A Design Framework for Limited Feedback MIMO Systems with Zero-Forcing DFE

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

Abstract—We consider the design of multiple-input multiple-output communication systems with a linear precoder at the transmitter, zero-forcing decision feedback equalization (ZF-DFE) at the receiver, and a low-rate feedback channel that enables communication from the receiver to the transmitter. The channel state information (CSI) available at the receiver is assumed to be perfect, and based on this information the receiver selects a suitable precoder from a codebook and feeds back the index of this precoder to the transmitter. Our approach to the design of the components of this limited feedback scheme is based on the development, herein, of a unified framework for the joint design of the precoder and the ZF-DFE under the assumption that perfect CSI is available at both the transmitter and the receiver. The framework is general and embraces a wide range of design criteria. This framework enables us to characterize the statistical distribution of the optimal precoder in a standard Rayleigh fading environment. Using this distribution, we show that codebooks constructed from Grassmann packings minimize an upper bound on an average distortion measure, and hence are natural candidates for the codebook in limited feedback systems. Our simulation studies show that the proposed limited feedback scheme can provide significantly better performance at a lower feedback rate than existing schemes in which the detection order is fed back to the transmitter.

Index Terms—Limited feedback, Decision feedback equalization (DFE), Grassmann packings, Majorization, Schur-convexity.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) communication schemes offer the potential for significant increases in spectral efficiency over their single-input single-output counterparts by enabling simultaneous transmission of independent data streams. MIMO schemes also offer the potential for significant performance gains in a variety of other metrics. Standard transceiver architectures for these schemes include linear precoding and equalization, and the combination of linear precoding and decision feedback equalization (DFE), which offers the potential for improved performance over the linear approach while maintaining comparable complexity. For scenarios in which accurate channel state information (CSI) is available at both the transmitter and the receiver, there is a well established framework that unifies the design of linear transceivers under many design criteria [1]. A counterpart for the design of systems with DFE has recently emerged

[2]–[5]. This framework was also extended to MIMO systems with pre-interference subtraction at the transmitter in [2]. However, in many scenarios, such as frequency division duplex systems, obtaining accurate CSI at the transmitter may require a considerable amount of feedback to the transmitter. An approach that allows the designer to limit the required amount of the feedback is to quantize the transmitter design. In these limited feedback schemes [6], the receiver uses its CSI to choose the best transmitter design from a codebook of available designs, and then feeds back the index of this precoder to the transmitter. This strategy has been considered for beamforming schemes (e.g., [7]–[13]), unitary precoding with linear equalization (e.g., [14]). and unitary precoding for orthogonal space time block codes [15], [16]. For zero-forcing DFE schemes, a limited feedback scheme in which the receiver feeds back the order of interference cancellation was proposed in [17], [18].

In this work, we consider the design of a limited feedback scheme for systems with a (general) linear precoder at the transmitter and zero-forcing DFE at the receiver. Our designs are based on a unified framework, developed herein, for the joint design of the precoder and the ZF-DFE in the presence of perfect CSI. This framework embraces a wide range of design criteria that can be expressed as functions of the mean square error (MSE) of each data stream, including minimization of the total MSE, minimization of the average bit error rate (BER), and maximization of the Gaussian mutual information. In particular, we show that the optimal precoder for systems with a zero-forcing DFE is the same for all these criteria; a property that cannot be achieved by a linear transceiver. Furthermore, we show that the optimal precoder for these objectives is a scaled unitary matrix that is isotropically distributed (over the Stiefel manifold of unitary matrices). Using this distribution, we show that codebooks constructed from Grassmann subspace packings minimize an upper bound on an average distortion measure, and hence are excellent candidates for the codebook in limited feedback schemes for systems with zero-forcing DFE. In contrast, the application of Grassmann codebooks in limited feedback schemes with linear receivers (e.g., [14]) involves an inherent compromise, because the optimal precoder in the presence of perfect CSI and a total power constraint is not unitary. Since the scheme that we propose involves the construction of codebooks for isotropically distributed unitary matrices, our scheme subsumes that in [17], [18], in which the precoder is, by construction, a permutation matrix. Our simulation studies suggest that the additional degrees of freedom available in our approach enable our scheme to provide significantly better performance than that in [17], [18] while using a lower feedback rate.

Our notation is as follows: Boldface type is used to denote matrices and vectors; $\mathbf{a}_i$ denotes the ith element of the vector $\mathbf{a}$, $\mathbf{A}_{ij}$ denotes the element at the intersection of the ith row and j^{th} column of the matrix $\mathbf{A}$, $\mathbf{A}^H$ denotes the conjugate transpose of $\mathbf{A}$, and $(\mathbf{A})^{\dagger}$ denotes the (minimum norm) pseudo-inverse of $\mathbf{A}$. The terms $\det(\mathbf{A})$, and $\|\mathbf{A}\|_2$ denote the determinant and the two-norm (maximum singular value) of $\mathbf{A}$, respectively. The notation $\text{Diag}(\mathbf{x})$ denotes the diagonal matrix whose elements are the elements of $\mathbf{x}$.

II. SYSTEM MODEL

We consider a point-to-point communication system with N_t transmit antennas and N_r receive antennas that transmits K data streams simultaneously, where K is no greater than the rank of the channel matrix $\mathbf{H}$. We adopt a narrow band block fading channel model, and we consider MIMO communication systems that use (generalized) zero-forcing decision feedback equalization, e.g., [19], [20], for spatial equalization. At the transmitter, the input data vector $\mathbf{s} \in \mathbb{C}^K$ is linearly precoded using $\mathbf{P}$ to generate the transmitted data vector $\mathbf{x} \in \mathbb{C}^{N_t}$,

$$\mathbf{x} = \mathbf{Ps}. \qquad (1)$$

Without loss of generality, we will assume that $\text{E}\{\mathbf{ss}^H\} = \mathbf{I}$, and hence the total transmitter power constraint can be written as $\text{E}\{\mathbf{x}^H\mathbf{x}\} = \text{tr}(\mathbf{P}^H\mathbf{P}) \leq P_{\text{total}}$.

The vector of received signals is given by

$$\mathbf{y} = \mathbf{HPs} + \mathbf{n}, \qquad (2)$$

where $\mathbf{H}$ is the channel matrix and $\mathbf{n}$ is the vector of additive noise which is assumed to have zero-mean and a covariance matrix $\text{E}\{\mathbf{nn}^H\} = \sigma_n^2\mathbf{I}$. As illustrated in Fig. 1, following linear processing using the feedforward matrix $\mathbf{G}$, the receiver makes successive decisions on each symbol by subtracting the effect of previously decided symbols. Hence, the feedback matrix $\mathbf{B}$ is strictly lower triangular. This system model embraces linear precoding and equalization as a special case when $\mathbf{B} = \mathbf{0}$. Assuming correct previous decisions, the vector of inputs to the quantizer is given by

$$\hat{\mathbf{s}} = (\mathbf{GHP} - \mathbf{B})\mathbf{s} + \mathbf{Gn}. \qquad (3)$$

By defining the error signal $\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}}$, the error covariance matrix (the "MSE" matrix) can be written as

$$\mathbf{E} = \text{E}\{\mathbf{ee}^H\} = \mathbf{CC}^H - \mathbf{CP}^H\mathbf{H}^H\mathbf{G}^H - \mathbf{GHPC}^H$$
$$+ \mathbf{GHPP}^H\mathbf{H}^H\mathbf{G}^H + \sigma_n^2\mathbf{GG}^H, \quad (4)$$

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

We will consider communication schemes in which perfect CSI is available only at the receiver. Based on its channel knowledge, the receiver selects a suitable precoding matrix from a codebook of precoders $\mathcal{P}$ of size $|\mathcal{P}|$, and feeds that index back to the transmitter using $\log_2 |\mathcal{P}|$ information bits; see Fig. 1. In order to develop effective methods for quantizing the precoding matrix, we first need to characterize the optimal precoding matrix for different design criteria in the presence of perfect CSI. We will then use the statistical distribution of this optimal precoder to define the distortion measures that are required to design the codebook for the limited feedback scheme.
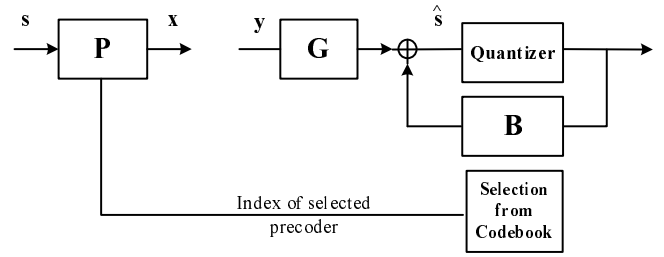

Fig. 1. MIMO transceiver with DFE using limited feedback.

III. UNIFIED FRAMEWORK FOR ZERO-FORCING DFE

In this section, we develop a general framework for the joint design of the transceiver matrices $\mathbf{G}$, $\mathbf{C} = \mathbf{I} + \mathbf{B}$, and $\mathbf{P}$ in the presence of perfect CSI. We consider system design criteria that are expressed as functions of the (logarithm of the) MSE of the individual data streams $\mathbf{E}_{ii}$. The proposed framework embraces a wide range of design objectives. It includes objectives for which optimal designs are already available (e.g., the total MSE, [20]), and several other objectives for which the optimal transceiver design has remained an open problem. The framework can be regarded as a counterpart for the existing framework of linear transceiver design [1]. Here, the framework is derived for DFEs with a zero-forcing constraint, but an analogous framework can be developed in the absence of this constraint [2]–[5].

A. ZF-DFE Receiver Design

The zero-forcing design criterion implies

$$\mathbf{GHP} - \mathbf{B} = \mathbf{I}. \qquad (5)$$

Given the assumption that $K \leq \text{rank}(\mathbf{H})$, the condition in (5) can be achieved so long as $\mathbf{P}$ is chosen such that $\text{rank}(\mathbf{HP}) = K$. In that case, the feedforward matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \mathbf{C}(\mathbf{HP})^{\dagger}. \qquad (6)$$

Since $\mathbf{HP}$ has full column rank, the pseudo-inverse in (6) can be written as

$$(\mathbf{HP})^{\dagger} = (\mathbf{P}^H\mathbf{H}^H\mathbf{HP})^{-1}\mathbf{P}^H\mathbf{H}^H. \qquad (7)$$

Using the expression for $\mathbf{G}$ in (6), the MSE matrix in (4) reduces to

$$\mathbf{E} = \mathbf{CNC}^H, \qquad (8)$$

where $\mathbf{N} = \sigma_n^2(\mathbf{P}^H\mathbf{H}^H\mathbf{HP})^{-1}$ is a positive definite Hermitian matrix. The optimal matrix $\mathbf{C}$, that minimizes the MSE of each individual data stream, subject to being unit diagonal and lower triangular, is given by [2]

$$\mathbf{C} = \text{Diag}(\mathbf{L}_{11}, \dots, \mathbf{L}_{KK})\mathbf{L}^{-1}, \qquad (9)$$

where $\mathbf{N} = \mathbf{LL}^H$ is the Cholesky factorization of $\mathbf{N}$, and $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries. Using this optimal $\mathbf{C}$, the MSE matrix can be rewritten as

$$\mathbf{E} = \text{Diag}(\mathbf{L}_{11}^2, \dots, \mathbf{L}_{KK}^2), \qquad (10)$$

where $\mathbf{L}_{ii}$ is the ith diagonal element of $\mathbf{L}$. Hence, the SNR of each data stream is

$$\text{SNR}_k = \frac{1}{\mathbf{E}_{kk}} = \frac{1}{\mathbf{L}_{kk}^2}. \qquad (11)$$

B. Transmitter Design

Given the optimal $\mathbf{G}$ and $\mathbf{C}$, our next step is to design a precoding matrix $\mathbf{P}$ so as to optimize design criteria that are expressed as functions of the (logarithm of the) MSE of each individual stream. To derive the optimal precoding matrix, we will first obtain some inequalities that involve the logarithm of the MSE of the individual data streams,

$$l = (\ln \mathbf{L}_{11}^2, \ldots, \ln \mathbf{L}_{KK}^2), \qquad (12)$$

using concepts from majorization theory.

Definition 1 (Additive Majorization [21]): Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^K$ and let $a_{[1]}, \ldots, a_{[K]}$ denote the re-ordering of the elements of $\mathbf{a}$ in a non-increasing order; i.e., $a_{[1]} \geq \ldots \geq a_{[K]}$. The vector $\mathbf{b}$ is said to majorize $\mathbf{a}$, $\mathbf{a} \prec \mathbf{b}$, if

$$\sum_{i=1}^{j} \mathbf{a}_{[i]} \leq \sum_{i=1}^{j} \mathbf{b}_{[i]} \qquad \text{for } j = 1, \ldots, K-1, \qquad (13)$$

$$\sum_{i=1}^{K} \mathbf{a}_{[i]} = \sum_{i=1}^{K} \mathbf{b}_{[i]}. \qquad (14)$$

$\square$

The following lemma will play a key role in our framework.

Lemma 1: For the Cholesky factorization $\mathbf{N} = \mathbf{LL}^H$, the following inequalities hold:

$$\frac{\ln \det(\mathbf{N})}{K}(1, \ldots, 1) \prec l \prec (\ln \lambda_1(\mathbf{N}), \ldots, \ln \lambda_K(\mathbf{N})),$$

where $\lambda_k(\mathbf{N})$ is the kth largest eigen value of $\mathbf{N}$. $\square$

Proof: To prove the first inequality, we observe that any vector $\mathbf{a} \in \mathbb{R}^K$ majorizes its mean vector $\overline{\mathbf{a}}$, whose elements are all equal to the mean $\overline{\mathbf{a}}_i = \frac{1}{K} \sum_{i=1}^{K} \mathbf{a}_i$. That is,

$$\overline{\mathbf{a}} \prec \mathbf{a}. \qquad (15)$$

Since $\mathbf{N} = \mathbf{LL}^H$, we have that $\prod_{k=1}^{K} \mathbf{L}_{kk}^2 = \det(\mathbf{LL}^H) = \det(\mathbf{N})$. Hence, the first inequality follows directly. The second inequality follows by applying Weyl's inequality [22] to the matrix $\mathbf{L}$. $\blacksquare$

It is worth observing that the second inequality in Lemma 1 holds with equality when $\mathbf{L}$ is normal [22]. Since $\mathbf{L}$ is a lower triangular matrix, in order to be normal it must be a diagonal matrix [23]. If $\mathbf{L}$ is diagonal, the matrix $\mathbf{C}$ will then be equal to $\mathbf{I}$ and decision feedback equalization will reduce to linear equalization.

The proposed designs will be based on the following classes of functions [21].

Definition 2 (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

$$\mathbf{a} \prec \mathbf{b} \text{ on } \mathcal{A} \Rightarrow f(\mathbf{a}) \leq f(\mathbf{b}) \qquad (16)$$

and is said to be Schur-concave if

$$\mathbf{a} \prec \mathbf{b} \text{ on } \mathcal{A} \Rightarrow f(\mathbf{a}) \geq f(\mathbf{b}). \qquad (17)$$

$\square$

In particular, we will consider communication objectives that can be expressed as the minimization of increasing functions of the MSEs of each data stream, $g(\mathbf{L}_{11}^2, \ldots, \mathbf{L}_{KK}^2) =$

$g(e^{l_1}, \ldots, e^{l_K}) = g(e^l)$, that are either Schur-convex or Schur-concave functions of the logarithm of the MSEs, l.

Let $\mathbf{H}^H \mathbf{H} = \mathbf{U} \mathbf{\Lambda}_{\mathbf{H}} \mathbf{U}^H$ be the eigen value decomposition of $\mathbf{H}^H \mathbf{H}$ such that the entries of the diagonal matrix $\mathbf{\Lambda}_{\mathbf{H}}$ are squared singular values of $\mathbf{H}$, $\sigma_k^2(\mathbf{H})$, in descending order. Let $\mathbf{U}_1$ and $\mathbf{\Lambda}_{\mathbf{H}1}$ be the first K columns of $\mathbf{U}$ and $\mathbf{\Lambda}_{\mathbf{H}}$, respectively. The optimal precoders for the above two classes of design criteria are given by the following theorem.

Theorem 1: The optimal precoder for the class of objectives for which $g(e^l)$ is a Schur-convex function of the logarithm of the MSEs is independent of the actual form of $g(\cdot)$ and is given by:

$$\mathbf{P} = \sqrt{\frac{P_{\text{total}}}{K}} \mathbf{U}_1 \mathbf{V}(\mathbf{\Lambda}_{\mathbf{H}1}), \qquad (18)$$

where $\mathbf{V}(\mathbf{\Lambda}_{\mathbf{H}1})$ is a unitary matrix that results in the QR decomposition of $\mathbf{\Lambda}_{\mathbf{H}1}^{-1/2} \mathbf{V}(\mathbf{\Lambda}_{\mathbf{H}1}) = \mathbf{QR}$ having an $\mathbf{R}$ factor with equal diagonal elements.

For the class of objectives for which $g(e^l)$ is a Schur-concave function of the logarithm of the MSEs, the optimal solution results in $\mathbf{B} = \mathbf{0}$, and hence the optimal zero-forcing linear transceiver is an optimal transceiver for a system with a zero-forcing DFE.

Proof: See the Appendix. $\blacksquare$

Algorithms for obtaining a matrix $\mathbf{\Phi}$ such that the R-factor of the QR decomposition of $\mathbf{A}\mathbf{\Phi}$ has equal diagonal elements were introduced in [24], [25], and $\mathbf{V}$ in (18) can be obtained by applying the algorithms therein to the matrix $\mathbf{\Lambda}_{\mathbf{H}1}^{-1/2}$.

As illustrated by the following examples, the developed framework embraces a wide range of design criteria:

- *Minimization of the sum of the individual MSEs:* In this case the objective is to minimize

$$g(e^l) = \sum_{k=1}^{K} e^{l_k}. \qquad (19)$$

Here, $g(e^l)$ takes the form $\sum_{k=1}^{K} h(l_k)$ for the convex function $h(l_k) = e^{l_k}$, and hence it is a Schur-convex function of l, [21].

- *Minimization of the maximum MSE / Maximization of minimum SNR:* In this case the objective is to minimize

$$g(e^l) = \max_k(e^{l_k}), \qquad (20)$$

which is the composition of the increasing Schur-convex function $\max_k \mathbf{x}_k$ with the increasing and convex function e^{l_k}. Hence, $g(e^l)$ is also a Schur-convex function [21, pp. 63].

- *Minimization of the average Bit Error Rate:* This corresponds to the minimization of the objective

$$g(e^l) = \frac{1}{K} \sum_{k=1}^{K} \text{BER}(\text{SNR}_k) = \sum_{k=1}^{K} \text{BER}(e^{-l_k}), \quad (21)$$

where the BER expression will depend on the constellation used, and we have assumed that the same constellation is used for each element of $\mathbf{s}$ in (1). It can be verified that under a mild constraint on the SNR, the BER expressions for BPSK and M-QAM constellations are convex functions of l_k. Hence, $g(e^l)$ is a Schur-convex function of l.

- *Maximization of Gaussian mutual information* This corresponds to the minimization of

$$g(e^{\boldsymbol{l}}) = \sum_{k=1}^{K} -\log(1 + e^{-\boldsymbol{l}_k}), \qquad (22)$$

which takes the form $\sum_{k=1}^{K} h(\boldsymbol{l}_k)$ for the convex function $h(\boldsymbol{l}_k) = -\log(1 + e^{-\boldsymbol{l}_k})$, and hence it is a Schur-convex function of $\boldsymbol{l}$.

- *Minimization of the product of MSEs:* Minimization of the product of the individual MSEs (or equivalently, the geometric mean of the MSEs) corresponds to the minimization of

$$g(e^{\boldsymbol{l}}) = \log \prod_{k=1}^{K} e^{\boldsymbol{l}_k} = \sum_{k=1}^{K} \boldsymbol{l}_k, \qquad (23)$$

which is both Schur-convex and Schur-concave. Furthermore, since $\sum_{k=1}^{K} \boldsymbol{l}_k = -\sum_{k=1}^{K} \log(\mathrm{SNR})$, at high SNR the minimization of the product of the MSEs corresponds to the maximization of the Gaussian mutual information.

As demonstrated by Theorem 1 and the above examples, the optimal precoder for a system with zero-forcing DFE and a design objective from the Schur-convex class simultaneously optimizes the total MSE, the average bit error rate, and the Gaussian mutual information. MIMO systems with linear precoding and equalization do not achieve this simultaneous optimality, and in the general case each of these objectives results in a different optimal precoder [1]. For design criteria that can be expressed as the minimization of objectives that are both Schur-convex and Schur-concave, both the optimal Schur-convex design in (18) and the optimal linear transceiver will yield the same objective value. In the following sections, we will consider the efficient design of codebooks for limited feedback systems with Schur-convex objectives. Our first step will be to obtain the statistical distribution of the optimal precoder.

IV. STATISTICAL DISTRIBUTION OF OPTIMAL PRECODER FOR SCHUR-CONVEX OBJECTIVES

The optimal precoder for the Schur-convex class of objectives can be written as

$$\mathbf{P} = \sqrt{\frac{P_{\text{total}}}{K}}\, \overline{\mathbf{P}}, \qquad (24)$$

where the matrix $\overline{\mathbf{P}} = \mathbf{U}_1 \mathbf{V}(\mathbf{\Lambda}_{\mathbf{H}1})$ belongs to the Stiefel manifold $\mathcal{S}(N_t, K)$ of complex $N_t \times K$ matrices with orthonormal columns. The statistical distribution of $\overline{\mathbf{P}}$ in (24) plays a key role in the design of the codebooks, and is established in Theorem 2 below. First, we establish an intermediate result.

Lemma 2: For an i.i.d. Rayleigh fading channel matrix $\mathbf{H}$, the matrices $\mathbf{U}_1$ and $\mathbf{V}(\mathbf{\Lambda}_{\mathbf{H}1})$ are statistically independent. Furthermore, $\mathbf{U}_1$ is isotropically distributed over the manifold $\mathcal{S}(N_t, K)$.

Proof: The proof follows directly from the isotropic distribution of the eigen vectors of the Wishart distributed matrix $\mathbf{H}^H \mathbf{H}$ and the independence of its eigen values. ∎

Theorem 2: For an i.i.d. Rayleigh fading channel matrix $\mathbf{H}$, the normalized optimal precoder matrix $\overline{\mathbf{P}}$ is isotropically distributed over the Stiefel manifold $\mathcal{S}(N_t, K)$.

Proof: We first observe from Lemma 2 that $\mathbf{U}_1$ is isotropically distributed over the manifold $\mathcal{S}(N_t, K)$. Hence, its probability distribution $p(\mathbf{U}_1)$ is unaffected by post-multiplication by any deterministic unitary matrix $\mathbf{Z}$; i.e., $p(\mathbf{U}_1) = p(\mathbf{U}_1 \mathbf{Z})$. Hence,

$$p(\overline{\mathbf{P}}) = \int p(\overline{\mathbf{P}}|\mathbf{V})\, p(\mathbf{V})\, d\mathbf{V} \qquad (25)$$

$$= \int p(\mathbf{U}_1)\, p(\mathbf{V})\, d\mathbf{V} = p(\mathbf{U}_1), \qquad (26)$$

Since $\mathbf{U}_1$ is isotropically distributed, then so is $\overline{\mathbf{P}}$. ∎

It is worth noting that for MIMO systems with linear precoding and equalization, the optimal precoder will not be isotropically distributed. That is true for a wide range of objectives under a total power constraint (e.g., [1] and the references therein), and holds for both zero-forcing and MMSE linear receivers. That said, some quantization methods for linear transceivers have been based on a suboptimal underlying scheme that selects the best unitary precoding matrix; e.g., [14]. In that case the distribution of the unquantized precoder is isotropic. In the case of systems with a zero-forcing DFE, we have shown that selection of the best unitary precoding matrix is optimal.

V. PRECODER SELECTION AND CODEBOOK DESIGN

In order to study the codebook design problem, we will first consider the selection method for choosing the best precoding matrix from a given codebook $\mathcal{P}$.

A. Precoding Matrix Selection

Given a codebook for quantizing the normalized optimal precoding matrix $\overline{\mathbf{P}}$, $\mathcal{P} = \{\overline{\mathbf{P}}^j, j = 1, \ldots, |\mathcal{P}|\}$, and a cost function $g(\cdot)$ associated with the design criterion, the receiver will select a normalized precoding matrix from the codebook that yields the minimum value for the cost function; i.e., the receiver will select the index

$$\arg \min_{j=1,\ldots,|\mathcal{P}|} g(e^{\boldsymbol{l}^j}), \qquad (27)$$

where $\boldsymbol{l}^j$ is the vector containing the logarithm of the diagonal elements of $\mathbf{L}^j$, the Cholesky factor of $\mathbf{N}^j = \sigma_n^2 \left(\frac{K}{P_{\text{total}}} \overline{\mathbf{P}}^{j\,H} \mathbf{H}^H \mathbf{H} \overline{\mathbf{P}}^j \right)^{-1}$. The quality of a given codebook can be measured in terms of the average degradation in the value of the objective that is incurred by using a precoder from the codebook rather than the optimal precoder in Theorem 1. Borrowing terminology from the source coding literature, we will refer to this degradation, and various bounds thereon, as distortion measures for the quantization scheme.

B. Grassmann Packing and Codebook Design

In the following section we will consider the design of codebooks to minimize distortion measures for the broad class of objectives $g(e^{\boldsymbol{l}})$ that are Schur-convex in $\boldsymbol{l}$. As shown in the previous section, for these objectives the optimal normalized precoder is uniformly distributed over the Stiefel manifold $\mathcal{S}(N_t, K)$. We observe that the range of the columns of any normalized precoding matrix $\overline{\mathbf{P}}$ represents a K dimensional subspace, $R_{\overline{\mathbf{P}}}$, of $\mathbb{C}^{N_t}$. Hence, the desired codebook

$\mathcal{P} = \{\overline{\mathbf{P}}^j, j = 1, \ldots, |\mathcal{P}|\}$ represents a set of subspaces $\mathcal{R} = \{R_{\overline{\mathbf{P}}^j}, j = 1, \ldots, |\mathcal{P}|\}$, and each of these subspaces can be represented as a point in the associated quotient space, namely the Grassmann manifold; e.g., [26], [27]. In the next section, we will relate the problem of designing codebooks that minimize suitable distortion measures to the Grassmann packing problem that selects a set of subspaces such that the minimum pairwise distance between any two subspaces in the packing is maximized. The distances between two subspaces $R_{\overline{\mathbf{P}}^1}$ and $R_{\overline{\mathbf{P}}^2}$ can be defined in different ways [28]. For example, the projection 2-norm is defined as

$$\text{dist}_{\text{proj2}}(\overline{\mathbf{P}}^1, \overline{\mathbf{P}}^2) = \left\| \overline{\mathbf{P}}^1 \overline{\mathbf{P}}^{1\,H} - \overline{\mathbf{P}}^2 \overline{\mathbf{P}}^{2\,H} \right\|_2, \quad (28)$$

while the Fubini-Study distance is defined as

$$\text{dist}_{\text{FS}}(\overline{\mathbf{P}}^1, \overline{\mathbf{P}}^2) = \arccos\left| \det(\overline{\mathbf{P}}^{1\,H} \overline{\mathbf{P}}^2) \right|. \quad (29)$$

For a given set or a packing of subspaces and a given distance measure, we will denote the minimum pairwise distance between any two subspaces in the packing by

$$d = \min_{1 \le i < j \le |\mathcal{P}|} \text{dist}(\overline{\mathbf{P}}^j, \overline{\mathbf{P}}^i). \quad (30)$$

In addition to the minimum distance of the packing d, we will also be interested in its density D; e.g., [28]. In our context, the density is the probability that the range space of an isotropically distributed unitary matrix falls within a distance $d/2$ of any of the subspaces of the packing, and is function of d, $|\mathcal{P}|$ and the volume of the manifold; see [28]. In the following two sections, we will show that codebooks from certain optimized Grassmann packings minimize distortion measures that are appropriate for two subclasses of the Schur-convex objectives: the strict Schur-convex objectives, and the objectives that are both Schur-convex and Schur-concave functions of l.

C. Codebook Designs for Strictly Schur-convex Objectives

In this section we will present suitable distortion measures for objectives $g(e^l)$ that are Schur-convex functions of l and are not Schur-concave; e.g., the sum of the MSEs, the maximum MSE and the BER. From first principles, we can obtain the following bounds on the these objectives:

• *Minimization of the sum of MSE:*

$$g(e^l) = \sum_{k=1}^{K} e^{l_k} \le K \max_k e^{l_k} = \frac{K}{\min_k e^{-l_k}}. \quad (31)$$

• *Minimization of the maximum MSE / Maximization of minimum SNR:*

$$g(e^l) = \max_k (e^{l_k}) = \frac{1}{\min_k e^{-l_k}}. \quad (32)$$

• *Minimization of the average Bit Error Rate:*

$$g(e^l) = \sum_{k=1}^{K} \text{BER}(e^{-l_k}) \le K \ \text{BER}(\min_k e^{-l_k}). \quad (33)$$

We observe that each of these bounds is expressed in terms of the minimum SNR over the K data streams, $\text{SNR}_{\min} = \min_k e^{-l_k}$.

Since each of these terms is bounded by the minimum SNR, a natural choice for the distortion measure for a given codebook is the average loss in the minimum SNR that one incurs by using a normalized precoder $\overline{\mathbf{P}}^{\text{quant}}$ chosen from the codebook $\mathcal{P}$ instead of using the optimal normalized precoder $\overline{\mathbf{P}}^{\text{opt}}$. That is,

$$\mathcal{E} = \mathbf{E_H}\left\{ \text{SNR}_{\min}(\overline{\mathbf{P}}^{\text{opt}}) - \text{SNR}_{\min}(\overline{\mathbf{P}}^{\text{quant}}) \right\}$$
$$= \frac{\mathbf{E_H}\{\sqrt[K]{\det \mathbf{\Lambda_{H1}}}\}}{\sigma_n^2} - \mathbf{E_H}\left\{ \max_{1 \le j \le |\mathcal{P}|} \min_{1 \le k \le K} e^{-l_k^j} \right\}, \quad (34)$$

where (34) follows by observing that the optimal $\overline{\mathbf{P}}$ results in $l_k = \frac{\ln \det(\mathbf{N})}{K}$ for every k. Consider the second term in the distortion measure in equation (34). From the definition of the majorization relation $\mathbf{a} \prec \mathbf{b}$, we have $\mathbf{a}_{[1]} \le \mathbf{b}_{[1]}$. Hence, from Lemma 1 we have

$$\max_{1 \le k \le K} l_k \le \ln \lambda_1(\mathbf{N}) = \ln\left(\frac{\sigma_n^2}{\sigma_{\min}^2(\mathbf{HP})} \right), \quad (35)$$

from which it follows that

$$\mathbf{E_H}\left\{ \max_{1 \le j \le |\mathcal{P}|} \min_{1 \le k \le K} e^{-l_k^j} \right\} \ge \mathbf{E_H}\left\{ \max_{1 \le j \le |\mathcal{P}|} \frac{\sigma_{\min}^2(\mathbf{HP}^j)}{\sigma_n^2} \right\}. \quad (36)$$

Hence, the distortion measure in (34) is upper bounded by

$$\mathcal{E} \le \frac{\mathbf{E_H}\{\sqrt[K]{\det \mathbf{\Lambda_{H1}}}\}}{\sigma_n^2} - \mathbf{E_H}\left\{ \max_{1 \le j \le |\mathcal{P}|} \frac{\sigma_{\min}^2(\mathbf{HP}^j)}{\sigma_n^2} \right\}. \quad (37)$$

When codebooks are designed from a Grassmann packing using the projection 2-norm distance in (28), the expectation on the right hand side of (36) satisfies [14],

$$\mathbf{E_H}\left\{ \max_{1 \le j \le |\mathcal{P}|} \sigma_{\min}^2(\mathbf{HP}^j) \right\} \ge \mathbf{E_H}\{\sigma_K^2(\mathbf{H})\} D_{\text{proj2}}\left(1 - \frac{d_{\text{proj2}}^2}{4} \right), \quad (38)$$

where d_{proj2} is the minimum pairwise distance of the packing (cf. (30)) for the projection 2-norm distance, and D_{proj2} is the corresponding packing density; cf. [28]. In addition, for a given $|\mathcal{P}|$ the right hand side of (38) is an increasing function of the packing distance d_{proj2}. Using the inequality in (37), we obtain the following upper bound on the distortion:

$$\mathcal{E} \le \frac{\mathbf{E_H}\{\sqrt[K]{\det \mathbf{\Lambda_{H1}}}\}}{\sigma_n^2} - \frac{\mathbf{E_H}\{\sigma_K^2(\mathbf{H})\}}{\sigma_n^2} D_{\text{proj2}}\left(1 - \frac{d_{\text{proj2}}^2}{4} \right), \quad (39)$$

which, for a given $|\mathcal{P}|$, is a decreasing function of the packing distance d_{proj2}. The bound on the right hand side of (39) can be easily manipulated by choosing the codebook from a Grassmann packing that is designed to maximize the packing distance d in (30) with the projection 2-norm as the distance metric. Such designs correspond to minimizing the bound on the distortion.

Since permutation matrices are special cases of unitary matrices, the limited feedback approach in [17], [18], in which the precoder is chosen from a codebook of permutation matrices, is a special case of our proposed design. However, the resulting codebooks do not necessarily have the maximum packing distance. Furthermore, the size of the codebook in the approaches in [17], [18] is fixed for a given N_t and K, while the Grassmann packings can be constructed for an arbitrary number of codewords.

D. Codebook Designs for Objectives that are Both Schur-convex and Schur-concave

For communication objectives $g(e^l)$ that are both Schur-convex and Schur-concave functions of l, such as the minimization of product of the MSEs, we observe that the design problem corresponds to maximization of $\det(\mathbf{P}^H \mathbf{H}^H \mathbf{H} \mathbf{P})/\sigma_n^2$. Hence, a suitable distortion measure for the codebook is

$$\mathcal{E} = \mathrm{E}_{\mathbf{H}}\Big\{\det(\overline{\mathbf{P}}^{\mathrm{opt}\,H}\mathbf{H}^H\mathbf{H}\overline{\mathbf{P}}^{\mathrm{opt}})$$
$$- \det(\overline{\mathbf{P}}^{\mathrm{quant}\,H}\mathbf{H}^H\mathbf{H}\overline{\mathbf{P}}^{\mathrm{quant}})\Big\}/\sigma_n^2 \quad (40)$$
$$= \mathrm{E}_{\mathbf{H}}\{\det \boldsymbol{\Lambda}_{\mathbf{H}1}\}/\sigma_n^2$$
$$- \mathrm{E}_{\mathbf{H}}\Big\{\max_{1\le j\le|\mathcal{P}|}\det(\overline{\mathbf{P}}^{j\,H}\mathbf{H}^H\mathbf{H}\overline{\mathbf{P}}^j)\Big\}/\sigma_n^2$$
$$\le \mathrm{E}_{\mathbf{H}}\{\det \boldsymbol{\Lambda}_{\mathbf{H}1}\}/\sigma_n^2$$
$$- \mathrm{E}_{\mathbf{H}}\{\det \boldsymbol{\Lambda}_{\mathbf{H}1}\}\mathrm{E}_{\mathbf{H}}\Big\{\max_{1\le j\le|\mathcal{P}|}\det(\overline{\mathbf{P}}^{j\,H}\mathbf{U}_1\mathbf{U}_1^H\overline{\mathbf{P}}^j)\Big\}/\sigma_n^2.$$
$$(41)$$

Here, (41) follows from the independence of $\mathbf{U}$ and $\boldsymbol{\Lambda}$. When codebooks are designed from a Grassmann packing using the Fubini-Study distance in (29), the last expectation on the right hand side of (41) satisfies the following inequality [14]:

$$\mathrm{E}_{\mathbf{H}}\Big\{\max_{1\le j\le|\mathcal{P}|}\det(\overline{\mathbf{P}}^{j\,H}\mathbf{U}_1\mathbf{U}_1^H\overline{\mathbf{P}}^j)\Big\} \ge D_{\mathrm{FS}}\cos^2(d_{\mathrm{FS}}/2).$$
$$(42)$$

Hence, we obtain the following upper bound on the distortion:

$$\mathcal{E} \le \mathrm{E}_{\mathbf{H}}\{\det \boldsymbol{\Lambda}_{\mathbf{H}1}\}\big(1 - D_{\mathrm{FS}}\cos^2(d_{\mathrm{FS}}/2)\big)/\sigma_n^2, \quad (43)$$

which, for a given $|\mathcal{P}|$, is a decreasing function of the packing distance d_{FS}. A similar upper bound was proposed for designing codebooks for MIMO systems with linear receivers [14].

E. Comparison with ZF-Linear Schemes

In this section, we will show that for a given codebook, the performance of the zero-forcing DFE with limited feedback provides an upper bound on the performance of its linear zero-forcing counterpart for any Schur-convex performance objective $g(e^l)$. As stated in the following lemma, this is true for any codebook, including those codebooks constructed from non-unitary matrices.

Lemma 3: Consider a codebook of precoding matrices, $\mathcal{P}$, and a Schur-convex performance $g(e^l)$. For any given channel $\mathbf{H}$, let l_{DFE}^j denote the vector l in (12) when the precoder $\mathbf{P}^j$ is used, and let the l_{Lin}^j denote the corresponding vector for the case of linear equalization. Then

$$\min_{j=1,\ldots,|\mathcal{P}|} g(e^{l_{\mathrm{DFE}}^j}) \le \min_{j=1,\ldots,|\mathcal{P}|} g(e^{l_{\mathrm{Lin}}^j}).$$

Proof: Consider a given channel $\mathbf{H}$ and any precoding matrix $\mathbf{P}^j \in \mathcal{P}$. For the linear zero-forcing receiver we have $\mathbf{C} = \mathbf{I}$. It follows from (9) that the corresponding matrix $\mathbf{N}^j$ and its Cholesky factor $\mathbf{L}^j$ are diagonal. Hence, $(\mathbf{L}_{ii}^j)^2 = \lambda_i(\mathbf{N}^j)$, or, equivalently,

$$l_{\mathrm{Lin}}^j = (\ln \lambda_1(\mathbf{N}^j), \ldots, \ln \lambda_K(\mathbf{N}^j)).$$

On the other hand, for the DFE receiver we have

$$l_{\mathrm{DFE}}^j = (\ln(\mathbf{L}_{11}^j)^2, \ldots, \ln(\mathbf{L}_{KK}^j)^2).$$

From Lemma 1, we have $l_{\mathrm{DFE}}^j \prec l_{\mathrm{Lin}}^j$, hence $g(e^{l_{\mathrm{DFE}}^j}) \le g(e^{l_{\mathrm{Lin}}^j})$ and

$$\min_{j=1,\ldots,|\mathcal{P}|} g(e^{l_{\mathrm{DFE}}^j}) \le \min_{j=1,\ldots,|\mathcal{P}|} g(e^{l_{\mathrm{Lin}}^j}).$$

∎

VI. SIMULATION STUDIES

In this section, we simulate the performance of the proposed limited feedback MIMO schemes over a standard i.i.d. Rayleigh block fading channel model.[1] For the error rate performance comparisons, we use 16-QAM signaling and we plot the average bit error rate (BER) of the K data streams against the signal-to-noise-ratio, which is defined as the ratio of the total average transmitted power P_{total} to the total receiver noise power $\mathrm{E}\{\mathbf{n}^H \mathbf{n}\}$. We compare the performance of the proposed codebook designs for systems with zero-forcing DFE with that of the optimal zero-forcing DFE transceiver for the case of perfect CSI that was presented in Section III. For the proposed limited-feedback schemes, the Grassmann codebooks are constructed using the design approach in [29]; see also [14]. (Grassmann codebooks could also be constructed using the optimization algorithms in [26], [27]). We also provide simulation-based comparisons with the two limited feedback schemes for zero-forcing DFE systems in [18]. In addition, we provide performance comparisons with limited feedback schemes for linear zero-forcing transceivers that use Grassmann codebooks [14], and with the optimal zero-forcing linear transceiver designs for the case of perfect CSI for minimum MSE and minimum bit error rate design criteria [30].

A. Comparisons with Limited Feedback Zero-forcing DFE Schemes

In Fig. 2, we consider a MIMO system with $N_t = 6$ transmit antennas and $N_r = 3$ receive antennas that transmits $K = 3$ independent data streams. We compare the performance of the proposed schemes with Grassmann codebook designs and precoder selection based on the minimization of the sum of the MSEs (Grassmann-6 bits- Sum MSE), minimization of the average BER (Grassmann-6 bits- Min BER), and the minimization of the maximum MSE (Grassmann-6 bits- Max MSE) which is equivalent to the maximization of minimum SINR. The codebooks consist of 64 unitary matrices, and hence 6 bits of feedback are used per block. We also make comparisons with the limited feedback schemes in [18] (Ordering Feedback ZF-DFE and Ordering Feedback2 ZF-DFE) in which the receiver feeds back the index of the selected permutation of the columns of $\mathbf{H}$ from the set of possible $P_K^{N_t} = N_t!/(N_t - K)!$ permutation matrices. For the system under consideration, the number of possible permutations matrices is 120, almost twice the size of the Grassmann

---

[1]The coefficients of the channel matrix $\mathbf{H}$ are modelled as independent circularly symmetric complex Gaussian random variables with zero mean and unit variance.
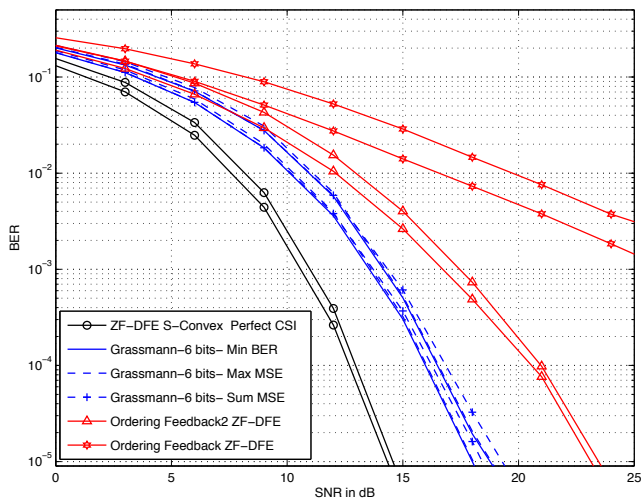
Fig. 2. BER performance of various MIMO transmission schemes with zero-forcing DFE for a system with $N_t = 6, N_r = 3$, and $K = 3$ simultaneously transmitted 16-QAM data streams. The schemes considered are: the proposed codebook designs for the objectives of minimization of the sum of MSEs (Grassmann-6 bits- Sum MSE), minimization of the average BER (Grassmann-6 bits- Min BER); the optimal zero-forcing design for any Schur-convex design objective with perfect CSI (ZF DFE - Perfect CSI); and the limited feedback schemes in [18], which are based on feeding back the detection ordering (Ordering Feedback - ZF DFE) and (Ordering Feedback2 - ZF DFE). The lower curve for each method represents the BER performance obtained under the assumption of correct previous decisions.
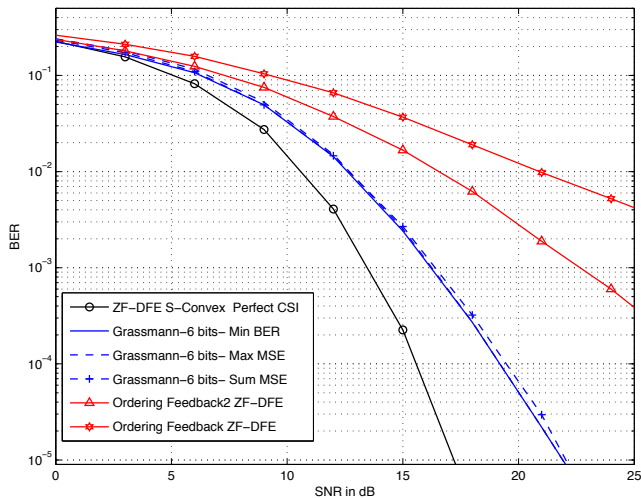

Fig. 4. Average of Gaussian mutual information in (22) for various MIMO transmission schemes with zero-forcing DFE for a system with $N_t = 5, N_r = 4$, and $K = 4$. The schemes considered are: the proposed codebook designs for Gaussian mutual information objective (Grassmann-6 bits- Mutual info); the optimal zero-forcing design for any Schur-convex design objective with perfect CSI (ZF DFE - Perfect CSI); and the limited feedback schemes in [18], which are based on feeding back the detection ordering (Ordering Feedback - ZF DFE) and (Ordering Feedback2 - ZF DFE).

the proposed codebooks with different Schur-convex selection criteria. This is to be expected, because in the limit of infinite feedback (i.e., perfect CSI), all these objectives result in the same optimal precoder design. We also observe that the Grassmann codebooks provide significantly better performance than the schemes that are based on precoding with permutation matrices, even though they employ fewer feedback bits. This is because codebooks constructed from permutation matrices are special cases of those constructed from unitary matrices, and they do not necessarily minimize the distortion measures. Note that for all error performance figures in this paper, the simulation results of all ZF-DFE methods include the effect of error propagation. For reference, in Fig. 2 we also provide the performance under the assumption of correct previous decisions; i.e., no error propagation. We observe that at high SNRs, the practical performance of the optimal zero-forcing DFE transceiver for the case of perfect CSI and the proposed designs based on Grassmann codebooks are close to their corresponding performance in absence of error propagation. This also holds for the permutation feedback scheme (Ordering Feedback2 ZF-DFE).

Analogous performance advantages to those in Fig. 2 are observed in Fig. 3, which shows the performance for a MIMO system with $N_t = 5$ transmit antennas and $N_r = 4$ receive antennas that transmits $K = 4$ data streams. The size of each permutation-based codebook is 120 matrices, while the size of each Grassmann codebook is 64 matrices.

In Fig. 4 we compare several different methods in terms of the Gaussian mutual information that they achieve. We consider a system with $N_t = 5$, $N_r = 4$, and $K = 4$, and we plot the average, over 1000 channel realizations, of the Gaussian mutual information achieved by the ZF-DFE transceiver with the quantized precoder; i.e., the average of the values of (22) achieved by the quantized precoder. For

Fig. 3. BER performance of various MIMO transmission schemes with zero-forcing DFE for a system with $N_t = 5, N_r = 4$, and $K = 4$ simultaneously transmitted 16-QAM data streams. The schemes considered are: the proposed codebook designs for the objectives of minimization of the sum of MSEs (Grassmann-6 bits- Sum MSE), minimization of the average BER (Grassmann-6 bits- Min BER); the optimal zero-forcing design for any Schur-convex design objective with perfect CSI (ZF DFE - Perfect CSI); and the limited feedback schemes in [18], which are based on feeding back the detection ordering (Ordering Feedback - ZF DFE) and (Ordering Feedback2 - ZF DFE).

codebook. In the scheme denoted Ordering Feedback ZF-DFE the permutation matrix is selected based on the norms of the columns of $\mathbf{H}$, while the scheme denoted Ordering Feedback2 ZF-DFE the permutation is selected based on a greedy ordering of the QR decomposition of the channel matrix $\mathbf{H}$. In Fig. 2, we observe the close performance of
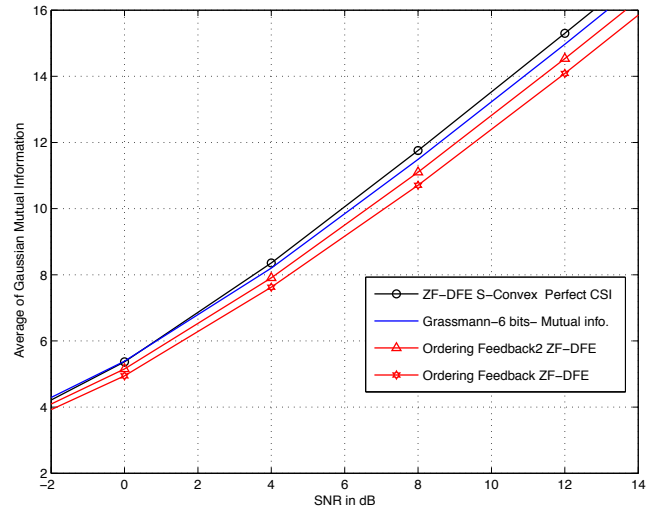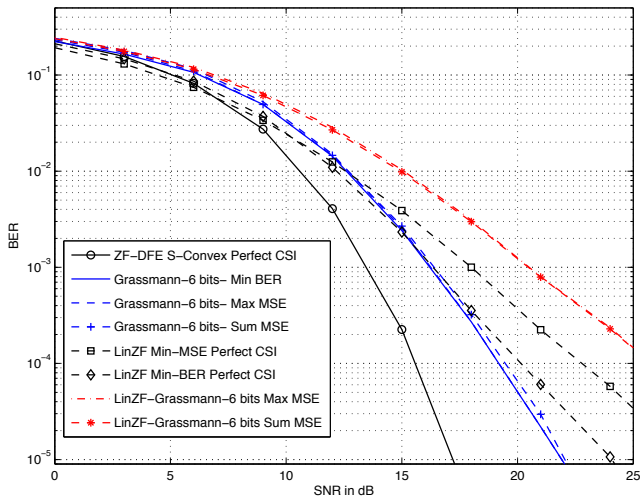
Fig. 5. BER performance of various MIMO transmission schemes with zero-forcing linear and DFE systems with $N_t = 5, N_r = 4$, and $K = 4$ simultaneously transmitted 16-QAM data streams. The schemes considered are: the proposed codebook designs for the objectives of minimization of the sum of MSEs (Grassmann-6 bits- Sum MSE), minimization of the average BER (Grassmann-6 bits- Min BER); the optimal zero-forcing design for any Schur-convex design objective with perfect CSI (ZF DFE - Perfect CSI); the optimal linear zero-forcing design for minimum MSE (LinZF Min-MSE Perfect CSI) and minimum average BER (LinZF Min-BER Perfect CSI) [30]; and the linear zero-forcing limited feedback schemes in [14] for minimum total MSE (LinZF-Grassmann-6 bits Sum MSE) and minimum maximum MSE (LinZF-Grassmann-6 bits Max MSE).

Fig. 6. BER performance of various MIMO transmission schemes with zero-forcing linear and DFE systems with $N_t = 4, N_r = 3$, and $K = 3$ simultaneously transmitted 16-QAM data streams. The schemes considered are: the proposed codebook designs for the objectives of minimization of the sum of MSEs (Grassmann-6 bits- Sum MSE), minimization of the average BER (Grassmann-6 bits- Min BER); the optimal zero-forcing design for any Schur-convex design objective with perfect CSI (ZF DFE - Perfect CSI); the optimal linear zero-forcing design for minimum MSE (LinZF Min-MSE Perfect CSI) and minimum average BER (LinZF Min-BER Perfect CSI) [30]; and the linear zero-forcing limited feedback schemes in [14] for minimum total MSE (LinZF-Grassmann-6 bits Sum MSE) and minimum maximum MSE (LinZF-Grassmann-6 bits Max MSE).

the proposed scheme we consider a Grassmann codebook design and precoder selection based on the maximization of the Gaussian mutual information (Grassmann-6 bits- Mutual info.), and a codebook that consists of 64 unitary matrices. We make comparisons with the limited feedback schemes in [18] (Ordering Feedback ZF-DFE and Ordering Feedback2 ZF-DFE), whose permutation-based codebooks contain 120 matrices. We observe that the proposed Grassmann codebook with precoder selection based on the maximization of the Gaussian mutual information provides the closest performance to the optimal ZF-DFE design for the case of perfect CSI, which was presented in Section III.

B. Comparisons with Limited Feedback Linear Zero-forcing Schemes

In Fig. 5, we consider a MIMO system with $N_t = 5$ transmit antennas and $N_r = 4$ receive antennas that transmits $K = 4$ independent data streams. We compare the performance of the proposed ZF-DFE schemes that use Grassmann codebooks with that of the corresponding linear zero-forcing schemes that use Grassmann codebooks with the same feedback rate [14]. We consider linear limited feedback schemes with different precoder selection criteria, namely minimization of the total MSE (LinZF-Grassmann-6 bits Sum MSE), and maximization of the minimum eigen value of the overall channel **HP** (LinZF-Grassmann-6 bits Max MSE), which corresponds to minimization of the maximum MSE [14]. We also provide performance comparisons with the zero-forcing DFE transceiver design for perfect CSI that simultaneously optimizes any Schur-Convex design criteria, and with the corresponding optimal zero-forcing linear transceiver designs for perfect CSI
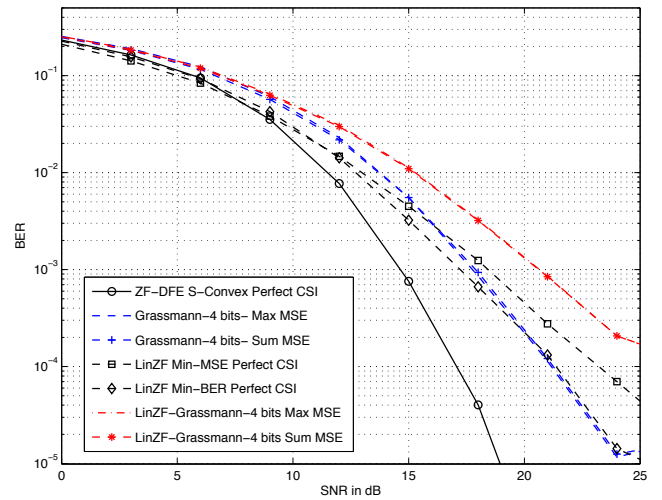
that minimize the total MSE or the average BER. Unlike the DFE case, these two design criteria result in different precoder designs [30]. In Fig. 5, we observe that the proposed zero-forcing DFE systems with limited feedback perform better than the corresponding linear schemes; as is to be expected, c.f. Lemma 3. Similar performance advantages are observed in Fig. 6 for a MIMO system with $N_t = 4$ transmit antennas and $N_r = 3$ receive antennas that transmits $K = 3$ independent data streams.

VII. CONCLUSION

We have considered the design of multiple-input multiple-output communication systems with zero-forcing decision feedback equalization (DFE) when only limited rate feedback from the receiver to the transmitter is available. We considered schemes in which the receiver uses its CSI to select the best available precoder from a codebook of precoders and then feeds back the index of this precoder to the transmitter using a small number of bits. To facilitate the development of the limited feedback scheme, we developed a unified design framework for the joint design of the precoder and DFE receiver when perfect channel state information is available. We then characterized the statistical distribution of the optimal precoder in a standard Rayleigh fading environment, and showed that codebooks constructed from Grassmann packings minimize an upper bound on an average distortion measure. Our simulation studies showed that the proposed limited feedback scheme can provide significantly better performance with a lower feedback rate than the existing schemes in which the detection order is fed back to the transmitter.

APPENDIX

A. Optimal Precoder for Schur-convex Functions

If $g(e^l)$ is a Schur convex function of l, then from Lemma 1 we have that

$$g(e^{\overline{l}}) \leq g(e^l), \tag{44}$$

and the optimal value is obtained when all l_i are equal to

$$l_i = \frac{1}{K} \ln \det(\mathbf{N}). \tag{45}$$

Hence, all MSEs are equal to $\mathbf{E}_{ii} = \mathbf{L}_{ii}^2 = \sqrt[K]{\det(\mathbf{N})}$. Since the objective is an increasing function of the individual MSEs, the design goal reduces to minimizing $\det \mathbf{N}$ subject to the power constraint on the precoder and to the constraint that diagonal elements of the Cholesky factor of $\mathbf{N}$ are all equal. We will start by characterizing the family of precoders that minimize $\det(\mathbf{N})$ subject to the power constraint, then we will show that there is a member of this family that yields a Cholesky factor of $\mathbf{N}$ with equal diagonal elements. Minimizing $\det(\mathbf{N})$ is equivalent to maximizing $\det(\mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P})$, and the family of optimal precoders is given by [31]:

$$\mathbf{P} = \sqrt{\frac{P_{\text{total}}}{K}} \mathbf{U}_1 \mathbf{V}, \tag{46}$$

where $\mathbf{U}_1 \in \mathbb{C}^{N_t \times K}$ contains the eigen vectors of $\mathbf{H}^H\mathbf{H}$ corresponding to the K largest eigen values, and $\mathbf{V} \in \mathbb{C}^{K \times K}$ is a unitary matrix degree of freedom. To complete the design of $\mathbf{P}$, we need to select $\mathbf{V}$ such that the Cholesky decomposition of $\mathbf{N} = \mathbf{L}\mathbf{L}^H$ yields an $\mathbf{L}$ factor with equal diagonal elements. Using (46) we have that

$$\begin{aligned}
\mathbf{N} &= \frac{K\sigma_n^2}{P_{\text{total}}} \left(\mathbf{V}^H \mathbf{\Lambda}_{\mathbf{H}1}^{-1/2} \right) \left(\mathbf{\Lambda}_{\mathbf{H}1}^{-1/2} \mathbf{V} \right) \\
&= \mathbf{L}\mathbf{L}^H = \mathbf{R}^H\mathbf{R} = (\mathbf{Q}\mathbf{R})^H(\mathbf{Q}\mathbf{R}),
\end{aligned} \tag{47}$$

where $\mathbf{\Lambda}_{\mathbf{H}1}$ is the diagonal matrix containing the largest K eigen values of $\mathbf{H}^H\mathbf{H}$, and $\mathbf{Q}$ is a matrix with orthonormal columns. Therefore, finding $\mathbf{V}$ is equivalent to finding a $\mathbf{V}$ such that QR decomposition of $\left(\mathbf{\Lambda}_{\mathbf{H}1}^{-1/2}\mathbf{V} \right)$ has an R-factor with equal diagonal. This problem was solved in [24], [25], and $\mathbf{V}$ can be obtained by applying the algorithms therein to the matrix $\mathbf{\Lambda}_{\mathbf{H}1}^{-1/2}$.

B. Optimal Precoder for Schur-concave Functions

If $g(e^l)$ is a Schur-concave function of l, then from Lemma 1 we have that $g(e^l)$ is minimized when $\mathbf{L}_{ii}^2 = \lambda_i(\mathbf{N})$, and that this equality holds when $\mathbf{L}$ is normal matrix. Since $\mathbf{L}$ is a lower triangular matrix, in order for it to be normal it must be a diagonal matrix [23]. The optimal $\mathbf{C}$ in that case is $\mathbf{I}$, and hence $\mathbf{B} = \mathbf{0}$. That is, in the case of Schur-concave functions of l, the optimal ZF-DFE design results in zero-forcing linear equalization.

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, no. 9, pp. 2381–2401, Sept. 2003.

[2] M. Botros Shenouda and T. N. Davidson, "A framework for designing MIMO systems with decision feedback equalization or Tomlinson-Harashima precoding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Honolulu, April 2007, pp. III-209 – III-212.

[3] ——, "A framework for designing MIMO systems with decision feedback equalization or Tomlinson Harashima precoding," *IEEE J. Select. Areas Commun.*, vol. 26, no. 2, pp. 401–411, Feb. 2008.

[4] Y. Jiang, D. Palomar, and M. K. Varanasi, "Precoder optimization for nonlinear MIMO transceiver based on arbitrary cost function," in *Proc. Conf. Information Sciences Systems*, Baltimore, March 2007, pp. 119–124.

[5] D. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4-5, pp. 331-551, 2006.

[6] D. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[7] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[8] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[9] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[10] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna system," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[11] W. Santipach and M. L. Honig, "Asymptotic performance of MIMO wireless channels with limited feedback," in *Proc. IEEE Mil. Commun. Conf.*, vol. 1, Boston, Oct. 2003, pp. 141–146.

[12] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[13] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

[14] D. J. Love and R. W. Heath Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[15] ——, "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 64–73, Jan. 2005.

[16] G. Jongren and M. Skoglund, "Quantized feedback information in orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2473–2486, Oct. 2004.

[17] Y. Bae and J. Lee, "Antenna selection for MIMO systems with sequential nulling and cancellation," in *Proc. Conf. Information Sciences Systems*, Princeton, March 2006, pp. 745–749.

[18] Y. Jiang and M. K. Varanasi, "A novel spatial multiplexing architecture with finite rate feedback," in *Proc. Conf. Information Sciences Systems*, Princeton, March 2006, pp. 755–760.

[19] 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.

[20] 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.

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

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

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

[24] 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.

[25] Y. Jiang, W. W. Hager, and J. Li, "The geometric mean decomposition," *Linear Algebra Appl.*, vol. 396, pp. 373–384, Feb. 2005.

[26] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Applicat.*, vol. 20, no. 2, pp. 303–353, Apr. 1999.

[27] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 635–650, Mar. 2002.

[28] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2450–2454, Sept. 2002.

[29] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.

[30] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2410–2423, Sept. 2003.

[31] M. Botros Shenouda and T. N. Davidson, "Minimum SER zero-forcing transmitter design for MIMO channels with interference pre-subtraction," in *Proc. IEEE Veh. Technol. Conf.*, Dublin, Apr. 2007, pp. 2109–2113.

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, robust and stochastic 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. He is also a Registered Professional Engineer in the Province of Ontario. Dr. Davidson's 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 as an Editor of the IEEE Transactions on Wireless Communications. He has also served as an Associate Editor of the IEEE Transactions on Circuits and Systems II, and 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.