

# MIMO LMMSE Transceiver Design with Imperfect CSI at Both Ends

Minhua Ding and Steven D. Blostein

Department of Electrical and Computer Engineering

Queen's University, Kingston, Ontario, Canada, K7L 3N6

Email: 2md18@qmlink.queensu.ca, steven.blostein@queensu.ca

**Abstract**—This paper presents a new result on minimum total mean squared error (MSE) joint transceiver design for multiple-input multiple-output (MIMO) systems, with imperfect channel state information (CSI) at both ends and subject to a total power constraint. The channel knowledge here is the channel mean and transmit correlation information. The joint design is formulated into an optimization problem, to which the closed-form optimum solution is found. The optimum precoder consists of a linear filter, a matrix which collects effective channel eigenmodes, and a diagonal power allocation matrix. The linear filter balances the suppression of channel noise and the additional noise induced by channel estimation error. Given the precoder, the optimum decoder is simply a linear minimum MSE (LMMSE) data estimator. The performance degradations due to imperfect channel estimation and/or transmit correlation are demonstrated by simulation results. The relation between the minimum total MSE design and the maximum mutual information design is determined as well, under the above mentioned imperfect CSI.

## I. INTRODUCTION

Previously, MIMO spatial multiplexing systems with joint linear processing at both ends have been studied from different perspectives under the minimum total MSE design criterion subject to a total power constraint [1]-[8].

In [2]-[3] and [4], Chapter 5, the optimum precoder and decoder with perfect CSI at both ends are derived. The optimum precoder suggests transmission along the effective (non-zero) channel eigenmodes with a proper diagonal power allocation across them.

In practice, CSI is often imperfect and robust designs which consider this fact are desired. In [5], the minimum total MSE design for MIMO systems has been studied with outdated CSI at the transmitter and perfect CSI at the receiver. Also in [6] and [7], imperfect CSI (covariance information only or channel mean and covariance information) is considered at the transmitter. The receiver, however, has perfect CSI. In [4], Chapter 7, and [8], channel estimation error at both ends has been taken into account, but no results have been obtained when there is channel correlation. More recently, in [9], channel estimation error as well as receive correlation information has been considered in the transceiver design. However, the optimum design with channel mean and transmit correlation information at both ends remains as an open problem.

In this paper, we address linear minimum total MSE precoding/decoding with channel mean and transmit correlation information at both ends of a MIMO link. We formulate this

transceiver design as an optimization problem. Our problem formulation is different from those in [5]-[8], due to different CSI assumptions. The optimum solution is found in closed form. In particular, the optimum precoder is composed of a linear filter, a matrix composed of effective channel eigenmodes and a diagonal power allocation matrix, whereas the optimum decoder is simply a LMMSE data estimator. To check our derivations of the closed-form solution, an iterative algorithm is introduced, which is an alternative to solving the optimization problem. Simulation results are presented, which verify our analysis. The maximum mutual information design is closely related to the minimum total MSE design [2][3]. With channel mean and transmit correlation information at both ends, the maximum mutual information design has been formulated as a capacity lower bound problem in [11]. We introduce the optimum closed-form transmit covariance matrix for this capacity lower bound. We also determine the relation between the maximum mutual information design and the minimum total MSE design under imperfect CSI.

Notation: Upper (lower) case boldface letters are for matrices (vectors).  $E\{\cdot\}$  denotes statistical expectation and  $tr(\cdot)$  stands for trace.  $|\mathbf{A}|$  means determinant of matrix  $\mathbf{A}$ , whereas  $|a|$  denotes the magnitude of complex scalar  $a$ .  $(\cdot)^*$  means complex conjugate and  $(\cdot)^H$  means complex conjugate transpose (Hermitian).  $(b)_+ = \max(b, 0)$ .  $\mathcal{N}_c(\cdot, \cdot)$  denotes the complex Gaussian distribution.  $\mathbf{I}$  is the identity matrix.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System model

We address linear precoding/decoding for a MIMO link. We assume that  $n_T$  antennas are used at the transmitter and  $n_R$  antennas are used at the receiver. The information streams 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 CSI into account. After the precoder, 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

$$\mathbf{y} = \mathbf{H}\mathbf{F}\mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\mathbf{n}$  is the AWGN with distribution  $\mathcal{N}_c(0, \sigma_n^2 \mathbf{I})$ . The input signal  $\mathbf{x}$  is assumed to be zero-mean and white ( $\mathbf{R}_{\mathbf{x}\mathbf{x}} = \mathbf{I}$ ), 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 the decoder, the signal vector  $\mathbf{r}$  is given by

$$\mathbf{r} = \mathbf{G}\mathbf{y} = \mathbf{G}(\mathbf{H}\mathbf{F}\mathbf{x} + \mathbf{n}).$$

We use the channel model as in [10], i.e.,  $\mathbf{H} = \mathbf{H}_w \mathbf{R}_T^{1/2}$ , 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. No receive correlation is considered.

### B. CSI model

As in [11], 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  $\sigma_E^2 = E\{|\mathbf{H}_{wji}|^2\} - E\{|\hat{\mathbf{H}}_{wji}|^2\}$ . Therefore, the CSI model is 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  $\mathbf{E} = \mathbf{E}_w \mathbf{R}_T^{1/2}$ . CSI at the transmitter is obtained either by using channel reciprocity (if available) and performing estimation at the transmitter, or by feedback of CSI at the receiver via a perfect link.

Below we assume that  $\hat{\mathbf{H}}$ ,  $\mathbf{R}_T$ ,  $\sigma_E^2$  and  $\sigma_n^2$  are known to both ends of the link. The CSI here is also referred to as channel mean and transmit correlation information.

### C. Problem formulation

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

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{F}\mathbf{x} + \underbrace{\mathbf{E}\mathbf{F}\mathbf{x} + \mathbf{n}}_{\text{total noise}} \quad (2)$$

The system mean squared error (MSE) matrix is calculated as

$$\begin{aligned} MSE(\mathbf{F}, \mathbf{G}) &= E\{(\mathbf{r} - \mathbf{x})(\mathbf{r} - \mathbf{x})^H\} \\ &= E\left\{[\mathbf{G}(\hat{\mathbf{H}} + \mathbf{E})\mathbf{F} - \mathbf{I}][\mathbf{G}(\hat{\mathbf{H}} + \mathbf{E})\mathbf{F} - \mathbf{I}]^H\right\} \\ &\quad + \sigma_n^2 \cdot \mathbf{G}\mathbf{G}^H. \end{aligned} \quad (3)$$

After some manipulations, we can show that (3) reduces to

$$\begin{aligned} MSE(\mathbf{F}, \mathbf{G}) &= \mathbf{G}\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H - \mathbf{G}\hat{\mathbf{H}}\mathbf{F} - \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H \\ &\quad + \mathbf{I} + [\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)]\mathbf{G}\mathbf{G}^H. \end{aligned} \quad (4)$$

In the above derivations, we have used the result  $E\{\mathbf{E}_w\mathbf{A}\mathbf{E}_w^H\} = \sigma_E^2 \cdot \text{tr}(\mathbf{A}) \cdot \mathbf{I}$ , if the entries of matrix  $\mathbf{E}_w$  are i.i.d.  $\mathcal{N}_c(0, \sigma_E^2)$ , and also the identity  $\text{tr}(\mathbf{A}_1\mathbf{A}_2) = \text{tr}(\mathbf{A}_2\mathbf{A}_1)$ .

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 (s.t.) a total power constraint  $P_T$ , i.e.,

$$\min_{\mathbf{F}, \mathbf{G}} \text{tr}\{MSE(\mathbf{F}, \mathbf{G})\}, \quad \text{s.t. } \text{tr}[\mathbf{F}\mathbf{F}^H] \leq P_T. \quad (5)$$

This is referred to as the minimum total MSE design. Furthermore, we are interested in finding out the effects of transmit correlation  $\mathbf{R}_T$  and channel estimation error  $\sigma_E^2$  on system performance.

The problem in (5) can be equivalently formulated as [12]

$$\min_{\mathbf{F}} \min_{\mathbf{G}(\mathbf{F})} \text{tr}[MSE(\mathbf{F}, \mathbf{G}(\mathbf{F}))], \quad \text{s.t. } \text{tr}[\mathbf{F}\mathbf{F}^H] \leq P_T. \quad (6)$$

The optimum  $\mathbf{G}$  as a function of  $\mathbf{F}$  can be found by treating (2) as a linear Bayesian model and then calculating the LMMSE data estimator given  $\hat{\mathbf{H}}$ , as shown below<sup>1</sup> [18]

$$\begin{aligned} \mathbf{G}_{LMMSE} \\ = \mathbf{F}^H \hat{\mathbf{H}}^H \left\{ \hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + [\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)] \cdot \mathbf{I} \right\}^{-1}. \end{aligned} \quad (7)$$

Substituting (7) into (6), the problem in (5) can be alternatively formulated as:

$$\min_{\mathbf{F}} \text{tr}\{MSE(\mathbf{F})\}, \quad \text{s.t. } \text{tr}[\mathbf{F}\mathbf{F}^H] \leq P_T, \quad (8)$$

where

$$MSE(\mathbf{F}) = \left[ \mathbf{I} + \frac{\mathbf{F}^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{F}}{\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)} \right]^{-1}. \quad (9)$$

Note that the optimum precoder and decoder pair is not unique. From (8), if  $\mathbf{F}_{opt}$  minimizes the total MSE, so does  $\mathbf{F}_{opt}\mathbf{U}$ , where  $\mathbf{U}$  is any  $B \times B$  unitary matrix. Below we refer to  $\mathbf{F}_{opt}$  as the optimum precoder up to a unitary transform.

When  $\sigma_E^2 = 0$ , problem formulations in (5) and (8) reduce to those in [2][3] or [4], Chapter 5. When  $\sigma_E^2 \neq 0$  and there is transmit correlation, previous results do not directly apply.

## III. OPTIMUM SOLUTION TO THE MINIMUM TOTAL MSE DESIGN

### A. Closed-form optimum solution

The Lagrangian associated with (5) is [12]

$$\mathcal{L}(\mathbf{F}, \mathbf{G}, \mu_1) = \text{tr}\{MSE(\mathbf{F}, \mathbf{G})\} + \mu_1 \cdot [\text{tr}\{\mathbf{F}\mathbf{F}^H\} - P_T],$$

where  $\mu_1$  is the Lagrangian multiplier. By taking the derivative of  $\mathcal{L}(\mathbf{F}, \mathbf{G}, \mu_1)$  with respect to  $\mathbf{F}^*$  and  $\mathbf{G}^*$  [19], respectively, the associated Karush-Kahn-Tucker (KKT) conditions can be obtained as follows:

$$\left\{ \hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + [\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T\mathbf{F}\mathbf{F}^H)] \cdot \mathbf{I} \right\} \mathbf{G}^H = \hat{\mathbf{H}}\mathbf{F}, \quad (10)$$

$$\mathbf{F}^H \left[ \hat{\mathbf{H}}^H \mathbf{G}^H \mathbf{G} \hat{\mathbf{H}} + \sigma_E^2 \cdot \text{tr}(\mathbf{G}\mathbf{G}^H) \cdot \mathbf{R}_T + \mu_1 \cdot \mathbf{I} \right] = \mathbf{G} \hat{\mathbf{H}}, \quad (11)$$

$$\mu_1 \geq 0, \text{tr}(\mathbf{F}\mathbf{F}^H) - P_T \leq 0, \quad (12)$$

$$\mu_1 \cdot [\text{tr}(\mathbf{F}\mathbf{F}^H) - P_T] = 0. \quad (13)$$

<sup>1</sup>This explains the title of this paper. We stress that it is the LMMSE data estimator instead of the MMSE estimator, since the total noise in (2) may not be Gaussian, even if the channel noise is.

We will solve (5) based on (10)-(13). Note that (7) can also be obtained from (10).

**Lemma 1:**  $\mu_1 = \sigma_n^2 \cdot \text{tr}(\mathbf{G}\mathbf{G}^H)/P_T$ .

**Proof:** We pre-multiply (10) by  $\mathbf{G}$  to get a new equation. We also post-multiply (11) by  $\mathbf{F}$  to get another equation. Taking the traces of both sides of these two new equations, and relying on (12)-(13), we can show that **Lemma 1** is true. Detailed calculations are omitted due to space limitation.

Consider the following eigen-value decomposition

$$\begin{aligned} & [\sigma_E^2 P_T \mathbf{R}_T + \sigma_n^2 \mathbf{I}]^{-1/2} \hat{\mathbf{H}}^H \hat{\mathbf{H}} [\sigma_E^2 P_T \mathbf{R}_T + \sigma_n^2 \mathbf{I}]^{-1/2} \\ &= (\mathbf{V} \tilde{\mathbf{V}}) \begin{pmatrix} \Lambda & 0 \\ 0 & \tilde{\Lambda} \end{pmatrix} (\mathbf{V} \tilde{\mathbf{V}})^H. \end{aligned} \quad (14)$$

Let  $r$  be equal to the rank of the matrix in (14), i.e.,  $r = \text{rank}(\Lambda)$ , the number of non-zero channel eigenmodes. Here the  $n_T \times (n_T - r)$  matrix  $\tilde{\mathbf{V}}$  consists of basis vectors for the null space of (14). The entries of the diagonal matrix  $\tilde{\Lambda}$  are all zero. The matrix  $\mathbf{V}$  is  $n_T \times r$ . Without loss of generality, we can assume that the entries of the diagonal matrix  $\Lambda$  are arranged in a non-increasing order.

**Lemma 2:** Assume that the number of data streams  $B$  is equal to  $r$ . Without loss of generality, the optimum precoder and decoder can be expressed as:

$$\mathbf{F} = [\sigma_E^2 \cdot P_T \cdot \mathbf{R}_T + \sigma_n^2 \cdot \mathbf{I}]^{-1/2} \mathbf{V} \Lambda_F, \quad (15)$$

$$\mathbf{G} = \Lambda_G \mathbf{V}^H [\sigma_E^2 \cdot P_T \cdot \mathbf{R}_T + \sigma_n^2 \cdot \mathbf{I}]^{-1/2} \hat{\mathbf{H}}^H, \quad (16)$$

where  $\Lambda_F$  and  $\Lambda_G$  are arbitrary  $B \times B$  matrices and  $\mathbf{V}$  comes from (14).

**Proof:** We use a method modified from that in [3]. The idea is to write the expressions of  $\mathbf{F}$  and  $\mathbf{G}$  in their general forms first, and then show that they can be expressed in simpler forms, as given by (15) and (16), respectively, without loss of generality. Details are omitted due to lack of space.

**Theorem 1:** Assume that the number of data streams  $B$  is equal to  $r$ . Without loss of generality, the optimum precoder and decoder for (5) have the following expressions, respectively:

$$\mathbf{F}_{opt} = [\sigma_E^2 \cdot P_T \cdot \mathbf{R}_T + \sigma_n^2 \cdot \mathbf{I}]^{-1/2} \mathbf{V} \Lambda_{Fopt}, \quad (17)$$

$$\mathbf{G}_{opt} = \Lambda_{Gopt} \mathbf{V}^H [\sigma_E^2 \cdot P_T \cdot \mathbf{R}_T + \sigma_n^2 \cdot \mathbf{I}]^{-1/2} \hat{\mathbf{H}}^H. \quad (18)$$

The  $B \times B$  matrices  $\Lambda_{Fopt}$  and  $\Lambda_{Gopt}$  are diagonal and are given below

$$\Lambda_{Fopt} = \left[ \tau_1^{1/2} \mu_1^{-1/2} \sigma_n \cdot \Lambda^{-1/2} - \tau_1 \cdot \Lambda^{-1} \right]_+^{1/2}, \quad (19)$$

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

Here

$$\tau_1 = \frac{a_2 \cdot P_T}{P_T \cdot a_3 + a_1 \cdot a_3 - a_2 \cdot a_4}, \quad (21)$$

$$\mu_1 = \frac{a_2 \cdot \sigma_n^2 (P_T \cdot a_3 + a_1 \cdot a_3 - a_2 \cdot a_4)}{(P_T + a_1)^2 \cdot P_T}. \quad (22)$$

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

**Proof:** We provide an outline of proof here. Based on **Lemma 1** and **Lemma 2**, using a method similar to that in [3], it can be shown that  $\Lambda_{Fopt}$  and  $\Lambda_{Gopt}$  are diagonal without loss of generality, and are given by (19) and (20), respectively. Here  $\tau_1 = \sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T \mathbf{F} \mathbf{F}^H)$ . Inserting (17)-(20) into the definition of  $\tau_1$  and the formula of  $\mu_1$  in **Lemma 1**, we obtain two equations with  $\tau_1$  and  $\mu_1$  as the variables. Solving these two equations, we can find  $\tau_1$  and  $\mu_1$  as given by (21) and (22), respectively.

We now describe the method to determine the number  $k$  in **Theorem 1**. Let  $\lambda_m$  be the  $m$ -th element of  $\Lambda$ . Recall that the diagonal elements of  $\Lambda$  are arranged in decreasing order. Initialize  $k = B$ .

- 1) Calculate  $\tau_1$  from (21) and  $\mu_1$  from (22), respectively. If  $\mu_1 \leq \lambda_k \sigma_n^2 / \tau_1$ , stop; else: go to step 2).
- 2) Let  $\Lambda_{Fopt,k} := 0$  and  $k := k - 1$ . Go to step 1).

This completes the outline of proof for **Theorem 1**.

**Remark 1:** When  $\sigma_E^2 = 0$ , **Theorem 1** reduces to the results in [2]-[4]. Compared with the results obtained under perfect CSI, a linear filter is added in the transmitter here, which balances the suppression of channel noise and the additional noise caused by channel estimation error.

**Remark 2:** When  $\mathbf{R}_T = \mathbf{I}$ , transmission along the eigenmodes of  $\hat{\mathbf{H}}^H \hat{\mathbf{H}}$  is optimum. Furthermore, the channel estimation error simply contributes additional noise ( $\sigma_E^2 \cdot P_T$ ). This result has been mentioned in [4], Chapter 7 and [9].

**Remark 3:** When  $P_T/\sigma_n^2$  goes to infinity, the filter  $[\sigma_E^2 \cdot P_T \cdot \mathbf{R}_T + \sigma_n^2 \cdot \mathbf{I}]^{-1/2}$  becomes a scaled  $\mathbf{R}_T^{-1/2}$ , i.e., the optimum precoder asymptotically cancels the effect of  $\mathbf{R}_T$ .

**Remark 4:** If the number of data streams  $B$  is chosen to be strictly smaller than the number of non-zero channel eigenmodes, the  $B$  strongest eigenmodes are used. In this case redundancy is introduced, which can be translated into improved diversity and thus performance improvement [3][16].

It can be shown that the **received signal-to-total-noise ratio matrix** is given by  $\Gamma = \mathbf{F}^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{F} / [\sigma_n^2 + \sigma_E^2 \cdot \text{tr}(\mathbf{R}_T \mathbf{F} \mathbf{F}^H)]$ , whose diagonal entries are the signal-to-total-noise ratios for all data streams. Note that  $MSE(\mathbf{F}) = [\mathbf{I} + \Gamma]^{-1}$ .

**Remark 5:** The analysis here can be directly applied to the minimum weighted MSE design [3] with channel mean and transmit correlation information at both ends.

*B. An iterative algorithm to obtain the optimum solution*

The KKT conditions (10)-(13) naturally lend themselves to an iterative algorithm for solving (5), which alternately updates  $\mathbf{F}$  and  $\mathbf{G}$ .<sup>2</sup> We introduce it here as a method to check our previous derivations. This algorithm is described below.

- 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 (10);
- 3) Update  $\mu_1$  from **Lemma 1**;
- 4) Update  $\mathbf{F}$  using (11);
- 5) If  $\text{tr}[(\mathbf{F}_i - \mathbf{F}_{i-1})(\mathbf{F}_i - \mathbf{F}_{i-1})^H]$  is sufficiently small<sup>3</sup> (say, less than  $10^{-3}$ ), stop; otherwise, go back to 2).

Note that any solution to the KKT conditions (10)-(13) must satisfy **Theorem 1** and leads to the same minimum total MSE. Furthermore, the above algorithm is guaranteed to converge [15]. Thus the solution from the iterative algorithm must be optimum and equivalent to the one from **Theorem 1**, up to a unitary transform.

IV. RELATION TO THE MAXIMUM MUTUAL INFORMATION DESIGN

The maximum mutual information (capacity) design is closely related to the minimum total MSE design [2]. In [11], it has been studied under imperfect CSI. With imperfect CSI, however, exact formulas for capacity are difficult to obtain. Instead, tight upper- or lower-bounds are often employed for system design. A capacity lower-bound has been formulated in Section IV-B of [11] with channel mean and transmit correlation information at both ends. Mathematically, the problem there is equivalent to minimizing the log-determinant of (9) subject to a total power constraint, i.e.,

$$\max_{\mathbf{Q}} \log_2 \left| \mathbf{I} + \frac{\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H}{\sigma_n^2 + \sigma_E^2 \text{tr}(\mathbf{R}_T \mathbf{Q})} \right|, \text{ s.t. } \text{tr}(\mathbf{Q}) \leq P_T, \tag{23}$$

where  $\mathbf{Q} = \mathbf{F}\mathbf{F}^H$ . In the above we have used the result  $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$ . In [11], the optimum closed-form  $\mathbf{Q}$  for (23) was thought to be unattainable and was not reported. However, it does exist and can be obtained using a method similar to that for the minimum total MSE design. The optimum covariance matrix can be written as  $\mathbf{Q}_{opt} = \mathbf{F}_{c,opt} \mathbf{F}_{c,opt}^H$ , where  $\mathbf{F}_{c,opt}$  has the same structure as  $\mathbf{F}_{opt}$  in (17), i.e., it also consists of

<sup>2</sup>In fact, similar algorithms have been developed for minimum total MSE designs in MIMO multiuser uplink [14] and downlink [15] communications. However, our calculation of the Lagrangian multiplier ( $\mu_1$ ) is much simpler and more reliable, based on **Lemma 1**, and is different from those in [14] or [15]. Simpler expressions for the Lagrangian multipliers in both [14] and [15] can be derived as well.

<sup>3</sup>Here  $\mathbf{F}_i$  ( $\mathbf{F}_{i-1}$ ) denotes  $\mathbf{F}$  in the  $i$ -th ( $(i-1)$ -st) iteration.

a linear filter, a matrix composed of channel eigenmodes and a diagonal power allocation matrix. Interestingly, the optimum transmissions for both designs under the above assumed imperfect CSI differ only in the diagonal power allocation [17], as they do under perfect CSI [2]. A detailed proof can be found in [17].

V. NUMERICAL RESULTS

In this section, we investigate, by simulations, the impacts of transmit correlation and channel estimation error on system performance under the minimum total MSE design.

A. Simulation scenario

We let  $n_T = n_R = 4$  and  $\sigma_n^2 = 1$ . In the following figures, the SNR is defined as  $P_T/\sigma_n^2$ .<sup>4</sup> The transmit correlation model is given by [11]

$$(\mathbf{R}_T)_{ij} = \begin{cases} \rho^{|i-j|}, & i \neq j, \quad i, j \in \{1, \dots, n_T\} \\ 1, & i = j. \end{cases}$$

We consider two cases of  $\sigma_E^2$ . One is as in [11]:  $\sigma_E^2 \leq n_T \cdot 2^{n_T} \cdot \epsilon + n_T/P_T$ , where  $\epsilon \approx 1.6 \times 10^{-4}$ . We will use  $\sigma_E^2 = 0.01 + 4/SNR$ . In the other case,  $\sigma_E^2 = 0.1$ , which implies as SNR increases, the quality of channel estimation does not improve.

QPSK (4-QAM) is used for each data stream. The system performance is shown in terms of average bit error probability (ABEP) and average MSE (AMSE) per data stream, which are defined as  $ABEP = \frac{1}{B} \sum_{j=1}^B BEP_j$  and  $AMSE = \frac{1}{B} \text{tr}\{MSE(\mathbf{F})\}$ , respectively.

B. Simulation results

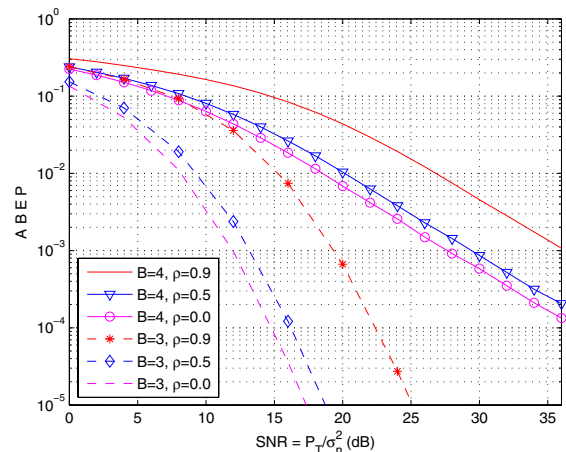


Fig. 1. Effect of transmit correlation;  $\sigma_E^2 = 0$ .

**Example 1:** The effect of transmit correlation  $\mathbf{R}_T$  ( $\sigma_E^2 = 0$ ;  $\rho = 0, 0.5, \text{ or } 0.9$ )

<sup>4</sup>Clearly, this is different from the received signal-to-total-noise ratio for each data stream.

The ABEP results are shown in Fig. 1. The data are obtained using results from [3], except that we consider transmit correlation here. We observe that if the number of data streams  $B$  is fixed, high transmit correlation has a large impact on system performance. This reminds us to choose  $B$  cautiously according to  $\mathbf{R}_T$ .<sup>5</sup>

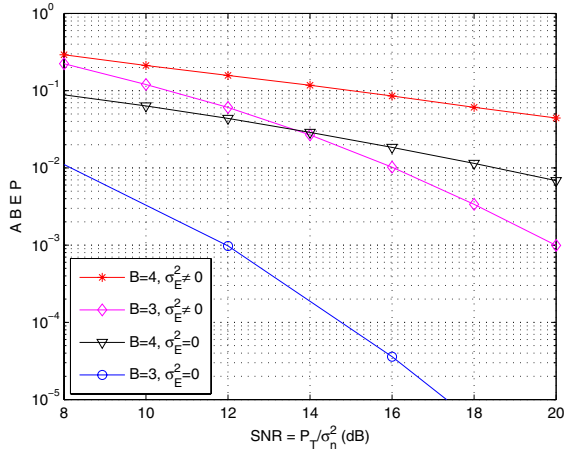


Fig. 2. Effect of channel estimation error; when  $\sigma_E^2 \neq 0$ ,  $\sigma_E^2 = 0.01 + 4/SNR$ .

**Example 2:** The effect of channel estimation error  $\sigma_E^2$  ( $\sigma_E^2 = 0.01 + 4/SNR$  and  $\mathbf{R}_T = \mathbf{I}$  ( $\rho = 0$ ))

This scenario corresponds to the special case discussed in **Remark 2** of Section III. The ABEP results are shown in Fig. 2. The channel estimation error incurs about 8 dB loss in SNR with the shown SNR range. Furthermore, we find that at medium to high SNR, the performance degradation caused by channel estimation error can be compensated by introducing diversity (by choosing  $B < n_T$ ), at the expense of reduced data rate.

**Example 3:** Performance of the optimum solution to the minimum total MSE design ( $\sigma_E^2 = 0.01 + 4/SNR$  and  $\mathbf{R}_T \neq \mathbf{I}$  ( $\rho = 0, 0.5, \text{ or } 0.9$ ))

Fig. 3 and 4 show the effects of channel estimation error and transmit correlation using the optimum transceiver for the minimum total MSE design. An error floor caused by channel estimation error is observed at high SNR. Here the same amount of transmit correlation causes smaller performance degradation than that in the case with no channel estimation error. However, high correlation still has a considerable impact on system performance.

**Example 4:** The optimum transceiver vs. a non-optimum transceiver ( $\rho = 0.7$  and  $B = 3$ )

<sup>5</sup>The choice of  $B$  according to channel conditions to guarantee a constant rate has been studied in [16]. However, in [16], perfect CSI is assumed and transmit correlation is not considered. Therefore, it is not clear how to choose the number of data streams  $B$  to guarantee a constant rate when there is channel estimation error and transmit correlation. This is a subject of future investigation.

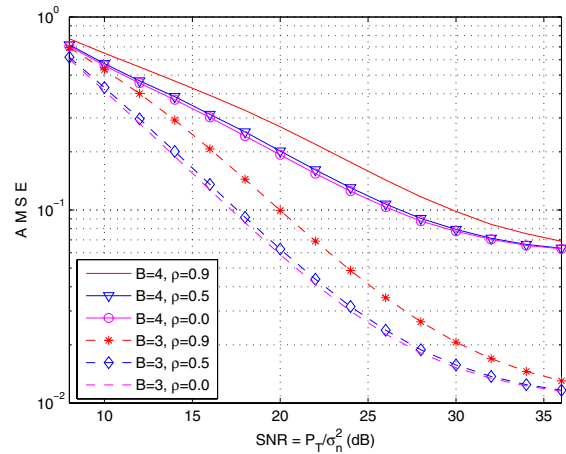


Fig. 3. AMSE performance of the optimum transceiver;  $\sigma_E^2 = 0.01 + 4/SNR$  and  $\rho = 0.0, 0.5, 0.9$ .

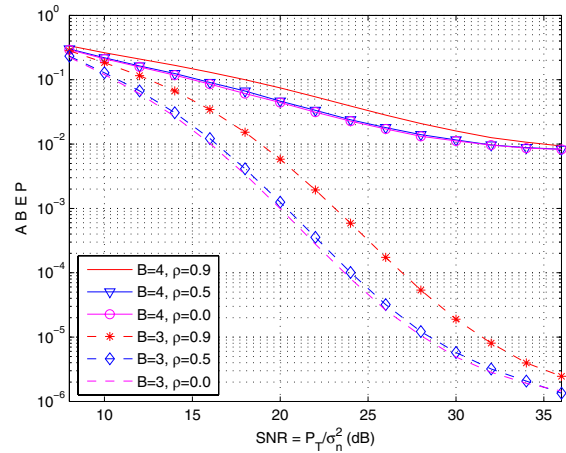


Fig. 4. ABEP performance of the optimum transceiver;  $\sigma_E^2 = 0.01 + 4/SNR$  and  $\rho = 0.0, 0.5, 0.9$ .

In this experiment, the non-optimum solution is obtained by treating  $\hat{\mathbf{H}}$  as if it was the true channel and then applying the results in [3]. The performance comparisons are shown in Fig. 5. We find that large performance gains can be obtained by using the optimum transceiver. This experiment also tells us that previous results in [2]-[3] should not be applied to the case with the imperfect CSI we have assumed.

**Example 5:** Closed-form optimum solution vs. numerical optimum solution from the iterative algorithm ( $\sigma_E^2 = 0.1$  and  $\rho = 0.7$ )

Fig. 6 presents the ABEP performance results obtained from the closed-form solution and from the iterative algorithm. It is clear that under the same set of parameters, the two curves

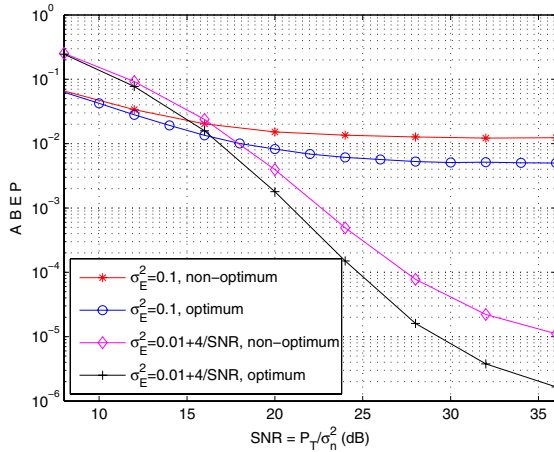


Fig. 5. The optimum transceiver vs. a non-optimum transceiver ( $\rho = 0.7$  and  $B = 3$ ).

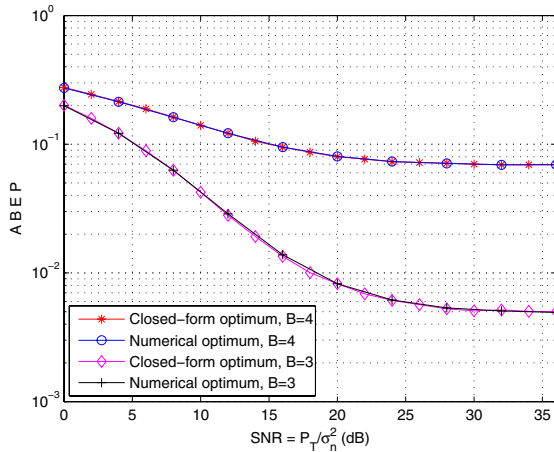


Fig. 6. Closed-form optimum solution vs. numerical optimum solution from the iterative algorithm ( $\sigma_E^2 = 0.1$  and  $\rho = 0.7$ ).

from both methods match well.<sup>6</sup>

## VI. CONCLUSION

We have formulated a new minimum total MSE joint transceiver design problem in MIMO systems with channel mean and transmit correlation information at both ends. The closed-form optimum precoder and decoder are derived. Our results gracefully fit those in literature with perfect CSI when channel estimation error diminishes. Based on simulation results, we have assessed the impact of transmit correlation on ABEP performance with and without channel estimation error, which has not been given much explicit attention in previous literature on MIMO LMMSE design. Channel estimation error causes a large performance degradation as well. For example,

<sup>6</sup>In fact, the optimum precoder from the iterative algorithm differs from that from **Theorem 1** by a unitary transform. We have considered this in our simulations.

from the ABEP curves, compared to the case without channel estimation error, an 8-dB loss in SNR is incurred when  $\sigma_E^2 = 0.01 + 4/SNR$  in the SNR range from 8dB to 20dB. This degradation can be remedied by introducing redundancy (diversity). Channel estimation error also causes an error floor at high SNR. We have determined the relationship between the minimum total MSE design and the maximum mutual information design under the same imperfect CSI.

## ACKNOWLEDGMENT

This work was supported by grants from Bell Mobility and by Communications and Information Technology Ontario.

## REFERENCES

- [1] J. Yang, S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Comm.*, vol. 42, no. 12, pp. 3221-3231, Dec. 1994.
- [2] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Sig. Proc.*, vol. 50, no. 5, pp. 1051-1064, May 2002.
- [3] H. Sampath, P. Stoica, A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Comm.*, vol. 49, no. 12, pp. 2198-2206, Dec. 2001.
- [4] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Technical University of Catalonia (UPC), Barcelona, Spain, May 2003. Available at: <http://www.ece.ust.hk/~palomar/publications.html>.
- [5] N. Khaled, G. Leus, C. Dessel, H. De Man, "A robust joint linear precoder and decoder MMSE design for slowly time-varying MIMO channels," *Proc. ICASSP 2004*, Vol. 4, pp. iv-485-iv-488, May 17-21, 2004.
- [6] X. Zhang, D. P. Palomar, B. Ottersten, "Robust design of linear MIMO transceiver for low SNR," *Proc. the 39th Asilomar Conference on Signals, Systems and Computers*, pp. 398-402, Oct. 28-Nov. 1, 2005.
- [7] X. Zhang, D. P. Palomar, B. Ottersten, "Robust design of linear MIMO transceiver under channel uncertainty," *Proc. ICASSP 2006*, Vol. 4, pp. IV-77-IV-80, May 14-19, 2006.
- [8] F. Rey, M. Lamarca, G. Vazquez, "Robust power allocation algorithms for MIMO OFDM Systems with imperfect CSI," *IEEE Trans. Sig. Proc.*, vol. 53, no. 3, pp. 1070-1085, Mar. 2005.
- [9] S. Serbetli, A. Yener, "MMSE transmitter design for correlated MIMO systems with imperfect channel estimates: power allocation trade-offs," *IEEE Trans. Wireless Comm.*, vol. 5, no. 8, pp. 2295-2304, Aug. 2006.
- [10] 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. Comm.*, vol. 48, no. 3, pp. 502-513, Mar. 2000.
- [11] 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.
- [12] S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [13] 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. Sig. Proc.*, vol. 51, no. 9, pp. 2381-2401, Sept. 2003.
- [14] S. Serbetli, A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Sig. Proc.*, vol. 52, no. 1, pp. 214-226, Jan. 2004.
- [15] J. Zhang, M. Xu, S. Zhou, J. Wang, "Minimum system-wide mean-squared error for downlink spatial multiplexing in multiuser MIMO channels," *Proc. IEEE Global Telecomm. Conf. 2005*, St. Louis, MO, Nov. 28-Dec. 2, 2005.
- [16] N. Khaled, S. Thoen, L. Deneire, "Optimizing the joint transmit and receive MMSE design using mode selection," *IEEE Trans. Comm.*, vol. 53, no. 4, pp. 730-737, April 2005.
- [17] M. Ding, S. D. Blostein, "MIMO LMMSE transceiver design with imperfect CSI at both ends," in preparation.
- [18] H. V. Poor, *Introduction to signal detection and estimation*, Springer-Verlag, 1994.
- [19] H. Lutkepohl, *Handbook of matrices*, Wiley, 1996.