A Framework for Designing MIMO Systems with Decision Feedback Equalization or Tomlinson-Harashima Precoding

Michael Botros Shenouda, *Student Member, IEEE,* and Timothy N. Davidson, *Member, IEEE,*

***Abstract*—We consider joint transceiver design for point-to-point Multiple-Input Multiple-Output communication systems that implement interference (pre-)subtraction; i.e., Decision Feedback Equalization (DFE) or Tomlinson-Harashima precoding (THP). We develop a unified framework for joint transceiver design of these two dual systems by considering design criteria that are expressed as functions of the (logarithm 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 broad classes of communication objectives, namely those that are Schur-convex and Schur-concave functions of these logarithms. These two classes embrace several design criteria for which the optimal transceiver design has remained an open problem. For Schur-convex objectives, the optimal design results in data streams with equal MSEs. In addition to other desirable properties, this design simultaneously minimizes the total MSE and the average bit error rate, and maximizes the Gaussian mutual information; a property that is not achieved by a linear transceiver. Moreover, we show that the optimal design yields objective values that are superior to the corresponding optimal objective value for a linear transceiver. For Schur-concave objectives, the optimal DFE design results in linear equalization and the optimal THP design results in linear precoding. The proposed design framework can be regarded as a counterpart of the existing framework for linear transceiver design.**

***Index Terms*—Non-linear MIMO transceiver design; unified design framework; majorization; Schur-convexity; convex optimization; Decision Feedback Equalization (DFE); Tomlinson-Harashima precoding (THP);**

I. 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. In point-to-point applications, such schemes typically involve processing of the data streams at the transmitter (precoding) to "match" the transmission to the channel and processing of the received signals (equalization) 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 the potential for performance improvements over the linear approach 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 (TH) precoding and linear equalization, and will be the focus of this paper.

A large number of joint design strategies have been proposed for the class of linear MIMO transceivers (e.g., [1]), and a unified framework that encompasses many of these designs was proposed in [2]. That framework is based on the classes of communication objectives that are Schur-convex or Schur-concave functions of the mean square error (MSE) of each data stream, and encompasses a broad range of design objectives. For DFE-based systems, joint transceiver designs based on a minimum MSE criterion were considered in [3]–[6], and designs subject to a zero-forcing constraint were considered in [7], [8]. However, for many of the design criteria for which (jointly) optimal linear transceivers are known, the jointly optimal DFE-based transceiver has remained an open problem. Furthermore, the development of a unifying design framework for DFE-based transceivers that encompasses these designs has appeared to be a challenging problem. For TH precoding schemes, designs based on minimum MSE criteria were considered in [5], [9], and designs subject to a zero-forcing constraint were considered in [9], [10]. However, the approach in [5] considers a lower bound on the MSE, and the approaches in [9], [10] do not use all the degrees of design freedom available in a single-user system. Hence, the approaches in [5], [9], [10] yield suboptimal designs. In addition to the absence of a minimum MSE transceiver, the design of (jointly) optimal TH-based transceivers for other design criteria, and the development of a unifying framework have remained open problems.

In this paper, we develop a broadly applicable framework for joint transmitter and receiver design for MIMO systems with DFE or TH precoding. (A related DFE-centric framework was developed, independently, in [11], [12].) 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 DFE designs based on the

Manuscript received 10 April 2007; revised 16 November 2007. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and an Ontario Graduate Scholarship in Science and Technology. The work of the second author is also supported by the Canada Research Chairs Program. A preliminary version of this manuscript appears in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* Honolulu, April 2007. See also: http://arxiv.org/abs/cs.IT/0701169

The authors are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada (e-mail: botrosmw@mcmaster.ca; davidson@mcmaster.ca)

Digital Object Identifier 10.1109/JSAC.2008.080216.

overall MSE, these classes of functions embrace other design criteria, such as minimizing the maximum of the individual MSEs, minimizing a general p-norm of the MSEs, and minimizing the product of the individual MSEs, which is equivalent to maximizing the Gaussian mutual information. Moreover, 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; e.g., maximizing the harmonic mean of the SINRs, maximizing a general p-norm of the SINRs, and minimizing the the total BER of all streams. Interestingly, the optimal design for both Schur-convex and Schur-convex objectives yields a diagonal MSE matrix. Hence, communication over the MIMO channel is decomposed into a number of uncorrelated subchannels. For Schur-convex objectives the optimal design results in data streams with equal MSEs. This property is not achieved by the previously proposed (suboptimal) designs for TH precoding systems (e.g., [5], [9]), and hence ordering the symbols prior to interference subtraction is necessary for those designs, as it is in multi-user schemes [13]. This ordering is unnecessary for the optimal transceiver designs derived herein. Another property of our optimal design for Schur-convex objectives is that it simultaneously minimizes the total MSE, minimizes the average bit error rate, and maximizes the Gaussian mutual information. This property is not achieved by the optimal linear transceiver. For any Schur-convex objective, our optimal design yields an objective value that is superior to the corresponding optimal objective value for a linear transceiver. For Schur-concave objectives, the optimal DFE design results in linear equalization and optimal TH precoding design results in linear precoding. From a broader prospective, the proposed framework can be viewed as a counterpart for the design of DFE-based and TH-precoding-based transceivers of the unified framework for the design of linear transceivers in [2].

Our notation is as follows: Boldface type is used to denote vectors and matrices ; $\mathbf{a}_i$ denotes the i^{th} element of the vector $\mathbf{a}$, $\mathbf{A}_{ij}$ denotes the element at the intersection of the i^{th} row and j^{th} column of the matrix $\mathbf{A}$; and $\mathbf{A}^H$ denotes the conjugate transpose of $\mathbf{A}$. The terms $\text{tr}(\mathbf{A})$, $\det(\mathbf{A})$, and $\|\mathbf{A}\|_F$ denote the trace, determinant and Frobenius norm of $\mathbf{A}$, respectively. The notation $\text{Diag}(\mathbf{x})$ denotes the diagonal matrix whose elements are the elements of $\mathbf{x}$.

II. Two System Models

We consider a generic MIMO communication system described by the channel matrix $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$, e.g., [14], and we denote by K the number of data streams transmitted simultaneously over the channel. We will consider the design of two communication architectures: systems with linear precoding (pre-equalization) at the transmitter and DFE at the receiver; and systems with Tomlinson-Harashima precoding at the transmitter and linear equalization at the receiver. We will assume that full channel state information (CSI) is available at both the transmitter and the receiver. However, the framework developed herein has recently been extended to scenarios with limited CSI at the transmitter; see [15].

A. Decision Feedback Equalization

As shown in the DFE model in Fig. 1, the vector $\mathbf{s} \in \mathbb{C}^K$ that contains the current data symbol of each stream is linearly precoded by the matrix $\mathbf{P} \in \mathbb{C}^{n_t \times K}$ to generate the transmitted vector

$$\mathbf{x} = \mathbf{P}\mathbf{s}, \tag{1}$$

where we assume, without loss of generality, that $\text{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$. Hence, the average transmitted power constraint can be written as $\text{E}_{\mathbf{s}}\{\mathbf{x}^H\mathbf{x}\} = \text{tr}(\mathbf{P}^H\mathbf{P}) \leq P_{\text{total}}$. The received vector $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{n}$ is the vector of additive noise samples which is assumed to have zero-mean and a covariance matrix $\text{E}\{\mathbf{n}\mathbf{n}^H\} = \mathbf{R}_n$. As shown in Fig. 1, the DFE is implemented using a feedforward matrix $\mathbf{G} \in \mathbb{C}^{K \times n_r}$ and a feedback matrix filter $\mathbf{B} \in \mathbb{C}^{K \times K}$. In this scenario, the detection of the k^{th} symbol is preceded by subtracting the effect of previously decoded symbols. Assuming correct previous decisions, the input to the quantizer, $\hat{\mathbf{s}}$, can be written as (e.g., [6])

$$\hat{\mathbf{s}}_{\text{DFE}} = (\mathbf{G}\mathbf{H}\mathbf{P} - \mathbf{B})\mathbf{s} + \mathbf{G}\mathbf{n}, \tag{3}$$

where $\mathbf{B}$ is a strictly lower triangular matrix.[1] Using the error signal $\mathbf{e} = \hat{\mathbf{s}}_{\text{DFE}} - \mathbf{s}$, we can define the Mean Square Error matrix,

$$\begin{aligned}\mathbf{E} = \text{E}_{\mathbf{s}}\{\mathbf{e}\mathbf{e}^H\} = \mathbf{C}\mathbf{C}^H - \mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H - \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}^H \\ + \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H + \mathbf{G}\mathbf{R}_n\mathbf{G}^H,\end{aligned} \tag{4}$$

where $\mathbf{C} = \mathbf{I} + \mathbf{B}$ is a unit diagonal lower triangular matrix.

B. Tomlinson-Harashima Precoding

As shown in Fig. 2(a), in a TH precoding system the transmitter performs successive interference pre-subtraction and 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 $\text{E}_{\mathbf{s}}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$. The Voronoi region, $\mathcal{V}$, of this constellation 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 $\mathbf{i}_k = \mathbf{i}_k^{re} D + j\,\mathbf{i}_k^{imag} D$ to $\mathbf{s}_k$, where $\mathbf{i}_k^{re}$, $\mathbf{i}_k^{imag} \in \mathbb{Z}$. Using this observation, we obtain the linearized model of the transmitter shown in Fig. 2(b), e.g., [9], in which

$$\mathbf{v} = (\mathbf{I} + \mathbf{B})^{-1}\mathbf{u} = \mathbf{C}^{-1}\mathbf{u}, \tag{5}$$

where $\mathbf{u} = \mathbf{i} + \mathbf{s}$ is the modified data symbol and $\mathbf{C} = \mathbf{I} + \mathbf{B}$. As a result of the modulo operation, the elements of $\mathbf{v}$ are almost uncorrelated and uniformly distributed over the Voronoi region $\mathcal{V}$ [9, Th. 3.1]. Therefore, the symbols of $\mathbf{v}$ will have slightly higher average energy than the input symbols $\mathbf{s}$. This slight increase in the average energy is termed precoding loss [9]. For example, for square M-ary QAM we have $\sigma_v^2 = \text{E}\{|\mathbf{v}_k|^2\} =$

[1] In general, the estimator in (3) is biased, but the effect of this bias can be mitigated by scaling the decision regions of the quantizer [16]. At operating points at which one can reasonably assume correct previous decisions, the effect of the bias is typically small [16].

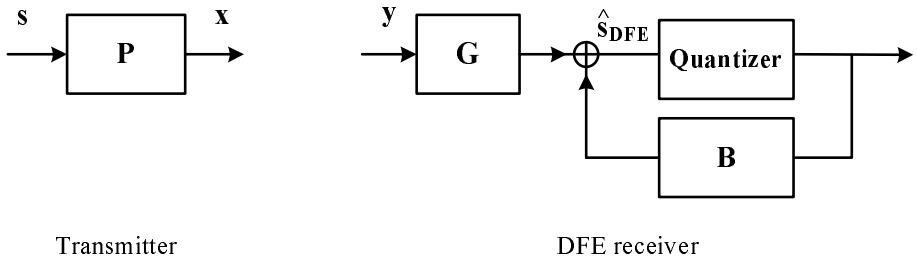

Fig. 1. MIMO transceiver using Decision Feedback Equalization.

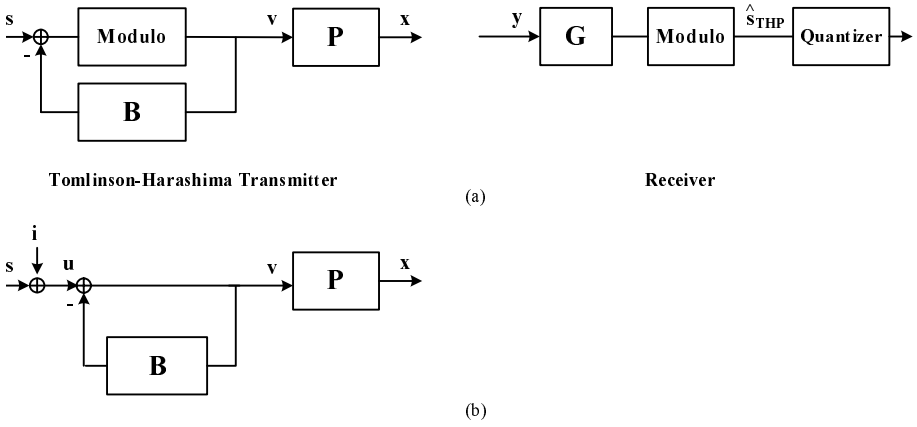

Fig. 2. (a) MIMO transceiver with Tomlinson-Harashima precoding (b) Equivalent linear transmitter model for Tomlinson-Harashima precoding system

$\frac{M}{M-1}\mathrm{E}\{|\mathbf{s}_k|^2\}$ for all k except the first one [9]. For moderate to large values of M this power increase can be neglected and the approximation $\mathrm{E}\{\mathbf{v}\mathbf{v}^H\} = \mathbf{I}$ is often used; e.g., [5], [10]. If we assume negligible precoding loss, the average transmitted power constraint can be written as $\mathrm{E}_{\mathbf{v}}\{\mathbf{x}^H\mathbf{x}\} = \mathrm{tr}(\mathbf{P}^H\mathbf{P}) \leq P_{\text{total}}$.

The vector of received signals in a TH precoded system can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{C}^{-1}\mathbf{u} + \mathbf{n}, \tag{6}$$

where $\mathbf{n}$ is the vector of additive noise which is assumed to have zero-mean and a covariance matrix $\mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \mathbf{R}_n$. At the receiver, the feedforward processing matrix $\mathbf{G}$ is used to obtain an estimate $\hat{\mathbf{u}} = \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}^{-1}\mathbf{u} + \mathbf{G}\mathbf{n}$ of the modified data symbols $\mathbf{u}$. Following this linear receive processing step, the modulo operation is used to obtain $\hat{\mathbf{s}}_{\text{THP}}$ by eliminating the effect of the periodic extension of the constellation caused by the integer vector $\mathbf{i}$. In terms of the modified data symbols, the error signal

$$\mathbf{e} = \hat{\mathbf{u}} - \mathbf{u} = \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{v} + \mathbf{G}^H\mathbf{n} - \mathbf{C}\mathbf{v} \tag{7}$$

can be used to define a Mean Square Error matrix

$$\begin{aligned}\mathbf{E} = \mathrm{E}_{\mathbf{v}}\{\mathbf{e}\mathbf{e}^H\} = \mathbf{C}\mathbf{C}^H - \mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H - \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}^H \\ + \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H + \mathbf{G}\mathbf{R}_n\mathbf{G}^H.\end{aligned} \tag{8}$$

Assuming negligible precoding loss and that the vector $\mathbf{i}$ is eliminated by the receiver modulo operation (which occurs with high probability, even at reasonably low SNRs), the error signal in (7) is equivalent to $\hat{\mathbf{s}}_{\text{THP}} - \mathbf{s}$. Hence, the mean square error matrix, $\mathbf{E}$, of the estimate $\hat{\mathbf{s}}_{\text{THP}}$ of the TH precoding model is the same as that of the estimate $\hat{\mathbf{s}}_{\text{DFE}}$ of the DFE model under the assumption of correct previous decisions in the DFE.

C. General Model

From (4) and (8), we observe that the MSE matrix of both systems can be rewritten as:

$$\mathbf{E} = \mathbf{C}\mathbf{C}^H - \mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H - \mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}^H + \mathbf{G}\mathbf{R}_y\mathbf{G}^H, \tag{9}$$

where $\mathbf{R}_y = \mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H + \mathbf{R}_n$. It can also be observed that linear transceivers are a special subclass of both system models with the feedback matrix $\mathbf{B} = \mathbf{0}$ (or, equivalently, $\mathbf{C} = \mathbf{I}$); see Figs 1 and 2. Our objective is to jointly design the matrices $\mathbf{G}, \mathbf{C}$ and $\mathbf{P}$ according to criteria that are functions of $\mathbf{E}$, subject to a constraint on the average transmitted power.

III. Optimal Feedforward and Feedback Matrices

We will consider the joint design of the transceiver matrices $\mathbf{G}, \mathbf{C}$ and $\mathbf{P}$ so as to optimize system design criteria that are expressed as (increasing) functions of the (logarithm of the) MSE of each individual data stream, $\mathbf{E}_{ii}$, subject to the transmitted power constraint $\mathrm{tr}(\mathbf{P}^H\mathbf{P}) \leq P_{\text{total}}$. We will adopt a three-step design approach. First, an expression for the optimal feedforward matrix $\mathbf{G}$ will be found as a function of $\mathbf{C}$ and $\mathbf{P}$. Second, using the expression for the optimal $\mathbf{G}$, an expression for the optimal $\mathbf{C}$ will be found as a function of $\mathbf{P}$. Finally, using the obtained expressions for the optimal $\mathbf{G}$ and $\mathbf{C}$, we will design the optimal precoder $\mathbf{P}$.

A. Optimal feedforward matrix $\mathbf{G}$

For given $\mathbf{C}$ and $\mathbf{P}$, the MSE of the i^{th} data stream, $\mathbf{E}_{ii}$, is a convex quadratic function of the i^{th} row of $\mathbf{G}$, and is independent of other rows. Therefore, the rows of $\mathbf{G}$ can be independently optimized to minimize the individual MSEs, and the resulting $\mathbf{G}$ is optimal for any transceiver objective that is an increasing function of the individual MSEs. (A similar property was observed in [2] for linear transceivers.) Since $\mathbf{G}$ is unconstrained and the MSE of the i^{th} data stream is a smooth convex function of the i^{th} row of $\mathbf{G}$, we can obtain an expression for optimal $\mathbf{G}$ by setting the gradient of $\mathbf{E}_{ii}$ with respect to the i^{th} row of $\mathbf{G}$ to zero. Hence, the optimal $\mathbf{G}$ can be written as

$$\mathbf{G} = \mathbf{C}\mathbf{P}^H\mathbf{H}^H\mathbf{R}_y^{-1}. \tag{10}$$

Using this expression, the MSE matrix for a system with the optimal $\mathbf{G}$ can be written as:

$$\mathbf{E} = \mathbf{C}(\mathbf{I} + \mathbf{P}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{P})^{-1}\mathbf{C}^H = \mathbf{C}\mathbf{M}\mathbf{C}^H, \tag{11}$$

where the matrix inversion lemma has been used, and $\mathbf{M} = (\mathbf{I} + \mathbf{P}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{P})^{-1}$.

B. Optimal feedback matrix $\mathbf{B}$

From (11) we observe that the MSE of each data stream, $\mathbf{E}_{ii}$, is a convex quadratic function of the i^{th} row of $\mathbf{C} = \mathbf{I}+\mathbf{B}$ and is independent of the other rows. Using a similar argument to that for $\mathbf{G}$ above, the matrix $\mathbf{C}$ whose rows independently minimize the individual MSEs is optimal for the transceiver objectives that we will consider. However, $\mathbf{C}$ is constrained to be a unit diagonal lower triangular matrix and these constraints must be incorporated in the design. To do so, we observe that the matrix $\mathbf{C}$ 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 $\text{tr}(\mathbf{C}\mathbf{M}\mathbf{C}^H)$ subject to $\mathbf{C}$ being unit diagonal lower triangular matrix. Using the Cholesky decomposition

$$\mathbf{M} = \mathbf{L}\mathbf{L}^H, \tag{12}$$

where $\mathbf{L}$ is a lower triangular matrix with positive real diagonal elements, we can rewrite the objective as $\text{tr}(\mathbf{C}\mathbf{M}\mathbf{C}^H) = \|\mathbf{C}\mathbf{L}\|_F^2$, where the product $\mathbf{C}\mathbf{L}$ is a positive definite lower triangular matrix [17]. Let $\lambda_1(\mathbf{C}\mathbf{L}) \geq \ldots \geq \lambda_K(\mathbf{C}\mathbf{L})$ and $\sigma_1(\mathbf{C}\mathbf{L}) \geq \ldots \geq \sigma_K(\mathbf{C}\mathbf{L})$ denote the ordered eigenvalues and singular values, respectively, of the matrix $\mathbf{C}\mathbf{L}$. Then the unit diagonal lower triangular $\mathbf{C}$ that minimizes $\text{tr}(\mathbf{C}\mathbf{M}\mathbf{C}^H)$ can be obtained using the following lower bound,

$$\|\mathbf{C}\mathbf{L}\|_F^2 = \sum_{i=1}^{K}\sigma_i^2(\mathbf{C}\mathbf{L}) \geq \sum_{i=1}^{K}\lambda_i^2(\mathbf{C}\mathbf{L}) \tag{13}$$

$$= \sum_{i=1}^{K}[\mathbf{C}\mathbf{L}]_{ii}^2 = \sum_{i=1}^{K}\mathbf{L}_{ii}^2, \tag{14}$$

where the bound in (13) is obtained by applying Weyl's inequality [18], and (14) follows from the fact that $\mathbf{C}\mathbf{L}$ is lower triangular and $\mathbf{C}$ is unit diagonal. The expression on the right hand side of (14) is a lower bound on $\|\mathbf{C}\mathbf{L}\|_F^2$ that is independent of $\mathbf{C}$. Furthermore, the inequality in (13) is satisfied with equality when the matrix is normal [18]. Since our matrix $\mathbf{C}\mathbf{L}$ is a triangular matrix, it can only be normal if it is diagonal [17, pp 103]. Therefore, the matrix $\mathbf{C}$ that attains the lower bound in (14), and hence is optimal, is

$$\mathbf{C} = \text{Diag}\left(\mathbf{L}_{11}, \ldots, \mathbf{L}_{KK}\right)\mathbf{L}^{-1}. \tag{15}$$

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

$$\mathbf{E} = \text{Diag}\left(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2\right). \tag{16}$$

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 $\mathbf{E}_{ii} = \mathbf{L}_{ii}^2$.

C. Optimality in the sense of maximizing individual SINRs

For any given channel and precoder, the minimum MSE design of the matrices $\mathbf{G}$ and $\mathbf{B}$ for a DFE system, is also optimal in sense of maximizing the signal-to-interference-plus-noise (SINR) of each stream [19]–[21]. Using this optimal minimum MSE design of the feedforward and feedback matrices, the SINR of the i^{th} stream is given by [19], [22]

$$\text{SINR}_i = \left(1/\mathbf{E}_{ii}\right) - 1. \tag{17}$$

Under the assumptions stated in Section II, the estimate vector $\hat{\mathbf{s}}_{\text{THP}}$ has the same covariance matrix as the vector $\hat{\mathbf{s}}_{\text{DFE}}$ at the input to the quantizer in the DFE system. Hence, the individual SINRs for both systems are the same for any given input covariance matrix, $\text{E}\{\mathbf{s}\mathbf{s}^H\}$, and noise covariance matrix, $\mathbf{R}_n$. An analogous relation between SINR_i and $\mathbf{E}_{ii}$ holds under a zero-forcing constraint for both the DFE model (e.g., [22]), and the TH precoding model under similar assumptions to those stated in Section II; e.g., [10]. (Similar relations also hold in the multiuser case; e.g., [23].) Since linear precoding is a special subclass of both models when $\mathbf{B} = \mathbf{0}$, the same relation between SINR_i and $\mathbf{E}_i$ holds for minimum MSE design of the receiver matrix $\mathbf{G}$; e.g., [2]. Using the expression for the individually minimized MSEs in (16), the individually maximized SINR of each data stream is given by

$$\text{SINR}_i = \left(1/\mathbf{L}_{ii}^2\right) - 1. \tag{18}$$

IV. Design of the Precoding Matrix: Preliminaries

Given the expressions for the optimal $\mathbf{G}$ and $\mathbf{C}$, the remaining step is to design a precoding matrix $\mathbf{P}$ to optimize design criteria that are expressed as functions of the individual MSE of each stream, $\mathbf{L}_{ii}^2$. We will first derive two inequalities involving $\mathbf{L}_{ii}$ that will enable us to characterize the optimal precoder. These inequalities will depend on the concepts of multiplicative and additive majorization [24].

A. A Multiplicative Majorization Inequality

The first inequality is derived using the concept of multiplicative majorization [18], [21], [24].

Definition 1 (Multiplicative Majorization): For a vector $\mathbf{a} \in \mathbb{R}^K$, let $a_{[1]}, \ldots, a_{[K]}$ denote the re-ordering of the elements of $\mathbf{a}$ in a non-decreasing order; i.e., $a_{[1]} \geq \ldots \geq a_{[K]}$. Let $\mathbb{R}_+$ denote the set of positive real numbers, and let

$\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^K$. The vector $\mathbf{b}$ is said to multiplicatively majorize $\mathbf{a}$, $\mathbf{a} \prec_\times \mathbf{b}$, if

$$\prod_{i=1}^{j} \mathbf{a}_{[i]} \leq \prod_{i=1}^{j} \mathbf{b}_{[i]} \qquad \text{for } j = 1, \ldots, K-1, \tag{19a}$$

$$\prod_{i=1}^{K} \mathbf{a}_{[i]} = \prod_{i=1}^{K} \mathbf{b}_{[i]}. \tag{19b}$$

□

An important example of the multiplicative majorization is the relation between the eigenvalues and singular values of a square matrix, and is given by the following lemma.

Lemma 1 (Weyl [18]): Let $\mathbf{A} \in \mathbb{C}^{K \times K}$ and let $\lambda_i(\mathbf{A})$ and $\sigma_i(\mathbf{A})$ denote the eigenvalues and singular values of $\mathbf{A}$, respectively. Then we have $\left[|\lambda_1(\mathbf{A})|^2, \ \ldots, \ |\lambda_K(\mathbf{A})|^2\right] \prec_\times \left[\sigma_1^2(\mathbf{A}), \ \ldots, \ \sigma_K^2(\mathbf{A})\right]$. 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

$$\left[\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2\right] \prec_\times \left[\sigma_1^2(\mathbf{L}), \ldots, \sigma_K^2(\mathbf{L})\right]. \tag{20}$$

B. An Additive Majorization Inequality

The second inequality involves the more common notion of additive majorization [24].

Definition 2 (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

$$\sum_{i=1}^{j} \mathbf{a}_{[i]} \leq \sum_{i=1}^{j} \mathbf{b}_{[i]} \qquad \text{for } j = 1, \ldots, K-1, \tag{21a}$$

$$\sum_{i=1}^{K} \mathbf{a}_{[i]} = \sum_{i=1}^{K} \mathbf{b}_{[i]}. \tag{21b}$$

□

We observe that if elements of $\mathbf{a}$ and $\mathbf{b}$ are positive, then $\mathbf{a} \prec_\times \mathbf{b} \Leftrightarrow \log(\mathbf{a}) \prec \log(\mathbf{b})$. Consequently, (20) can be written as:

$$\boldsymbol{l} \prec \mathbf{m}, \tag{22}$$

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

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 $\mathbf{M} = \mathbf{L}\mathbf{L}^H$, we know that $\prod_{i=1}^{K} \mathbf{L}_{ii}^2 = \det(\mathbf{L}\mathbf{L}^H) = \det(\mathbf{M})$. As a result, we have $\sum_{i=1}^{K} \boldsymbol{l}_i = \log\det(\mathbf{M})$ and hence

$$\overline{\boldsymbol{l}} \prec \boldsymbol{l}, \tag{23}$$

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

C. Schur-convex and Schur-concave functions

The proposed designs will be based on the following classes of functions [24].

Definition 3 (Schur-convex and Schur-concave 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 increasing functions of the MSEs of each data stream, $g(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2) = g(e^{\boldsymbol{l}_1}, \ldots, e^{\boldsymbol{l}_K}) = g(e^{\boldsymbol{l}})$, that are either Schur-convex or Schur-concave functions of $\boldsymbol{l}$.

V. Optimal Precoding Matrix: Schur-convex objectives

In this section, we will present a closed-form expression for the optimal precoding matrix $\mathbf{P}$ for the class of Schur-convex objectives. We will also study the properties of the optimal solution and compare it to optimal linear transceiver designs. Finally, we will present examples of design objectives $g(e^{\boldsymbol{l}})$ that are Schur-convex functions of $\boldsymbol{l}$.

A. Optimal Precoding Matrix

If $g(e^{\boldsymbol{l}})$ is a Schur-convex function of $\boldsymbol{l}$, then from (23) we have that $g(e^{\overline{\boldsymbol{l}}}) \leq g(e^{\boldsymbol{l}})$, and that equality is obtained if the elements of $\boldsymbol{l}$ are equal. Our approach to finding the optimal precoder is to characterize the family of procoders that minimize the lower bound $g(e^{\overline{\boldsymbol{l}}})$ subject to the power constraint, and then to show that within this family there is a precoder that results in all of the elements of $\boldsymbol{l}$ being equal, and hence attains the minimized lower bound.

Since the objective is an increasing function of the individual MSEs, and since $\overline{\boldsymbol{l}}_i = \frac{1}{K} \log\det(\mathbf{M})$, where $\mathbf{M}$ was defined following (11), the problem of minimizing the lower bound subject to the power constraint can be formulated as:

$$\max_{\mathbf{P}} \quad \log\det(\mathbf{I} + \sigma^2 \mathbf{P}^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}\mathbf{P}) \tag{24a}$$

$$\text{subject to} \quad \operatorname{tr}(\mathbf{P}^H\mathbf{P}) \leq P_{\text{total}}. \tag{24b}$$

This formulation is equivalent to maximizing the Gaussian mutual information, and hence the family of optimal precoders is obtained using a standard water-filling algorithm [25]. To state this family, we use the eigenvalue decomposition

$$\mathbf{R}_{\mathbf{H}} = \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} = \mathbf{U}_{\mathbf{H}} \mathbf{\Lambda}_{\mathbf{H}} \mathbf{U}_{\mathbf{H}}^H, \tag{25}$$

where $\mathbf{\Lambda}_{\mathbf{H}} = \operatorname{Diag}(\lambda_{\mathbf{H},1}, \ldots)$, and $\lambda_{\mathbf{H},i}$ are eigenvalues of $\mathbf{R}_{\mathbf{H}}$ in descending order. In the water-filling algorithm, power is allocated to K_{wf} eigenvalues of $\mathbf{R}_{\mathbf{H}}$, where K_{wf} is the maximum integer j satisfying $(P_{\text{total}} + \sum_{i=1}^{j} \lambda_{\mathbf{H},i}^{-1}) \geq j/\lambda_{\mathbf{H},j}$, [25]. If we define $\hat{K} = \min(K_{\text{wf}}, K)$, the family of optimal precoders can be written as

$$\mathbf{P} = \mathbf{U}_{\mathbf{H},1} \hat{\mathbf{\Phi}} \mathbf{V} = \mathbf{U}_{\mathbf{H},1} [\mathbf{\Phi} \quad \mathbf{0}] \mathbf{V}, \tag{26}$$

where $\mathbf{U}_{\mathbf{H},1} \in \mathbb{C}^{N_t \times \hat{K}}$ contains the eigenvectors of $\mathbf{R}_{\mathbf{H}}$ corresponding to the largest $\hat{K}$ eigenvalues, $\mathbf{V} \in \mathbb{C}^{K \times K}$ is a unitary matrix degree of freedom, and the diagonal matrix $\mathbf{\Phi}$ is

$$\mathbf{\Phi}_{ii} = \mu - 1/\lambda_{\mathbf{H},i}, \tag{27}$$

where the "water" level μ is given by $\frac{1}{\hat{K}}(P + \sum_{i=1}^{\hat{K}} \lambda_{\mathbf{H},i}^{-1})$.

To complete the design of $\mathbf{P}$, we need to select the unitary matrix $\mathbf{V}$ in (26) so that the minimized lower bound is attained; i.e., so that the Cholesky decomposition of $\mathbf{M} = \mathbf{L}\mathbf{L}^H$ yields an $\mathbf{L}$ factor with equal diagonal elements. Using (26),

$$\begin{aligned}\mathbf{M} &= \left(\mathbf{V}^H(\mathbf{I}+\hat{\mathbf{\Phi}}^T\mathbf{\Lambda}_{\mathbf{H},1}\hat{\mathbf{\Phi}})^{-1/2}\right)\left((\mathbf{I}+\hat{\mathbf{\Phi}}^T\mathbf{\Lambda}_{\mathbf{H},1}\hat{\mathbf{\Phi}})^{-1/2}\mathbf{V}\right)\\ &= \mathbf{L}\mathbf{L}^H = \mathbf{R}^H\mathbf{R} = (\mathbf{Q}\mathbf{R})^H(\mathbf{Q}\mathbf{R}),\end{aligned} \tag{28}$$

where $\mathbf{\Lambda}_{\mathbf{H},1}$ is the diagonal matrix containing the largest $\hat{K}$ eigenvalues of $\mathbf{R}_{\mathbf{H}}$, 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 [7] and $\mathbf{V}$ can be obtained by applying the algorithm in [7] to the matrix $(\mathbf{I}+\hat{\mathbf{\Phi}}^T\mathbf{\Lambda}_{\mathbf{H},1}\hat{\mathbf{\Phi}})^{-1/2}$; see also [6], [26], [27].

B. Properties of the optimal design

In this section we describe some interesting properties of the optimal transceiver design for Schur-convex objectives.

1) Independence of the optimal transceiver design from the design objective $g(e^{\boldsymbol{l}_i})$: We observe that the above derivation of the optimal precoder design is independent of the actual design objective, $g(e^{\boldsymbol{l}})$. (A similar property holds for linear transceiver design, but with objectives that are Schur-convex functions of the individual MSEs themselves.) Therefore, the desirable properties of the DFE transceiver that minimizes the total MSE generalize to other Schur-convex objectives for both DFE and TH models. For example, the DFE transceiver that minimizes the total MSE has asymptotically the same symbol error rate as the transceiver that employs the optimal precoder with maximum likelihood detection [27]. This property is now applicable to all DFE and TH transceivers with Schur-convex objectives.

2) For any Schur-convex objective $g(e^{\boldsymbol{l}_i})$, the optimal transceiver is information lossless: Since maximizing the Gaussian mutual information is a Schur-convex objective, it follows that the optimal design for any Schur-convex objective is information lossless, in the sense that optimizing the chosen objective does not incur any reduction of the Gaussian mutual information. This result generalizes the information lossless property of MMSE-DFE receivers (e.g., [3], [20]), and that of minimum MSE DFE-based transceivers [6], to designs for DFE and TH transceivers with an arbitrary Schur-convex objective, $g(e^{\boldsymbol{l}})$. This property does not hold in general for the linear transceiver designs because the precoder that maximizes the Gaussian mutual information does not necessarily optimize other criteria.

3) Relation to linear transceiver designs: Using the majorization results in (22) and (23), we can show the following interesting result for any Schur-convex objective $g(e^{\boldsymbol{l}})$.

Proposition 1: For design criteria with a Schur-convex objective $g(e^{\boldsymbol{l}})$, the optimal THP or DFE design yields a lower bound on the objective value obtained by any linear transceiver. □

Proof: For any linear transceiver, $\mathbf{C} = \mathbf{I}$. It follows from (15) that $\mathbf{L}$ is diagonal and hence $\mathbf{L}_{ii}^2 = \sigma_i^2(\mathbf{L})$, or equivalently $\boldsymbol{l} = \mathbf{m}$. Since the optimal THP or DFE transceiver corresponds to $\boldsymbol{l} = \overline{\boldsymbol{l}}$ and we have $\overline{\boldsymbol{l}} \prec \mathbf{m}$, it follows that $g(\overline{\boldsymbol{l}}) \leq g(\mathbf{m})$, for any Schur-convex objective $g(\cdot)$. ■

This result shows that the optimal DFE or THP transceiver for any Schur-convex objective $g(e^{\boldsymbol{l}})$ will yield an objective value that is less than or equal to the objective value achieved by the optimal linear transceiver for the same objective. Furthermore, a stronger results can be obtained by considering the subclass of strictly Schur-convex objectives. For this class of objectives, $f(\mathbf{a}) < f(\mathbf{b})$, whenever $\mathbf{a} \prec \mathbf{b}$ and $\mathbf{a}$ is not a permutation of $\mathbf{b}$. Since the optimal transceiver corresponds to $\boldsymbol{l} = \overline{\boldsymbol{l}}$, and any linear transceiver corresponds to $\boldsymbol{l} = \mathbf{m}$, it follows from $\overline{\boldsymbol{l}} \prec \mathbf{m}$ that $g(e^{\overline{\boldsymbol{l}}}) < g(e^{\mathbf{m}})$, for every strictly Schur-convex function $g(\cdot)$ whenever $\mathbf{m}$ is not equal to a permutation of $\overline{\boldsymbol{l}}$. Since all elements of $\overline{\boldsymbol{l}}$ are equal, it follows that $g(e^{\overline{\boldsymbol{l}}}) < g(e^{\mathbf{m}})$ whenever $\overline{\boldsymbol{l}} \neq \mathbf{m}$. The case $\overline{\boldsymbol{l}} = \mathbf{m}$ corresponds to the optimal design of $\mathbf{L}$ being a diagonal matrix with equal diagonal elements; i.e., a scaled identity matrix. This case can arise from water-filling over $K \leq K_{\text{wf}}$ equal eigenvalues of the matrix $\mathbf{R}_{\mathbf{H}}$.

C. Examples of Schur-convex objectives

In this section we present examples of design objectives that are Schur-convex functions of $\boldsymbol{l}$, the vector of logarithms of the individual MSEs. (Sketches of the proofs are provided in Appendix A.) Before we do so, we point out that by using the composition properties of Schur-convex functions [24] one can prove the following result.

Lemma 2: Let $\mathbf{y} = \mathbf{e}^{\boldsymbol{l}}$. If $g(\mathbf{y})$ is Schur-convex in $\mathbf{y}$, then $g(e^{\boldsymbol{l}})$ is Schur-convex in $\boldsymbol{l}$.

Using this lemma and the results in [2], functions such as the total MSE and the average BER can be shown to be to Schur-convex functions of $\boldsymbol{l}$. However, by analyzing $g(e^{\boldsymbol{l}})$ directly, we will obtain stronger results. For example, we will show that the total MSE is *strictly* Schur-convex in $\boldsymbol{l}$. (It is not strictly Schur-convex in the MSEs themselves.) We will also show that the average BER of certain constellations, including 16-QAM, is a Schur-convex function of $\boldsymbol{l}$ for the entire range of the MSE, whereas it is a Schur-convex function of the MSEs only for limited ranges of the MSE [2]. In addition, by taking the direct approach we will be able to show that several objectives that are not Schur-convex functions of the MSEs are Schur-convex functions of the logarithm of the MSEs; e.g., the Gaussian mutual information and the geometric mean of the SINRs.

1) Minimization of the total MSE: Minimization of total MSE (or the arithmetic mean of the MSEs) corresponds to minimization of

$$g(e^{\boldsymbol{l}}) = \sum_{i=1}^{K} e^{\boldsymbol{l}_i}, \tag{29}$$

which is a strictly Schur-convex function of $\boldsymbol{l}$. Hence, the optimal precoder is given by the closed-form expression derived in Section V-A. For the DFE model, transceiver design based on minimization of the total MSE was considered in [6], and the solution therein is, as expected, the same as that in Section V-A. For the TH precoding model, a design approach based on a bound on the total MSE was presented in [5], but that approach does not necessarily minimize the total MSE. Furthermore, the TH designs in [9], [10] do not exploit all the available degrees of design freedom. Using the approach

presented in this section, we obtain a jointly optimal design for TH precoding model for the total MSE objective.

2) Minimization of product of MSEs and maximization of Gaussian mutual information: Given the diagonal structure of the matrix $\mathbf{E}$ in (16), minimization of the product of the MSEs (or the geometric mean of the MSEs) is equivalent to minimization of the determinant of $\mathbf{E}$. Furthermore, maximization of the Gaussian mutual information is equivalent to minimization of $\log\det(\mathbf{E})$, [3]. Therefore, these three objectives are equivalent and correspond to minimization of

$$g(e^{\boldsymbol{l}}) = \log \prod_{i=1}^{K} e^{\boldsymbol{l}_i} = \sum_{i=1}^{K} \boldsymbol{l}_i. \tag{30}$$

In Appendix A, we show that $g(e^{\boldsymbol{l}})$ is both a Schur-convex and a Schur-concave function of $\boldsymbol{l}$. Hence, the optimal design in (26) is information lossless for both the DFE and TH precoding models. (Examples of existing designs that apply these criteria to DFE-based transceivers appear in [3], [4], [6].) Since the expression in (30) is also Schur-concave, a design that maximizes the Gaussian mutual information can also be obtained using the Schur-concave approach in Section VI, below. That approach results in a linear transceiver with a standard water-filling power allocation [25]. (Of course, both approaches yield the same maximized Gaussian mutual information.)

3) Minimization of maximum MSE (Maximization of minimum SINR): Minimization of the maximum MSE corresponds to minimization of the following Schur-convex function of $\boldsymbol{l}$

$$g(e^{\boldsymbol{l}}) = \max_{1\le i\le K} (e^{\boldsymbol{l}_i}). \tag{31}$$

According to (17), the stream with the maximum MSE is the one with the minimum SINR. Hence, this objective is equivalent to maximization of the minimum SINR.

4) Minimization of p-norm of MSEs: In this case, the objective is to minimize

$$g(e^{\boldsymbol{l}}) = \Big(\sum_{i=1}^{K} (e^{\boldsymbol{l}_i})^p \Big)^{1/p}, \quad p \ge 1. \tag{32}$$

This design criteria includes the minimization of total MSE, $p = 1$, and the minimization of the maximum MSE, $p = \infty$, among several other norms of the vector of MSEs of each data stream.

5) Maximization of the harmonic mean of SINRs: In this case, the objective is to minimize

$$g(e^{\boldsymbol{l}}) = \sum_{i=1}^{K} \frac{1}{\mathrm{SINR}_i} = \sum_{i=1}^{K} \frac{1}{e^{-\boldsymbol{l}_i} - 1}, \quad \boldsymbol{l}_i < 0. \tag{33}$$

6) Maximization of product of SINRs: Maximization of the product of the SINRs (or the geometric mean of the SINRs) can be expressed as the minimization of

$$g(e^{\boldsymbol{l}}) = -\log \prod_{i=1}^{K} \big(e^{-\boldsymbol{l}_i} - 1\big) = -\sum_{i=1}^{K} \log\big(e^{-\boldsymbol{l}_i} - 1\big). \tag{34}$$

7) Minimization of average BER: Assuming that each each data stream employs the same constellation, the average BER is given by

$$g(e^{\boldsymbol{l}}) = \frac{1}{K} \sum_{i=1}^{K} \mathrm{BER}(\mathrm{SINR}_i) = \frac{1}{K} \sum_{i=1}^{K} \mathrm{BER}(e^{-\boldsymbol{l}_i} - 1), \tag{35}$$

where $\mathrm{BER}(\cdot)$ is the bit error rate of the chosen constellation as function of the SINR. For many constellations, such as M-ary QAM, the bit error rate function $\mathrm{BER}(\mathrm{SINR})$ can be closely approximated by [28, eq. 18], [29, eq. 13]:

$$\mathrm{BER}(\mathrm{SINR}) = c_2\, Q(\sqrt{c_1\ \mathrm{SINR}}), \tag{36}$$

where c_1 and c_2 are constants that depend on the size of constellation M, and $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-z^2/2}\, dz$. For BPSK and QPSK, we have $c_1 = c_2 = 1$ and the approximation becomes exact. In Appendix A we show that the objective in (36) is a Schur-convex function of $\boldsymbol{l}$ for BPSK and M-ary QAM up to $M = 16$, and that for higher-order QAM it is Schur-convex under the mild constraint that the SINR is above a small threshold. (The design of DFE-based systems with an average BER objective was considered in [6].)

VI. Optimal Precoding Matrix: Schur-concave objectives

A. Optimal Precoding Matrix

If $g(e^{\boldsymbol{l}})$ is a Schur-concave function of $\boldsymbol{l}$, then from (22) we have $g(e^{\mathbf{m}}) \le g(e^{\boldsymbol{l}})$, and the optimal value is obtained when

$$\mathbf{L}_{ii} = \sigma_i(\mathbf{L}). \tag{37}$$

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 [17]. The optimal $\mathbf{C}$ in that case is $\mathbf{I}$, and hence the optimal feedback matrix is $\mathbf{B} = \mathbf{0}$. 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.

This result shows that for Schur-concave objectives the design problem reduces to that for the special subclass of linear transceivers; e.g., [1], [2]. What remains is to compare the direct linear designs with those that we have derived from the optimization of DFE and TH transceivers with Schur-concave objectives of the logarithm of the individual MSEs, $g(e^{\boldsymbol{l}})$. Using the composition properties of Schur-concave functions [24] the following counterpart to Lemma 2 can be established.

Lemma 3: Let $\mathbf{y} = e^{\boldsymbol{l}}$. If $g(e^{\boldsymbol{l}})$ is Schur-concave in $\boldsymbol{l}$, then $g(\mathbf{y})$ is Schur-concave in $\mathbf{y}$.

A consequence of this result is that the optimal DFE or TH transceiver design for an objective that is Schur-concave in the logarithm of the individual MSEs is the optimal linear transceiver for the corresponding Schur-concave function of the individual MSEs themselves. As shown in [2], that optimal precoder will depend on the objective. This is in contrast to the Schur-convex designs, which are independent of the objective; see Section V.

B. Examples of Schur-concave objectives

We now briefly present some examples of design objectives that are Schur-concave functions of $\boldsymbol{l}$. (Sketches of the proofs are provided in Appendix B.)

1) Minimization of harmonic mean of MSEs: This objective corresponds to the minimization of

$$g(e^{\boldsymbol{l}}) = \frac{1}{\sum_{i=1}^{K} e^{-\boldsymbol{l}_i}}. \tag{38}$$

2) Maximization of p-norm of SINRs: In this case, the objective is to minimize

$$g(e^{\boldsymbol{l}}) = -\Big(\sum_{i=1}^{K}(e^{-\boldsymbol{l}_i}-1)^p\Big)^{1/p}, \quad p \geq 1. \tag{39}$$

3) Minimization of a subclass of weighted products of MSEs (weighted geometric mean of MSEs): The minimization of the weighted product of MSEs is equivalent to minimization of

$$g(e^{\boldsymbol{l}}) = \log \prod_{i=1}^{K} \left(e^{\boldsymbol{l}_i}\right)^{a_i} = \sum_{i=1}^{K} a_i \boldsymbol{l}_i, \tag{40}$$

where, without loss of generality, we may assume that the MSEs are arranged in a decreasing order; i.e. $\boldsymbol{l}_1 \geq \cdots, \geq \boldsymbol{l}_K$. For this ordering, $g(e^{\boldsymbol{l}})$ is Schur-concave whenever the weights are in ascending order.

VII. Simulation Studies

In this section, we provide some simulation results for systems designed using the proposed framework. We consider systems that transmit vectors of 16-QAM symbols over an independent Rayleigh fading channel (with perfect channel state information at both the receiver and transmitter). The coefficients of the $N_r \times N_t$ channel matrix $\mathbf{H}$ are modelled as being independent rotationally-symmetric complex Gaussian random variables with zero mean and unit variance, and the elements of the additive noise vector $\mathbf{n}$ are modelled as being independent rotationally-symmetric complex Gaussian random variables with zero mean and equal variance. For each design we will plot the average bit error rate (BER) of the K data streams against the signal-to-noise ratio (SNR), which is defined as the ratio of the total average transmitted power, $\mathrm{E}\{\mathbf{x}^H\mathbf{x}\}$, to the total receiver noise power, $\mathrm{E}\{\mathbf{n}^H\mathbf{n}\}$.

A. Validation of the design assumptions

In this section, we validate the assumptions that we made in the development of the proposed designs. For DFE systems we made the standard assumption that the previously detected symbols were correctly detected, and for TH precoding systems we made the assumption of no precoding loss; see Section II. To validate these assumptions, we consider the case of systems optimized for Schur-convex objectives. These designs minimize the total MSE, as well as minimizing the average BER and maximizing the Gaussian mutual information. In Fig. 3 we compare the actual performance of the proposed designs to the performance that would have been achieved if the assumptions held precisely, in the case of a system with $N_t = N_r = K = 4$. In Fig. 3 the practical performance of the proposed jointly optimal TH transceiver is very close to that of a system that assumes no precoding loss, and the impact of the standard assumption of correct decisions in a DFE system is quite mild, especially at high SNRs. Indeed, the four curves coalesce at high SNRs. The slight advantage of the TH transceiver in Fig. 3 over the DFE transceiver can be attributed to the fact that interference subtraction at the transmitter is, inherently, free from error propagation.

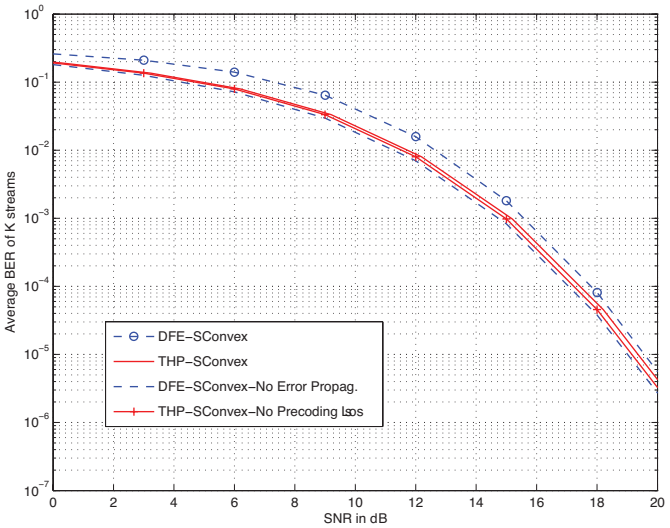

Fig. 3. BERs of the optimal Schur-convex design of a DFE transceiver (DFE-SConvex), and a TH transceiver (THP-SConvex) for a system with $N_t = N_r = K = 4$. Also plotted is the BER of the optimal Schur-Convex DFE design in the absence of error propagation (DFE-SConvex-No Error Propag.), and the BER of the optimal Schur-Convex TH design with no precoding loss (THP-SConvex-No Precoding Loss).

B. Comparisons with linear transceivers

In this section, we compare the performance of the proposed (jointly optimal) DFE and TH transceiver designs to that of (jointly-optimized) linear transceivers. We compare the performance of the optimal Schur-convex design for the DFE and TH transceivers, which simultaneously minimizes the total MSE, minimizes the average BER and maximizes the Gaussian mutual information, with that of the (different) optimal linear transceivers that: minimize the total MSE, e.g., [1]; minimize the average BER [2], [30]; and maximize the Gaussian mutual information, e.g., [2], [25]. We compare the performance of these five methods in an $N_t = N_r = K = 4$ scenario in Fig. 4. By comparing the curves for the DFE and TH transceivers with that of the minimum BER linear transceiver, one can quantify the statement in Proposition 1 that for Schur-convex design objectives, the DFE and TH transceivers provide provably better performance than the corresponding linear transceiver.

C. Comparisons with other designs for interference (pre-)subtraction transceivers

In this section, we compare the performance of the proposed jointly optimal DFE and TH transceiver designs to that of some existing suboptimal designs for systems that employ MMSE interference (pre-)subtraction. In particular, we will provide comparisons to systems with an identity precoder

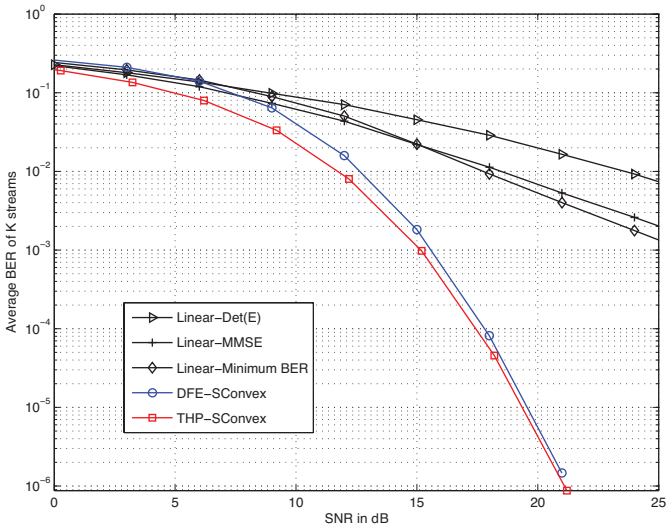

Fig. 4. BERs of the optimal Schur-convex designs of DFE (DFE-SConvex) and TH (THP-SConvex) and the optimal linear transceivers: minimum MSE (Linear-MMSE) e.g., [1], minimum average BER (Linear-Minimum BER) [2], [30], and maximum mutual information (Linear-Det(E)) e.g., [2], [25], for a system with $N_t = N_r = K = 4$.

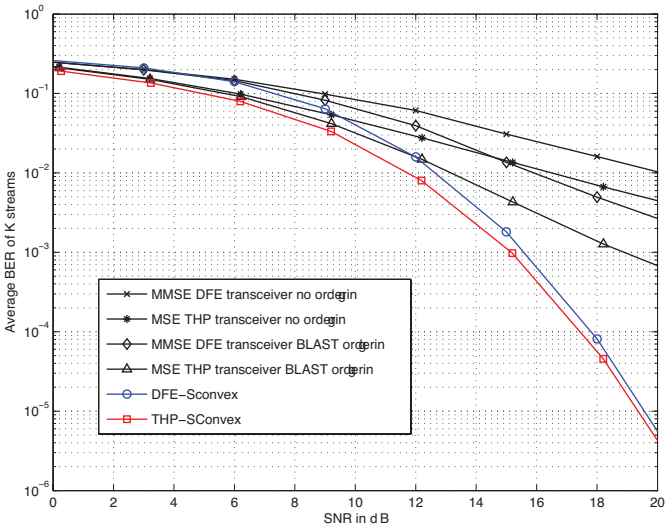

Fig. 5. BERs of the optimal Schur-convex designs for DFE (DFE-SConvex) and TH (THP-SConvex) transceivers and other interference (pre-)subtraction approaches: MMSE DFE with BLAST ordering [9], [32], and MMSE DFE with no ordering, TH transceiver MMSE design in [9] with BLAST ordering and with no ordering, for a system with $N_t = N_r = K = 4$.

at the transmitter and an MMSE-DFE receiver with the 'BLAST' [31] detection ordering [9], [32], or an unordered MMSE-DFE receiver. We will also provide comparisons with the performance of the MMSE-TH transceiver design in [9], with both BLAST ordering and the natural ordering. We compare the performance of these six methods in an $N_t = N_r = K = 4$ scenario in Fig. 5, and in an $N_t = K = 4$, $N_r = 5$ scenario in Fig. 6. These comparisons are appropriate because the MMSE-DFE approach in [9], [32] and the MMSE-TH design in [9] can be represented by special cases of our system model in which the precoder $\mathbf{P}$ is restricted to be a permutation matrix. The significantly lower BERs of the proposed designs demonstrate that the exploitation of all the available degrees of design freedom in the proposed approach can have a substantial impact on performance. Moreover, the permutation-based approaches in [9], [32] result in data streams with different MSEs (and SINRs), and hence different ordering algorithms are required for different performance objectives. For example, for error performance criteria the BLAST ordering [31] is appropriate, as it attempts to maximize the SINR of the weakest data stream, but maximizing the Gaussian mutual information requires a different ordering [33]. In contrast to these permutation-based approaches, the proposed approach exploits all the degrees of design freedom in the system and results in data streams with equal SINRs, and hence no ordering algorithm is necessary. It is worth pointing out that while precoding generalizes ordering for point-to-point DFE or TH models, in the corresponding multiuser models ordering must be considered in conjunction with precoder design because on the uplink the transmitters cannot cooperate, and on the downlink the receivers cannot cooperate; cf. [13].

VIII. Conclusion

We have developed a unified framework for joint transceiver design for interference (pre-)subtraction schemes for communication over generic point-to-point MIMO channels, and we have obtained optimal designs for two broad classes of communication objectives, namely those that are Schur-convex and Schur-concave functions of the logarithms of the (individual) MSEs of each data stream. For Schur-convex objectives, the optimal transceiver results in equal individual MSEs, and simultaneously minimizes the total MSE, minimizes the average bit error rate, and maximizes the Gaussian mutual information. Furthermore, that design yields objective values that are superior to the corresponding optimal objective value for a linear transceiver. For the class Schur-concave objectives, the optimal DFE design results in linear equalization and the optimal TH precoding design results in linear precoding.

Appendix A
Proofs of Schur-convex objectives

a) Minimization of total MSE: The objective here is to minimize $g(e^{\boldsymbol{l}}) = \sum_{i=1}^{K} e^{\boldsymbol{l}_i}$, which has the form of $g(e^{\boldsymbol{l}}) = \sum_{i=1}^{K} f(\boldsymbol{l}_i)$ for the strictly convex function $f(\mathbf{x}_i) = e^{\mathbf{x}_i}$. Hence, $g(e^{\boldsymbol{l}})$ is a strictly Schur-convex function of $\boldsymbol{l}$, [24, p. 64].

b) Minimization of product of MSEs: This objective can be written as: minimize $g(e^{\boldsymbol{l}}) = \sum_{i=1}^{K} \boldsymbol{l}_i$. Since this is the sum of each $\boldsymbol{l}_i$, it is both a Schur-convex and a Schur-concave function of $\boldsymbol{l}$, [24].

c) Minimization of p-norm of MSEs: In this case, the objective is to minimize $g(e^{\boldsymbol{l}}) = \left(\sum_{i=1}^{K} (e^{\boldsymbol{l}_i})^p\right)^{1/p}$, $p \geq 1$, which has the form $g(e^{\boldsymbol{l}}) = h(f(\boldsymbol{l}_1), \ldots, f(\boldsymbol{l}_K))$, where $h(\mathbf{x}_1, \ldots, \mathbf{x}_K) = \left(\sum_{i=1}^{K} |\mathbf{x}_i|^p\right)^{1/p}$ is Schur-convex and is an increasing function of each argument, and $f(x) = e^x$ is a convex function. If follows from the composition properties of Schur-convex functions [24] that $g(e^{\boldsymbol{l}})$ is a Schur-convex function. Although minimization of the total MSE is a special case of the p-norm minimization for $p = 1$, the proof used for the total MSE case provides the stronger result of strict Schur-convexity.

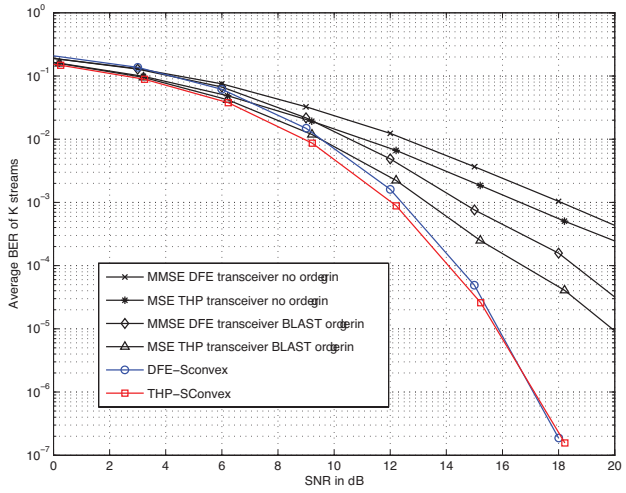

Fig. 6. BERs of the optimal Schur-convex designs for DFE (DFE-SConvex) and TH (THP-SConvex) transceivers and other interference (pre-)subtraction approaches: MMSE DFE with BLAST ordering [9], [32], and MMSE DFE with no ordering, TH transceiver MMSE design in [9] with BLAST ordering and with no ordering, for a system with $N_t = K = 4$, $N_r = 5$.

d) Maximization of product of SINRs: This objective can be written as: minimize $g(e^{\boldsymbol{l}}) = -\sum_{i=1}^{K} \log\left(e^{-\boldsymbol{l}_i} - 1\right)$. Since $-g(e^{\boldsymbol{l}})$ is the sum of the concave function $f(x) = \log(e^{-x} - 1)$ applied to each $\boldsymbol{l}_i$, $-g(e^{\boldsymbol{l}})$ is a Schur-concave function of $\boldsymbol{l}$ [24, p. 64], and it follows that $g(e^{\boldsymbol{l}})$ is Schur-convex.

e) Maximization of harmonic mean of SINRs: In this case the objective is to minimize $g(e^{\boldsymbol{l}}) = \sum_{i=1}^{K} \frac{1}{\text{SINR}_i} = \sum_{i=1}^{K} \frac{1}{e^{-\boldsymbol{l}_i} - 1}$, $\boldsymbol{l}_i < 0$. Since each MSE satisfies $0 \leq e^{\boldsymbol{l}_i} < 1$, we will restrict our proof to the case of $\boldsymbol{l}_i < 0$. We observe that $g(e^{\boldsymbol{l}})$ is a sum of the strictly convex function $f(x) = 1/(e^{-x} - 1)$, for $x < 0$, applied to each $\boldsymbol{l}_i$. Hence, $g(e^{\boldsymbol{l}})$ is a strictly Schur-convex function.

f) Minimization of average BER: Assuming that each each data stream employs the same constellation, the average BER is $g(e^{\boldsymbol{l}}) = \frac{1}{K}\sum_{i=1}^{K} \text{BER}(\text{SINR}_i)$, where $\text{BER}(\cdot)$ is the bit error rate of the chosen constellation as a function of the SINR, and $\text{SINR}_i = e^{-\boldsymbol{l}_i} - 1$. As pointed out in Section V-C, for many constellations the bit error rate function $\text{BER}(\text{SINR})$ can be closely approximated by

$$\text{BER}(\text{SINR}) = c_2\, Q(\sqrt{c_1\, \text{SINR}}), \tag{41}$$

where c_1 and c_2 are constants that depend on the constellation. If each $\text{BER}(e^{-\boldsymbol{l}_i} - 1)$ is a (strictly) convex function of $\boldsymbol{l}_i$, it follows that their sum $g(e^{\boldsymbol{l}})$ is (strictly) Schur-convex. To show the convexity of $\text{BER}(e^{-\boldsymbol{l}_i} - 1)$, we obtain the second derivative of (41) with respect to $\boldsymbol{l}_i$:

$$\frac{d^2\, \text{BER}}{d\, \boldsymbol{l}_i^2} = \frac{c_2\, c_1^{1/2}\, e^{-\frac{c_1}{2}(y^{-1}-1)}}{4\sqrt{2\pi}\, y^{3/2}\, (1-y)^{3/2}} \left(2y^2 - (c_1+1)y + c_1\right), \tag{42}$$

where $y = e^{\boldsymbol{l}_i}$. Since the first term is non-negative for all values of the MSE, the sign of the second derivative is determined by the quadratic term $\left(2y^2 - (c_1+1)y + c_1\right)$. To check the sign of this term, we have to consider two cases:

- For values of the constellation constant c_1 such that the discriminant of the quadratic equation is negative, the second derivative is non-negative for all the range of the MSE. Hence, the expression for BER in (41) is convex function of $\boldsymbol{l}_i$. This case includes BPSK and M-ary QAM with $M \leq 16$.
- For values of the constellation constant c_1 such that discriminant of the quadratic equation is non-negative, the second derivative is non-negative for the range of MSE $y \leq y_r$, where $y_r = \left(c_1 + 1 - \sqrt{c_1^2 - 6c_1 + 1}\right)/4$ is a root of the quadratic equation. In this case, which applies to M-ary constellations with $M \geq 16$, the BER expression in (41) will be convex for all SINRs above the small threshold $1/y_r - 1$.

APPENDIX B
PROOFS OF SCHUR-CONCAVE OBJECTIVES

a) Minimization of harmonic mean of MSEs: This corresponds to the minimization of $g(e^{\boldsymbol{l}}) = \frac{1}{\sum_{i=1}^{K} e^{-\boldsymbol{l}_i}}$, where the denominator is the sum of a convex function $f(x) = e^{-x}$ applied to each $\boldsymbol{l}_i$. Hence, the denominator is a Schur-convex function [24, p. 64]. Since $g(e^{\boldsymbol{l}})$ is a decreasing function of a Schur-convex function, it follows that $g(e^{\boldsymbol{l}})$ is Schur-concave [24, p. 61].

b) Maximization of p-norm of SINRs: In this case, the objective is to minimize: $g(e^{\boldsymbol{l}}) = -\left(\sum_{i=1}^{K} (e^{-\boldsymbol{l}_i} - 1)^p\right)^{1/p}$, $p \geq 1$. We observe that $-g(e^{\boldsymbol{l}})$ has the form $g(e^{\boldsymbol{l}}) = h(f(\boldsymbol{l}_1), \ldots, f(\boldsymbol{l}_K))$, where $h(\mathbf{x}_1, \ldots, \mathbf{x}_K) = \left(\sum_{i=1}^{K} |\mathbf{x}_i|^p\right)^{1/p}$ is Schur-convex and is an increasing function of each argument [24], and that $f(x) = e^{-x} - 1$ is a convex function. It follows from composition rules of Schur-convex functions [24, p. 63] that $-g(e^{\boldsymbol{l}})$ is a Schur-convex function. Hence, $g(e^{\boldsymbol{l}})$ is Schur-concave.

c) Minimization of a subclass of weighted product of MSEs: Minimization of the weighted product of the individual MSEs (or, equivalently, the weighted geometric mean of the MSEs) corresponds to minimization of the objective $g(e^{\boldsymbol{l}}) = \log \prod_{i=1}^{K} \left(e^{\boldsymbol{l}_i}\right)^{a_i} = \sum_{i=1}^{K} a_i \boldsymbol{l}_i$. Assuming that $\boldsymbol{l}_i$ are in decreasing order, then $g(e^{\boldsymbol{l}})$ is a Schur-concave function when the weights a_i are in ascending order [2], [24]. A special case of this objective is the unweighted product, for which all $a_i = 1$. That function is both Schur-concave and Schur-convex; see Appendix A.

REFERENCES

[1] 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, no. 7, pp. 1988–2006, July 1999.

[2] 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, no. 9, pp. 2381–2401, Sept. 2003.

[3] J. M. Cioffi and G. D. Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing*, A. Paulraj, V. Roychowdhury, and C. Schaper, Eds. Kluwer, 1997, ch. 4, pp. 79–127.

[4] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1334–1347, Sept. 1994.

[5] 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, no. 2, pp. 373–378, Mar. 2004.

[6] 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, no. 3, pp. 965–978, Mar. 2006.
[7] 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, no. 1, pp. 154–172, Jan. 2005.
[8] Y. Jiang, J. Li, and W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.
[9] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.
[10] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, Jul. 2004.
[11] Y. Jiang, D. P. Palomar, and M. K. Varanasi, "Precoder optimization for nonlinear MIMO transceiver based on arbitrary cost function," in *Proc. Conf. Inform. Sci. Syst.*, Baltimore, Mar. 2007.
[12] Y. Jiang and D. P. Palomar, "MIMO transceiver design via majorization theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4–5, pp. 331–551, 2006.
[13] M. Schubert and H. Boche, "User ordering and power allocation for optimal multiantenna precoding/decoding," in *Proc. ITG Wkshp Smart Antennas*, Munich, Mar. 2004, pp. 174–181.
[14] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.
[15] M. Botros Shenouda and T. N. Davidson, "Limited feedback design of MIMO systems with zero-forcing DFE using Grassmann codebooks," in *Proc. IEEE Canadian Wkshp Inform. Theory*, Edmonton, June 2007, pp. 118–123.
[16] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding—Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.
[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1985.
[18] H. Weyl, "Inequalities between the two kinds of eigenvalues of a linear transformation," *Proc. Nat. Acad. Sci.*, vol. 35, pp. 408–411, July 1949.
[19] T. Guess and M. K. Varanasi, "Multiuser decision-feedback receivers for the general Gaussian multiple-access channel," in *Proc. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Oct. 1996.
[20] ——, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 173–187, 2005.
[21] T. Guess, "Optimal sequences for CDMA with decision-feedback receivers," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 886–900, April 2003.
[22] L. Li, Y. Yao, and H. Li, "Transmit diversity and linear and decision-feedback equalizations for frequency-selective fading channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 5, pp. 1217–1231, Sept. 2003.
[23] M. Schubert and S. Shi, "MMSE transmit optimization with interference pre-compensation," in *Proc. Veh. Tech. Conf.*, vol. 2, Stockholm, May 2005, pp. 845–849.
[24] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic Press, 1979.
[25] H. S. Witsenhausen, "A determinant maximization problem occurring in the theory of data communication," *SIAM J. Appl. Math.*, vol. 29, pp. 515–522, 1975.
[26] Y. Jiang, J. Li, and W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.
[27] J. Zhang, T. N. Davidson, and K. M. Wong, "Uniform decomposition of mutual information using MMSE decision feedback detection," in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Sept. 2005, pp. 714–718.
[28] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, July 2002.
[29] L. Yang and L. Hanzo, "A recursive algorithm for the error probability evaluation of M-QAM," *IEEE Commun. Lett.*, vol. 4, no. 10, pp. 304–306, Oct. 2000.
[30] S. S. Chan, T. N. Davidson, and K. M. Wong, "Asymptotically minimum BER linear block precoders for MMSE equalisation," *IEE Proc. Commun.*, vol. 151, no. 4, pp. 297–304, Aug. 2004.
[31] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14–16, 7 Jan. 1999.
[32] G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Commun. Letters*, vol. 5, no. 9, pp. 364–366, 2001.
[33] C. Windpassinger, T. Vencel, and R. F. H. Fischer, "Precoding and loading for BLAST-like systems," in *Proc. IEEE Int. Conf. Commun.*, vol. 5, Anchorage, May 2003, pp. 3061–3065.

Michael Botros Shenouda received the B.Sc. (Hons. 1) degree in 2001 and the M.Sc. degree in 2003, both in electrical engineering and both from Cairo University, Egypt. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, McMaster University, Canada. His main areas of interest include wireless and MIMO communication, convex and robust optimization, and signal processing algorithms. He is also interested in majorization theory, and its use in the development of design frameworks for non-linear MIMO transceivers. Mr. Botros Shenouda was awarded an IEEE Student Paper Award at ICASSP 2006, and was a finalist in the IEEE Student Paper Award competition at ICASSP 2007.

Tim Davidson (M'96) received the B.Eng. (Hons. I) degree in Electronic Engineering from the University of Western Australia (UWA), Perth, in 1991 and the D.Phil. degree in Engineering Science from the University of Oxford, U.K., in 1995.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Ontario, Canada, where he holds the (Tier II) Canada Research Chair in Communication Systems, and is currently serving as Acting Director of the School of Computational Engineering and Science. His research interests lie in the general areas of communications, signal processing and control. He has held research positions at the Communications Research Laboratory at McMaster University, the Adaptive Signal Processing Laboratory at UWA, and the Australian Telecommunications Research Institute at Curtin University of Technology, Perth, Western Australia.

Dr. Davidson was awarded the 1991 J. A. Wood Memorial Prize [for "the most outstanding (UWA) graduand" in the pure and applied sciences] and the 1991 Rhodes Scholarship for Western Australia. He is currently serving as an Associate Editor of the IEEE Transactions on Signal Processing and the IEEE Transactions on Circuits and Systems II, and he recently served as a Guest Co-editor of issues of the IEEE Journal on Selected Areas in Communications and the IEEE Journal on Selected Topics in Signal Processing.