}_{g_{\rm opt}}$ is from **Theorem 1**.

C. Special Case: $\mathbf{R}_T \neq \mathbf{I}_{n_T}$ and $\mathbf{R}_R = \mathbf{I}_{n_R}$

In MIMO downlink transmissions, base stations are usually located at high elevation with relatively few local scatterers around. As a result, the channels arising from the BS antennas (transmitter) are often correlated. In contrast, the channels arising from the mobile station antennas (receiver) are often uncorrelated due to sufficient local scattering [10], [23]. Thus, we consider (5) in this special case. It turns out that the closed-form optimum transmit covariance matrix can be found, unlike in the general case where undetermined parameters appear in the eigenvalue decomposition [see (15)].

TABLE I
ITERATIVE ALGORITHM FOR SOLVING (8) IN THE GENERAL CASE

1)	Initialize $\mathbf{F} = \mathbf{F}_0$; the upper $r_g \times r_g$ sub-matrix of $\mathbf{F}_0$ is chosen to be a scaled identity and to satisfy the power constraint with equality, while the remaining entries of $\mathbf{F}_0$ are set to zero.
2)	Update $\mathbf{R}_{n_{total}}$ using (2), with $\mathbf{Q} = \mathbf{F}\mathbf{F}^H$; update $\mathbf{G}$ using (29), i.e., $\mathbf{G} = \mathbf{F}^H\hat{\mathbf{H}}^H[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \mathbf{R}_{n_{total}}]^{-1}$.
3)	Update μ_g and α_g using (37) and (33), respectively;
4)	Update $\mathbf{F}$ using (34), i.e., $\mathbf{F} = \left[\mu_g\mathbf{I} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-1}\hat{\mathbf{H}}^H\mathbf{G}^H$. Scale $\mathbf{F}$ if needed, such that $\mathrm{tr}(\mathbf{F}\mathbf{F}^H) = P_T$.
5)	If $\|\mathbf{F}_i - \mathbf{F}_{i-1}\|_F \leq \epsilon$ [e.g., $\epsilon = 0.01$], stop; otherwise, go back to 2). Here $\mathbf{F}_i$ denotes $\mathbf{F}$ in the i-th iteration.

Corollary 2: For the case with $\mathbf{R}_R = \mathbf{I}_{n_R}$, the *closed-form* optimal precoder and decoder for (8) are given by

$$\mathbf{F}_{\rm opt} = \left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\mathbf{V}\mathbf{\Phi}_{Fo} \quad (16)$$

$$\mathbf{G}_{\rm opt} = \mathbf{\Phi}_{Go}\mathbf{V}^H\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\hat{\mathbf{H}}^H \quad (17)$$

respectively, where

$$\mathbf{\Phi}_{Fo} = \left(\mu/\sigma_n^2\right)^{-\frac{1}{2}}\left[\mathbf{I}_r - \left[(\mu\beta)/\sigma_n^2\right]\mathbf{\Lambda}^{-1}\right]_+^{\frac{1}{2}} \quad (18)$$

$$\mathbf{\Phi}_{Go} = \left(\mu/\sigma_n^2\right)^{\frac{1}{2}}\left[\mathbf{I}_r - \left[(\mu\beta)/\sigma_n^2\right]\mathbf{\Lambda}^{-1}\right]_+^{\frac{1}{2}}\mathbf{\Lambda}^{-1}. \quad (19)$$

Matrices $\mathbf{V}$ and $\mathbf{\Lambda}$ are obtained from the following eigenvalue decomposition

$$\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \quad (20)$$

and r $(= \mathrm{rank}(\hat{\mathbf{H}}) = \mathrm{rank}(\mathbf{\Lambda}))$ denotes the number of nonzero channel eigenmodes. $\mathbf{\Lambda}$ is the $r \times r$ diagonal matrix whose entries are the nonzero eigenvalues arranged in nonincreasing order. The $n_T \times r$ matrix $\mathbf{V}$ is composed of the eigenvectors corresponding to the nonzero eigenvalues. Scalars μ and β are given by

$$\mu = \frac{\sigma_n^2(b_1P_T + b_1b_3 - mb_2)}{P_T(P_T + b_3)} \quad (21)$$

$$\beta = \frac{mP_T}{b_1P_T + b_1b_3 - mb_2} \quad (22)$$

where the integer m $(m \leq r)$ denotes the number of the nonzero entries of the diagonal matrix $\mathbf{\Phi}_{Fo}$. b_1, b_2 and b_3 are traces of the $m \times m$ top-left sub-matrices of $\mathbf{V}^H\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-1}\mathbf{V}$, $\mathbf{\Lambda}^{-1}\mathbf{V}^H\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-1}\mathbf{V}$, and $\mathbf{\Lambda}^{-1}$, respectively. The optimum solution is unique up to a unitary transform. The unique, optimum covariance matrix is determined using $\mathbf{Q}_{\rm opt} = \mathbf{F}_{\rm opt}\mathbf{F}_{\rm opt}^H$.

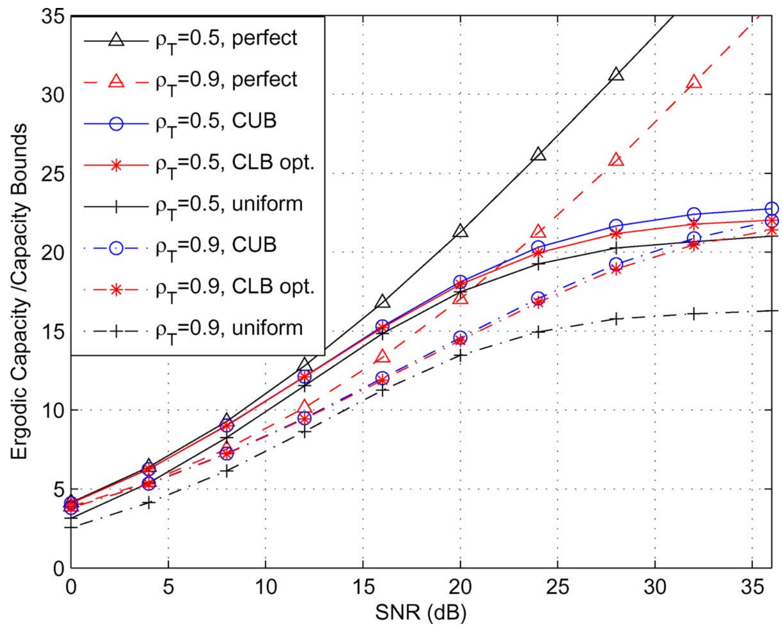

Fig. 2. Ergodic capacity or capacity bounds, with different amounts of CSI and with different transmit strategies. $\rho_R = 0$ for all curves here. $\sigma_E^2 = 0.01$ for the case with imperfect CSI. "CUB" stands for the capacity upper-bound, and "CLB" for the lower-bound. Those marked with "uniform" are the lower-bounds obtained using a uniform power allocation at the transmitter under imperfect CSI.

Proof: Here μ is the Lagrangian associated with the optimization problem for the special case, and

$$\beta \stackrel{\text{def}}{=} \sigma_n^2 + \sigma_E^2 \text{tr}\left(\mathbf{R}_T \mathbf{F}\mathbf{F}^H\right).$$

The main procedure is similar to the proof for **Theorem 1**. Therefore, details are omitted for brevity. By substituting (16)–(19) into the expression for β and the power constraint $\text{tr}\left\{\mathbf{F}\mathbf{F}^H\right\} = P_T$, two equations with μ and β being the only two unknowns can be obtained, from which we finally obtain (21) and (22). The method to determine m is included in Appendix C.

Remark 2: When $\sigma_E^2 = 0$, **Corollary 2** reduces to the capacity results obtained in [1]–[3], and [20].

Remark 3: In [10, Th. 5], it has been pointed out that when P_T/σ_n^2 goes to infinity, the optimum transmit strategy is to perform a water-filling procedure over the eigenmodes of $\hat{\mathbf{H}}_w$ and then invert the effect of $\mathbf{R}_T$. This agrees with **Corollary 2**.

Remark 4: With the same imperfect CSI considered here (i.e., transmit correlation only), the optimal transmitters for the maximum mutual information design and the minimum total MSE design share the same structure, and differ only in the power allocations [16], [17].

IV. Numerical Results

The transmit correlation model is given by [25, eq. (13)],[10]: $(\mathbf{R}_T)_{ij} = \rho_T^{|i-j|}$ for $i, j \in \{1, \ldots, n_T\}$. The receive correlation matrix $\mathbf{R}_R$ is similarly defined with the exception that ρ_T is replaced by ρ_R and that the indices range from 1 to n_R. Four antennas are employed at each end of the channel. Below the ergodic capacity and the ergodic capacity bounds will be shown in bits per MIMO channel use (or bits per MIMO transmission), which are calculated by averaging the instantaneous mutual information over the fading distribution (similar to (6)). The signal-to-noise ratio (SNR) here is defined as P_T/σ_n^2.

To obtain the optimum transmit covariance matrix for the lower-bound, in the case without receive correlation (i.e., $\mathbf{R}_R = \mathbf{I}_{n_R}$), the closed-form result in Section III-C is applied. When there is both transmit and receive correlation, we use the algorithm in Table I.

A. Comparison of the Ergodic Capacity and Different Capacity Bounds

Fig. 2 shows a comparison between the ergodic capacity and upper- and lower-bounds of capacity. For each of two different channel spatial correlation values, the bounds (marked with "CUB" or "CLB") are calculated using (4) and (3), respectively, and then averaged over the fading distribution. The optimum covariance matrix for the lower-bound derived in this paper is used as the transmit strategy in the calculations of (4) and (3). The two bounds obtained this way are seen to be very close to each other, especially in the low to medium SNR region (≤ 20 dB), and thus are both tight in our cases. This justifies the use of the lower-bound as a design criterion to maximize the mutual information. Asymptotic analysis of the difference between the upper- and lower-bounds can be found in [8], [11].

When there is no channel estimation error (perfect CSI at both ends), the optimum strategy is to transmit along the nonzero channel eigenmodes and then perform a water-filling type of power allocation among the channel eigenmodes. The uniform power allocation scheme simply ignores the CSIT and allocates the same power to all antennas. In the case of imperfect CSI considered here, the optimum transmit strategy refers to that for the capacity lower-bound. The case with imperfect CSI is shown in Fig. 2 (those two curves marked with "uniform"). As seen in

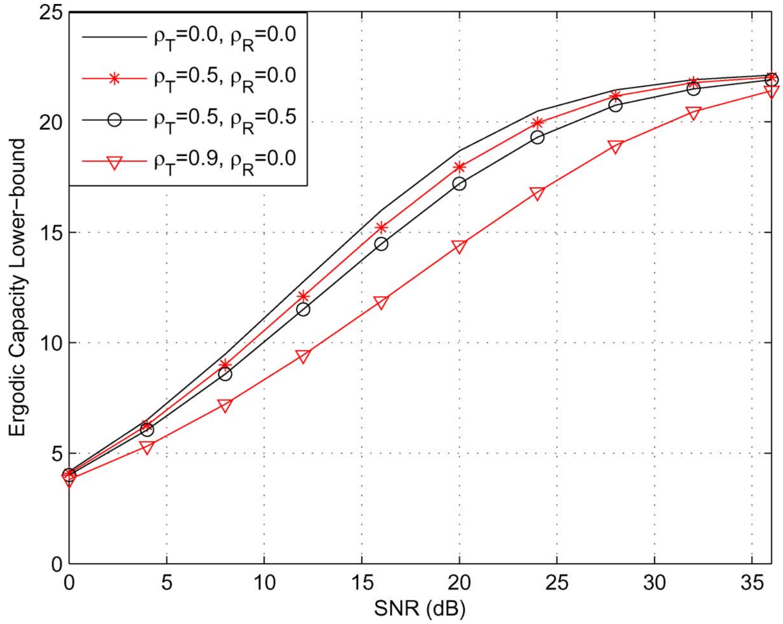

Fig. 3. Ergodic capacity lower-bound of MIMO channels for the case with imperfect CSI ($\sigma_E^2 = 0.01$).

Fig. 2, with imperfect CSI, the uniform power allocation (used in [11]) performs strictly worse over the entire SNR range when the channel correlation is high. This, in turn, shows the advantage (in terms of accuracy) of using the optimum covariance matrix for the lower-bound. From Fig. 2, this advantage increases with the amount of channel correlation.

B. Effects of Channel Estimation Error and Channel Correlation

Fig. 3 shows more curves of the capacity lower-bound with imperfect CSI using the optimum covariance matrix. Channel estimation error causes a significant loss in ergodic capacity. From Figs. 2 and 3, at high SNR the capacity saturates due to channel estimation error. Channel correlation can also significantly reduce the ergodic capacity, based on Figs. 2 and 3.

C. Comparison of the Effects of Transmit and Receive Correlation

In [11], based on the capacity lower-bound derived from the suboptimum uniform transmit power allocation, the effects of the same amount of transmit and receive correlation are shown to be different (see [11, Figs. 4 and 5]). In fact, by using the closed-form optimal transmit covariance matrices for the cases with transmit correlation only and with receive correlation only, we can show that, when $n_T = n_R$ and $\mathbf{R}_T = \mathbf{R}_R$, the effective eigenchannels and the optimal MSE matrices in both cases are identical. This implies that the same amount of transmit and receive correlation should have the same effect on the optimum capacity lower-bound, as shown in Fig. 4.

V. Conclusions

We have used a tight capacity lower-bound as a design criterion to maximize the mutual information of a MIMO channel with imperfect channel knowledge. The expression of the optimum transmit covariance matrix (or the precoder matrix) for this lower-bound has been determined, which clearly gives the transmitter structure and is shown to be advantageous over the suboptimum uniform power allocation scheme. The effects of channel estimation error and channel correlation have also been assessed.

Appendix A
Proof of Lemma 1

Proof: We first equivalently formulate (8) as [18, p. 130, Sec. 4.1.3]

$$\min_{\mathbf{F},\ \mathrm{tr}(\mathbf{F}\mathbf{F}^H)\le P_T}\ \min_{\mathbf{G}(\mathbf{F})} \ln|\mathrm{MSE}[\mathbf{F},\mathbf{G}(\mathbf{F})]| \tag{23}$$

where $\ln$ denotes the natural logarithm. It can be shown that the inner minimization is given by (see, for example, [20, Sec. IV-A]): $\mathbf{G} = \mathbf{F}^H\hat{\mathbf{H}}^H[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \mathbf{R}_{n_{\mathrm{total}}}]^{-1}$, where $\mathbf{R}_{n_{\mathrm{total}}}$ is defined in (2). Note that this optimum $\mathbf{G}$ is the same as that from (29) in Appendix B, derived using a different method. Substituting this formula into the third equation in (7), using the matrix inversion lemma [22], following the same steps as (29)–(32) in Appendix B, we can see that (8) is equivalent to

$$\min_{\mathbf{F},\ \mathrm{tr}(\mathbf{F}\mathbf{F}^H)\le P_T} -\ln\left|\mathbf{I}_{n_R} + \hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{R}_{n_{\mathrm{total}}}^{-1}\right| \tag{24}$$

where we have used the identity $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$ [22]. Let $\mathbf{Q} = \mathbf{F}\mathbf{F}^H$, and then the equivalence between (24) and (5) becomes clear.

The problem (24) has a compact feasible set [16], while its objective function is continuous at all points of the feasible set. Thus, by Weierstrass' Theorem [19, p. 654, Prop. A.8], a global minimum exists for (24). By equivalence [18, p. 130, Sec. 4.1.3], the same global maximum exists for (8). Again, by equivalence, a global maximum exists for (5) and the maximizing $\mathbf{Q}_{g_{\mathrm{opt}}}$ for (5) is related to the minimizing $\mathbf{F}_{g_{\mathrm{opt}}}$ for (24) or (8) through $\mathbf{Q}_{g_{\mathrm{opt}}} = \mathbf{F}_{g_{\mathrm{opt}}}\mathbf{F}_{g_{\mathrm{opt}}}^H$. This concludes the proof of **Lemma 1**.

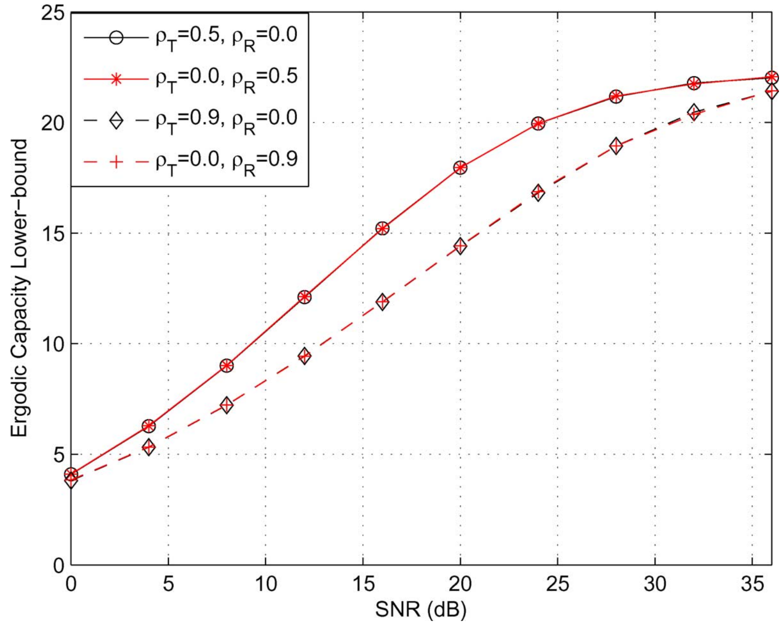

Fig. 4. Effect of transmit correlation versus that of receive correlation for the case of imperfect CSI ($\sigma_E^2 = 0.01$).

Appendix B
Proof of Theorem 1

1) The Lagrangian, KKT Conditions, and Simplifications: The Lagrangian associated with (8) is

$$\mathcal{L}_c(\mathbf{F}, \mathbf{G}, \mu) = \ln|\mathrm{MSE}(\mathbf{F}, \mathbf{G})| + \mu_g[\mathrm{tr}(\mathbf{F}\mathbf{F}^H) - P_T]$$

where μ_g is the Lagrange multiplier. By taking the derivative of the Lagrangian with respect to $(\mathbf{F}^*, \mathbf{G}^*)$ [21], [22], respectively, we obtain the KKT necessary conditions for (8)

$$[\mathrm{MSE}(\mathbf{F}, \mathbf{G})]^{-1}[\mathbf{G}\mathbf{H}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H - \mathbf{F}^H\hat{\mathbf{H}}^H + \mathbf{G}\mathbf{R}_{n_{\mathrm{total}}}] = 0 \tag{25}$$

$$\hat{\mathbf{H}}^H\mathbf{G}^H[\mathrm{MSE}(\mathbf{F}, \mathbf{G})]^{-1}(\mathbf{G}\hat{\mathbf{H}}\mathbf{F} - \mathbf{I}_{r_g}) + \sigma_E^2\mathrm{tr}\left\{[\mathrm{MSE}(\mathbf{F}, \mathbf{G})]^{-1}\mathbf{G}\mathbf{R}_R\mathbf{G}^H\right\}\mathbf{R}_T\mathbf{F} + \mu_g\mathbf{F} = 0 \tag{26}$$

$$\mu_g \geq 0, \mathrm{tr}(\mathbf{F}\mathbf{F}^H) - P_T \leq 0 \tag{27}$$

$$\mu_g\left[\mathrm{tr}(\mathbf{F}\mathbf{F}^H) - P_T\right] = 0 \tag{28}$$

where $\mathbf{R}_{n_{\mathrm{total}}}$, being used throughout Appendix B, is the same as given in (2) except that there $\mathbf{Q}$ is replaced by $\mathbf{F}\mathbf{F}^H$.

An obvious solution that satisfies the above conditions is $\mathbf{F} = 0, \mathbf{G} = 0, \mu_g = 0$, which will be compared to other solutions presented later. For now, we consider the case where $\mathbf{F} \neq 0$.

Since (8) is assumed to be well-defined, $|\mathrm{MSE}(\mathbf{F}, \mathbf{G})|$ should be larger than 0, and then (25) is equivalent to

$$\mathbf{F}^H\hat{\mathbf{H}}^H = \mathbf{G}[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \mathbf{R}_{n_{\mathrm{total}}}]. \tag{29}$$

Based on (29), by writing $\mathbf{G}$ as a function of $\mathbf{F}$, and after some algebra, we obtain

$$\mathrm{MSE}(\mathbf{F}, \mathbf{G}) = \mathbf{I}_{r_g} - \mathbf{G}\hat{\mathbf{H}}\mathbf{F}. \tag{30}$$

Again based on (29), using the matrix inversion lemma [22], we can show that

$$\mathbf{G} = [\mathbf{I} + \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{R}_{n_{\mathrm{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}]^{-1}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{R}_{n_{\mathrm{total}}}^{-1} \tag{31}$$

while

$$\mathrm{MSE}(\mathbf{F}, \mathbf{G}) = [\mathbf{I} + \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{R}_{n_{\mathrm{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}]^{-1} \tag{32}$$

and

$$\alpha_g \stackrel{\mathrm{def}}{=} \mathrm{tr}\left\{[\mathrm{MSE}(\mathbf{F}, \mathbf{G})]^{-1}\mathbf{G}\mathbf{R}_R\mathbf{G}^H\right\} = \mathrm{tr}\left\{\mathbf{G}\mathbf{R}_R\mathbf{R}_{n_{\mathrm{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}\right\}. \tag{33}$$

Substituting (30) and (33) into (26), we obtain

$$\hat{\mathbf{H}}^H\mathbf{G}^H = \left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]\mathbf{F}. \tag{34}$$

From now on, we will focus on (34), (29), (27), and (28).

2) Relation Between μ_g and $(\mathbf{F}, \mathbf{G})$: From (33), using the fact that

$$\mathbf{R}_R = \left[\mathbf{R}_{n_{\mathrm{total}}} - \sigma_n^2\mathbf{I}_{n_R}\right] / \left[\sigma_E^2\mathrm{tr}\left(\mathbf{R}_T\mathbf{F}\mathbf{F}^H\right)\right]$$

we can show that

$$\sigma_E^2\alpha_g\mathrm{tr}\left(\mathbf{R}_T\mathbf{F}\mathbf{F}^H\right) = \mathrm{tr}(\mathbf{G}\hat{\mathbf{H}}\mathbf{F}) - \sigma_n^2\mathrm{tr}(\mathbf{G}\mathbf{R}_{n_{\mathrm{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}). \tag{35}$$

Premultiplying both sides of (34) by $\mathbf{F}^H$, we have

$$\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H = \mathbf{F}^H\left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]\mathbf{F}. \tag{36}$$

Then taking the trace of both sides, using (35) and the fact that $\mathbf{G}\hat{\mathbf{H}}\mathbf{F}$ is Hermitian, we obtain

$$\mu_g\mathrm{tr}\left(\mathbf{F}\mathbf{F}^H\right) = \sigma_n^2\mathrm{tr}\left(\mathbf{G}\mathbf{R}_{n_{\mathrm{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}\right).$$

Due to (28), if $\mu_g > 0$, $\mathrm{tr}(\mathbf{F}\mathbf{F}^H)$ must be equal to P_T, which yields

$$\mu_g = \frac{\sigma_n^2}{P_T}\,\mathrm{tr}(\mathbf{G}\mathbf{R}_{n_{\text{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}). \tag{37}$$

Assume that $\mathbf{F} \neq 0$ (i.e., $\mathrm{tr}(\mathbf{F}\mathbf{F}^H) > 0$). Then if $\mu_g = 0$, we must have $\mathrm{tr}(\mathbf{G}\mathbf{R}_{n_{\text{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}) = 0$, and (37) is still valid. Furthermore, (37) holds even when $\mathbf{F} = 0$, $\mathbf{G} = 0$, and $\mu_g = 0$.

3) Determining the Optimal Structures of $\mathbf{F}$ *and* $\mathbf{G}$*:* Parallel to the calculations in [16], we first write $\mathbf{F}$ in a general form[5]

$$\begin{aligned}\mathbf{F} &= \left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-\frac{1}{2}}\left[\mathbf{V}_g\tilde{\mathbf{V}}_g\right]\left[\boldsymbol{\Phi}_F^H\tilde{\boldsymbol{\Phi}}_F^H\right]^H \\ &= \left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-\frac{1}{2}}\left[\underbrace{\mathbf{V}_g\boldsymbol{\Phi}_F}_{\mathbf{F}_{\|}} + \underbrace{\tilde{\mathbf{V}}_g\tilde{\boldsymbol{\Phi}}_F}_{\mathbf{F}_{\perp}}\right]\end{aligned} \tag{38}$$

where $\mathbf{V}_g$ and $\tilde{\mathbf{V}}_g$ are defined in (15), and $\boldsymbol{\Phi}_F$ and $\tilde{\boldsymbol{\Phi}}_F$ are arbitrary $r_g \times r_g$ and $(n_T - r_g) \times r_g$ matrices, respectively. It can be verified that

$$\mathbf{F}_{\|}^H\mathbf{F}_{\perp} = 0, \mathbf{F}_{\perp}^H\mathbf{F}_{\|} = 0, \hat{\mathbf{H}}\left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-\frac{1}{2}}\mathbf{F}_{\perp} = 0. \tag{39}$$

Substituting (38) into (36), using (39), after some calculations, we can show that $\mathbf{F}_{\perp}^H\mathbf{F}_{\perp} = 0$. Therefore, without loss of generality, here $\mathbf{F}$ can be expressed as

$$\mathbf{F} = \left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-\frac{1}{2}}\mathbf{V}_g\boldsymbol{\Phi}_F. \tag{40}$$

Based on (31) and (40), using the matrix inversion lemma [22]

$$\begin{aligned}\mathbf{G} &= [\mathbf{I}_{r_g} + \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{R}_{n_{\text{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}]^{-1}\boldsymbol{\Phi}_F^H\mathbf{V}_g^H \\ &\quad\times\left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-\frac{1}{2}}\hat{\mathbf{H}}^H\mathbf{R}_{n_{\text{total}}}^{-1} \\ &= \boldsymbol{\Phi}_G\mathbf{V}_g^H\left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]^{-\frac{1}{2}}\hat{\mathbf{H}}^H\mathbf{R}_{n_{\text{total}}}^{-1}\end{aligned} \tag{41}$$

where $\boldsymbol{\Phi}_G \overset{\text{def}}{=} [\mathbf{I}_{r_g} + \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{R}_{n_{\text{total}}}^{-1}\hat{\mathbf{H}}\mathbf{F}]^{-1}\boldsymbol{\Phi}_F^H$, also an arbitrary matrix due to $\boldsymbol{\Phi}_F$.

At this point, one might argue that the sizes of $\boldsymbol{\Phi}_F$ and $\boldsymbol{\Phi}_G$ should be $r_g \times B$ and $B \times r_g$ matrices, respectively, where B should be allowed to take any integer value satisfying $r_g \leq B \leq n_T$. Since our ultimate goal is to determine $\mathbf{Q}$, which is given by $\mathbf{Q} = \mathbf{F}\mathbf{F}^H$, the quantity of interest is the $r_g \times r_g$ matrix $\boldsymbol{\Phi}_F\boldsymbol{\Phi}_F^H$. We can then safely choose $B = r_g$ without incurring any loss of generality for our problem. This explains our earlier choice of $\mathbf{F}$ as a $n_T \times r_g$ matrix.

Assume $\mathbf{F} \neq 0$. Post-multiply both sides of (29) by $\mathbf{G}^H$, and then we obtain

$$\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H = \mathbf{G}\left[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \mathbf{R}_{n_{\text{total}}}\right]\mathbf{G}^H. \tag{42}$$

[5]Our proof here is similar but not identical to that in [3]. Since $\left[\mu_g\mathbf{I}_{n_T} + \alpha_g\sigma_E^2\mathbf{R}_T\right]$ is a nonsingular matrix, (38) is general.

Substituting (40) and (41) into (36) and (42), using (15), the following two formulas hold:

$$\boldsymbol{\Phi}_G\boldsymbol{\Lambda}_g\boldsymbol{\Phi}_F = \boldsymbol{\Phi}_G\boldsymbol{\Lambda}_g\boldsymbol{\Phi}_F\boldsymbol{\Phi}_F^H\boldsymbol{\Lambda}_g\boldsymbol{\Phi}_G^H + \boldsymbol{\Phi}_G\boldsymbol{\Lambda}_g\boldsymbol{\Phi}_G^H \tag{43}$$

$$\boldsymbol{\Phi}_G\boldsymbol{\Lambda}_g\boldsymbol{\Phi}_F = \boldsymbol{\Phi}_F^H\boldsymbol{\Phi}_F. \tag{44}$$

Based on (43)-(44), using the same method as in [3, App. I, B and C], we obtain

$$\boldsymbol{\Phi}_F = \left[\mathbf{I}_{r_g} - \boldsymbol{\Lambda}_g^{-1}\right]_+^{\frac{1}{2}} \tag{45}$$

$$\boldsymbol{\Phi}_G = \left[\mathbf{I}_{r_g} - \boldsymbol{\Lambda}_g^{-1}\right]_+^{\frac{1}{2}}\boldsymbol{\Lambda}_g^{-1}. \tag{46}$$

Based on (40), (41), (45), and (46), we have proved (9) and (10). Equations (13) and (14) follow from (33) and (37), respectively. In addition, the specific values of α_g and μ_g need to be determined numerically.

Up to now, we have shown that the nonzero solutions satisfying the KKT conditions (25)–(28) with $\mathbf{F} \neq 0$ are given by (9)–(14), up to a unitary transform. By direct calculations, all these nonzero solutions lead to the same value of the objective function in (8), which is lower than that from using ($\mathbf{F} = 0, \mathbf{G} = 0, \mu = 0$). Therefore, we conclude that they are equivalently optimum and **Theorem 1** holds.

Appendix C
Determining m in Corollary 2

In order to determine the number m in the expressions of (21) and (22), we use an iterative procedure. Let $\boldsymbol{\lambda}_m$ be the mth element of $\boldsymbol{\Lambda}$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$. Initialize $m = r$.

1) Calculate β from (22) and μ from (21). If $\mu \leq \boldsymbol{\lambda}_m\sigma_n^2/\beta$, stop; else: go to step 2).
2) Let $\boldsymbol{\Phi}_{Fo,m} := 0$ and $m := m - 1$. Go to step 1).

Acknowledgment

The authors would like to thank the anonymous reviewers and the editor for their helpful comments on the manuscript.

References

[1] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecommun. (ETT)*, vol. 10, no. 6, pp. 585–596, Nov./Dec. 1999.
[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] 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.
[4] S. A. Jafar and A. Goldsmith, “Transmitter optimization and optimality of beamforming for multiple antenna systems,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.
[5] T. L. Marzetta and B. M. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
[6] S. A. Jafar and A. Goldsmith, “Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[7] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas. Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
[8] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
[9] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
[10] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?," in *Proc. IEEE Globecom*, Dallas, TX, Dec. 2004, pp. 96–100.
[11] L. Musavian, M. R. Nakhai, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 2798–2806, Sep. 2007.
[12] 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.
[13] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
[14] P. Stoica, Y. Jiang, and J. Li, "On MIMO channel capacity: An intuitive discussion," *IEEE Signal Proccess. Mag.*, vol. 5, pp. 83–84, May 2005, lecture notes.
[15] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 214–226, Jan. 2004.
[16] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, pp. 1141–1150, Mar. 2009.
[17] M. Ding, "Multiple-Input Multiple-Output Wireless System Designs with Imperfect Channel Knowledge," Ph.D. dissertation, Queen's Univ., Kingston, ON, Canada, 2008.
[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[19] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Boston, MA: Athena Scientific, 1999.
[20] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamformingdesign for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
[21] A. Hjørungnes, D. Gesbert, and D. P. Palomar, "Unified theory of complex-valued matrix differentiation," in *Proc. ICASSP*, Apr. 2007, vol. 3, pp. III-345–III-348.
[22] H. Lutkepohl, *Handbook of Matrices*. New York: Wiley, 1996.
[23] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[24] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[25] V. A. Aalo, "Performance of maximal-ratio diversity systems in a correlated Nakagami-fading environment," *IEEE Trans. Commun.*, vol. 43, pp. 2360–2369, Aug. 1995.
[26] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

Minhua Ding (S'00–M'08) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2002, respectively, and the Ph.D. degree from Queen's University, Kingston, ON, Canada, in 2008, all in electrical engineering.

From 2002 to 2003, she was a Research Engineer in Datang Mobile Communications Equipment Co., Ltd., Beijing. From 2003 to 2008, she was a Research and Teaching Assistant in the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada. From 2008 to 2009, she held research positions at the Hong Kong University of Science and Technology as well as at Queen's University Belfast, Belfast, U.K. She is currently a Research Associate at the Department of Electronic Engineering, City University of Hong Kong. Her research interests include general areas of wireless communications and signal processing.

Steven D. Blostein (S'83–M'88–SM'96) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1983, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 1985 and 1988, respectively.

He has been on the faculty in the Department of Electrical and Computer Engineering at Queen's University since 1988. Since 2001, he has held the position of Professor and from 2004–2009 he served as Head of Department. From 1999 to 2003, he was the leader of the Multi-Rate Wireless Data Access Major Project sponsored by the Canadian Institute for Telecommunications Research. He has also been a consultant to industry and government in the areas of image compression, target tracking, radar imaging and wireless communications. He spent 1994–1995 at Lockheed Martin Electronic Systems and 2006 at Communications Research Centre in Ottawa. His current interests lie in the application of signal processing to wireless communications, including MIMO systems, cooperative systems as well as multimedia transmission. He has been a member of the Samsung 4G Wireless Forum as well as an invited distinguished speaker at Ryerson University and at Samsung Advanced Institute of Technology.

Dr. Blostein served as Chair of IEEE Kingston Section (1994), Chair of the Biennial Symposium on Communications (2000, 2006, and 2008), Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996–2000), Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–present), Publications Chair for IEEE ICASSP 2004, and has chaired the committee for the Canadian Award for Telecommunications Research from 2004 to 2010. He is a registered Professional Engineer in Ontario.