MIMO Minimum Total MSE Transceiver Design With Imperfect CSI at Both Ends

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

***Abstract*—This paper presents new results on joint linear transceiver design under the minimum total mean-square error (MSE) criterion, with channel mean as well as both transmit and receive correlation information at both ends of a multiple-input multiple-output (MIMO) link. The joint design is formulated into an optimization problem. The optimum closed-form precoder and decoder are derived. Compared to the case with perfect channel state information (CSI), linear filters are added at both ends to balance the suppression of channel noise and the noise from imperfect channel estimation. The impact of channel estimation error as well as channel correlation on system performance is assessed, based on analytical and simulation results.**

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

I. Introduction

MULTIPLE-INPUT MULTIPLE-OUTPUT (MIMO) systems have been subject of recent extensive research, due to their capability of providing high data rates through spatial multiplexing [1], [2] and their capability of providing diversity through space-time coding [3] or beamforming [4], [5].

Previously, precoder designs or joint precoder-decoder designs for MIMO spatial multiplexing systems have been proposed to improve data rates or to enhance link reliability. Linear designs are often preferred due to complexity constraints, especially for mobile terminals. Various performance measures have been considered as the design criteria, e.g., minimum total mean-square error (MSE) from all data streams [6], [7], minimum weighted MSE [8], maximum mutual information (capacity) [1], [7], [8], minimum Euclidean distance between received signal points [9] and minimum bit error rate (BER) [10]. A comprehensive study of joint precoder-decoder designs under MSE-based, signal-to-interference-plus-noise-ratio (SINR)-based, or BER-based criteria has been presented in [11]. Among those, the minimum total MSE criterion aims at improving system BER performance. It leads to convenient analysis, balances interference and noise suppression, and has been widely adopted in single-user (point-to-point) MIMO, multiuser MIMO, or MIMO-OFDM systems [12]–[14].

Precoder designs or joint precoder-decoder designs require channel state information (CSI) at the transmitter (CSIT) and at the receiver (CSIR). In coherent communications (the case considered throughout this paper), CSIR is obtained by channel estimation. To enable signal processing at the transmitter, CSIR is transferred to the transmitter side [15]. CSIT thus acquired is usually imperfect, due to channel estimation error and/or limitations of the feedback link.

The minimum total MSE transceiver designs for MIMO systems have been studied in [6]–[8], [11] and [15]–[19], under different CSI assumptions. In [6]–[8] and [11], the optimum precoder and decoder are derived assuming that perfect CSI is available at both ends. Later there have been more practical designs that consider imperfect CSIT. In [16], the minimum total MSE design has been studied with outdated CSIT and perfect CSIR. In [17], [18], [39, Sec. VII], it is assumed that the CSIT is the channel mean information (CMI) and/or channel correlation information (CCI) [20], whereas the receiver has perfect CSI. The imperfectness in CSIR has also been addressed. In [19, Ch. 7], the same imperfect CSI is assumed at both ends, but there the channel correlation has not been accounted for. In [39, Sec. VI], closed-form robust designs (including the minimum total MSE design) have been derived assuming that the same imperfect CSI, including *channel mean and receive correlation information*, is available to both ends. The same CSI assumption is also used in [15], where the minimum total MSE design has been specifically studied. However, to the best of our knowledge, little attention has been paid to the joint design where the same imperfect CSI, including *channel mean and transmit correlation information*, is available at both ends. This case is very interesting, since in practical downlink systems, the mobile is often surrounded by many local scatterers, whereas the base station is often situated at high enough elevation to limit scattering and thus the channels from different transmit antennas are correlated. The more general case, when the CSI at both ends is imperfect and both transmit and receive antennas exhibit some correlation, also remains as an open problem.

In this paper, we address the problem of linear precoding/decoding to minimize the total MSE with imperfect CSI at both ends of a single-user MIMO link. The CSIR here is composed of the estimated channel (channel mean) as well as transmit (and more generally, transmit and receive) correlation information. To simplify the analysis, we assume that the feedback is error-free and instantaneous, as in [15], [21], [22], and [39, Sec.

Manuscript received April 11, 2007; accepted September 23, 2008. First version published October 31, 2007; current version published February 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Timothy N. Davidson. This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant 41731. This work was presented in part at the International Conference on Communications, Glasgow, U.K., June 2007.

M. Ding was with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada. She is now with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: minhua.ding@ieee.org).

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

Digital Object Identifier 10.1109/TSP.2008.2008542

VI], which implies that the CSIT is the same as CSIR.[1] The assumption of instantaneous feedback is partly justified since, as will be shown by our simulations, the system has acceptable performance with a reasonably low feedback delay. The design under the above assumption is a step forward from that assuming perfect CSI at both ends [6]–[8], [11]. It can also serve as a basis for comparison to future system designs which explicitly take into account the errors and/or delays in the feedback link. Our contributions in this paper can be summarized as follows.

1) A detailed model for channel estimation is presented and the CSI is described. A minimum total MSE design is then formulated as a nonconvex optimization problem under a total transmit power constraint.
2) The closed-form optimum linear precoder and decoder are determined by solving this nonconvex optimization problem. Our analysis and results are extensions of those in [6]–[8], [11] to the case with imperfect CSI. The results obtained here gracefully fit those in [6]–[8], [11] when channel estimation error diminishes, and match the discussions in [15], [19, Ch. 7] when there is no channel correlation (or those in [39, Sec. VI] when there is no transmit correlation). The analysis here can also be extended to other quality-of-service (QoS)-based designs through the minimum weighted MSE design. The relationship between the minimum total MSE design and the maximum mutual information design is discussed under the above assumed imperfect CSI.
3) Based on analytical and simulation results, the impact of channel estimation error as well as the effect of transmit and receive correlation is assessed.

The rest of the paper is organized as follows. Section II describes the system model, the CSI model, the minimum total MSE problem formulation and the basic methodology. For ease of presentation, in Section III the minimum total MSE design problem is solved assuming channel mean and transmit correlation at both ends, and then in Section IV, the analysis is extended to the more general case with both transmit and receive correlation as well as channel mean information at both ends. Numerical results are presented in Section V and conclusions are given in Section VI.

Notation: Upper (lower) case boldface letters are for matrices (vectors); $E\{\cdot\}$ denotes statistical expectation and $tr(\cdot)$ stands for trace; $\det(\mathbf{A})$ means the determinant of matrix $\mathbf{A}$; $(\cdot)^*$ and $(\cdot)^H$ denote complex conjugate and complex conjugate transpose (Hermitian), respectively; $(b)_+ = \max(b, 0)$; $\mathbf{I}_a$ is the $a \times a$ identity matrix; $\mathcal{N}_c(\cdot, \cdot)$ denotes the circularly symmetric complex Gaussian distribution; $vec(\mathbf{A})$ denotes the vectorization operation which stacks the columns of $\mathbf{A}$ into a column vector; $\otimes$ is used for the Kronecker product; $\mathbf{A} \succeq 0$ means that matrix $\mathbf{A}$ is positive semidefinite.

II. System Model, Problem Formulation, and Methodology

A. System Model

Linear precoding/decoding for a MIMO link is addressed. It is assumed that n_T antennas are used at the transmitter and n_R antennas are used at the receiver. The information symbols to be sent are denoted by a $B \times 1$ vector $\mathbf{x}$, where the number of data streams $B(\leq n_T)$ is chosen and fixed. The data vector is then fed into the precoder, denoted by $\mathbf{F}$, which is a $n_T \times B$ linear matrix processor and takes the available CSIT into account. After the precoder, the data vector is transmitted across the slowly varying, flat, Rayleigh fading MIMO channel, described by the $n_R \times n_T$ matrix $\mathbf{H}$. The $n_R \times 1$ received signal vector at the receive antennas is $\mathbf{y} = \mathbf{HFx} + \mathbf{n}$, where $\mathbf{n}$ is the spatially and temporally white additive Gaussian noise with distribution $\mathcal{N}_c(0, \sigma_n^2 \cdot \mathbf{I}_{n_R})$. The input signal $\mathbf{x}$ is assumed to be zero-mean and white $(\mathbf{R_{xx}} = \mathbf{I}_B)$, and independent of channel realization. In the receiver, a linear decoder described by the $B \times n_R$ matrix $\mathbf{G}$ is employed to recover the original information. At the output of the decoder, the signal vector $\mathbf{r}$ is given by $\mathbf{r} = \mathbf{Gy} = \mathbf{G}(\mathbf{HFx} + \mathbf{n})$.

We use the channel model in [23], i.e., $\mathbf{H} = \mathbf{R}_R^{1/2}\mathbf{H}_w\mathbf{R}_T^{1/2}$, where $\mathbf{H}_w$ is a spatially white matrix whose entries are independent and identically distributed (i.i.d.) $\mathcal{N}_c(0, 1)$. The matrices $\mathbf{R}_T$ and $\mathbf{R}_R$ represent normalized transmit and receive correlation (i.e., with unit diagonal entries), respectively. Both $\mathbf{R}_T$ and $\mathbf{R}_R$ are assumed to be *full-rank*.

B. Modeling Channel Estimation

It is assumed that channel estimation is performed on $\mathbf{H}_w$ using the well-established orthogonal training method [15], [24]. At the receive antennas, the signal matrix $\mathbf{Y}_{tr} = \mathbf{HX}_{tr} + \mathbf{N}_{tr}$ is received in n_T successive time slots, where $\mathbf{X}_{tr}$ is a known $n_T \times n_T$ training signal matrix and $\mathbf{N}_{tr}$ is the collection of channel noise vectors. Thus

$$\mathbf{Y}_{tr} = \mathbf{R}_R^{1/2}\mathbf{H}_w\mathbf{R}_T^{1/2}\mathbf{X}_{tr} + \mathbf{N}_{tr}. \tag{1}$$

Let P_{tr} denote the total training power, i.e., $tr(\mathbf{X}_{tr}\mathbf{X}_{tr}^H) = P_{tr}$. Choose $\mathbf{X}_{tr} = \mathbf{R}_T^{-1/2}\mathbf{X}$, where $\mathbf{X}$ is a $n_T \times n_T$ unitary matrix scaled by $\sqrt{P_{tr}/tr(\mathbf{R}_T^{-1})}$. Premultiplying both sides of (1) by $\mathbf{R}_R^{-1/2}$ and then postmultiplying the resultant formula by $\mathbf{X}^{-1}$, we obtain

$$\tilde{\mathbf{H}}_w = \mathbf{R}_R^{-1/2}\mathbf{Y}_{tr}\mathbf{X}^{-1} = \mathbf{H}_w + \mathbf{R}_R^{-1/2}\mathbf{N}_{tr}\mathbf{X}^{-1}.$$

Let $\mathbf{N}_0 = \mathbf{N}_{tr}\mathbf{X}^{-1}$. The entries of $\mathbf{N}_0$ are i.i.d. $\mathcal{N}_c(0, \sigma_{ce}^2)$ with $\sigma_{ce}^2 = tr(\mathbf{R}_T^{-1}) \cdot \sigma_n^2/P_{tr}$. Then

$$\tilde{\mathbf{H}}_w = \mathbf{H}_w + \mathbf{R}_R^{-1/2}\mathbf{N}_0. \tag{2}$$

To obtain a better channel estimation performance, the minimum MSE (MMSE) channel estimation of $\mathbf{H}_w$ is performed based on (2) [15], [21], [22], [25], which yields

$$\hat{\mathbf{H}}_w = E[\mathbf{H}_w|\tilde{\mathbf{H}}_w] = \left[\mathbf{I}_{n_R} + \sigma_{ce}^2 \cdot \mathbf{R}_R^{-1}\right]^{-1}\tilde{\mathbf{H}}_w. \tag{3}$$

Furthermore, $\mathbf{H}_w$ is expressed as the sum of $\hat{\mathbf{H}}_w$ and the estimation error matrix, i.e.

$$\mathbf{H}_w = \hat{\mathbf{H}}_w + \mathbf{R}_R^{-1/2}\left[\mathbf{I}_{n_R} + \sigma_{ce}^2 \cdot \mathbf{R}_R^{-1}\right]^{-1/2}\mathbf{E}_w \tag{4}$$

[1] One can also assume that the system is implemented offline, and the precoding matrix is calculated at the receiver and then fed back to the transmitter [12].

where the entries of $\mathbf{E}_w$ are i.i.d. $\mathcal{N}_c(0,\sigma_{ce}^2)$. Detailed derivations of (3) and (4) are provided in Appendix A. Let $\mathbf{R}_{e,R} = \left[\mathbf{I}_{n_R} + \sigma_{ce}^2 \cdot \mathbf{R}_R^{-1}\right]^{-1}$. The CSI model is described by

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E} \tag{5}$$

where $\mathbf{H}$ is the true channel matrix, $\hat{\mathbf{H}} = \mathbf{R}_R^{1/2}\hat{\mathbf{H}}_w\mathbf{R}_T^{1/2}$ is the estimated channel matrix (i.e., the channel mean), and $\mathbf{E} = \mathbf{R}_{e,R}^{1/2}\mathbf{E}_w\mathbf{R}_T^{1/2}$ is the channel estimation error matrix.

In summary, the CSI is given by (3)–(5). In subsequent sections, we assume that $\hat{\mathbf{H}}$, $\mathbf{R}_R$, $\mathbf{R}_T$, σ_{ce}^2, and σ_n^2 are known to both ends of the link, which is referred to as *the channel mean as well as both transmit and receive correlation information*. (Note that this implies knowledge of $\mathbf{R}_{e,R}$.)

We point out that in [22], a different CSI model for the channel $\mathbf{H} = \mathbf{R}_R^{1/2}\mathbf{H}_w\mathbf{R}_T^{1/2}$ is employed, which is given by $\mathbf{H} = \underline{\hat{\mathbf{H}}} + \underline{\mathbf{E}}$, with $\underline{\hat{\mathbf{H}}} = \mathbf{R}_R^{1/2}\underline{\hat{\mathbf{H}}}_w\mathbf{R}_T^{1/2}$ and $\underline{\mathbf{E}} = \mathbf{R}_R^{1/2}\underline{\mathbf{E}}_w\mathbf{R}_T^{1/2}$, where the entries of $\underline{\hat{\mathbf{H}}}_w$ and $\underline{\mathbf{E}}_w$ are assumed to be i.i.d.. However, in [22] it is assumed that a genie-provided estimate of $\mathbf{H}_w$ (i.e. $\underline{\hat{\mathbf{H}}}_w$) is available at the receiver. In comparison, this is not required in our channel estimation method or CSI model. Also, in [39, Sec. VI], it has been assumed that $\mathbf{H} = \bar{\mathbf{H}} + \mathbf{R}_R^{1/2}\underline{\underline{\mathbf{E}}}_w\mathbf{R}_T^{1/2}$, where $\bar{\mathbf{H}}$ is the channel mean, and $\underline{\underline{\mathbf{E}}}_w$ is spatially white (i.e., with i.i.d. entries). It is important to note that the analysis to be presented in this paper can be applied exactly the same way when using the CSI model in [22] or [39, Sec. VI].

C. Problem Formulation

With the above CSI, the received signal vector $\mathbf{y}$ can be written as

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{F}\mathbf{x} + \underbrace{\mathbf{E}\mathbf{F}\mathbf{x} + \mathbf{n}}_{\text{total noise}}. \tag{6}$$

The system MSE matrix is calculated as

$$\begin{aligned}\text{MSE}(\mathbf{F},\mathbf{G}) &= E\left[(\mathbf{r}-\mathbf{x})(\mathbf{r}-\mathbf{x})^H\right] \\ &= E\left\{\left[\mathbf{G}(\hat{\mathbf{H}}+\mathbf{E})\mathbf{F} - \mathbf{I}_B\right]\mathbf{x}\mathbf{x}^H \right. \\ &\qquad \left. \times \left[\mathbf{G}(\hat{\mathbf{H}}+\mathbf{E})\mathbf{F} - \mathbf{I}_B\right]^H\right\} + \sigma_n^2 \cdot \mathbf{G}\mathbf{G}^H.\end{aligned} \tag{7}$$

Using our assumptions on the statistics of the channel, noise, and data, with some manipulations, we can simplify (7) as

$$\begin{aligned}\text{MSE}(\mathbf{F},\mathbf{G}) = \mathbf{G}\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H - \mathbf{G}\hat{\mathbf{H}}\mathbf{F} - \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H + \mathbf{I}_B \\ + \left[\sigma_{ce}^2 \cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)\right]\mathbf{G}\mathbf{R}_{e,R}\mathbf{G}^H + \sigma_n^2\mathbf{G}\mathbf{G}^H.\end{aligned} \tag{8}$$

In the above, we have used the result $E\{\mathbf{E}_w\mathbf{A}\mathbf{E}_w^H\} = \sigma_{ce}^2 \cdot tr(\mathbf{A}) \cdot \mathbf{I}_{n_R}$, if the entries of matrix $\mathbf{E}_w$ are i.i.d. $\mathcal{N}_c(0,\sigma_{ce}^2)$, as well as the identity $tr(\mathbf{A}_1\mathbf{A}_2) = tr(\mathbf{A}_2\mathbf{A}_1)$.

Our goal is to find a pair of appropriate $\mathbf{F}$ and $\mathbf{G}$, such that the sum of MSEs from different data streams is minimized subject to a total power constraint P_T, i.e.

$$\begin{aligned}&\min_{\mathbf{F},\mathbf{G}} \quad tr\left[\text{MSE}(\mathbf{F},\mathbf{G})\right] \\ &\text{subject to} \quad tr(\mathbf{F}\mathbf{F}^H) \le P_T.\end{aligned} \tag{9}$$

This is referred to as the minimum total MSE design. We are also interested in determining the effects of channel correlation and channel estimation error on system performance.

Note that when $\sigma_{ce}^2 = 0$, our problem in (9) reduces to those treated in [6]–[8] and [11]. Also, when $\sigma_{ce}^2 \neq 0$ and $\mathbf{R}_T = \mathbf{I}_{n_T}$, $tr(\mathbf{F}\mathbf{F}^H)$ is replaced by P_T as in [15], [19, Ch. 7], [39, Sec. VI], and the problem (9) becomes mathematically equivalent to the perfect CSI case. However, the case with $\sigma_{ce}^2 \neq 0$ and $\mathbf{R}_T \neq \mathbf{I}_{n_T}$ has not yet been considered in the literature.

D. Methodology

The objective function of (9), i.e. $tr[\text{MSE}(\mathbf{F},\mathbf{G})]$, is nonconvex in $(\mathbf{F},\mathbf{G})$, and thus the methods designed for convex problems are not applicable here. Fortunately, it can be shown that a global minimum exists for the problem in (9) (see Appendix B). Furthermore, the objective and constraint functions of (9) are *continuously differentiable* (with respect to $\mathbf{G}$ and/or $\mathbf{F}$). Since there is only one inequality constraint and no equality constraints in (9), any feasible precoder-decoder pair is regular whether the inequality constraint is active or inactive [26, pp. 309, 310].[2] The global-minimum-achieving $(\mathbf{F},\mathbf{G})$ therefore satisfies the first-order Karush-Kuhn-Tucker (KKT) necessary conditions for optimality [26, p. 310, Proposition 3.3.1]. Our method is then to find all the solutions which satisfy the KKT conditions and identify the optimum $(\mathbf{F},\mathbf{G})$ among them.

Note that if $(\mathbf{F},\mathbf{G})$ minimizes the total MSE, so does $(\mathbf{F}\mathbf{U},\mathbf{U}^H\mathbf{G})$, with $\mathbf{U}$ being an arbitrary $B \times B$ unitary matrix. Below we refer to a specific optimum precoder-decoder pair as the optimum solution *up to a unitary transform*.

III. A Special Case: $\mathbf{R}_T \neq \mathbf{I}_{n_T}$, $\mathbf{R}_R = \mathbf{I}_{n_R}$

From (2)–(4), it can be shown that when $\mathbf{R}_R = \mathbf{I}_{n_R}$, $\mathbf{H}_w = \hat{\mathbf{H}}_{w0} + \mathbf{E}_{w0}$, with both $\hat{\mathbf{H}}_{w0}$ and $\mathbf{E}_{w0}$ being spatially white. The entries of $\hat{\mathbf{H}}_{w0}$ and $\mathbf{E}_{w0}$ are mutually uncorrelated, and are i.i.d. $\mathcal{N}_c(0,1-\sigma_E^2)$ and $\mathcal{N}_c(0,\sigma_E^2)$, respectively, with $\sigma_E^2 = \sigma_{ce}^2/(1+\sigma_{ce}^2)$. The CSI model is described by $\mathbf{H} = (\hat{\mathbf{H}}_{w0} + \mathbf{E}_{w0})\mathbf{R}_T^{1/2} = \hat{\mathbf{H}} + \mathbf{E}$, where $\hat{\mathbf{H}} = \hat{\mathbf{H}}_{w0}\mathbf{R}_T^{1/2}$ is the channel mean, and $\mathbf{E} = \mathbf{E}_{w0}\mathbf{R}_T^{1/2}$. Here we assume that $\hat{\mathbf{H}}$, $\mathbf{R}_T$, σ_E^2 and σ_n^2 are known to both ends, which is referred to as *the channel mean and transmit correlation information*.

When $\mathbf{R}_R = \mathbf{I}_{n_R}$, the MSE matrix in (8) reduces to

$$\begin{aligned}\text{MSE}(\mathbf{F},\mathbf{G}) = \mathbf{G}\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H - \mathbf{G}\hat{\mathbf{H}}\mathbf{F} - \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H + \mathbf{I}_B \\ + \left[\sigma_n^2 + \sigma_E^2 \cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)\right]\mathbf{G}\mathbf{G}^H.\end{aligned} \tag{10}$$

The problem formulation here is the same as (9), except that the MSE matrix is given by (10). The associated Lagrangian is $\mathcal{L}(\mathbf{F},\mathbf{G},\mu_1) = tr[\text{MSE}(\mathbf{F},\mathbf{G})] + \mu_1 \cdot [tr(\mathbf{F}\mathbf{F}^H) - P_T]$, where μ_1 is the Lagrange multiplier. By taking the derivatives of $\mathcal{L}(\mathbf{F},\mathbf{G},\mu_1)$ with respect to $\mathbf{F}$ and $\mathbf{G}$ [27], [33], together with the power constraint and complementary slackness, the associated KKT conditions can be obtained as follows:

$$\left\{\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \left[\sigma_n^2 + \sigma_E^2 tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)\right]\mathbf{I}_{n_R}\right\}\mathbf{G}^H = \hat{\mathbf{H}}\mathbf{F} \tag{11}$$

[2] According to [26, pp. 309, 310], a feasible point is said to be *regular* if either: (i) all equality constraint gradients and active inequality constraint gradients at this point are linearly independent; or (ii) in the case of no equality constraints, all the inequality constraints are inactive at this point.

$$\mathbf{F}^H\left[\hat{\mathbf{H}}^H\mathbf{G}^H\mathbf{G}\hat{\mathbf{H}}+\sigma_E^2 tr(\mathbf{G}\mathbf{G}^H)\cdot\mathbf{R}_T+\mu_1\mathbf{I}_{n_T}\right]=\mathbf{G}\hat{\mathbf{H}} \quad (12)$$

$$\mu_1\geq 0, tr(\mathbf{F}\mathbf{F}^H)-P_T\leq 0 \quad (13)$$

$$\mu_1\cdot\left[tr(\mathbf{F}\mathbf{F}^H)-P_T\right]=0. \quad (14)$$

Clearly, if $\mathbf{F}=0$, then an obvious solution satisfying the KKT condition is: $\mathbf{F}=0$, $\mathbf{G}=0$, and $\mu_1=0$. Note that ($\mathbf{F}=0$, $\mathbf{G}=0$) is a regular feasible point at which the only inequality constraint is inactive [26, pp. 309, 310]. However, this case is not interesting to us in practice. Therefore, we proceed to search for those solutions with $\mathbf{F}\neq 0$ (referred to as *the nonzero solutions*).

Lemma 1: For any solution satisfying the KKT conditions (11)–(14), $\mu_1=\sigma_n^2\cdot tr(\mathbf{G}\mathbf{G}^H)/P_T$.

Proof: See Appendix C.

Consider the following eigenvalue decomposition:

$$\left[\sigma_E^2 P_T\cdot\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\left[\sigma_E^2 P_T\cdot\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}=(\mathbf{V}\quad\tilde{\mathbf{V}})\begin{pmatrix}\mathbf{\Lambda}&0\\0&\tilde{\mathbf{\Lambda}}\end{pmatrix}(\mathbf{V}\ \tilde{\mathbf{V}})^H. \quad (15)$$

Let r be equal to the rank of the matrix in (15), i.e., $r=\mathrm{rank}(\mathbf{\Lambda})$, the number of nonzero channel eigenmodes. Here the $n_T\times(n_T-r)$ matrix $\tilde{\mathbf{V}}$ consists of basis vectors for the null space of (15). The entries of the diagonal matrix $\tilde{\mathbf{\Lambda}}$ are all zero. The $n_T\times r$ matrix $\mathbf{V}$ is composed of the eigenvectors corresponding to nonzero eigenvalues. Without loss of generality, the entries of the diagonal matrix $\mathbf{\Lambda}$ are arranged in decreasing order.

Lemma 2: Assume that the number of data streams B is equal to r. The precoder and decoder satisfying the KKT conditions (11)–(14) can be expressed as

$$\mathbf{F}=\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\mathbf{V}\mathbf{\Lambda}_F \quad (16)$$

$$\mathbf{G}=\mathbf{\Lambda}_G\mathbf{V}^H\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\hat{\mathbf{H}}^H \quad (17)$$

where $\mathbf{\Lambda}_F$ and $\mathbf{\Lambda}_G$ are arbitrary $r\times r$ matrices, and $\mathbf{V}$ comes from (15).

Proof: See Appendix D.

Theorem 1: Assume that the number of data streams B is equal to r. Without loss of generality, the optimum precoder and decoder for (9) have the following expressions, respectively:

$$\mathbf{F}_{\mathrm{opt}}=\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\mathbf{V}\mathbf{\Lambda}_{Fopt} \quad (18)$$

$$\mathbf{G}_{\mathrm{opt}}=\mathbf{\Lambda}_{Gopt}\mathbf{V}^H\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\hat{\mathbf{H}}^H. \quad (19)$$

The diagonal $r\times r$ matrices $\mathbf{\Lambda}_{Fopt}$ and $\mathbf{\Lambda}_{Gopt}$ are given by

$$\mathbf{\Lambda}_{Fopt}=\left[\tau_1^{1/2}\mu_1^{-1/2}\sigma_n\cdot\mathbf{\Lambda}^{-1/2}-\tau_1\cdot\mathbf{\Lambda}^{-1}\right]_+^{1/2} \quad (20)$$

$$\mathbf{\Lambda}_{Gopt}=\left[\mu_1^{1/2}\tau_1^{-1/2}\frac{1}{\sigma_n}\cdot\mathbf{\Lambda}^{-1/2}-\frac{\mu_1}{\sigma_n^2}\cdot\mathbf{\Lambda}^{-1}\right]_+^{1/2}\mathbf{\Lambda}^{-1/2} \quad (21)$$

where

$$\tau_1=\frac{a_2\cdot P_T}{P_T\cdot a_3+a_1\cdot a_3-a_2\cdot a_4} \quad (22)$$

$$\mu_1=\frac{a_2\cdot\sigma_n^2(P_T\cdot a_3+a_1\cdot a_3-a_2\cdot a_4)}{(P_T+a_1)^2\cdot P_T}. \quad (23)$$

Let the integer k denote the number of nonzero entries of $\mathbf{\Lambda}_{Fopt}(k\leq r)$. Scalars a_1, a_2, a_3, and a_4 are traces of the $k\times k$ top-left submatrices of $\mathbf{\Lambda}^{-1}$, $\mathbf{\Lambda}^{-1/2}$, $\mathbf{\Lambda}^{-1/2}\mathbf{V}^H[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]^{-1}\mathbf{V}$, and $\mathbf{\Lambda}^{-1}\mathbf{V}^H[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]^{-1}\mathbf{V}$, respectively. The optimum precoder-decoder pair obtained here is unique up to a unitary transform.

Proof: As derived in Appendix E, all the nonzero solutions of $(\mathbf{F},\mathbf{G},\mu_1)$ satisfying the KKT conditions are given by (18)–(23) up to a unitary transform. The method to determine the number k is also included there. Thus, we have obtained all the solutions satisfying the KKT conditions (11)–(14), including ($\mathbf{F}=0$, $\mathbf{G}=0$, $\mu_1=0$), and the nonzero solutions given by (18)–(23) up to a unitary transform. It can be readily shown that all the nonzero solutions lead to the same total MSE, which is lower than the MSE yielded by ($\mathbf{F}=0$, $\mathbf{G}=0$, $\mu_1=0$) [33]. Therefore, we conclude that the nonzero solutions [(18)–(23), up to a unitary transform] are equivalent global MSE-minimizers.[3] □

Remark 1: When $\sigma_E^2=0$ ($\sigma_{ce}^2=0$), Theorem 1 reduces to the results in [6]–[8] or [11]. Compared to the results obtained under perfect CSI, we observe from (18) and (19) that a linear filter is added to both the transmitter and receiver here, to balance the suppression of channel noise and the noise from imperfect channel estimation. Furthermore, the estimation error variance σ_E^2 is coupled with the transmit correlation matrix $\mathbf{R}_T$.

Remark 2: From (18), when $\mathbf{R}_T=\mathbf{I}_{n_T}$, transmission along the eigenmodes of $\hat{\mathbf{H}}^H\hat{\mathbf{H}}$ is optimum, and the channel estimation error simply contributes additional noise ($\sigma_E^2\cdot P_T$). This result has been mentioned in [15], [19, Ch. 7], [39, Sec. VI]. When $P_T/\sigma_n^2\to\infty$, the filter $[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]^{-1/2}$ becomes a scaled version of $\mathbf{R}_T^{-1/2}$. This implies that the optimum precoder asymptotically cancels the effect of $\mathbf{R}_T$ and transmits along channel eigenmodes of the white part of the channel estimate ($\hat{\mathbf{H}}_{w0}$).

IV. The General Case: $\mathbf{R}_T\neq\mathbf{I}_{n_T}$ and $\mathbf{R}_R\neq\mathbf{I}_{n_R}$

Consider the general problem formulated in (9). Applying the same method as in the special case, the associated Lagrangian is $\mathcal{L}_g(\mathbf{F},\mathbf{G},\mu)=tr\{\mathrm{MSE}(\mathbf{F},\mathbf{G})\}+\mu\cdot[tr(\mathbf{F}\mathbf{F}^H)-P_T]$, where μ is the Lagrange multiplier. Correspondingly, the KKT conditions associated with (9) can be derived, as given by (24)–(27)

$$\left[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H+\sigma_{ce}^2 tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)\mathbf{R}_{e,R}+\sigma_n^2\mathbf{I}_{n_R}\right]\mathbf{G}^H=\hat{\mathbf{H}}\mathbf{F} \quad (24)$$

$$\mathbf{F}^H\left[\hat{\mathbf{H}}^H\mathbf{G}^H\mathbf{G}\hat{\mathbf{H}}+\sigma_{ce}^2 tr(\mathbf{G}\mathbf{R}_{e,R}\mathbf{G}^H)\mathbf{R}_T+\mu\mathbf{I}_{n_T}\right]=\mathbf{G}\hat{\mathbf{H}} \quad (25)$$

$$\mu\geq 0, tr(\mathbf{F}\mathbf{F}^H)-P_T\leq 0 \quad (26)$$

[3]One can verify the results in Theorem 1 using the Saddle Point Theorem [26, p. 491, Proposition 5.1.6]. Basically, the optimum solution obtained here satisfies a necessary and sufficient condition for global optimality.

$$\mu \cdot \left[tr(\mathbf{F}\mathbf{F}^H) - P_T \right] = 0. \tag{27}$$

It can be shown that Lemma 1 still holds here, i.e., $\mu = \sigma_n^2 \cdot tr(\mathbf{G}\mathbf{G}^H)/P_T$ for any solution which satisfies the KKT conditions (24)–(27). Define

$$\alpha = tr(\mathbf{G}\mathbf{R}_{e,R}\mathbf{G}^H), \quad \text{and} \quad \beta = \sigma_{ce}^2 \cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H). \tag{28}$$

Consider the following eigenvalue decomposition:

$$\begin{aligned}[\mathbf{V}_g \;\; \tilde{\mathbf{V}}_g] \begin{pmatrix} \mathbf{\Lambda}_g & 0 \\ 0 & \tilde{\mathbf{\Lambda}}_g \end{pmatrix} [\mathbf{V}_g \;\; \tilde{\mathbf{V}}_g]^H \\ = \left[\alpha \cdot \sigma_{ce}^2 \cdot \mathbf{R}_T + \mu \cdot \mathbf{I}_{n_T}\right]^{-1/2} \hat{\mathbf{H}}^H \left[\beta \cdot \mathbf{R}_{e,R} + \sigma_n^2 \cdot \mathbf{I}_{n_R}\right]^{-1} \\ \cdot \hat{\mathbf{H}} \left[\alpha \cdot \sigma_{ce}^2 \cdot \mathbf{R}_T + \mu \cdot \mathbf{I}_{n_T}\right]^{-1/2}\end{aligned} \tag{29}$$

where the subscript "g" means the general case, the matrices $\mathbf{V}_g$, $\tilde{\mathbf{V}}_g$, $\mathbf{\Lambda}_g$, and $\tilde{\mathbf{\Lambda}}_g$ are similarly defined as those in (15), and r_g denotes the rank of the matrix in (29), i.e., $r_g = \mathrm{rank}(\mathbf{\Lambda}_g)$. The diagonal entries of $\mathbf{\Lambda}_g$ are arranged in decreasing order.

Theorem 2: Assume that the number of data streams B is equal to r_g. Without loss of generality, the optimum precoder and decoder for (9) can be expressed as

$$\mathbf{F}_{gopt} = \left[\alpha \cdot \sigma_{ce}^2 \cdot \mathbf{R}_T + \mu \cdot \mathbf{I}_{n_T}\right]^{-1/2} \mathbf{V}_g \mathbf{\Lambda}_{F,gopt} \tag{30}$$

$$\begin{aligned}\mathbf{G}_{gopt} = \mathbf{\Lambda}_{G,gopt} \mathbf{V}_g^H \left[\alpha \cdot \sigma_{ce}^2 \cdot \mathbf{R}_T + \mu \cdot \mathbf{I}_{n_T}\right]^{-1/2} \\ \cdot \hat{\mathbf{H}}^H \left[\beta \cdot \mathbf{R}_{e,R} + \sigma_n^2 \cdot \mathbf{I}_{n_R}\right]^{-1}\end{aligned} \tag{31}$$

where $\mathbf{V}_g$ is from (29), $\mathbf{\Lambda}_{F,gopt}$ and $\mathbf{\Lambda}_{G,gopt}$ are $r_g \times r_g$ diagonal matrices, as given

$$\mathbf{\Lambda}_{F,gopt} = \left[\mathbf{\Lambda}_g^{-1/2} - \mathbf{\Lambda}_g^{-1}\right]_+^{1/2} \tag{32}$$

$$\mathbf{\Lambda}_{G,gopt} = \left[\mathbf{\Lambda}_g^{-1/2} - \mathbf{\Lambda}_g^{-1}\right]_+^{1/2} \mathbf{\Lambda}_g^{-1/2}. \tag{33}$$

Inserting (30)–(33) into the power constraint $tr\{\mathbf{F}\mathbf{F}^H\} = P_T$ as well as the definitions of α and β in (28) results in three equations with α, β, and μ being the unknowns. By solving these equations numerically, the values of α, β, and μ can be determined. The optimum solution obtained using this method is unique up to a unitary transform.

Proof: The proof is similar to that for Theorem 1 and it is omitted for brevity. Details are available in [33].

Theorem 2, hence, provides the structures of the optimum precoder and decoder. However, the scalars α, β and μ need to be determined numerically, which can be a difficult task. Alternatively, (9) can be solved using an iterative algorithm developed from the KKT conditions [12], [13], [28], which is given in Table I. This algorithm converges according to [13]. Furthermore, starting from a nonzero feasible $\mathbf{F}$, this algorithm obtains a nonzero solution satisfying the KKT conditions. Since all the nonzero solutions satisfying the KKT conditions lead to the same minimum total MSE, we conclude that the iterative algorithm obtains a solution equivalent to the one obtained from Theorem 2 up to a unitary transform. Therefore, Table I presents a convenient method for the general case. Note that our method to update the Lagrange multiplier μ (see Table I) is different from that in [12] or [13]. In fact, our method is much simpler and more reliable in simulations, as the method in [12] and [13] requires an eigenvalue decomposition and a solution to a nonlinear equation for each update of the Lagrange multiplier.

TABLE I
AN ITERATIVE ALGORITHM FOR SOLVING (9) IN THE GENERAL CASE

1)	Initialize $\mathbf{F} = \mathbf{F}_0$; the upper $B \times B$ 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 $\mathbf{G}$ using (24);
3)	Update μ using $\mu = \sigma_n^2 \cdot tr(\mathbf{G}\mathbf{G}^H)/P_T$;
4)	Update $\mathbf{F}$ using (25); Scale if needed such that $tr(\mathbf{F}\mathbf{F}^H) = P_T$;
5)	If $tr\{(\mathbf{F}_i - \mathbf{F}_{i-1})(\mathbf{F}_i - \mathbf{F}_{i-1})^H\}$ is sufficiently small (say, less than 10^{-4}), stop; otherwise, go back to 2). Here $\mathbf{F}_i$ ($\mathbf{F}_{i-1}$) denotes $\mathbf{F}$ in the i-th [(i-1)-st] iteration.

Remark 3: We have assumed $B = r$ in Theorem 1 and $B = r_g$ in Theorem 2. If the number of data streams B is chosen to be strictly smaller than the number of nonzero channel eigenmodes, i.e., the B strongest eigenmodes are used, then redundancy is introduced, which can be translated into improved diversity and, thus, performance improvement [8], [29]. However, the diversity effect is achieved at the cost of a reduced number of data streams (and thus reduced data rate). Therefore, there is a diversity-multiplexing tradeoff here [30]. The choice of B according to channel conditions to guarantee a constant rate has been studied in [29] assuming perfect CSI at both ends. Also, the optimum choice of B to guarantee that the MSE in every data stream is lower than a given target (in addition to the total transmit power constraint) has been studied in [15] where channel estimation error and receive correlation only are considered. It is possible to extend the method in [15] to the more general setting presented in this paper.

Remark 4: In [8], the minimum weighted MSE joint transceiver design has been studied under perfect CSI at both ends. It includes several other designs as special cases, e.g., the QoS-based designs which achieve different SNRs on different subchannels by adjusting the weighting matrix. It is straightforward to extend our analysis to the minimum weighted MSE design with the imperfect CSI assumed earlier [33]. The impact of channel estimation error and channel correlation on these designs can then be investigated.

Remark 5: (Relation between the minimum total MSE design and the maximum mutual information design) The minimum total MSE design and the maximum mutual information design are closely related. In [7] and [11], it is pointed out that, under perfect CSI, the minimum total MSE design minimizes the trace of a MSE matrix, whereas the maximum mutual information (capacity) design minimizes its log-determinant. Furthermore, the optimum transmitters for both designs differ only in a diagonal power allocation [7], [8], [11]. With imperfect CSI, exact capacity formulas are difficult to obtain. Instead, tight upper- and lower-bounds on mutual information are derived to determine appropriate transmission strategies [21], [31], [32]. In [31,

Sec. IV-B], a lower-bound problem on maximum mutual information (capacity) has been formulated with channel mean and transmit correlation information at both ends, i.e.

$$\max_{\mathbf{Q}} \quad \log_2 \det \left[\mathbf{I}_{n_R} + \frac{\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H}{\sigma_n^2 + \sigma_E^2 tr(\mathbf{R}_T\mathbf{Q})} \right]$$
$$\text{subject to} \quad \mathbf{Q} \succeq 0 \text{ and } tr(\mathbf{Q}) \le P_T,$$

and a numerical search method has been used to determine the optimum transmit covariance matrix $\mathbf{Q}$. However, using a method similar to that for our minimum total MSE design, the closed-form optimum transmit covariance matrix $\mathbf{Q}$ can be determined [33], and can be written as $\mathbf{Q}_{\text{opt}} = \mathbf{F}_{c,opt}\mathbf{F}_{c,opt}^H$. The matrix $\mathbf{F}_{c,opt}$ has the same structure as $\mathbf{F}_{\text{opt}}$ in (18), i.e., it also consists of a linear filter, a matrix composed of channel eigenmodes and a diagonal power allocation matrix. Interestingly, with channel mean and transmit correlation at both ends, the optimum transmitters for both designs differ only in the diagonal power allocation, as they do under perfect CSI. The relation between these two designs in the general case (with both transmit and receive correlation) has also been determined in [33].

V. Numerical Results

A. Simulation Scenario

Let $n_T = n_R = 4$. The exponential model is used for both transmit and receive correlation [31], [34]. Specifically, the transmit correlation model is given by: $(\mathbf{R}_T)_{ij} = \rho_T^{|i-j|}$ 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. QPSK is used for each data stream. The optimum matrices $\mathbf{F}_{\text{opt}}$ and $\mathbf{G}_{\text{opt}}$ or $\mathbf{F}_{g,opt}$ and $\mathbf{G}_{g,opt}$ are chosen specifically as (18) and (19) or (30) and (31), respectively. For the minimum total MSE design, a relevant performance measure is the average bit error probability (ABEP) per data stream [8], defined as $\text{ABEP} = (1/B)\sum_{j=1}^{B} \text{BEP}_j$, where BEP_j is the BEP of data stream j. Define SNR as P_T/σ_n^2. For fair comparisons, we fix P_{tr}/σ_n^2 in the training stage and let $\sigma_{ce}^2 = tr(\mathbf{R}_T^{-1}) \cdot \sigma_n^2/P_{tr}$ vary with $\mathbf{R}_T$. In our simulations, P_{tr}/σ_n^2 is chosen to be 16.016 or 26.016 dB, which corresponds to $\sigma_{ce}^2 = 0.1$ or 0.01 if $\mathbf{R}_T = \mathbf{I}_{n_T}$.

B. Simulation Results

Example 1: Effects of transmit and receive correlation when there is no channel estimation error [$\sigma_{ce}^2 = 0$ ($\sigma_E^2 = 0$); $\rho_T = 0.0, 0.5, 0.9$; $\rho_R = 0.0, 0.5$; $B = 3, 4$]

Fig. 1 is obtained when there is no channel estimation error ($\sigma_{ce}^2 = 0$). High correlation is observed to have a large impact on system performance. It is clear that reducing the number of data streams B introduces diversity and thus compensates for the loss caused by channel correlation. Consequently, B should be chosen carefully according to channel correlation information.

Example 2: Effects of channel estimation error and channel correlation [$\sigma_{ce}^2 \neq 0$ with $P_{tr}/\sigma_n^2 = 26.016$ dB; $\sigma_E^2 = \sigma_{ce}^2/(1+\sigma_{ce}^2)$; $\rho_T = 0.0, 0.5, 0.9$; $\rho_R = 0.0, 0.5$; $B = 3, 4$]

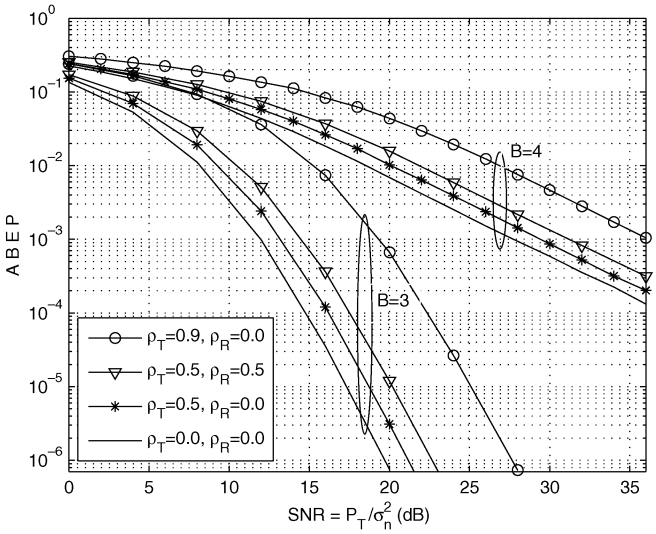

Fig. 1. The ABEP from minimum total MSE design with perfect CSI. $n_T = n_R = 4$, $B = 3$ or 4. Different amounts of channel correlation are considered: $\rho_T = 0.0, 0.5, 0.9$ and $\rho_R = 0.0, 0.5$.

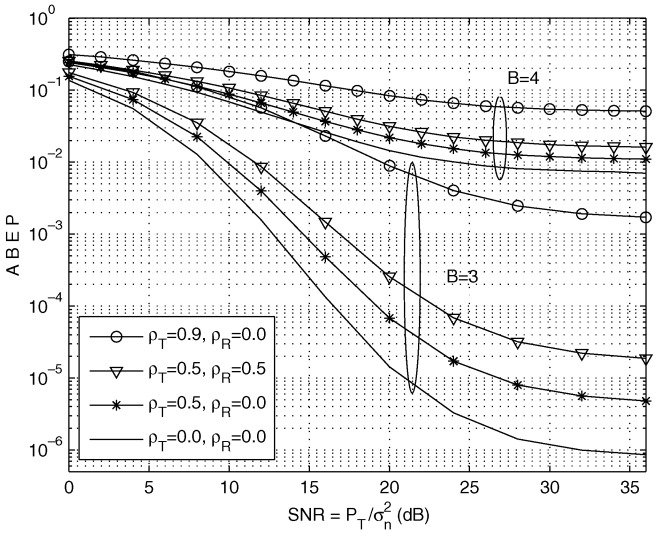

Fig. 2. The ABEP from minimum total MSE design with imperfect CSI. $n_T = n_R = 4$, $B = 3$ or 4. Different amounts of channel correlation are considered. $P_{tr}/\sigma_n^2 = 26.016$ dB. The values of σ_{ce}^2 are 0.01, 0.015, and 0.0739, for $\rho_T = 0.0, 0.5$, and 0.9, respectively. Correspondingly, the values of σ_E^2 are 0.0099, 0.0148, and 0.0689, for $\rho_T = 0.0, 0.5$, and 0.9.

Fig. 2 shows the ABEPs from using the optimum precoder and decoder when the CSI is imperfect. From Fig. 2, we observe that channel estimation error alone has a tremendously detrimental effect on system ABEP performance (see the curves in Fig. 2 with $\rho_T = \rho_R = 0$ and compare them with those in Fig. 1). At medium to high SNR, the performance degradation caused by channel estimation error can be compensated by introducing diversity (i.e., reducing the number of data streams), at the expense of reduced data rate. Also, channel estimation error causes an irreducible error floor at high SNR. High channel correlation further deteriorates system performance.

Example 3: Optimum precoder versus two suboptimum precoders with asymptotic optimality [$\rho_T = 0.7$, $\rho_R = 0$; $B = 3$; $\sigma_{ce}^2 \neq 0$ with $P_{tr}/\sigma_n^2 = 16.016$ or 26.016 dB].

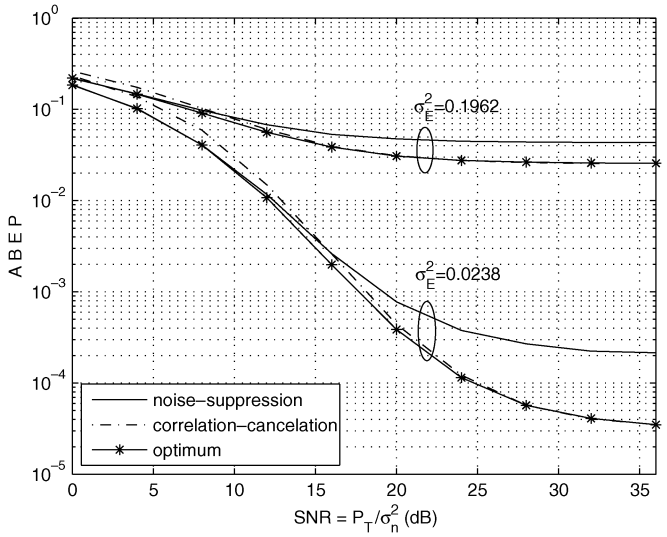

Fig. 3. Comparison of the optimum and two suboptimum transceivers. $n_T = n_R = 4$, $B = 3$, $\rho_T = 0.7$, $\rho_R = 0.0$. Two values of the training power are used: $P_{tr}/\sigma_n^2 = 26.016$ dB (corresponding to $\sigma_E^2 = 0.0238$), and $P_{tr}/\sigma_n^2 = 16.016$ dB (corresponding to $\sigma_E^2 = 0.1962$).

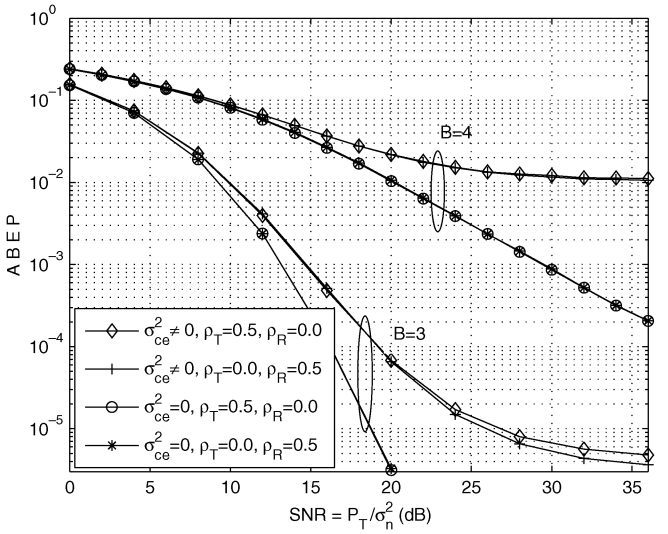

Fig. 4. Effect of transmit correlation versus effect of receive correlation. $n_T = n_R = 4$, $B = 3$ or 4, $(\rho_T, \rho_R) = (0.5, 0.0)$ or (0.0, 0.5). In the case of imperfect CSI, $P_{tr}/\sigma_n^2 = 26.016$ dB. The values of σ_{ce}^2 are 0.01 and 0.015, for $\rho_T = 0.0$ and 0.5, respectively.

Consider the special case without receive correlation in Section III. A suboptimum transceiver structure can be obtained by ignoring the channel correlation information, treating $\hat{\mathbf{H}}$ as if it were the true channel and then applying the results in [8]. In this way, the precoder is restricted to be of the form $\mathbf{F}_{NS} = \mathbf{V}_1\mathbf{\Lambda}_{FNS}$, where the matrix $\mathbf{V}_1$ consists of effective eigenvectors (corresponding to nonzero eigenvalues) of $\hat{\mathbf{H}}^H\hat{\mathbf{H}}$ and $\mathbf{\Lambda}_{FNS}$ is a diagonal matrix for power allocation. The decoder is obtained from (11) [or (35)]. Based on Remark 1 and Remark 2, this structure is asymptotically optimum when $\sigma_{ce}^2 \to 0$ (and, thus, the additive noise is the dominant source of error), or when $\mathbf{R}_T \to \mathbf{I}_{n_T}$. We refer to it as the noise-suppression structure.

On the other hand, we obtain another transceiver structure by restricting the precoder as $\mathbf{F}_{CC} = \mathbf{R}_T^{-1/2} \cdot \mathbf{V}_2 \cdot \mathbf{\Lambda}_{FCC}$, where $\mathbf{V}_2$ is composed of the effective eigenvectors of $\hat{\mathbf{H}}_{w0}^H\hat{\mathbf{H}}_{w0}$ and $\mathbf{\Lambda}_{FCC}$ is a diagonal power allocation matrix. Again, the decoder is determined from (11) [or (35)]. Based on Remark 2, this structure is asymptotically optimum when P_T/σ_n^2 goes to infinity so that channel estimation error becomes the dominant source of error, or when $\mathbf{R}_T \to \mathbf{I}_{n_T}$. We will refer to this as the correlation-cancelation structure.

It is interesting to compare the optimum precoder structure in Theorem 1 and the above two suboptimum ones. The performance comparisons shown in Fig. 3 remind us of the relationship between the matched filter (the noise-suppression precoder), the zero-forcing filter (the correlation-cancelation precoder) and the optimum linear MMSE filter in multiuser detection. From Fig. 3, we observe again the tremendous effect of channel estimation error on system performance.

The noise-suppression structure represents the direct application of previous results assuming perfect CSI at both ends [7], [8], [11] to the case with imperfect CSI. Clearly, this causes a large performance loss at medium to high SNR. We also notice that the correlation-cancelation structure has performance close to that of the optimum design, since it also utilizes the transmit correlation information. However, inverting the transmit correlation is not optimum at low SNR, and thus the correlation-cancelation structure performs differently from the optimum transmitter in a lower SNR region.

Example 4: Effect of transmit correlation versus effect of receive correlation [$\rho_T = 0, 0.5$; $\rho_R = 0, 0.5$; $\sigma_{ce}^2 = 0$, or $\sigma_{ce}^2 \neq 0$ with $P_{tr}/\sigma_n^2 = 26.016$ dB; $B = 3, 4$]

Fig. 4 is a comparison of the effects of transmit and receive correlation. It is clear that the same amount of transmit or receive correlation has exactly the same effect on system performance when the CSI is perfect.

In the channel estimation described in Section II-B, the knowledge of $\mathbf{R}_T$ is explicitly used in the training signal design, but the knowledge of $\mathbf{R}_R$ is not. This means the knowledge of $\mathbf{R}_T$ and $\mathbf{R}_R$ is not exploited the same way. Therefore, when there is channel estimation error, the same amount of correlation at the transmitter and the receiver affects system performance differently.[4]

Example 5: Effect of feedback delay and channel estimation error ($\rho_T = 0.7$, $\rho_R = 0$; $B = 3$; $\sigma_{ce}^2 \neq 0$ with $P_{tr}/\sigma_n^2 = 26.016$ dB)

We now simulate the effect of feedback delay on the joint precoder-decoder design. Both spatial and temporal channel correlation are considered. At any time instant, the spatial correlation is modeled in the same way as before. Assuming the Jakes' model [37], the temporal magnitude correlation of two channel realizations separated η seconds apart is given by $\rho = J_0^2(2\pi f_d \eta)$, where $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind, and f_d is the maximum Doppler frequency. As in [38], we consider the system working at a carrier frequency of $f_c = 2$ GHz. The data rate R_s on each data stream is set to be 400 kilo-symbols per second (ks/s), which implies

[4]As mentioned in Section II-B, a different CSI model appears in [22], in which the functions of $\mathbf{R}_T$ and $\mathbf{R}_R$ are symmetric. It is interesting to note that after applying our analysis with the CSI model used there, the effects of $\mathbf{R}_T$ and $\mathbf{R}_R$ are the same whether there is channel estimation error or not. The same comments also apply to the CSI model used in [39, Sec. VI].

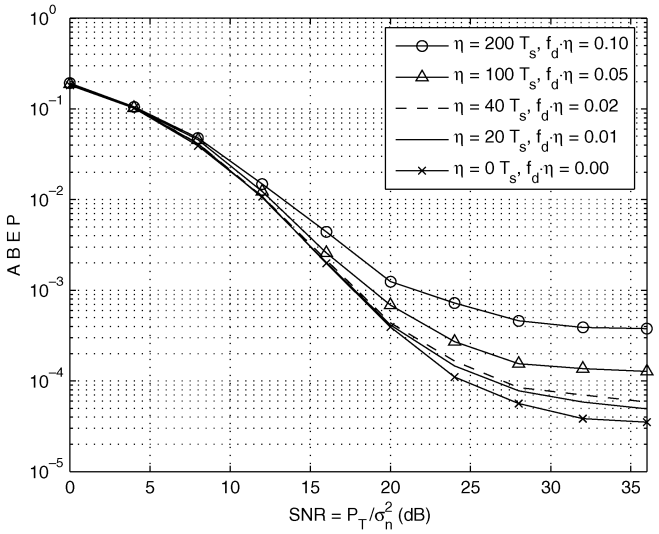

Fig. 5. The effect of feedback delay on system performance. $n_T = n_R = 4$, $B = 3$, $\rho_T = 0.7$, $\rho_R = 0.0$, $P_{tr}/\sigma_n^2 = 26.016$ dB (corresponding to $\sigma_E^2 = 0.0238$).

the symbol duration $T_s = 1/R_s = 2.5 \times 10^{-6}$ s. The terminal speed is $v = 30$ m/s. Then f_d is calculated to be 200 Hz.

At time t, the receiver obtains $\hat{\mathbf{H}}_t$, an estimate of the true channel $\mathbf{H}_t$. Here the subscript denotes the time index. Due to the feedback delay η, the transmitter only knows $\hat{\mathbf{H}}_{t-\eta}$, i.e., the estimate of $\mathbf{H}_{t-\eta}$. Therefore, at time t, the precoder $\mathbf{F}$ can only be calculated according to $\hat{\mathbf{H}}_{t-\eta}$. The decoder $\mathbf{G}$ is matched to the precoder, and is calculated from (11) or (35) (with $\hat{\mathbf{H}}$ there being replaced by $\hat{\mathbf{H}}_t$). The resulting performances for such a scenario are shown in Fig. 5 with different amounts of delay. As shown, as long as the normalized delay $f_d \cdot \eta$ is smaller than 0.01 (i.e., $\eta \leq 20T_s$ in the example above), the performance degradation is small. Note that $f_d \cdot \eta = 0.01$ is observed as a critical normalized delay also in [38], below which the system still operates satisfactorily. Thus, our system design is robust against reasonably small delays. Similar results can also be observed using other values of system parameters ($\rho_T, P_{tr}/\sigma_n^2, B$, etc.).

VI. Conclusion

We have formulated and solved a minimum total MSE transceiver design problem for MIMO systems with channel mean as well as both transmit and receive correlation information at both ends. The closed-form optimum precoder and decoder matrices are obtained. Our results gracefully fit those in the existing literature as channel estimation error diminishes. Simulation results are provided for the minimum total MSE design. We have observed that channel estimation error causes an error floor at high SNR and a large performance degradation across the whole SNR range. At medium to high SNR, this degradation can be compensated by introducing diversity. High correlation has a large impact on system performance as well. The same amounts of transmit and receive correlation impact the system performance equivalently under the assumption of perfect channel estimation, whereas under imperfect channel estimation, they might show different effects on the system performance, depending on the specific channel estimation method employed.

Appendix A
Derivations of (3) and (4)

Perform vectorization operation on (2) to obtain $vec(\tilde{\mathbf{H}}_w) = vec(\mathbf{H}_w) + (\mathbf{I}_{n_T} \otimes \mathbf{R}_R^{-1/2})vec(\mathbf{N}_0)$. Then the minimum MSE (MMSE) estimate of $vec(\mathbf{H}_w)$ is given by [35, p. 156, Eq. (IV.B.53)]

$$\begin{aligned} vec(\hat{\mathbf{H}}_w) &= \left[\mathbf{I}_{n_T n_R} + \mathbf{I}_{n_T} \otimes \mathbf{R}_R^{-1} \cdot \sigma_{ce}^2\right]^{-1} vec(\tilde{\mathbf{H}}_w) \\ &= \left\{\mathbf{I}_{n_T} \otimes \left[\mathbf{I}_{n_R} + \mathbf{R}_R^{-1} \cdot \sigma_{ce}^2\right]^{-1}\right\} vec(\tilde{\mathbf{H}}_w). \end{aligned}$$

Then (3) is obtained by converting $vec(\hat{\mathbf{H}}_w)$ back to its matrix version. The resulting estimation error covariance matrix is [35, p. 156, Eq. (IV.B.54)]

$$\begin{aligned} \mathbf{\Psi} &= E\left\{\left[vec(\mathbf{H}_w) - vec(\hat{\mathbf{H}}_w)\right]\left[vec(\mathbf{H}_w) - vec(\hat{\mathbf{H}}_w)\right]^H\right\} \\ &= \mathbf{I}_{n_T n_R} - \left[\mathbf{I}_{n_T n_R} + \sigma_{ce}^2 \cdot \left(\mathbf{I}_{n_T} \otimes \mathbf{R}_R^{-1}\right)\right]^{-1} \\ &= \mathbf{I}_{n_T} \otimes \left[\mathbf{I}_{n_R} - \left(\mathbf{I}_{n_R} + \sigma_{ce}^2 \mathbf{R}_R^{-1}\right)^{-1}\right] \\ &= \sigma_{ce}^2 \cdot \underbrace{\mathbf{I}_{n_T} \otimes \left[\mathbf{R}_R^{-1}\left(\mathbf{I}_{n_R} + \sigma_{ce}^2 \mathbf{R}_R^{-1}\right)^{-1}\right]}_{\mathbf{\Psi}_0}. \end{aligned}$$

The estimation error vector can be represented by $\mathbf{\Psi}_0^{1/2} vec(\mathbf{E}_w)$, where the entries of $vec(\mathbf{E}_w)$ are i.i.d. $\mathcal{N}_c(0, \sigma_{ce}^2)$ [15], [21], [22]. The matrix version of $\mathbf{\Psi}_0^{1/2} vec(\mathbf{E}_w)$ is given by $\mathbf{R}_R^{-1/2}(\mathbf{I}_{n_R} + \sigma_{ce}^2 \cdot \mathbf{R}_R^{-1})^{-1/2}\mathbf{E}_w$, and (4) follows. □

Appendix B
Existence of a Global Minimum for (9)

The problem in (9) can be equivalently formulated as [36, p. 130, Sec. 4.1.3]

$$\min_{\mathbf{F}, tr(\mathbf{F}\mathbf{F}^H) \leq P_T} \; \min_{\mathbf{G}(\mathbf{F})} tr\left\{\text{MSE}\left[\mathbf{F}, \mathbf{G}(\mathbf{F})\right]\right\}. \tag{34}$$

The minimizing $\mathbf{G}$ for the inner unconstrained minimization is readily shown to be

$$\mathbf{G} = \mathbf{F}^H \hat{\mathbf{H}}^H \cdot \left[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \sigma_{ce}^2 tr(\mathbf{R}_T \mathbf{F}\mathbf{F}^H)\mathbf{R}_{e,R} + \sigma_n^2 \mathbf{I}_{n_R}\right]^{-1} \tag{35}$$

which is the linear MMSE data estimator (Wiener filter) given $\hat{\mathbf{H}}$ and $\mathbf{F}$ [35]. Substituting (35) into (34), the problem in (9) can be equivalently formulated as

$$\min_{\mathbf{F}, tr(\mathbf{F}\mathbf{F}^H) \leq P_T} tr\left\{\text{MSE}(\mathbf{F})\right\} \tag{36}$$

where

$$\begin{aligned} \text{MSE}(\mathbf{F}) = \Big\{&\mathbf{I}_B + \mathbf{F}^H \hat{\mathbf{H}}^H \\ &\times \left[\sigma_{ce}^2 tr(\mathbf{R}_T \mathbf{F}\mathbf{F}^H)\mathbf{R}_{e,R} + \sigma_n^2 \mathbf{I}_{n_R}\right]^{-1} \hat{\mathbf{H}}\mathbf{F}\Big\}^{-1}. \end{aligned}$$

The feasible set of (36) is $\{\mathbf{F} | tr(\mathbf{F}\mathbf{F}^H) \leq P_T\}$, a (closed and bounded) Frobenius norm ball of radius $\sqrt{P_T}$ [36, p. 30, Sec. 2.2.3]. This means the feasible set is compact [26, p. 653, Proposition A.6 (g)]. The objective function of (36) is continuous at all points of the feasible set. Thus, according to Weierstrass' Theorem [26, p. 654, Proposition A.8], there exists a global minimum for the problem given by (36). Since (9) and

(36) are equivalent, the same global minimum also exists for (9) [36, p. 130, Sec. 4.1.3]. In addition, the minimizing $\mathbf{F}$ is the same for both problems, while the minimizing $\mathbf{G}$ for (9) is calculated according to (35). □

APPENDIX C
PROOF OF LEMMA 1

Proof: Premultiplying both sides of (11) by $\mathbf{G}$, we obtain

$$\mathbf{G}\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H+\left[\sigma_n^2+\sigma_E^2\cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)\right]\cdot\mathbf{G}\mathbf{G}^H=\mathbf{G}\hat{\mathbf{H}}\mathbf{F}. \tag{37}$$

Similarly, postmultiplying both sides of (12) by $\mathbf{F}$, to get

$$\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H\mathbf{G}\hat{\mathbf{H}}\mathbf{F}+\sigma_E^2 tr(\mathbf{G}\mathbf{G}^H)\mathbf{F}^H\mathbf{R}_T\mathbf{F}+\mu_1\mathbf{F}^H\mathbf{F}=\mathbf{G}\hat{\mathbf{H}}\mathbf{F}. \tag{38}$$

Clearly, $\mathbf{G}\hat{\mathbf{H}}\mathbf{F}$ must be Hermitian. From (37) and (38)

$$\left[\sigma_n^2+\sigma_E^2\cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)\right]\mathbf{G}\mathbf{G}^H = \sigma_E^2\cdot tr(\mathbf{G}\mathbf{G}^H)\cdot\mathbf{F}^H\mathbf{R}_T\mathbf{F}+\mu_1\cdot\mathbf{F}^H\mathbf{F}.$$

Taking the traces at both sides, we obtain $\sigma_n^2\cdot tr(\mathbf{G}\mathbf{G}^H)=\mu_1\cdot tr(\mathbf{F}^H\mathbf{F})$. Due to (14), if $\mu_1>0$, $tr(\mathbf{F}^H\mathbf{F})$ must be equal to P_T, which yields $\mu_1=\sigma_n^2\cdot tr(\mathbf{G}\mathbf{G}^H)/P_T$, the desired result. Assume $\mathbf{F}\neq 0$ [i.e., $tr(\mathbf{F}\mathbf{F}^H)>0$]. Then, if $\mu_1=0$, we must have $tr(\mathbf{G}\mathbf{G}^H)=0$, and $\mu_1=\sigma_n^2\cdot tr(\mathbf{G}\mathbf{G}^H)/P_T$ is still valid. Finally, it is easy to see that it holds even when $\mathbf{F}=0$, $\mathbf{G}=0$, and $\mu_1=0$. This concludes the proof of Lemma 1. □

APPENDIX D
PROOF OF LEMMA 2

Proof: Assumed $B=r$. Let the matrix $\mathbf{F}$ be expressed as

$$\begin{aligned}\mathbf{F}&=\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\left[\mathbf{V}\ \tilde{\mathbf{V}}\right]\left[\mathbf{\Lambda}_F^H\ \tilde{\mathbf{\Lambda}}_F^H\right]^H\\&=\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\left[\mathbf{V}\mathbf{\Lambda}_F+\tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}_F\right]\end{aligned} \tag{39}$$

where $\mathbf{V}$ and $\tilde{\mathbf{V}}$ come from (15), and $\mathbf{\Lambda}_F$ and $\tilde{\mathbf{\Lambda}}_F$ are arbitrary $r\times r$ and $(n_T-r)\times r$ matrices, respectively. Since $[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]$ is of full rank, $[\mathbf{V}\ \tilde{\mathbf{V}}]$ is a $n_T\times n_T$ unitary matrix and $[\mathbf{\Lambda}_F^H\ \tilde{\mathbf{\Lambda}}_F^H]^H$ denotes an arbitrary $n_T\times r$ matrix, (39) is a general expression for $\mathbf{F}$. Define $\mathbf{F}_{||}=\mathbf{V}\mathbf{\Lambda}_F$ and $\mathbf{F}_{\perp}=\tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}_F$. It can be verified that

$$\mathbf{F}_{\perp}^H\mathbf{F}_{||}=0,\ \ \mathbf{F}_{||}^H\mathbf{F}_{\perp}=0,\ \ \hat{\mathbf{H}}\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]^{-\frac{1}{2}}\mathbf{F}_{\perp}=0. \tag{40}$$

By using Lemma 1, (12) can be rewritten as

$$\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H\mathbf{G}\hat{\mathbf{H}}+\frac{tr(\mathbf{G}\mathbf{G}^H)}{P_T}\mathbf{F}^H\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]=\mathbf{G}\hat{\mathbf{H}}. \tag{41}$$

Postmultiplying both sides of (41) by $[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]^{-1/2}\mathbf{F}_{\perp}$, using (40), we get $\mathbf{F}_{\perp}^H\mathbf{F}_{\perp}=0$, i.e., $\mathbf{F}_{\perp}=0$. This means $\mathbf{F}=[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]^{-1/2}\mathbf{V}\mathbf{\Lambda}_F$ and (16) holds. Now let $\tau_1=\sigma_n^2+\sigma_E^2\cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)$. The matrix $\mathbf{G}$ satisfying (11) is given by

$$\begin{aligned}\mathbf{G}&=\mathbf{F}^H\hat{\mathbf{H}}^H\left[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H+\tau_1\cdot\mathbf{I}_{n_R}\right]^{-1}\\&=[\mathbf{F}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{F}+\tau_1\cdot\mathbf{I}_r]^{-1}\mathbf{F}^H\hat{\mathbf{H}}^H\\&=\mathbf{J}\mathbf{V}^H\left[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}\right]^{-1/2}\hat{\mathbf{H}}^H\end{aligned}$$

where the second equation is obtained using the matrix inversion lemma [27], the third equation is obtained by substituting (16), and $\mathbf{J}\stackrel{\text{def}}{=}[\mathbf{F}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{F}+\tau_1\cdot\mathbf{I}_r]^{-1}\mathbf{\Lambda}_F^H$ is an arbitrary $r\times r$ matrix. Let $\mathbf{\Lambda}_G$ denote an arbitrary $r\times r$ matrix. The general $\mathbf{G}$ satisfying (11) can thus be expressed as $\mathbf{G}=\mathbf{\Lambda}_G\mathbf{V}^H[\sigma_E^2\cdot P_T\cdot\mathbf{R}_T+\sigma_n^2\cdot\mathbf{I}_{n_T}]^{-1/2}\hat{\mathbf{H}}^H$. Therefore, (17) holds. □

APPENDIX E
DERIVING (18)–(23) AND DETERMINING k IN THEOREM 1

Assume $\mathbf{F}\neq 0$. We first show how to obtain (18)–(23). Postmultiplying both sides of (41) [which is from (12)] by $\mathbf{F}$, we get

$$\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H\mathbf{G}\hat{\mathbf{H}}\mathbf{F}+\frac{tr(\mathbf{G}\mathbf{G}^H)}{P_T}\mathbf{F}^H\times\left[\sigma_E^2P_T\mathbf{R}_T+\sigma_n^2\mathbf{I}_{n_T}\right]\mathbf{F}=\mathbf{G}\hat{\mathbf{H}}\mathbf{F}. \tag{42}$$

Let $\tau_1=\sigma_n^2+\sigma_E^2\cdot tr(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)$ as in Appendix D. Per Lemma 2, substituting (16) and (17) into (37) and (42), using Lemma 1, we obtain the following identities, respectively:

$$\begin{aligned}\mathbf{\Lambda}_G\mathbf{\Lambda}\mathbf{\Lambda}_F\mathbf{\Lambda}_F^H\mathbf{\Lambda}\mathbf{\Lambda}_G^H+\tau_1\mathbf{\Lambda}_G\mathbf{\Lambda}\mathbf{\Lambda}_G^H&=\mathbf{\Lambda}_G\mathbf{\Lambda}\mathbf{\Lambda}_F\\\mathbf{\Lambda}_F^H\mathbf{\Lambda}\mathbf{\Lambda}_G^H\mathbf{\Lambda}_G\mathbf{\Lambda}\mathbf{\Lambda}_F+\frac{\mu_1}{\sigma_n^2}\mathbf{\Lambda}_F^H\mathbf{\Lambda}_F&=\mathbf{\Lambda}_G\mathbf{\Lambda}\mathbf{\Lambda}_F.\end{aligned}$$

Based on the above two equations, the optimum $\mathbf{\Lambda}_F$ and $\mathbf{\Lambda}_G$ can be shown to be diagonal without loss of generality as in [8], and are given by $\mathbf{\Lambda}_{Fopt}$ and $\mathbf{\Lambda}_{Gopt}$ in (20) and (21), respectively. Finally, insert (18)–(21) into $tr\{\mathbf{F}\mathbf{F}^H\}=P_T$ and $\mu_1=\sigma_n^2\cdot tr(\mathbf{G}\mathbf{G}^H)/P_T$, to obtain two equations with τ_1 and μ_1 being the unknown variables. Solving these two equations, we can find τ_1 and μ_1 as given by (22) and (23), respectively.

We further describe an iterative procedure for calculating the number k in Theorem 1, which is similar to that in [8]. Let λ_l be the lth entry on the main diagonal of $\mathbf{\Lambda}$ ($l=1,\ldots,B=r$). Recall that the diagonal elements of $\mathbf{\Lambda}$ are arranged in decreasing order. Initialize $k=B$.

1) Calculate τ_1 and μ_1 from (22) and (23), respectively. If $\mu_1\leq\lambda_k\sigma_n^2/\tau_1$, stop; else: go to step 2).
2) Let $\mathbf{\Lambda}_{Fopt,k}:=0$ and $k:=k-1$. Go to step 1).

□

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments on the paper.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun. (ETT)*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, 1996.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[4] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Sel. Areas. Commun.*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.
[5] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[6] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.
[7] 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.
[8] 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, no. 12, pp. 2198–2206, Dec. 2001.
[9] L. Collin, O. Berder, P. Rostaing, and G. Burel, "Optimal minimum distance-based precoder for MIMO spatial multiplexing systems," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 617–627, Mar. 2004.
[10] Z. Yan, K. M. Wong, and Z.-Q. Luo, "Optimal diagonal precoder for multiantenna communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2089–2100, Jun. 2005.
[11] 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.
[12] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 214–226, Jan. 2004.
[13] J. Zhang, Y. Wu, S. Zhou, and J. Wang, "Joint linear transmitter and receiver design for the downlink of multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 991–993, Nov. 2005.
[14] 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.
[15] S. Serbetli and A. Yener, "MMSE transmitter design for correlated MIMO systems with imperfect channel estimates: Power allocation trade-offs," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2295–2304, Aug. 2006.
[16] N. Khaled, G. Leus, C. Desset, and H. De Man, "A robust joint linear precoder and decoder MMSE design for slowly time-varying MIMO channels," in *Proc. IEEE ICASSP 2004*, May 17–21, 2004, vol. 4, pp. 485–488.
[17] X. Zhang, D. P. Palomar, and B. Ottersten, "Robust design of linear MIMO transceiver for low SNR," in *Proc. 39th Asilomar Conf. Signals, Syst. Comput.*, Oct. 28–Nov. 1 2005, pp. 398–402.
[18] X. Zhang, D. P. Palomar, and B. Ottersten, "Robust design of linear MIMO transceiver under channel uncertainty," in *Proc. ICASSP 2006*, May 14–19, 2006, vol. 4, pp. 77–80.
[19] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Univ. Politécnica de Catalunya (UPC), Barcelona, Spain, May 2003.
[20] 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, June 2003.
[21] 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.
[22] 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, no. 5, pp. 2798–2806, Sep. 2007.
[23] 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, no. 3, pp. 502–513, Mar. 2000.
[24] B. Hassibi and B. M. 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] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. 45th IEEE Veh. Technol. Conf. (VTC)*, Jul. 25–28, 1995, vol. 2, pp. 815–819.
[26] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. New York: Athena Scientific, 1999.
[27] H. Lutkepohl, *Handbook of Matrices*. New York: Wiley, 1996.
[28] M. Ding and S. D. Blostein, "MIMO LMMSE transceiver design with imperfect CSI at both ends," in *Proc. IEEE Int. Conf. Commun. (ICC) 2007*, Glasgow, Scotland, Jun. 2007, pp. 4369–4374.
[29] N. Khaled, S. Thoen, and L. Deneire, "Optimizing the joint transmit and receive MMSE design using mode selection," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 730–737, Apr. 2005.
[30] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[31] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?," in *Proc. IEEE Globecom 2004*, Dallas, TX, Dec. 2004, pp. 96–100.
[32] M. Medard, "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.
[33] M. Ding, "Multiple-input multiple-out wireless system designs with imperfect channel knowledge," Ph.D. thesis, Queen's Univ., Kingston, Ontario, Canada, 2008.
[34] A. B. Gershman and N. Sidiropoulos, *Space-Time Processing for MIMO Communications*. New York: Wiley, 2005.
[35] H. V. Poor, *Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[37] *Microwave Mobile Communications*, W. C. Jakes, Ed. New York: Wiley, 1974.
[38] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, Jul. 2004.
[39] 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.

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, Ontario, Canada, in 2008, all in electrical engineering.

From 2002 to 2003, she was a Research Engineer with Datang Mobile Communications Equipment Co., Ltd., Beijing. From 2003 to 2008, she was a Research and Teaching Assistant with the Department of Electrical and Computer Engineering, Queen's University. She is currently a Research Associate with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. 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 of Queen's University, Kingston, ON, Canada, since 1988 and currently holds the position of Professor and Head of the Department of Electrical and Computer Engineering. From 1999 to 2003, he was 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 image compression, target tracking, radar imaging, and wireless communications. He spent 1994–1995 with Lockheed Martin Electronic Systems in Montreal and 2006 with the Communications Research Centre, Ottawa, Canada. His current interests lie in the application of signal processing to wireless communications systems, including smart antennas, MIMO systems, and space-time-frequency processing for MIMO-OFDM systems.

Dr. Blostein has been a member of the Samsung 4G Wireless Forum, as well as an invited distinguished speaker at Ryerson University and at the Samsung Advanced Institute of Technology. He served as Chair of IEEE Kingston Section (1994), Chair of the Biennial Symposium on Communications (2000, 2006, 2008), an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996–2000), and Publications Chair for IEEE ICASSP 2004. He is currently serving as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a registered Professional Engineer in Ontario.