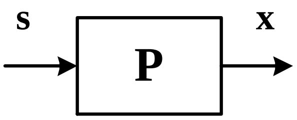
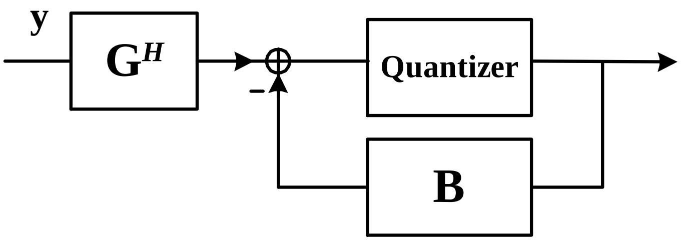

s
P
x


y
$\mathbf{G}^H$
−
Quantizer
B

A FRAMEWORK FOR DESIGNING MIMO SYSTEMS WITH DECISION FEEDBACK EQUALIZATION OR TOMLINSON-HARASHIMA PRECODING

M. Botros Shenouda and T. N. Davidson

Department of Electrical and Computer Engineering
McMaster University, Hamilton, Ontario, L8S4K1, Canada

ABSTRACT

We consider joint transceiver design for general Multiple-Input Multiple-Output communication systems that implement interference (pre-)subtraction, such as those based on Decision Feedback Equalization (DFE) or Tomlinson-Harashima precoding (THP). We develop a unified framework for joint transceiver design by considering design criteria that are expressed as functions of the Mean Square Error (MSE) of the individual data streams. By deriving two inequalities that involve the logarithms of the individual MSEs, we obtain optimal designs for two classes of communication objectives, namely those that are Schur-convex and Schur-concave functions of these logarithms. For Schur-convex objectives, the optimal design results in data streams with equal MSEs. This design simultaneously minimizes the total MSE and maximizes the mutual information for the DFE-based model. For Schur-concave objectives, the optimal DFE design results in linear equalization and the optimal THP design results in linear precoding. The proposed framework embraces a wide range of design objectives and can be regarded as a counterpart of the existing framework of linear transceiver design.

Index Terms— Decision Feedback Equalization, Tomlinson-Harashima precoding, transceiver design, MIMO channels.

1. INTRODUCTION

One of the key advantages of Multiple-Input Multiple-Output (MIMO) communications schemes is that they facilitate the simultaneous transmission of multiple data streams. Typically, such schemes involve processing of the data streams at the transmitter (precoding) and processing of the received signals (equalization) to "match" the transmission to the channel and to mitigate the interference between the received streams at reasonable computational cost. One approach to the design of such a scheme is to focus on linear precoding and linear equalization; e.g. [1, 2]. An alternative approach that offers some advantages is to allow interference (pre-)subtraction at either the transmitter or the receiver. This approach includes schemes with linear precoding and Decision Feedback Equalization (DFE), and schemes with Tomlinson-Harashima precoding (THP) and linear equalization, and will be the focus of this paper.

A large number of design strategies have been proposed for the class of linear MIMO transceivers (e.g., [2]), and a uniform framework that encompasses many of these designs was proposed in [1]. This framework consists of functions that capture a broad range of communication objectives, namely those that are Schur-convex and Schur-concave functions of the mean square error (MSE) of each data stream. For the class of interference (pre-)subtraction, designs for DFE based schemes using an MMSE criterion receiver were considered in [3, 4], and designs subject to a zero-forcing constraint were considered in [5, 6]. Some THP counterparts of these designs were presented in [3] and [7], respectively.

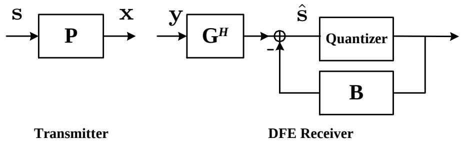


Fig. 1. MIMO transceiver with Decision Feedback Equalization.

In this paper, we develop a broadly applicable framework for joint transmitter and receiver design for MIMO systems with a DFE or a THP. We consider the broad range of design criteria that can be expressed as either Schur-convex or Schur-concave functions of the logarithm of the MSE of each data stream, and we provide optimal transceiver designs for these two classes. In addition to providing a generalization of existing designs based on the overall MSE, these classes of functions embrace other design criteria such as minimizing the maximum of the individual MSEs, or minimizing a weighted geometric mean of the MSEs. Moreover, for the DFE model, design criteria expressed in terms of the signal to interference-plus-noise ratio (SINR) and bit error rate (BER) of each stream are included in the set of objectives covered by these classes. Interestingly, the optimal design for both Schur-convex and Schur-convex objectives yields a diagonal MSE matrix. For Schur-convex objectives, the optimal design results in data streams with equal MSEs. Furthermore, for the DFE model, the optimal design for this class simultaneously minimizes the total MSE and maximizes the mutual information. For Schur-concave objectives, the optimal design results in linear precoding and equalization. From a boarder prospective, the proposed framework can be viewed as a counterpart for the design of DFE-based and THP-based transceivers of the unified framework for the design of linear transceivers in [1].

2. TWO SYSTEM MODELS

We consider a generic MIMO communication system in which the received signal can be written as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ represents the channel, the transmitted vector $\mathbf{x}$ is synthesized from a vector $\mathbf{s} \in \mathbb{C}^K$ of data symbols, and the additive noise has zero-mean and covariance matrix $\mathrm{E}_n\{\mathbf{n}\mathbf{n}^H\} = \mathbf{R}_n$. We will consider a general design approach that encompasses several design criteria for two communication systems, namely those systems with linear precoding at the transmitter and a DFE at the receiver, and those

This work was supported in part by Natural Sciences and Engineering Research Council of Canada. The work of the second author is also supported by the Canada Research Chairs Program.

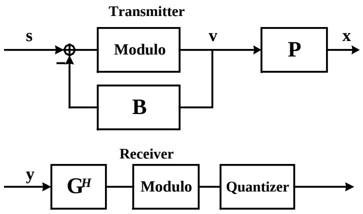


Fig. 2. MIMO transceiver with Tomlinson-Harashima precoding

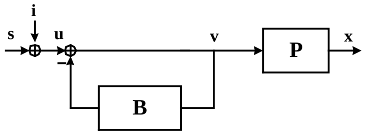


Fig. 3. Equivalent linear transmitter model for THP-based system

systems with THP at the transmitter and linear equalization at the receiver. (The linear transceiver is a special case of both systems with the feedback matrix $\mathbf{B} = \mathbf{0}$; see Figs 1 and 2.)

2.1. Decision Feedback Equalization

As shown in Fig. 1, the transmitted vector is generated by linear precoding, $\mathbf{x} = \mathbf{P}\mathbf{s}$, and hence the received vector $\mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{s} + \mathbf{n}$. The DFE is implemented using a feedforward matrix $\mathbf{G}^H$ and a strictly lower triangular feedback matrix $\mathbf{B} \in \mathbb{C}$. Assuming correct previous decisions, the vector of inputs to the quantizer is $\hat{\mathbf{s}} = (\mathbf{G}^H\mathbf{H}\mathbf{P} - \mathbf{B})\mathbf{s} + \mathbf{G}^H\mathbf{n}$. Defining the error signal $\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}}$, and using the assumption $\mathrm{E}_{\mathbf{s}}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$, the mean square error matrix can be written as:

$$\begin{aligned}\mathbf{E} = \mathrm{E}_{\mathbf{s}}\{\mathbf{e}\mathbf{e}^H\} = \mathbf{C}\mathbf{C}^H - \mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{G} - \mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{C}^H \\ + \mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\mathbf{G} + \mathbf{G}^H\mathbf{R}_n\mathbf{G}, \quad (1)\end{aligned}$$

where $\mathbf{C} = \mathbf{I} + \mathbf{B}$ is a unit diagonal lower triangular matrix. The objective is to design the $\mathbf{G}, \mathbf{C}, \mathbf{P}$ for different design criteria, subject to the transmitter power constraint $\mathrm{E}_{\mathbf{s}}\{\mathbf{x}\mathbf{x}^H\} = \mathrm{tr}(\mathbf{P}\mathbf{P}^H) \leq P_{\text{total}}$.

2.2. Tomlinson-Harashima Precoding

As shown in Fig. 2, in THP the transmitter performs successive interference pre-subtraction and spatial precoding using the strictly lower triangular matrix $\mathbf{B}$ and the precoding matrix $\mathbf{P}$, respectively. We assume that the elements of $\mathbf{s}$ are chosen from a square QAM constellation $\mathcal{S}$ with cardinality M and that $\mathrm{E}_{\mathbf{s}}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$. The Voronoi region of this constellation, $\mathcal{V}$, is a square whose side length is D. Following pre-subtraction of the effect of previously precoded symbols, the transmitter uses the modulo operation so that the symbols of $\mathbf{v}$ lie within the boundaries of $\mathcal{V}$. The effect of the modulo operation is equivalent to the addition of $\mathbf{i}_k = \mathbf{i}_k^{re}D + \mathbf{i}_k^{imag}D$ to $\mathbf{s}_k$, where $\mathbf{i}_k^{re}$, $\mathbf{i}_k^{imag} \in \mathbb{Z}$. Using this observation, we obtain the standard linearized model of the transmitter shown in Fig. 3 (e.g. [7]), in which $\mathbf{v} = (\mathbf{I} + \mathbf{B})^{-1}\mathbf{u} = \mathbf{C}^{-1}\mathbf{u}$. As a result of the modulo operation, the elements of $\mathbf{v}$ are almost uncorrelated and uniformly distributed over the Voronoi region $\mathcal{V}$ [7, Th. 3.1]. Therefore, the symbols of $\mathbf{v}$ will have slightly higher average energy than the input symbols $\mathbf{s}$. For a square QAM, we have $\sigma_v^2 = \mathrm{E}\{|\mathbf{v}_k|^2\} = \frac{M}{M-1}\mathrm{E}\{|\mathbf{s}_k|^2\}$ for all k except the first one [7]. For moderate to large values of M this power increase is negligible and the approximation $\mathrm{E}\{\mathbf{v}\mathbf{v}^H\} = \mathbf{I}$ can be used. We will use the more accurate approximation $\mathrm{E}\{\mathbf{v}\mathbf{v}^H\} = \sigma_v^2\mathbf{I}$; e.g., [3, 7].

For the THP scheme, the received signal vector can be written as $\mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{C}^{-1}\mathbf{u} + \mathbf{n}$, and hence the receiver's estimate of the of the modified data symbols is $\hat{\mathbf{u}} = \mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{C}^{-1}\mathbf{u} + \mathbf{G}^H\mathbf{n}$. Following linear equalization, the modulo operation is used to eliminate the effect of the periodic extension of the constellation induced at the transmitter. In terms of the modified data symbols, the error signal $\mathbf{e} = \hat{\mathbf{u}} - \mathbf{u} = \mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{v} + \mathbf{G}^H\mathbf{n} - \mathbf{C}\mathbf{v}$ can be used to define the Mean Square Error matrix $\mathbf{E} = \mathrm{E}_{\mathbf{v}}\{\mathbf{e}\mathbf{e}^H\}$:

$$\begin{aligned}\mathbf{E} = \sigma_v^2\mathbf{C}\mathbf{C}^H - \sigma_v^2\mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{G} - \sigma_v^2\mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{C}^H \\ + \sigma_v^2\mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\mathbf{G} + \mathbf{G}^H\mathbf{R}_n\mathbf{G}. \quad (2)\end{aligned}$$

For the TH precoding model, the transmitter power constraint is given by $\mathrm{E}\{\mathbf{x}\mathbf{x}^H\} = \sigma_v^2\mathrm{tr}(\mathbf{P}\mathbf{P}^H) \leq P_{\text{total}}$.

2.3. General Model

From equations (1) and (2), we observe that the MSE matrix $\mathbf{E}$ of both systems has a common form:

$$\mathbf{E} = \sigma^2\mathbf{C}\mathbf{C}^H - \sigma^2\mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{G} - \sigma^2\mathbf{G}^H\mathbf{H}\mathbf{P}\mathbf{C}^H + \mathbf{G}^H\mathbf{R}_y\mathbf{G}, \quad (3)$$

where $\mathbf{R}_y = \sigma^2\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H + \mathbf{R}_n$. For the DFE model $\sigma^2 = 1$ while for the TH precoding model $\sigma^2 = \sigma_v^2$. The average transmitter power constraint can be rewritten as $\mathrm{tr}(\mathbf{P}\mathbf{P}^H) \leq P_{\text{total}}/\sigma^2 = P$.

3. OPTIMAL FEEDFORWARD AND FEEDBACK MATRICES

We consider the joint design of the transceiver matrices $\mathbf{G}, \mathbf{C}, \mathbf{P}$ in order to optimize system design criteria that are expressed as functions of the MSE of the individual data streams $\mathbf{E}_{ii}$. We will adopt three-step design approach. First, an expression for the optimal feedforward matrix $\mathbf{G}^H$ will be found as a function of $\mathbf{C}$ and $\mathbf{P}$. Second, using the expression of the optimal $\mathbf{G}$, an expression of the optimal $\mathbf{C}$ will be found as a function of $\mathbf{P}$. Finally, using the obtained expressions of $\mathbf{G}$ and $\mathbf{C}$, we will design the optimal precoder $\mathbf{P}$.

3.1. Optimal feedforward matrix $\mathbf{G}^H$

For given $\mathbf{C}$ and $\mathbf{P}$, the MSE of the i^{th} data stream, $\mathbf{E}_{ii}$, is a convex function of the i^{th} column of $\mathbf{G}$, denoted $\mathbf{g}_i$, and is independent of other columns. Therefore, the columns of $\mathbf{G}$ can be independently optimized to minimize the individual MSEs. A similar property was observed in [1] for linear transceivers. Setting the gradient of $\mathbf{E}_{ii}$ with respect to $\mathbf{g}_i$ to zero, we obtain following expression for the optimal $\mathbf{G}$:

$$\mathbf{G} = \sigma^2\mathbf{R}_y^{-1}\mathbf{H}\mathbf{P}\mathbf{C}^H. \quad (4)$$

Since each $\mathbf{g}_i$ independently minimizes the MSE of the i^{th} data stream, the expression of $\mathbf{G}$ in (4) is also the optimal feedforward matrix in the sense of the sum of MSEs, $\mathrm{tr}(\mathbf{E})$. Using this expression, the MSE matrix can be written as:

$$\mathbf{E} = \sigma^2\mathbf{C}(\mathbf{I} + \sigma^2\mathbf{P}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{P})^{-1}\mathbf{C}^H = \mathbf{C}\mathbf{M}\mathbf{C}^H, \quad (5)$$

where the matrix inversion lemma has been used.

3.2. Optimal feedback matrix B

From (5) we observe that the MSE of each data stream $\mathbf{E}_{ii}$ is convex function of the i^{th} row of $\mathbf{C} = \mathbf{I} + \mathbf{B}$ and is independent of the other rows. Therefore, the optimal $\mathbf{C}$ that minimizes the individual MSEs can be obtained by minimizing any convex combination of $\mathbf{E}_{ii}$. By choosing that convex combination to be the sum, our goal reduces to minimizing tr$(\mathbf{CMC}^H)$ subject to $\mathbf{C}$ being unit diagonal lower triangular matrix. Using the Cholesky decomposition $\mathbf{M} = \mathbf{LL}^H$, where $\mathbf{L}$ is a lower triangular matrix with positive diagonal elements, we can rewrite the objective as tr$(\mathbf{CMC}^H) = \|\mathbf{CL}\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm and the product $\mathbf{CL}$ is positive definite lower triangular matrix [8]. Let $\lambda_1(\mathbf{CL}) \geq \ldots, \geq \lambda_K(\mathbf{CL})$ and $\sigma_1(\mathbf{CL}) \geq \ldots \geq \sigma_K(\mathbf{CL})$ denote the ordered eigen values and singular values, respectively, of the matrix $\mathbf{CL}$. Then the unit diagonal lower triangular $\mathbf{C}$ that minimizes tr$(\mathbf{CMC}^H)$ can be obtained using the following lower bound:

$$\begin{aligned} \|\mathbf{CL}\|_F^2 = \textstyle\sum_{i=1}^K \sigma_i^2(\mathbf{CL}) \quad &\geq \quad \sum_{i=1}^K \lambda_i^2(\mathbf{CL}) \qquad (6) \\ &= \quad \sum_{i=1}^K (\mathbf{CL})_{ii}^2 = \sum_{i=1}^K \mathbf{L}_{ii}^2, \qquad (7) \end{aligned}$$

where the bound in (6) is obtained by applying Weyl's inequality [9], and (7) follows from the fact that $\mathbf{CL}$ is lower triangular and $\mathbf{C}$ is unit diagonal. The inequality in (6) is satisfied with equality when the matrix is normal [9]. Since our matrix $\mathbf{CL}$ is a triangular matrix, it can only be normal if it is diagonal [8, pp 103]. Therefore, the matrix $\mathbf{C}$ that attains the lower bound is:

$$\mathbf{C} = \mathrm{Diag}\left(\mathbf{L}_{11}, \ldots, \mathbf{L}_{KK}\right) \mathbf{L}^{-1}. \tag{8}$$

Using this optimal $\mathbf{C}$, the MSE matrix can be rewritten as:

$$\mathbf{E} = \mathrm{Diag}\left(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2\right). \tag{9}$$

We observe that for any given precoding matrix $\mathbf{P}$, the optimal feedforward and feedback matrices will yield a diagonal MSE matrix, with the individual MSEs being $\mathbf{E}_{ii} = \mathbf{L}_{ii}^2$.

4. OPTIMAL PRECODING MATRIX P

Given the optimal $\mathbf{G}$ and $\mathbf{C}$, the last step is to design a precoding matrix $\mathbf{P}$ to optimize design criteria expressed as functions of individual MSE of each stream, $\mathbf{L}_{ii}^2$. We will first derive two inequalities involving $\mathbf{L}_{ii}$ that enable us to characterize the optimal precoder.

4.1. Preliminaries

To derive the first inequality, we will use the concept of multiplicative majorization:
Multiplicative Majorization [9, 10]: Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^K$ and let $a_{[1]} \geq \ldots \geq a_{[K]}$ denote the elements of $\mathbf{a}$ in descending order. The vector $\mathbf{b}$ is said to multiplicatively majorize $\mathbf{a}$, $\mathbf{a} \prec_\times \mathbf{b}$, if $\prod_{i=1}^j \mathbf{a}_{[i]} \leq \prod_{i=1}^j \mathbf{b}_{[i]}$, for $j = 1, \ldots, K-1$ and $\prod_{i=1}^K \mathbf{a}_{[i]} = \prod_{i=1}^K \mathbf{b}_{[i]}$.
An important example of this definition is:

Lemma 1 Weyl [9]: *Let $\mathbf{A} \in \mathbb{C}^{K\times K}$ and let $\lambda_i(\mathbf{A})$ and $\sigma_i(\mathbf{A})$ denote the eigen values and singular values of $\mathbf{A}$, respectively. Then we have $(|\lambda_1(\mathbf{A})|^2,\ \ldots,\ |\lambda_K(\mathbf{A})|^2) \prec_\times (\sigma_1^2(\mathbf{A}),\ \ldots,\ \sigma_K^2(\mathbf{A}))$. If $\mathbf{A}$ is normal, then $|\lambda_i(\mathbf{A})| = \sigma_i(\mathbf{A})$.*

Applying the above lemma to the positive definite lower triangular matrix $\mathbf{L}$, we obtain out first inequality:

$$(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2) \prec_\times (\sigma_1^2(\mathbf{L}), \ldots, \sigma_K^2(\mathbf{L})). \tag{10}$$

The second inequality involves the more common notation of additive majorization:
Additive Majorization [10]: Let $\mathbf{a},\ \mathbf{b} \in \mathbb{R}^K$. The vector $\mathbf{b}$ is said to majorize $\mathbf{a}$, $\mathbf{a} \prec \mathbf{b}$, if $\sum_{i=1}^j \mathbf{a}_{[i]} \leq \sum_{i=1}^j \mathbf{b}_{[i]}$, for $j = 1, \ldots, K-1$ and $\sum_{i=1}^K \mathbf{a}_{[i]} = \sum_{i=1}^K \mathbf{b}_{[i]}$

We observe that if elements of $\mathbf{a}$ and $\mathbf{b}$ are positive, then $\mathbf{a} \prec_\times \mathbf{b} \Leftrightarrow \ln(\mathbf{a}) \prec \ln(\mathbf{b})$. Consequently, (10) can be written as:

$$\boldsymbol{l} \prec \mathbf{m}, \tag{11}$$

where $\boldsymbol{l} = (\ln \mathbf{L}_{11}^2, \ldots, \ln \mathbf{L}_{KK}^2)$ and $\mathbf{m} = (\ln \sigma_1^2(\mathbf{L}), \ldots, \ln \sigma_K^2(\mathbf{L}))$.

To derive the second inequality, we will use the following consequence of additive majorization: Any vector $\mathbf{a} \in \mathbb{R}^K$ majorizes its mean vector $\overline{\mathbf{a}}$ whose elements are all equal to the mean; i.e., $\overline{\mathbf{a}}_i = \frac{1}{K} \sum_{i=1}^K \mathbf{a}_i$. That is, $\overline{\mathbf{a}} \prec \mathbf{a}$. Now, since $\mathbf{M} = \mathbf{LL}^H$, we know that $\prod_{i=1}^K \mathbf{L}_{ii}^2 = \det(\mathbf{LL}^H) = \det(\mathbf{M})$. As a result, we have $\sum_{i=1}^K \boldsymbol{l}_i = \ln\det(\mathbf{M})$ and our second inequality is:

$$\overline{\boldsymbol{l}} \prec \boldsymbol{l}, \tag{12}$$

where $\overline{\boldsymbol{l}}_i = \frac{1}{K} \ln\det(\mathbf{M})$.

The proposed designs will be based on the following classes of functions [10]: A real-valued function $f(\mathbf{x})$ defined on a subset $\mathcal{A}$ of $\mathbb{R}^K$ is said to be Schur-convex if $\mathbf{a} \prec \mathbf{b}$ on $\mathcal{A} \Rightarrow f(\mathbf{a}) \leq f(\mathbf{b})$, and is said to be Schur-concave if $\mathbf{a} \prec \mathbf{b}$ on $\mathcal{A} \Rightarrow f(\mathbf{a}) \geq f(\mathbf{b})$. In particular, we will consider communication objectives that can be expressed as the minimization of a functions of the MSEs of each data stream, $g(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2) = g(e^{\boldsymbol{l}_1}, \ldots, e^{\boldsymbol{l}_K})) = g(e^{\boldsymbol{l}})$.

4.2. Schur-convex functions

Examples of objectives that result in $g(e^{\boldsymbol{l}})$ being a Schur-convex function of $\boldsymbol{l}$ include: minimization of the maximum of individual MSEs: $g(e^{\boldsymbol{l}}) = \max_i e^{\boldsymbol{l}_i}$; minimization of the total MMSE: $g(e^{\boldsymbol{l}}) = \sum_i e^{\boldsymbol{l}_i}$; and minimization of the (log) determinant MSE matrix: $\det(\mathbf{E}) = \prod_i e^{\boldsymbol{l}_i}$, which is also Schur-concave function of $\boldsymbol{l}$. For the DFE model, the SINR of the i^{th} stream is given by SINR$_i = (1/\text{MSE}_i) - 1 = e^{-\boldsymbol{l}_i} - 1$. Hence, many objectives in terms of SINR and BER can be expressed as Schur-convex functions of $\boldsymbol{l}$. As we will show below, the optimal transceiver design is identical for all these objectives.

If $g(e^{\boldsymbol{l}})$ is a Schur convex function of $\boldsymbol{l}$, then from (12) we have that $g(e^{\overline{\boldsymbol{l}}}) \leq g(e^{\boldsymbol{l}})$ and the optimal value is obtained when all $\boldsymbol{l}_i$ are equal to $\boldsymbol{l}_i = \frac{1}{K} \ln\det(\mathbf{M})$; i.e., $\mathbf{E}_{ii} = \mathbf{L}_{ii}^2 = \sqrt[K]{\det(\mathbf{M})}$. Since the objective is an increasing function of the individual MSE, the design goal reduces to minimizing $\det \mathbf{M}$ subject to the power constraint and to the constraint that diagonal elements of the Cholesky factor of $\mathbf{M}$ are all equal. We will start by characterizing the family of solutions that minimize $\det(\mathbf{M})$ subject to the power constraint, then we will show that there is a member of this family that yields a Cholesky factor of $\mathbf{M}$ with equal diagonal elements. Minimizing $\det(\mathbf{M})$ is equivalent to maximizing the Gaussian mutual information, and the family of optimal precoders is obtained using a standard water-filling algorithm [11]. In particular, if $\mathbf{R}_H = \sigma^2 \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}_{\mathbf{H}}\mathbf{U}^H$, the family of optimal precoders takes the form:

$$\mathbf{P} = \mathbf{U}_1 \hat{\mathbf{\Phi}} \mathbf{V} = \mathbf{U}_1 [\mathbf{\Phi} \quad \mathbf{0}] \mathbf{V}, \tag{13}$$

where $\mathbf{U}_1 \in \mathbb{C}^{N_t \times \hat{K}}$ contains the eigen vectors of $\mathbf{R_H}$ corresponding to the $\hat{K} \leq K$ largest eigen values, $\hat{K}$ and the diagonal positive definite matrix $\mathbf{\Phi}$ are obtained from the water- filling algorithm [11], and $\mathbf{V} \in \mathbb{C}^{K \times K}$ is a unitary matrix degree of freedom. This result shows that for DFE based systems designed according to any Schur-convex function of $\boldsymbol{l}$, the optimal solution is information lossless. To complete the design of $\mathbf{P}$, we need to select $\mathbf{V}$ such that the Cholesky decomposition of $\mathbf{M} = \mathbf{L}\mathbf{L}^H$ yields an $\mathbf{L}$ factor with equal diagonal elements. Using (13):

$$\begin{aligned} \mathbf{M} &= \left(\mathbf{V}^H (\mathbf{I} + \hat{\mathbf{\Phi}}^T \mathbf{\Lambda}_{\mathbf{H}1} \hat{\mathbf{\Phi}})^{-1/2}\right) \left((\mathbf{I} + \hat{\mathbf{\Phi}}^T \mathbf{\Lambda}_{\mathbf{H}1} \hat{\mathbf{\Phi}})^{-1/2} \mathbf{V}\right) \\ &= \mathbf{L}\mathbf{L}^H = \mathbf{R}^H \mathbf{R} = (\mathbf{Q}\mathbf{R})^H (\mathbf{Q}\mathbf{R}), \end{aligned} \tag{14}$$

where $\mathbf{\Lambda}_{\mathbf{H}1}$ is the diagonal matrix containing the largest $\hat{K}$ eigen values of $\mathbf{R_H}$, and $\mathbf{Q}$ is a matrix with orthonormal columns. Hence, finding $\mathbf{V}$ is equivalent to finding a $\mathbf{V}$ such that QR decomposition of $(\mathbf{I} + \hat{\mathbf{\Phi}}^T \mathbf{\Lambda}_{\mathbf{H}1} \hat{\mathbf{\Phi}})^{-1/2} \mathbf{V}$ has an R-factor with equal diagonal. This problem was solved in [6] and $\mathbf{V}$ can be obtained by applying the algorithm in [6] to the matrix $(\mathbf{I} + \hat{\mathbf{\Phi}}^T \mathbf{\Lambda}_{\mathbf{H}1} \hat{\mathbf{\Phi}})^{-1/2}$; see also [4, 5].

4.3. Schur-concave functions

If $g(e^{\boldsymbol{l}})$ is a Schur-concave function of $\boldsymbol{l}$, then from (11) we have $g(e^{\mathbf{m}}) \leq g(e^{\boldsymbol{l}})$ and the optimal value is obtained when $\mathbf{L}_{ii} = \sigma_i(\mathbf{L})$. According to Lemma 1, this equality holds when $\mathbf{L}$ is normal matrix. Since $\mathbf{L}$ is a lower triangular matrix, in order to be normal it must be a diagonal matrix [8]. The optimal $\mathbf{C}$ in that case is $\mathbf{I}$. That is, in the case of Schur-concave functions of $\boldsymbol{l}$, the optimal DFE design results in linear equalization and optimal TH precoding design results in linear precoding. Examples of this class of objectives include minimization of product of the MSEs and general (weighted) geometrical mean of MSEs.

5. SIMULATION STUDY

We consider a system that transmits $K = 4$ streams of 16-QAM symbols over a 4×4 slowly fading independent Rayleigh channel with additive white Gaussian noise. We plot the average bit error rate (BER) against the signal to ratio $P_{\text{total}}/\text{tr}(\mathbf{R}_n)$. We compare the performance of the proposed Schur-convex designs for THP and DFE (which minimize the total MSE among other objectives), with the corresponding linear transceiver design that minimizes total MSE [1, 2], and the optimal linear transceiver that maximizes the mutual information (minimizes $\log\det(\mathbf{E})$) [1]. The performance advantages of interference cancellation are quite clear from Fig. 4.

6. CONCLUSION

We developed a unified framework for joint transceiver design of interference (pre-)subtraction schemes for MIMO channels. We obtained optimal designs for two classes of communication objectives, namely those that are Schur-convex and Schur-concave functions of the logarithms of the individual MSEs. For Schur-convex objectives, the optimal transceiver results in equal individual MSEs. For the DFE model, it optimizes both the total MSE and mutual information. For the class Schur-concave objectives, the optimal DFE design results in linear equalization and the optimal TH precoding design results in linear precoding.

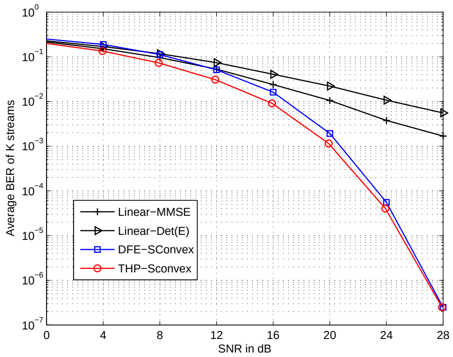


Fig. 4. BERs of the proposed Schur-convex designs and the optimal linear transceivers: minimum MSE (Linear-MMSE), and maximum mutual information (Linear-Det(E)), for $N_t = N_r = K = 4$.

7. REFERENCES

[1] 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 Processing*, vol. 51, pp. 2381–2401, Sept. 2003.

[2] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.

[3] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Linear and nonlinear preequalization/equalization for MIMO systems with long-term channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 373–378, Mar. 2004.

[4] F. Xu, T. N. Davidson, J. Zhang, and K. M. Wong, "Design of block transceivers with decision feedback detection," *IEEE Trans. Signal Processing*, vol. 54, pp. 965–978, March 2006.

[5] Y. Jiang, J. Li, and W.W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Processing*, vol. 53, pp. 3791–3803, Oct. 2005.

[6] J. Zhang, A. Kavcic, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inform. Theory*, vol. 51, pp. 154–172, Jan. 2005.

[7] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*, Wiley, New York, 2002.

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1985.

[9] H. Weyl, "Inequalities between the two kinds of eigenvalues of a linear transformation," *Proc. Nat. Acad. Sci.*, vol. 35, pp. 408–411, July 1949.

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

[11] H. S. Witsenhausen, "A determinant maximization problem occurring in the theory of data communication," *SIAM J. Appl. Math.*, vol. 29, pp. 515–522, 1975.

A FRAMEWORK FOR DESIGNING MIMO SYSTEMS WITH DECISION FEEDBACK EQUALIZATION OR TOMLINSON-HARASHIMA PRECODING

M. Botros Shenouda and T. N. Davidson

Department of Electrical and Computer Engineering
McMaster University, Hamilton, Ontario, L8S4K1, Canada

ABSTRACT

We consider joint transceiver design for general Multiple-Input Multiple-Output communication systems that implement interference (pre-)subtraction, such as those based on Decision Feedback Equalization (DFE) or Tomlinson-Harashima precoding (THP). We develop a unified framework for joint transceiver design by considering design criteria that are expressed as functions of the Mean Square Error (MSE) of the individual data streams. By deriving two inequalities that involve the logarithms of the individual MSEs, we obtain optimal designs for two classes of communication objectives, namely those that are Schur-convex and Schur-concave functions of these logarithms. For Schur-convex objectives, the optimal design results in data streams with equal MSEs. This design simultaneously minimizes the total MSE and maximizes the mutual information for the DFE-based model. For Schur-concave objectives, the optimal DFE design results in linear equalization and the optimal THP design results in linear precoding. The proposed framework embraces a wide range of design objectives and can be regarded as a counterpart of the existing framework of linear transceiver design.

***Index Terms*— Decision Feedback Equalization, Tomlinson-Harashima precoding, transceiver design, MIMO channels.

1. INTRODUCTION

One of the key advantages of Multiple-Input Multiple-Output (MIMO) communications schemes is that they facilitate the simultaneous transmission of multiple data streams. Typically, such schemes involve processing of the data streams at the transmitter (precoding) and processing of the received signals (equalization) to "match" the transmission to the channel and to mitigate the interference between the received streams at reasonable computational cost. One approach to the design of such a scheme is to focus on linear precoding and linear equalization; e.g. [**?**, **?**]. An alternative approach that offers some advantages is to allow interference (pre-)subtraction at either the transmitter or the receiver. This approach includes schemes with linear precoding and Decision Feedback Equalization (DFE), and schemes with Tomlinson-Harashima precoding (THP) and linear equalization, and will be the focus of this paper.

A large number of design strategies have been proposed for the class of linear MIMO transceivers (e.g., [**?**]), and a uniform framework that encompasses many of these designs was proposed in [**?**]. This framework consists of functions that capture a broad range of communication objectives, namely those that are Schur-convex and Schur-concave functions of the mean square error (MSE) of each data stream. For the class of interference (pre-)subtraction, designs for DFE based schemes using an MMSE criterion receiver were considered in [**?**, **?**], and designs subject to a zero-forcing constraint were considered in [**?**, **?**]. Some THP counterparts of these designs were presented in [**?**] and [**?**], respectively.

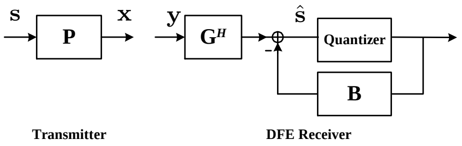


Fig. 1. MIMO transceiver with Decision Feedback Equalization.

In this paper, we develop a broadly applicable framework for joint transmitter and receiver design for MIMO systems with a DFE or a THP. We consider the broad range of design criteria that can be expressed as either Schur-convex or Schur-concave functions of the logarithm of the MSE of each data stream, and we provide optimal transceiver designs for these two classes. In addition to providing a generalization of existing designs based on the overall MSE, these classes of functions embrace other design criteria such as minimizing the maximum of the individual MSEs, or minimizing a weighted geometric mean of the MSEs. Moreover, for the DFE model, design criteria expressed in terms of the signal to interference-plus-noise ratio (SINR) and bit error rate (BER) of each stream are included in the set of objectives covered by these classes. Interestingly, the optimal design for both Schur-convex and Schur-convex objectives yields a diagonal MSE matrix. For Schur-convex objectives, the optimal design results in data streams with equal MSEs. Furthermore, for the DFE model, the optimal design for this class simultaneously minimizes the total MSE and maximizes the mutual information. For Schur-concave objectives, the optimal design results in linear precoding and equalization. From a boarder prospective, the proposed framework can be viewed as a counterpart for the design of DFE-based and THP-based transceivers of the unified framework for the design of linear transceivers in [**?**].

2. TWO SYSTEM MODELS

We consider a generic MIMO communication system in which the received signal can be written as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ represents the channel, the transmitted vector $\mathbf{x}$ is synthesized from a vector $\mathbf{s} \in \mathbb{C}^K$ of data symbols, and the additive noise has zero-mean and covariance matrix $\mathrm{E}_n\{\mathbf{n}\mathbf{n}^H\} = \mathbf{R}_n$. We will consider a general design approach that encompasses several design criteria for two communication systems, namely those systems with linear precoding at the transmitter and a DFE at the receiver, and those

This work was supported in part by Natural Sciences and Engineering Research Council of Canada. The work of the second author is also supported by the Canada Research Chairs Program.

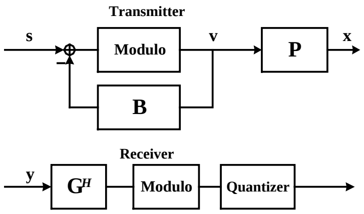


Fig. 2. MIMO transceiver with Tomlinson-Harashima precoding

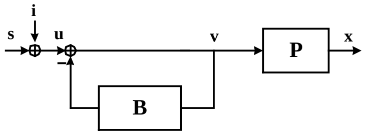


Fig. 3. Equivalent linear transmitter model for THP-based system

systems with THP at the transmitter and linear equalization at the receiver. (The linear transceiver is a special case of both systems with the feedback matrix $\mathbf{B} = \mathbf{0}$; see Figs 1 and 2.)

2.1. Decision Feedback Equalization

As shown in Fig. 1, the transmitted vector is generated by linear precoding, $\mathbf{x} = \mathbf{Ps}$, and hence the received vector $\mathbf{y} = \mathbf{HPs} + \mathbf{n}$. The DFE is implemented using a feedforward matrix $\mathbf{G}^H$ and a strictly lower triangular feedback matrix $\mathbf{B} \in \mathbb{C}$. Assuming correct previous decisions, the vector of inputs to the quantizer is $\hat{\mathbf{s}} = (\mathbf{G}^H\mathbf{HP} - \mathbf{B})\mathbf{s} + \mathbf{G}^H\mathbf{n}$. Defining the error signal $\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}}$, and using the assumption $\mathrm{E}_{\mathbf{s}}\{\mathbf{ss}^H\} = \mathbf{I}$, the mean square error matrix can be written as:

$$\begin{aligned}\mathbf{E} = \mathrm{E}_{\mathbf{s}}\{\mathbf{ee}^H\} = \mathbf{CC}^H - \mathbf{CP}^H\mathbf{H}^H\mathbf{G} - \mathbf{G}^H\mathbf{HPC}^H \\ + \mathbf{G}^H\mathbf{HPP}^H\mathbf{H}^H\mathbf{G} + \mathbf{G}^H\mathbf{R}_n\mathbf{G}, \quad (1)\end{aligned}$$

where $\mathbf{C} = \mathbf{I} + \mathbf{B}$ is a unit diagonal lower triangular matrix. The objective is to design the $\mathbf{G}, \mathbf{C}, \mathbf{P}$ for different design criteria, subject to the transmitter power constraint $\mathrm{E}_{\mathbf{s}}\{\mathbf{xx}^H\} = \mathrm{tr}(\mathbf{PP}^H) \leq P_{\text{total}}$.

2.2. Tomlinson-Harashima Precoding

As shown in Fig. 2, in THP the transmitter performs successive interference pre-subtraction and spatial precoding using the strictly lower triangular matrix $\mathbf{B}$ and the precoding matrix $\mathbf{P}$, respectively. We assume that the elements of $\mathbf{s}$ are chosen from a square QAM constellation $\mathcal{S}$ with cardinality M and that $\mathrm{E}_{\mathbf{s}}\{\mathbf{ss}^H\} = \mathbf{I}$. The Voronoi region of this constellation, $\mathcal{V}$, is a square whose side length is D. Following pre-subtraction of the effect of previously precoded symbols, the transmitter uses the modulo operation so that the symbols of $\mathbf{v}$ lie within the boundaries of $\mathcal{V}$. The effect of the modulo operation is equivalent to the addition of $\mathbf{i}_k = \mathbf{i}_k^{re}D + \mathbf{i}_k^{imag}D$ to $\mathbf{s}_k$, where $\mathbf{i}_k^{re}$, $\mathbf{i}_k^{imag} \in \mathbb{Z}$. Using this observation, we obtain the standard linearized model of the transmitter shown in Fig. 3 (e.g. [**?**]), in which $\mathbf{v} = (\mathbf{I} + \mathbf{B})^{-1}\mathbf{u} = \mathbf{C}^{-1}\mathbf{u}$. As a result of the modulo operation, the elements of $\mathbf{v}$ are almost uncorrelated and uniformly distributed over the Voronoi region $\mathcal{V}$ [**?**, Th. 3.1]. Therefore, the symbols of $\mathbf{v}$ will have slightly higher average energy than the input symbols $\mathbf{s}$. For a square QAM, we have $\sigma_v^2 = \mathrm{E}\{|\mathbf{v}_k|^2\} = \frac{M}{M-1}\mathrm{E}\{|\mathbf{s}_k|^2\}$ for all k except the first one [**?**]. For moderate to large values of M this power increase is negligible and the approximation $\mathrm{E}\{\mathbf{vv}^H\} = \mathbf{I}$ can be used. We will use the more accurate approximation $\mathrm{E}\{\mathbf{vv}^H\} = \sigma_v^2\mathbf{I}$; e.g., [**?**, **?**].

For the THP scheme, the received signal vector can be written as $\mathbf{y} = \mathbf{HPC}^{-1}\mathbf{u} + \mathbf{n}$, and hence the receiver's estimate of the of the modified data symbols is $\hat{\mathbf{u}} = \mathbf{G}^H\mathbf{HPC}^{-1}\mathbf{u} + \mathbf{G}^H\mathbf{n}$. Following linear equalization, the modulo operation is used to eliminate the effect of the periodic extension of the constellation induced at the transmitter. In terms of the modified data symbols, the error signal $\mathbf{e} = \hat{\mathbf{u}} - \mathbf{u} = \mathbf{G}^H\mathbf{HPv} + \mathbf{G}^H\mathbf{n} - \mathbf{Cv}$ can be used to define the Mean Square Error matrix $\mathbf{E} = \mathrm{E}_{\mathbf{v}}\{\mathbf{ee}^H\}$:

$$\begin{aligned}\mathbf{E} = \sigma_v^2\mathbf{CC}^H - \sigma_v^2\mathbf{CP}^H\mathbf{H}^H\mathbf{G} - \sigma_v^2\mathbf{G}^H\mathbf{HPC}^H \\ + \sigma_v^2\mathbf{G}^H\mathbf{HPP}^H\mathbf{H}^H\mathbf{G} + \mathbf{G}^H\mathbf{R}_n\mathbf{G}. \quad (2)\end{aligned}$$

For the TH precoding model, the transmitter power constraint is given by $\mathrm{E}\{\mathbf{xx}^H\} = \sigma_v^2\mathrm{tr}(\mathbf{PP}^H) \leq P_{\text{total}}$.

2.3. General Model

From equations (1) and (2), we observe that the MSE matrix $\mathbf{E}$ of both systems has a common form:

$$\mathbf{E} = \sigma^2\mathbf{CC}^H - \sigma^2\mathbf{CP}^H\mathbf{H}^H\mathbf{G} - \sigma^2\mathbf{G}^H\mathbf{HPC}^H + \mathbf{G}^H\mathbf{R}_y\mathbf{G}, \quad (3)$$

where $\mathbf{R}_y = \sigma^2\mathbf{HPP}^H\mathbf{H}^H + \mathbf{R}_n$. For the DFE model $\sigma^2 = 1$ while for the TH precoding model $\sigma^2 = \sigma_v^2$. The average transmitter power constraint can be rewritten as $\mathrm{tr}(\mathbf{PP}^H) \leq P_{\text{total}}/\sigma^2 = P$.

3. OPTIMAL FEEDFORWARD AND FEEDBACK MATRICES

We consider the joint design of the transceiver matrices $\mathbf{G}, \mathbf{C}, \mathbf{P}$ in order to optimize system design criteria that are expressed as functions of the MSE of the individual data streams $\mathbf{E}_{ii}$. We will adopt three-step design approach. First, an expression for the optimal feedforward matrix $\mathbf{G}^H$ will be found as a function of $\mathbf{C}$ and $\mathbf{P}$. Second, using the expression of the optimal $\mathbf{G}$, an expression of the optimal $\mathbf{C}$ will be found as a function of $\mathbf{P}$. Finally, using the obtained expressions of $\mathbf{G}$ and $\mathbf{C}$, we will design the optimal precoder $\mathbf{P}$.

3.1. Optimal feedforward matrix $\mathbf{G}^H$

For given $\mathbf{C}$ and $\mathbf{P}$, the MSE of the i^{th} data stream, $\mathbf{E}_{ii}$, is a convex function of the i^{th} column of $\mathbf{G}$, denoted $\mathbf{g}_i$, and is independent of other columns. Therefore, the columns of $\mathbf{G}$ can be independently optimized to minimize the individual MSEs. A similar property was observed in [**?**] for linear transceivers. Setting the gradient of $\mathbf{E}_{ii}$ with respect to $\mathbf{g}_i$ to zero, we obtain following expression for the optimal $\mathbf{G}$:

$$\mathbf{G} = \sigma^2\mathbf{R}_y^{-1}\mathbf{HPC}^H. \quad (4)$$

Since each $\mathbf{g}_i$ independently minimizes the MSE of the i^{th} data stream, the expression of $\mathbf{G}$ in (4) is also the optimal feedforward matrix in the sense of the sum of MSEs, $\mathrm{tr}(\mathbf{E})$. Using this expression, the MSE matrix can be written as:

$$\mathbf{E} = \sigma^2\mathbf{C}(\mathbf{I} + \sigma^2\mathbf{P}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HP})^{-1}\mathbf{C}^H = \mathbf{CMC}^H, \quad (5)$$

where the matrix inversion lemma has been used.

3.2. Optimal feedback matrix B

From (5) we observe that the MSE of each data stream $\mathbf{E}_{ii}$ is convex function of the i^{th} row of $\mathbf{C} = \mathbf{I} + \mathbf{B}$ and is independent of the other rows. Therefore, the optimal $\mathbf{C}$ that minimizes the individual MSEs can be obtained by minimizing any convex combination of $\mathbf{E}_{ii}$. By choosing that convex combination to be the sum, our goal reduces to minimizing tr$(\mathbf{CMC}^H)$ subject to $\mathbf{C}$ being unit diagonal lower triangular matrix. Using the Cholesky decomposition $\mathbf{M} = \mathbf{LL}^H$, where $\mathbf{L}$ is a lower triangular matrix with positive diagonal elements, we can rewrite the objective as tr$(\mathbf{CMC}^H) = \|\mathbf{CL}\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm and the product $\mathbf{CL}$ is positive definite lower triangular matrix [?]. Let $\lambda_1(\mathbf{CL}) \geq \ldots, \geq \lambda_K(\mathbf{CL})$ and $\sigma_1(\mathbf{CL}) \geq \ldots \geq \sigma_K(\mathbf{CL})$ denote the ordered eigen values and singular values, respectively, of the matrix $\mathbf{CL}$. Then the unit diagonal lower triangular $\mathbf{C}$ that minimizes tr$(\mathbf{CMC}^H)$ can be obtained using the following lower bound:

$$\begin{aligned} \|\mathbf{CL}\|_F^2 = \textstyle\sum_{i=1}^K \sigma_i^2(\mathbf{CL}) \quad &\geq \quad \sum_{i=1}^K \lambda_i^2(\mathbf{CL}) && (6) \\ &= \quad \sum_{i=1}^K (\mathbf{CL})_{ii}^2 = \sum_{i=1}^K \mathbf{L}_{ii}^2, && (7) \end{aligned}$$

where the bound in (6) is obtained by applying Weyl's inequality [?], and (7) follows from the fact that $\mathbf{CL}$ is lower triangular and $\mathbf{C}$ is unit diagonal. The inequality in (6) is satisfied with equality when the matrix is normal [?]. Since our matrix $\mathbf{CL}$ is a triangular matrix, it can only be normal if it is diagonal [?, pp 103]. Therefore, the matrix $\mathbf{C}$ that attains the lower bound is:

$$\mathbf{C} = \text{Diag}\,(\mathbf{L}_{11}, \ldots, \mathbf{L}_{KK})\,\mathbf{L}^{-1}. \tag{8}$$

Using this optimal $\mathbf{C}$, the MSE matrix can be rewritten as:

$$\mathbf{E} = \text{Diag}\left(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2\right). \tag{9}$$

We observe that for any given precoding matrix $\mathbf{P}$, the optimal feedforward and feedback matrices will yield a diagonal MSE matrix, with the individual MSEs being $\mathbf{E}_{ii} = \mathbf{L}_{ii}^2$.

4. OPTIMAL PRECODING MATRIX P

Given the optimal $\mathbf{G}$ and $\mathbf{C}$, the last step is to design a precoding matrix $\mathbf{P}$ to optimize design criteria expressed as functions of individual MSE of each stream, $\mathbf{L}_{ii}^2$. We will first derive two inequalities involving $\mathbf{L}_{ii}$ that enable us to characterize the optimal precoder.

4.1. Preliminaries

To derive the first inequality, we will use the concept of multiplicative majorization:
Multiplicative Majorization [?, ?]: Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^K$ and let $a_{[1]} \geq \ldots \geq a_{[K]}$ denote the elements of $\mathbf{a}$ in descending order. The vector $\mathbf{b}$ is said to multiplicatively majorize $\mathbf{a}$, $\mathbf{a} \prec_\times \mathbf{b}$, if $\prod_{i=1}^j \mathbf{a}_{[i]} \leq \prod_{i=1}^j \mathbf{b}_{[i]}$, for $j = 1, \ldots, K-1$ and $\prod_{i=1}^K \mathbf{a}_{[i]} = \prod_{i=1}^K \mathbf{b}_{[i]}$.
An important example of this definition is:

Lemma 1 Weyl [?]: *Let* $\mathbf{A} \in \mathbb{C}^{K \times K}$ *and let* $\lambda_i(\mathbf{A})$ *and* $\sigma_i(\mathbf{A})$ *denote the eigen values and singular values of* $\mathbf{A}$*, respectively. Then we have* $(|\lambda_1(\mathbf{A})|^2, \ \ldots, \ |\lambda_K(\mathbf{A})|^2) \prec_\times (\sigma_1^2(\mathbf{A}), \ \ldots, \ \sigma_K^2(\mathbf{A}))$. *If* $\mathbf{A}$ *is normal, then* $|\lambda_i(\mathbf{A})| = \sigma_i(\mathbf{A})$.

Applying the above lemma to the positive definite lower triangular matrix $\mathbf{L}$, we obtain out first inequality:

$$(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2) \prec_\times (\sigma_1^2(\mathbf{L}), \ldots, \sigma_K^2(\mathbf{L})). \tag{10}$$

The second inequality involves the more common notation of additive majorization:
Additive Majorization [?]: Let $\mathbf{a}, \ \mathbf{b} \in \mathbb{R}^K$. The vector $\mathbf{b}$ is said to majorize $\mathbf{a}$, $\mathbf{a} \prec \mathbf{b}$, if $\sum_{i=1}^j \mathbf{a}_{[i]} \leq \sum_{i=1}^j \mathbf{b}_{[i]}$, for $j = 1, \ldots, K-1$ and $\sum_{i=1}^K \mathbf{a}_{[i]} = \sum_{i=1}^K \mathbf{b}_{[i]}$

We observe that if elements of $\mathbf{a}$ and $\mathbf{b}$ are positive, then $\mathbf{a} \prec_\times \mathbf{b} \Leftrightarrow \ln(\mathbf{a}) \prec \ln(\mathbf{b})$. Consequently, (10) can be written as:

$$\boldsymbol{l} \prec \mathbf{m}, \tag{11}$$

where $\boldsymbol{l} = (\ln \mathbf{L}_{11}^2, \ldots, \ln \mathbf{L}_{KK}^2)$ and $\mathbf{m} = (\ln \sigma_1^2(\mathbf{L}), \ldots, \ln \sigma_K^2(\mathbf{L}))$.

To derive the second inequality, we will use the following consequence of additive majorization: Any vector $\mathbf{a} \in \mathbb{R}^K$ majorizes its mean vector $\bar{\mathbf{a}}$ whose elements are all equal to the mean; i.e., $\bar{\mathbf{a}}_i = \frac{1}{K}\sum_{i=1}^K \mathbf{a}_i$. That is, $\bar{\mathbf{a}} \prec \mathbf{a}$. Now, since $\mathbf{M} = \mathbf{LL}^H$, we know that $\prod_{i=1}^K \mathbf{L}_{ii}^2 = \det(\mathbf{LL}^H) = \det(\mathbf{M})$. As a result, we have $\sum_{i=1}^K \boldsymbol{l}_i = \ln\det(\mathbf{M})$ and our second inequality is:

$$\bar{\boldsymbol{l}} \prec \boldsymbol{l}, \tag{12}$$

where $\bar{\boldsymbol{l}}_i = \frac{1}{K}\ln\det(\mathbf{M})$.

The proposed designs will be based on the following classes of functions [?]: A real-valued function $f(\mathbf{x})$ defined on a subset $\mathcal{A}$ of $\mathbb{R}^K$ is said to be Schur-convex if $\mathbf{a} \prec \mathbf{b}$ on $\mathcal{A} \Rightarrow f(\mathbf{a}) \leq f(\mathbf{b})$, and is said to be Schur-concave if $\mathbf{a} \prec \mathbf{b}$ on $\mathcal{A} \Rightarrow f(\mathbf{a}) \geq f(\mathbf{b})$. In particular, we will consider communication objectives that can be expressed as the minimization of a functions of the MSEs of each data stream, $g(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2) = g(e^{\boldsymbol{l}_1}, \ldots, e^{\boldsymbol{l}_K})) = g(e^{\boldsymbol{l}})$.

4.2. Schur-convex functions

Examples of objectives that result in $g(e^{\boldsymbol{l}})$ being a Schur-convex function of $\boldsymbol{l}$ include: minimization of the maximum of individual MSEs: $g(e^{\boldsymbol{l}}) = \max_i e^{\boldsymbol{l}_i}$; minimization of the total MMSE: $g(e^{\boldsymbol{l}}) = \sum_i e^{\boldsymbol{l}_i}$; and minimization of the (log) determinant MSE matrix: $\det(\mathbf{E}) = \prod_i e^{\boldsymbol{l}_i}$, which is also Schur-concave function of $\boldsymbol{l}$. For the DFE model, the SINR of the i^{th} stream is given by $\text{SINR}_i = (1/\text{MSE}_i) - 1 = e^{-\boldsymbol{l}_i} - 1$. Hence, many objectives in terms of SINR and BER can be expressed as Schur-convex functions of $\boldsymbol{l}$. As we will show below, the optimal transceiver design is identical for all these objectives.

If $g(e^{\boldsymbol{l}})$ is a Schur convex function of $\boldsymbol{l}$, then from (12) we have that $g(e^{\bar{\boldsymbol{l}}}) \leq g(e^{\boldsymbol{l}})$ and the optimal value is obtained when all $\boldsymbol{l}_i$ are equal to $\boldsymbol{l}_i = \frac{1}{K}\ln\det(\mathbf{M})$; i.e., $\mathbf{E}_{ii} = \mathbf{L}_{ii}^2 = \sqrt[K]{\det(\mathbf{M})}$. Since the objective is an increasing function of the individual MSE, the design goal reduces to minimizing $\det \mathbf{M}$ subject to the power constraint and to the constraint that diagonal elements of the Cholesky factor of $\mathbf{M}$ are all equal. We will start by characterizing the family of solutions that minimize $\det(\mathbf{M})$ subject to the power constraint, then we will show that there is a member of this family that yields a Cholesky factor of $\mathbf{M}$ with equal diagonal elements. Minimizing $\det(\mathbf{M})$ is equivalent to maximizing the Gaussian mutual information, and the family of optimal precoders is obtained using a standard water-filling algorithm [?]. In particular, if $\mathbf{R}_H = \sigma^2\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H} =$

$\mathbf{U}\mathbf{\Lambda_H}\mathbf{U}^H$, the family of optimal precoders takes the form:

$$\mathbf{P} = \mathbf{U}_1\hat{\mathbf{\Phi}}\mathbf{V} = \mathbf{U}_1[\mathbf{\Phi} \quad \mathbf{0}]\mathbf{V}, \tag{13}$$

where $\mathbf{U}_1 \in \mathbb{C}^{N_t \times \hat{K}}$ contains the eigen vectors of $\mathbf{R_H}$ corresponding to the $\hat{K} \leq K$ largest eigen values, $\hat{K}$ and the diagonal positive definite matrix $\mathbf{\Phi}$ are obtained from the water- filling algorithm [**?**], and $\mathbf{V} \in \mathbb{C}^{K \times K}$ is a unitary matrix degree of freedom. This result shows that for DFE based systems designed according to any Schur-convex function of $\boldsymbol{l}$, the optimal solution is information lossless. To complete the design of $\mathbf{P}$, we need to select $\mathbf{V}$ such that the Cholesky decomposition of $\mathbf{M} = \mathbf{L}\mathbf{L}^H$ yields an $\mathbf{L}$ factor with equal diagonal elements. Using (13):

$$\begin{aligned} \mathbf{M} &= \left(\mathbf{V}^H(\mathbf{I} + \hat{\mathbf{\Phi}}^T\mathbf{\Lambda_{H1}}\hat{\mathbf{\Phi}})^{-1/2}\right)\left((\mathbf{I} + \hat{\mathbf{\Phi}}^T\mathbf{\Lambda_{H1}}\hat{\mathbf{\Phi}})^{-1/2}\mathbf{V}\right) \\ &= \mathbf{L}\mathbf{L}^H = \mathbf{R}^H\mathbf{R} = (\mathbf{Q}\mathbf{R})^H(\mathbf{Q}\mathbf{R}), \end{aligned} \tag{14}$$

where $\mathbf{\Lambda_{H1}}$ is the diagonal matrix containing the largest $\hat{K}$ eigen values of $\mathbf{R_H}$, and $\mathbf{Q}$ is a matrix with orthonormal columns. Hence, finding $\mathbf{V}$ is equivalent to finding a $\mathbf{V}$ such that QR decomposition of $(\mathbf{I} + \hat{\mathbf{\Phi}}^T\mathbf{\Lambda_{H1}}\hat{\mathbf{\Phi}})^{-1/2}\mathbf{V}$ has an R-factor with equal diagonal. This problem was solved in [**?**] and $\mathbf{V}$ can be obtained by applying the algorithm in [**?**] to the matrix $(\mathbf{I} + \hat{\mathbf{\Phi}}^T\mathbf{\Lambda_{H1}}\hat{\mathbf{\Phi}})^{-1/2}$; see also [**?**, **?**].

4.3. Schur-concave functions

If $g(e^{\boldsymbol{l}})$ is a Schur-concave function of $\boldsymbol{l}$, then from (11) we have $g(e^{\mathbf{m}}) \leq g(e^{\boldsymbol{l}})$ and the optimal value is obtained when $\mathbf{L}_{ii} = \sigma_i(\mathbf{L})$. According to Lemma 1, this equality holds when $\mathbf{L}$ is normal matrix. Since $\mathbf{L}$ is a lower triangular matrix, in order to be normal it must be a diagonal matrix [**?**]. The optimal $\mathbf{C}$ in that case is $\mathbf{I}$. That is, in the case of Schur-concave functions of $\boldsymbol{l}$, the optimal DFE design results in linear equalization and optimal TH precoding design results in linear precoding. Examples of this class of objectives include minimization of product of the MSEs and general (weighted) geometrical mean of MSEs.

5. SIMULATION STUDY

We consider a system that transmits $K = 4$ streams of 16-QAM symbols over a 4×4 slowly fading independent Rayleigh channel with additive white Gaussian noise. We plot the average bit error rate (BER) against the signal to ratio $P_{\text{total}}/\text{tr}(\mathbf{R}_n)$. We compare the performance of the proposed Schur-convex designs for THP and DFE (which minimize the total MSE among other objectives), with the corresponding linear transceiver design that minimizes total MSE [**?**, **?**], and the optimal linear transceiver that maximizes the mutual information (minimizes $\log\det(\mathbf{E})$) [**?**]. The performance advantages of interference cancellation are quite clear from Fig. 4.

6. CONCLUSION

We developed a unified framework for joint transceiver design of interference (pre-)subtraction schemes for MIMO channels. We obtained optimal designs for two classes of communication objectives, namely those that are Schur-convex and Schur-concave functions of the logarithms of the individual MSEs. For Schur-convex objectives, the optimal transceiver results in equal individual MSEs. For the DFE model, it optimizes both the total MSE and mutual information. For the class Schur-concave objectives, the optimal DFE design results in linear equalization and the optimal TH precoding design results in linear precoding.

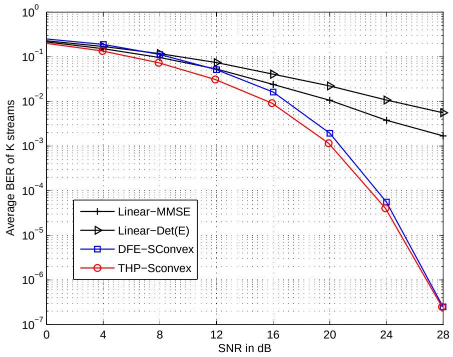
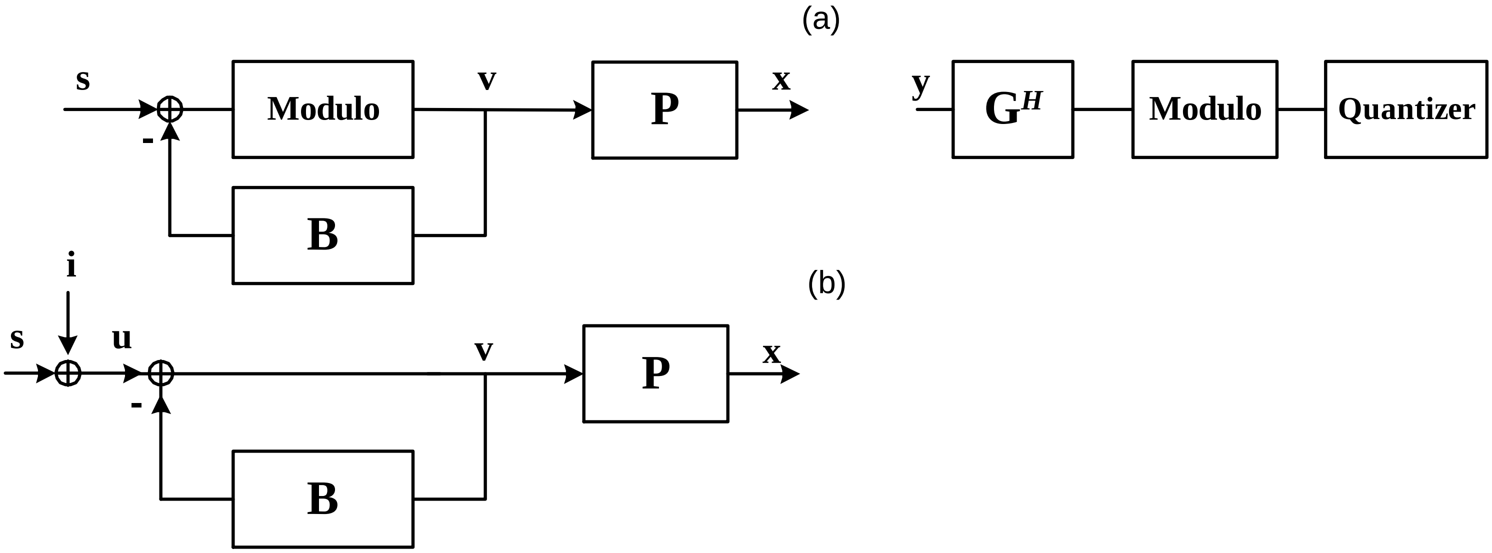
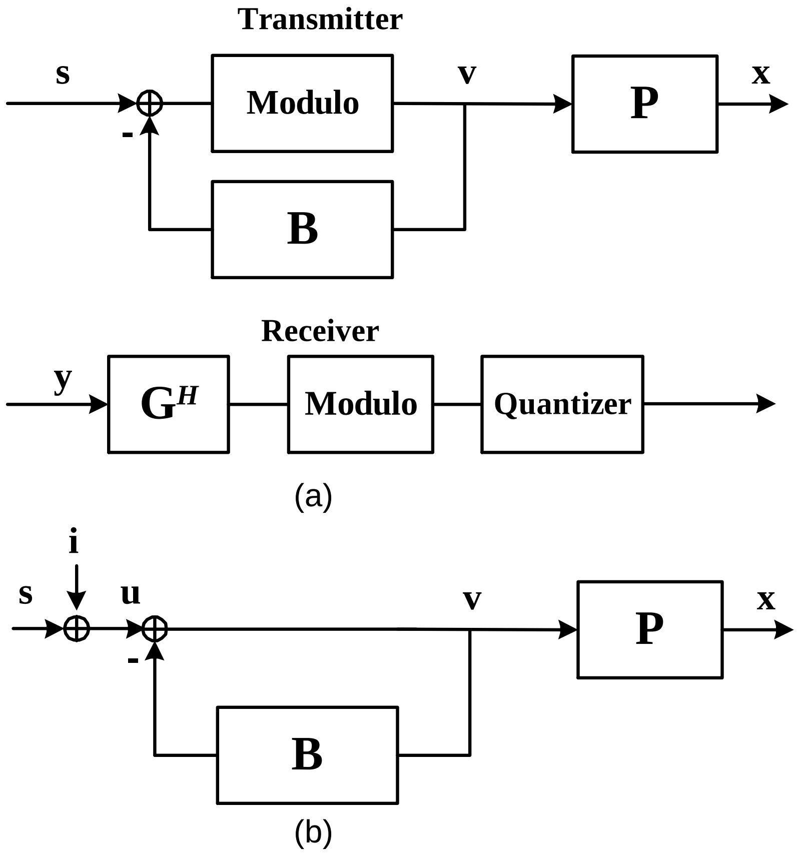


Fig. 4. BERs of the proposed Schur-convex designs and the optimal linear transceivers: minimum MSE (Linear-MMSE), and maximum mutual information (Linear-Det(E)), for $N_t = N_r = K = 4$.


(a)
s
Modulo
v
P
x
y
G^H
Modulo
Quantizer
-
B
(b)
i
s
u
v
P
x
-
B


Transmitter
s
-
Modulo
v
P
x
B
Receiver
y
G^H
Modulo
Quantizer
i
s
u
-
v
P
x
B

(a)

(b)