A general framework for MIMO transceiver design with imperfect CSI and transmit correlation

Steven Blostein, Minhua Ding

2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications

A General Framework for MIMO Transceiver Design with Imperfect CSI and Transmit Correlation

Minhua Ding†, Steven D. Blostein‡, Wai Ho Mow†, Constantin Siriteanu#
†Hong Kong University of Science and Technology, Hong Kong, P. R. China
‡Queen's University, Kingston, Ontario, Canada #Seoul National University, Seoul, Korea
Email: †{eemding, eewhmow}@ust.hk, ‡steven.blostein@queensu.ca, #costi@cse.snu.ac.kr

***Abstract*— Assuming perfect channel state information (CSI), linear precoding/decoding for multiple-input multiple-output (MIMO) systems has been considered in the literature under various performance criteria, such as minimum total mean-square error (MSE), maximum mutual information, and minimum average bit error rate (BER). It has been shown that these criteria belong to a set of reasonable Schur-concave or Schur-convex objective functions of the diagonal entries of the system mean-square error (MSE) matrix. In this paper, assuming only the knowledge of *channel mean and transmit correlation* at both ends, a general theoretical framework is presented to derive the optimum precoder and decoder for MIMO systems using these objective functions. It is shown that for all these objective functions the optimum transceivers share a similar structure. Compared to the case with perfect CSI, a linear filter is added to both ends to balance the suppression of channel noise and the additional noise induced from channel estimation error. Simulation results are provided.**[1]

I. Introduction

The performance of multiple-input multiple-output (MIMO) systems depends on the availability of channel state information (CSI) at the transmitter (CSIT) and/or at the receiver (CSIR) [1]. Previously, optimum precoding or joint precoding/decoding for MIMO spatial multiplexing systems has been obtained using mean-square error (MSE)-related design criteria under different CSI assumptions [2]-[13]. Assuming perfect CSI at both ends, optimum transceivers are derived for minimizing total MSE or for maximizing capacity [2][3][4]. In [5], assuming perfect CSI, the optimum transceivers are obtained for a set of MSE, signal-to-interference-plus-noise (SINR), or bit error rate (BER)-related design criteria, which are Schur-convex or Schur-concave functions of the diagonal entries of the MIMO system MSE matrix and include the minimum total MSE and maximum capacity design criteria as special cases.

CSI is imperfect in practice, and there have been robust designs which take this fact into account. Transceiver optimization has been considered assuming *perfect CSIR* and imperfect CSIT (channel mean and/or channel covariance information) (see [6, Sec. VII], and references therein). In [7][8], the same imperfect CSI is assumed at both ends of a MIMO link without explicit consideration of channel correlation. In [9][6, Sec. VI], transceiver designs have been studied assuming, at both ends, the imperfect CSI composed of *channel mean and receive correlation information*. In [10][11][13], optimum signaling for a capacity lower-bound (i.e., minimizing the determinant of the system MSE matrix [13]) has been studied assuming imperfect CSI at both ends with *channel mean and transmit correlation information*, where the closed-form transmit covariance matrix has been found in [13]. Under the same CSI assumption, optimum transceivers to minimize the total MSE (trace of the system MSE matrix) have been found in [12, Sec. III][13]. It is worth pointing out that, with *channel mean and transmit correlation information at both ends*, the transceiver optimization problem is nontrivial compared to the perfect CSI case.

In this paper, we consider the MIMO transceiver design with channel mean and transmit correlation information at both ends as in [12][13]. This scenario is particularly interesting in practical downlink transmissions, where the channels arising from base station antennas are correlated. With this assumption of CSI, the optimum precoder-decoder pairs for the minimum total MSE design and the maximum capacity lower-bound design have been derived based on the associated optimality conditions [12][13]. However, this approach involves matrix differentiation and has to be applied individually for different objective functions. On the other hand, the optimum transceivers derived share the same structure, which implies that a unified approach might be possible.

In light of the results from [5], here we present *a general theoretical framework* to derive the optimum transceivers for various practical designs (as summarized in [5], including those in [12][13] as special cases) under the same imperfect CSI. The approach taken here is to equivalently reformulate the original design problem using the notion of "reasonable functions", and then apply majorization theory [5][17]. We obtain the optimum transceiver for the whole set of design criteria which are Schur-convex or Schur-concave functions of the diagonal entries of the MIMO system MSE matrix. Assuming imperfect CSI, the analysis can also be extended for transceiver optimization for MIMO-OFDM systems using cyclic prefix (CP) and without subcarrier cooperation.

Notation: $\mathbb{E}\{\cdot\}$ stands for statistical expectation, $\mathrm{tr}(\cdot)$ for trace, and $\det(\cdot)$ for determinant. $(\cdot)^H$ means complex conjugate transpose (Hermitian). $\mathbf{A} \succ \mathbf{B}$ means that $(\mathbf{A} - \mathbf{B})$ is positive definite. $(b)_+ = \max(b, 0)$. $\mathcal{N}_c(\cdot, \cdot)$ denotes the complex Gaussian distribution. $\mathbf{I}$ is the identity matrix and $\mathbf{1}$

[1]The work in this paper was supported by the Hong Kong Research Grants Councils under project number 617087.

is reserved for the all-one vector. diag($\mathbf{A}$) and eig($\mathbf{A}$) denote vectors whose entries are the diagonal entries and eigenvalues of a positive semidefinite matrix $\mathbf{A}$, respectively. For square $\mathbf{B}$, $[\mathbf{B}]_{ii}$ denotes the i-th diagonal entry of $\mathbf{B}$.

II. System model and problem formulation

A. System model

It is assumed that n_T (n_R) antennas are used at the transmitter (receiver). The information streams to be sent are denoted by a $B \times 1$ vector $\mathbf{s}$, where the number of data streams, B ($\leq n_T$), is chosen and fixed. A $n_T \times B$ precoder, denoted by $\mathbf{F}$, is employed at the transmitter, taking the available CSI into account. After precoding, the data vector is transmitted across a slowly-varying flat-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

$$\mathbf{y} = \mathbf{HFs} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is the AWGN with distribution $\mathcal{N}_c(0, \sigma_n^2 \cdot \mathbf{I})$. The input signal $\mathbf{s}$ is assumed to be zero-mean and white ($\mathbf{R_{ss}} = \mathbf{I}$), and independent of channel realizations. In the receiver, a linear decoder, described by the $B \times n_R$ matrix $\mathbf{G}$, is employed to recover the original information. After decoding, the signal vector $\mathbf{r}$ is given by $\mathbf{r} = \mathbf{Gy} = \mathbf{G}(\mathbf{HFs} + \mathbf{n})$.

The MIMO channel is modeled as in [14]: $\mathbf{H} = \mathbf{H}_w \mathbf{R}_T^{1/2}$, where $\mathbf{H}_w$ is a matrix whose entries are independent and identically distributed (i.i.d.) $\mathcal{N}_c(0, 1)$. The matrix $\mathbf{R}_T$ represents normalized transmit correlation with diagonal entries all equal to one. We assume that $\mathbf{R}_T$ is invertible.

B. Description of the CSI

As in [10][12], MMSE estimation of $\mathbf{H}_w$ is performed at the receiver, which yields $\mathbf{H}_w = \hat{\mathbf{H}}_w + \mathbf{E}_w$, with $\hat{\mathbf{H}}_w$ being the estimate of $\mathbf{H}_w$ and $\mathbf{E}_w$ being the error matrix. $\hat{\mathbf{H}}_w$ and $\mathbf{E}_w$ are mutually uncorrelated, and are both spatially white with entries $\mathcal{N}_c(0, 1 - \sigma_E^2)$ and $\mathcal{N}_c(0, \sigma_E^2)$, respectively. Variance $\sigma_E^2 = \mathbb{E}\{|\mathbf{H}_{wji}|^2\} - \mathbb{E}\{|\hat{\mathbf{H}}_{wji}|^2\}$. The CSI model is thus described by $\mathbf{H} = (\hat{\mathbf{H}}_w + \mathbf{E}_w)\mathbf{R}_T^{1/2} = \hat{\mathbf{H}} + \mathbf{E}$, where $\hat{\mathbf{H}} = \hat{\mathbf{H}}_w \mathbf{R}_T^{1/2}$ is the estimated channel matrix (channel mean) and $\mathbf{E} = \mathbf{E}_w \mathbf{R}_T^{1/2}$. Below we assume that $\hat{\mathbf{H}}$, $\mathbf{R}_T$, σ_E^2 and σ_n^2 are known to both ends of the link, which is also referred to as *channel mean and transmit correlation information.* It is assumed that CSIT is obtained by perfect feedback of CSIR via a dedicated link.

With the above CSI model, the received signal vector $\mathbf{y}$ is given by $\mathbf{y} = \hat{\mathbf{H}}\mathbf{Fs} + \mathbf{EFs} + \mathbf{n} = \hat{\mathbf{H}}\mathbf{Fs} + \mathbf{E}_w \mathbf{R}_T^{1/2}\mathbf{Fs} + \mathbf{n}$, and $\mathbf{r} = \mathbf{Gy}$. The system MSE matrix is calculated as

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

Note that $\mathbb{E}\left\{\mathbf{E}_w \mathbf{A} \mathbf{E}_w^H\right\} = \sigma_E^2 \cdot \text{tr}(\mathbf{A}) \cdot \mathbf{I}$, if the entries of matrix $\mathbf{E}_w$ are i.i.d. $\mathcal{N}_c(0, \sigma_E^2)$. The optimum linear MMSE data estimator [15] is used at the receiver, i.e.,

$$\mathbf{G}_{opt} = \mathbf{F}^H\hat{\mathbf{H}}^H\{\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + [\sigma_n^2 + \sigma_E^2\text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)]\mathbf{I}\}^{-1}. \tag{3}$$

Substituting (3) into (2), we obtain the MSE matrix in terms of $\mathbf{F}$ alone:

$$\text{MSE}(\mathbf{F}) = \left[\mathbf{I}_B + \frac{\mathbf{F}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{F}}{\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)}\right]^{-1}. \tag{4}$$

C. Problem formulation

Our goal here is to find the optimum $\mathbf{F}$ which minimizes a set of reasonable[2] Schur-convex or Schur-concave objective functions [denoted as $g(\cdot)$] [5] of the diagonal entries of $\text{MSE}(\mathbf{F})$ subject to a total power constraint:

$$\min_{\mathbf{F}} g\left(\text{diag}[\text{MSE}(\mathbf{F})]\right), \text{ subject to } \text{tr}(\mathbf{F}\mathbf{F}^H) \leq P_T. \tag{5}$$

It can be shown that a global minimum exists for continuous g functions, since the feasible set is a finite-dimension Frobenius norm ball [16]. Based on the optimized $\mathbf{F}$, we can evaluate the performance of different designs with imperfect CSI. When $\sigma_E^2 = 0$, the problem formulation in (5) reduces to that in [5], or those in [2][3][4] when the objective function is the trace or determinant of $\text{MSE}(\mathbf{F})$. Furthermore, when $\sigma_E^2 \neq 0$ and the objective function is the trace or determinant of $\text{MSE}(\mathbf{F})$, the optimum $\mathbf{F}$ has also been determined in [12][13]. However, the methodology used in [12][13] depends on the differentiation of the objective function with respect to the precoder and decoder matrices, and has to be applied to each objective function individually. Here we will provide a general framework to find the optimum $\mathbf{F}$ for a set of objective functions (different g's) without matrix differentiation.

III. General results

For convenience, define

$$\mathbf{T} = [\sigma_n^2 \cdot \mathbf{I}_{n_T} + \sigma_E^2 \cdot P_T \cdot \mathbf{R}_T]. \tag{6}$$

Below we assume that the number of data stream, B, is equal to r, the rank of the estimated channel $\hat{\mathbf{H}}$.

A. General results

Proposition 1: Assume that $g : \mathcal{R}_+^B \to \mathcal{R}$ is reasonable (i.e., it is an increasing function in each of its arguments).

- If g is Schur-concave, then the optimum $\mathbf{F}$ for (5) is given by:

$$\mathbf{F} = [\sigma_n^2 \cdot \mathbf{I}_{n_T} + \sigma_E^2 \cdot P_T \cdot \mathbf{R}_T]^{-\frac{1}{2}}\mathbf{V}\mathbf{\Phi}_{F1}, \tag{7}$$

where $\mathbf{\Phi}_{F1}$ is a diagonal matrix satisfying the power constraint with equality, and $\mathbf{V}$ is obtained from the following eigen-value decomposition:

$$\mathbf{T}^{-\frac{1}{2}}\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{T}^{-\frac{1}{2}} = [\mathbf{V}\ \tilde{\mathbf{V}}]\begin{pmatrix}\mathbf{\Lambda} & 0 \\ 0 & \tilde{\mathbf{\Lambda}}\end{pmatrix}[\mathbf{V}\ \tilde{\mathbf{V}}]^H. \tag{8}$$

In (8), $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal entries are the non-zero eigenvalues arranged in decreasing order. $\tilde{\mathbf{\Lambda}}$ is a zero matrix, and $\tilde{\mathbf{V}}$ consists of basis vectors of the

[2] A function $g : \mathcal{R}_+^B \to \mathcal{R}$ is reasonable if it is increasing in each of its arguments [5]. This definition fits in the context of linear precoding/decoding design for MIMO systems.

null space. $\mathbf{V}$ is composed of eigenvectors corresponding to the nonzero eigenvalues.

- If g is Schur-convex, then the optimum $\mathbf{F}$ for (5) is of the form:

$$\mathbf{F} = [\sigma_n^2 \cdot \mathbf{I}_{n_T} + \sigma_E^2 \cdot P_T \cdot \mathbf{R}_T]^{-\frac{1}{2}} \mathbf{V} \mathbf{\Phi}_{F2} \mathbf{U}, \tag{9}$$

where $\mathbf{\Phi}_{F2}$ is diagonal, and $\mathbf{U}$ is a unitary matrix chosen to make the diagonal entries of the resulting MSE($\mathbf{F}$) equal.

Proof: First, we show in Appendix A that if g is reasonable, then the minimum of (5) is achieved when the constraint is satisfied with equality, i.e., $\text{tr}(\mathbf{F}\mathbf{F}^H) = P_T$. Then (5) can be equivalently formulated as

$$\min_{\mathbf{F}} \; g\left(\text{diag}\left[\mathbf{I} + \frac{P_T \cdot \mathbf{F}^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{F}}{\text{tr}\{\mathbf{F}^H[\sigma_n^2 \mathbf{I}_{n_T} + \sigma_E^2 P_T \mathbf{R}_T]\mathbf{F}\}}\right]^{-1}\right)$$
$$\text{subject to } \; \text{tr}(\mathbf{F}\mathbf{F}^H) = P_T. \tag{10}$$

Without loss of generality, $\mathbf{F}$ can be expressed as

$$\begin{aligned} \mathbf{F} &= \mathbf{T}^{-\frac{1}{2}} [\mathbf{V} \;\; \tilde{\mathbf{V}}][\mathbf{\Phi}_F^H \;\; \tilde{\mathbf{\Phi}}_F^H]^H \\ &= \mathbf{T}^{-\frac{1}{2}} [\mathbf{V}\mathbf{\Phi}_F + \tilde{\mathbf{V}}\tilde{\mathbf{\Phi}}_F], \end{aligned} \tag{11}$$

where $\mathbf{V}$ and $\tilde{\mathbf{V}}$ are both from (8), and $\mathbf{\Phi}_F$ and $\tilde{\mathbf{\Phi}}_F$ are ***arbitrary*** $r \times r$ and $(n_T - r) \times r$ matrices, respectively. Define $\mathbf{F}_{\parallel} = \mathbf{V}\mathbf{\Phi}_F$ and $\mathbf{F}_{\perp} = \tilde{\mathbf{V}}\tilde{\mathbf{\Phi}}_F$. It is shown in Appendix B that, to achieve the minimum, $\mathbf{F}_{\perp} = 0$, i.e., $\mathbf{F} = \mathbf{T}^{-1/2}\mathbf{V}\mathbf{\Phi}_F$. Substituting this into (10), after some algebra, we can show that (10) is equivalent to

$$\min_{\mathbf{\Phi}_F} \; g\left(\text{diag}\left[\mathbf{I} + \frac{P_T \mathbf{\Phi}_F^H \mathbf{\Lambda} \mathbf{\Phi}_F}{\text{tr}\{\mathbf{\Phi}_F^H \mathbf{\Phi}_F\}}\right]^{-1}\right)$$
$$\text{subject to } \; \text{tr}(\mathbf{\Phi}_F^H \mathbf{V}^H \mathbf{T}^{-1} \mathbf{V} \mathbf{\Phi}_F) = P_T, \tag{12}$$

where $\mathbf{V}$ and $\mathbf{\Lambda}$ are from (8).

From (12), it is the structure of $\mathbf{\Phi}_F$ that determines the value of the objective function. The norm of $\mathbf{\Phi}_F$ does not affect it.

To proceed, we need the following results from [5][17]. Let $\mathbf{M}$ be a $n \times n$ positive semidefinite matrix. Let the entries of diag($\mathbf{M}$) and eig($\mathbf{M}$) be arranged in decreasing order, respectively. Then diag($\mathbf{M}$) is majorized by eig($\mathbf{M}$), and so is $\frac{\text{tr}(\mathbf{M})}{n}\mathbf{1}$ by diag($\mathbf{M}$). If $f : \mathcal{R}^n \to \mathcal{R}$ is Schur-concave, then $f[\text{eig}(\mathbf{M})] \le f[\text{diag}(\mathbf{M})]$. If f is Schur-convex, then $f[\frac{\text{tr}(\mathbf{M})}{n}\mathbf{1}] \le f[\text{diag}(\mathbf{M})]$.

Thus, for a Schur-concave g, the minimum of (12) is achieved when $\mathbf{\Phi}_F$ is diagonal with its diagonal entries properly arranged, denoted as $\mathbf{\Phi}_{F1}$. This gives us (7)[3].

On the other hand, for a Schur-convex g, the minimum is achieved when all the diagonal entries are made equal and their sum [trace of the matrix appeared in the objective of (12) as the argument of the diag function] is minimized. Since the trace function is itself a Schur-concave objective function of the diagonal entries of the MSE matrix, the optimum $\mathbf{\Phi}_F$ for (12) has the form $\mathbf{\Phi}_{F2}\mathbf{U}$, where $\mathbf{\Phi}_{F2}$ is diagonal and $\mathbf{U}$ is a unitary matrix which renders the diagonal entries of the resulting MSE matrix equal. Therefore, (9) is proved. □

[3]It can be shown that if a non-diagonal matrix $\mathbf{\Phi}_F$ (satisfying the power constraint) achieves a certain value of the Schur-concave objective function in (12), then there exists a diagonal matrix (satisfying the power constraint) that achieves a value of the objective function no greater than that achieved by the non-diagonal matrix. Therefore, the optimum $\mathbf{\Phi}_F$ for (12) must be diagonal. Due to space limitations, we do not elaborate on this here.

Remark 1: **Proposition 1** reduces to the results in [2][3][4][5] when $\sigma_E^2 = 0$. Compared to the perfect CSI case, a linear filter $\mathbf{T}^{-\frac{1}{2}}$ [see (6)] is added in the transceiver, which balances the suppression of channel noise and the additional noise caused by channel estimation error. The effect of σ_E^2 is coupled with transmit correlation $\mathbf{R}_T$. When **Proposition 1** is applied to (5), it remains to determine the entries of $\mathbf{\Phi}_{F1}$ or those of $\mathbf{\Phi}_{F2}$, and the original matrix optimization problem (5) is now scalarized.

B. Applications

We consider examples of g which satisfy the requirements of **Proposition 1**. For brevity, define $[\mathbf{V}^H\mathbf{T}^{-1}\mathbf{V}]_{ii} = \beta_i, i = 1, \ldots, r = B$. Also recall that the entries of $\mathbf{\Lambda}$ [see (8)] are arranged **in decreasing order**.

i. Examples of Schur-concave functions

Let $[\mathbf{\Phi}_{F1}]_{ii} = \phi_i, i = 1, \ldots, r$. [$\mathbf{\Phi}_{F1}$ is defined in (7).] Define $x_i = \phi_i^2$.

(a) Minimization of weighted arithmetic mean of the MSEs

Let $g_1(\{[\text{MSE}(\mathbf{F})]_{ii}\}_{i=1}^r) = \sum_{i=1}^r (w_i \cdot [\text{MSE}(\mathbf{F})]_{ii})$, where $\{w_i\}_{i=1}^r$ are positive weights. This objective function has been considered in [4] assuming perfect CSI, which incudes the unweighted MMSE design and the maximum capacity design as special cases. By choosing different weights, one can also design the transceiver to achieve different SNRs on different subchannels [4]. Clearly, g_1 is reasonable. Per [5], g_1 is Schur-concave. The problem in (5) is scalarized as

$$\min_{\{x_i\}_{i=1}^r} \sum_{i=1}^r w_i \cdot \frac{1}{1 + \frac{P_T \lambda_i x_i}{\sum_{m=1}^r x_m}}$$
$$\text{subject to } \; \sum_{i=1}^r x_i \beta_i = P_T, \;\; x_i \ge 0, \forall i. \tag{13}$$

Solving this problem using the method of Lagrange multipliers, we obtain

$$x_i = \left[\frac{w_i^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}} P_T (P_T + a_1) - a_2 P_T \lambda_i^{-1}}{(P_T + a_1) a_3 - a_2 a_4}\right]_+. \tag{14}$$

Let the integer k ($k \le r$) denote the number of non-zero $x_i's$. Note that k can be readily determined using a procedure as in [12]. Then $a_1 = \sum_{i=1}^k \lambda_i^{-1}$, $a_2 = \sum_{i=1}^k \lambda_i^{-\frac{1}{2}}$, $a_3 = \sum_{i=1}^k \lambda_i^{-\frac{1}{2}} \beta_i$ and $a_4 = \sum_{i=1}^k \lambda_i^{-1} \beta_i$. Note that this result coincides with that obtained in [13, Sec. 3.6] using a different approach.

After obtaining x_i [as in (13)], we can obtain $\phi_i = \sqrt{x_i}, \forall i$, and thus $\mathbf{\Phi}_{F1}$.

(b) Minimization of the geometric mean of the MSEs

Let $g_2(\{[\text{MSE}(\mathbf{F})]_{ii}\}_{i=1}^r) = \Pi_{i=1}^r [\text{MSE}(\mathbf{F})]_{ii}^{w_i}$. This design problem is shown to be related to the minimization of the determinant of the MSE matrix (or maximization of the mutual information) with perfect CSI [5]. Again, g_2 is reasonable. It is shown in [5] that g_2 is Schur-concave. The problem in (5) is reduced to

$$\min_{\{x_i\}_{i=1}^r} \Pi_{i=1}^r \left[\frac{1}{1+\frac{P_T\lambda_i x_i}{\sum_{m=1}^r x_m}}\right]^{w_i}$$
$$\text{subject to } \sum_{i=1}^r x_i\beta_i = P_T, \quad x_i \geq 0, \forall i. \tag{15}$$

Solving this problem, we obtain

$$x_i = \left[\frac{P_T\{w_i(P_T+b_3) - b_0\lambda_i^{-1}\}}{(P_T+b_3)b_1 - b_2 b_0}\right]_+. \tag{16}$$

Let the integer m ($m \leq r$) denote the number of non-zero $x_i's$. Similar to k for (14), m can be readily determined (see [13]). Then $b_1 = \sum_{i=1}^m \beta_i$, $b_2 = \sum_{i=1}^m \lambda_i^{-1}\beta_i$, $b_3 = \sum_{i=1}^m \lambda_i^{-1}$, and $b_0 = \sum_{i=1}^m w_i$. In this case, $[\mathbf{\Phi}_{F1}]_{ii} = \sqrt{x_i}$.

The optimum precoder obtained here is the same as that used to maximize a mutual information *lower-bound* [13][10][11] when $w_i = 1, \forall i$:

$$\max_{\substack{\mathbf{F} \\ \text{tr}\{\mathbf{F}\mathbf{F}^H\}\leq P_T}} \log_2 \det\left[\mathbf{I} + \frac{\mathbf{F}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{F}}{\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)}\right]. \tag{17}$$

The relationship between the minimization of the geometric mean and (17) is similar to that in the perfect CSI case, where minimizing the geometric mean is equivalent to maximizing the *exact* mutual information [5].

Remark 2: [Relationship between minimization of the unweighted geometric mean of MSEs (maximization of the capacity *lower-bound* in (17)) and minimization of the weighted arithmetic mean of MSEs] Let $w_i = \lambda_i, \forall i$ in (13) and let $w_i = 1, \forall i$ in (15), then (14) is equal to (16).

ii. Schur-convex functions

Schur-convex functions are involved in many system designs of interest, e.g., minimization of the maximum MSE from all data streams [5]. For all reasonable Schur-convex functions, the solutions to (5) are the same[4]. Further, the optimum $\mathbf{\Phi}_{F2}$ in (9) is the same as the optimum $\mathbf{\Phi}_{F1}$ in (7) to minimize the trace of MSE($\mathbf{F}$), which has been obtained when minimizing the arithmetic mean of the MSEs with $w_i = 1, \forall i$ [see (13) (14) with $w_i = 1, \forall i$].

Remark 3: For single-carrier MIMO, despite different design criteria for particular applications, the minimizations of the weighted arithmetic mean and geometric mean of the MSEs [Sec. III-B–(**i**)] are the core of all designs. This is true for both perfect and imperfect CSI cases.

iii. Extension to CP-based MIMO-OFDM systems with imperfect channel estimation and transmit correlation

Assuming imperfect channel estimation, it is straightforward to extend our results in Sec. III-A to a CP-based MIMO-OFDM system [19] with individual processing[5] and power constraints on each subcarrier.

On the other hand, assuming individual processing, when a sum power constraint is imposed on all subcarriers, we employ a two-stage processing (primal decomposition). First, we initialize the power for each subcarrier, and apply our results on transceiver optimization to each subcarrier. Then an outer power allocation is performed among all subcarriers. Iteration is performed until the globally optimum transceivers are obtained for all subcarriers.

Note that the outer power optimization problem is nontrivial in the case of imperfect CSI, and remains to be solved. However, the structures of the optimum precoders can be readily shown.

IV. Numerical Results

Per **Remark 3**, for single-carrier MIMO, all the designs discussed here are related to minimization of the arithmetic or geometric mean of the MSEs. Thus we refer the readers to [12][13], where numerical examples for the results in Sec. III-B–(**i**) of this paper can be found. For single-carrier MIMO, the corresponding BER results for all Schur-convex functions are the same. An example is given below (see Fig. 1). Let $n_T = n_R = 4$, the number of data streams $B = 3$. The transmit correlation model is given by: $(\mathbf{R}_T)_{ij} = \rho^{|i-j|}$ for $i, j \in \{1, \ldots, n_T\}$. Here $\rho = 0.5$. The SNR in Fig. 1 is defined as P_T/σ_n^2. QPSK (4-QAM) is used for each data stream. The system performance is shown in terms of the arithmetic mean of BERs (ABER $= \frac{1}{B}\sum_{j=1}^B \text{BER}_j$), and is obtained from Monte Carlo simulations. For the imperfect CSI case, the error variance is modeled in the same way as in [12, Sec. II-B, Sec. III], and is set to be $\sigma_E^2 = 0.01478$ for $\rho = 0.5$. The two designs shown in Fig. 1 differ only in a unitary rotation, and the comparison results are as they are designed to be.

V. Conclusions

Assuming channel mean and transmit correlation information at both ends, optimum transceiver structures for MIMO systems have been determined for a set of reasonable Schur-convex or Schur-concave objective functions of the diagonal entries of the system MSE matrix. Compared to the case with perfect CSI, a linear filter is added to both ends to balance the suppression of channel noise and the additional noise induced from channel estimation error. Results can also be applied to the transceiver design assuming imperfect CSI for CP-based MIMO-OFDM systems with noncooperative subcarriers.

[4]It is important to note that our design (5) takes an averaging approach [see (2)]. In addition, we should be cautious when applying some design criteria with imperfect CSI. For example, in the design to minimize the arithmetic mean of BERs from all data streams, by simulations, we have found that, with imperfect CSI, $\mathcal{Q}$-function cannot be used to describe the BER of each data stream at high SNR. This is different from the perfect CSI case [5]. Therefore, some Schur-convex (or Schur-concave) functions established in the perfect CSI case have to be re-examined.

[5]This means that the subcarriers are not cooperating. Each subcarrier has its own transceiver (precoder-decoder pair).

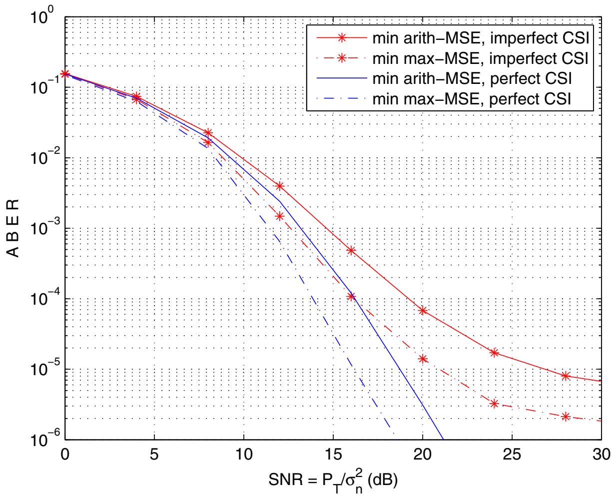

Fig. 1. ABER performance; $\sigma_E^2 = 0$ (perfect CSI) or $\sigma_E^2 = 0.01478$ (imperfect CSI). $n_T = n_R = 4$, $B = 3$, $\rho = 0.5$. Here min arith-MSE (or max-MSE) denotes the design to minimize the arithmetic mean (or the maximum) of the MSEs from all data streams. Note that max-MSE is Schur-convex.

Appendix A

The entries of g are the diagonal entries of MSE($\mathbf{F}$) and can be represented as $\mathbf{e}_i^H[\text{MSE}(\mathbf{F})]\mathbf{e}_i$, where $\mathbf{e}_i$ is the i-th column of the identity matrix, $i = 1, \ldots, B$. Suppose that $\mathbf{F}_A$ is the optimum of (5) when the power constraint is P_A and denote corresponding minimum of (5) as v_A. Let $P_B > P_A$ and $\breve{\mathbf{F}}_B = \sqrt{\frac{P_B}{P_A}}\mathbf{F}_A$. Then,

$$\begin{aligned}
&\mathbf{e}_i^H \left[\mathbf{I} + \frac{\breve{\mathbf{F}}_B^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \breve{\mathbf{F}}_B}{\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T \breve{\mathbf{F}}_B \breve{\mathbf{F}}_B^H)}\right]^{-1} \mathbf{e}_i \\
&= \mathbf{e}_i^H \left[\mathbf{I} + \frac{\mathbf{F}_A^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{F}_A}{\frac{P_A}{P_B}\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T \mathbf{F}_A \mathbf{F}_A^H)}\right]^{-1} \mathbf{e}_i \\
&< \mathbf{e}_i^H \left[\mathbf{I} + \frac{\mathbf{F}_A^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{F}_A}{\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T \mathbf{F}_A \mathbf{F}_A^H)}\right]^{-1} \mathbf{e}_i, \quad \forall i. \qquad (18)
\end{aligned}$$

The last inequality follows from the fact that if $\mathbf{A} \succ \mathbf{B}$, then $\mathbf{B}^{-1} \succ \mathbf{A}^{-1}$ [18, p. 586, A.8, (vii)], and $\mathbf{e}_i^H(\mathbf{B}^{-1} - \mathbf{A}^{-1})\mathbf{e}_i > 0$. Consequently, if $\mathbf{A} \succ \mathbf{B}$, $\mathbf{e}_i^H\mathbf{A}^{-1}\mathbf{e}_i < \mathbf{e}_i^H\mathbf{B}^{-1}\mathbf{e}_i$. Denote the value of the objective in (5) corresponding to $\breve{\mathbf{F}}_B$ as $\breve{v}_B$. Since g is increasing in each of its arguments, based on (18), we have $\breve{v}_B < v_A$. Define the global minimum of (5) corresponding to P_B as v_B. Clearly, $v_B \leq \breve{v}_B$, since v_B is the global minimum whereas $\breve{v}_B$ is simply the cost of using one feasible point. Therefore, $v_B < v_A$ for $P_B > P_A$. This shows that the minimum of (5) must be achieved when the constraint is satisfied with equality. □

Appendix B

Due to space limitations, we can only present an outline here. Substituting (11) into (10), using the fact that $\mathbf{F}_\perp^H\mathbf{F}_\parallel = 0$, $\mathbf{F}_\parallel^H\mathbf{F}_\perp = 0$, $\mathbf{F}_\perp^H\mathbf{V} = 0$, $\mathbf{F}_\parallel^H\mathbf{F}_\parallel = \mathbf{\Phi}_F^H\mathbf{\Phi}_F$ and $\mathbf{F}_\perp^H\mathbf{F}_\perp = \tilde{\mathbf{\Phi}}_F^H\tilde{\mathbf{\Phi}}_F$, after some calculations, we obtain the following equivalent problem

$$\begin{aligned}
&\min_{\mathbf{\Phi}_F, \tilde{\mathbf{\Phi}}_F} \; g\left(\text{diag}\left[\mathbf{I} + \frac{P_T \cdot \mathbf{\Phi}_F^H \mathbf{\Lambda} \mathbf{\Phi}_F / \text{tr}\{\mathbf{\Phi}_F^H\mathbf{\Phi}_F\}}{1 + \text{tr}\{\tilde{\mathbf{\Phi}}_F^H\tilde{\mathbf{\Phi}}_F\}/\text{tr}\{\mathbf{\Phi}_F^H\mathbf{\Phi}_F\}}\right]^{-1}\right) \\
&\text{subject to } \; \text{tr}\left[\mathbf{T}^{-1}(\mathbf{F}_\parallel + \mathbf{F}_\perp)(\mathbf{F}_\parallel + \mathbf{F}_\perp)^H\right] = P_T. \qquad (19)
\end{aligned}$$

Using the same technique as in Appendix A, one can show that the objective in (19) is decreased if $\text{tr}\{\tilde{\mathbf{\Phi}}_F^H\tilde{\mathbf{\Phi}}_F\}$ is decreased (assuming that g is reasonable, i.e., increasing in each of its arguments). Therefore, the objective is minimized when $\text{tr}\{\tilde{\mathbf{\Phi}}_F^H\tilde{\mathbf{\Phi}}_F\} = 0$, i.e., $\mathbf{F}_\perp = 0$. □

References

[1] A. Paulraj, R. Nabar, D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[2] J. Yang, 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.

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

[4] H. Sampath, P. Stoica, 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.

[5] D. P. Palomar, J. M. Cioffi, M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381-2401, Sept. 2003.

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

[7] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Technical University of Catalonia (UPC), Barcelona, Spain, May 2003.

[8] T. Yoo, A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Info. Theory*, vol. 52, no. 5, pp. 2203-2214, May 2006.

[9] S. Serbetli, 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.

[10] T. Yoo, A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?" in *Proc. IEEE Global Telecomm. Conf. 2004*, pp. 96-100, Dallas, Texas, Dec. 2004.

[11] L. Musavian, M. R. Nakhai, M. Dohler, 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, Sept. 2007.

[12] M. Ding, S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 1141-1150, Mar. 2009.

[13] M. Ding, "Multiple-input multiple-output wireless system designs with imperfect channel knowledge," Ph.D. thesis, Queen's University, Kingston, Ontario, Canada, July 2008.

[14] D. Shiu, G. J. Foschini, M. J. Gans, 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.

[15] H. V. Poor, *Introduction to Signal Detection and Estimation*, Springer-Verlag, second edition, 1994.

[16] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, second edition, 1999.

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

[18] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, 1982.

[19] E. G. Larsson, P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, 2003.