MMSE OPTIMAL FEEDBACK OF CORRELATED CSI FOR MULTI-USER PRECODING

Michael Joham[1], Paula M. Castro[2], Luis Castedo[2], and Wolfgang Utschick[1]*

1) Technische Universität München, Theresienstr. 90, 80290 Munich, Germany
2) Universidad de A Coruña, Campus de Elviña s/n, 15071. A Coruña, Spain
Email: {joham,utschick}@tum.de, {pcastro,luis}@udc.es

ABSTRACT

For the separation of the signals for multiple users in the vector *broadcast channel* (BC), *channel state information* (CSI) is necessary at the transmitter. Since the transmitter has no access to this information in many cases, the CSI must be fed back from the receivers to the transmitter. Before the feedback, the receivers estimate the CSI and apply a rank reduction possible due to the channel correlations. We propose a joint optimization of the estimation, the rank reduction, and the codebook used for the feedback. Interestingly, the estimator and the rank reduction resulting from this monolithic formulation are independent of the used codebook which can be computed with the generalized Lloyd algorithm. Applying the proposed feedback design to a system with multi-user precoding based on CSI feedback shows the clear superiority of the optimized codebook compared to previous designs.

Index Terms— Feedback systems, mean square methods, transceivers, multi-user channel.

1. INTRODUCTION

For the transmission in the vector BC, the transmitter must have some CSI knowledge, if the capabilities of the BC should be exploited better than with a TDMA scheme [1]. Unfortunately, the transmitter cannot obtain this knowledge in many systems (such as for frequency division duplex) and the CSI must be fed back from the receivers to the transmitter.

Most papers on feedback for precoding concentrated on the single-user case (see e.g., [2]). In [3], the impact of feedback on the outage rate of a single-user MISO system was studied. The connection of the optimization of a single-user MIMO system with single stream transmission to Grassmannian line packing was shown and exploited in [4]. Spatial multiplexing in single-user MIMO was considered in [5] and the analysis of a single-user MIMO system for a large number of transmit antennas was reported in [6]. The feedback of unitary beamforming matrices was considered in [7].

For multi-user scenarios, mainly zero-forcing approaches have been proposed. The rate of a vector BC system with random vector quantization was investigated in [8], where it was shown that the necessary number of bits is proportional to the SNR and the number of transmit antennas. The feedback for a multi-user system with zero-forcing dirty paper coding was proposed in [9]. In [10], the sum rate performance of zero-forcing techniques for the vector BC was analyzed, where also shape feedback was considered.

*This work was supported by the Ministerio de Educacion y Ciencia of Spain (grant number TEC2004-06451-C05-01 and integrated action HA2006-0112), FEDER funds of the European Union (grant numberTEC2004-06451-C05-01), and DAAD of Germany (integrated action number D/06/12809).

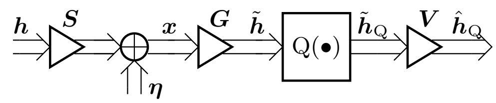

Fig. 1. Feedback with estimation and rank reduction

We propose a feedback design for correlated channels including the estimation, the rank reduction, and the quantizer. The new formulation is a considerable extension to that of [11], where we only optimized the estimation and rank reduction. We obtain the very useful result that the optimal estimator and rank reduction only depend on the channel statistics and are independent of the used quantizer. The distortion is a diagonally weighted squared error and thus, the Lloyd algorithm can be employed to compute the quantizer. The fed back CSI is employed to design a robust non-zero-forcing precoder, i.e., robust *minimum MSE* (MMSE) precoder.

2. SYSTEM MODEL

For the sake of notational brevity, our formulation is without a user index. The N_{tr} N-dimensional training symbols are transmitted over the channel $\boldsymbol{h} \in \mathbb{C}^N \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{h}})$ and perturbed with some noise. We collect the N_{tr} scalar outputs of the training channel in

$$\boldsymbol{x} = \boldsymbol{S}\boldsymbol{h} + \boldsymbol{\eta} \in \mathbb{C}^{N_{\text{tr}}} \tag{1}$$

where $\boldsymbol{S} \in \mathbb{C}^{N_{\text{tr}} \times N}$ contains the training symbols and the noise is $\boldsymbol{\eta} \in \mathbb{C}^{N_{\text{tr}}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}})$. The filter $\boldsymbol{G} \in \mathbb{C}^{d \times N_{\text{tr}}}$ performs estimation and rank reduction in one step. The coefficients of the rank-reduced representation are collected in

$$\tilde{\boldsymbol{h}} = \boldsymbol{G}\left(\boldsymbol{S}\boldsymbol{h} + \boldsymbol{\eta}\right) \in \mathbb{C}^d \tag{2}$$

which is the input to the quantizer

$$\mathrm{Q}(\boldsymbol{y}) = \sum_{i=1}^{M} \boldsymbol{y}_i \, \mathrm{S}_i(\boldsymbol{y}) \tag{3}$$

with M codebook entries $\boldsymbol{y}_i, i = 1, \ldots, M$. The selector function $\mathrm{S}_i(\boldsymbol{y})$ is 1, if $\boldsymbol{y} \in \mathcal{R}_i$, and 0 else. The disjoint convex partition cells $\mathcal{R}_i$ fulfill $\bigcup_{i=1}^{M} \mathcal{R}_i = \mathbb{C}^d$, i.e., $\mathrm{Q}(\boldsymbol{y})$ is regular (e.g., [12]). The output $\tilde{\boldsymbol{h}}_{\mathrm{Q}} = \mathrm{Q}(\tilde{\boldsymbol{h}})$ is transmitted over the error-free feedback channel. With the basis $\boldsymbol{V} \in \mathbb{C}^{N \times d}$ of the rank reduction, the CSI at the transmitter reads as (see Fig. 1)

$$\hat{\boldsymbol{h}}_{\mathrm{Q}} = \boldsymbol{V}\,\mathrm{Q}\left(\boldsymbol{G}\boldsymbol{S}\boldsymbol{h} + \boldsymbol{G}\boldsymbol{\eta}\right) \in \mathbb{C}^N. \tag{4}$$

3. MMSE BASED FEEDBACK DESIGN

Our goal is the joint optimization of the orthonormal basis $\boldsymbol{V}$, the equalizer $\boldsymbol{G}$, the codebook entries $\boldsymbol{y}_i$, and the partition cells $\mathcal{R}_i$, $i = 1, \ldots, M$, by minimizing the MSE $\varepsilon = \mathrm{E}[\|\boldsymbol{h} - \hat{\boldsymbol{h}}_{\mathrm{Q}}\|_2^2]$, i.e.,

$$\{\boldsymbol{V}, \boldsymbol{G}, \boldsymbol{y}_i, \mathcal{R}_i\}_{\text{opt}} = \underset{\{\boldsymbol{V},\boldsymbol{G},\boldsymbol{y}_i,\mathcal{R}_i\}}{\operatorname{argmin}} \ \varepsilon \quad \text{s.t.: } \boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \mathbf{I}. \tag{5}$$

As we will see in the following, the main difficulty is the derivation of $\boldsymbol{V}$ and $\boldsymbol{G}$. The conditions for $\boldsymbol{y}_i$ and $\mathcal{R}_i$ are standard. Note that $\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}} \neq \mathbf{I}$ although $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \mathbf{I}$, since $d < N$.

3.1. Codebook Entries

Substituting (4) and the definition of $\mathrm{Q}(\bullet)$ into the MSE ε and setting the derivative with respect to $\boldsymbol{y}_i$ to zero yields

$$\boldsymbol{y}_i = \left(\mathrm{E}\left[\mathrm{S}_i(\tilde{\boldsymbol{h}})\right]\right)^{-1} \boldsymbol{V}^{\mathrm{H}}\,\mathrm{E}\left[\mathrm{S}_i(\tilde{\boldsymbol{h}})\boldsymbol{h}\right] \tag{6}$$

which is the well known centroid condition (e.g., [12]). In (2), the rank-reduced estimate $\tilde{\boldsymbol{h}}$ can be found. Thus, the MSE is

$$\varepsilon = \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{h}}) - \sum_{i=1}^{M} \frac{\mathrm{E}[\mathrm{S}_i(\tilde{\boldsymbol{h}})\boldsymbol{h}^{\mathrm{H}}]\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}\,\mathrm{E}[\mathrm{S}_i(\tilde{\boldsymbol{h}})\boldsymbol{h}]}{\mathrm{E}[\mathrm{S}_i(\tilde{\boldsymbol{h}})]}. \tag{7}$$

The channel $\boldsymbol{h}$ and the noise $\boldsymbol{\eta}$ are Gaussian. Therefore, $\boldsymbol{h}$ and $\tilde{\boldsymbol{h}}$ are jointly Gaussian due to (2) and it is not difficult to find the mean of $\boldsymbol{h}$ conditioned on $\tilde{\boldsymbol{h}}$ (e.g., [13]):

$$\mathrm{E}\left[\boldsymbol{h}|\tilde{\boldsymbol{h}}\right] = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1}\tilde{\boldsymbol{h}} \tag{8}$$

where $\boldsymbol{C}_{\tilde{\boldsymbol{h}}}$ is the covariance matrix of $\tilde{\boldsymbol{h}}$. With $\boldsymbol{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$ and $\mathrm{E}[\mathrm{S}_i(\tilde{\boldsymbol{h}})\boldsymbol{h}] = \mathrm{E}[\mathrm{S}_i(\tilde{\boldsymbol{h}})\,\mathrm{E}[\boldsymbol{h}|\tilde{\boldsymbol{h}}]]$, we obtain

$$\mathrm{E}\left[\mathrm{S}_i(\tilde{\boldsymbol{h}})\boldsymbol{h}\right] = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1/2}\,\mathrm{E}\left[\mathrm{S}_i\left(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{w}\right)\boldsymbol{w}\right].$$

Substituting this result into (7) yields for the MSE

$$\varepsilon = \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{h}}) - \operatorname{tr}\left(\boldsymbol{V}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{C}_{\tilde{h}}^{-1/2}\boldsymbol{C}_{\mathrm{Q}}\boldsymbol{C}_{\tilde{h}}^{-1/2}\boldsymbol{G}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{V}\right)$$

with $\boldsymbol{C}_{\mathrm{Q}} = \sum_{i=1}^{M} \frac{\mathrm{E}[\mathrm{S}_i(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{w})\boldsymbol{w}]\,\mathrm{E}[\mathrm{S}_i(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{w})\boldsymbol{w}^{\mathrm{H}}]}{\mathrm{E}[\mathrm{S}_i(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{w})]}$.

3.2. Estimator and Rank Reduction Basis

Due to [cf. (2)]

$$\boldsymbol{C}_{\tilde{\boldsymbol{h}}} = \boldsymbol{G}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}})\boldsymbol{G}^{\mathrm{H}}$$

we get for the estimator

$$\boldsymbol{G} = \boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{X}^{\mathrm{H}}\left(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)^{-1/2} \in \mathbb{C}^{d\times N_{\mathrm{tr}}}. \tag{9}$$

The unknown $\boldsymbol{X} \in \mathbb{C}^{N_{\mathrm{tr}}\times d}$ has orthonormal columns. Let us define $\boldsymbol{A} = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}})^{-1/2} \in \mathbb{C}^{N\times N_{\mathrm{tr}}}$. We must solve

$$\{\boldsymbol{V}_{\text{opt}}, \boldsymbol{X}_{\text{opt}}\} = \underset{\{\boldsymbol{V},\boldsymbol{X}\}}{\operatorname{argmax}} \operatorname{tr}\left(\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}\boldsymbol{C}_{\mathrm{Q}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{V}\right) \tag{10}$$

subject to $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \mathbf{I}$ and $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X} = \mathbf{I}$. From the derivative of the Lagrangian function with respect to $\boldsymbol{X}$, we conclude that

$$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}\boldsymbol{C}_{\mathrm{Q}} = \boldsymbol{C}_{\mathrm{Q}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}.$$

With the EVD $\boldsymbol{C}_{\mathrm{Q}} = \boldsymbol{U}\boldsymbol{\Xi}\boldsymbol{U}^{\mathrm{H}}$, it can be shown that $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}\boldsymbol{U}$ is diagonal, i.e., $\boldsymbol{\Psi} = \boldsymbol{Q}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}\boldsymbol{U}$ is diagonal with some unitary $\boldsymbol{Q} \in \mathbb{C}^{d\times d}$. In order to maximize the resulting objective $\operatorname{tr}(\boldsymbol{\Xi}\boldsymbol{\Psi}^2)$, we must choose the i-th column of $\boldsymbol{V}$ and $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{U}$ to be the i-th dominant left and right singular vector of $\boldsymbol{A}$, respectively. Thus, $\boldsymbol{\Psi}$ has the d dominant singular values of $\boldsymbol{A}$ on its diagonal. Note that we set $\boldsymbol{Q} = \mathbf{I}$, since the objective is independent of $\boldsymbol{Q}$. We see that the optimal basis $\boldsymbol{V}_{\text{opt}}$ contains the d dominant eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$. Interestingly, we obtained this result also for the case, where only $\boldsymbol{V}$ and $\boldsymbol{G}$ are optimized [11]. Note that $\boldsymbol{V}_{\text{opt}}$ and $\boldsymbol{W}_{\text{opt}}$ are fixed for given statistics $\boldsymbol{C}_{\boldsymbol{h}}$ and $\boldsymbol{C}_{\boldsymbol{\eta}}$. Therefore, the maximization (10) is solved by $\boldsymbol{X}_{\text{opt}} = \boldsymbol{W}_{\text{opt}}\boldsymbol{U}^{\mathrm{H}}$, i.e., the MSE minimized, where $\boldsymbol{U}$ is the modal matrix of $\boldsymbol{C}_{\mathrm{Q}}$.

Since $\boldsymbol{W}_{\text{opt}}$ contains the principal right singular vectors of $\boldsymbol{A}$, the estimator can be written as [cf. (9)]

$$\boldsymbol{G}_{\text{opt}} = \boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{U}\boldsymbol{\Psi}^{-1}\boldsymbol{V}_{\text{opt}}^{\mathrm{H}}\boldsymbol{G}_{\text{MMSE-estim}} \in \mathbb{C}^{d\times N_{\mathrm{tr}}}. \tag{11}$$

It can be seen that the conventional linear MMSE estimator

$$\boldsymbol{G}_{\text{MMSE-estim}} = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}})^{-1} \in \mathbb{C}^{N\times N_{\mathrm{tr}}}$$

is applied followed by the rank reduction with $\boldsymbol{V}_{\text{opt}}^{\mathrm{H}}$. These two stages constitute the solution of [11] for the estimator. However, when the quantizer is included in the optimization as in (5), an additional transformation with $\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{U}\boldsymbol{\Psi}^{-1}$ is appended to ensure that $\boldsymbol{C}_{\tilde{\boldsymbol{h}}} = \boldsymbol{G}_{\text{opt}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}})\boldsymbol{G}_{\text{opt}}^{\mathrm{H}}$.

3.3. Partition Cells

The MSE is the average distortion, i.e.,

$$\varepsilon = \mathrm{E}\left[\|\boldsymbol{h} - \hat{\boldsymbol{h}}_{\mathrm{Q}}\|_2^2\right] = \mathrm{E}\left[\mathrm{d}\left(\tilde{\boldsymbol{h}}, \mathrm{Q}(\tilde{\boldsymbol{h}})\right)\right]$$

with the distortion

$$\mathrm{d}\left(\tilde{\boldsymbol{h}}, \mathrm{Q}(\tilde{\boldsymbol{h}})\right) = \mathrm{E}\left[\|\boldsymbol{h} - \hat{\boldsymbol{h}}_{\mathrm{Q}}\|_2^2|\tilde{\boldsymbol{h}}\right].$$

Incorporating (4) and (8) leads to

$$\mathrm{d}\left(\tilde{\boldsymbol{h}}, \mathrm{Q}(\tilde{\boldsymbol{h}})\right) = c + \left\|\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1}\tilde{\boldsymbol{h}} - \boldsymbol{V}\,\mathrm{Q}(\tilde{\boldsymbol{h}})\right\|_2^2$$

where $c = \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1}\boldsymbol{G}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}})$ is the trace of the covariance matrix of $\boldsymbol{h}$ conditioned on $\tilde{\boldsymbol{h}}$. Due to (11), we have that $\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}_{\text{opt}}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1} = \boldsymbol{V}_{\text{opt}}\boldsymbol{\Psi}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1/2}$. Hence, we get for the distortion under the assumption that $\boldsymbol{G}_{\text{opt}}$ and $\boldsymbol{V}_{\text{opt}}$ are used

$$\mathrm{d}\left(\tilde{\boldsymbol{h}}, \mathrm{Q}(\tilde{\boldsymbol{h}})\right) = c + \left\|\boldsymbol{\Psi}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1/2}\tilde{\boldsymbol{h}} - \mathrm{Q}(\tilde{\boldsymbol{h}})\right\|_2^2. \tag{12}$$

The partition cells must be chosen to minimize $\mathrm{d}(\tilde{\boldsymbol{h}}, \mathrm{Q}(\tilde{\boldsymbol{h}}))$ for every $\tilde{\boldsymbol{h}}$, i.e., $\mathcal{R}_i = \{\boldsymbol{x} \in \mathbb{C}^d |\, \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}_i) \leq \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}_j), \forall j\}$, which is the nearest neighbor condition (e.g., [12]).

3.4. Suggested Codebook Design

The expression (12) for the distortion can be further simplified. Recall that $\mathrm{Q}(\boldsymbol{y}) = \sum_{i=1}^{M} \boldsymbol{y}_i\,\mathrm{S}_i(\boldsymbol{y})$. After incorporating (8) and (11), the i-th codebook entry reads as [cf. (6)]

$$\boldsymbol{y}_i = \left(\mathrm{E}\left[\mathrm{S}_i(\tilde{\boldsymbol{h}})\right]\right)^{-1}\boldsymbol{\Psi}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1/2}\,\mathrm{E}\left[\mathrm{S}_i(\tilde{\boldsymbol{h}})\tilde{\boldsymbol{h}}\right].$$

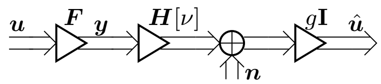

Fig. 2. System with Linear Precoding

It is useful to redefine the quantizer as

$$\mathrm{Q}(\boldsymbol{y}) = \boldsymbol{\Psi}\,\mathrm{Q}'\left(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1/2}\boldsymbol{y}\right) \tag{13}$$

with $\mathrm{Q}'(\boldsymbol{y}) = \sum_{i=1}^{M} \boldsymbol{y}'_i\,\mathrm{S}'_i(\boldsymbol{y})$, $\mathrm{S}'_i(\boldsymbol{y}) = \mathrm{S}_i(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{U}\boldsymbol{y})$, and

$$\boldsymbol{y}'_i = \boldsymbol{\Psi}^{-1}\boldsymbol{y}_i = \left(\mathrm{E}\left[\mathrm{S}'_i\left(\boldsymbol{w}\right)\right]\right)^{-1}\mathrm{E}\left[\mathrm{S}'_i\left(\boldsymbol{w}\right)\boldsymbol{w}\right]. \tag{14}$$

As above, $\boldsymbol{w} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I})$. Then, we get for the distortion

$$\mathrm{d}\left(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{U}\check{\boldsymbol{h}}, \mathrm{Q}(\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{1/2}\boldsymbol{U}\check{\boldsymbol{h}})\right) = c + \left\|\boldsymbol{\Psi}\left(\check{\boldsymbol{h}} - \mathrm{Q}'(\check{\boldsymbol{h}})\right)\right\|_2^2 \tag{15}$$

where $\check{\boldsymbol{h}} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}}^{-1/2}\tilde{\boldsymbol{h}} \in \mathbb{C}^d$. Remember that $\boldsymbol{\Psi}$ is diagonal. Thus, the distortion to be minimized for the design of $\mathrm{Q}'(\bullet)$ has a very simple structure. Additionally, $\check{\boldsymbol{h}} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I})$ which leads to the simple centroid condition in (14).

Note that we can concentrate on the design of $\mathrm{Q}'(\bullet)$, because $\mathrm{Q}(\tilde{\boldsymbol{h}}) = \boldsymbol{\Psi}\,\mathrm{Q}'(\check{\boldsymbol{h}})$ with the output $\check{\boldsymbol{h}}$ of

$$\boldsymbol{G}'_{\text{opt}} = \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{\text{opt}}^{\mathrm{H}}\boldsymbol{G}_{\text{MMSE-estim}}. \tag{16}$$

The resulting CSI of the transmitter is $\boldsymbol{V}_{\text{opt}}\boldsymbol{\Psi}\,\mathrm{Q}'(\boldsymbol{G}'_{\text{opt}}\boldsymbol{x})$. Also note that $\boldsymbol{G}'_{\text{opt}}$ and $\boldsymbol{V}_{\text{opt}}$ only depend on the channel statistics. Hence, they can be computed independently of the choice for $\mathrm{Q}'(\bullet)$.

To summarize, we obtained from the joint optimization in (5) that the received training symbols $\boldsymbol{x}$ are passed through the ordinary MMSE estimator $\boldsymbol{G}_{\text{MMSE-estim}}$, rank reduced with $\boldsymbol{V}_{\text{opt}}^{\mathrm{H}}$, and weighted with $\boldsymbol{\Psi}^{-1}$ to get uncorrelated unit-variance entries. Then, the index ℓ found by the quantizer $\mathrm{Q}'(\bullet)$ is fed back and the CSI at the transmitter is $\boldsymbol{V}_{\text{opt}}\boldsymbol{\Psi}\boldsymbol{y}'_\ell$. Note that $\boldsymbol{A} = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}})^{-1/2}$ only depends on the channel statistics $\boldsymbol{C}_{\boldsymbol{h}}$ and $\boldsymbol{C}_{\boldsymbol{\eta}}$ that change very slowly. Therefore, $\boldsymbol{V}_{\text{opt}}$ can be communicated to the transmitter with negligible overhead (with the scheme of [7] for example) and we assume a perfect knowledge of $\boldsymbol{V}_{\text{opt}}$ and $\boldsymbol{\Psi}$ at the transmitter.

Despite of the simplicity of (15), we suggest to restrict to separate scalar quantization for every entry (also real and imaginary part are split), i.e., the partition cells $\mathcal{R}'_i$ are hyperrectangles (transform coding, e.g., [12]). With this restriction, the design of $\mathrm{Q}'(\bullet)$ is independent of $\boldsymbol{\Psi}$ or any other quantity related to our system and the scalar quantizer for any of the $2d$ real-valued scalars is the MMSE optimal scalar quantizer for a real-valued Gaussian random variable with variance 0.5. Due to this property, it is not necessary to compute the parameters for $\mathrm{Q}'(\bullet)$ in real-time. Instead, they can be computed in advance (with the Lloyd algorithm, e.g., [12]) and stored. Moreover, the restriction to separate scalar quantization enables closed-form expressions for the conditional moments for the precoder design (see next section).

4. ROBUST MULTI-USER PRECODER

We discuss the principle of employing the fed back CSI for precoder design with the simple example of linear precoding as in Fig. 2 with the data signal $\boldsymbol{u} \in \mathbb{C}^K$, its estimate $\hat{\boldsymbol{u}} \in \mathbb{C}^K$, the linear precoder $\boldsymbol{F} \in \mathbb{C}^{N \times K}$, the noise $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{n}})$, and the common scalar weight g used by all receivers (e.g., [14]). The vector channels $\boldsymbol{h}_k[\nu] \in \mathbb{C}^N, k = 1, \ldots, K$, for the K users are collected in the channel matrix

$$\boldsymbol{H}[\nu] = [\boldsymbol{h}_1[\nu], \ldots, \boldsymbol{h}_K[\nu]]^{\mathrm{T}} \in \mathbb{C}^{K \times N}$$

where ν denotes the index of the time slot.

The k-th receiver feeds back the index $\ell_k[\nu] \in \{1, \ldots, M\}$ to the transmitter D slots before the transmitter uses this information for precoding at time slot $\nu + D$. For the sake of brevity, we set $\boldsymbol{l}[\nu] = [\ell_1[\nu], \ldots, \ell_K[\nu]]^{\mathrm{T}} \in \{1, \ldots, M\}^K$. To take into account the CSI errors, the precoder is found by minimizing the mean of the MSE $\varepsilon_{\text{p}} = \mathrm{E}[\|\boldsymbol{u} - \hat{\boldsymbol{u}}\|_2^2 | \boldsymbol{H}[\nu + D]]$ conditioned on the indices $\boldsymbol{l}[\nu]$:

$$\{\boldsymbol{F}_{\text{rob}}, g_{\text{rob}}\} = \underset{\{\boldsymbol{F}, g\}}{\operatorname{argmin}}\,\mathrm{E}\left[\varepsilon_{\text{p}} | \boldsymbol{l}[\nu]\right] \quad \text{s.t.: } \mathrm{E}\left[\|\boldsymbol{F}\boldsymbol{u}\|_2^2\right] = E_{\text{tx}}. \tag{17}$$

The optimal precoder can be written as

$$\boldsymbol{F}_{\text{rob}} = g_{\text{rob}}^{-1}\left(\boldsymbol{R}_{\boldsymbol{H}[\nu+D]|\boldsymbol{l}[\nu]} + \xi\mathbf{I}\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{Q}}^{\mathrm{H}} \tag{18}$$

with $\boldsymbol{R}_{\boldsymbol{H}[\nu+D]|\boldsymbol{l}[\nu]} = \sum_{k=1}^{K}\boldsymbol{R}^*_{\boldsymbol{h}_k[\nu+D]|\ell_k[\nu]}$, g_{rob} follows from $\mathrm{E}[\|\boldsymbol{F}\boldsymbol{u}\|_2^2] = E_{\text{tx}}$, $\hat{\boldsymbol{H}}_{\mathrm{Q}} = [\hat{\boldsymbol{h}}_{\mathrm{Q},1}, \ldots, \hat{\boldsymbol{h}}_{\mathrm{Q},K}]^{\mathrm{T}} \in \mathbb{C}^{K \times N}$, and $\xi = \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{n}})/E_{\text{tx}}$. The necessary conditional moments are

$$\begin{aligned}
\hat{\boldsymbol{h}}_{\mathrm{Q},k} &= \mathrm{E}\left[\boldsymbol{h}_k[\nu + D] | \ell_k[\nu]\right] = r_k\boldsymbol{V}_{\text{opt},k}\boldsymbol{\Psi}_k\boldsymbol{y}'_{\ell_k[\nu]} \\
\boldsymbol{R}_{\boldsymbol{h}_k[\nu+D]|\ell_k[\nu]} &= \mathrm{E}\left[\boldsymbol{h}_k[\nu + D]\boldsymbol{h}_k^{\mathrm{H}}[\nu + D] | \ell_k[\nu]\right] \\
&= \hat{\boldsymbol{h}}_{\mathrm{Q},k}\hat{\boldsymbol{h}}_{\mathrm{Q},k}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{h}_k} - r_k^2\boldsymbol{V}_{\text{opt},k}\boldsymbol{\Psi}_k^2\boldsymbol{\Upsilon}_k\boldsymbol{V}_{\text{opt},k}^{\mathrm{H}}
\end{aligned}$$

where r_k is the temporal correlation (resulting from the Jakes model for example). Note that the non-zero elements of the diagonal matrix $\boldsymbol{\Upsilon}_k \in \mathbb{R}_{0,+}^{d \times d}$ only depend on the properties of $\mathrm{Q}'(\bullet)$. Therefore, they can be computed in advance and stored as the parameters of $\mathrm{Q}'(\bullet)$. Due to space limitations, we omit the expression for $\boldsymbol{\Upsilon}_k$ that can be found in closed form for the special case that $\mathrm{Q}'(\bullet)$ performs separate scalar quantization.

Note that the above mentioned solution for the scalar g_{rob} is only used by the transmitter. Although not considered in the optimization (17), every receiver performs an MMSE design of its scalar weight [15]. The resulting weights will be different from g_{opt}.

The design of other types of precoding, e.g., Tomlinson Harashima precoding (THP, e.g., [16, 11]), follows the same line, i.e., we take the conditional mean of the MSE and the resulting filter expressions depend on above conditional moments.

5. SIMULATION RESULTS

We present the results for a system with $N = 4$ antennas, $K = 4$ users, and QPSK modulation. The slot duration was $T_{\text{slot}} = 6.67\,\text{ms}$ at a carrier frequency of $2\,\text{GHz}$. The Doppler frequency was 0.0123 ($v = 10\,\text{km/h}$) normalized to T_{slot} and the temporal correlations obeyed the Jakes model. The results were the mean of 5000 channel realizations and 200 symbols were transmitted per channel realization. We used a macro-cell environment with an offset of 5 degrees (see [17]). The training sequence had $N_{\text{tr}} = 16$ symbols and the dimensionality after rank reduction was $d = 2$. As suggested in Subsection 3.4, we employed separate scalar quantizers with m bits each, i.e., $2dm$ bits were fed back in total and the overall number of codebook entries of $\mathrm{Q}'(\bullet)$ was $M = 2^{2dm}$. A feedback delay of $D = 1$ slots was assumed.

In Fig. 3, the BER results are shown for non-robust and robust linear precoders based on fed back CSI, where $m = 2$ bits per real

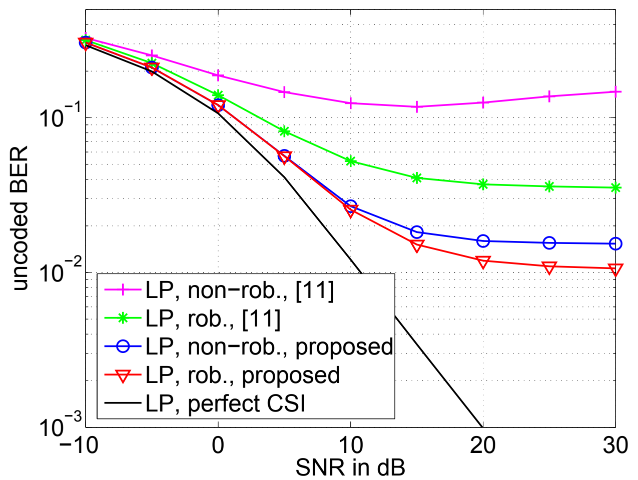

Fig. 3. BER vs. SNR for linear precoders with overall $2dm = 8$ bits feedback per user

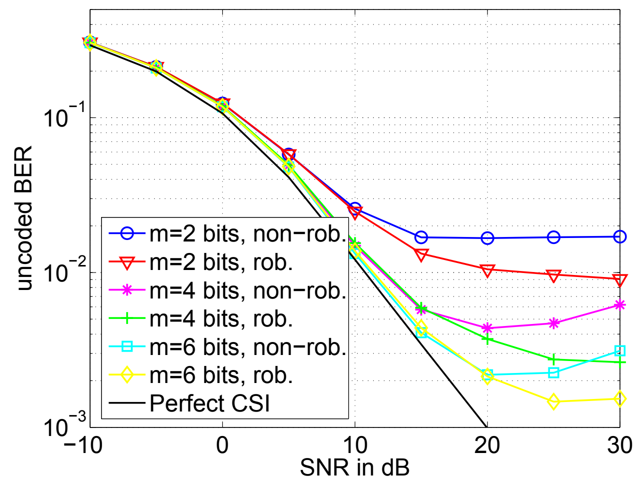

Fig. 4. BER vs. SNR for linear precoders with overall $2dm = 8, 16, 24$ bits feedback per user.

and imaginary part of every scalar coefficient are used. In total, every user feeds back $2dm = 8$ bits per slot. It can be seen in Fig. 3 that a robust design is necessary, since the robust designs achieve a better performance in terms of BER compared to the non-robust schemes due to the lower saturation at high SNR. The optimized quantizer resulting from the joint optimization proposed in this paper clearly outperforms the uniform quantizer employed in [11].

The dependence of the BER performance on the number of bits used for feedback is depicted in Fig. 4. Obviously, the performance is improved, if the number of bits is increased, because the errors due to the quantization process are smaller. For $m = 4$ bits and $m = 6$ bits, it can be seen in Fig. 4 that the BER of the non-robust schemes even increases for high SNR due to the reduced regularization term $\xi\mathbf{I}$ with increasing SNR, i.e., zero-forcing precoding is performed with erroneous CSI leading to a higher BER than for medium SNR. The respective robust schemes do not show such a behavior, since the precoder solution is regularized with $\boldsymbol{C}_{\boldsymbol{h}_k} - r_k^2 \boldsymbol{V}_{\text{opt},k} \boldsymbol{\Psi}_k^2 \boldsymbol{\Upsilon}_k \boldsymbol{V}_{\text{opt},k}^{\mathrm{H}}$ that is independent of the SNR (we assumed a constant SNR for the training symbols).

6. REFERENCES

[1] N. Jindal and A. Goldsmith, "Dirty-Paper Coding versus TDMA for MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[2] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What Is the Value of Limited Feedback for MIMO Channels?," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, October 2004.

[3] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback Gain in Multiple Antenna Systems," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 785–798, May 2002.

[4] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, October 2003.

[5] V. Lau, Y. Liu, and T. Chen, "On the Design of MIMO Block-Fading Channels With Feedback-Link Capacity Constraint," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 62–70, January 2004.

[6] W. Santipach and M. L. Honig, "Capacity of Beamforming with Limited Training and Feedback," in *Proc. ISIT 2006*, July 2006, pp. 376–380.

[7] J. C. Roh and B. D. Rao, "Efficient Feedback Methods for MIMO Channels Based on Parameterization," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 282–292, January 2007.

[8] N. Jindal, "MIMO Broadcast Channels with Finite Rate Feedback," in *Proc. Globecom 2005*, November 2005, vol. 3, pp. 1520–1524.

[9] B. Mielczarek and W. A. Krzymień, "Vector Quantization of Channel Information in Linear Multi-User MIMO Systems," in *Proc. ISSSTA 2006*, August 2006, pp. 302–306.

[10] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple Antenna Broadcast Channels With Shape Feedback and Limited Feedback," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3417–3428, July 2007.

[11] P. M. Castro, M. Joham, W. Utschick, and L. Castedo, "Optimized CSI Feedback for Robust THP Design," *Accepted for presentation at 41st Asilomar Conference on Signals, Systems and Computers*, 2007.

[12] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer, 1993.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing – Estimation Theory*, Prentice Hall, 1993.

[14] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, August 2005.

[15] J. G. Proakis, *Digital Communications*, McGraw-Hill, Inc., 1995.

[16] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Cholesky Factorization with Symmetric Permutation Applied to Detecting and Precoding Spatially Multiplexed Data Streams," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 3089–3103, June 2007.

[17] 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, "Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (release 6)," 2003.