This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

Tomlinson-Harashima Precoding: A Continuous Transition From Complete to Statistical Channel Knowledge

Frank A. Dietrich, Peter Breun, 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, Breun, Utschick}@nws.ei.tum.de
web: www.nws.ei.tum.de

***Abstract*— Tomlinson-Harashima precoding (THP) for a system with multiple transmit antennas and non-cooperative receivers is considered (broadcast channel). Design of THP for this channel is typically based on complete CSI at the transmitter, which is not available in mobile wireless systems. For larger Doppler frequencies it even performs significantly worse than linear precoding or simple beamforming due to its high sensitivity to parameter errors. We apply a novel optimization criterion based on partial CSI to THP. For this robust design it is shown that a continuous transition from complete to statistical CSI is achieved and that THP is now guaranteed to perform always better than or equal to linear precoding.**

I. Introduction

Tomlinson-Harashima precoding is a well established transmit processing technique in wireline communications [1], [2]. In contrast to the wireline channel the mobile wireless channel may be highly time-varying. In time-division duplex (TDD) systems channel state information (CSI) is available from training signals in the uplink or—in other systems—due to (limited) feedback from the receivers. The CSI at the transmitter is already significantly outdated at low Doppler frequencies, as the slot structure is generally asymmetric in TDD systems or only a small feedback rate is available. This results in a large performance degradation for THP [3].

We consider a system with M transmit antennas and K non-cooperating (decentralized) one-antenna receivers, i.e., a broadcast channel. For this channel THP can be considered the one-dimensional implementation of "writing-on-dirty-paper" [4], [5]. Moreover, reducing interference by preprocessing the transmitted signal results in simple, power-efficient (mobile) receivers. Applicability of THP in this scenario highly depends on its robustness w.r.t. erroneous CSI.

Conventionally THP is designed assuming *complete CSI* (C-CSI): Zero-forcing THP was presented in [6] and the minimum mean square error (MMSE) optimization was introduced in [7], [8], [9]. A *robust optimization* for zero-forcing THP assuming erroneous CSI and no specific receiver processing was first presented in [3], which includes linear minimum mean square error (LMMSE) prediction of the channel parameters. For THP in case of *cooperative* receivers, where the linear feedforward filter is still at the receiver, a similar approach is given in [10] for a SISO system and a heuristic solution was proposed by [6]; both do not include prediction of the parameters.

We define *partial CSI* (P-CSI) as knowledge of a conditional probability density function (PDF) for the random channel parameters given outdated and noisy observations of the training sequence. *Statistical CSI* (S-CSI)—as asymptotic case of P-CSI—is the knowledge of the PDF of the random channel parameters, e.g., the observation of the training sequence is statistically independent of the current channel state.

In case of P-CSI the channel is a random variable from the perspective of the transmitter. Consequently, the cost function is also a random variable, as it depends on the channel and is described by a conditional PDF. The traditional solution is to estimate the channel, e.g., using an LMMSE estimator, and apply the estimates to the cost function as if they were error-free. We show in Sec. V that bit error rate (BER) performance of this traditional THP design degrades for increasing parameter errors and, finally, saturates at a BER of 0.5. We show that it even performs worse than a simple beamforming solution based on S-CSI.

This is not acceptable and a robust solution incorporating the knowledge about the PDF is given in Sec. III. Based on the MSE criterion [7] we propose a novel THP optimization based on P-CSI: Taking a Bayesian approach we perform a conditional mean estimate of the MSE relying on observed training signals from the uplink (Sec. III). This general paradigm has already been applied successfully to equalizer optimization in [11]. In contrast to [3] the transmitter assumes receivers with CSI performing a simple phase correction. This optimization problem including the optimum ordering of the data streams can be solved explicitly with an average transmit power constraint. Numerical complexity is similar to the traditional THP optimization. In Sec. IV we show that our solution achieves a smooth transition between complete and S-CSI. For S-CSI, i.e., only knowledge about the channel parameters' PDF, THP reduces to linear processing in most scenarios, where it performs a channel covariance matrix based beamforming. In these cases, from information theoretic considerations it is intuitively clear that successive non-linear precoding in a "writing-on-dirty-paper" fashion—as performed by THP—is not possible for S-CSI. Our THP solution confirms this intuition. Thus, our THP optimization for P-CSI performs always better than a linear precoder as confirmed by simulations (Sec. V).

In summary *our contributions* are: 1) THP optimization with P-CSI including receivers' processing. 2) A THP design enabling the transition from complete to S-CSI. 3) Interpretation of THP based on S-CSI as MSE based linear precoding, i.e., beamforming. 4) Performance of THP based on P-CSI is always better than linear precoding in the region of interest.

Notation: Random vectors and matrices are denoted by lower and upper case sans serif bold letters (e.g. $\boldsymbol{\mathsf{b}}$, $\boldsymbol{\mathsf{B}}$), 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 ith column of an $N \times N$ identity matrix $\boldsymbol{I}_N$

II. System Model

Downlink Data Channel: Data symbols $\boldsymbol{\mathsf{s}}_{\mathrm{d}}[n] \in \mathbb{B}^K$ with $\mathrm{E}[\boldsymbol{\mathsf{s}}_{\mathrm{d}}[n]\boldsymbol{\mathsf{s}}_{\mathrm{d}}[n]^{\mathrm{H}}] = \boldsymbol{I}_K$ and modulation alphabet $\mathbb{B}$ are first reordered before being sequentially precoded (Fig. 1). Reordering is performed by the *permutation matrix* $\boldsymbol{\Pi}^{(\mathcal{O})} \in \{0,1\}^{K\times K}$, whose (i,k)th element is one, if user k is precoded in the i-step, and zero elsewhere. The dependency on the specific ordering $\mathcal{O}$ is denoted by the superscript $(\mathcal{O})$. Non-linear precoding requires a modulo operator $\mathrm{M}(\bullet)$ at the transmitter and receivers, which is defined as

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

with $\tau = 2\sqrt{2}$ for QPSK symbols and taken element-wise for a vector. The *feedback filter* $\boldsymbol{F} \in \mathbb{C}^{K\times K}$ with columns $\boldsymbol{f}_k$ is lower triangular with zero diagonal to ensure spatial causality for a realizable feedback loop. The output $\boldsymbol{\mathsf{w}}[n]$ of the modulo operator is linearly precoded with $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K] \in \mathbb{C}^{M\times K}$ 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 $\boldsymbol{G} = \mathrm{diag}[g_k]_{k=1}^K \in \mathbb{C}^{K\times K}$. Including white additive complex Gaussian noise $\boldsymbol{\mathsf{n}}[n] \sim \mathcal{N}_{\mathrm{c}}(\boldsymbol{0}, \sigma_{\boldsymbol{\mathsf{n}}}^2 \boldsymbol{I}_K)$ the estimate of the signal $\boldsymbol{\mathsf{s}}_{\mathrm{d}}[n]$ before decision is

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

Additionally, a common real-valued scaling β^{-1} is introduced to allow for a power constraint at the transmitter in the optimization (Sec. III).

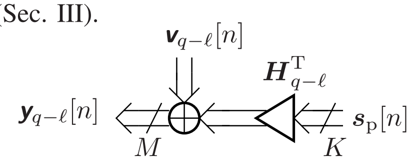

Figure 2. Uplink training channel

Downlink Training Channel: There are two main concepts for training symbol based channel estimation in the downlink: Providing K receiver specific (dedicated) training sequences (in some systems restricted to M "antenna specific" sequences) or transmitting one common training sequence to the receivers. For the former case a different transmit filter $\boldsymbol{q}_k$ can be used for every training sequence, e.g., $\boldsymbol{q}_k$ is the kth column of $\boldsymbol{P}$. In the latter case it is transmitted with the same filter $\boldsymbol{q}_k = \boldsymbol{q}$, e.g., over the first antenna $\boldsymbol{q}_k = \boldsymbol{e}_1$. Thus, we assume that receiver k knows $\boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k$ ($\boldsymbol{h}_{k,q}$ is the kth column of $\boldsymbol{H}_q^{\mathrm{T}}$) and corrects the phase based on this CSI:

$$\boldsymbol{G} = \mathrm{diag}\left[g_k\right]_{k=1}^K = \mathrm{diag}\left[\left(\boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k\right)^* / \left|\boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k\right|\right]_{k=1}^K. \tag{3}$$

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* slot. We assume alternating up-/downlink slots and a 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)

$$\boldsymbol{\mathsf{y}}_q[n] = \boldsymbol{\mathsf{H}}_q^{\mathrm{T}}\boldsymbol{s}_{\mathrm{p}}[n] + \boldsymbol{\mathsf{v}}_q[n] \in \mathbb{C}^M,\ n \in \{1, \ldots, N\} \tag{4}$$

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

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

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

$$\boldsymbol{\mathsf{y}}_q = \boldsymbol{S}\boldsymbol{\mathsf{h}}_{\mathrm{T},q} + \boldsymbol{\mathsf{v}}_q \in \mathbb{C}^{MNQ} \tag{5}$$

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

The channel coefficients $\boldsymbol{\mathsf{h}}_q = \mathrm{vec}[\boldsymbol{\mathsf{H}}_q^{\mathrm{T}}]$ are modeled as a stationary zero mean complex Gaussian random vector with covariance matrix $\boldsymbol{C}_{\boldsymbol{\mathsf{h}}} = \mathrm{E}[\boldsymbol{\mathsf{h}}_q\boldsymbol{\mathsf{h}}_q^{\mathrm{H}}]$, which is block diagonal assuming $\mathrm{E}[\boldsymbol{\mathsf{h}}_{k,q}\boldsymbol{\mathsf{h}}_{k',q}^{\mathrm{H}}] = \boldsymbol{C}_{\boldsymbol{\mathsf{h}}_k}\delta_{k,k'}$. For simplicity, we assume identical autocorrelation $r[i]$ (normalized to $r[0] = 1$) for all elements of $\boldsymbol{\mathsf{h}}_q$ 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 $\boldsymbol{C}_{\boldsymbol{\mathsf{h}}_{\mathrm{T}}} = \boldsymbol{C}_{\mathrm{T}} \otimes \boldsymbol{C}_{\boldsymbol{\mathsf{h}}}$.

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

III. Optimization with Partial CSI at the Transmitter

The modulo operators at the transmitter and receivers can be expressed by the summation of $\boldsymbol{\mathsf{a}}[n]$ and $\tilde{\boldsymbol{\mathsf{a}}}[n]$ [2], [9]. Optimization of THP is based on this linear representation (Fig. 1). MMSE optimization of THP was shown to be superior to zero-forcing [7], where the MSE

$$\mathrm{C}_{\mathrm{T}}(\boldsymbol{P}, \boldsymbol{F}, \beta; \boldsymbol{H}_q) = \mathrm{E}[\|\boldsymbol{\mathsf{d}}[n] - \tilde{\boldsymbol{\mathsf{d}}}[n]\|_2^2] \tag{6}$$

depends on the current channel parameters $\boldsymbol{\mathsf{H}}_q$.

Complete knowledge about the realization $\boldsymbol{H}_q$ is never available at the transmitter, but obtained via the observations in $\boldsymbol{y}_q$ (5). Thus, from the point of view of optimization the channel $\boldsymbol{\mathsf{H}}_q$ is a random variable, which is described by its conditional probability density function (PDF) $\mathrm{p}_{\boldsymbol{\mathsf{H}}|\boldsymbol{\mathsf{y}}_q}(\boldsymbol{H}_q|\boldsymbol{y}_q)$. Based on our assumptions (Sec. II) it is a (complex) Gaussian

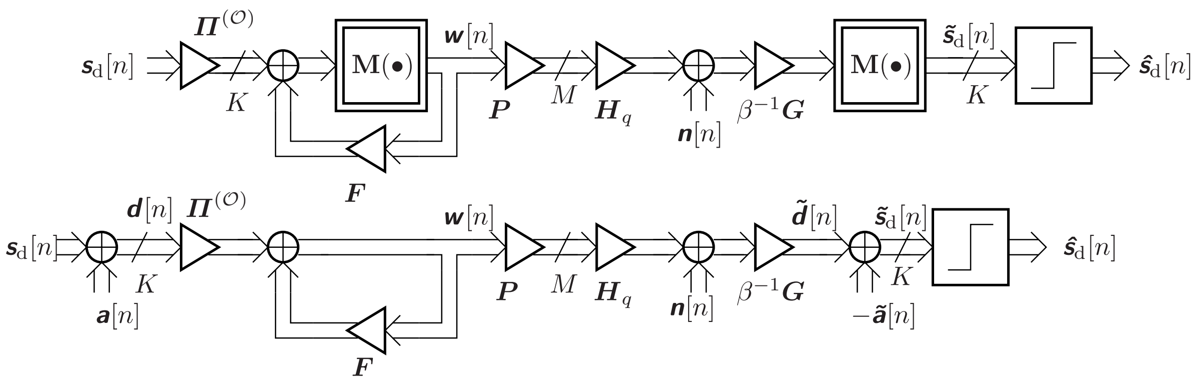

Figure 1. THP for downlink data transmission with M transmit antennas and K non-cooperative receivers: Representation with modulo operators (upper diagram) and equivalent linear representation (lower diagram)

PDF with mean $\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q} = \mathrm{E}[\boldsymbol{h}_q|\boldsymbol{y}_q]$ and covariance matrix $\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}_q} = \mathrm{E}[(\boldsymbol{h}_q - \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q})(\boldsymbol{h}_q - \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q})^{\mathrm{H}}|\boldsymbol{y}_q]$:

$$\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_q} = \hat{\boldsymbol{h}}_q = \boldsymbol{W}\boldsymbol{y}_q, \boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}} + \sigma_{\boldsymbol{v}}^2\boldsymbol{I}_{MNQ})^{-1}$$
$$\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}_q} = \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{W}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}}^{\mathrm{H}}, \quad (7)$$

where $\boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}} = \mathrm{E}[\boldsymbol{h}_q\boldsymbol{h}_{\mathrm{T},q}^{\mathrm{H}}] = [r[3], r[5], \ldots, r[2Q+1]] \otimes \boldsymbol{C}_{\boldsymbol{h}}$, $\boldsymbol{h}_q = \mathrm{vec}(\boldsymbol{H}_q^{\mathrm{T}})$, and $\boldsymbol{W}$ is equivalent to the LMMSE estimator [12]. Due to the partial knowledge about $\boldsymbol{H}_q$ via the conditional PDF the cost function (6) is a random variable, too. In the sequel, the traditional approach for dealing with this random cost function and our new approach are presented.

For the subsequent derivations the widespread assumption $\boldsymbol{C}_{\boldsymbol{w}} = \mathrm{diag}[\sigma_{w_k}^2]_{k=1}^K$ is made.

A. Traditional Optimization

Conventionally, the errors in $\hat{\boldsymbol{H}}_q$ are assumed to be negligible and the estimate $\hat{\boldsymbol{H}}_q$ is used as if it was the true one, i.e., we set $\boldsymbol{H}_q = \hat{\boldsymbol{H}}_q$ in (6), leading to the optimization problem

$$\min_{\boldsymbol{P},\boldsymbol{F},\beta} \mathrm{C}_{\mathrm{T}}(\boldsymbol{P},\boldsymbol{F},\beta;\hat{\boldsymbol{H}}_q) \quad \text{s.t. } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\right) \leq P_{\mathrm{T}},$$
$$\boldsymbol{F}: \text{ lower triangular, zero diagonal} \quad (8)$$

with average transmit power constrained by P_{T} and the constraint on $\boldsymbol{F}$ to ensure implementability. The solution using the Lagrange approach and the KKT conditions is [9]

$$\boldsymbol{p}_{\mathrm{T},k} = \beta_{\mathrm{T}}\left(\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O})} + \frac{K\sigma_{\boldsymbol{n}}^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k \quad (9)$$

$$\boldsymbol{f}_{\mathrm{T},k} = -\beta_{\mathrm{T}}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \hat{\boldsymbol{B}}_{\mathrm{T},k}^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{T},k}, \quad (10)$$

where β_{T} is chosen to satisfy the power constraint with equality. $\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O})}$ denotes the first k rows and $\hat{\boldsymbol{B}}_{\mathrm{T},k}^{(\mathcal{O})}$ the last $K-k$ rows of the reordered and estimated effective channel matrix $\boldsymbol{\Pi}^{(\mathcal{O})}\hat{\boldsymbol{G}}_{\mathrm{T}}\hat{\boldsymbol{H}}_q$. The estimate is obtained from the LMMSE estimator in (7) and our assumption about the receivers' processing (3), i.e., $\hat{\boldsymbol{G}}_{\mathrm{T}} = \mathrm{diag}\left[\left(\hat{\boldsymbol{h}}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k\right)^* / \left|\hat{\boldsymbol{h}}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k\right|\right]_{k=1}^K$. Optimization of the ordering is similar to Sec. III-D [9].

B. Conditional Mean Estimate of the Cost Function

In our systematic approach we do not estimate the channel parameters, but the MSE cost function (6). Employing the Bayesian paradigm the best estimator in the mean-square sense is the conditional mean (CM) estimator. The CM estimate of (6) is

$$\begin{aligned}\mathrm{C}_{\mathrm{P}}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_q) &= \mathrm{E}[\mathrm{C}_{\mathrm{T}}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{H}_q)|\boldsymbol{y}_q] \quad (11)\\ &= \mathrm{tr}\Big((\boldsymbol{I}_K - \boldsymbol{F})\boldsymbol{C}_{\boldsymbol{w}}(\boldsymbol{I}_K - \boldsymbol{F})^{\mathrm{H}}\\ &+\beta^{-2}\boldsymbol{C}_{\boldsymbol{n}} + \beta^{-2}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q]\boldsymbol{P}\Big)\\ &-2\beta^{-1}\mathrm{Re}\left\{\mathrm{tr}\left(\boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}[\boldsymbol{G}\boldsymbol{H}_q|\boldsymbol{y}_q]\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}(\boldsymbol{I}_K - \boldsymbol{F})^{\mathrm{H}}\right)\right\}.\end{aligned}$$

This approach results in a large difference in the quality of the solution, as the MSE is a non-linear function of $\boldsymbol{H}_q$, i.e., its PDF is not Gaussian anymore. The impact of this paradigm and conclusion on THP design will become clear from the discussion of the solution (Sec. IV).

The novel cost function (11) was simplified using $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G} = \boldsymbol{I}_K$. Now, it depends on $\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q]$ and $\mathrm{E}[\boldsymbol{G}\boldsymbol{H}_q|\boldsymbol{y}_q]$, which can be computed explicitly in terms of the moments in (7). The CM estimate of the channel Gramian $\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q$ is

$$\mathrm{E}\left[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q\right] = \hat{\boldsymbol{H}}_q^{\mathrm{H}}\hat{\boldsymbol{H}}_q + \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q}, \ \hat{\boldsymbol{H}}_q = \mathrm{E}[\boldsymbol{H}_q|\boldsymbol{y}_q]. \quad (12)$$

The conditional covariance matrix can be computed using (7) $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q} = \mathrm{E}[(\boldsymbol{H}_q - \hat{\boldsymbol{H}}_q)^{\mathrm{H}}(\boldsymbol{H}_q - \hat{\boldsymbol{H}}_q)|\boldsymbol{y}_q] = \sum_{k=1}^K \boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_q}^*$ and is identical to the covariance matrix of the estimation error $\mathrm{E}[(\boldsymbol{H}_q - \hat{\boldsymbol{H}}_q)^{\mathrm{H}}(\boldsymbol{H}_q - \hat{\boldsymbol{H}}_q)]$ due to the orthogonality property of the LMMSE estimator *and* the jointly (complex) Gaussian distribution of $\boldsymbol{y}_q$ and $\boldsymbol{h}_q$ [12].

The CM estimate of the effective channel $\boldsymbol{G}\boldsymbol{H}_q$ reads

$$\mathrm{E}[\boldsymbol{G}\boldsymbol{H}_q|\boldsymbol{y}_q] = \hat{\boldsymbol{G}}\hat{\boldsymbol{H}}_q + \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_q}, \quad \hat{\boldsymbol{G}} = \mathrm{E}[\boldsymbol{G}|\boldsymbol{y}_q] = \mathrm{diag}\left[\hat{g}_k\right]_{k=1}^K,$$

where the kth row of $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_q} \in \mathbb{C}^{K\times M}$ is given by

$$\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_q} = \boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_q}^* c_{\mathsf{x}_k|\boldsymbol{y}_q}^{-1}\left(\mathrm{E}\left[|\mathsf{x}_k|\ |\boldsymbol{y}_q\right] - \mu_{\mathsf{x}_k|\boldsymbol{y}_q}\hat{g}_k\right).$$

The channel estimated by the kth receiver is $\mathsf{x}_k = \boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k$ (Sec. II) with first and second order moments

$$\mu_{\mathsf{x}_k|\boldsymbol{y}_q} = \mathrm{E}[\mathsf{x}_k|\boldsymbol{y}_q] = \hat{\boldsymbol{h}}_{k,q}^{\mathrm{T}}\boldsymbol{q}_k,$$
$$c_{\mathsf{x}_k|\boldsymbol{y}_q} = \mathrm{E}[|\mathsf{x}_k - \mu_{\mathsf{x}_k|\boldsymbol{y}_q}|^2|\boldsymbol{y}_q] = \boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_q}^*\boldsymbol{q}_k.$$

With [13] the remaining terms, i.e., the CM estimate of the receivers' processing g_k and of the magnitude $|x_k|$ are

$$\hat{g}_k = \mathrm{E}[g_k|\boldsymbol{y}_q] \quad (13)$$
$$= \frac{\sqrt{\pi}}{2}\frac{|\mu_{x_k|\boldsymbol{y}_q}|}{c^{1/2}_{x_k|\boldsymbol{y}_q}}\frac{\mu^*_{x_k|\boldsymbol{y}_q}}{|\mu_{x_k|\boldsymbol{y}_q}|}\ {}_1F_1\left(\frac{1}{2},2,-\frac{|\mu_{x_k|\boldsymbol{y}_q}|^2}{c_{x_k|\boldsymbol{y}_q}}\right),$$

$$\mathrm{E}[|x_k|\ |\boldsymbol{y}_q] = \frac{\sqrt{\pi}}{2}c^{1/2}_{x_k|\boldsymbol{y}_q}\ {}_1F_1\left(-\frac{1}{2},1,-\frac{|\mu_{x_k|\boldsymbol{y}_q}|^2}{c_{x_k|\boldsymbol{y}_q}}\right), \quad (14)$$

where ${}_1F_1(\alpha,\beta,z)$ is the confluent hypergeometric function.

C. Solution for CM Estimate of the Cost Function

With the CM estimate of the MSE expressed in terms of the moments (7) of the conditional PDF, which determines the transmitter's P-CSI, the new optimization problem is

$$\min_{\boldsymbol{P},\boldsymbol{F},\beta} \mathrm{C_P}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_q) \ \text{ s.t. } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\right) \leq P_{\mathrm{T}},$$
$$\boldsymbol{F}: \text{ lower triangular, zero diagonal.} \quad (15)$$

The solution is again obtained based on the Lagrange function and KKT conditions following the steps in [9], where the derivative of the Lagrange function w.r.t. $\boldsymbol{P}$ yields

$$\boldsymbol{P} = \beta\left(\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q] + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_q]\times$$
$$\times\,\boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}}(\boldsymbol{I}_K-\boldsymbol{F}), \quad (16)$$

which is needed below for interpretation. The complete solution for a given ordering $\mathcal{O}$ is

$$\boldsymbol{p}_{\mathrm{P},k} = \beta_{\mathrm{P}}\left(\boldsymbol{L}_{\boldsymbol{y}_q} + \hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})} + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k$$
$$\boldsymbol{f}_{\mathrm{P},k} = -\beta_{\mathrm{P}}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{P},k}, \quad (17)$$

where β_{P} is chosen to satisfy the power constraint with equality. It is expressed in terms of the first k and last $K-k$ rows $\hat{\boldsymbol{A}}_k^{(\mathcal{O})} \in \mathbb{C}^{k\times M}$ and $\hat{\boldsymbol{B}}_k^{(\mathcal{O})} \in \mathbb{C}^{K-k\times M}$, respectively, of the ordered *CM estimate* of the effective channel $\boldsymbol{GH}_q$

$$\begin{bmatrix}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix} = \boldsymbol{\Pi}^{(\mathcal{O})}\begin{bmatrix}\hat{\boldsymbol{A}}_k\\ \hat{\boldsymbol{B}}_k\end{bmatrix} = \boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}[\boldsymbol{GH}_q|\boldsymbol{y}_q] \in \mathbb{C}^{K\times M}.$$

The difference to the traditional solution (9) and (10) is twofold: A structured loading

$$\boldsymbol{L}_{\boldsymbol{y}_q} = \mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q] - \mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_q]\mathrm{E}[\boldsymbol{GH}_q|\boldsymbol{y}_q] \quad (18)$$

is added in the inverse of $\boldsymbol{p}_{\mathrm{P},k}$ (17) and the CM estimate of the overall channel $\boldsymbol{GH}_q$ is used instead of simply plugging the CM estimate of $\boldsymbol{H}_q$ into the product $\boldsymbol{GH}_q$ as in Sec. III-A. Obviously, the additional complexity of this solution is small compared to the complexity of the traditional THP solution assuming an LMMSE estimator is used and the confluent hypergeometric function is given by a look-up table.

D. Ordering Optimization

Applying the solution of (15), which is given in terms of a fixed ordering $\mathcal{O}$, to the cost function (11) we obtain

$$\mathrm{C_P}(\mathcal{O};\boldsymbol{y}_q) = \mathrm{tr}(\boldsymbol{C}_{\boldsymbol{w}}) -$$
$$-\sum_{k=1}^{K}\boldsymbol{e}_k^{\mathrm{T}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}\left(\boldsymbol{L}_{\boldsymbol{y}_q} + \hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})} + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k.$$

To avoid the high complexity of $O(K!K^3)$ for a full search among all possible orderings the standard suboptimum approach [9] minimizes each term in the sum separately starting with the Kth term. Thus, the user to precode in the ith step, where $i = K,\ldots,1$, is determined by

$$\max_k \boldsymbol{e}_k^{\mathrm{T}}\hat{\boldsymbol{A}}_i^{(\mathcal{O})}\left(\boldsymbol{L}_{\boldsymbol{y}_q} + \hat{\boldsymbol{A}}_i^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_i^{(\mathcal{O})} + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_i^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k.$$

Finally, this yields the permutation matrix $\boldsymbol{\Pi}^{(\mathcal{O}_{\mathrm{P}})}$.

IV. Transition from Complete to Statistical CSI

A. Complete CSI at the Transmitter

For C-CSI, i.e., $\hat{\boldsymbol{H}}_q = \boldsymbol{H}_q$, the error covariance matrix of the channel estimate is zero (7) and the CM estimate $\mathrm{E}[\boldsymbol{GH}_q|\boldsymbol{y}_q]$ is equivalent to the (true) effective channel $\boldsymbol{GH}_q$ (Sec. III-B):

$$\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_q} \to \boldsymbol{0}_{M\times M},\ \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_q} \to \boldsymbol{0}_{K\times M},\ \hat{\boldsymbol{H}}_q \to \boldsymbol{H}_q,\ \hat{\boldsymbol{G}} \to \boldsymbol{G}.$$

Convergence is achieved for $\sigma_v^2 \to 0$ and $r[i] \to 1\,\forall i$. The MSE (6) is no random variable anymore, as the conditional PDF $\mathrm{p}_{\boldsymbol{H}|\boldsymbol{y}_q}(\boldsymbol{H}_q|\boldsymbol{y}_q) = \delta(\boldsymbol{h}_q - \boldsymbol{W}\boldsymbol{y}_q)$ is a Dirac distribution centered at $\boldsymbol{W}\boldsymbol{y}_q$. The optimum filters $\boldsymbol{p}_{\mathrm{P},k}$ and $\boldsymbol{f}_{\mathrm{P},k}$ converge to

$$\boldsymbol{p}_{\mathrm{C},k} = \beta_{\mathrm{C}}\left(\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{A}_k^{(\mathcal{O})} + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k \quad (19)$$

$$\boldsymbol{f}_{\mathrm{C},k} = -\beta_{\mathrm{C}}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \boldsymbol{B}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{C},k} \quad (20)$$

with the reordered effective channel

$$\left[\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{T}},\ \boldsymbol{B}_k^{(\mathcal{O}),\mathrm{T}}\right]^{\mathrm{T}} = \boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{GH}_q. \quad (21)$$

B. Statistical CSI at the Transmitter

The critical case for THP performance occurs for large errors in $\hat{\boldsymbol{H}}_q$, where it performs worse than a linear precoder, which is less sensitive to an inaccurate CSI than THP. Statistical CSI is the limiting case, where the observation $\boldsymbol{y}_q$ is statistically independent of the current channel $\boldsymbol{H}_q$, i.e., $\mathrm{p}_{\boldsymbol{H}|\boldsymbol{y}_q}(\boldsymbol{H}_q|\boldsymbol{y}_q) \to \mathrm{p}_{\boldsymbol{H}}(\boldsymbol{H}_q)$. This means no information about the current channel realization is available if $\sigma_v^2 \to \infty$ or for a temporally uncorrelated channel $r[i] \to 0$ for $i > 0$, e.g., at very high Doppler frequency. If $\boldsymbol{H}_q$ is zero mean, the LMMSE estimate is $\hat{\boldsymbol{H}}_q = \boldsymbol{0}_{K\times M}$ in this limit.

One consequence of this observation is that $\boldsymbol{P}_{\mathrm{T}} \to \boldsymbol{0}_{M\times K}$ (cf. Eqn 9). Although this case is not achieved exactly in the simulations as $\|\hat{\boldsymbol{H}}_q\|_{\mathrm{F}} > 0$ numerically, the BER saturates at 0.5. This is not acceptable as any simple beamforming solution

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

based on S-CSI would perform better as will be shown in Sec. V.

The solution based on the CM estimate of the MSE (11) has much more appealing properties. It is determined by the asymptotic behavior of

$$\begin{aligned} \mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q] &\rightarrow \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}} = \mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q] \\ \mathrm{E}[\boldsymbol{G}\boldsymbol{H}_q|\boldsymbol{y}_q] &\rightarrow \boldsymbol{U}_{\boldsymbol{H}} \end{aligned} \tag{22}$$

with kth row $\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{H}} = \sqrt{\pi}\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k}^*/(\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k}^*\boldsymbol{q}_k)^{1/2}/2$. Applying these limits to (17) we obtain the solution for THP in case only S-CSI is available. Its interpretation follows in the next two subsections.

C. Inactivity of Modulo Operation at Transmitter

In case of S-CSI our THP solution still seems to be non-linear, as $\boldsymbol{F}$ is non-zero. But the modulo operator (1), which is piecewise linear, is only non-linear if its input's real- or imaginary part is larger than $\tau/2$ in magnitude. This is mainly determined by $\boldsymbol{F}$. The relative frequency of the modulo operator to be *inactive* for each reordered data stream (in the order it is precoded) vs. maximum normalized Doppler frequency f_{d} in the scenario of Sec. V is shown in Fig. 3. We see that the modulo operator is always inactive for the first data stream to be precoded due to the lower triangular structure of $\boldsymbol{F}$. For larger Doppler frequencies the modulo operator is inactive for all users. Thus, THP is linear and reduces to linear precoding which is briefly derived in the next section. This observation was confirmed by more extensive simulations in other scenarios with "well"-conditioned $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}}$ (see Sec. IV-E).

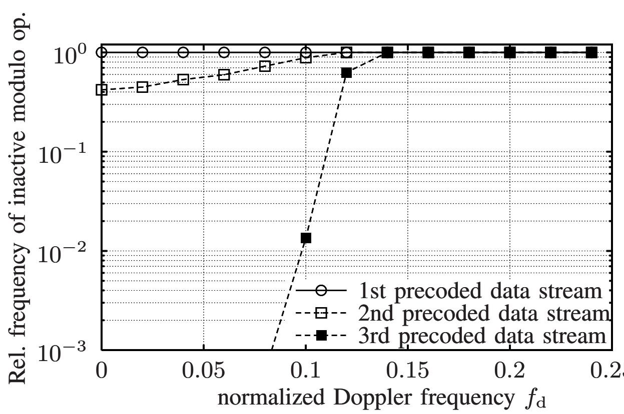

Figure 3. Relative frequency of inactive modulo operation at transmitter vs. max. Doppler frequency at $P_{\mathrm{T}}/\sigma_n^2$ of 20 dB.

D. Linear Precoding

If the modulo operator is inactive in the case of S-CSI, the THP is a linear precoder with parameters (22). Thus, $\boldsymbol{w}[n]$ depends linearly on $\boldsymbol{s}_{\mathrm{d}}[n]$ (Fig. 1)

$$\boldsymbol{w}[n] = (\boldsymbol{I}_K - \boldsymbol{F})^{-1}\boldsymbol{s}_{\mathrm{d}}[n] \tag{23}$$

and we set $\boldsymbol{\Pi}^{(\mathcal{O})} = \boldsymbol{I}_K$ as ordering is meaningless now. The linear precoder for P-CSI is identical to THP optimization with $\boldsymbol{F} = \boldsymbol{0}_{K\times K}$ or can be derived using (15), (16), and (23) assuming an *inactive* modulo operator

$$\boldsymbol{p}_{\mathrm{lin},k} = \beta_{\mathrm{lin}}\left(\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{H}_q|\boldsymbol{y}_q] + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\mathrm{E}[\boldsymbol{H}_q^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_q]\boldsymbol{e}_k. \tag{24}$$

For S-CSI with (22) it converges to

$$\boldsymbol{p}_{\mathrm{lin},k} = \beta_{\mathrm{lin}}\left(\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}} + \frac{K\sigma_n^2}{P_{\mathrm{T}}}\boldsymbol{I}_M\right)^{-1}\boldsymbol{U}_{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{e}_k, \tag{25}$$

which shows a close relation to traditional beamforming solutions. Note, that the transmitter's channel model is determined largely by the system's concept for transmitting the downlink training sequence with spatial filters $\boldsymbol{q}_k$ (Sec. II).

E. Convergence of THP to Linear Precoding

The previous two subsections show that THP for S-CSI can be identical to beamforming (25). Thus, with the novel THP design paradigm (11) it is ensured that the performance is always better than or—in the worst case—equal to a linear design of the transmitter. This has important consequences for the relevance of THP in wireless systems: For low mobile velocities (low Doppler frequencies) we obtain the excellent THP performance, and for high velocities performance is upper bounded by a linear design or beamformer.

THP is considered an implementable one-dimensional version of "writing-on-dirty-paper" [4], [5] in case of known interference at the transmitter, which is given for C-CSI. In the non-linear part of THP previously precoded data streams, which present known interference in case of C-CSI, are fed back, subtracted, and modulo reduced. For S-CSI the interference at the receivers is unknown besides its statistic (similar to additive noise at the receiver) and cannot be removed. Interference at the receivers can only be *avoided* by linear precoding. In the light of this relation convergence of THP to linear precoding for S-CSI could be expected as "writing-on-dirty-paper" is not possible anymore. Our THP for P-CSI is the first implementation of this (ad-hoc) information theoretic conjecture. Surprisingly, in scenarios (results not shown here) with a bad condition number of $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}}$ THP remains non-linear and gains in performance over linear precoding, i.e., is a non-linear "beamformer" for S-CSI.

V. Performance Evaluation

Simulation parameters: QPSK data symbols, $M = 4$ transmit antennas in a uniform linear array (half wavelength spacing), and $K = 3$ receivers are used. Transmitter and receivers are modeled as in Sec. II. All complex Gaussian channel coefficients have the same Jakes power spectrum with maximum Doppler frequency f_{d} (normalized to the slot period). The azimuth directions of the receivers' channels are uniformly distributed within 3° centered at the means $[-15°, 0°, 15°]$. Walsh-Hadamard sequences of length $N = 32$ are used for training in the uplink and the received training sequences from $Q = 5$ previous uplink slots are considered for prediction. $\boldsymbol{q}_k^*$ is given by the principal eigenvector of the conditional correlation matrix $\mathrm{E}\left[\boldsymbol{h}_{k,q}\boldsymbol{h}_{k,q}^{\mathrm{H}}|\boldsymbol{y}_q\right]$ (12). For all THP results suboptimum spatial (re)ordering is applied.

Results: Figures 4 and 5 show the average uncoded BER vs. Doppler frequency f_{d} and $P_{\mathrm{T}}/\sigma_n^2$, respectively. Aiming at a THP design which *always* performs better than linear precoding also the performance for linear precoding based on P-CSI (24) and for S-CSI (25) are given as a reference. From Fig. 4 we see that performance of THP with C-CSI (19), (20) is only achievable for very small Doppler frequencies (up to $f_{\mathrm{d}} = 0.04$). Traditional THP design (9), (10) degrades quickly with increasing Doppler frequency, i.e., larger parameter errors, and already performs worse than linear precoding (S-CSI) for $f_{\mathrm{d}} \geq 0.09$. It saturates at a BER of 0.5. For our novel THP optimization based on P-CSI and a CM estimate of the cost function (cf. Eq. 17) we show two versions: one as described above, the second *omits* the modulo operation at a receiver (modulo adaptation = "mod. adapt."), if its own and preceding precoding operations were performed linearly, i.e., with inactive modulo operation. We see that the novel THP optimization—including adaptation of the modulo operator at the receivers ("mod. adapt.")— always performs better than linear precoding and *converges* to linear precoding with S-CSI at high f_{d}. If the modulo operators are not adapted, performance is worse than linear precoding with S-CSI due to the increased number of neighbors (a.k.a. "modulo-loss").

Here, we considered a system with alternating up- and downlink slots. Degradation of THP is shifted to lower Doppler frequencies in a system with asymmetric slot structure, i.e., with more down- than uplink slots. In such a scenario it is even more important to be able to design THP with a smooth transition to linear precoding in case of S-CSI. Note that the alternative to a continuous transition would be switching between linear precoding and traditional THP design, which performs worse (cf. Fig. 4) and is usually much harder to design with the problem of defining and adapting switching points.

At a Doppler frequency of $f_{\mathrm{d}} = 0.08$ a BER of 10^{-2} is achieved with $2.5\,\mathrm{dB}$ less transmitter power for THP based on P-CSI compared to linear precoding with P-CSI and a $5\,\mathrm{dB}$ gain over the traditional design at 10^{-2} (Fig. 5). To achieve the same BER traditional design results in an increased transmit power compared to linear precoding.

VI. Conclusions

A novel Tomlinson-Harashima precoding optimization for partial CSI was introduced, which results in a continuous transition from the case of complete CSI to statistical CSI. In case of statistical CSI THP reduces to linear precoding. Thus, we ensure a THP performance, which is always better than linear precoding and continuously adapts to the amount of CSI available. With previously known designs THP performs significantly worse than linear precoding for medium to large parameter errors, i.e., high Doppler frequencies.

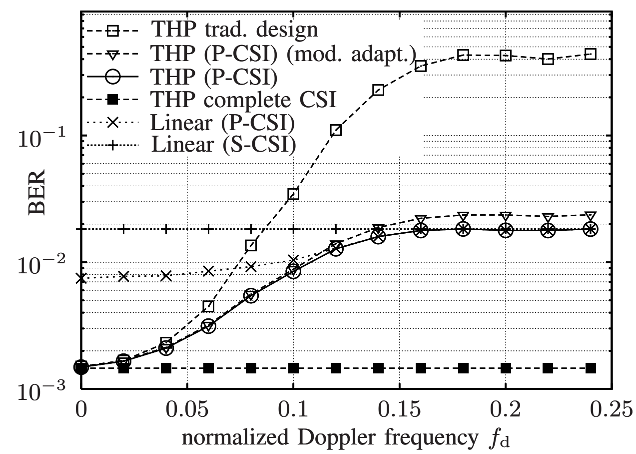

Figure 4. Uncoded BER vs. Doppler freq. at $P_{\mathrm{T}}/\sigma_n^2$ of $20\,\mathrm{dB}$

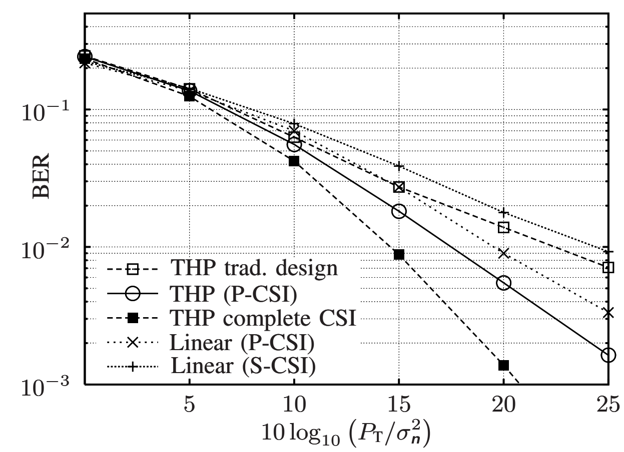

Figure 5. Uncoded BER vs. $P_{\mathrm{T}}/\sigma_n^2$ at max. Doppler frequency of $f_{\mathrm{d}} = 0.08$

References

[1] G. D. Forney and M. V. Eyuboglu. Combined equalization and coding using precoding. *IEEE Com. Mag.*, 29(12):25–34, Dec. 1991.

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

[3] 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.

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

[5] U. Erez, S. Shamai, and R. Zamir. Capacity and Lattice-Strategies for Cancelling Known Interference. In *Int. Symp. Inform. Theory and its Appl.*, Nov. 2000.

[6] 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.

[7] 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.

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

[9] 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*.

[10] A. P. Liavas. Tomlinson-Harashima Precoding with Partial Channel Knowledge. *IEEE Trans. Commun.*, 53(1):5–9, Jan. 2005.

[11] F. A. Dietrich, M. Joham, and W. Utschick. Joint Optimization of Pilot Assisted Channel Estimation and Equalization applied to Space-Time Decision Feedback Equalization. In *Proc. of ICC*, South Korea, May 2005.

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

[13] K. S. Miller. *Complex Stochastic Processes*. Addison-Wesley, 1st edition, 1974.