Optimized CSI Feedback for Robust THP Design

Paula M. Castro[1], Michael Joham[2], Luis Castedo[1], and Wolfgang Utschick[2]
1) University of A Coruña, Campus de Elviña s/n, A Coruña, Spain, 15071, Email: {pcastro,luis}@udc.es
2) Munich University of Technology, Theresienstr. 90, 80290 Munich, Germany, Email: {joham,utschick}@tum.de

***Abstract*—In the *broadcast channel* (BC), the signals of the multiple users can be separated by means of precoding. For the design of precoding, *channel state information* (CSI) is necessary at the transmitter side. In many cases (e.g., for *frequency division duplex* (FDD) systems), the transmitter cannot estimate this information and the CSI has to be communicated from the receivers to the transmitter via a feedback channel that is assumed to be error-free but introduces a delay. Every user estimates the channel and reduces it to a low-dimensional representation for data compression that is possible due to the channel correlations. Before the feedback, the CSI is quantized and only the index of the codebook entry is sent to the transmitter, since the data rate of the feedback channel is limited.**

We propose a joint MSE optimization of the channel estimation and the rank reduction basis, where the quantizer is modeled as a data independent additive noise source. Based on the fed back codebook index, robust multi-user precoding schemes, viz., linear precoding and *Tomlinson Harashima precoding* (THP), are designed that clearly outperform non-robust schemes.

I. Introduction

Dirty paper coding (DPC, [1]) must be used to achieve the capacity of the vector and the MIMO broadcast channel [2]–[4]. With THP [5], [6], the high complexity of DPC can be circumvented, but THP suffers from the shaping loss, the power loss, and the modulo loss (e.g., [7]). The design of THP systems is well known for the ideal case where the CSI is perfectly known at the transmitter [8]–[10]. However, the situation is different for the case with erroneous CSI. Since no DPC has been proposed for this case, the application of the SINR criterion as in [11] is questionable for erroneous CSI. Consequently, it is inevitable to resort to an MMSE criterion together with THP for the case of partial CSI, since a THP design based on the sum MSE criterion is possible (e.g., [12], [13]). We address the case, where the CSI must be fed back to transmitter for the precoder design, since the transmitter is unable to estimate the CSI as in an FDD system, for example.

In the system proposed in this paper, we start by estimating the channel at the receivers using the observations of different pilot symbols sent from the transmit antennas. Then, the estimate is reduced to a low-dimensional representation of the channel by projecting the estimate onto a basis which only depends on the statistics of the channel. The coefficients are then quantized prior to transmission over the feedback channel which is assumed to be error-free but introduces a delay. The contribution of this paper is the joint optimization of the estimator, the basis for the rank reduction, and the inherent prediction of the estimator by minimizing the MSE. Interestingly, the resulting reduction basis is different from the Karhunen-Loève basis, i.e., the eigenbasis of the channel covariance matrix.

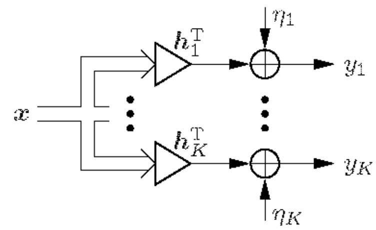

Fig. 1. Vector BC with K receivers.

Besides the design of the components of the feedback system, the joint MSE optimization also delivers the error covariance matrix which is necessary for a robust precoder design. We employ the paradigm of stochastic programming for the precoder design, i.e., the erroneous CSI is modeled to be the deterministic mean of the unknown and random channel and the expectation w.r.t. the channel of the MSE is minimized.

This paper is organized as follows. Section II describes the system model and Section III shows the proposed joint MMSE optimization. Section IV presents the robust designs obtained taking into account the proposed errors model. The simulation results are presented in Section V and some concluding remarks are made in Section VI.

II. System Model

The final goal is the design of a precoder for the broadcast channel shown in Fig. 1. We consider the downlink of a *Multiuser Multiple Input Single Output* (MU-MISO) system where a centralized transmitter equipped with N_{t} antennas communicates with K single antenna users . The output of the precoder is the transmit signal $\boldsymbol{x} \in \mathbb{C}^{N_{\mathrm{t}}}$ which propagate over the vector channel $\boldsymbol{h}_k \in \mathbb{C}^{N_{\mathrm{t}}}$ to receiver k and is perturbed by the noise $\eta_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_k}^2)$ to form the received signal y_k. For the sake of notational brevity, we use the model

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{y} = [y_1, \ldots, y_K]^{\mathrm{T}} \in \mathbb{C}^K$, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_K] \in \mathbb{C}^K$ with $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{C}_\eta)$, and $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{T}} \in \mathbb{C}^{K \times N_{\mathrm{t}}}$.

Since the CSI must be fed back to the transmitter, the k-th receiver first has to estimate the channel by means of the training channel

$$y_k(t) = \boldsymbol{h}_k^{\mathrm{T}}[n]\boldsymbol{s}_{\mathrm{tr}}(t) + \eta_k(t)$$

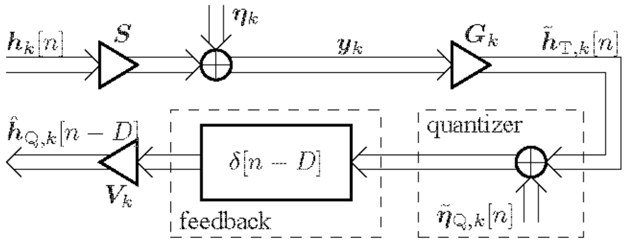

Fig. 2. Model of training channel, quantization, and feedback.

where n denotes the slot index and t the time index inside a slot. Collecting the $N_{\rm tr}$ received training symbols leads to

$$\boldsymbol{y}_k[n] = \boldsymbol{S}\boldsymbol{h}_k[n] + \boldsymbol{\eta}_k[n] \tag{2}$$

with $\boldsymbol{y}_k[n] = [y_k(1), \ldots, y_k(N_{\rm tr})]^{\rm T}$. The training symbols are comprised in $\boldsymbol{S} = [\boldsymbol{s}(1), \ldots, \boldsymbol{s}(N_{\rm tr})]^{\rm T} \in \mathbb{C}^{N_{\rm tr}\times N_{\rm t}}$ and the noise is $\boldsymbol{\eta}_k[n] = [\eta_k(1), \ldots, \eta_k(N_{\rm tr})]^{\rm T} \in \mathbb{C}^{N_{\rm tr}}$ with $\boldsymbol{\eta}_k[n] \sim \mathcal{N}_{\rm C}(0, \boldsymbol{C}_{\eta_k})$. The above received signal $\boldsymbol{y}_k[n]$ is passed through a channel estimator $\boldsymbol{G}_k \in \mathbb{C}^{d\times N_{\rm tr}}$ which also performs a rank reduction at the same time, i.e.,

$$\tilde{\boldsymbol{h}}_{{\rm T},k}[n] = \boldsymbol{G}_k\boldsymbol{y}_k[n] \in \mathbb{C}^d. \tag{3}$$

where $d \leq N_{\rm t}$ is the dimension of the low-dimensional representation prior to quantization. After estimation and rank reduction, each receiver quantizes the CSI, that is, a search is performed to find the element in the codebook closest to the channel coefficients obtained in every time slot. The real and imaginary part of every coefficient in $\tilde{\boldsymbol{h}}_{{\rm T},k}[n]$ is quantized separately with a uniform quantizer, where we assume that the input is bounded (see Section III). Then, the corresponding codebook index is fed back to the transmitter. We assume that the transmission of the index is error-free, but the feedback channel introduces a delay of D slots. Finally, the transmitter finds the quantized coefficients in the codebook. With the reduction basis, CSI is obtained at the transmitter that is used for the robust precoder design.

The channel is assumed to be zero-mean complex Gaussian and has temporal and spatial correlations. Therefore,

$$\mathrm{E}\left[\boldsymbol{h}_k[n]\boldsymbol{h}_k^{\rm H}[n]\right] = \boldsymbol{C}_{h_k} \in \mathbb{C}^{N_{\rm t}\times N_{\rm t}} \tag{4}$$

and

$$\mathrm{E}[\boldsymbol{h}_k[n]\boldsymbol{h}_k^{\rm H}[n+D]] = \mathrm{J}_0\left(2\pi\frac{f_{{\rm d},k}}{f_{\rm slot}}D\right)\boldsymbol{C}_{h_k} = r\boldsymbol{C}_{h_k} \tag{5}$$

where r is implicitly defined, J_0 denotes the zero-th order Bessel function of the first kind, $f_{{\rm d},k}$ is the *maximum* Doppler frequency of user k, and $f_{\rm slot}$ is the slot rate [16].

Each coefficient of the rank reduced channel $\tilde{\boldsymbol{h}}_{{\rm T},k}[n] \in \mathbb{C}^d$ is quantized with a uniform quantizer with step size γ. In the following, we make the simplifying assumption that the additive error introduced by the quantizer is independent of the input, i.e.,

$$\tilde{\boldsymbol{h}}_{{\rm Q},k}[n] = \mathrm{Q}\left(\tilde{\boldsymbol{h}}_{{\rm T},k}[n]\right) = \tilde{\boldsymbol{h}}_{{\rm T},k}[n] + \tilde{\boldsymbol{\eta}}_{{\rm Q},k}[n] \in \mathbb{C}^d.$$

Additionally, we assume that the quantization error is uniformly distributed inside the cell corresponding to a codebook entry. The resulting error variance is $\gamma^2/12$ for the real or imaginary part of a coefficient [17]. Assuming uncorrelated outputs, we get

$$\boldsymbol{C}_{\tilde{\eta}_{{\rm Q},k}} = \mathrm{E}\left[\tilde{\boldsymbol{\eta}}_{{\rm Q},k}[n]\tilde{\boldsymbol{\eta}}_{{\rm Q},k}^{\rm H}[n]\right] = \frac{\gamma^2}{6}\mathbf{I}_d \in \mathbb{R}^{d\times d}$$

for the covariance matrix of the quantization noise $\tilde{\boldsymbol{\eta}}_{{\rm Q},k}[n]$ of user k. Since the feedback channel introduces a delay of D slots, the CSI at the transmitter can be written as

$$\begin{aligned}\hat{\boldsymbol{h}}_{{\rm Q},k}[n] &= \boldsymbol{V}_k\tilde{\boldsymbol{h}}_{{\rm Q},k}[n-D] \\ &= \boldsymbol{V}_k\left(\tilde{\boldsymbol{h}}_{{\rm T},k}[n-D] + \tilde{\boldsymbol{\eta}}_{{\rm Q},k}[n-D]\right) \in \mathbb{C}^{N_{\rm t}}\end{aligned} \tag{6}$$

with the reduction basis $\boldsymbol{V}_k \in \mathbb{C}^{N_{\rm t}\times d}$ known to the transmitter and the quantized coefficients $\tilde{\boldsymbol{h}}_{{\rm Q},k}[n] \in \mathbb{C}^d$ for user k. For notational brevity, we introduce $\boldsymbol{\eta}_{{\rm Q},k}[n] = \boldsymbol{V}_k\tilde{\boldsymbol{\eta}}_{{\rm Q},k}[n]$ with $\boldsymbol{C}_{\eta_{{\rm Q},k}} = \mathrm{E}[\boldsymbol{\eta}_{{\rm Q},k}\boldsymbol{\eta}_{{\rm Q},k}^{\rm H}] = \boldsymbol{V}_k\boldsymbol{C}_{\tilde{\eta}_{{\rm Q},k}}\boldsymbol{V}_k^{\rm H}$. The model of training channel, quantization and feedback is depicted in Fig. 2.

III. Joint MMSE Optimization of Feedback

Combining (6), (3), and (2), the quantized estimate for $\boldsymbol{h}_k[n]$ can be expressed as

$$\hat{\boldsymbol{h}}_{{\rm Q},k}[n] = \boldsymbol{V}_k\boldsymbol{G}_k\boldsymbol{S}\boldsymbol{h}_k[n-D] + \boldsymbol{V}_k\boldsymbol{G}_k\boldsymbol{\eta}_k[n-D] + \boldsymbol{\eta}_{{\rm Q},k}[n-D]. \tag{7}$$

The channel estimation and rank reduction with $\boldsymbol{G}_k$ and the basis $\boldsymbol{V}_k$ are jointly optimized to end up with a channel estimate at the transmitter with minimum MSE:

$$\begin{aligned}\{\boldsymbol{G}_{{\rm MMSE},k}, \boldsymbol{V}_{{\rm MMSE},k}\} &= \underset{\{\boldsymbol{G}_k,\boldsymbol{V}_k\}}{\mathrm{argmin}}\,\mathrm{MSE}_k(\boldsymbol{G}_k, \boldsymbol{V}_k) \\ &\quad \text{s.t.: } \boldsymbol{V}_k^{\rm H}\boldsymbol{V}_k = \mathbf{I}_d\end{aligned} \tag{8}$$

with the MSE of user k

$$\begin{aligned}\mathrm{MSE}_k(\boldsymbol{G}_k, \boldsymbol{V}_k) &= \mathrm{E}\left[\left\|\boldsymbol{h}_k[n] - \hat{\boldsymbol{h}}_{{\rm Q},k}[n]\right\|_2^2\right] \\ &= \mathrm{tr}(\boldsymbol{C}_{h_k}) - 2\,\mathrm{Re}\left(\mathrm{tr}\left(\boldsymbol{V}_k\boldsymbol{G}_k\boldsymbol{S}\boldsymbol{C}_{h_k}\right)\right) \\ &\quad + \mathrm{tr}\left(\boldsymbol{V}_k\boldsymbol{G}_k\boldsymbol{S}\boldsymbol{C}_{h_k}\boldsymbol{S}^{\rm H}\boldsymbol{G}_k^{\rm H}\boldsymbol{V}_k^{\rm H}\right) \\ &\quad + \mathrm{tr}\left(\boldsymbol{V}_k\boldsymbol{G}_k\boldsymbol{C}_{\eta_k}\boldsymbol{G}_k^{\rm H}\boldsymbol{V}_k^{\rm H}\right) + \mathrm{tr}\left(\boldsymbol{C}_{\eta_{{\rm Q},k}}\right).\end{aligned}$$

In the optimization problem given by (8), we included the constraint that the columns of $\boldsymbol{V}_k$ are orthonormal. The filter $\boldsymbol{G}_k$ is readily found by setting the derivative of the cost function with respect to $\boldsymbol{G}_k$ to zero:

$$\boldsymbol{G}_{{\rm MMSE},k} = r\boldsymbol{V}_k^{\rm H}\boldsymbol{C}_{h_k}\boldsymbol{S}^{\rm H}\left(\boldsymbol{S}\boldsymbol{C}_{h_k}\boldsymbol{S}^{\rm H} + \boldsymbol{C}_{\eta_k}\right)^{-1}. \tag{9}$$

We see that $\boldsymbol{G}_{{\rm MMSE},k} = r\boldsymbol{V}_k^{\rm H}\boldsymbol{G}_{\text{MMSE-estim},k}$, i.e., the joint estimation and rank reduction can be decomposed into the ordinary MMSE channel estimator $\boldsymbol{G}_{\text{MMSE-estim},k}$ followed by the projection onto the basis $\boldsymbol{V}_k$. The factor r is due to the inherent prediction, since we receive the pilots in slot n and estimate the channel in slot ν. The weight r can be applied at the receiver or the transmitter.

Substituting the optimum $\boldsymbol{G}_{\text{MMSE},k}$ into the cost function of (8) yields:

$$\text{MSE}_k(\boldsymbol{G}_{\text{MMSE},k}, \boldsymbol{V}_k) = \operatorname{tr}(\boldsymbol{C}_{h_k}) + \operatorname{tr}\left(\boldsymbol{C}_{\eta_{\text{Q}},k}\right) - \operatorname{tr}\left(r^2 \boldsymbol{V}_k^{\text{H}} \boldsymbol{G}_{\text{MMSE-estim},k} \boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{V}_k\right).$$

Now, the optimization (8) only depends on $\boldsymbol{V}_k$ and can be solved using Lagrangian multipliers. One of the KKT conditions (i.e., set the derivative of the Lagrangian function with respect to $\boldsymbol{V}_k^*$ to zero) is

$$r^2 \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} \left(\boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} + \boldsymbol{C}_{\eta_k}\right)^{-1} \boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{V}_k = \boldsymbol{V}_k \boldsymbol{\Delta}_k$$

where $\boldsymbol{\Delta}_k \in \mathbb{C}^{d \times d}$ is the Lagrangian multiplier for the constraint of (8). Multiplying by $\boldsymbol{V}_k^{\text{H}}$ from the left leads to

$$r^2 \boldsymbol{V}_k^{\text{H}} \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} \left(\boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} + \boldsymbol{C}_{\eta_k}\right)^{-1} \boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{V}_k = \boldsymbol{\Delta}_k. \quad (10)$$

We see that $\boldsymbol{\Delta}_k \in \mathbb{C}^{d \times d}$ must be positive definite, i.e., its *eigenvalue decomposition* (EVD) is $\boldsymbol{\Delta}_k = \boldsymbol{Q}_k \boldsymbol{\Lambda}_k \boldsymbol{Q}_k^{\text{H}}$ with unitary $\boldsymbol{Q}_k \in \mathbb{C}^{d \times d}$ and $\boldsymbol{\Lambda}_k$ is diagonal with positive diagonal elements. Multiplying (10) with $\boldsymbol{Q}_k^{\text{H}}$ from the left and $\boldsymbol{Q}_k$ from the right, we obtain

$$r^2 \boldsymbol{Q}_k^{\text{H}} \boldsymbol{V}_k^{\text{H}} \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} \left(\boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} + \boldsymbol{C}_{\eta_k}\right)^{-1} \boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{V}_k \boldsymbol{Q}_k = \boldsymbol{\Lambda}_k.$$

Thus,

$$\boldsymbol{A}_k = r^2 \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} (\boldsymbol{S} \boldsymbol{C}_{h_k} \boldsymbol{S}^{\text{H}} + \boldsymbol{C}_{\eta_k})^{-1} \boldsymbol{S} \boldsymbol{C}_{h_k} \in \mathbb{C}^{N_t \times N_t} \quad (11)$$

is diagonalized by $\boldsymbol{V}_k \boldsymbol{Q}_k$, that is, the columns of $\boldsymbol{V}_k \boldsymbol{Q}_k$ are the eigenvectors of $\boldsymbol{A}_k$. Note that this optimal basis is different from the Karhunen-Loève basis (eigenbasis of $\boldsymbol{C}_{h_k}$) as was intuitively used in [14].

With this intermediate result for the rank reduction basis $\boldsymbol{V}_k$, the cost function of (8) is given by

$$\text{MSE}\left(\boldsymbol{G}_{\text{MMSE},k}, \boldsymbol{V}_k\right) = \operatorname{tr}(\boldsymbol{C}_{h_k}) + \operatorname{tr}\left(\boldsymbol{C}_{\eta_{\text{Q}},k}\right) - \sum_{i \in \mathbb{I}} \varphi_{k,i} \quad (12)$$

where $\mathbb{I}$ denotes the set of eigenvectors indices collected in $\boldsymbol{V}_k \boldsymbol{Q}_k$ and $\varphi_{k,i}$ is the i-th eigenvalue of $\boldsymbol{A}_k$. Clearly, $\text{MSE}_k(\boldsymbol{G}_{\text{MMSE},k}, \boldsymbol{V}_k)$ is independent of $\boldsymbol{Q}_k$. Therefore, we can set $\boldsymbol{Q}_k = \mathbf{I}_d$ and $\boldsymbol{V}_k \in \mathbb{C}^{N_t \times d}$ contains d eigenvectors of $\boldsymbol{A}_k$. Moreover, the indices $\mathbb{I}$ must be chosen such that the sum in (12) is maximized, that is, $\boldsymbol{V}_{\text{MMSE},k}$ contains the d dominant eigenvectors of $\boldsymbol{A}_k$. For this optimized basis $\boldsymbol{V}_{\text{MMSE},k}$, the MSE matrix can be written as

$$\boldsymbol{M}_k = \boldsymbol{C}_{h_k} + \boldsymbol{C}_{\eta_{\text{Q},k}} - \boldsymbol{V}_{\text{MMSE},k} \boldsymbol{\Phi}_{\text{dom},k} \boldsymbol{V}_{\text{MMSE},k}^{\text{H}}. \quad (13)$$

Here, the diagonal matrix

$$\boldsymbol{\Phi}_{\text{dom},k} = \boldsymbol{V}_{\text{MMSE},k}^{\text{H}} \boldsymbol{A}_k \boldsymbol{V}_{\text{MMSE},k} \in \mathbb{R}^{d \times d} \quad (14)$$

has the d dominant eigenvalues of $\boldsymbol{A}_k$ on its diagonal.

Due to (2), (3), and (9), the covariance matrix of $\tilde{\boldsymbol{h}}_{\text{T},k}[n]$ reads as

$$\boldsymbol{C}_{\tilde{h}_{\text{T},k}} = \text{E}\left[\tilde{\boldsymbol{h}}_{\text{T},k}[n] \tilde{\boldsymbol{h}}_{\text{T},k}^{\text{H}}[n]\right] = \boldsymbol{\Phi}_{\text{dom},k} \in \mathbb{R}^{d \times d}. \quad (15)$$

So, the entries of $\tilde{\boldsymbol{h}}_{\text{T},k}[n]$ are uncorrelated and the variance of the i-th entry of $\tilde{\boldsymbol{h}}_{\text{T},k}[n]$ is $\varphi_{k,i}$, where $\varphi_{k,i}$ is the i-th largest eigenvalue of $\boldsymbol{A}_k$.

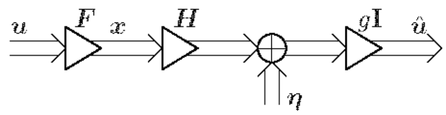

Fig. 3. System with linear precoding.

For the design of the uniform quantizer, we make the assumption that the input is bounded, i.e., the real and imaginary part of the i-th entry of $\tilde{\boldsymbol{h}}_{\text{T},k}[n]$ lie in the interval $[-\sqrt{2\varphi_{k,i}}, +\sqrt{2\varphi_{k,i}}]$. Since $\tilde{\boldsymbol{h}}_{\text{T},k}[n]$ is Gaussian, this interval selection ensures that the overload probability be less than $5\,\%$.

IV. Robust design

For the robust precoder design, we interpret the channel as a random variable and the given fed back CSI as being deterministic, i.e.,

$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \boldsymbol{\Theta}$$

where $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{h}}_{\text{Q},1}, \ldots, \hat{\boldsymbol{h}}_{\text{Q},K}]^{\text{T}} \in \mathbb{C}^{K \times N_t}$ comprises the estimates obtained from the feedback of the quantized coefficients of the rank reduced channel. The covariance of the error $\boldsymbol{\Theta}$ is the sum of the MSE matrices in (13):

$$\begin{aligned} \boldsymbol{C}_{\Theta} &= \text{E}\left[\boldsymbol{\Theta}^{\text{H}} \boldsymbol{\Theta}\right] \\ &= \sum_{k=1}^{K} \text{E}\left[\left(\boldsymbol{h}_k^*[n] - \hat{\boldsymbol{h}}_{\text{Q},k}^*[n]\right)\left(\boldsymbol{h}_k^{\text{T}}[n] - \hat{\boldsymbol{h}}_{\text{Q},k}^{\text{T}}[n]\right)\right] \\ &= \sum_{k=1}^{K} \boldsymbol{M}_k^{\text{T}}. \end{aligned} \quad (16)$$

A. Robust Linear Precoding

For the standard design of linear precoding shown in Fig. 3, the total MSE $\text{E}[\|\boldsymbol{u} - \hat{\boldsymbol{u}}\|_2^2 | \boldsymbol{H}]$ between the uncorrelated unit variance symbols $\boldsymbol{u}$ and the received signal given by $\hat{\boldsymbol{u}} = g\boldsymbol{H}\boldsymbol{F}\boldsymbol{u} + g\boldsymbol{\eta}$ is minimized under a constraint of total transmit power, i.e., $\operatorname{tr}(\boldsymbol{F}\boldsymbol{F}^{\text{H}}) = E_{\text{tx}}$ [10], where a perfect knowledge of $\boldsymbol{H}$ is assumed. We get the robust optimization by minimizing the expectation of this MSE with respect to the channel $\boldsymbol{H}$ instead, i.e.,

$$\begin{aligned} \{\boldsymbol{F}_{\text{Rlin}}, g_{\text{Rlin}}\} &= \underset{\{F,g\}}{\operatorname{argmin}}\, \text{E}\left[\text{E}\left[\|\boldsymbol{u} - \hat{\boldsymbol{u}}\|_2^2 \,|\boldsymbol{H}\right]\right] \\ \text{s.t.:}\ \text{E}\left[\|\boldsymbol{x}\|_2^2\right] &= E_{\text{tx}} \end{aligned} \quad (17)$$

The solution to this optimization problem yields to the linear precoder robust design [13]

$$\boldsymbol{F}_{\text{Rlin}} = \frac{1}{g_{\text{Rlin}}} \left(\hat{\boldsymbol{H}}^{\text{H}} \hat{\boldsymbol{H}} + \boldsymbol{C}_{\Theta} + \xi \mathbf{I}_N\right)^{-1} \hat{\boldsymbol{H}}^{\text{H}} \quad (18)$$

where $\xi = \operatorname{tr}(\boldsymbol{C}_{\eta})/E_{\text{tx}}$ and where E_{tx} is the average total transmit power. Note that the real scalar g_{Rlin} is directly obtained from the transmit power constraint. Therefore, for the robust design the solution is regularized by means of $\boldsymbol{C}_{\Theta}$, as it can be seen in (18).

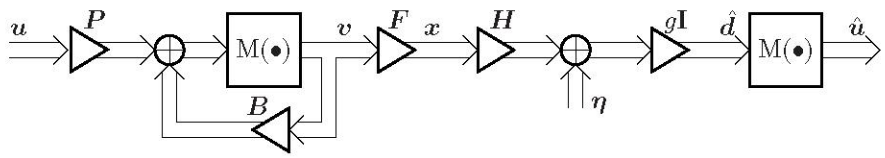

Fig. 4. System with Tomlinson Harashima precoding.

B. Robust THP

The TH precoder (see Fig. 4) consists of the permutation matrix $\boldsymbol{P} \in \{0,1\}^{K\times K}$ that depends on the precoding order, the modulo operator $\mathrm{M}(\bullet)$ with the modulo constant τ (e.g., [7]), the strictly lower triangular feedback filter $\boldsymbol{B}$, and the feedforward filter $\boldsymbol{F}$. For the THP design, the linear representation of the modulo operator is used, i.e., $\mathrm{M}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{a}$ with the perturbation signal $\boldsymbol{a} \in \tau\mathbb{R}^K + \mathrm{j}\,\tau\mathbb{R}^K$. Moving the addition of $\boldsymbol{a}$ to the input of the permutation matrix $\boldsymbol{P}$ gives the virtual desired signal

$$\boldsymbol{d} = \boldsymbol{u} + \boldsymbol{a} \in \mathbb{C}^K.$$

For the standard THP design, the MSE

$$\varepsilon(\boldsymbol{F},\boldsymbol{B},\boldsymbol{P},g) = \mathrm{E}\left[\left\|\boldsymbol{d}-\hat{\boldsymbol{d}}\right\|_2^2 |\boldsymbol{H}\right] \tag{19}$$

is minimized under the total transmit power constraint $\mathrm{E}[\|\boldsymbol{x}\|_2^2] = E_{\mathrm{tx}}$ (e.g., [10]). The assumption of perfect knowledge of $\boldsymbol{H}$ is given up for the robust design and the expected value of the MSE w.r.t. the channel is minimized instead:

$$\{\boldsymbol{F}_{\mathrm{RTHP}},\boldsymbol{B}_{\mathrm{RTHP}},\boldsymbol{P}_{\mathrm{RTHP}},g_{\mathrm{RTHP}}\} = \underset{\{F,B,P,g\}}{\operatorname{argmin}}\ \mathrm{E}[\varepsilon(\boldsymbol{F},\boldsymbol{B},\boldsymbol{P},g)]$$
$$\text{s.t.:}\quad \mathrm{E}\left[\|\boldsymbol{x}\|_2^2\right] = E_{\mathrm{tx}}. \tag{20}$$

The standard assumption for the THP design is the assumption that the output of the modulo operator at the transmitter is zero-mean and uncorrelated, i.e., $\boldsymbol{C}_v = \mathrm{E}[\boldsymbol{v}\boldsymbol{v}^{\mathrm{H}}]$ is diagonal (e.g., [7]). Therefore, the MSE $\epsilon(\boldsymbol{F},\boldsymbol{B},\boldsymbol{P},g)$ and the constraint $\mathrm{E}[\|\boldsymbol{x}\|_2^2] = E_{\mathrm{tx}}$ can be expressed in terms of $\boldsymbol{C}_v$. With steps similar to that in [10], the solution for the above robust THP optimization can be obtained. We define the matrix

$$\boldsymbol{T} = \xi^{-1}\boldsymbol{C}_{\Theta_{\mathrm{T}}} + \mathbf{I}_N$$

and the positive definite

$$\boldsymbol{\Phi} = \hat{\boldsymbol{H}}\boldsymbol{T}^{-1}\hat{\boldsymbol{H}}^{\mathrm{H}} + \xi\mathbf{I}_K \in \mathbb{C}^{K\times K}. \tag{21}$$

With the permuted Cholesky decomposition,

$$\boldsymbol{P}\boldsymbol{\Phi}^{-1}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{L} \tag{22}$$

where $\boldsymbol{L}$ is unit lower triangular and $\boldsymbol{D}$ is diagonal with positive diagonal elements, and the algorithm described in [10] and [14], the robust THP solutions are

$$\boldsymbol{F}_{\mathrm{RTHP}} = \frac{1}{g_{\mathrm{RTHP}}}\boldsymbol{T}^{-1}\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D} \tag{23}$$

$$\boldsymbol{B}_{\mathrm{RTHP}} = \mathbf{I} - \boldsymbol{L}^{-1}. \tag{24}$$

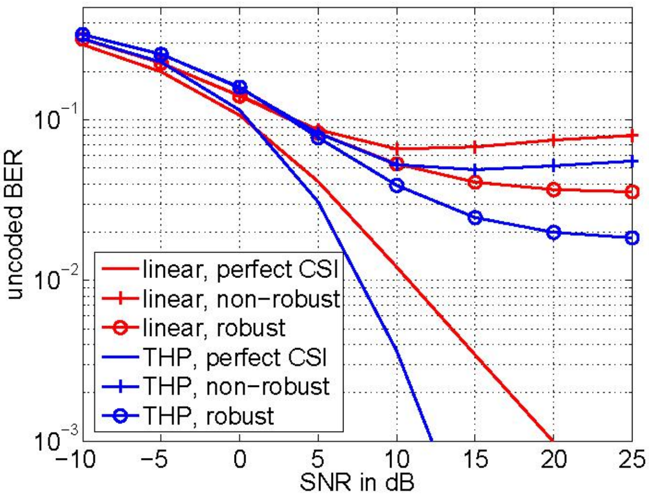

Fig. 5. BER vs. SNR of linear precoding and THP for $N_{\mathrm{t}} = K = 4$ and $\mathrm{v} = 10\,\mathrm{kmph}$. Robust design with linear precoding and THP.

The weight at the receiver results from the transmit power constraint and reads as

$$g_{\mathrm{RTHP}} = \sqrt{\frac{\mathrm{tr}\left(\boldsymbol{T}^{-2}\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}^2\boldsymbol{C}_v\boldsymbol{L}\boldsymbol{P}\hat{\boldsymbol{H}}\right)}{E_{\mathrm{tx}}}}. \tag{25}$$

C. Receiver Weights

Although the weights g_{Rlin} and g_{RTHP} result from the robust optimizations (17) and (20) of the respective precoders, we use MMSE receiver weights instead. The main reason is the phase of g_{Rlin} and g_{RTHP}, i.e., zero phase, that is only correct for CSI without errors. Otherwise, the erroneous CSI leads to a phase of the precoder combined with the channel that is different from zero. To enable a coherent detection, a phase correction by a receiver weight is necessary. For the design of the MMSE receiver weights and the precoding for the training signals, see [14].

V. Simulation Results

We show the results of some computer simulations that we carried out to validate the proposed system. The input bits are QPSK modulated. The centralized transmitter has $N_{\mathrm{t}} = 4$ transmit antennas and $K = 4$ receivers are served. The results are the mean of 10000 channel realizations and 200 symbols were transmitted per channel realization. We consider $d = 2$ coefficients for the reduced rank approximation and a delay of $D = 2$ time slots. The channel estimation is based on $N_{\mathrm{tr}} = 16$ pilot symbols and the carrier frequency is $1.5\,\mathrm{GHz}$.

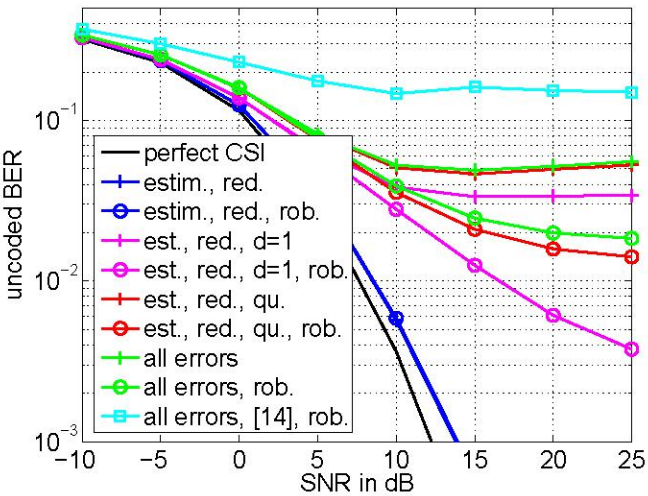

Fig. 6. BER vs. SNR of robust and non-robust THP for $N_t = K = 4$ and v = 10 kmph. Different amount of errors: estimation and rank reduction; estimation and rank reduction with $d = 1$; estimation, rank reduction, and quantization; all errors.

We employ the channel model described in [15], where an offset of 5 degrees is considered for obtaining the simulation results. A uniform codebook of only 4 entries (2 bits) is used for coding the real and imaginary part of each coefficient of the reduced rank approximation, i.e., only 8 bits are fed back from each user to the transmitter side by means of the feedback channel.

In Fig. 5, we can see that a robust design of both, linear precoding and THP, is crucial. Interestingly, the non-robust designs show an increase of BER for increasing SNR at high SNR values. This behavior can be explained by the reduction of the regularization term $\xi\mathbf{I}$ with increasing SNR and the convergence of the MMSE designs to the zero-forcing precoders that are highly non-robust to CSI errors. The robust designs do not show such a behavior and saturate at a lower BER than their non-robust counterparts. Additionally, it can be clearly seen in Fig. 5 that non-robust and robust THP outperforms the linear counterpart for a BER below 10^{-1}. The disadvantage of THP for low SNR is mainly due to the power loss of THP (e.g., [7]).

Fig. 6 plots the BER performance for the THP scheme when different types of errors are introduced. We observe a considerable improvement in performance when the new approach for joint optimization of the CSI feedback is employed. Note the results for the case where only estimation and rank reduction errors are simulated. With $d = 2$, the non-robust and robust schemes are only slightly worse than the precoder based on error-free CSI. Contrary, if only one coefficient ($d = 1$) per user is fed back to the transmitter, robust THP with optimized feedback clearly outperforms the non-robust THP. However, as the results for $d = 1$ are clearly inferior to that for $d = 2$, we can follow that the channel effectively is of rank two and we set $d = 2$. Additionally, we can see in Fig. 6 that the proposed scheme clearly outperforms the robust THP of [14], where no optimization of the feedback was performed and the Karhunen-Loève basis was used.

Thanks to the used compression techniques, the feedback channel overhead is strongly reduced, and with the proposed robust design for THP, we are capable of adapting the precoder parameters to channel variations with a limited feedback channel.

VI. Conclusions

We have demonstrated in this paper how the joint optimization of the channel estimation and the rank reduction basis leads to a robust precoding design that clearly outperforms non-robust designs for high SNR scenarios.

References

[1] M. Costa, "Writing on Dirty Paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[2] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Transactions on Information Theory*, 49(8):1912–1921, August 2003.

[3] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, 49(10):2658–2668, October 2003.

[4] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Transactions on Information Theory*, 52(9):3936–3964, September 2006.

[5] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Letters*, vol. 7, no. 5/6, pp. 138–139, March 1971.

[6] H. Harashima and H. Miyakawa, "Matched-Transmission Technique for Channels With Intersymbol Interference," *IEEE Transactions on Communications*, vol. 20, no. 4, pp. 774–780, August 1972.

[7] R. F. H. Fischer. *Precoding and Signal Shaping for Digital Transmission.* John Wiley & Sons, 2002.

[8] G. Ginis and J. M. Cioffi, "A Multi-user Precoding Scheme achieving Crosstalk Cancellation with Application to DSL Systems," *Proc. Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 1627–1631, October 2000.

[9] C. Windpassinger, T. Vencel, and R. F. H. Fischer, "Precoding and Loading for BLAST-like Systems," *Proc. ICC'03*, vol. 5, pp. 3061–3065, May 2003.

[10] 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. 5, no. 6, pp. 3089–3103, June 2007.

[11] M. Schubert and H. Boche. "Iterative Multiuser Uplink and Downlink Beamforming Under SINR Constraints," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2324–2334, July 2005.

[12] A. P. Liavas, "Tomlinson-Harashima Precoding With Partial Channel Knowledge," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 5–9, January 2005.

[13] F. A. Dietrich and W. Utschick, "Robust Tomlinson-Harashima Precoding," *Proc. of the 16th IEEE Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.

[14] P. M. Castro, M. Joham, L. Castedo, and W. Utschick, "Robust Precoding for Multi-User MISO Systems with Limited-Feedback Channels," *Proc. of IEEE/ITG Workshop on Smart Antennas*, Vienna, 2007.

[15] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 6)*, 2003.

[16] T. Croft, P. Dent, and G. E. Bottomley. *Jakes fading model revisited.* IEE Electronics Letter, vol. 29, 1993.

[17] A. Gersho and R. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.