ROBUST TOMLINSON-HARASHIMA PRECODING

Frank A. Dietrich and Wolfgang Utschick

Institute for Circuit Theory and Signal Processing, Munich University of Technology
Arcisstrasse 21, 80290, Munich, Germany
phone: + (49) 89-289 28511, fax: + (49) 89-289 28504, email: {Dietrich, Utschick}@nws.ei.tum.de
web: www.nws.ei.tum.de

ABSTRACT

Tomlinson-Harashima Precoding (THP) for a wireless system with multiple cooperative transmitters and *non-cooperative* receivers is considered (downlink channel). In the literature THP has been optimized for this channel assuming complete channel state information (CSI). But quality of CSI is generally poor at the transmitter due to the time-varying channel. We present a robust optimization of THP and a *combined optimization* of THP *and* channel estimation, which take into account the quality of CSI. They yield a significantly improved robustness towards errors in CSI.

1. INTRODUCTION

The downlink of a wireless communication system with multiple transmit antennas and non-cooperative receivers with one antenna each is considered. Precoding (pre-equalization) is applied at the transmitter to reduce interference (space division multiple access). *Tomlinson-Harashima Precoding* (THP)—a generalization of the work Tomlinson and Harashima (see references in [1, 2]) for intersymbol-interference—is a non-linear precoder. It can be considered a one-dimensional implementation of "writing-on-dirty paper" [3] or successive encoding. Due to its non-linearity it achieves very good performance, but it is also very sensitive to errors in channel state information (CSI).

CSI is available at the transmitter from the uplink in time-division duplex (TDD) systems or from (limited) feedback from the receivers. In both cases the—generally—asymmetric assignment of time slots to the up- and downlink and a time-variant channel lead to significantly outdated (partial) CSI at the transmitter.

In spite of such harsh constraints THP is conventionally designed assuming *complete CSI*: Zero-forcing THP was presented in [4] and the minimum mean square error (MMSE) optimization was introduced in [5, 2, 6].

But performance of THP degrades tremendously due to outdated CSI at the transmitter. A robust optimization assuming partial CSI for zero-forcing THP was first presented in [7]. For THP in case of *cooperative* receivers, where the linear prefiltering is still performed at the receiver, a similar approach is given in [8] for a SISO system (without prediction) and a heuristic solution was proposed by [4].

Applying the channel estimates to the THP solution as if they were the true parameters, i.e., assuming small errors, is the standard approach in case of partial CSI (Sec. 3.1). Firstly, we optimize THP based on a mean square error (MSE) criterion [5] and a paradigm from static stochastic programming [9] assuming a stochastic model for the parameter errors. This yields an average robustness (Sec. 3.2). The second approach performs a *conditional mean* (CM) estimate of the MSE cost function given the observation of a training sequence from previous uplink slots. This is a *general* approach to a *combined optimization* of precoding (equalization) and channel estimation (Sec. 3.3). Its solution shows, that linear minimum mean square error (LMMSE) channel estimation is optimum together with a robust design, which communicates size and structure of the parameter errors to THP optimization. These methods can also be applied to THP for frequency selective channels. The additional numerical complexity is small compared to the complexity of the THP and LMMSE estimator (Sec. 3.4). In Sec. 4 we briefly emphasize different aspects of the new robust solutions, e.g., their relation to regularization, channel estimation, and compensation of imperfect feedback. Finally, performance results in terms of uncoded bit error rate (BER) show the gains obtained by a robust optimization of THP (Sec. 5).

Our original contributions are: The CM estimate of the cost function and combined optimization, different interpretations of robust design for THP, and the extension of our previous work [7] to correlated channels and MMSE-THP.

Notation: Random vectors and matrices are denoted by lower and upper case sans serif bold letters (e.g. b, B), whereas the realizations or deterministic variables are, e.g., $\boldsymbol{b}$, $\boldsymbol{B}$. The operators $\mathrm{E}[\bullet]$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, and $\mathrm{tr}(\bullet)$ stand for expectation, transpose, Hermitian transpose, and trace of a matrix, respectively. $\otimes$ and $\delta_{k,k'}$ denote the Kronecker product and function, $\mathrm{vec}(\boldsymbol{B})$ stacks the columns of $\boldsymbol{B}$ in a vector. $\boldsymbol{e}_i$ is the i-th column of an $N \times N$ identity matrix $\boldsymbol{I}_N$.

2. SYSTEM MODEL

Downlink Data Channel: Data symbols $\boldsymbol{s}_\mathrm{d}[n] \in \mathbb{B}^K$ with $\mathrm{E}[\boldsymbol{s}_\mathrm{d}[n]\boldsymbol{s}_\mathrm{d}[n]^\mathrm{H}] = \boldsymbol{I}_K$ and modulation alphabet $\mathbb{B}$ are first reordered before being sequentially precoded (Fig. 1). Nonlinear precoding requires a modulo operator $\mathrm{M}(\bullet)$ at the transmitter and receivers defined as

$$\mathrm{M}(z) = z - \left\lfloor \frac{\mathrm{Re}(z)}{\tau} + \frac{1}{2} \right\rfloor \tau - \mathrm{j} \left\lfloor \frac{\mathrm{Im}(z)}{\tau} + \frac{1}{2} \right\rfloor \qquad (1)$$

with $\tau = 2\sqrt{2}$ for QPSK symbols and defined elementwise for a vector. The *feedback filter* $\boldsymbol{F} \in \mathbb{C}^{K \times K}$ is lower triangular with zero diagonal to ensure spatial causality for a realizable feedback loop. The output of the modulo operator is linearly precoded with $\boldsymbol{P} \in \mathbb{C}^{M \times K}$ and transmitted using M antennas over the channel $\boldsymbol{H}_q \in \mathbb{C}^{K \times M}$ to K receivers in the *downlink* (time slot q). The channel is assumed constant during one time slot ("block-fading"). The non-cooperative receivers are modeled as $\boldsymbol{G} = \beta^{-1}\boldsymbol{I}_K$. Including white additive complex Gaussian noise $\boldsymbol{n}[n] \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}_K)$ the estimate of $\boldsymbol{s}_\mathrm{d}[n]$ before decision is

$$\tilde{\boldsymbol{s}}_\mathrm{d}[n] = \mathrm{M}\left(\beta^{-1}\boldsymbol{H}_q\boldsymbol{P}\boldsymbol{w}[n] + \beta^{-1}\boldsymbol{n}[n]\right) \in \mathbb{C}^K. \qquad (2)$$

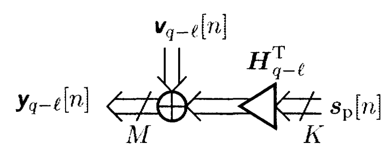

Figure 2. Uplink training channel

Uplink Training Channel: In a TDD system the *channel parameters for optimizing* the THP parameters $\boldsymbol{F}$ and $\boldsymbol{P}$ can be estimated from N training symbols (per receiver) $\boldsymbol{s}_\mathrm{p}[n] \in \mathbb{C}^K$ ($n \in \{1, \ldots, N\}$) in an *uplink* link slot. As an example we assume alternating up-/downlink slots and a worst-case delay of 3 slots (due to processing the training sequence) to the first uplink slot available with a training sequence. The receive training signal is (Fig. 2)

$$\boldsymbol{y}_q[n] = \boldsymbol{H}_q^\mathrm{T}\boldsymbol{s}_\mathrm{p}[n] + \boldsymbol{v}_q[n] \in \mathbb{C}^M,\ n \in \{1, \ldots, N\} \qquad (3)$$

with additive white noise $\boldsymbol{v}_q[n] \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0}, \sigma_v^2 \boldsymbol{I}_M)$. Collecting all N training symbols in the columns of $\boldsymbol{S}'_\mathrm{p} \in \mathbb{C}^{K \times N}$ we obtain

$$\boldsymbol{Y}_q = \boldsymbol{H}_q^\mathrm{T}\boldsymbol{S}'_\mathrm{p} + \boldsymbol{V}_q \in \mathbb{C}^{M \times N}$$
$$\bar{\boldsymbol{y}}_q = \mathrm{vec}[\boldsymbol{Y}_q] = (\boldsymbol{S}'^{\mathrm{T}}_\mathrm{p} \otimes \boldsymbol{I}_M)\boldsymbol{h}_q + \bar{\boldsymbol{v}}_q \in \mathbb{C}^{MN},$$

where $\boldsymbol{h}_q = \mathrm{vec}[\boldsymbol{H}_q^\mathrm{T}]$. Considering training signals from Q previous uplink slots the total observation is

$$\boldsymbol{y}_q = \boldsymbol{S}\boldsymbol{h}_{\mathrm{T},q} + \boldsymbol{v}_q \in \mathbb{C}^{MNQ} \qquad (4)$$

with $\boldsymbol{h}_{\mathrm{T},q} = [\boldsymbol{h}_{q-3}^\mathrm{T}, \boldsymbol{h}_{q-5}^\mathrm{T}, \ldots, \boldsymbol{h}_{q-(2Q+1)}^\mathrm{T}]^\mathrm{T} \in \mathbb{C}^{QMK}$, $\boldsymbol{y}_q = [\bar{\boldsymbol{y}}_{q-3}^\mathrm{T}, \bar{\boldsymbol{y}}_{q-5}^\mathrm{T}, \ldots, \bar{\boldsymbol{y}}_{q-(2Q+1)}^\mathrm{T}]^\mathrm{T}$, and $\boldsymbol{S} = \boldsymbol{I}_Q \otimes \boldsymbol{S}'^{\mathrm{T}}_\mathrm{p} \otimes \boldsymbol{I}_M$.

The channel coefficients $\boldsymbol{h}_q = \mathrm{vec}[\boldsymbol{H}_q^\mathrm{T}]$ are modeled as a stationary zero mean complex Gaussian random vector with covariance matrix $\boldsymbol{C}_{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h}_q\boldsymbol{h}_q^\mathrm{H}]$, which is block diagonal assuming $\mathrm{E}[\boldsymbol{h}_{k,q}\boldsymbol{h}_{k',q}^\mathrm{H}] = \boldsymbol{C}_{\boldsymbol{h}_k}\delta_{k,k'}$. The k-th column of $\boldsymbol{H}_q^\mathrm{T}$ is $\boldsymbol{h}_{k,q} \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$. For simplicity, we assume identical autocorrelation $r[i]$ (normalized to $r[0] = 1$) for all elements of $\boldsymbol{h}_q$ and a time-difference of i slots: $\boldsymbol{C}_\mathrm{T}$ is Toeplitz with first column $[r[0], r[2], \ldots, r[2Q-2]]^\mathrm{T}$ and $\boldsymbol{C}_{\boldsymbol{h}_\mathrm{T}} = \boldsymbol{C}_\mathrm{T} \otimes \boldsymbol{C}_{\boldsymbol{h}}$.

Based on this model the current channel parameters $\boldsymbol{h}_q$ may be estimated (predicted) by

$$\hat{\boldsymbol{h}}_q = \boldsymbol{W}\boldsymbol{y}_q, \qquad (5)$$

where $\boldsymbol{W}$ is an LMMSE estimator [10], for example.

Throughout the article we assume that first and second order channel and noise statistics are given.

3. ROBUST TOMLINSON-HARASHIMA PRECODING

THP is optimized based on its linear representation (Fig. 1), where $\boldsymbol{a}[n]$ and $\tilde{\boldsymbol{a}}[n]$ describe the modulo operations at the transmitter and receivers, respectively (see references in [1, 2]). MMSE optimization, which is superior to a zero-forcing design, is considered. The MSE based on knowledge of the channel parameters $\boldsymbol{H}_q$ is defined as

$$\begin{aligned} &\mathrm{f_S}(\boldsymbol{P}, \boldsymbol{F}, \beta; \boldsymbol{H}_q, \mathcal{M}_\mathrm{S}) = \mathrm{E}[\|\boldsymbol{d}[n] - \tilde{\boldsymbol{d}}[n]\|_2^2] = \\ &\mathrm{tr}\left((\boldsymbol{I}_K - \boldsymbol{F})\boldsymbol{C}_{\boldsymbol{w}}(\boldsymbol{I}_K - \boldsymbol{F})^\mathrm{H}\right) + \beta^{-2}\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^\mathrm{H}\boldsymbol{H}_q^\mathrm{H}\boldsymbol{H}_q\boldsymbol{P}\right) \\ &+ \beta^{-2}K\sigma_n^2 - 2\beta^{-1}\mathrm{tr}\left(\mathrm{Re}\left((\boldsymbol{I}_K - \boldsymbol{F})\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^\mathrm{H}\boldsymbol{H}_q^\mathrm{H}\right)\right). \end{aligned} \qquad (6)$$

Spatial ordering is not considered here to simplify our notation, but the extension is straight forward [6, 2] and applied in the simulations. For the derivations the widespread assumption $\boldsymbol{C}_{\boldsymbol{w}} = \mathrm{diag}[\sigma_{w_k}^2]_{k=1}^K$ is made.

To emphasize the generality of our approaches we introduce the set of parameters to be optimized $\mathbb{P}$, which includes the constraints of the optimization: $\{\boldsymbol{P}, \boldsymbol{F}, \beta\} \in \mathbb{P}$. Additionally, required model parameters, which we assume to be perfectly known, are collected in $\mathcal{M}_\mathrm{S} = \{\sigma_{w_i}^2|_{i=1}^K, \sigma_n^2\}$.

3.1. Standard Solution

The optimization problem for MMSE-THP based on a complete knowledge of $\boldsymbol{H}_q$ is

$$\begin{aligned} &\min_{\boldsymbol{P}, \boldsymbol{F}, \beta} \mathrm{f_S}(\boldsymbol{P}, \boldsymbol{F}, \beta; \boldsymbol{H}_q, \mathcal{M}_\mathrm{S}) \text{ s.t. } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^\mathrm{H}\right) \le P_\mathrm{T}, \\ &\boldsymbol{F}: \text{ lower triangular, zero diagonal} \end{aligned} \qquad (7)$$

with average transmit power constrained by P_T. The standard approach treats an estimate $\hat{\boldsymbol{H}}_q$ (5) as if it was error-free, and sets $\boldsymbol{H}_q = \hat{\boldsymbol{H}}_q$ (Fig. 3). Under this assumption

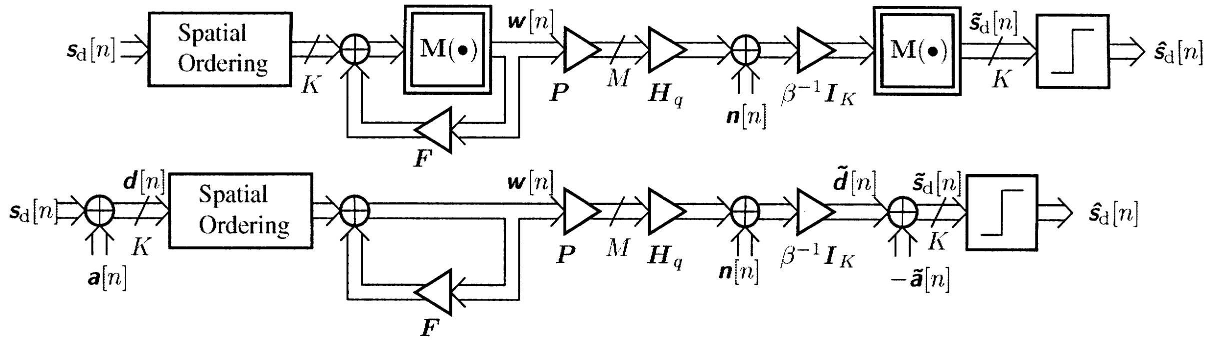

Figure 1. THP for downlink data transmission with M transmit antennas and K non-cooperative receivers: Representation with modulo operators $\mathrm{M}(\bullet)$ (upper diagram) and equivalent linear representation (lower diagram)

the solution of (7) is[1] [6, 2]

$$P_{\mathrm{S}} = [p_{\mathrm{S},1}, \dots, p_{\mathrm{S},K}] \quad \text{with} \tag{8}$$

$$p_{\mathrm{S},k} = \beta_{\mathrm{S}} \left(\hat{H}_q^{(k),\mathrm{H}} \hat{H}_q^{(k)} + \frac{K\sigma_n^2}{P_{\mathrm{T}}} I_M \right)^{-1} \hat{H}_q^{(k),\mathrm{H}} e_k,$$

$$F_{\mathrm{S}} = [f_{\mathrm{S},1}, \dots, f_{\mathrm{S},K}] \quad \text{with} \tag{9}$$

$$f_{\mathrm{S},k} = -\beta_{\mathrm{S}}^{-1} \begin{bmatrix} 0_{k\times M} \\ \hat{\mathcal{H}}_q^{(k)} \end{bmatrix} p_{\mathrm{S},k},$$

where β_{S} is chosen to satisfy the power constraint with equality, $H_q^{(k)}$ denotes the first k rows, and $\mathcal{H}_q^{(k)}$ the last $K-k$ rows of H_q: $H_q = [H_q^{(k),\mathrm{T}}, \mathcal{H}_q^{(k),\mathrm{T}}]^{\mathrm{T}}$.

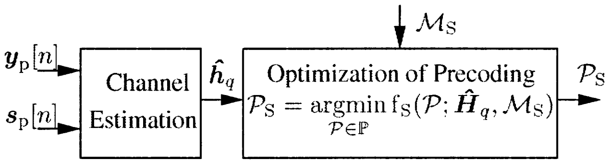

Figure 3. Standard optimization of THP

3.2. Robust Optimization

In the standard approach only the channel estimate was communicated to the optimization (Fig. 3). Now, we extend this interface passing also information about the first and second order error statistics to THP optimization (Fig. 4).

From the point of view of optimization the true channel $h_q = \mathrm{vec}(H_q^{\mathrm{T}})$ is a random variable, with the linear model

$$h_q = \hat{h}_q + \theta \tag{10}$$

for the estimate $\hat{h}_q$ (5). The estimation error θ is described by its mean $\mathrm{E}[\theta] = 0$ (for simplicity) and covariance matrix

$$C_\theta = \mathrm{E}[(h_q - \hat{h}_q)(h_q - \hat{h}_q)^{\mathrm{H}}] \tag{11}$$

$$= C_h + W C_y W^{\mathrm{H}} - W C_{yh} - C_{yh}^{\mathrm{H}} W^{\mathrm{H}}. \tag{12}$$

[1] Although it is a non-convex problem the solution can be found *explicitly* via the Lagrange function and KKT conditions [6, 2].

As $h_q = \mathrm{vec}(H_q^{\mathrm{T}})$ we have a model of the error Θ in $H_q = \hat{H}_q + \Theta$ with $C_{\Theta^{\mathrm{H}}} = \mathrm{E}[\Theta^{\mathrm{H}}\Theta]$ as needed for robust THP.

Based on this error model we define the MSE on average given $\hat{H}_q$:

$$\mathrm{f}_{\mathrm{R}}(P, F, \beta; \hat{H}_q, \mathcal{M}_{\mathrm{R}}) = \mathrm{E}_\Theta[\mathrm{f}_{\mathrm{S}}(P, F, \beta; \hat{H}_q + \Theta), \mathcal{M}_{\mathrm{S}}]. \tag{13}$$

This yields a robustness on average in contrast to worst case robustness as obtained by a min-max formulation. The robust optimization problem is similar to (7)

$$\min_{P,F,\beta} \mathrm{f}_{\mathrm{R}}(P, F, \beta; \hat{H}_q, \mathcal{M}_{\mathrm{R}}) \ \text{s.t.}\ \mathrm{tr}\left(P C_w P^{\mathrm{H}}\right) \le P_{\mathrm{T}},$$
$$F: \text{ lower triangular, zero diagonal} \tag{14}$$

with the solution following the steps in [6, 2]

$$P_{\mathrm{R}} = [p_{\mathrm{R},1}, \dots, p_{\mathrm{R},K}] \quad \text{with} \tag{15}$$

$$p_{\mathrm{R},k} = \beta_{\mathrm{R}} \left(\hat{H}_q^{(k),\mathrm{H}} \hat{H}_q^{(k)} + C_{\Theta^{\mathrm{H}}} + \frac{K\sigma_n^2}{P_{\mathrm{T}}} I_M \right)^{-1} \hat{H}_q^{(k),\mathrm{H}} e_k,$$

$$F_{\mathrm{R}} = [f_{\mathrm{R},1}, \dots, f_{\mathrm{R},K}] \quad \text{with} \tag{16}$$

$$f_{\mathrm{R},k} = -\beta_{\mathrm{R}}^{-1} \begin{bmatrix} 0_{k\times M} \\ \hat{\mathcal{H}}_q^{(k)} \end{bmatrix} p_{\mathrm{R},k}.$$

The additional statistical information about the estimation error leads to a structured loading of the inverse in $p_{\mathrm{R},k}$. The set of model parameters (Fig. 4) is extended to $\mathcal{M}_{\mathrm{R}} = \{\mathcal{M}_{\mathrm{S}}, \mathrm{E}[\Theta], C_{\Theta^{\mathrm{H}}}\}$.

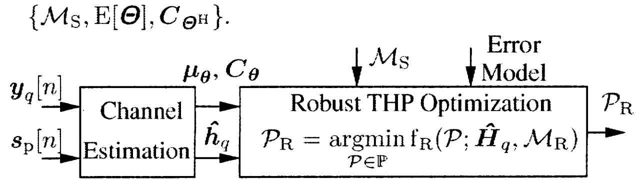

Figure 4. Robust optimization of THP

3.3. Conditional Mean Estimate of Cost Function

The robust optimization allows for a THP design, which takes into account size and structure of the estimation errors.

It remains to choose/optimize the channel estimator $\boldsymbol{W}$ (5). Although many alternatives for $\boldsymbol{W}$ are available [10], we would like to find the optimum estimator for the THP. Any choice would be heuristic and a performance comparison would be necessary. The goal is to optimize the precoder directly based on the available observation $\boldsymbol{y}_q$ (4) about the channel coefficients. Additionally, for our model (Sec. 2) the conditional probability density function $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_q}(\boldsymbol{h}_q|\boldsymbol{y}_q)$ (PDF) is complex Gaussian with moments $\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q} = \mathrm{E}[\boldsymbol{h}_q|\boldsymbol{y}_q]$ and $\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}_q} = \mathrm{E}[(\boldsymbol{h}_q - \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q})(\boldsymbol{h}_q - \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q})^{\mathrm{H}}|\boldsymbol{y}_q]$

$$\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q} = \hat{\boldsymbol{h}}_q = \boldsymbol{W}\boldsymbol{y}_q, \boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}} + \sigma_{\mathrm{v}}^2\boldsymbol{I}_{MNQ})^{-1}$$
$$\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}_q} = \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{W}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}\boldsymbol{h}}, \tag{17}$$

where $\boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}} = \mathrm{E}[\boldsymbol{h}_q\boldsymbol{h}_{\mathrm{T},q}^{\mathrm{H}}] = [r[3], r[5], \ldots, r[2Q+1]] \otimes \boldsymbol{C}_{\boldsymbol{h}}$ and $\boldsymbol{W}$ is equivalent to the LMMSE estimator [10].

Given $\boldsymbol{y}_q$ the CM estimate of the cost function (6) is:

$$\mathrm{f_C}(\boldsymbol{P}, \boldsymbol{F}, \beta; \boldsymbol{y}_q, \mathcal{M}_{\mathrm{C}}) = \mathrm{E}_{\boldsymbol{H}_q}[\mathrm{f_S}(\boldsymbol{P}, \boldsymbol{F}, \beta; \boldsymbol{H}_q, \mathcal{M}_{\mathrm{S}})|\boldsymbol{y}_q]. \tag{18}$$

The new cost function depends explicitly only on the observation $\boldsymbol{y}_q$. It can be easily shown that the cost function is equivalent to $\mathrm{f_R}(\boldsymbol{P}, \boldsymbol{F}, \beta; \hat{\boldsymbol{H}}_q, \mathcal{M}_{\mathrm{R}})$ with the LMMSE estimate $\hat{\boldsymbol{H}}_q$ and $\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}} = \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q} = \sum_{k=1}^{K} \boldsymbol{C}^*_{\boldsymbol{h}_k|\boldsymbol{y}_q}$. Thus, the solution is given by (15) and (16) with $\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}} = \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q}$ ($\mathcal{M}_{\mathrm{C}} = \{\mathcal{M}_{\mathrm{S}}, \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q}\}$).

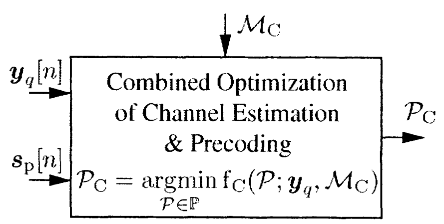

Figure 5. Combined optimization of THP and channel estimation based on the observed training sequence $\boldsymbol{y}_q$

3.4. Numerical Complexity

For comparing numerical complexity we assume an LMMSE estimator (17) for standard and robust (combined) optimization. The comparison also holds in case ordering optimization is included. The additional complexity is given by $\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}}$, which is based on (17), and of $O(M^3K^2NQ)$.

4. INTERPRETATION OF RESULTS

4.1. Tikhonov Regularization

The robust cost function (13) can be expressed in terms of the standard cost function (6)

$$\mathrm{f_R}(\boldsymbol{P}, \boldsymbol{F}, \beta; \hat{\boldsymbol{H}}_q, \mathcal{M}_{\mathrm{R}}) = \mathrm{f_S}(\boldsymbol{P}, \boldsymbol{F}, \beta; \hat{\boldsymbol{H}}_q, \mathcal{M}_{\mathrm{S}}) + \beta^{-2}\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}}\boldsymbol{P}\right). \tag{19}$$

The second term penalizes the norm of the linear prefilter $\boldsymbol{P}$, i.e., it regularizes the solution. Decreasing the norm of $\boldsymbol{P}$ leads to smaller amplification of the estimation errors and reduced sensitivity in performance. Similar solutions are known from least squares [11, 12] with a stochastic error in the model. The second term can be written in terms of a sum of weighted norms of the columns $\boldsymbol{p}_k$ of $\boldsymbol{P}$

$$\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}}\boldsymbol{P}\right) = \sum_{k=1}^{K} \sigma_{w_k}^2 \boldsymbol{p}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}}\boldsymbol{p}_k \tag{20}$$

The weighting matrix $\boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}}$ incorporates the error structure in the solution.

4.2. Improved Channel Estimation

The CM estimate of the Gramian of $\boldsymbol{H}_q$ is [13]

$$\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q] = \hat{\boldsymbol{H}}_q^{\mathrm{H}}\hat{\boldsymbol{H}}_q + \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q}. \tag{21}$$

This term also appears, when writing the CM estimate of the cost (18) explicitly. As computation of the Gramian changes the statistics, this has to be taken considered when designing the estimator. The standard approach intuitively applies the maximum likelihood invariance principle [10] to the case of a stochastic parameter model, where this principle does not hold. If $\hat{\boldsymbol{H}}_q$ is not an LMMSE estimate, we can think of

$$\widehat{\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q} = \hat{\boldsymbol{H}}_q^{\mathrm{H}}\hat{\boldsymbol{H}}_q + \boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}}, \tag{22}$$

which occurs in the robust cost function (13), as an improved estimate of the Gramian $\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q$.

4.3. Compensation for Imperfect Feedback Filter

The structured loading term in the inverse (15) incorporates knowledge about estimation errors in $\hat{\boldsymbol{H}}_q^{(k)}$ *and* $\hat{\mathcal{H}}_q^{(k)}$. As $\boldsymbol{P}_{\mathrm{R}}$ and $\boldsymbol{P}_{\mathrm{C}}$ only depend on $\hat{\boldsymbol{H}}_q^{(k)}$, the forward filter $\boldsymbol{P}_{\mathrm{R}}$ compensates for imperfect interference cancellation in the feedback. This observation is confirmed, when the solution $\boldsymbol{F}_{\mathrm{S}}$ from (9), which depends on $\hat{\boldsymbol{H}}_q$, is applied to (6) and the resulting cost function is optimized for $\boldsymbol{P}$.

4.4. Relation of Robust to Combined Optimization

As discussed above the solution from robust (13) and combined optimization (18) are identical for

$$\hat{\boldsymbol{h}}_q = \mathrm{vec}(\hat{\boldsymbol{H}}_q^{\mathrm{T}}) = \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q}, \quad \boldsymbol{C}_{\boldsymbol{\Theta}^{\mathrm{H}}} = \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q}. \tag{23}$$

This shows that LMMSE estimation is optimum for MMSE-THP design, if its error covariance matrix is considered in the inverse in (15). A heuristic choice of the estimator $\boldsymbol{W}$ is needed for robust optimization. Moreover, for robust optimization no assumption about the PDF of $\boldsymbol{H}_q$ or its conditional PDF is made, only knowledge about the 1st and 2nd order moments is assumed. On the other hand, for the solution of the combined optimization joint Gaussianity of the channel and observation is required.

978-3-8007-2909-8/05/$20.00 ©2005 IEEE

5. PERFORMANCE EVALUATION

Simulation parameters: QPSK data symbols, $M = 4$ transmit antennas in a uniform linear array (half wavelength spacing), and $K \in \{2, 4\}$ receivers are used. Transmitter and receivers are modeled as in Sec. 2. All complex Gaussian channel coefficients have the same Jakes power spectrum with maximum Doppler frequency $f_{\mathrm{d}} = 0.08$ (normalized to the slot period), i.e., a max. speed of $\sim 65\,\mathrm{km/h}$ for slot frequency $1500\,\mathrm{Hz}$ at $2\,\mathrm{GHz}$ as in TDD-CDMA. The azimuth directions of the receivers' channels are Laplace distributed with means $[-15°, 0°, 15°]$ and standard deviation $10°$. Random QPSK sequences of length $N = 30$ are used for training in the uplink. (Robust) spatial ordering (Fig. 1) is included similarly to [5]. An LMMSE channel estimator $\boldsymbol{W}$ is employed, i.e., the solutions for robust and combined optimization are identical (23) with this choice. The symmetric slot allocation from Sec. 2 is chosen. A comparison with other robust methods (e.g. min-max) is not possible in this article as—to our knowledge—this is the *first robust* THP design for non-cooperative receivers.

Results: Figures 6 and 7 show the uncoded average BER for $K = 2$ and $K = 4$. If *no* LMMSE channel estimation/prediction is performed, i.e., channel estimates outdated by 3 time slots are used ($Q = 1$), the BER saturates at 0.5. Performing LMMSE channel estimation based on the received training sequences from $Q = 5$ previous uplink slots and applying the standard solution (7) yields a significant improvement. But still the performance degradation compared to complete CSI is large. Robust or combined optimization of THP (cf. Eqs. 13 and 18) results in an improvement of about $3\,\mathrm{dB}$ for $K = 2$ and $4\,\mathrm{dB}$ for $K = 4$ at a BER of 0.01. The gain is larger for $K = 4$ as the sensitivity of THP increases with the system load, i.e., the BER degradation is larger than for $K = 2$. There is no saturation for robust and standard THP in the shown SNR range due to prediction and same SNR in training and data channel. The performance gains are mainly due to the rich structure in $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q}$.

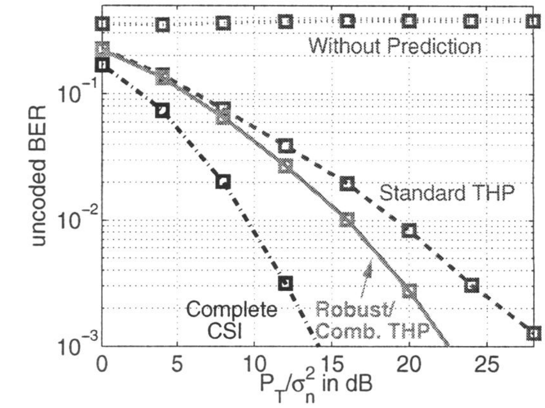

Figure 6. Uncoded average BER vs. $P_{\mathrm{T}}/\sigma_n^2$ at Doppler frequency $f_{\mathrm{d}} = 0.08$ ($K = 2$).

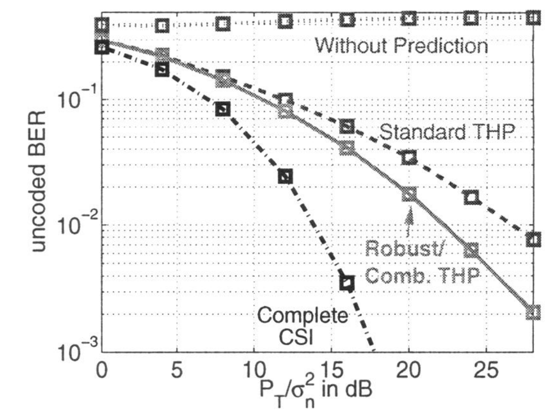

Figure 7. Uncoded average BER vs. $P_{\mathrm{T}}/\sigma_n^2$ at Doppler frequency $f_{\mathrm{d}} = 0.08$ ($K = 4$ – fully loaded system).

6. REFERENCES

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

[2] M. Joham. *Optimization of Linear and Nonlinear Transmit Signal Processing*. Dissertation, Technische Universität München, June 2004.

[3] W. Yu and J. M. Cioffi. Trellis Precoding for the Broadcast Channel. In *Proc. Globecom*, pages 1344–1348, Nov. 2001.

[4] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber. Tomlinson-Harashima Precoding in Space-Time Transmission for Low-Rate Backward Channel. In *Proc. Int. Zurich Seminar on Broadband Communications*, Feb. 2002.

[5] M. Joham, J. Brehmer, and W. Utschick. MMSE Approaches to Multiuser Spatio-Temporal Tomlinson-Harashima Precoding. In *Proc. ITG SCC'04*, pages 387–394, January 2004.

[6] M. Joham and W. Utschick. Ordered Spatial Tomlinson Harashima Precoding. In *Smart Antennas - State-of-the-Art*, chapter 3: Transmitter. EURASIP, Hindawi Publishing Corporation, 2005. *To appear.*

[7] R. Hunger, F. A. Dietrich, M. Joham, and W. Utschick. Robust Transmit Zero-Forcing Filters. In *Proc. of the ITG Workshop on Smart Antennas*, Munich, Germany, March 2004.

[8] A. P. Liavas. Tomlinson-Harashima Precoding with Partial Channel Knowledge. *IEEE Trans. Commun.*, (1):5–9, Jan. 2005.

[9] András Prékopa. *Stochastic Programming*. Kluwer, 1995.

[10] Steven M. Kay. *Fundamentals of Statistical Signal Processing - Estimation Theory*. PTR Prentice Hall, 1st edition, 1993.

[11] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[12] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of Ill-Posed Problems*. V. H. Winston & Sons, 1977.

[13] F. A. Dietrich, F. Hoffmann, and W. Utschick. Conditional Mean Estimator for the Gramian Matrix of Complex Gaussian Random Variables. In *Proc. of ICASSP*, Philadelphia, March 2005.