Robust Precoding with Bayesian Error Modeling for Limited Feedback MU-MISO Systems

Paula M. Castro, *Member, IEEE*, Michael Joham, *Member, IEEE*, Luis Castedo, *Member, IEEE*, and Wolfgang Utschick, *Senior Member, IEEE*

Abstract

We consider the robust precoder design for *Multi-User Multiple Input Single Output* (MU-MISO) systems, where the *Channel State Information* (CSI) is fed back from the single antenna receivers to the centralized transmitter equipped with multiple antennas. We propose to compress the feedback data by projecting the channel estimates onto a vector basis, known at the receivers and the transmitter, and quantizing the resulting coefficients. The channel estimator and the basis for the rank reduction are jointly optimized by minimizing the *Mean Square Error* (MSE). Expressions for the conditional mean and the conditional covariance of the channel are derived which are necessary for the robust precoder design. These expressions take into account the following sources of error: channel estimation, truncation for rank reduction, quantization, and feedback channel delay. Three well-known precoder types, namely *Linear Precoding* (LP), *Vector Precoding* (VP), and *Tomlinson-Harashima Precoding* (THP), are designed based on the expectation of the MSE conditioned on the fed–back CSI. Our results show that robust precoding based on fed–back CSI clearly outperforms conventional precoding designs which do not take into account the errors in the CSI. Additionally, we observe that a robust design is especially crucial for systems employing non-linear precoding with scarce feedback rate.

Index Terms

Feedback channel, Bayesian approach, imperfect CSI, robust precoding.

I. Introduction

We consider a MU-MISO system, i.e. a multiple antennas transmitter and several single-antenna receivers, since the centralized access point in a cellular system admits more complexity and cost than the mobiles. A MU-MISO system is a prominent example of a vector broadcast channel [1]. Recently, it has been shown that the *Dirty Paper*

Michael Joham and Wolfgang Utschick are with the Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany, e-mail:{joham,utschick}@tum.de.

Paula M. Castro and Luis Castedo are with the Department of Electronics and Systems, University of A Coruña, A Coruña, 15071 Spain, e-mail: {pcastro,luis}@udc.es.

Coding (DPC) [2] signaling techniques designed according to *Signal-to-Interference-plus-Noise Ratio* (SINR) criteria are able to approach the sum capacity of a broadcast channel [3], [4]. These contributions, however, only consider the ideal case where the CSI at the transmitter is perfectly known, similar to [5]–[7]. In the more practical case, where only an estimate of the CSI is available at the transmitter, the capacity region of the vector broadcast channel has not been found yet. First, the application of DPC is questionable, since it is unclear up to now how DPC can be used with erroneous CSI. Second, it is unclear how to systematically include the uncertainties in the SINR criterion (see the discussion in [8] and the attempt in [9] for the case of statistical CSI).

As shown in [10], the SINR and the MSE achievable regions for MU-MISO systems are tightly related. Additionally, *minimum MSE* (MMSE) allows for a robust precoder design by considering a conditional expectation of the cost function [11]–[15]. Hence, we concentrate on the MMSE precoder design. Based on the MMSE design for linear precoding as in [16], [17], for THP in [7], [18], and for VP in [19], we develop robust linear precoding, robust THP, and robust VP, where we take the expectation of the MSE conditioned on the available CSI.

Most of the work on precoding with erroneous CSI was motivated by a *Time Division Duplex* (TDD) setup, where the transmitter can estimate the CSI during the transmission in the opposite direction [13], [14]. This approach however is difficult due to the necessity of very good calibration [20]. Contrarily, we focus on the more difficult case, where the CSI is obtained by the receivers and fed back to the transmitter. In this case, calibration errors are estimated as being part of the CSI and, therefore, no special problems arise from calibration. Additionally, the feedback of CSI enables precoding in *Frequency Division Duplex* (FDD) systems, where the transmitter is unable to obtain the CSI during reception, because the channels are not reciprocal.

Since the data rate of the feedback channels is limited [21], the CSI must be compressed to ensure that the tight scheduling constraints are satisfied. Moreover, when the CSI is not perfectly known by the receiver, it is a matter of discussion what kind of information has to be sent from the receiver to the transmitter and the way of recovering it at the transmitter side.

In the system proposed in this paper, we start by estimating the channel at the receivers using the observations of pilot symbols sent from all the transmit antennas. This enables the receivers to estimate their respective vector channels. Then, we reduce the estimates to a low-dimensional representation by projecting them onto a basis which only depends on the channel statistics. We assume that the channel statistics are also known to the transmitter. The coefficients are quantized prior to transmission over the feedback channel which also introduces a delay. For simplicity, we use a uniform quantizer.

The estimator and the basis for the rank reduction are jointly optimized by minimizing the MSE, where the optimization is formulated such that the estimator additionally performs the rank reduction (see [22]). The resulting estimator can be decomposed into an ordinary MMSE estimator followed by a projection on the basis. Interestingly, the resulting basis is different from that of the Karhunen-Loève expansion [23], i.e. the eigenbasis of the channel covariance matrix.

In order to properly design robust precoders, it is necessary to obtain an adequate statistical characterization of the errors in the fed–back CSI. The following sources of error are considered: channel estimation, truncation

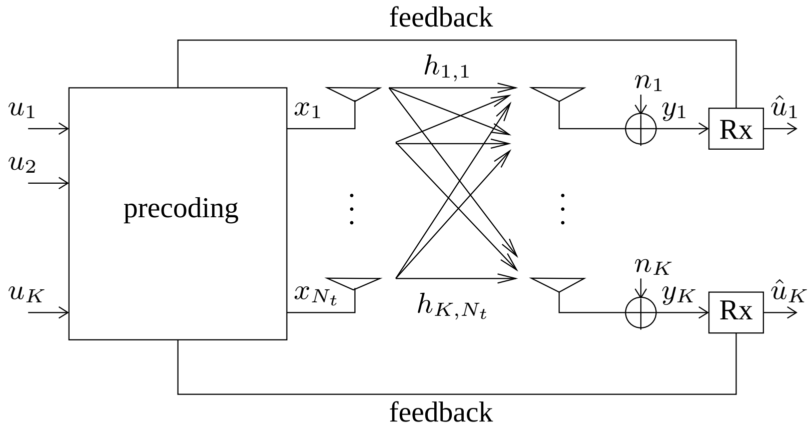

Fig. 1. Multi-User System with Precoding over Flat MISO Channels.

(rank reduction), quantization, and feedback channel delay. Channel estimation and truncation errors are Gaussian and their analysis follows a conventional MSE approach [24]. Since the delayed channel versions fed back to the transmitter after estimation and truncation are also Gaussian, we can easily obtain their statistical properties. On the other hand, quantization errors are often assumed to be uniformly distributed [25] which makes the analysis of their impact on the imperfect CSI difficult. Nevertheless, we obtain an expression for the probability density function of the channel vector according to a Bayesian framework, i.e. conditioned on the delayed, truncated, and quantized channel estimate. The resulting expression for this conditional channel PDF enables us to find closed-form expressions for the robust precoders. Compared to our previous work in [26], where we assumed uncorrelated and Gaussian distributed quantization errors, the exact analysis presented herein enables the design of robust precoding schemes with considerably better performance.

This paper is organized as follows. Section II describes the signal model of a MU-MISO system with correlated channels. In Section III, the Bayesian model for the CSI error sources is developed and Section IV contains the robust precoder design. The MSE receivers are derived and the used training data are discussed in Section V. Computer simulations are presented in Section VI. Finally, concluding remarks are given in Section VII.

Vectors and matrices are denoted by lower case bold and capital bold letters, respectively. The $K \times K$ identity matrix is denoted by $\mathbf{I}_K$ and $\mathbf{0}_K$ is a K-dimensional zero vector. We use $\mathrm{E}[\bullet]$, $\Re(\bullet)$, $\Im(\bullet)$, $\mathrm{tr}(\bullet)$, $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, $\det(\bullet)$, $\otimes$, $*$, and $\|\bullet\|_2$ for expectation, real and imaginary part of the argument, trace of a matrix, complex conjugation, transposition, conjugate transposition, determinant of a matrix, Kronecker product, convolution, and Euclidean norm, respectively. The i-th element of a vector $\boldsymbol{x}$ is x_i. With $f_{\mathrm{G}}\left(\boldsymbol{x}, \boldsymbol{\mu}_x, \boldsymbol{C}_x\right)$, we refer to a circularly symmetric complex Gaussian distribution of $\boldsymbol{x} \in \mathbb{C}^m$ with the mean $\boldsymbol{\mu}_{\boldsymbol{x}} \in \mathbb{C}^m$ and the covariance matrix $\boldsymbol{C}_{\boldsymbol{x}} \in \mathbb{C}^{m \times m}$, i.e.

$$f_{\mathrm{G}}\left(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{C}_{\boldsymbol{x}}\right) = \frac{\exp\left(-\left(\boldsymbol{x} - \boldsymbol{\mu}_x\right)^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{x}}^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}_x\right)\right)}{\pi^m \det(\boldsymbol{C}_{\boldsymbol{x}})}.$$

II. MU-MISO Signal Model

Let us consider a MU-MISO system with N transmit antennas and K single antenna receivers as depicted in Fig. 1. The precoder generates the transmit signal $\boldsymbol{x}$ from all data symbols $\{u_1, \ldots, u_K\}$ belonging to the different users $1, \ldots, K$. The signal x_ℓ from transmit antenna ℓ propagates over the channel with the coefficient $h_{k,\ell}$ to the k-th receiver, superimposes with the signals of the other transmit antennas, and is perturbed by the additive white Gaussian noise η_k with variance σ_η^2, i.e.

$$y_k = \sum_{\ell=1}^{N} h_{k,\ell} x_\ell + \eta_k = \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{x} + \eta_k \tag{1}$$

where $\boldsymbol{h}_k = [h_{k,1}, \ldots, h_{k,N}]^{\mathrm{T}} \in \mathbb{C}^N$ represents the flat fading vector channel corresponding to the k-th user and $\boldsymbol{x} = [x_1, \ldots, x_N]^{\mathrm{T}} \in \mathbb{C}^N$ is the transmit signal. The transmit signal $\boldsymbol{x}$ must satisfy an average total transmit power constraint, i.e. $\mathrm{E}[\|\boldsymbol{x}\|_2^2] = E_{\mathrm{tx}}$. Combining (1) for $k = 1, \ldots, K$, we get

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\eta} \tag{2}$$

with the $K \times N$ channel matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{T}}$, the received vector $\boldsymbol{y} = [y_1, \ldots, y_K]^{\mathrm{T}} \in \mathbb{C}^K$, and the noise vector $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_K]^{\mathrm{T}} \in \mathbb{C}^K$ with $f_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = f_{\mathrm{G}}(\boldsymbol{\eta}, \mathbf{0}_K, \boldsymbol{C}_{\boldsymbol{\eta}})$.

We model the k-th user's channel vector $\boldsymbol{h}_k$ as a zero-mean circularly symmetric complex Gaussian random vector with covariance matrix $\boldsymbol{C}_{\boldsymbol{h}_k}$, i.e.

$$f_{\boldsymbol{h}_k}(\boldsymbol{h}_k) = f_{\mathrm{G}}\left(\boldsymbol{h}_k, \mathbf{0}_N, \boldsymbol{C}_{\boldsymbol{h}_k}\right). \tag{3}$$

Additionally, the channel has temporal correlations according to the Jakes model [27], [28] described in [29]. Thus, the channel vector for user k in the time slot n can be written as

$$\boldsymbol{h}_k[n] = \boldsymbol{C}_{\boldsymbol{h}_k}^{1/2} \boldsymbol{h}_{\mathrm{w},k}[n] \tag{4}$$

with the stationary white Gaussian vector process $\boldsymbol{h}_{\mathrm{w},k}[n]$ (with elements of unit variance) and $(\bullet)^{1/2}$ represents matrix root operation computed via the Cholesky decomposition for example. The covariance matrix $\boldsymbol{C}_{\boldsymbol{h}_k}$ results from the model in [30].

Notice that, according to our model, the channel $\boldsymbol{h}_k[n]$ is stationary because $\boldsymbol{h}_{\mathrm{w},k}[n]$ is stationary. Realistic channels are often non-stationary, e.g., either the location of the receiver or the scenario geometry can change. Thus, the channel covariance matrix has to be tracked in real situations. However, since the channel statistics change very slowly compared to the channel itself, it is realistic to assume that they remain constant and are perfectly known at both the receiver and the transmitter. Nevertheless, the feedback rate is limited and the feedback of the channel realizations for the precoder design must thus be optimized.

III. Bayesian Model for Imperfect CSI

In systems with CSI feedback, the CSI errors result not only from the estimation but also from the compression (projection onto a basis of lower dimensionality), the quantization, and the delay due to the feedback. In the

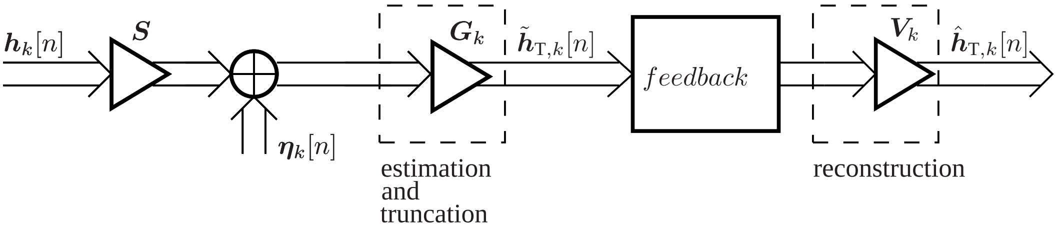

Fig. 2. Feedback Design including Channel Estimation and Truncation Errors.

following subsections, we will describe and model the sources of error. Our goal is to find the PDF of the channel vector conditioned on the fed–back coefficients that is the basis of our robust precoder design.

A. Channel Estimation and Rank Reduction Error

Fig. 2 depicts the feedback model based on CSI MSE which takes into account channel estimation and rank reduction errors as described in this subsection. We use linear estimators at the receivers based on N_{tr} pilot symbols per time slot n to enable the channel vector estimation for the k-th user. The vector comprising the N_{tr} received symbols for the k–th user reads as

$$\boldsymbol{y}_k[n] = \boldsymbol{S}\boldsymbol{h}_k[n] + \boldsymbol{\eta}_k[n] \tag{5}$$

with $\boldsymbol{S} \in \mathbb{C}^{N_{\mathrm{tr}} \times N}$ containing the training symbols [31], [32] and $\boldsymbol{\eta}_k[n] \in \mathbb{C}^{N_{\mathrm{tr}}}$ is the zero-mean additive Gaussian noise with the covariance matrix $\boldsymbol{C}_{\boldsymbol{\eta}_k} = \mathrm{E}[\boldsymbol{\eta}_k[n]\boldsymbol{\eta}_k^{\mathrm{H}}[n]]$. The above received signal $\boldsymbol{y}_k[n]$ is passed through a channel estimator $\boldsymbol{G}_k \in \mathbb{C}^{d \times N_{\mathrm{tr}}}$ which also performs a rank reduction at the same time, i.e.

$$\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n] = \boldsymbol{G}_k \boldsymbol{y}_k[n] \in \mathbb{C}^d. \tag{6}$$

Here, $d \leq N$ denotes the dimensionality of the rank reduction. The rank reduced channel can be written as

$$\hat{\boldsymbol{h}}_{\mathrm{T},k}[n] = \boldsymbol{V}_k \tilde{\boldsymbol{h}}_{\mathrm{T},k}[n] \in \mathbb{C}^N \tag{7}$$

with the orthonormal reduction basis $\boldsymbol{V}_k \in \mathbb{C}^{N \times d}$ and the reduced rank coefficients $\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n] \in \mathbb{C}^d$ for user k. Combining (7), (6), and (5), the truncated estimate for $\boldsymbol{h}_k[n]$ can be expressed as

$$\hat{\boldsymbol{h}}_{\mathrm{T},k}[n] = \boldsymbol{V}_k \boldsymbol{G}_k \boldsymbol{S}\boldsymbol{h}_k[n] + \boldsymbol{V}_k \boldsymbol{G}_k \boldsymbol{\eta}_k[n]. \tag{8}$$

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

$$\{\boldsymbol{G}_{\mathrm{MMSE},k}, \boldsymbol{V}_{\mathrm{MMSE},k}\} = \underset{\{\boldsymbol{G}_k, \boldsymbol{V}_k\}}{\operatorname{argmin}} \mathrm{MSE}_k\left(\boldsymbol{G}_k, \boldsymbol{V}_k\right)$$
$$\text{s.t.: } \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{V}_k = \mathbf{I}_d \tag{9}$$

with the MSE of user k

$$\mathrm{MSE}_k(\boldsymbol{G}_k, \boldsymbol{V}_k) = \mathrm{E}\left[\left\|\boldsymbol{h}_k[n] - \hat{\boldsymbol{h}}_{\mathrm{T},k}[n]\right\|_2^2\right].$$

In the above optimization (9), we included the constraint that the columns of $\boldsymbol{V}_k$ are orthonormal. The filter $\boldsymbol{G}_k$ is readily found by setting to zero the derivative of the cost function with respect to $\boldsymbol{G}_k$, i.e.

$$\begin{aligned}\boldsymbol{G}_{\text{MMSE},k} &= \boldsymbol{V}_k^{\text{H}} \boldsymbol{C}_{\boldsymbol{h}_k} \boldsymbol{S}^{\text{H}} \left(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{S}^{\text{H}} + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1} \\ &= \boldsymbol{V}_k^{\text{H}} \boldsymbol{G}_{\text{MMSE-estim},k}\end{aligned} \tag{10}$$

where it can be seen that $\boldsymbol{G}_{\text{MMSE},k}$ is decomposed into the ordinary MMSE channel estimator $\boldsymbol{G}_{\text{MMSE-estim},k}$, and the term due to the projection onto the basis, $\boldsymbol{V}_k^{\text{H}}$. Substituting the optimum $\boldsymbol{G}_{\text{MMSE},k}$ into the cost function of (9) yields

$$\text{MSE}_k(\boldsymbol{G}_{\text{MMSE},k}, \boldsymbol{V}_k) = \text{tr}\left(\boldsymbol{C}_{\boldsymbol{h}_k}\right) - \text{tr}\left(\boldsymbol{V}_k^{\text{H}} \boldsymbol{W}_k \boldsymbol{V}_k\right) \tag{11}$$

with the $N \times N$ non-negative definite matrix

$$\boldsymbol{W}_k = \boldsymbol{C}_{\boldsymbol{h}_k} \boldsymbol{S}^{\text{H}} \left(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{S}^{\text{H}} + \boldsymbol{C}_{\boldsymbol{\eta}_k}\right)^{-1} \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_k}. \tag{12}$$

Now, the optimization (9) only depends on $\boldsymbol{V}_k$ and can be solved using Lagrangian multipliers. One of the KKT conditions is

$$\boldsymbol{W}_k \boldsymbol{V}_k = \boldsymbol{V}_k \boldsymbol{\Delta}_k \tag{13}$$

where $\boldsymbol{\Delta}_k \in \mathbb{C}^{d \times d}$ is the Lagrangian multiplier for the constraint of (9). After multiplying by $\boldsymbol{V}_k^{\text{H}}$ from the left, we see that $\boldsymbol{\Delta}_k$ is non-negative definite. Thus, the EVD of $\boldsymbol{\Delta}_k$ is $\boldsymbol{\Delta}_k = \boldsymbol{Q}_k \boldsymbol{\Phi}_k \boldsymbol{Q}_k^{\text{H}}$ with the unitary matrix $\boldsymbol{Q}_k$ and the non-negative diagonal matrix $\boldsymbol{\Phi}_k$. Then, (13) can be rewritten as

$$\boldsymbol{W}_k \boldsymbol{V}_k' = \boldsymbol{V}_k' \boldsymbol{\Phi}_k \tag{14}$$

where $\boldsymbol{V}_k' = \boldsymbol{V}_k \boldsymbol{Q}_k$ is a matrix with orthonormal columns as $\boldsymbol{V}_k$, since $\boldsymbol{Q}_k$ is unitary. Thus, we see that $\boldsymbol{\Delta}_k$ in (13) can be replaced by a diagonal matrix $\boldsymbol{\Phi}_k$ without loss of generality. After multiplying (13) by $\boldsymbol{V}_k^{\text{H}}$ from the left, we have that

$$\boldsymbol{V}_k^{\text{H}} \boldsymbol{W}_k \boldsymbol{V}_k = \boldsymbol{\Phi}_k \tag{15}$$

i.e. $\boldsymbol{V}_k$ diagonalizes $\boldsymbol{W}_k$. Thus, the columns of $\boldsymbol{V}_k$ are eigenvectors of $\boldsymbol{W}_k$ and not of $\boldsymbol{C}_{\boldsymbol{h}_k}$ as we intuitively used in [26]. With this intermediate result for the rank reduction basis $\boldsymbol{V}_k$, the cost function of (9) is given by

$$\text{MSE}_k\left(\boldsymbol{G}_{\text{MMSE},k}, \boldsymbol{V}_k\right) = \text{tr}\left(\boldsymbol{C}_{\boldsymbol{h}_k}\right) - \sum_{i \in \mathbb{I}} \varphi_{k,i}$$

where $\mathbb{I}$ denotes the set of indices of eigenvectors collected in $\boldsymbol{V}_k$ and $\varphi_{k,i}$ is the i-th eigenvalue of $\boldsymbol{W}_k$. Consequently, the indices $\mathbb{I}$ must be chosen such that the sum is maximized, that is, $\boldsymbol{V}_{\text{MMSE},k} \in \mathbb{C}^{N \times d}$ contains the d dominant eigenvectors of $\boldsymbol{W}_k$. Note that no errors due to rank reduction are added to the channel estimation if all the eigenvectors are employed.

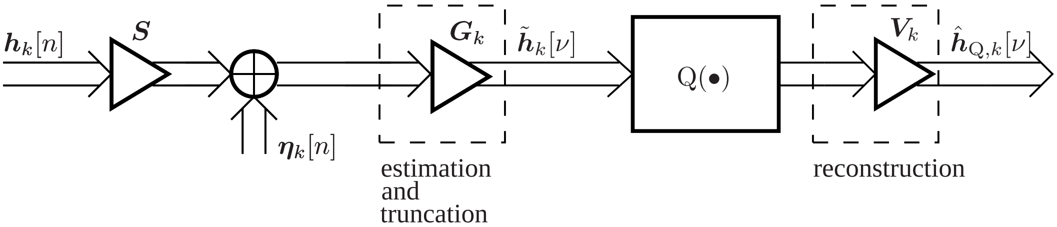

Fig. 3. Feedback Design including Channel Estimation, Truncation, Feedback Delay, and Quantization Errors.

B. Feedback Delay Error

The transmission over the feedback channel introduces a delay of $\nu - n$ slots, where the precoder is designed during the time slot ν and the most recent channel estimate was obtained during the time slot n. Fig. 3 depicts the feedback model which takes into account all the errors considered along this work, i.e. channel estimation, truncation, feedback delay, and quantization errors.

From (6) and (5), our model for the estimation and truncation is written as

$$\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n] = \boldsymbol{G}_{\mathrm{MMSE},k}\boldsymbol{S}\boldsymbol{h}_k[n] + \boldsymbol{G}_{\mathrm{MMSE},k}\boldsymbol{\eta}_k[n] \tag{16}$$

whose covariance matrix $\boldsymbol{\Phi}_{\mathrm{dom},k} = \mathrm{E}\left[\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n]\tilde{\boldsymbol{h}}_{\mathrm{T},k}^{H}[n]\right] = \boldsymbol{V}_{\mathrm{MMSE},k}^{\mathrm{H}}\boldsymbol{W}_k\boldsymbol{V}_{\mathrm{MMSE},k}$ is diagonal with $\boldsymbol{W}_k$ from (12). When the transmitter processes multiple feedback vectors, the available channel information is given by

$$\tilde{\boldsymbol{h}}_k[\nu] = \left[\tilde{\boldsymbol{h}}_{\mathrm{T},k}[\nu - D_1]^{\mathrm{T}}, \cdots, \tilde{\boldsymbol{h}}_{\mathrm{T},k}[\nu - D_L]^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{dL} \tag{17}$$

where L is the number of delayed vectors processed at the transmitter, and $D_i, i = 1, \ldots, L$, is the delay expressed as the number of slots for the i-th vector.

With the properties of $\boldsymbol{h}_k$ and $\boldsymbol{h}_{\mathrm{w},k}$ described in Section II, we have that

$$\mathrm{E}\left[\boldsymbol{h}_k[n]\boldsymbol{h}_k^{\mathrm{H}}[\nu]\right] = \mathrm{J}_0\left(\alpha_k\left(\nu - n\right)\right)\boldsymbol{C}_{\boldsymbol{h}_k} \tag{18}$$

where J_0 denotes the zero–th order Bessel function of the first kind and $\alpha_k = 2\pi\frac{f_{\mathrm{d},k}}{f_{\mathrm{slot}}}$, where $f_{\mathrm{d},k}$ is the *maximum* Doppler frequency of user k and f_{slot} is the slot rate [28]. Consequently, considering (16), (18), and (10), we find

$$\mathrm{E}\left[\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n]\tilde{\boldsymbol{h}}_{\mathrm{T},k}^{\mathrm{H}}[\nu]\right] = \begin{cases}\mathrm{J}_0\left(\alpha_k\left(\nu - n\right)\right)\boldsymbol{\Psi}_k & n \neq \nu, \\ \boldsymbol{\Phi}_{\mathrm{dom},k} & n = \nu\end{cases} \tag{19}$$

with $\boldsymbol{\Psi}_k = \boldsymbol{\Phi}_{\mathrm{dom},k}\boldsymbol{V}_{\mathrm{MMSE},k}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k}^{-1}\boldsymbol{V}_{\mathrm{MMSE},k}\boldsymbol{\Phi}_{\mathrm{dom},k}$. Therefore, we have for the processed feedback information

$$f_{\tilde{\boldsymbol{h}}_k[\nu]}\left(\tilde{\boldsymbol{h}}_k[\nu]\right) = f_{\mathrm{G}}\left(\tilde{\boldsymbol{h}}_k[\nu], \mathbf{0}_{dL}, \boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}\right) \tag{20}$$

where we introduced

$$\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k} = \boldsymbol{C}_{\mathrm{temp}} \otimes \boldsymbol{\Psi}_k + \mathbf{I}_L \otimes \boldsymbol{\Phi}_{\mathrm{dom},k}$$

and the matrix $\boldsymbol{C}_{\text{temp}}$ comprises the temporal correlations and its i-th element in the j-th column is

$$[\boldsymbol{C}_{\text{temp},k}]_{i,j} = \begin{cases} \mathrm{J}_0\left(\alpha_k\left(D_i - D_j\right)\right) & j \neq i, \\ 0 & j = i. \end{cases} \tag{21}$$

With (17) and (16), we obtain for the crosscorrelation between the channel and the feedback information

$$\begin{aligned} \mathrm{E}\left[\tilde{\boldsymbol{h}}_k[\nu]\boldsymbol{h}_k^{\mathrm{H}}[\nu]\right] &= \begin{bmatrix} \boldsymbol{G}_{\text{MMSE},k}\boldsymbol{S}\,\mathrm{E}\left[\boldsymbol{h}_k[\nu-D_1]\boldsymbol{h}_k^{\mathrm{H}}[\nu]\right] \\ \vdots \\ \boldsymbol{G}_{\text{MMSE},k}\boldsymbol{S}\,\mathrm{E}\left[\boldsymbol{h}_k[\nu-D_L]\boldsymbol{h}_k^{\mathrm{H}}[\nu]\right] \end{bmatrix} \\ &= \boldsymbol{\beta}_k \otimes \boldsymbol{\Phi}_{\text{dom},k}\boldsymbol{V}_{\text{MMSE},k}^{\mathrm{H}} \end{aligned} \tag{22}$$

with $\boldsymbol{\beta}_k = \left[\mathrm{J}_0\left(\alpha_k\left(D_1\right)\right), \ldots, \mathrm{J}_0\left(\alpha_k\left(D_L\right)\right)\right]^{\mathrm{T}} \in \mathbb{R}^L$.

According to the Theorem 10.2 of [24], given the zero-mean joint Gaussian vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ with covariance matrices $\boldsymbol{C}_{\boldsymbol{x}}$ and $\boldsymbol{C}_{\boldsymbol{y}}$, respectively, and the crosscovariance matrix $\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{x}} = \mathrm{E}[\boldsymbol{y}\boldsymbol{x}^{\mathrm{H}}]$, the mean and the covariance matrix describing $f_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x}) = f_{\mathrm{G}}(\boldsymbol{y}, \boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{x}}, \boldsymbol{C}_{\boldsymbol{y}|\boldsymbol{x}})$ are

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{x}} &= \mathrm{E}[\boldsymbol{y}|\boldsymbol{x}] = \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}{}^{-1}\boldsymbol{x} \\ \boldsymbol{C}_{\boldsymbol{y}|\boldsymbol{x}} &= \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}|\boldsymbol{x}] - \boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{x}}^{\mathrm{H}} = \boldsymbol{C}_{\boldsymbol{y}} - \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}{}^{-1}\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}} \end{aligned} \tag{23}$$

respectively. In our case, $\boldsymbol{x} = \boldsymbol{h}_k[\nu]$ and $\boldsymbol{y} = \tilde{\boldsymbol{h}}_k[\nu]$. Hence, applying (23) yields for the conditional mean

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{h}}_k[\nu]|\boldsymbol{h}_k[\nu]} = \mathrm{E}\left[\tilde{\boldsymbol{h}}_k[\nu]|\boldsymbol{h}_k[\nu]\right] = \boldsymbol{A}_k\boldsymbol{h}_k[\nu] \tag{24}$$

where we defined

$$\boldsymbol{A}_k = \boldsymbol{\beta}_k \otimes \boldsymbol{\Phi}_{\text{dom},k}\boldsymbol{V}_{\text{MMSE},k}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k}^{-1} \tag{25}$$

and for the conditional covariance matrix

$$\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k[\nu]|\boldsymbol{h}_k[\nu]} = \left(\boldsymbol{C}_{\text{temp},k} - \boldsymbol{\beta}_k\boldsymbol{\beta}_k^{\mathrm{T}}\right) \otimes \boldsymbol{\Psi}_k + \mathbf{I}_L \otimes \boldsymbol{\Phi}_{\text{dom},k}. \tag{26}$$

In the following, we will denote $\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k[\nu]|\boldsymbol{h}_k[\nu]}$ as $\boldsymbol{B}_k$ for brevity. From (24) and (26), we obtain that the conditional PDF

$$f_{\tilde{\boldsymbol{h}}_k[\nu]|\boldsymbol{h}_k[\nu]}\left(\tilde{\boldsymbol{h}}_k[\nu]|\boldsymbol{h}_k[\nu]\right) = f_{\mathrm{G}}\left(\tilde{\boldsymbol{h}}_k[\nu], \boldsymbol{A}_k\boldsymbol{h}_k[n], \boldsymbol{B}_k\right). \tag{27}$$

C. Quantization Error

Under the assumption that the channel statistics do not depend on time, the modal matrix obtained from the eigenvalue decomposition of the matrix $\boldsymbol{W}_k$ (see (12)) is also constant over time. With this assumption, only the coefficients of the reduced rank approximation have to be sent from the receiver to the transmitter due to the fast variations of the channel (so referred to as *short–term* variations).

We employ the uniform quantizer which is the most common of the scalar quantizers and whose principle is rather simple (see [25]). Furthermore, we make the simplifying assumption that the input is bounded, i.e. we assume that the real and imaginary parts of every entry of $\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n]$ lie in the the interval $[-\sqrt{2\varphi_{k,i}}, +\sqrt{2\varphi_{k,i}}]$, where $\varphi_{k,i}$ denotes the i-th principal eigenvalue of $\boldsymbol{W}_k$ (see (12)). The overload region has a very low probability (less than $5\,\%$) of containing an input sample. Thus, we choose representants between $-\sqrt{2\varphi_{k,i}}$ and $+\sqrt{2\varphi_{k,i}}$ to construct a codebook that is stored at both transmitter and receiver. After transmission, every receiver performs a search to find for each coefficients component (real and imaginary part) the element in the corresponding codebook that is closest. Then, the respective codebook index is fed back to the transmitter. Finally, the transmitter simply looks into its codebook and builds the precoder parameters from the selected codeword [33].

As shown in Fig. 3, we have the following error model,

$$\tilde{\boldsymbol{h}}_{\mathrm{Q},k}[\nu] = \tilde{\boldsymbol{h}}_k[\nu] + \tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}[\nu] \tag{28}$$

where $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}[\nu] \in \mathbb{C}^{dL}$ comprises the representants (codebook entries) and $\tilde{\boldsymbol{h}}_k[\nu]$ is given by (17). Note that $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}[\nu]$ is the quantized version of $\tilde{\boldsymbol{h}}_k[\nu]$ and it thus contains L fed–back vectors. The quantization noise of user k can be written as

$$\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}[\nu] = \left[\tilde{\boldsymbol{n}}_{\mathrm{Q},k}^{\mathrm{T}}[\nu - D_1], \cdots, \tilde{\boldsymbol{n}}_{\mathrm{Q},k}^{\mathrm{T}}[\nu - D_L]\right]^{\mathrm{T}} \in \mathbb{C}^{dL}. \tag{29}$$

The i-th coefficient of the rank reduced channel estimate $\tilde{\boldsymbol{h}}_{\mathrm{T},k}[n] \in \mathbb{C}^d$ is quantized with a uniform quantizer with step size γ_i (the choice of γ_i depends on the number of bits for the feedback of the i-th coefficient or, equivalently, the number of entries in the i-th codebook). Under the assumption that a high resolution quantizer is used, we have a uniform distribution over the cell corresponding to a codebook entry [25]. Additionally, the errors $\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}[\nu]$ are assumed to be mutually independent and independent with the truncated channel estimates $\tilde{\boldsymbol{h}}_k[\nu]$.

For the robust precoder design, we must find the conditional probability density function $f_{\boldsymbol{h}_k[\nu]|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}[\nu]}(\boldsymbol{h}_k[\nu]|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}[\nu])$, since the transmitter only knows $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}[\nu]$, but the cost function depends on $\boldsymbol{h}_k[\nu]$. From now on, we will drop the index ν for notational brevity. According to Bayesian theory, we have that

$$f_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}\left(\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \varepsilon\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k\right) f_{\boldsymbol{h}_k}(\boldsymbol{h}_k) \tag{30}$$

with $\varepsilon(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}) = 1/f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$. From (28), we see that $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$ is the sum of $\tilde{\boldsymbol{h}}_k$ and $\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}$. The PDF $f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ is thus the convolution of the PDFs of $\tilde{\boldsymbol{h}}_k$ and $\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}$. The PDF of $\tilde{\boldsymbol{h}}_k$ can be found in (20) and, as mentioned above, $\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}$ is uniformly distributed over the hyperrectangle $\mathbb{S}$ around the origin with sidelengths γ_i, i.e.

$$\begin{aligned}\mathbb{S} = \Big\{ & \boldsymbol{x} = \left[\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_L^{\mathrm{T}}\right]^{\mathrm{T}},\ \forall i : \boldsymbol{x}_i \in \mathbb{C}^d, \\ & \forall j : |\Re(x_{i,j})| \leq \gamma_j/2, |\Im(x_{i,j})| \leq \gamma_j/2 \Big\}.\end{aligned}$$

So, we obtain for the PDF of $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$

$$f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \frac{1}{\prod_{i=1}^d \gamma_i^{2L}} \int_{\mathbb{S}} f_{\tilde{\boldsymbol{h}}_k}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} - \tilde{\boldsymbol{\eta}}_k\right) \mathrm{d}\tilde{\boldsymbol{\eta}}_k.$$

For the special case that $L = 1$ (i.e. $\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k} = \boldsymbol{\Phi}_{\mathrm{dom},k}$), we get (see Appendix A)

$$f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} = \boldsymbol{\omega}_k + \mathrm{j}\,\boldsymbol{\chi}_k\right) = f_{\boldsymbol{\omega}_k}(\boldsymbol{\omega}_k) f_{\boldsymbol{\omega}_k}(\boldsymbol{\chi}_k) \tag{31}$$

with $\boldsymbol{\omega}_k = \Re(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$, $\boldsymbol{\chi}_k = \Im(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$,

$$f_{\boldsymbol{\omega}_k}(\boldsymbol{a}) = \prod_{i=1}^{d} \frac{1}{\gamma_i} \left(Q\left(\frac{-\gamma_i - 2a_i}{\sqrt{2\varphi_{k,i}}}\right) - Q\left(\frac{\gamma_i - 2a_i}{\sqrt{2\varphi_{k,i}}}\right)\right) \tag{32}$$

and $Q(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\,\mathrm{d}t$.

The PDF of $\boldsymbol{h}_k$ is $f_{\boldsymbol{h}_k}(\boldsymbol{h}_k) = f_{\mathrm{G}}(\boldsymbol{h}_k, \boldsymbol{0}_N, \boldsymbol{C}_{\boldsymbol{h}_k})$ (see (3)). The last missing term of (30) is the convolution of the PDFs of two random variables [cf. (28)]:

$$f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k\right) = \left(f_{\tilde{\boldsymbol{h}}_k|\boldsymbol{h}_k} * f_{\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}|\boldsymbol{h}_k}\right)\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k\right).$$

The Gaussian conditional PDF $f_{\tilde{\boldsymbol{h}}_k|\boldsymbol{h}_k}(\tilde{\boldsymbol{h}}_k|\boldsymbol{h}_k)$ can be found in (27) and

$$f_{\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}|\boldsymbol{h}_k}(\boldsymbol{\omega} + \mathrm{j}\,\boldsymbol{\chi}|\boldsymbol{h}_k) = f_{\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}}(\boldsymbol{\omega} + \mathrm{j}\,\boldsymbol{\chi}) = \prod_{i=1}^{dL} f_{\mathrm{U}}(\omega_i) f_{\mathrm{U}}(\chi_i)$$

with $\boldsymbol{\omega}, \boldsymbol{\chi} \in \mathbb{R}^{dL}$, $f_{\mathrm{U}}(a) = 1/\gamma$ for $|a| \leq \gamma/2$ and $f_{\mathrm{U}}(a) = 0$ else. Note that we used the assumption that the quantization noise is independent of the quantity which is quantized and, hence, $\tilde{\boldsymbol{\eta}}_{\mathrm{Q},k}$ is independent of $\boldsymbol{h}_k$. Consequently, we have that

$$f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}|\boldsymbol{h}_k\right) = \int_{\mathbb{S}} \frac{f_{\mathrm{G}}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} - \boldsymbol{w}, \boldsymbol{A}_k\boldsymbol{h}_k, \boldsymbol{B}_k\right)}{\prod_{i=1}^{d} \gamma_i^{2L}}\,\mathrm{d}\boldsymbol{w}.$$

Substituting this result and (27) into (30) leads to

$$\begin{aligned} f_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}\left(\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = &\frac{\varepsilon(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})}{\prod_{i=1}^{d}\gamma_i^{2L}} \int_{\mathbb{S}} f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} - \boldsymbol{w}\right) \\ &\times f_{\mathrm{G}}\left(\boldsymbol{h}_k, \boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}^{-1}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} - \boldsymbol{w}\right), \boldsymbol{C}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_k}\right)\mathrm{d}\boldsymbol{w} \end{aligned}$$

with $\boldsymbol{C}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_k} = (\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{B}_k^{-1}\boldsymbol{A}_k + \boldsymbol{C}_{\boldsymbol{h}_k}^{-1})^{-1}$. With this result, the conditional mean $\boldsymbol{\mu}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}} = \mathrm{E}[\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}]$ and the conditional correlation matrix $\boldsymbol{R}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_k} = \mathrm{E}[\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}|\tilde{\boldsymbol{h}}_k]$ can be respectively written as

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}} &= \boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}^{-1}\boldsymbol{m}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) \\ \boldsymbol{R}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}} &= \boldsymbol{C}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_k} + \boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}^{-1}\boldsymbol{M}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k}^{-1}\boldsymbol{A}_k\boldsymbol{C}_{\boldsymbol{h}_k}. \end{aligned} \tag{33}$$

Here, we introduced

$$\begin{aligned} \boldsymbol{m}_k(\boldsymbol{a}) &= \frac{\varepsilon(\boldsymbol{a})}{\prod_{i=1}^{d}\gamma_i^{2L}} \int_{\mathbb{S}} (\boldsymbol{a} - \boldsymbol{w})\, f_{\tilde{\boldsymbol{h}}_k}(\boldsymbol{a} - \boldsymbol{w})\,\mathrm{d}\boldsymbol{w} \\ \boldsymbol{M}_k(\boldsymbol{a}) &= \frac{\varepsilon(\boldsymbol{a})}{\prod_{i=1}^{d}\gamma_i^{2L}} \int_{\mathbb{S}} (\boldsymbol{a} - \boldsymbol{w})\,(\boldsymbol{a} - \boldsymbol{w})^{\mathrm{H}}\, f_{\tilde{\boldsymbol{h}}_k}(\boldsymbol{a} - \boldsymbol{w})\,\mathrm{d}\boldsymbol{w} \end{aligned}$$

which, for $L = 1$ ($\boldsymbol{C}_{\tilde{\boldsymbol{h}}_k} = \boldsymbol{\Phi}_{\mathrm{dom},k}$), can be expressed as

$$\boldsymbol{m}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \boldsymbol{\mu}_k(\boldsymbol{\omega}_k) + \mathrm{j}\,\boldsymbol{\mu}_k(\boldsymbol{\chi}_k) \tag{34}$$

$$\boldsymbol{M}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \boldsymbol{m}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)\boldsymbol{m}_k^{\mathrm{H}}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) + \boldsymbol{\Sigma}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) \tag{35}$$

with $\boldsymbol{\mu}_k(\boldsymbol{\omega}_k) = [\mu_{k,1}(\omega_{k,1}), \ldots, \mu_{k,d}(\omega_{k,d})]^{\mathrm{T}}$. In Appendix B, it is shown that

$$\mu_{k,i}(a) = \frac{\sqrt{\varphi_{k,i}}}{2\sqrt{\pi}} \frac{\exp\left(-\frac{(2a-\gamma_i)^2}{4\varphi_{k,i}}\right) - \exp\left(-\frac{(2a+\gamma_i)^2}{4\varphi_{k,i}}\right)}{Q\left(\frac{-\gamma_i-2a}{\sqrt{2\varphi_{k,i}}}\right) - Q\left(\frac{\gamma_i-2a}{\sqrt{2\varphi_{k,i}}}\right)}. \tag{36}$$

The second term of $\boldsymbol{M}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)$ is diagonal, i.e.

$$\boldsymbol{\Sigma}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \mathrm{diag}\left(\sigma_{k,1}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right), \ldots, \sigma_{k,d}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)\right)$$

whose i-th diagonal element can be expressed as (see Appendix C)

$$\sigma_{k,i}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \tau_{k,i}(\omega_{k,i}) + \tau_{k,i}(\chi_{k,i}) \tag{37}$$

with

$$\begin{aligned}\tau_{k,i}(a) = &\frac{\varphi_{k,i}}{2} - \mu_{k,i}^2(a) + \frac{\sqrt{\varphi_{k,i}}}{4\sqrt{\pi}} \\ &\times \frac{(2a-\gamma_i)\exp\left(-\frac{(2a-\gamma_i)^2}{4\varphi_{k,i}}\right) - (2a+\gamma_i)\exp\left(-\frac{(2a+\gamma_i)^2}{4\varphi_{k,i}}\right)}{Q\left(\frac{-\gamma_i-2a}{\sqrt{2\varphi_{k,i}}}\right) - Q\left(\frac{\gamma_i-2a}{\sqrt{2\varphi_{k,i}}}\right)}.\end{aligned}$$

The above results enable us to compute the conditional covariance matrix (for $L=1$)

$$\boldsymbol{C}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}} = \boldsymbol{C}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_k} + \mathrm{J}_0^2(\alpha_k D_1)\boldsymbol{V}_{\text{MMSE},k}\boldsymbol{\Sigma}_k\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)\boldsymbol{V}_{\text{MMSE},k}^{\mathrm{H}}. \tag{38}$$

The first term comes from the erroneous knowledge about $\boldsymbol{h}_k$, if we had $\tilde{\boldsymbol{h}}_k$. But since we only have $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$ available, the variance of the error is increased by the second term.

IV. Robust Precoder Design

The uncertain knowledge about the channel at the transmitter is modeled by the conditioned probability density function obtained in the previous section. Therefore, we consider the channel as being random but we can exploit the statistical dependence between the channel and the fed–back information. This goal can be achieved by extending the classical precoder optimizations with a mean with respect to the channel conditioned on the fed–back information. A similar problem was considered for THP design in [12], [14] and for linear precoder design in [15], where the reciprocity of the channel in a TDD system was exploited. We will see in the sequel how the conditional mean introduces a regularization of the solution which makes it more robust to CSI errors.

When taking the conditional mean of the MSE, we always encounter the conditional mean of the channel and the conditional mean of the channel Gram which can be respectively written as (see (33) and (38))

$$\mathrm{E}\left[\boldsymbol{H}\,\middle|\,\tilde{\boldsymbol{H}}_{\mathrm{Q}}\right] = \left[\boldsymbol{\mu}_{\boldsymbol{h}_1|\tilde{\boldsymbol{h}}_{\mathrm{Q},1}}, \ldots, \boldsymbol{\mu}_{\boldsymbol{h}_K|\tilde{\boldsymbol{h}}_{\mathrm{Q},K}}\right]^{\mathrm{T}} = \hat{\boldsymbol{H}}_{\mathrm{Q}} \tag{39}$$

$$\mathrm{E}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\,\middle|\,\tilde{\boldsymbol{H}}_{\mathrm{Q}}\right] = \hat{\boldsymbol{H}}_{\mathrm{Q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{Q}} + \boldsymbol{C}_{\text{error}} \tag{40}$$

where $\tilde{\boldsymbol{H}} = [\tilde{\boldsymbol{h}}_{\mathrm{Q},1}, \tilde{\boldsymbol{h}}_{\mathrm{Q},2}, \ldots, \tilde{\boldsymbol{h}}_{\mathrm{Q},k}]$ and $\boldsymbol{C}_{\text{error}} = \sum_{k=1}^{K} \boldsymbol{C}_{\boldsymbol{h}_k|\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}^{\mathrm{T}}$. Notice that for MMSE designs, no other channel conditional moments are necessary.

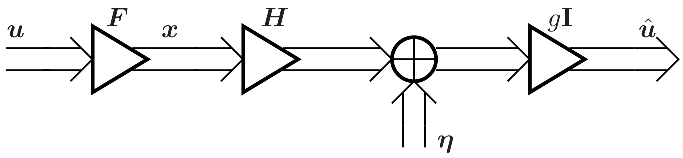

Fig. 4. MU–MISO System with Linear Precoding.

To simplify the presentation and to ensure closed form solutions for the precoder design, we only consider the case where the receivers use a common real weight [14]. Nevertheless, as explained in Section V, the receivers apply an MMSE weight afterwards to correct the phase and the amplitude of the received signals prior to detection. Therefore, there is a slight mismatch between the receivers model for the precoder design and the final system receivers.

A. Linear Precoding (LP)

In the case of linear precoding [17], the transmit signal $\boldsymbol{x} \in \mathbb{C}^N$ results from a linear transformation of the uncorrelated unit variance symbols $\boldsymbol{u} \in \mathbb{C}^K$, i.e. $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{u}$ as in Fig. 4. For robust linear MMSE precoding, $\boldsymbol{F} \in \mathbb{C}^{N\times K}$ together with the common weight $g \in \mathbb{R}$ minimizes the conditional mean of the MSE under a transmit power constraint

$$\begin{aligned} \{\boldsymbol{F}_{\text{Rlin}}, g_{\text{Rlin}}\} &= \underset{\{\boldsymbol{F},g\}}{\operatorname{argmin}}\, \mathrm{E}\left[\varepsilon_{\text{lin}}(\boldsymbol{F},g)\,\middle|\,\tilde{\boldsymbol{H}}_{\text{Q}}\right] \\ \text{s.t.:}\quad & \mathrm{E}\left[\|\boldsymbol{x}\|_2^2\right] = E_{\text{tx}} \end{aligned} \tag{41}$$

where the MSE is defined as

$$\varepsilon_{\text{lin}}(\boldsymbol{F},g) = \mathrm{E}\left[\,\|\boldsymbol{u} - g\boldsymbol{H}\boldsymbol{F}\boldsymbol{u} - g\boldsymbol{\eta}\|_2^2\,\middle|\,\boldsymbol{H}\right].$$

With Lagrangian multipliers, the above optimization (41) can be solved with similar steps as for the standard MMSE precoder in [17]. Substituting (39) and (40) into the solution, we get

$$\boldsymbol{F}_{\text{Rlin}} = \frac{1}{g_{\text{Rlin}}}\left(\hat{\boldsymbol{H}}_{\text{Q}}^{\text{H}}\hat{\boldsymbol{H}}_{\text{Q}} + \boldsymbol{C}_{\text{error}} + \xi\mathbf{I}_N\right)^{-1}\hat{\boldsymbol{H}}_{\text{Q}}^{\text{H}} \tag{42}$$

$$= \frac{1}{g_{\text{Rlin}}}\boldsymbol{T}^{-1}\hat{\boldsymbol{H}}_{\text{Q}}^{\text{H}}\boldsymbol{\Phi}^{-1} \tag{43}$$

where we obtained the second line with the matrix inversion lemma [23]. Additionally, we defined $\xi = \operatorname{tr}(\boldsymbol{C}_{\boldsymbol{\eta}})/E_{\text{tx}}$, $\boldsymbol{T} = \xi^{-1}\boldsymbol{C}_{\text{error}} + \mathbf{I}_N$, and the positive definite matrix

$$\boldsymbol{\Phi} = \hat{\boldsymbol{H}}_{\text{Q}}\boldsymbol{T}^{-1}\hat{\boldsymbol{H}}_{\text{Q}}^{\text{H}} + \xi\mathbf{I}_K \in \mathbb{C}^{K\times K}. \tag{44}$$

Note from (42) that the solution is regularized with $\boldsymbol{C}_{\text{error}}$. With the transmit power constraint, i.e. $\operatorname{tr}(\boldsymbol{F}_{\text{Rlin}}\boldsymbol{F}_{\text{Rlin}}^{\text{H}}) = E_{\text{tx}}$, the real scalar g_{Rlin} is readily found.

We see from (43) and (44) that the structure and the amount of error have a strong influence on the final precoder. For very small error, i.e. $\boldsymbol{C}_{\text{error}} \to \mathbf{0}$, we obtain the classical linear MMSE precoder as in [17] and for very large error, we get $\boldsymbol{\Phi} \to \xi\mathbf{I}_K$ and $\boldsymbol{F}_{\text{Rlin}}$ acts like a matched filter which is inherently the most robust precoder.

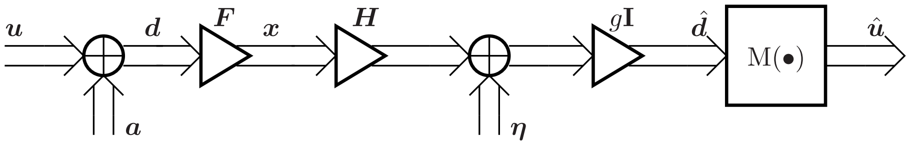

Fig. 5. MU–MISO System with Vector Precoding.

B. Vector Precoding (VP)

When the receivers are equipped with modulo operators as in Fig. 5, the transmitter has the freedom to add a perturbation signal $\boldsymbol{a} \in \tau\mathbb{Z}^K + \mathrm{j}\,\tau\mathbb{Z}^K$ to the data signal $\boldsymbol{u}$ prior to the linear transformation with the filter $\boldsymbol{F}$ (see Fig. 5). Here, τ denotes the constant associated with the modulo operator $\mathrm{M}(\bullet)$. The freedom of adding $\boldsymbol{a}$ is optimally exploited by VP [19], [34], whose robust MMSE optimization reads as [cf. [19]]

$$\begin{aligned}\{\boldsymbol{x}_{\mathrm{RVP}}(m), g_{\mathrm{RVP}}, \boldsymbol{a}_{\mathrm{RVP}}(m)\} &= \underset{\{\boldsymbol{x}(m),g,\boldsymbol{a}(m)\}}{\operatorname{argmin}} \ \mathrm{E}\left[\varepsilon_{\mathrm{VP}}(\boldsymbol{x}(m), g, \boldsymbol{a}(m)) \,\middle|\, \tilde{\boldsymbol{H}}_{\mathrm{Q}}\right] \\ \text{s.t.:} \quad & \frac{1}{N_{\mathrm{B}}}\sum_{m=1}^{N_{\mathrm{B}}} \|\boldsymbol{x}(m)\|_2^2 = E_{\mathrm{tx}}.\end{aligned} \tag{45}$$

Here, m is the symbol index in a block of N_{B} symbols. The MSE for VP is the variance of the difference between the signal $\boldsymbol{d}(m)$ and the modulo operator input $\hat{\boldsymbol{d}}(m)$

$$\varepsilon_{\mathrm{VP}} = \mathrm{E}\left[\|\boldsymbol{d}(m) - g\boldsymbol{H}\boldsymbol{x}(m) - g\boldsymbol{\eta}(m)\|_2^2 \,\middle|\, \boldsymbol{H}, \boldsymbol{u}(m)\right].$$

Note that the expectation is neither taken with respect to the symbols $\boldsymbol{u}(m)$ nor the transmit signal $\boldsymbol{x}(m)$, since the transmitter has full knowledge of the data signal $\boldsymbol{u}(m)$. With similar steps as in [19], it can be shown that the transmit signal for robust MMSE VP is

$$\boldsymbol{x}_{\mathrm{RVP}}(m) = \frac{1}{g_{\mathrm{RVP}}}\boldsymbol{T}^{-1}\hat{\boldsymbol{H}}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{\Phi}^{-1}\left(\boldsymbol{u}(m) + \boldsymbol{a}_{\mathrm{RVP}}(m)\right). \tag{46}$$

The real scalar g_{RVP} follows from the transmit power constraint and the perturbation signal can be found via following closest point search in a lattice

$$\boldsymbol{a}_{\mathrm{RVP}}(m) = \underset{\boldsymbol{a}(m)\in\tau\mathbb{Z}^K+\mathrm{j}\,\tau\mathbb{Z}^K}{\operatorname{argmin}} \ \boldsymbol{d}(m)^{\mathrm{H}}\boldsymbol{\Phi}^{-1}\boldsymbol{d}(m) \tag{47}$$

with $\boldsymbol{d}(m) = \boldsymbol{u}(m) + \boldsymbol{a}(m)$. For small errors, the above search becomes the standard MMSE VP rule to compute the perturbation vector as in [19]. For large errors, $\boldsymbol{\Phi}^{-1}$ is a weighted identity matrix leading to $\boldsymbol{a}_{\mathrm{RVP}}(m) = \boldsymbol{0}$, i.e. robust VP converges to linear precoding.

C. Tomlinson–Harashima Precoding (THP)

To avoid the high complexity of the robust VP rule in (47), we can employ THP as depicted in Fig. 6. For the THP design, the standard assumption is that the output covariance matrix of the modulo operator at the transmitter,

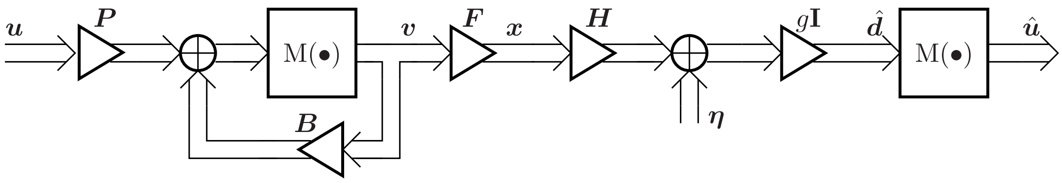

Fig. 6. MU–MISO System with Tomlinson–Harashima Precoding.

$\boldsymbol{C}_{\boldsymbol{v}} = \mathrm{E}[\boldsymbol{v}\boldsymbol{v}^{\mathrm{H}}]$, is diagonal [7]. Additionally, the feedback filter $\boldsymbol{B}$ must be strictly lower triangular to avoid a delay-free loop. The optimization for robust THP can be expressed as

$$\{\boldsymbol{F}_{\mathrm{RTHP}}, \boldsymbol{B}_{\mathrm{RTHP}}, g_{\mathrm{RTHP}}\} = \underset{\{\boldsymbol{F},\boldsymbol{B},g\}}{\operatorname{argmin}} \mathrm{E}\left[\varepsilon_{\mathrm{THP}}(\boldsymbol{F}, \boldsymbol{B}, g) \middle| \tilde{\boldsymbol{H}}_{\mathrm{Q}}\right]$$
$$\text{s.t.:} \quad \mathrm{E}\left[\|\boldsymbol{x}\|_2^2\right] = E_{\mathrm{tx}} \quad \text{and} \tag{48}$$
$$\boldsymbol{B} \text{ is strictly lower triangular}$$

with the MSE for THP [cf. [7]] being

$$\varepsilon_{\mathrm{THP}}(\boldsymbol{F}, \boldsymbol{B}, g) = \mathrm{E}\left[\|(\mathbf{I} - \boldsymbol{B})\,\boldsymbol{v} - g\boldsymbol{H}\boldsymbol{F}\boldsymbol{v} - g\boldsymbol{\eta}\|_2^2 \middle| \boldsymbol{H}\right]$$

where $(\mathbf{I} - \boldsymbol{B})\boldsymbol{v}$ is the desired value for the inputs of the modulo operators at the receivers that is the sum of the permuted symbols $\boldsymbol{P}\boldsymbol{u}$ and the perturbation added by the modulo operator at the transmitter (see [7]). With the symmetrically permuted Cholesky factorization

$$\boldsymbol{P}\boldsymbol{\Phi}^{-1}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{L} \tag{49}$$

where $\boldsymbol{P}$ is a permutation matrix, $\boldsymbol{L}$ is unit lower triangular, and $\boldsymbol{D}$ is non-negative diagonal, the solution to (48) can be concisely written as

$$\boldsymbol{F}_{\mathrm{RTHP}} = \frac{1}{g_{\mathrm{RTHP}}}\boldsymbol{T}^{-1}\hat{\boldsymbol{H}}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D} \tag{50}$$

$$\boldsymbol{B}_{\mathrm{RTHP}} = \mathbf{I} - \boldsymbol{L}^{-1} \tag{51}$$

and g_{RTHP} follows from $\mathrm{tr}(\boldsymbol{F}_{\mathrm{RTHP}}\boldsymbol{C}_{\boldsymbol{v}}\boldsymbol{F}_{\mathrm{RTHP}}^{\mathrm{H}}) = E_{\mathrm{tx}}$. For the algorithm to compute the symmetrically permuted factorization (49), we refer to [7].

V. MMSE RECEIVER AND TRAINING SYMBOLS

As shown in [14], phase correction at the receivers is particularly crucial for a system with erroneous CSI at the transmitter. However, contrary to [14], in this work we do not restrict ourselves to modulation formats with constant modulus alphabets. Indeed, the joint robust design of the transmitter and the receivers based on the receivers model in [14] where only the phase is corrected is not possible.

As discussed in the previous section, our MU-MISO system uses a very simple receiver model for the precoder design in which all the receivers apply the same real scalar weight. This assumption is necessary to obtain closed-form solutions for the precoders. Nevertheless, in practice, the receivers must correct the received signals wrong amplitudes and phases caused by the errors in the CSI at the transmitter. This goal can be achieved by selecting the receiver coefficients according to the MMSE criterion. For the k-th receiver, the MMSE weight is found via

$$g_{\text{MMSE},k} = c_k^* c_{y,k}^{-1} \tag{52}$$

where $c_{y,k} = \mathrm{E}[|y_k|^2]$ is the variance of the received signal, y_k, and $c_k = \mathrm{E}[u_k^* y_k]$ is the crosscorrelation between the received signal, y_k, and the desired signal, u_k. The estimation of c_{y_k} is straightforward, i.e. it can be found via averaging over time, but the estimation of c_k is more delicate because it depends on the precoder type being used. According to our signal model, the values of c_k for the robust linear, Tomlinson-Harashima and vector precoding are the following

$$\begin{aligned}
c_{\text{Rlin},k} &= \mathrm{E}\left[u_k^* y_k\right] = \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{F}_{\text{Rlin}} \boldsymbol{e}_k \\
c_{\text{RVP},k} &= \frac{1}{N_{\mathrm{B}}} \sum_{m=1}^{N_{\mathrm{B}}} \mathrm{E}\left[\left(u_k(m) + a_{\text{RVP},k}(m)\right)^* y_k(m) \middle| \boldsymbol{u}(m) \right] \\
&= \frac{1}{N_{\mathrm{B}}} \sum_{m=1}^{N_{\mathrm{B}}} \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{x}_{\text{RVP}}(m) \left(u_k(m) + a_{\text{RVP},k}(m)\right)^* \\
c_{\text{RTHP},k} &= \mathrm{E}\left[\boldsymbol{v}^{\mathrm{H}} \left(\mathbf{I}_K - \boldsymbol{B}^{\mathrm{H}}\right) \boldsymbol{P} \boldsymbol{e}_k y_k\right] \\
&= \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{F}_{\text{RTHP}} \boldsymbol{C}_{\boldsymbol{v}} \left(\mathbf{I}_K - \boldsymbol{B}^{\mathrm{H}}\right) \boldsymbol{P} \boldsymbol{e}_k.
\end{aligned}$$

It is apparent from these expressions that the receivers are unable to directly estimate c_k because neither the precoder nor the perturbation signal is known at reception. However, it can be estimated via a time average if the transmitter sends a dedicated training sequence to each receiver. The key idea is that the training symbols are precoded such that the overall channel, i.e. the combination of the channel and the precoder, is equal to c_k. For example, the dedicated pilot symbols for receiver k should be precoded with $\boldsymbol{F}_{\text{RTHP}} \boldsymbol{C}_{\boldsymbol{v}} (\mathbf{I}_K - \boldsymbol{B}) \boldsymbol{P} \boldsymbol{e}_k$ in the case of robust THP.

As a consequence, the proposed system with robust precoding assumes the transmission of two training signals that must be sent frequently. First, distinct common pilot signals must be transmitted from the transmit antennas to enable an estimation of the channel vectors at the single-antenna receivers. With these channel vector estimates, the receivers find the channel covariance matrices via time averaging, with some forgetting factor to account for possible channel non stationarities. Since the covariance matrices only change slowly, the feedback of the dominant eigenvectors of $\boldsymbol{W}_k$, which depends on the channel covariance matrix (see (12)), does not cost much data rate. Whenever the CSI must be fed back to the transmitter, the receiver computes the coefficients via the projection onto the dominant eigenvectors of $\boldsymbol{W}_k$ and transmits the index found by the quantizer to the transmitter.

Second, dedicated pilot signals must be sent to each receiver to allow an estimation of the channel and precoder combination. This estimate is necessary for the receivers MMSE design which correct the phase and the amplitude of the received signal. Notice that phase correction is particularly crucial in a system with erroneous CSI (see [14]).

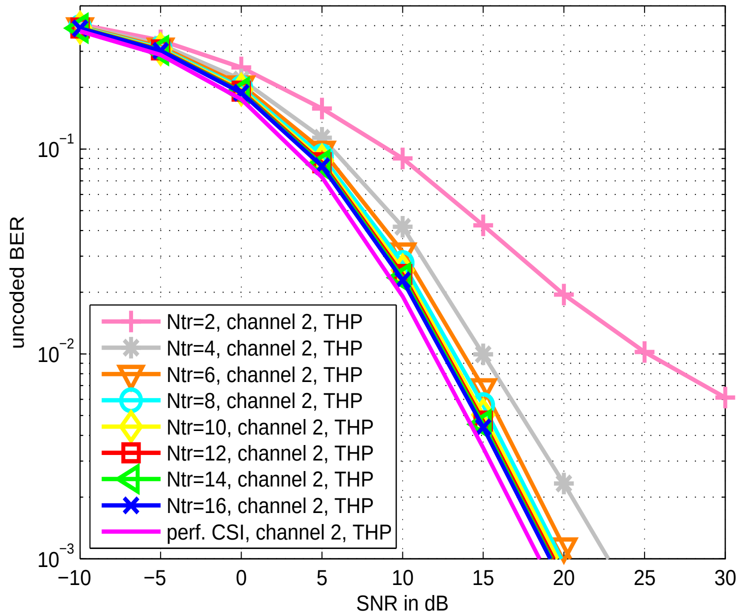

Fig. 7. Effect of Estimation Error on the Proposed Robust Scheme as a Function of Different Training Lengths in an Urban Macrocell Environment.

VI. Simulations

This section presents the results of several computer simulations carried out to assess the proposed MU-MISO system with precoding and limited feedback channel. We considered a MU-MISO system with $N = 4$ antennas at the transmitter and $K = 4$ single antenna users. Performance is evaluated in terms of uncoded Bit Error Rate (BER) versus Signal to Noise Ratio (SNR). The results are the mean of $5{,}000$ channel realizations and 50 QPSK modulated symbols were transmitted in each channel realization. A delay of $D = 2$ slots is considered for all the users which are not fixed-located but moving at a given speed. The Doppler frequency is normalized with respect to the slot period and it is calculated by taking into account that f_{slot} is $1{,}500$ Hz and that the center frequency is 2 GHz. We considered three different environments following the 3GPP Spatial Channel Model (SCM) [30]. The first one corresponds to a suburban macrocell environment (channel 1); the second one is an urban macrocell environment (channel 2); and the last one is an urban microcell scenario (channel 3). We considered channel 2 in most of the results presented in this section due to its intermediate BER performance and diversity. The BER curves were obtained after averaging 100 channel covariance matrices. Finally, we assume, for simplicity reasons, perfect CSI at the receiver for calculating the MMSE coefficients.

We carried out some preliminary simulations to select the size of the training sequence. Fig. 7 shows the uncoded BER for robust THP over urban macrocell environments (channel 2) and different training sequence lengths in order to illustrate the performance degradation caused by channel estimation errors. In this computer experiment, this is the

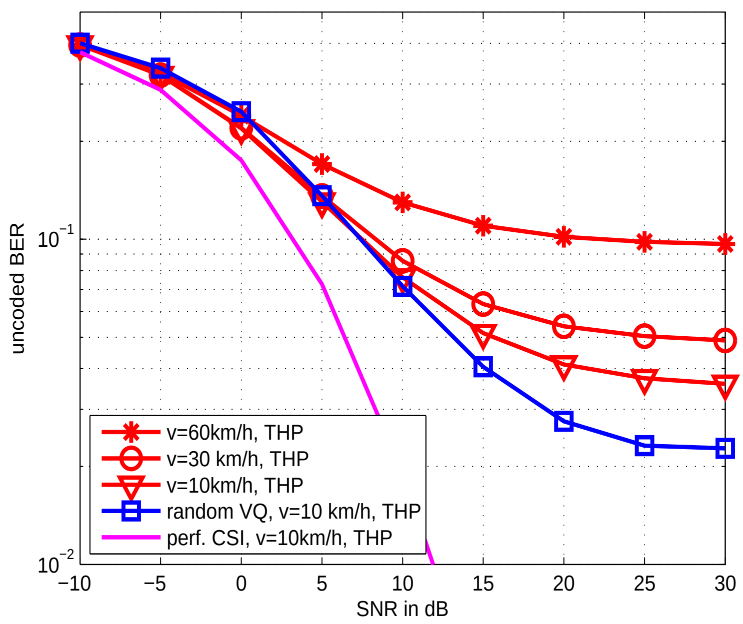

Fig. 8. Effect of User Speed on the Proposed Robust Scheme in an Urban Macrocell Environment with All Errors and 12 Bits per User.

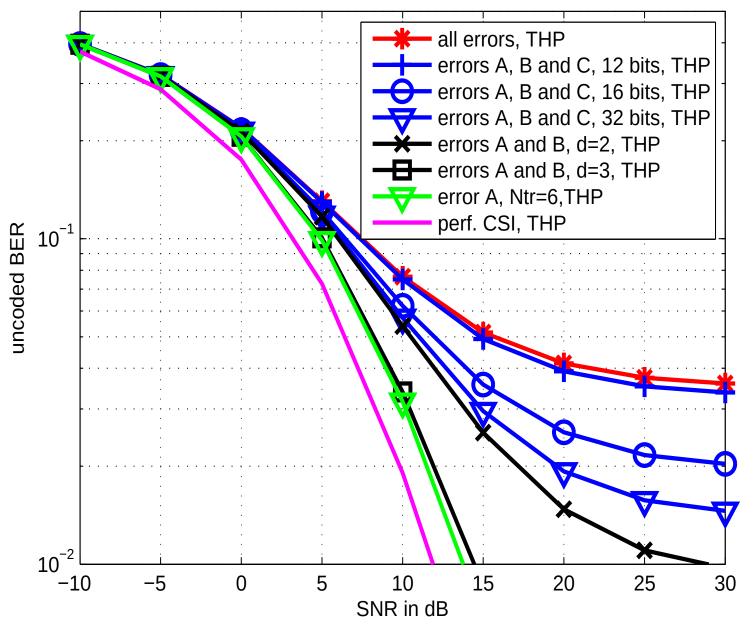

Fig. 9. Effect of Different Types of Errors on the Proposed Robust Scheme in an Urban Macrocell Environment. Error A: Estimation; Error B: Rank Reduction; Error C: Quantization; All Errors: Estimation, Rank Reduction, Quantization, and Delay.

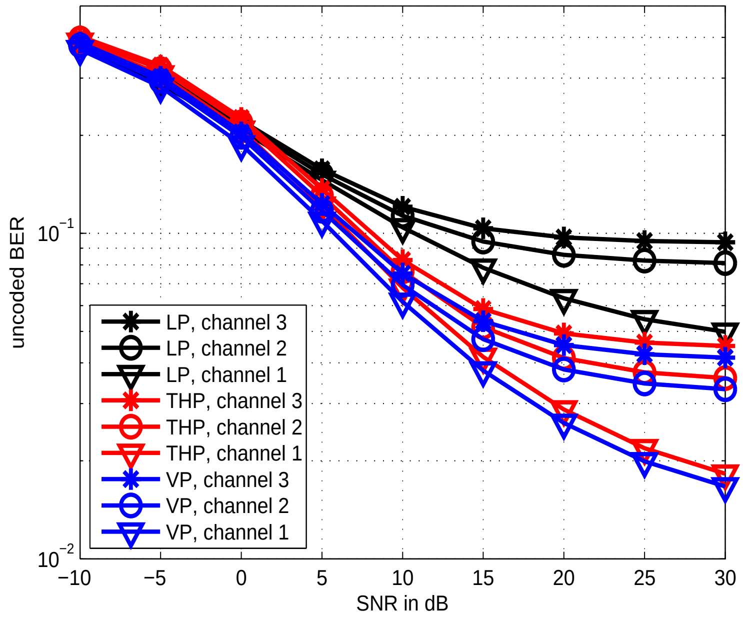

Fig. 10. BER Performance for Different Types of 3GPP Channel Models with the Proposed Robust Precoding and 12 Bits per User.

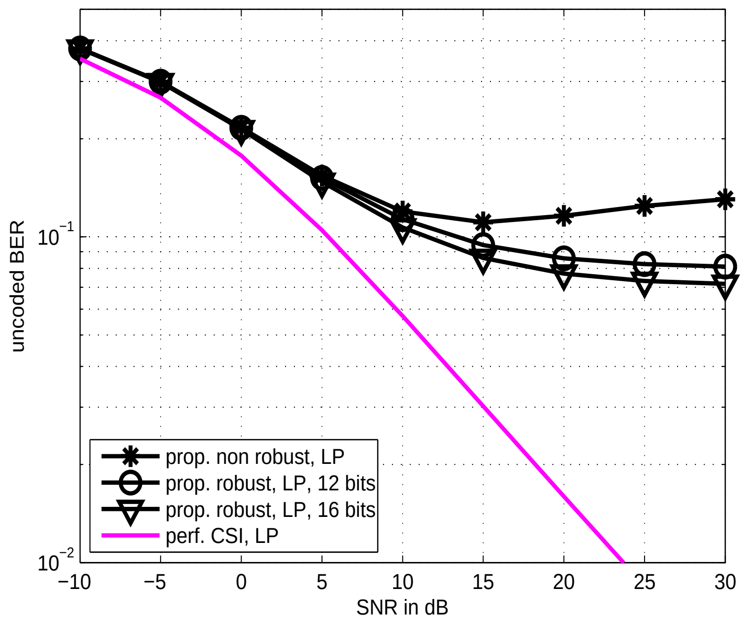

Fig. 11. BER vs. SNR for Linear Precoding and Urban Macrocell Environment.

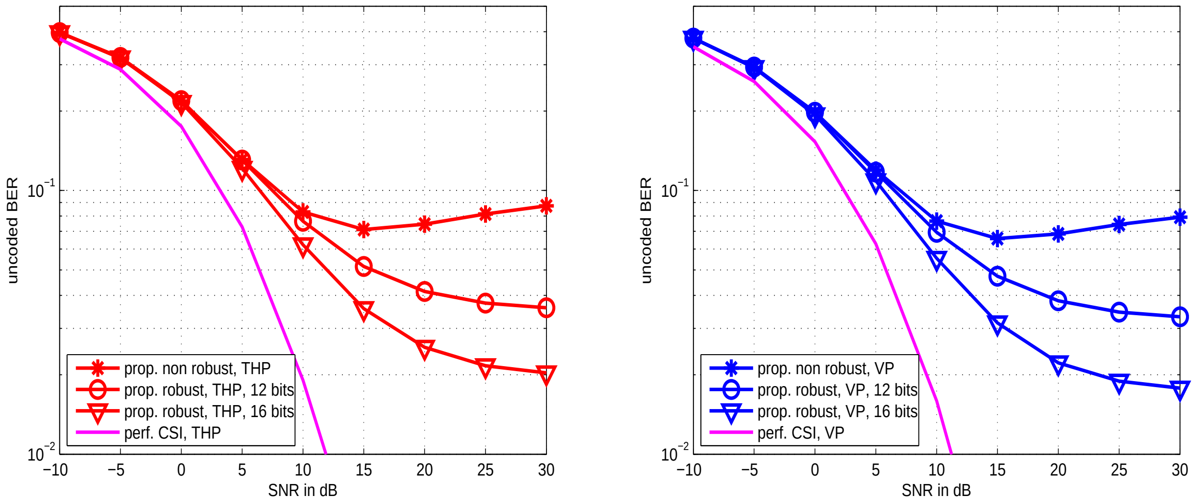

Fig. 12. BER vs. SNR for THP and VP with an Urban Macrocell Environment.

only errors source in the system. As a compromise between training sequence length and performance degradation, we pick for our subsequent simulations the value $N_{\text{tr}} = 6$ which introduces a $2\,$dB loss with respect to the case of perfect CSI.

The performance of robust THP schemes in channel 2 for different user speeds is plotted in Fig. 8. Rank reduction is applied and only $d = 2$ complex coefficients per user are transmitted through the feedback channel. These coefficients are scalarly quantized using 6 bits (3 bits per complex dimension) which yields 12 bits per user. Fig. 8 considers the speed values of 10, 30 and 60 km/h which correspond to normalized Doppler frequencies of 0.0123, 0.0370, and 0.0741, respectively. It is apparent that, as expected, the performance degrades more the faster the fading is.

Fig. 8 also plots the uncoded BER when the users speed is 10 km/h and Random Vector Quantization (RVQ) is applied instead of scalar quantization with the same number of 12 bits per user. Note that in RVQ the stored user's codebook contains channel vectors. Obviously, the errors model that we have developed in section III for scalar quantization is not adequate for RVQ. Indeed, in RVQ, the regularization error matrix used for the robust design is the error matrix $\boldsymbol{C}_{\text{error},k} = \alpha(\boldsymbol{C}_{\hat{\boldsymbol{h}}_k} + \boldsymbol{C}_{\boldsymbol{h}_k})$, where $\boldsymbol{C}_{\hat{\boldsymbol{h}}_k}$ is the covariance matrix of the MMSE estimate. The factor α results from selecting the codebook entry and is the ratio of the MSE with selection over the MSE without selection, where the mean squared error is given by $\mathrm{E}\left[\left\|\hat{\boldsymbol{h}}_{\boldsymbol{k}} - \boldsymbol{y}_i\right\|_2^2\right]$, with $\hat{\boldsymbol{h}}_{\boldsymbol{k}}$ being the output of the MMSE estimator and $\boldsymbol{y}_i$ one of the M codebook entries. In scalar quantization, the error matrix will be given by $\boldsymbol{C}_{\text{error}} = \sum_{k=1}^{K} \boldsymbol{C}_{\text{error},k}^{\mathrm{T}}$. As expected, the system performance is better when RVQ is used. This is because RVQ carries out a joint quantization that uses a much larger codebook ($2^{12} = 4{,}096$ entries per user) and compares a N-dimensional vector with $4{,}096$ complex vectors to choose the closest one for each channel realization and each channel covariance matrix. Therefore, its computational complexity is much higher than scalar quantization, where the search is reduced to a comparison with $2^3 = 8$ scalar values for the real and imaginary parts of each fed–back coefficient. For the considered number of 12 fed–back bits per user, it is clear that the performance of RVQ for medium and high SNR must be better than that obtained with scalar quantization.

Fig. 9 shows the influence on the uncoded BER of the different errors sources considered along this work. Again, robust THP over channel 2 with a user speed of 10 km/h is considered. Obviously, each new error source adds a greater degradation in performance to the previous one. Indeed, note the performance degradation when moving from $d = 3$ to $d = 2$ truncated coefficients. Also, note the performance loss as the number of bits per user decreases. Nevertheless, truncation to $d = 2$ coefficients and $L = 12$ fed–back bits per user ensures a suitable system performance (BER below 4×10^{-2}) with the enormous advantage of noticeably reducing the feedback channel overhead. This overhead reduction is more appreciated the larger the number of transmitting antennas is. In the subsequent computer experiments in this section, we will use $d = 2$ and $L = 12$ as system parameters.

Fig. 10 plots the performance of Linear Precoding (LP), Tomlinson-Harashima Precoding (THP) and Vector Precoding (VP) robust schemes for the three different scenarios described in [30]. All errors sources are considered, i.e. estimation, quantization, truncation, and delay errors inherent to the fed–back sending. Obviously, the performance for channel 1 (suburban macrocell) is much better than that for channel 2 (urban macrocell). And the performance for channel 2 is again better than that for channel 3 (urban microcell). This is because the spatial correlation in channel 1 is considerably larger than in channel 3 (with channel 2 in between) which causes that the third and fourth channel eigenvalues are negligible in channel 1 whereas they have significative values in channel 3 or even in channel 2. Thus, the performance degradation due to truncation to $d = 2$ is more severe in channel 3 than in channel 1. When comparing the three considered precoding schemes, LP exhibits the worst performance for the robust design, as it also occurs in the case of perfect CSI. The achieved performance of VP is always better than that of THP but it is quite similar. Note that the complexity of VP is considerably larger (due to the search in the lattice) which motivates the utilization of suboptimum robust THP schemes instead.

Figs. 11 and 12 show the improvement of our robust schemes with respect to the non–robust ones. It is apparent from these figures that the non–robust curves go up for high SNR due to the sensitivity of these schemes to imperfect CSI. It is also shown the advantage of using the robust schemes, which provide a performance improvement and compensate the CSI imperfect knowledge caused by the different error sources. In these simulation results, a scalar codebook of size $m = 8$ and $m = 16$ has been used, i.e. we are employing $L = 3$ and $L = 4$ bits, respectively, for coding the real and imaginary part of each coefficient. Clearly, if the number of bits is increased, the BER reduces because the errors due to the quantization process are smaller. However, with a codebook of reasonable size, we are obtaining good BER performance. Obviously, larger codebook sizes improve the performance but at the cost of decreasing the compression rate for the CSI sent through the feedback channel and considerably increasing the storage capability at the receivers [33].

Finally, Fig. 13 shows the performance improvement for THP when considering the error modeling Bayesian approach described in section III with respect to the non-Bayesian approach developed in [26]. The Bayesian scheme implements a joint optimization of the estimator, the rank reduction basis, and the inherent predicition of

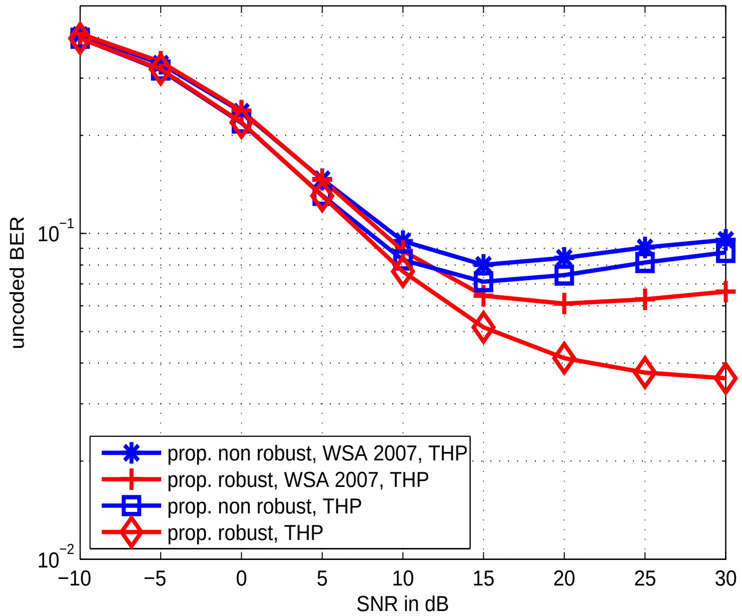

Fig. 13. Performance Comparison with Previous Schemes without Bayesian Formulation.

the estimator by minimizing the overall channel MSE. One of the major contributions of this article is to have found the channel vector PDF conditoned on the fed–back coefficients which is the basis of our robust precoding, i.e. to achieve a Bayesian approach for the errors modeling. Moreover, the advantage of the proposed robust design could be even higher if we could exploit in an adequate procedure the Gaussian input assumption to design the initial codebook according to the Lloyd algorithm.

VII. Conclusion

In this paper, we have investigated the compression of Channel State Information (CSI) data in a MU-MISO system with precoding and limited feedback channel. Three different type of precoders have been considered: Linear Precoding (LP), Tomlinson-Harashima Precoding (THP) and Vector Precoding (VP). We have followed a Bayesian approach to obtain an adequate statistical characterization of the errors in the compressed CSI. Four sources of errors have been considered: channel estimation, truncation for rank reduction, coefficient quantization, and feedback delay. The error modeling has allowed us to formulate robust designs for each precoding scheme with a performance considerably better than that of conventional non–robust schemes. Simulation results show that it is possible to implement these techniques in MU-MISO time-varying channels while transmitting a minimum amount of information through the feedback channel

Appendix A
Rectangular Multivariate Gaussian Probability

With $\boldsymbol{\omega}_k = \Re(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ and $\boldsymbol{\chi}_k = \Im(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$, the PDF of $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$ can be decomposed as follows (see (31))

$$f_{\tilde{\boldsymbol{h}}_{\mathrm{Q},k}}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} = \boldsymbol{\omega}_k + \mathrm{j}\,\boldsymbol{\chi}_k\right) = f_{\boldsymbol{\omega}_k}(\boldsymbol{\omega}_k) f_{\boldsymbol{\omega}_k}(\boldsymbol{\chi}_k)$$

where we have for $L = 1$ that

$$f_{\boldsymbol{\omega}_k}(\boldsymbol{\omega}_k) = \prod_{i=1}^{d} \frac{1}{\gamma_i} \int_{-\gamma_i/2}^{\gamma_i/2} \frac{1}{\sqrt{\pi \varphi_{k,i}}} \exp\left(-\frac{(\zeta_i - \omega_{k,i})^2}{\varphi_{k,i}}\right) \mathrm{d}\zeta_i.$$

After substituting $u_i = (\zeta_i - \omega_{k,i})\sqrt{2/\varphi_{k,i}}$, we obtain

$$f_{\boldsymbol{\omega}_k}(\boldsymbol{\omega}_k) = \prod_{i=1}^{d} \frac{1}{\gamma_i} \int_{\frac{-\gamma_i - 2\omega_{k,i}}{\sqrt{2\varphi_{k,i}}}}^{\frac{\gamma_i - 2\omega_{k,i}}{\sqrt{2\varphi_{k,i}}}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_i^2}{2}\right) \mathrm{d}u_i.$$

With $\int_a^b \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\,\mathrm{d}t = Q(a) - Q(b)$, we finally reach the result of (32).

Appendix B
Rectangular Multivariate Gaussian Centroid

Due to the symmetry of the real and imaginary part of $\boldsymbol{m}_k(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$, it suffices to find the real part of $\boldsymbol{m}_k(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ to proof (34). Let us split up $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$ into its real and imaginary part, i.e. $\boldsymbol{\omega}_k = \Re(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ and $\boldsymbol{\chi}_k = \Im(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$. Then, the real part of $\boldsymbol{m}_k(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ that we denote as $\boldsymbol{\mu}_k$ reads as

$$\boldsymbol{\mu}_k = \frac{\varepsilon(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})}{\prod_{i=1}^{d} \gamma_i^2} \int_{\mathbb{S}} (\boldsymbol{\omega}_k - \boldsymbol{\zeta}) \frac{1}{\pi^d \det(\boldsymbol{\Phi}_{\mathrm{dom},k})} \times \exp\left(-\sum_{i=1}^{d} \frac{(\omega_{k,i} - \zeta_i)^2 + (\chi_{k,i} - \xi_i)^2}{\varphi_{k,i}}\right) \mathrm{d}\boldsymbol{\zeta}\,\mathrm{d}\boldsymbol{\xi}.$$

Here, $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ denote the real and imaginary part of $\boldsymbol{w}$, respectively. We see that we deal with nearly the same integral as the one considered in Appendix A—except the integration with respect to ζ_i, when we compute the i-th entry of $\boldsymbol{\mu}_k$. So, all terms of $\varepsilon(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ drop out except the one corresponding to the integration with respect to ζ_i

$$\mu_{k,i} = \frac{\int_{-\gamma_i/2}^{\gamma_i/2} (\omega_{k,i} - \zeta_i) \frac{1}{\sqrt{\pi\varphi_{k,i}}} \exp\left(-\frac{(\omega_{k,i} - \zeta_i)^2}{\varphi_{k,i}}\right) \mathrm{d}\zeta_i}{Q\left(\frac{-\gamma_i - 2\omega_{k,i}}{\sqrt{2\varphi_{k,i}}}\right) - Q\left(\frac{\gamma_i - 2\omega_{k,i}}{\sqrt{2\varphi_{k,i}}}\right)}.$$

Note that $\mu_{k,i}$ only depends on $\omega_{k,i}$. Therefore, the real part of $m_k(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ only depends on the real part of $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$. With similar steps, it can be shown that the imaginary part of $m_k(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ has the same dependency on the imaginary part of $\tilde{\boldsymbol{h}}_{\mathrm{Q},k}$. Thus, (34) holds. With

$$\int \frac{a-b}{\sqrt{\pi\varphi}} \exp\left(-\frac{(a-b)^2}{\varphi}\right) \mathrm{d}b = \frac{\sqrt{\varphi}}{2\sqrt{\pi}} \exp\left(-\frac{(a-b)^2}{\varphi}\right)$$

which can be obtained with the substitution $u = (a-b)^2/\varphi$, we also get (36).

Appendix C

Rectangular Multivariate Gaussian Covariance

That (35) holds for the off-diagonal elements can be easily shown with similar steps as in Appendix B. So, we only have to obtain the expression for $\sigma_{k,i}(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ that can be found in (37). Due to (35), we have that

$$\sigma_{k,i}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = [\boldsymbol{M}_k]_{i,i} - \left|m_{k,i}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)\right|^2$$

where $[\boldsymbol{M}_k]_{i,i}$ denotes the i-th diagonal element of $\boldsymbol{M}_k(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$ and reads as

$$[\boldsymbol{M}_k]_{i,i} = \frac{\varepsilon(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})}{\prod_{i=1}^d \gamma_i^2} \int_{\mathbb{S}} \left|\tilde{h}_{\mathrm{Q},k,i} - w_i\right|^2 f_{\tilde{\boldsymbol{h}}_k}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k} - \boldsymbol{w}\right) \mathrm{d}\boldsymbol{w}.$$

Similar to Appendix B, $\varepsilon(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})/\prod_{i=1}^d \gamma_i^2$ drops out and we get

$$[\boldsymbol{M}_k]_{i,i} = \lambda_{k,i}(\omega_{k,i}) + \lambda_{k,i}(\chi_{k,i})$$

with $\boldsymbol{\omega}_k = \Re(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$, $\boldsymbol{\chi}_k = \Im(\tilde{\boldsymbol{h}}_{\mathrm{Q},k})$, and

$$\lambda_{k,i}(a) = \frac{\int_{-\gamma_i/2}^{\gamma_i/2} (a-\zeta)^2 \frac{1}{\sqrt{\pi\varphi_{k,i}}} \exp\left(-\frac{(a-\zeta)^2}{\varphi_{k,i}}\right) \mathrm{d}\zeta}{Q\left(\frac{-\gamma_i-2a}{\sqrt{2\varphi_{k,i}}}\right) - Q\left(\frac{\gamma_i-2a}{\sqrt{2\varphi_{k,i}}}\right)}.$$

The enumerator is an integral of the form

$$\int (a-b)^2 \frac{1}{\sqrt{\pi\varphi}} \exp\left(-\frac{(a-b)^2}{\varphi}\right) \mathrm{d}b =$$
$$= \frac{\sqrt{\varphi}}{2\sqrt{\pi}}(a-b)\exp\left(-\frac{(a-b)^2}{\varphi}\right) - \frac{\varphi}{2} Q\left(\frac{\sqrt{2}(b-a)}{\sqrt{\varphi}}\right).$$

To obtain this expression, we used the last integral of Appendix B and employed the substitution $u = (b-a)\sqrt{2/\varphi}$.

From (34), we can follow that

$$\left|m_{k,i}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right)\right|^2 = \mu_{k,i}^2(\omega_{k,i}) + \mu_{k,i}^2(\chi_{k,i})$$

and thus,

$$\sigma_{k,i}\left(\tilde{\boldsymbol{h}}_{\mathrm{Q},k}\right) = \lambda_{k,i}(\omega_{k,i}) - \mu_{k,i}^2(\omega_{k,i}) + \lambda_{k,i}(\chi_{k,i}) - \mu_{k,i}^2(\chi_{k,i}).$$

Defining $\tau_{k,i}(a) = \lambda_{k,i}(a) - \mu_{k,i}^2(a)$ gives (37).

Acknowledgment

This work was supported by Xunta de Galicia, Ministerio de Educacion y Ciencia of Spain and FEDER funds of the European Union under grants number PGIDT06TIC10501PR, TEC2007-68020-C04-01 and CSD2008-00010. Additionally, this work was prepared through an integrated action funded by the Ministerio de Educacion y Ciencia of Spain (fund HA2006-0112) and DAAD of Germany (fund D/06/12809).

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

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

[3] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, August 2003.

[4] 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.

[5] G. Ginis and J. M. Cioffi, "Vectored Transmission for Digital Subscriber Line Systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[6] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in Multiantenna and Multiuser Communications," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[7] 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.

[8] R. E. Assir, F. A. Dietrich, M. Joham, and W. Utschick, "Min-Max MSE Precoding for Broadcast Channels based on Statistical Channel State Information," in *Proc. SPAWC 2006*, July 2006.

[9] B. Chun, "A Downlink Beamforming in Consideration of Common Pilot and Phase Mismatch," in *Proc. European Conference on Circuit Theory and Design*, September 2005.

[10] M. Schubert and S. Shi, "MMSE Transmit Optimization with Interference Pre-Compensation," in *Proc. VTC 2005 Spring*, vol. 2, May 2005, pp. 845–849.

[11] F. Rey, M. Lamarca, and G. Vázquez, "Optimal Power Allocation with Partial Channel Knowledge for MIMO Multicarrier Systems," in *Proc. VTC 2002-Fall*, vol. 4, September 2002, pp. 2121–2125.

[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, W. Utschick, and P. Breun, "Linear Precoding Based on a Stochastic MSE Criterion," in *Proc. of the 13th European Signal Processing Conference*, September 2005.

[14] F. A. Dietrich, P. Breun, and W. Utschick, "Robust Tomlinson-Harashima Precoding for the Wireless Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 631–644, February 2007.

[15] A. D. Dabbagh and D. J. Love, "Multiple Antenna MMSE Based Downlink Precoding with Quantized Feedback or Channel Mismatch," *IEEE Transactions on Communications*, vol. 56, no. 11, pp. 1859–1868, November 2008.

[16] R. L. Choi and R. D. Murch, "New Transmit Schemes and Simplified Receiver for MIMO Wireless Communication Systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1217–1230, November 2003.

[17] 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.

[18] M. Joham and W. Utschick, "Ordered Spatial Tomlinson Harashima Precoding," in *Smart Antennas — State-of-the-Art*, ser. EURASIP Book Series on Signal Processing and Communications, T. Kaiser, A. Bourdoux, H. Boche, J. R. Fonollosa, J. B. Andersen, and W. Utschick, Eds. EURASIP, Hindawi Publishing Corporation, 2006, vol. 3, ch. III. Transmitter, pp. 401–422.

[19] D. A. Schmidt, M. Joham, and W. Utschick, "Minimum Mean Square Error Vector Precoding," 2007, *Accepted for publication in European Transactions on Telecommunications*.

[20] W. Keusgen, C. M. Walke, and B. Rembold, "A System Model Considering the Influence of Front-End Imperfections on the Reciprocity of Up- and Downlink System Impulse Responses," in *Proc. ASST 2001*, September 2001, pp. 243–248.

[21] D. Love, R. Heath, W. Santipach, and M. L. Honig, "What is the Value of Limited Feedback for MIMO Channels?" vol. 42, no. 10, October 2004, pp. 54–59.

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

[23] L. L. Scharf, *Statistical Signal Processing*. Addison-Wesley, 1991.

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

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

[26] P. M. Castro, M. Joham, W. Utschick, and L. Castedo, "Robust Precoding for Multi-User MISO Systems with Limited-Feedback Channels," in *Proc. ITG/IEEE WSA 2007*, February 2007.

[27] W. Jakes, *Microwave Mobile Communications*. John Wiley & Sons, 1974.

[28] P. Dent, G. E. Bottomley, and T. Croft, "Jakes fading model revisited," *Electronics Letters*, vol. 29, no. 13, pp. 1162–1163, June 1993.

[29] T. Zemen and C. F. Mecklenbrauker, "Time-Variant Channel Estimation Using Discrete Prolate Spheroidal Sequences," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3597–3607, September 2005.

[30] 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, "Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (release 6)," 2003.

[31] F. A. Dietrich and W. Utschick, "Pilot-Assisted Channel Estimation Based on Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 1178–1193, March 2005.

[32] J. Balakrishnan, M. Rupp, and H. Viswanathan, "Optimal Channel Training for Multiple Antenna Systems," in *Proc. Conference on Multiaccess, Mobility and Teletraffic for Wireless Communications*, December 2000.

[33] P. M. Castro and L. Castedo, "Adaptive Vector Quantization for Precoding Using Blind Channel Prediction in Frequency Selective MIMO Mobile Channels," in *Proc. ITG WSA 2005*, April 2005.

[34] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication—Part II: Perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 195–202, March 2005.