Robust Tomlinson–Harashima Precoding for the Wireless Broadcast Channel

Frank A. Dietrich, *Student Member, IEEE*, Peter Breun, and Wolfgang Utschick, *Senior Member, IEEE*

***Abstract*—Previous designs of Tomlinson–Harashima precoding (THP) for the broadcast channel with multiple (cooperative) transmitters and noncooperative (decentralized) scalar receivers assume complete channel state information (CSI) at the transmitter. Due to its nonlinearity, THP is more sensitive to errors in CSI at the transmitter than linear precoding, which may be large in wireless communications. Thus, THP performance degrades severely—and it is even outperformed by corresponding linear precoders. We propose a novel robust optimization for THP with partial CSI based on mean-square error (MSE), which performs a conditional mean (CM) estimate of the cost function including a novel model of the receivers. This enables a smooth transition of robust THP from complete, via partial, to statistical CSI. The performance is now always superior or in the worst case equal to corresponding linear precoding techniques. With these features, the proposed robust THP is very attractive for application in the wireless broadcast channel.**

***Index Terms*—Broadcast channel, channel estimation, robust optimization, Tomlinson–Harashima precoding.**

I. Introduction

A communication scenario with multiple cooperating transmitters, which can perform a joint preprocessing of the signals to be transmitted, and multiple decentralized receivers, which can only process the received signals independently, is referred to as "(MIMO) broadcast channel." We consider receivers with a scalar receive signal, e.g., from one antenna element. The broadcast channel scenario occurs in wireline (e.g., [3]; far-end crosstalk) as well as in wireless communications (communication from an access point to multiple mobile terminals), which is considered here. Precoding the signals at the transmitter aims at reducing interference between the parallel transmitted data streams.

A. Channel State Information At the Transmitter

Regarding the choice of a precoding technique, the key issue is the degree of channel state information (CSI) available at the transmitter. *Complete* CSI (C-CSI) refers to perfect knowledge of the channel realization $\boldsymbol{h}$ in the current time slot. Theoretically, it requires a reciprocal communication channel, which is available in time-division duplex (TDD) systems, or a (high-rate) feedback from the receivers (closed-loop mode). If the channel is time variant, as in wireless communication systems with mobile receivers or environment, complete CSI is never available at the transmitter and the CSI is outdated. Moreover, it is noisy due to a limited number of training symbols. We refer to this situation as *partial* CSI (P-CSI), which we model with the conditional probability density function (pdf) $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_\mathrm{T}}(\boldsymbol{h}|\boldsymbol{y}_\mathrm{T})$ of the channel parameters $\boldsymbol{h}$ given observations $\boldsymbol{y}_\mathrm{T}$ of the training sequence from the uplink (TDD system). In the extreme case, the channel parameters are statistically independent of the observations $\boldsymbol{y}_\mathrm{T}$, i.e., $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_\mathrm{T}}(\boldsymbol{h}|\boldsymbol{y}_\mathrm{T}) = \mathrm{p}_{\boldsymbol{h}}(\boldsymbol{h})$, which we call *statistical* CSI (S-CSI). This situation occurs for high Doppler frequencies, i.e., large velocities. These three categories of CSI are summarized in Table I.

TABLE I
CATEGORIES OF CHANNEL STATE INFORMATION AT THE TRANSMITTER AND THEIR DEFINITIONS

CSI Category	Knowledge of	Characterization by
Complete CSI (C-CSI)	channel realization	$\boldsymbol{h}$
Partial CSI (P-CSI)	conditional channel PDF	$\mathrm{p}_{\boldsymbol{h}\|\boldsymbol{y}_\mathrm{T}}(\boldsymbol{h}\|\boldsymbol{y}_\mathrm{T})$
Statistical CSI (S-CSI)	channel PDF	$\mathrm{p}_{\boldsymbol{h}}(\boldsymbol{h})$

B. Previous and Related Work

Linear precoding techniques have a long history and are often referred to as pre-equalization in case of C-CSI and beamforming for S-CSI. A survey of mean-square error (MSE)-based techniques is given in [4] and references to signal-to-interference-and-noise ratio (SINR)-based techniques are given in, e.g., [5] and [6].

Nonlinear precoding shows large performance gains for C-CSI. It is motivated by information-theoretic considerations: Writing-on-dirty-paper precoding [7], [8] promises communication without performance loss compared with an additive white Gaussian noise (AWGN) channel if the interference is known at the transmitter and appropriate coding is used. *Tomlinson–Harashima precoding* (THP) is considered as the "one-dimensional implementation" of writing-on-dirty-paper [9]. It was first introduced by Tomlinson [10] and Harashima [11] for equalization of a frequency-selective channel and later for point-to-point MIMO communication (for more references, see [12]–[14]).

THP was proposed for the *broadcast* channel, i.e., point-to-multipoint communication, by Ginis *et al.* [3], Yu *et al.* [9], and

Manuscript received August 11, 2005; revised March 27, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athina Petropulu. This work was supported by the Deutsche Forschungsgemeinschaft under grant Ut-36/3. Parts of the manuscript were published in *Proceedings of the Sixteenth IEEE Symposium on Personal, Indoor, and Mobile Radio Communications*, Berlin, Germany, 2005, and *Proceedings of the IEEE Global Telecommunications (IEEE GLOBECOM) Conference*, St. Louis, MO, 2005.

The authors are with the Associate Institute for Signal Processing, Munich University of Technology (TUM), 80290 Munich, Germany (e-mail: Dietrich@nws.ei.tum.de; Breun@nws.ei.tum.de; utschick@tum.de).

Digital Object Identifier 10.1109/TSP.2006.885713

Fischer *et al.* [13], [15]. All three describe zero-forcing techniques for frequency-flat channels. THP minimizing the MSE (MMSE-THP) based on finite impulse response filters for frequency-selective channels—including optimization of the precoding order and latency time—was first introduced by Joham *et al.* [16]. More details and references can be found in [17] and [18]. Efficient algorithms for reducing the computational complexity of filter and precoding order computation based on the symmetrically permuted Cholesky factorization are given in [19] and [20].

In *wireless* communications, particularly, in a system with mobility, i.e., a *time-variant* channel, the assumption of C-CSI or complete knowledge of the channel state with a small (negligible) error at the transmitter is not valid (Section I-A). Conventionally, THP is still optimized assuming C-CSI. However, with increasing time variance of the channel, the performance of this traditional THP design degrades significantly as shown in Section VI.

To incorporate P-CSI in the optimization a Bayesian framework using the conditional pdf (Table I) can be applied. A similar approach is static stochastic programming based on the statistics of the estimation error [21]. These methods have been used successfully in general signal processing (e.g., [22] and [23]) and signal processing for communications (e.g., [24]–[27]).

A first *robust optimization* for zero-forcing THP in the broadcast channel assuming erroneous CSI and no specific receiver processing was first presented in [28], including an MMSE prediction of the channel parameters. For THP based on P-CSI with *cooperative* receivers (not considered here), a similar approach is given in [29] for a single-input single-output system, and a heuristic solution was proposed by [30] for a MIMO system; both do not include prediction of the parameters; for S-CSI, a solution is proposed by [31].

To optimize MMSE linear precoding based on S-CSI Joham *et al.* [32], [33] proposed to incorporate the rake receiver in direct-sequence code-division multiple-access (DS-CDMA) (channel matched filter) into the system model at the transmitter and find a solution based on statistically equivalent signals for rank-one correlation matrices (extended to general rank signals in [34]). A more systematic direct optimization of this problem for P-CSI (and S-CSI) was introduced in [35] by the authors modeling the receivers' signal processing with a simple phase correction, which is also presented by [36] for S-CSI in a more heuristic fashion. This novel model is introduced in more detail in this paper for THP (Section III).

C. Contributions and Outline

Our main contribution is the robust optimization of THP based on MSE for P-CSI. It relies on two paradigms:

1) From the point of view of optimization, the true channel is a random variable in case of P-CSI. Thus, a Bayesian paradigm is employed for optimization and the conditional mean (CM) estimate of the random cost function is performed. The solution can be given explicitly.
2) A novel model for the receivers' signal processing capabilities at the transmitter is presented. It models the receivers' CSI explicitly, which is more accurate than the transmitter's CSI in general. An analytical solution is given for a suboptimum choice of this novel receiver model.

Combining both paradigms yields a performance of THP, which is always superior to corresponding linear precoding (based on MSE), whereas traditional THP's uncoded bit error rate (BER) saturates at 0.5 for high Doppler frequency—much higher than for linear precoding (e.g., classical beamforming). Moreover, a framework to deal with the asymptotic cases of C-CSI and S-CSI is provided: Optimization criteria for pre-equalization, which are typically based on C-CSI, can now be applied to the case of S-CSI. In the regime of S-CSI, traditionally beamforming approaches are considered, which are based on different optimization criteria. Thus, this framework can be summarized by the motto "When pre-equalization meets beamforming." Although a frequency flat channel model is considered here, the extension to frequency-selective channels is rather straightforward following [37].

In Section II, the system models of the downlink data, downlink training, and uplink training channel are described. Section III introduces a novel approach for the transmitter to model the receivers' signal processing capabilities. The robust THP optimization based on P-CSI is presented in Section IV, including optimization of the precoding order and with explicit solutions for different receiver models. Section V discusses the transition of robust THP to the asymptotic cases of C-CSI and S-CSI. The sensitivity and the performance of traditional THP and robust THP is compared in Section VI. In the Appendixes, a detailed derivation of robust THP is given, including a derivation of the CM estimate of the cascade of channel and a simple choice of the novel receiver model.

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 $\mathrm{E}_{\boldsymbol{\mathsf{b}}}[\bullet]$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^*$, $(\bullet)^{\mathrm{H}}$, and $\mathrm{tr}(\bullet)$ stand for expectation with regard to $\boldsymbol{\mathsf{b}}$, transpose, complex conjugate, 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, and $\mathrm{diag}[a_i]_{i=1}^N$ is the diagonal matrix with a_i being the ith diagonal element. $\boldsymbol{e}_i$ is the ith column of an $N \times N$ identity matrix $\boldsymbol{I}_N$. The $M \times N$ matrix of zeros is $\boldsymbol{0}_{M\times N}$. The covariance matrix of $\boldsymbol{\mathsf{b}}$ is $\boldsymbol{C}_{\boldsymbol{\mathsf{b}}} = \mathrm{E}_{\boldsymbol{\mathsf{b}}}[(\boldsymbol{\mathsf{b}} - \mathrm{E}_{\boldsymbol{\mathsf{b}}}[\boldsymbol{\mathsf{b}}])(\boldsymbol{\mathsf{b}} - \mathrm{E}_{\boldsymbol{\mathsf{b}}}[\boldsymbol{\mathsf{b}}])^{\mathrm{H}}]$. The cross-covariance matrix is defined as $\boldsymbol{C}_{\boldsymbol{\mathsf{ab}}} = \mathrm{E}_{\boldsymbol{\mathsf{a}},\boldsymbol{\mathsf{b}}}[(\boldsymbol{\mathsf{a}} - \mathrm{E}_{\boldsymbol{\mathsf{a}}}[\boldsymbol{\mathsf{a}}])(\boldsymbol{\mathsf{b}} - \mathrm{E}_{\boldsymbol{\mathsf{b}}}[\boldsymbol{\mathsf{b}}])^{\mathrm{H}}]$.

II. System Model

A vector broadcast channel with M cooperative transmit antennas and K noncooperative (decentralized) receivers (users) is considered. Communication in uplink and downlink takes place in the TDD mode (Fig. 1). Thus, we consider data transmission in the downlink, where all K receivers are served simultaneously in one downlink time slot and are precoded spatially using THP. Reciprocity is assumed for uplink and downlink (UL and DL) channels, i.e., $\boldsymbol{H}_{\mathrm{DL}} = \boldsymbol{H}_{\mathrm{UL}}^{\mathrm{T}} \in \mathbb{C}^{K\times M}$. We consider data transmission in a fixed time slot with index q^{tx}, whereas the symbol index n always denotes a relative index within a downlink or uplink slot. Thus, assuming time slots with N_{slot}

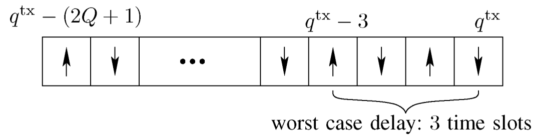


Fig. 1. Example for a TDD slot structure: Symmetric allocation of time slots to uplink ("↑") and downlink ("↓") for data transmission in a fixed time slot with index q^{tx} and assumption of worst-case delay of three slots for this allocation (see Section VI).

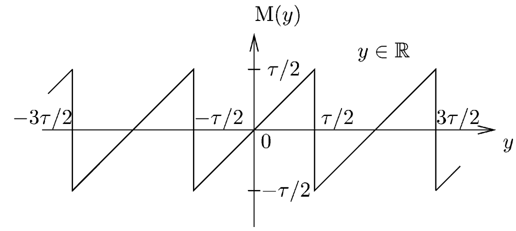


Fig. 2. Graph of the modulo-operator for real-valued arguments.

symbols, the absolute symbol index in time slot q is given by $N_{\mathrm{slot}}(q-1)+n$.

A. Downlink Data Channel

Data symbols $\boldsymbol{d}[n] = [d_1[n],\ldots,d_K[n]]^{\mathrm{T}} \in \mathbb{B}^K$ with $\mathrm{E}_{\boldsymbol{d}}[\boldsymbol{d}[n]\boldsymbol{d}[n]^{\mathrm{H}}] = \boldsymbol{I}_K$ and modulation alphabet $\mathbb{B}$ are precoded with THP (Fig. 3).[1] First, they are reordered using 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 of parameters on the specific precoding order $\mathcal{O}$ is denoted by the superscript $(\mathcal{O})$. Next, they are sequentially precoded starting with data stream $d_k[n]$, if the $(1,k)$th element of $\boldsymbol{\Pi}^{(\mathcal{O})}$ is one. Nonlinear precoding requires a modulo operator $\mathrm{M}(\bullet)$ at the transmitter and receivers, which is defined as (Fig. 2)

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

It is defined elementwise for a vector argument. Common choices for τ are $\tau = 2\sqrt{2}$ when $\mathbb{B}$ is the set of quadrature phase-shift keying (QPSK) symbols and $\tau = 8/\sqrt{10}$ in case of rectangular 16-QAM and the symbol variance is one. To ensure spatial causality for a realizable feedback loop, the *feedback filter* $\boldsymbol{F} = [\boldsymbol{f}_1,\ldots,\boldsymbol{f}_K] \in \mathbb{C}^{K\times K}$ with columns $\boldsymbol{f}_k$ is lower triangular with zero diagonal. Then, the output of the modulo operator $\boldsymbol{w}[n] \in \mathbb{C}^K$ is linearly precoded with $\boldsymbol{P} = [\boldsymbol{p}_1,\ldots,\boldsymbol{p}_K] \in \mathbb{C}^{M\times K}$. The data symbols in $\boldsymbol{d}[n]$ are precoded independently for each time step within a time slot with fixed index q^{tx}. The resulting signal $\boldsymbol{P}\boldsymbol{w}[n]$ is transmitted over the downlink channel $\boldsymbol{H} \triangleq \boldsymbol{H}[q^{\mathrm{tx}}]$ to K receivers, where the individual vector channels $\boldsymbol{h}_k^{\mathrm{T}}$ form the rows of

$$\boldsymbol{H} = [\boldsymbol{h}_1,\ldots,\boldsymbol{h}_K]^{\mathrm{T}} \triangleq \boldsymbol{H}[q^{\mathrm{tx}}] \in \mathbb{C}^{K\times M}. \quad (2)$$

The channel is assumed constant during one time slot ("block-fading"). The (noncooperative) receivers are modeled as $\tilde{\boldsymbol{G}} = \mathrm{diag}[\tilde{g}_k]_{k=1}^K \in \mathbb{C}^{K\times K}$. Including additive white complex Gaussian noise $\boldsymbol{n}[n] = [n_1[n],\ldots,n_K[n]]^{\mathrm{T}} \sim \mathcal{N}_{\mathrm{c}}(\boldsymbol{0},\sigma_n^2\boldsymbol{I}_K)$, the output of the modulo operator at the receivers before decision reads as (Fig. 3)

$$\tilde{\boldsymbol{d}}[n] = \mathrm{M}\left(\tilde{\boldsymbol{G}}\boldsymbol{H}\boldsymbol{P}\boldsymbol{w}[n] + \tilde{\boldsymbol{G}}\boldsymbol{n}[n]\right) \in \mathbb{C}^K. \quad (3)$$

[1]To have a concise notation, we define $\boldsymbol{d}[n] \triangleq \boldsymbol{d}[N_{\mathrm{slot}}(q^{\mathrm{tx}}-1)+n]$. This applies also to all other signals related to data transmission in Figs. 3 and 6.

B. Downlink Training Channel

The receivers' channel knowledge is determined by the linear precoding $\boldsymbol{Q} = [\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_P] \in \mathbb{C}^{M\times P}, P \leq K$ (Fig. 4) used for the downlink training sequences, which are transmitted orthogonally to the data (e.g., time-multiplexed). Precoding $\boldsymbol{Q}$ for the training sequences offers additional degrees of freedom in system design, which is not discussed here. As an example, a simple choice is introduced and discussed in the next Section. As every receiver only needs to estimate one complex coefficient $\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k$, we assume perfect knowledge of $\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k$ at the receivers.

This precoding is mainly determined by the system's standardized specification, i.e., the number of training sequences P available. There are different concepts for training symbol based channel estimation in the downlink. If $P = K$ receiver specific (dedicated) training sequences are available, each sequence is precoded with an individual spatial filter $\boldsymbol{q}_k \in \mathbb{C}^M$. Obviously, choosing $\boldsymbol{q}_k$ as a function of the THP parameters $\boldsymbol{P}$ and $\boldsymbol{F}$ leads to a more accurate knowledge at the receiver about the compound channel consisting of THP and the channel $\boldsymbol{H}[q^{\mathrm{tx}}]$. Thus, the receivers could get information about the preprocessing of the data signal at the transmitter. In some systems, a fixed number (independent of K) of P sequences are defined, e.g., in a "grid-of-beams" approach, where P distinct vector precoders $\boldsymbol{q}_k$, $k \in \{1,\ldots,P\}$, are given by predefined fixed beamforming vectors and the user may select the most relevant estimate (cf. references in [38] and [39]). If only one *common* training sequence ($P = 1$) is provided, it is transmitted with $\boldsymbol{q}_k = \boldsymbol{q}$, e.g., over the first antenna $\boldsymbol{q} = \boldsymbol{e}_1$.

C. Uplink Training Channel

In a TDD system, the *channel parameters for optimizing* the THP parameters $\boldsymbol{F}$ and $\boldsymbol{P}$ for transmitting data in time slot q^{tx} (Fig. 1) can be estimated from the received training signal in previous uplink slots.

N training symbols per receiver $\boldsymbol{s}[n] \in \mathbb{C}^K$ are available in one *uplink* slot. The receive training signal is (Fig. 5)

$$\boldsymbol{y}[N_{\mathrm{slot}}(q-1)+n] = \boldsymbol{H}[q]^{\mathrm{T}}\boldsymbol{s}[n] + \boldsymbol{v}[N_{\mathrm{slot}}(q-1)+n] \in \mathbb{C}^M \quad (4)$$

for $n \in \{1,\ldots,N\}$ with additive white complex Gaussian noise $\boldsymbol{v}[N_{\mathrm{slot}}(q-1)+n] \sim \mathcal{N}_{\mathrm{c}}(\boldsymbol{0},\sigma_v^2\boldsymbol{I}_M)$. Here, reciprocity of the channel in a TDD system is assumed. Collecting all N training symbols in the columns of $\boldsymbol{S}'_{\mathrm{p}} \in \mathbb{C}^{K\times N}$, we obtain

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

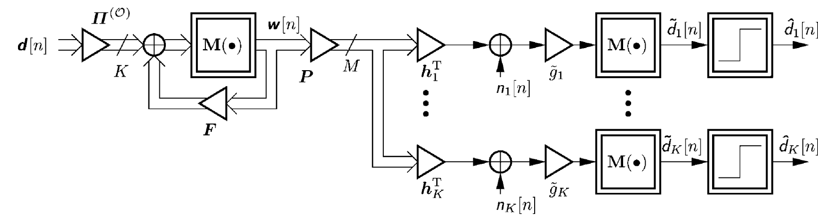


Fig. 3. THP for downlink data transmission with M transmit antennas and K noncooperative receivers: Representation with modulo operators in the time slot with fixed index q^{tx} and channel realization $\boldsymbol{H}$ (cf. (2)).

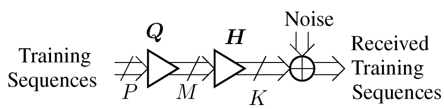


Fig. 4. Downlink training channel: Linear Precoding with $\boldsymbol{Q}$ for P training sequences defined in the communication system ($\boldsymbol{H} \triangleq \boldsymbol{H}[q^{\mathrm{tx}}]$).

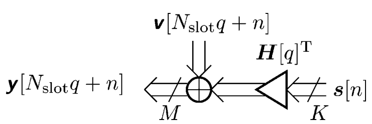


Fig. 5. Uplink training channel (4) in time slot $q = q^{\mathrm{tx}} - \ell$, i.e., ℓ slots before data transmission in time slot q^{tx}.

where $\boldsymbol{h}[q] = \mathrm{vec}[\boldsymbol{H}[q]^{\mathrm{T}}] = [\boldsymbol{h}_1[q]^{\mathrm{T}}, \ldots, \boldsymbol{h}_K[q]^{\mathrm{T}}]^{\mathrm{T}}$. To obtain a more accurate estimate of the channel realization at time q^{tx}, training signals from Q previous uplink slots are considered, which are outdated by $\ell_i, i \in \{1, \ldots, Q\}$, slots. Thus, the total observation is (Fig. 1)

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

with $\boldsymbol{h}_{\mathrm{T}} = [\boldsymbol{h}[q^{\mathrm{tx}} - \ell_1]^{\mathrm{T}}, \boldsymbol{h}[q^{\mathrm{tx}} - \ell_2]^{\mathrm{T}}, \ldots, \boldsymbol{h}[q^{\mathrm{tx}} - \ell_Q]^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{QMK}$, $\boldsymbol{y}_{\mathrm{T}} = [\bar{\boldsymbol{y}}[q^{\mathrm{tx}} - \ell_1]^{\mathrm{T}}, \bar{\boldsymbol{y}}[q^{\mathrm{tx}} - \ell_2]^{\mathrm{T}}, \ldots, \bar{\boldsymbol{y}}[q^{\mathrm{tx}} - \ell_Q]^{\mathrm{T}}]^{\mathrm{T}}$ ($\boldsymbol{v}_{\mathrm{T}}$ equivalently), and $\boldsymbol{S} = \boldsymbol{I}_Q \otimes \boldsymbol{S}_{\mathrm{p}}'^{\mathrm{T}} \otimes \boldsymbol{I}_M$.

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 $\boldsymbol{C}_{\boldsymbol{h}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}[q]\boldsymbol{h}[q]^{\mathrm{H}}]$, which is block diagonal assuming channels from different receivers are uncorrelated, i.e., $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k[q]\boldsymbol{h}_{k'}[q]^{\mathrm{H}}] = \boldsymbol{C}_{\boldsymbol{h}_k}\delta_{k,k'}$. The kth column of $\boldsymbol{H}[q]^{\mathrm{T}}$ is $\boldsymbol{h}_k[q] \sim \mathcal{N}_{\mathrm{c}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$. As an example, we assume identical autocorrelation $r[i]$ for all elements of $\boldsymbol{h}[q]$: $r[i]$ is defined by $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k[q]\boldsymbol{h}_k[q-i]^{\mathrm{H}}] = r[i]\boldsymbol{C}_{\boldsymbol{h}_k}$ for all k and a time difference of i slots and scaled to $r[0] = 1$. Thus, we can write $\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_{\mathrm{T}}\boldsymbol{h}_{\mathrm{T}}^{\mathrm{H}}] = \boldsymbol{C}_{\mathrm{T}} \otimes \boldsymbol{C}_{\boldsymbol{h}}$. For a symmetric slot structure as depicted in Fig. 1, $\boldsymbol{C}_{\mathrm{T}}$ is Toeplitz with first column $[r[0], r[2], \ldots, r[2Q-2]]^{\mathrm{T}}$.

Throughout the paper, we assume that first- and second-order channel and noise statistics are given.

III. Models for Signal Processing at Receivers

A. Conventional Receiver Model

Conventionally, for THP optimization, the receivers are assumed to perform only a real-valued scaling [17]. Generally, every receiver can be modeled with a different scaling factor [40], [41], which leads to a closed-form solution for zero-forcing THP, but not for the MMSE-THP design [40]. It can be seen as a model for the automatic gain control at the receivers.

As we consider MMSE-THP in Section IV and aim at an explicit solution, we simplify this model to $\tilde{g}_k = \beta^{-1}$ as in [17], [37]

$$\tilde{\boldsymbol{G}} = \beta^{-1}\boldsymbol{G} = \beta^{-1}\boldsymbol{I}_K. \tag{6}$$

From the point of view of THP optimization, due to a real-valued scaling at the receivers, an average transmit power constraint can be included into the optimization problem.

Using this receiver model for THP optimization, we implicitly assume that the receivers have the same CSI as the transmitter, i.e., no additional channel knowledge is available, e.g., about the phase of $\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k$. This is not a restriction in case of C-CSI at the transmitter but yields a very conservative design of THP in case of P-CSI (cf. Section IV).

B. Novel Receiver Model

The conventional model makes a rather restrictive assumption about the receivers' channel knowledge, although the transmitter is aware that training sequences are transmitted to the receivers (Section II-B). Of course, the transmitter also knows the filter vector $\boldsymbol{q}_k$ used for precoding the training sequence to receiver k. Thus, a safe assumption is that the receiver will also exploit this knowledge about the kth receiver's channel $\boldsymbol{h}_k$ projected onto $\boldsymbol{q}_k^*$, i.e., knowledge about $\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k$. Moreover, training sequences are usually provided to enable coherent detection at the receivers, which requires phase and amplitude of the channel.

Generally, the receivers should be modeled as $\tilde{g}_k = \beta^{-1}\mathrm{g}(\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k)$, where the real-valued scaling β is an amplitude scaling and provides the necessary degree of freedom to allow for the transmit power constraint. The (novel) scalar function $\mathrm{g}(\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k)$ models the kth receiver's CSI. This yields a novel receiver model explicitly depending on the channel realization $\boldsymbol{H}$

$$\tilde{\boldsymbol{G}} = \beta^{-1}\boldsymbol{G} = \beta^{-1}\mathrm{diag}\left[\mathrm{g}\left(\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k\right)\right]_{k=1}^{K}. \tag{7}$$

We assume perfect knowledge of $\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k$ at the receivers for simplicity, although Gaussian channel estimation errors can be easily included in optimization knowing the number of training symbols and the noise variance at receiver. This general receiver model is used in the sequel until Section IV-C.

As an example, the following choice of this general receiver model is used in Section IV-E to simplify the general approach and to obtain further insights in the solution. The *simplest* and clearly suboptimum model for the receivers' CSI is a correction of the phase choosing the scalar function g as

$$g_k = \mathrm{g}(\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k) = \frac{\left(\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k\right)^*}{\left|\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k\right|}, \text{ i.e., } \boldsymbol{G} = \mathrm{diag}\,[g_k]_{k=1}^K. \quad (8)$$

In Section IV-E, we show that an analytical solution for this simple choice (8) of the novel receiver model (7) can be obtained.

Having introduced the receiver model, we can now further comment on the choice of $\boldsymbol{q}_k$. Although choosing $\boldsymbol{q}_k$ as a function of $\boldsymbol{P}$ and $\boldsymbol{F}$ is reasonable for precoding the downlink training sequence (Section II-B), we will assume precoders $\boldsymbol{q}_k$, which do not depend on $\boldsymbol{P}$ and $\boldsymbol{F}$. Else the receivers also depend on $\boldsymbol{P}$ or $\boldsymbol{F}$ and the optimization problems in Section IV-B can be solved in very special cases only—due to the difficult mathematical structure. Thus, in this case, the implemented precoding for the training channel needs to be approximated for THP optimization, e.g., choosing $\boldsymbol{q}_k$ as the complex conjugate principal eigenvector of the conditional correlation matrix $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}]$. This approximation provides enough information for THP optimization to ensure its robust performance and a meaningful solution for S-CSI.

In general, the actual receiver concept will differ from the transmitter's model. For example, if the receivers' actual signal processing g_k is more sophisticated than the transmitter's model, a conservative design of THP is obtained due to this mismatch. We want to emphasize that in the asymptotic case of S-CSI, it is essential to model the receivers' CSI explicitly to obtain a meaningful precoder solution.

IV. THP Optimization With Partial CSI at the Transmitter

THP optimization is based on the linear representation (Fig. 6) of the system in Fig. 3. In this standard approach, the modulo operators are substituted by the additive signals $\boldsymbol{a}[n] \in \tau(\mathbb{Z}^K + \mathrm{j}\mathbb{Z}^K)$ and $\tilde{\boldsymbol{a}}[n] \in \tau(\mathbb{Z}^K + \mathrm{j}\mathbb{Z}^K)$, which are functions of $\boldsymbol{w}[n]$ and $\tilde{\boldsymbol{d}}[n]$, respectively [12], [17]. In MMSE optimization of THP, which was shown to be superior to zero-forcing [16], the MSE between the new signals $\boldsymbol{b}[n]$ and $\tilde{\boldsymbol{b}}[n]$ in the linear representation (Fig. 6) is

$$\begin{aligned} \mathrm{Z}_{\mathrm{T}}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{H}) &= \mathrm{E}_{\boldsymbol{w},\boldsymbol{n}}[\|\boldsymbol{b}[n] - \tilde{\boldsymbol{b}}[n]\|_2^2] \\ \text{with } \boldsymbol{b}[n] &= \boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}}(\boldsymbol{I}_K - \boldsymbol{F})\boldsymbol{w}[n] \\ \tilde{\boldsymbol{b}}[n] &= \beta^{-1}\boldsymbol{G}\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{w}[n] + \boldsymbol{n}[n]\right) \end{aligned} \quad (9)$$

based on the current realization $\boldsymbol{H}$ of the channel parameters.[2]

As the realization $\boldsymbol{H}$ is not perfectly known to the transmitter, but previous (outdated) realizations are observed via $\boldsymbol{y}_{\mathrm{T}}$ (5), a stochastic model is employed to capture the transmitter's P-CSI. Hence, from the point of view of optimization, the channel $\boldsymbol{H}$ is a random variable, which is described by its conditional pdf $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}}(\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}})$. In Section II-C, the observation $\boldsymbol{y}_{\mathrm{T}}$ and channel parameters $\boldsymbol{h}$ are assumed to be complex Gaussian. Thus, the conditional pdf is also (complex) Gaussian with mean $\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}]$ and covariance matrix $\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}} = \mathrm{E}_{\boldsymbol{h}}[(\boldsymbol{h} - \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}})(\boldsymbol{h} - \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}})^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}]$.

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}} &= \hat{\boldsymbol{h}} = \boldsymbol{W}\boldsymbol{y}_{\mathrm{T}}, \\ \boldsymbol{W} &= \boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}} + \sigma_v^2\boldsymbol{I}_{MNQ})^{-1} \\ \boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}} &= \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{h}} = \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{W}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}}^{\mathrm{H}} \end{aligned} \quad (10)$$

where $\boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}^{\mathrm{H}}] = [r[\ell_1], r[\ell_2], \ldots, r[\ell_Q]] \otimes \boldsymbol{C}_{\boldsymbol{h}}$, and $\boldsymbol{h} = \mathrm{vec}(\boldsymbol{H}^{\mathrm{T}})$. $\boldsymbol{W}$ is the LMMSE estimator [42] for $\boldsymbol{h}$ given $\boldsymbol{y}_{\mathrm{T}}$.

Due to the P-CSI via the conditional pdf, the cost function (9) is a random variable, as well. In the next section, we first present the traditional approach for dealing with this random cost function and our new approach, which is applied together with the receiver models from Section III.

A. Traditional Optimization

Conventionally, the errors in $\hat{\boldsymbol{H}}$ are assumed to be negligible, and the estimate $\hat{\boldsymbol{H}}$ is used as if it was the true one, i.e., we set $\boldsymbol{H} = \hat{\boldsymbol{H}}$ in (9), leading to the optimization problem (Fig. 7)

$$\min_{\boldsymbol{P},\boldsymbol{F},\beta} \mathrm{Z}_{\mathrm{T}}(\boldsymbol{P},\boldsymbol{F},\beta;\hat{\boldsymbol{H}}) \ \text{ s.t. } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\right) \le P_{\mathrm{Tx}}, \quad \boldsymbol{F}: \text{ lower triangular, zero diagonal} \quad (11)$$

with average transmit power constrained by P_{Tx} and the constraint on $\boldsymbol{F}$ to ensure implementability. The solution using the Lagrange approach and the Karush–Kuhn–Tucker (KKT) conditions [17] is given by (with $\hat{G}_{\mathrm{N}} = \mathrm{tr}\left(\hat{\boldsymbol{G}}_{\mathrm{T}}\hat{\boldsymbol{G}}_{\mathrm{T}}^{\mathrm{H}}\right)$; cf. (14))[3]

$$\boldsymbol{p}_{\mathrm{T},k} = \beta_{\mathrm{T}}\left(\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O})} + \frac{\hat{G}_{\mathrm{N}}\sigma_n^2}{P_{\mathrm{Tx}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_{\mathrm{T},k}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k \quad (12)$$

$$\boldsymbol{f}_{\mathrm{T},k} = -\beta_{\mathrm{T}}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M} \\ \hat{\boldsymbol{B}}_{\mathrm{T},k}^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{T},k}. \quad (13)$$

The scalar β_{T} is chosen to satisfy the transmit power constraint $\mathrm{tr}(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}) \le P_{\mathrm{Tx}}$ 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 $\boldsymbol{\Pi}^{(\mathcal{O})}\hat{\boldsymbol{G}}_{\mathrm{T}}\hat{\boldsymbol{H}}$. The estimate is obtained from the linear MMSE (LMMSE) estimator in (10) and an assumption about the receivers' processing (Section III). For the

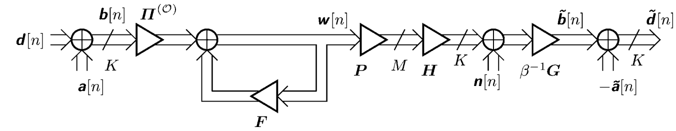


Fig. 6. Equivalent linear representation of Fig. 3 for THP optimization with $\boldsymbol{a}[n] \in \tau(\mathbb{Z}^K + \mathrm{j}\mathbb{Z}^K)$ and $\tilde{\boldsymbol{a}}[n] \in \tau(\mathbb{Z}^K + \mathrm{j}\mathbb{Z}^K)$.

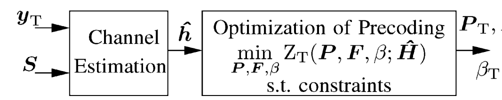


Fig. 7. Traditional optimization: Separate optimization of channel estimation and THP.

[2]The precoding order of the data streams described by $\boldsymbol{\Pi}^{(\mathcal{O})}$ is fixed in the sequel. Its optimization is discussed in Section IV-C.

[3]They are a special case of the derivations for the approach in Section IV-B as described in Appendix I, when omitting the conditional expectation and setting $\boldsymbol{H} = \hat{\boldsymbol{H}}$.

general novel receiver model (7) and its specific choice from (8), we have

$$\hat{\boldsymbol{G}}_{\mathrm{T}} = \mathrm{diag}\left[\mathrm{g}\left(\hat{\boldsymbol{h}}_k^{\mathrm{T}}\boldsymbol{q}_k\right)\right]_{k=1}^{K} = \mathrm{diag}\left[\frac{\left(\hat{\boldsymbol{h}}_k^{\mathrm{T}}\boldsymbol{q}_k\right)^*}{\left|\hat{\boldsymbol{h}}_k^{\mathrm{T}}\boldsymbol{q}_k\right|}\right]_{k=1}^{K}. \quad (14)$$

Optimization of the precoding order is similar to Section IV-C [17].

Remark: To obtain this solution (cf. (12) and (13)), it is common in the literature to make the assumption $\boldsymbol{C_w} = \mathrm{diag}[\sigma_{w_k}^2]_{k=1}^K$ [12], which is based on the observation that the outputs of the modulo operator are independent and uniformly distributed.

B. Conditional Mean Estimate of the Cost Function

As the value of the cost function (9) is a random variable, the true MSE is not known for optimization. Thus, the MSE has to be estimated from the observation $\boldsymbol{y}_{\mathrm{T}}$ using the *a priori* information $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}}(\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}})$. Employing the Bayesian paradigm, the best estimator in the mean-square sense is the CM estimator [42]. Thus, the CM estimate of (9), i.e., the expected MSE, is

$$\begin{aligned}
&\mathrm{Z_P}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_{\mathrm{T}})\\
&=\mathrm{E}_{\boldsymbol{h}}[\mathrm{Z_T}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{H})|\boldsymbol{y}_{\mathrm{T}}]\\
&=\mathrm{tr}\Big((\boldsymbol{I}_K-\boldsymbol{F})\boldsymbol{C_w}(\boldsymbol{I}_K-\boldsymbol{F})^{\mathrm{H}}\\
&\quad+\beta^{-2}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}]\sigma_n^2+\beta^{-2}\boldsymbol{P}\boldsymbol{C_w}\boldsymbol{P}^{\mathrm{H}}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]\Big)\\
&\quad-2\beta^{-1}\mathrm{Re}\Big\{\mathrm{tr}\Big(\boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]\boldsymbol{P}\boldsymbol{C_w}(\boldsymbol{I}_K-\boldsymbol{F})^{\mathrm{H}}\Big)\Big\}. \quad (15)
\end{aligned}$$

There is a significant difference between the traditional approach (11), which ignores estimation errors and simply plugs the CM estimate of $\boldsymbol{h}$ (10) into the cost function (9), and this novel approach. Assuming a complex Gaussian distribution $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}}(\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}})$, they are equivalent if and only if the cost function depends linearly on $\boldsymbol{h}$. In this case, the cost would also be complex Gaussian distributed, and due to linearity of the expectation, the solution is to plug the CM estimate of $\boldsymbol{h}$ into the cost function.

Here, the MSE cost function (9) is a nonlinear function of $\boldsymbol{h}$, i.e., its pdf is not Gaussian anymore. Applying this novel approach, which is systematic from an estimation-theoretic perspective, yields a solution with very different and desirable properties, as shown in Sections V and VI.

For a given receiver model (Section III), the CM estimate of the MSE can be expressed in terms of the moments (10) of the conditional pdf. In general, the new optimization problem for P-CSI at the transmitter reads (Fig. 8)

$$\begin{aligned}
&\min_{\boldsymbol{P},\boldsymbol{F},\beta}\mathrm{Z_P}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_{\mathrm{T}})\ \text{ s.t. } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{C_w}\boldsymbol{P}^{\mathrm{H}}\right)\le P_{\mathrm{Tx}},\\
&\qquad\qquad \boldsymbol{F}:\ \text{lower triangular, zero diagonal.} \quad (16)
\end{aligned}$$

In Appendix I, the complete solution for a given precoding order $\mathcal{O}$ is derived and can be written as ($\bar{G}_{\mathrm{N}}=\mathrm{tr}(\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}])$)

$$\boldsymbol{p}_{\mathrm{P},k}=\beta_{\mathrm{P}}\left(\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}}+\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}+\frac{\bar{G}_{\mathrm{N}}\sigma_n^2}{P_{\mathrm{Tx}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k$$

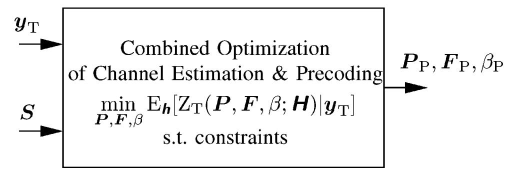


Fig. 8. Optimization based on partial CSI: Combined optimization of channel estimation and THP.

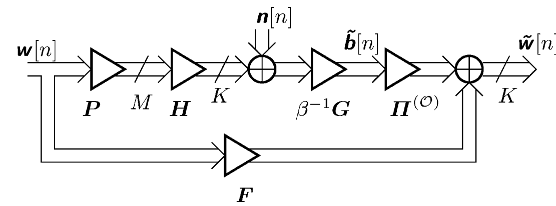


Fig. 9. Linear representation of THP optimization, which is equivalent to Fig. 6 if the error signal $\boldsymbol{w}[n]-\tilde{\boldsymbol{w}}[n]$ is used instead of $\boldsymbol{b}[n]-\tilde{\boldsymbol{b}}[n]$ for optimization.

$$\boldsymbol{f}_{\mathrm{P},k}=-\beta_{\mathrm{P}}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{P},k} \quad (17)$$

where β_{P} is chosen to satisfy the power constraint in (16) 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 $\boldsymbol{GH}$

$$\begin{bmatrix}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix}=\boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]\in\mathbb{C}^{K\times M}. \quad (18)$$

The nondiagonal loading matrix is the conditional covariance matrix of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}$

$$\begin{aligned}
\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}}&=\mathrm{E}_{\boldsymbol{h}}\left[(\boldsymbol{G}\boldsymbol{H}-\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}])^{\mathrm{H}}(\boldsymbol{G}\boldsymbol{H}-\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}])\,|\boldsymbol{y}_{\mathrm{T}}\right]\\
&=\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]-\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}]\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]. \quad (19)
\end{aligned}$$

Remark 1: Exploiting the statistics of the stochastic MSE (9) and optimally estimating the MSE—instead of optimally estimating the channel parameters—results in two significant differences of the novel solution for P-CSI (17) compared with the traditional solution (12) and (13): A nondiagonal loading with $\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}}$ is added in the inverse of $\boldsymbol{p}_{\mathrm{P},k}$ (17), and the CM estimate (18) of the overall channel $\boldsymbol{GH}$ is used instead of simply plugging the CM estimate $\hat{\boldsymbol{h}}$ of $\boldsymbol{h}$ (10) into the product $\boldsymbol{GH}$, as in Section IV-A. The nondiagonal loading is due to the regularization term in the cost function (43): $\beta^{-2}\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}}\boldsymbol{P}\boldsymbol{C_w}\right)$ regularizes a sum of weighted 2-norms of the columns of $\boldsymbol{P}$.

Remark 2: The block diagram in Fig. 9 yields the same optimization problem with cost function $\mathrm{E}_{\boldsymbol{h}}[\mathrm{E}_{\boldsymbol{w},\boldsymbol{n}}[\|\boldsymbol{w}[n]-\tilde{\boldsymbol{w}}[n]\|_2^2]|\boldsymbol{y}_{\mathrm{T}}]$ as in (16) based on Fig. 6, which can easily be seen by reinterpretation of (43). From this representation, we can directly interpret the effect of the feedback filter $\boldsymbol{f}_{\mathrm{P},k}$: The interference created by data stream k to the succeeding (determined by the precoding order $\mathcal{O}$) data streams is reduced; this interference, which is only partially known at the transmitter, is subtracted based on the CM estimate of $\boldsymbol{GH}$. This is implemented with the nonlinear modulo operation in

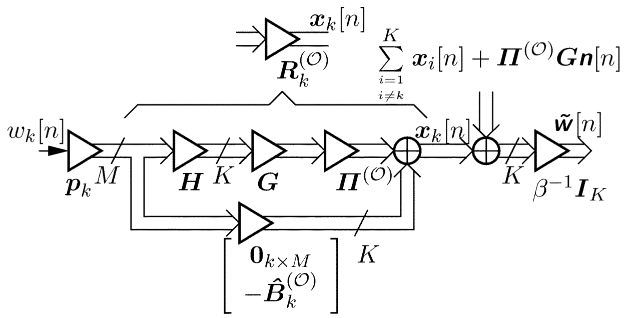


Fig. 10. Linear representation of THP optimization, which is equivalent to Fig. 9.

the feedback loop (Fig. 3). Applying $\boldsymbol{f}_{\mathrm{P},k}$ to $\mathrm{Z_P}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_\mathrm{T})$ yields the cost function for optimization of the linear part of THP, i.e., $\boldsymbol{P}$ and β,

$$\mathrm{Z_P}(\boldsymbol{P},\boldsymbol{F}_\mathrm{P},\beta;\boldsymbol{y}_\mathrm{T}) = \beta^{-2}\sum_{k=1}^{K}\sigma_{w_k}^2\boldsymbol{p}_k^\mathrm{H}\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{R}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{R}_k^{(\mathcal{O})}|\boldsymbol{y}_\mathrm{T}\right]\boldsymbol{p}_k$$
$$-2\beta^{-1}\mathrm{Re}\left[\sum_{k=1}^{K}\sigma_{w_k}^2\boldsymbol{e}_k^\mathrm{T}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}\boldsymbol{p}_k\right] + \mathrm{tr}[\boldsymbol{C}_{\boldsymbol{w}}] + \beta^{-2}\bar{G}_\mathrm{N}\sigma_n^2. \quad (20)$$

The residual channel (including the precoding order) after taking into account the effect of $\boldsymbol{F}_\mathrm{P}$ is

$$\boldsymbol{R}_k^{(\mathcal{O})} = \boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{G}\boldsymbol{H} - \begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_k^{(\mathcal{O})}\\ \boldsymbol{B}_k^{(\mathcal{O})} - \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix} \quad (21)$$

where $\boldsymbol{A}_k^{(\mathcal{O})}$ and $\boldsymbol{B}_k^{(\mathcal{O})}$ are defined similarly to (18) based on the true channel $\boldsymbol{G}\boldsymbol{H}$. The CM estimate of its Gram matrix is needed in the cost function (20) and given as

$$\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{R}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{R}_k^{(\mathcal{O})}|\boldsymbol{y}_\mathrm{T}\right] = \mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{A}_k^{(\mathcal{O})}|\boldsymbol{y}_\mathrm{T}\right] +$$
$$+\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{B}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{B}_k^{(\mathcal{O})}|\boldsymbol{y}_\mathrm{T}\right] - \hat{\boldsymbol{B}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{B}}_k^{(\mathcal{O})} \quad (22)$$

which is identical to $\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}} + \hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}$ (cf. (19)) in the solution (17). Thus, $\boldsymbol{P}$ and β are optimized based on the residual channel $\boldsymbol{R}_k^{(\mathcal{O})}$ (compare Fig. 10). Hence, they take into account the imperfect interference compensation by $\boldsymbol{F}_\mathrm{P}$ due to P-CSI.

Remark 3: Channel estimation and precoding are optimized jointly using the CM estimate of the cost function (15). With the statistical assumptions from above it is clear that it is equivalent to LMMSE channel estimation followed by a robust optimization of precoding (see [1]).

Remark 4: An alternative point of view on the solution is possible interpreting the new cost function $\mathrm{Z_P}$ (15) as the original cost function (9) modified by a regularization term, which results in the nondiagonal loading with $\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}}$ (cf. (43)). A nondiagonal loading is well known from the solution of robust least squares [43], when solved with methods from static stochastic programming [21].

C. Optimization of Precoding Order

In the previous subsection, THP is optimized assuming a fixed precoding order $\mathcal{O}$ of the K data streams. The precoding order is described by the permutation matrix $\boldsymbol{\Pi}^{(\mathcal{O})} \in \{0,1\}^{K\times K}$ as defined in Section II-A. Applying the solution (17) for P-CSI to the cost function (15), we obtain

$$\mathrm{Z_P}(\mathcal{O};\boldsymbol{y}_\mathrm{T}) = \mathrm{tr}(\boldsymbol{C}_{\boldsymbol{w}}) -$$
$$-\sum_{k=1}^{K}\boldsymbol{e}_k^\mathrm{T}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}\left(\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}} + \hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})} + \frac{\bar{G}_\mathrm{N}\sigma_n^2}{P_\mathrm{Tx}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k. \quad (23)$$

To avoid the high complexity of $O(K!K^3)$ for a full search among all possible precoding orders, the standard suboptimum approach (see [17] for details) 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

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

Finally, this yields the permutation matrix $\boldsymbol{\Pi}^{(\mathcal{O}_\mathrm{P})}$. The complexity for computing the precoding order can be reduced to $O(K^3)$ using a symmetrically permuted Cholesky factorization as described in [19], [20], and [44].

D. Solution for Conventional Receiver Model

For the conventional receiver model with $\boldsymbol{G} = \boldsymbol{I}_K$ (6), the conditional expectations in the solution (17) for P-CSI can simply be written in terms of the moments in (10). Applying the properties of a correlation matrix, the CM estimate of the channel Gram matrix $\boldsymbol{H}^\mathrm{H}\boldsymbol{H}$ reads

$$\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{H}^\mathrm{H}\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}\right] = \hat{\boldsymbol{H}}^\mathrm{H}\hat{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}|\boldsymbol{y}_\mathrm{T}},\ \hat{\boldsymbol{H}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}]. \quad (24)$$

The conditional covariance matrix can be computed using (10), as $\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}|\boldsymbol{y}_\mathrm{T}} = \mathrm{E}_{\boldsymbol{h}}[(\boldsymbol{H}-\hat{\boldsymbol{H}})^\mathrm{H}(\boldsymbol{H}-\hat{\boldsymbol{H}})|\boldsymbol{y}_\mathrm{T}] = \sum_{k=1}^{K}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_\mathrm{T}}^*$. It is identical to the covariance matrix of the estimation error $\mathrm{E}_{\boldsymbol{h},\hat{\boldsymbol{H}}}[(\boldsymbol{H}-\hat{\boldsymbol{H}})^\mathrm{H}(\boldsymbol{H}-\hat{\boldsymbol{H}})]$ due to the orthogonality property of the LMMSE estimator *and* the jointly (complex) Gaussian distribution of $\boldsymbol{y}$ and $\boldsymbol{h}$ [42], [45].

Thus, the THP solution in (17) is given analytically with

$$\begin{bmatrix}\hat{\boldsymbol{A}}_k^{(\mathcal{O})}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix} = \boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}] = \boldsymbol{\Pi}^{(\mathcal{O})}\hat{\boldsymbol{H}} \in \mathbb{C}^{K\times M} \quad (25)$$

and the loading matrix

$$\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}} = \boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}|\boldsymbol{y}_\mathrm{T}}. \quad (26)$$

Remark 1: Defining $\boldsymbol{A}_k^{(\mathcal{O})}$ and $\boldsymbol{B}_k^{(\mathcal{O})}$ equivalently to (18) (see also (36) with $\boldsymbol{G} = \boldsymbol{I}_K$), the CM estimate of $\boldsymbol{H}^\mathrm{H}\boldsymbol{H}$ can be written

$$\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}|\boldsymbol{y}_\mathrm{T}} = \boldsymbol{C}_{\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}|\boldsymbol{y}_\mathrm{T}} + \boldsymbol{C}_{\boldsymbol{B}_k^{(\mathcal{O}),\mathrm{H}}|\boldsymbol{y}_\mathrm{T}}. \quad (27)$$

This allows for the following interpretation of the terms in the inverse in $\boldsymbol{p}_k$ (17), which differ from the traditional solution:

$$\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})} + \boldsymbol{C}_{\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}|\boldsymbol{y}_\mathrm{T}}$$

is the CM estimate of the Gram matrix $\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{A}_k^{(\mathcal{O})}$ [45]. $\boldsymbol{C}_{\boldsymbol{B}_k^{(\mathcal{O}),\mathrm{H}}|\boldsymbol{y}_\mathrm{T}}$ is the covariance matrix of the remaining interference, which is due to the imperfect feedback filter $\boldsymbol{F}$. $\boldsymbol{F}$ does not cancel interference from previously encoded data streams completely. The statistical properties of the remaining interference are now taken into account by $\boldsymbol{P}_{\mathrm{P}}$, whereas the traditional solution (12) assumed perfect operation of the feedback $\boldsymbol{F}_{\mathrm{T}}$ (13).

Remark 2: As the only difference to traditional THP (12) is $\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}}$, the additional complexity of this solution for the conventional receiver model is small compared to the complexity of the traditional THP solution (Section IV-A) assuming an LMMSE estimator is used.

E. Solution for the Simple Choice of Novel Receiver Model

For the derivations until Section IV-D, the general version of the novel receiver model (7) was assumed. In the robust THP solutions, it remains to evaluate the conditional expectations in (17), (18), and (19) analytically. To obtain a closed-form solution, we assume the particular choice of the novel receiver model (8) in the sequel.

With the receiver model (8), the CM estimate of the effective channel $\boldsymbol{GH}$ is derived in Appendix II and reads

$$\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{GH}|\boldsymbol{y}_{\mathrm{T}}] = \hat{\boldsymbol{G}}\hat{\boldsymbol{H}} + \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}} \tag{28}$$

where the kth row of $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}} \in \mathbb{C}^{K\times M}$ is given in (58). The first term is the product of the *separate* CM estimates of the receivers

$$\hat{\boldsymbol{G}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}|\boldsymbol{y}_{\mathrm{T}}] = \mathrm{diag}\,[\hat{g}_k]_{k=1}^{K} \tag{29}$$

and the channel $\hat{\boldsymbol{H}}$. The second term $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}}$ can be seen as the correction term for estimating the *product* $\boldsymbol{GH}$.

Due to (8), we have $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G} = \boldsymbol{I}_K$. Thus, the correlation matrices

$$\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{GH}|\boldsymbol{y}_{\mathrm{T}}\right] = \mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}\right] \tag{30}$$

are equal and given by (24). Note, that the corresponding *covariance* matrices $\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}}$ and $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}}$ are different for the novel receiver model in contrast to the conventional model (26).

Remark 1: Now, the CM estimate of the Gram matrix of the residual channel (22) simplifies to (Fig. 10)

$$\begin{aligned}\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{R}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{R}_k^{(\mathcal{O})}|\boldsymbol{y}_{\mathrm{T}}\right] = {}& \hat{\boldsymbol{H}}_{\mathrm{A},k}^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{A},k}^{(\mathcal{O})} + \boldsymbol{C}_{\boldsymbol{H}_{\mathrm{A},k}^{(\mathcal{O}),\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}} + \\ &+\hat{\boldsymbol{H}}_{\mathrm{B},k}^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{B},k}^{(\mathcal{O})} + \boldsymbol{C}_{\boldsymbol{H}_{\mathrm{B},k}^{(\mathcal{O}),\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}} - \hat{\boldsymbol{B}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{aligned} \tag{31}$$

where the first two terms are the CM estimate of the Gram matrix of $\boldsymbol{H}_{\mathrm{A},k}^{(\mathcal{O})}$ [45], the next two terms the CM estimate of the Gram matrix of $\boldsymbol{H}_{\mathrm{B},k}^{(\mathcal{O})}$, and finally the last term represents the amount of interference, which has already been canceled by $\boldsymbol{F}_{\mathrm{P}}$, is subtracted. ($\boldsymbol{H}_{\mathrm{A},k}^{(\mathcal{O})}$ contains the first k rows of $\boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{H}$ and $\boldsymbol{H}_{\mathrm{B},k}^{(\mathcal{O})}$ the last $K-k$ rows.)

Remark 2: As it is possible to give an explicit solution for the novel receiver model, the additional complexity is again small compared with the complexity of the traditional THP solution (Section IV-A) assuming an LMMSE estimator is used and the confluent hypergeometric function (Appendix II) is given by a lookup table.

F. Special Case: Linear Precoding

Linear precoding with P-CSI, which is a special case of THP, can be simply derived assuming $\boldsymbol{F} = \boldsymbol{0}_{K\times K}$ and $\boldsymbol{\Pi}^{(\mathcal{O})} = \boldsymbol{I}_K$ in the cost function (15) and solving (16). For the specific choice of receiver model in (8), the solution $\boldsymbol{P}_{\mathrm{lin}} = [\boldsymbol{p}_{\mathrm{lin},1}, \ldots, \boldsymbol{p}_{\mathrm{lin},K}]$ can be obtained from (47) and reads [35]

$$\boldsymbol{p}_{\mathrm{lin},k} = \beta_{\mathrm{lin}}\left(\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}] + \frac{K\sigma_n^2}{P_{\mathrm{Tx}}}\boldsymbol{I}_M\right)^{-1}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}]\boldsymbol{e}_k \tag{32}$$

where the conditional expectations are given in (24), (28), and Appendix II.

V. Transition From Complete to Statistical CSI

From the definitions of the different categories of CSI in terms of the (conditional) probability distribution of the channel parameters (Table I), it is evident that a continuous transition from THP designed for P-CSI to C-CSI and S-CSI is possible. In this section, we discuss these asymptotic cases of the THP solution for P-CSI with (8) (Section IV-B).

A. Complete CSI at the Transmitter

Convergence to C-CSI is achieved in scenarios with no noise in the observations $\boldsymbol{y}_{\mathrm{T}}$ from the training sequence received in the uplink and for a channel, whose parameters are time invariant over multiple time slots, i.e.,

$$\sigma_v^2 \to 0 \quad \text{and} \quad r[i] \to 1\,\forall i. \tag{33}$$

In the case of C-CSI, i.e., $\hat{\boldsymbol{H}} = \boldsymbol{H}$, the error covariance matrix of the channel estimate is zero (10), and the CM estimate $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{GH}|\boldsymbol{y}_{\mathrm{T}}]$ is equivalent to the (true) effective channel $\boldsymbol{GH}$

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

This leads to $\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}} = \boldsymbol{0}_M$. The conditional pdf $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}}(\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}) = \delta(\boldsymbol{h} - \boldsymbol{W}\boldsymbol{y}_{\mathrm{T}})$ is a Dirac distribution centered at $\boldsymbol{W}\boldsymbol{y}_{\mathrm{T}}$, and the MSE (9) is no random variable anymore. The optimum filters $\boldsymbol{p}_{\mathrm{P},k}$ and $\boldsymbol{f}_{\mathrm{P},k}$ (17) converge to (cf. Table II)

$$\boldsymbol{p}_{\mathrm{C},k} = \beta_{\mathrm{C}}\left(\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{A}_k^{(\mathcal{O})} + \frac{K\sigma_n^2}{P_{\mathrm{Tx}}}\boldsymbol{I}_M\right)^{-1}\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k \tag{34}$$

$$\boldsymbol{f}_{\mathrm{C},k} = -\beta_{\mathrm{C}}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \boldsymbol{B}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{C},k} \tag{35}$$

with the reordered effective channel

$$\left[\boldsymbol{A}_k^{(\mathcal{O}),\mathrm{T}},\ \boldsymbol{B}_k^{(\mathcal{O}),\mathrm{T}}\right]^{\mathrm{T}} = \boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{GH}. \tag{36}$$

The traditional approach (cf. (12) and (13)) leads to the same solution for C-CSI.

B. Statistical CSI at the Transmitter

If the observation $\boldsymbol{y}_{\mathrm{T}}$ is statistically independent of the current channel $\boldsymbol{H}$, i.e., $\mathrm{p}_{\boldsymbol{h}|\boldsymbol{y}_{\mathrm{T}}}(h|\boldsymbol{y}_{\mathrm{T}}) \to \mathrm{p}_{\boldsymbol{h}}(h)$, THP can be opti-

TABLE II
PARAMETERS OF THP SOLUTION FOR COMPLETE, PARTIAL, AND STATISTICAL CSI AT THE TRANSMITTER AND DIFFERENT RECEIVER MODELS

	C-CSI	P-CSI	S-CSI
Conventional Receiver Model (Eq. 6): $\boldsymbol{G} = \boldsymbol{I}_K$			
$\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}}$	$\boldsymbol{0}_{M\times M}$	$\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}\vert\boldsymbol{y}_\mathrm{T}}$	$\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}}$
$\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}\vert\boldsymbol{y}_\mathrm{T}]$	$\boldsymbol{H}$	$\hat{\boldsymbol{H}}$	$\boldsymbol{0}_{K\times M}$
Specific Choice of the Novel Receiver Model (Eq. 8):			
$\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}}$ (Eqs. 19, 38)	$\boldsymbol{0}_{M\times M}$	$\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}\vert\boldsymbol{y}_\mathrm{T}} + \hat{\boldsymbol{H}}^\mathrm{H}\hat{\boldsymbol{H}} - \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^\mathrm{H}\boldsymbol{G}^\mathrm{H}\vert\boldsymbol{y}_\mathrm{T}]\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}\vert\boldsymbol{y}_\mathrm{T}]$	$\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} - \boldsymbol{U}_{\boldsymbol{H}}^\mathrm{H}\boldsymbol{U}_{\boldsymbol{H}}$
$\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}\vert\boldsymbol{y}_\mathrm{T}]$ (Eqs. 28, 38)	$\boldsymbol{G}\boldsymbol{H}$	$\hat{\boldsymbol{G}}\hat{\boldsymbol{H}} + \boldsymbol{U}_{\boldsymbol{H}\vert\boldsymbol{y}_\mathrm{T}}$	$\boldsymbol{U}_{\boldsymbol{H}}$

mized based on S-CSI only, as no information about the current channel realization is available. Statistical independence is achieved for

$$\sigma_v^2 \to \infty \quad \text{or} \quad r[i] \to 0 \ \text{ for } i \neq 0 \tag{37}$$

i.e., for very low signal-to-noise ratio (SNR) or a temporally uncorrelated channel, e.g., at very high Doppler frequency. If $\boldsymbol{H}$ is zero mean, the LMMSE estimate is $\hat{\boldsymbol{H}} = \boldsymbol{0}_{K\times M}$ in this limit (Table II).

If $\hat{\boldsymbol{H}} = \boldsymbol{0}_{K\times M}$, we obtain $\beta_\mathrm{T}^{-1} = 0$ for traditional THP (11) and $\beta_\mathrm{P}^{-1} = 0$ for THP with P-CSI based on the conventional receiver model (16), i.e., the receivers are switched off. It can be shown that the solution for the linear part $\boldsymbol{P}$ of THP is not unique: It can be chosen arbitrarily as long as it does not exceed the transmit power constraint. The most reasonable choice is $\boldsymbol{P}_\mathrm{T} = \boldsymbol{P}_\mathrm{P} = \boldsymbol{0}_{M\times K}$, as the transmitter should stop transmitting data, if it assumes that the receivers are switched off. Although this case is not achieved exactly in the simulations as $\|\hat{\boldsymbol{H}}\|_\mathrm{F} > 0$ numerically, the BER saturates at 0.5 (Section VI).

Practically, the transmitter stops transmitting data, as the CSI is not reliable enough. This behavior is not acceptable, as CSI in terms of the statistical channel properties is available and could be used to perform simple beamforming with reasonable performance in many scenarios as will be shown in Section VI.

In the case of traditional THP optimization, this property is a consequence of the fact that the *a priori* information about the signal processing application is not included into channel estimation: The mean-square error of the channel estimate is not the best criterion with regard to THP performance. Although the CM channel estimate, which is zero in this case, is the best channel estimate on average with regard to the task of channel identification, it is highly suboptimum for THP.

For optimization based on the conventional receiver model (6), the transmitter "believes" that the receiver does not have any CSI at all and votes for stopping data transmission.

The disadvantages of traditional optimization and a too-simple receiver model are overcome when combining the CM estimate of the MSE (15) with the novel receiver model. The solution is determined by the asymptotic behavior of (Table II)

$$\begin{aligned}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^\mathrm{H}\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}] &\to \boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^\mathrm{H}\boldsymbol{H}]\\ \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}] &\to \boldsymbol{U}_{\boldsymbol{H}}\end{aligned} \tag{38}$$

with kth row $\boldsymbol{e}_k^\mathrm{T}\boldsymbol{U}_{\boldsymbol{h}} = \sqrt{\pi}\boldsymbol{q}_k^\mathrm{H}\boldsymbol{C}_{\boldsymbol{h}_k}^*/(\boldsymbol{q}_k^\mathrm{H}\boldsymbol{C}_{\boldsymbol{h}_k}^*\boldsymbol{q}_k)^{1/2}/2$ (cf. (56) and (58)). Applying these limits to (17), we obtain the solution for THP in case only S-CSI is available (note that $\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}} + \hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})} = \boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} - \hat{\boldsymbol{B}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{B}}_k^{(\mathcal{O})}$), as follows:

$$\begin{aligned}\boldsymbol{p}_{\mathrm{S},k} &= \beta_\mathrm{S}\left(\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} - \hat{\boldsymbol{B}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{B}}_k^{(\mathcal{O})} + \frac{K\sigma_n^2}{P_\mathrm{Tx}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k\\ \boldsymbol{f}_{\mathrm{S},k} &= -\beta_\mathrm{S}^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_{\mathrm{S},k}.\end{aligned} \tag{39}$$

Here, for S-CSI, the effective channel (18) is

$$\left[\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{T}},\ \hat{\boldsymbol{B}}_k^{(\mathcal{O}),\mathrm{T}}\right]^\mathrm{T} = \boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{U}_{\boldsymbol{H}}. \tag{40}$$

The covariance matrix of the interference (21), which remains from imperfect feedback (Fig. 10), is

$$\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{R}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{R}_k^{(\mathcal{O})}\right] = \underbrace{\boldsymbol{C}_{\boldsymbol{H}_{\mathrm{A},k}^{(\mathcal{O}),\mathrm{H}}} + \boldsymbol{C}_{\boldsymbol{H}_{\mathrm{B},k}^{(\mathcal{O}),\mathrm{H}}}}_{\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}}} - \hat{\boldsymbol{B}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{B}}_k^{(\mathcal{O})}. \tag{41}$$

Certainly, in case of S-CSI, the interference from previously encoded data streams at the receivers is unknown and cannot be canceled anymore. However, as $\hat{\boldsymbol{B}}_k^{(\mathcal{O})}$ is not zero, the feedback filter $\boldsymbol{F}_\mathrm{S}$ effectively reduces the residual interference with covariance matrix described by (41). Thus, under the assumption that the magnitude of real or imaginary part of the input signal to the modulo operator at the transmitter are larger than $\tau/2$, i.e., the modulo operation changes its input (Fig. 2), still some interference can be reduced by the feedback in case of S-CSI and improves performance [2]. Note that if the input of the modulo operator is small enough in magnitude, THP reduces to linear precoding, which is

$$\boldsymbol{p}_{\mathrm{lin},k} = \beta_\mathrm{lin}\left(\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} + \frac{K\sigma_n^2}{P_\mathrm{Tx}}\boldsymbol{I}_M\right)^{-1}\boldsymbol{U}_{\boldsymbol{H}}^\mathrm{H}\boldsymbol{e}_k \tag{42}$$

in case of S-CSI (cf. (32)).

VI. Performance Results

As an example, we assume alternating uplink/downlink slots (Fig. 1) and a worst-case delay of three slots[4] to the first uplink slot available with a training sequence. Thus, the vector collecting the available observations of Q time slots reads as $\boldsymbol{y}_\mathrm{T} = [\bar{\boldsymbol{y}}[q^\mathrm{tx}-3]^\mathrm{T}, \bar{\boldsymbol{y}}[q^\mathrm{tx}-5]^\mathrm{T}, \ldots, \bar{\boldsymbol{y}}[q^\mathrm{tx}-(2Q+1)]^\mathrm{T}]^\mathrm{T} \in \mathbb{C}^{MNQ}$, whereas the corresponding collection of channel vectors as $\boldsymbol{h}_\mathrm{T} =$

[4]This conservative choice is made to show the worst case effect in case of a symmetric slot structure (alternating uplink/downlink slots, Fig. 1). It may be due to limited processing power at the transmitter such that the estimate from the previous uplink slot cannot be incorporated in the current filter optimization in time. In addition, in a possible closed-loop system, there would always be a minimum delay of two slots (or more due to limited feedback). Note that for systems with an asymmetric slot structure, where the distance between two consecutive uplink slots is increased compared to a symmetric structure, the worst-case delay can be even larger.

$[\boldsymbol{h}[q^{\mathrm{tx}}-3]^{\mathrm{T}}, \boldsymbol{h}[q^{\mathrm{tx}}-5]^{\mathrm{T}}, \ldots, \boldsymbol{h}[q^{\mathrm{tx}}-(2Q+1)]^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{QMK}$. Due to the symmetric slot structure, $\boldsymbol{C}_{\mathrm{T}}$ is Toeplitz with first column $[r[0], r[2], \ldots, r[2Q-2]]^{\mathrm{T}}$ (cf. Section II-C).

For the following simulations, we consider a system with $M = 4$ transmit antennas in a uniform linear array with half wavelength spacing and $K = 3$ receivers with one antenna each. The temporal autocorrelation of the complex Gaussian channel coefficients is identical for all coefficients and corresponds to a Jakes power spectrum with maximum Doppler frequency f_{d}, which is normalized to the slot period.[5] System parameters for UMTS UTRA TDD systems taken from [46] are, e.g., a carrier frequency of 2 GHz and a slot period of 0.675 ms. For these parameters, a normalized Doppler frequency of $f_{\mathrm{d}} = 0.08$ corresponds to a velocity of 64 km/h, for example. Increasing the carrier frequency to 5 GHz, this Doppler frequency corresponds to a velocity of 25.6 km/h.

The azimuth directions of $W = 100$ wave fronts per receiver channel are uniformly distributed within 3° centered around their mean. A spatial scenario with mean angles $[-15°, 0°, 15°]$ for the three receivers is considered. It is selected to have acceptable performance in case of S-CSI (see below), i.e., users are spatially well separated.

Walsh–Hadamard sequences of length $N = 32$ are used for the training in the uplink and received training sequences from $Q = 5$ previous uplink slots are used for the prediction of the channel. The filter vector $\boldsymbol{q}_k$ (cf. Section II-B) for the transmission of training sequences in the downlink for receiver k is the complex conjugate principal eigenvector of the conditional correlation matrix of the channel $\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{T}}\right]$. As in Section II-A, we assume $\boldsymbol{C}_{\boldsymbol{d}} = \boldsymbol{I}_K$ and $\boldsymbol{C}_{\boldsymbol{n}} = \sigma_n^2 \boldsymbol{I}_K$.

With "conventional (conv.) receiver model," we refer to the model in (6), with "novel receiver model" to (7) and (8). In the simulations, the actual receivers use the transmitter's model.

For the results shown in Figs. 11 and 12, 300 QPSK data symbols were transmitted per channel realization and averaged over 10^4 independent channel realizations. For the normalized Doppler frequency $f_{\mathrm{d}} = 0$, the channel is constant over all time slots. Due to the high SNR of 20 dB and the large number of $NQ = 160$ training symbols, the LMMSE prediction of the channel is almost perfect for this case. Thus, all THP solutions converge to the THP solution based on C-CSI (cf. Section V-A).

For higher Doppler frequencies, the quality of the available CSI gets worse as the time variance of the channel increases. The traditional THP (12), (13), with conventional receive processing (6) which still assumes C-CSI (solid line), suffers from the imperfect prediction and finally saturates at an uncoded BER of 0.5. The performance degradation is smaller at low Doppler frequencies if the simple novel receiver model (14) is considered (dashed line with circles). In the high Doppler region, the performance is improved by considering P-CSI in the optimization (16) and maintaining the conventional receiver (dotted line with crosses). These two curves do not saturate at 0.5 due to the reasons mentioned in Section V-B.

Note that traditional THP with the simple choice of the novel receiver model performs better than the robust THP with

[5]The Doppler frequency normalized to the symbol period is obtained dividing f_{d} by the number of symbols per slot N_{slot}.

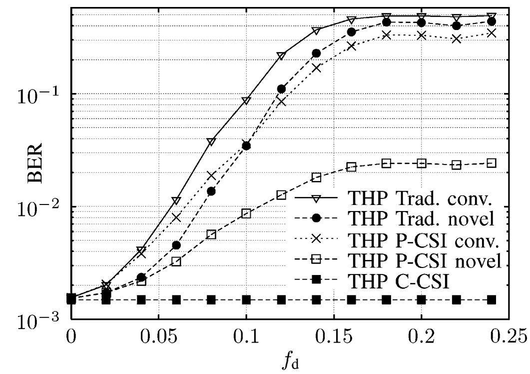


Fig. 11. Uncoded BER versus normalized Doppler frequency for different THP designs (THP with C-CSI as reference) and QPSK symbols, $10\log_{10}(P_{\mathrm{Tx}}/\sigma_n^2) = 20$ dB.

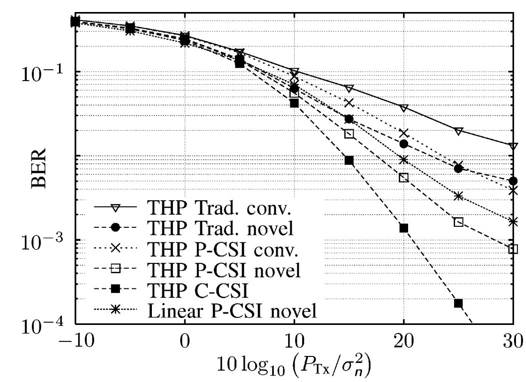


Fig. 12. Uncoded BER versus SNR for different THP designs (THP with C-CSI and linear precoder as reference) and QPSK symbols, $f_{\mathrm{d}} = 0.08$.

conventional receive processing for low Doppler frequencies, whereas the situation is vice versa for high values of f_{d}. In this Doppler region, the remaining interference due to imperfect channel estimation dominates the BER. Thus, robust THP performs better since it uses the second-order moment of the channel to reduce interference.

Only the combination of the CM estimate of the cost function with the simple choice of the novel receiver (cf. Section IV-E) leads to an improvement for all Doppler frequencies (dashed line with boxes): especially for high values of f_{d}, i.e., S-CSI (cf. Section V-B), the performance is remarkably better than for the other solutions. The systematic approach of estimating the MSE cost function and employing receivers which are capable of performing a phase correction based on channel knowledge allows for a smooth transition between C-CSI and S-CSI with acceptable BER for the latter case.

As mentioned above, the chosen spatial scenario ensures good performance also for S-CSI. For a spatially less correlated channel—$\boldsymbol{C}_{\boldsymbol{h}_k} = \boldsymbol{I}_M$ in the extreme case—the BER for high Doppler frequency of all algorithms saturates at 0.5 as no statistical, i.e., spatial, structure is available.

Fig. 12 shows results for the same THP designs for a fixed normalized Doppler frequency of 0.08 versus SNR. Note that the robust linear precoder using the simple novel receiver (cf. Section IV-F) outperforms all THP solutions without this receive processing. Only traditional THP with the simple novel

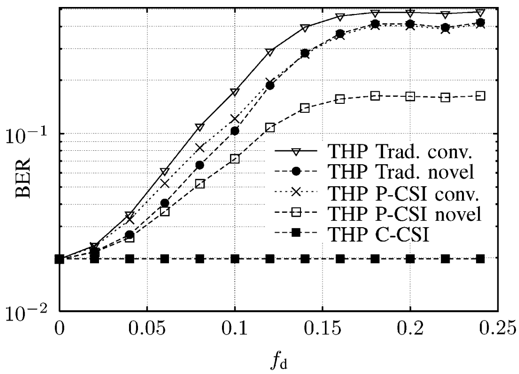


Fig. 13. Uncoded BER versus normalized Doppler frequency for different THP designs (THP with C-CSI as reference) and 16 QAM symbols, $10\log_{10}(P_{\mathrm{Tx}}/\sigma_n^2) = 20$ dB.

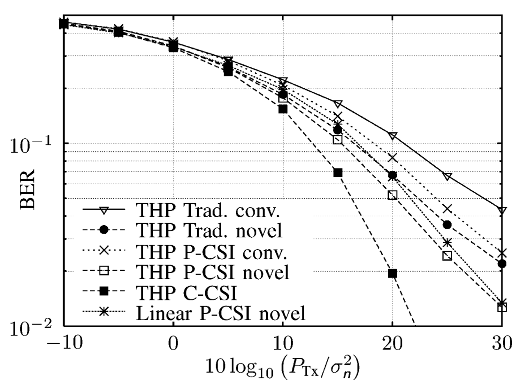


Fig. 14. Uncoded BER versus SNR for different THP designs (THP with C-CSI and linear precoder as reference) and 16 QAM symbols, $f_\mathrm{d} = 0.08$.

receiver performs slightly better between 5 and 15 dB, but at the expense of a higher BER floor. Robust THP with the novel receiver model performs better than all other linear and nonlinear precoders.

Figs. 13 and 14 show the results for the same simulation parameters as before, but with the modulation alphabet chosen to be 16-QAM. The results are qualitatively comparable with those shown in Figs. 11 and 12, but on a generally higher BER level due to the employed modulation alphabet. Note that the gap between robust linear and robust nonlinear precoding with the novel receiver is smaller at high SNR values (for $f_\mathrm{d} = 0.08$) compared with the transmission of QPSK symbols (cf. Fig. 14). In addition, traditional THP with novel receiver performs better than robust THP with conventional receive processing for the whole simulated SNR range.

VII. Conclusion

Traditional MSE design of Tomlinson–Harashima precoding explicitly (or sometimes implicitly) assumes the same CSI at the receiver as at the transmitter and complete CSI at the transmitter. When applying THP to the wireless broadcast channel, complete channel state information is not available at the transmitter. This leads to a large degradation in performance.

We presented a novel robust design for Tomlinson–Harashima precoding. It is based on two paradigms: It performs a conditional mean estimate of the cost function and includes a new model of the receiver, which describes the receivers' perfect (or at least more accurate) CSI explicitly. Thus, using this robust design paradigm a meaningful solution for Tomlinson–Harashima precoding for statistical CSI is obtained and Tomlinson–Harashima precoding based on the MSE is now ready in *principle* for the application "wireless broadcast channel."

Of course, not all extensions and practical aspects of *robust* Tomlinson–Harashima precoding with partial CSI could be explored here, where our focus was on the presentation of the principal ideas: As one important aspect, future work should address the choice of different receiver models and their integration in the results given here. Furthermore, estimation methods for temporal and spatial channel correlations have to be integrated into robust THP.

Appendix I
Derivation of MMSE-THP With P-CSI

The optimization of MMSE-THP for C-CSI has been performed in [17] by means of Lagrangian multipliers to ensure the lower triangular structure of $\boldsymbol{F}$ with zero main diagonal. In this paper, we follow a different approach minimizing the cost function in (15) for the more general case of P-CSI.

The cost function in (15) can be rewritten as

$$\begin{aligned}
&\mathrm{Z_P}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_\mathrm{T})\\
&=\mathrm{E}_{\boldsymbol{h}}[\mathrm{Z_T}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{H})|\boldsymbol{y}_\mathrm{T}]\\
&=\mathrm{E}_{\boldsymbol{w},\boldsymbol{n}}\left[\left\|\left(\beta^{-1}\boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}]\boldsymbol{P}-(\boldsymbol{I}_K-\boldsymbol{F})\right)\boldsymbol{w}[n]\right\|_2^2\right]\\
&\quad+\beta^{-2}\bar{G}_\mathrm{N}\sigma_n^2+\beta^{-2}\mathrm{tr}\left(\boldsymbol{P}^\mathrm{H}\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}}\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\right)
\end{aligned}\tag{43}$$

with $\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}}$ as defined in (19) and $\bar{G}_\mathrm{N} = \mathrm{tr}(\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{G}^\mathrm{H}|\boldsymbol{y}_\mathrm{T}])$.[6] First, this cost function is minimized with respect to $\boldsymbol{F}$ assuming a fixed but unknown matrix $\boldsymbol{P}$. Since the signal $\boldsymbol{w}[n]$ and the noise $\boldsymbol{n}[n]$ are assumed to be uncorrelated, only the first summand of the expected value in (43) depends on $\boldsymbol{F}$. Using the widespread assumption of a diagonal covariance matrix [12] $\boldsymbol{C}_{\boldsymbol{w}} = \mathrm{diag}\left[\sigma_{w_k}^2\right]_{k=1}^K$, its contribution to the MSE reads as

$$\mathrm{E}_{\boldsymbol{w}}\left[\left\|\left(\beta^{-1}\boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}]\boldsymbol{P}-(\boldsymbol{I}_K-\boldsymbol{F})\right)\boldsymbol{w}[n]\right\|_2^2\right] = \\ = \sum_{k=1}^K \|\boldsymbol{\theta}_k\|_2^2\,\sigma_{w_k}^2 \tag{44}$$

with $\boldsymbol{\theta}_k = \beta^{-1}\boldsymbol{\Pi}^{(\mathcal{O})}\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_\mathrm{T}]\boldsymbol{p}_k - \boldsymbol{e}_k + \boldsymbol{f}_k$. Each summand can be minimized separately taking into account that only the last $K-k$ elements of $\boldsymbol{f}_k$ may contain nonzero entries. The Euclidean norm of $\boldsymbol{\theta}_k$ is minimized while setting the first k elements of $\boldsymbol{f}_k$ to zero. Thus, the kth column of the feedback filter is chosen as

$$\boldsymbol{f}_k = -\beta^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M}\\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_k \tag{45}$$

with $\hat{\boldsymbol{B}}_k^{(\mathcal{O})}$ defined in (18), which fulfills the constraint on the structure of $\boldsymbol{F}$ and still depends on $\boldsymbol{p}_k$.

[6]Note, that (43) without conditional expectation suffices for derivation of (12) and (13), where $\boldsymbol{L}_{\boldsymbol{y}_\mathrm{T}} = \boldsymbol{0}_{M\times M}$.

To determine the parameters $\boldsymbol{P}$ and β, the transmit power constraint $\operatorname{tr}\left(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\right) \leq P_{\mathrm{Tx}}$ has to be taken into account. Thus, we set up the Lagrangian function

$$L(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_{\mathrm{T}},\lambda) = \\ = \mathrm{Z}_{\mathrm{P}}(\boldsymbol{P},\boldsymbol{F},\beta;\boldsymbol{y}_{\mathrm{T}}) + \lambda\left(\operatorname{tr}\left(\boldsymbol{P}\boldsymbol{C}_{\boldsymbol{w}}\boldsymbol{P}^{\mathrm{H}}\right) - P_{\mathrm{Tx}}\right) \quad (46)$$

where $\lambda \in \mathbb{R}_{+,0}$ is the Lagrangian multiplier, and we assume a fixed but unknown feedback filter $\boldsymbol{F}$. The partial derivative of (46) with respect to $\boldsymbol{P}$ yields

$$\boldsymbol{P} = \beta\left(\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}] + \xi\boldsymbol{I}_M\right)^{-1} \times \\ \times \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}]\boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}}(\boldsymbol{I}_K - \boldsymbol{F}) \quad (47)$$

where $\xi = \bar{G}_{\mathrm{N}}\sigma_n^2/P_{\mathrm{Tx}}$ is found using the partial derivative of (46) with respect to β. Note that $\boldsymbol{P}$ is a linear precoder for the vector $\boldsymbol{b}[n]$ since $\boldsymbol{w}[n] = (\boldsymbol{I}_K - \boldsymbol{F})^{-1}\boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{b}[n]$ (cf. Fig. 6).

In order to find the solution for $\boldsymbol{P}$, which is independent of the feedback filter, we insert (45) in the expression of the kth column of (47) and get

$$\boldsymbol{p}_k = \beta\left(\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}\right] + \frac{\bar{G}_{\mathrm{N}}\sigma_n^2}{P_{\mathrm{Tx}}}\boldsymbol{I}_M\right)^{-1}\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}|\boldsymbol{y}_{\mathrm{T}}\right] \times \\ \times \boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}}\left(\boldsymbol{e}_k + \beta^{-1}\begin{bmatrix}\boldsymbol{0}_{k\times M} \\ \hat{\boldsymbol{B}}_k^{(\mathcal{O})}\end{bmatrix}\boldsymbol{p}_k\right). \quad (48)$$

Solving for $\boldsymbol{p}_k$ yields

$$\boldsymbol{p}_{\mathrm{P},k} = \beta_{\mathrm{P}}\left(\boldsymbol{L}_{\boldsymbol{y}_{\mathrm{T}}} + \hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\hat{\boldsymbol{A}}_k^{(\mathcal{O})} + \frac{\bar{G}_{\mathrm{N}}\sigma_n^2}{P_{\mathrm{Tx}}}\boldsymbol{I}_M\right)^{-1}\hat{\boldsymbol{A}}_k^{(\mathcal{O}),\mathrm{H}}\boldsymbol{e}_k \quad (49)$$

with $\hat{\boldsymbol{A}}_k^{(\mathcal{O})}$ as defined in (18) and where β_{P} is found by fulfilling the transmit power constraint with equality. The corresponding column of the feedback filter is found by inserting (49) in (45).

The expression for the optimization of precoding order shown in (23) can be obtained by inserting (45) and (49) in the cost function (cf. (15)).

APPENDIX II
DERIVATION OF CONDITIONAL MEAN $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]$ FOR THE SIMPLE NOVEL RECEIVER ((8) IN SECTION IV-E)

To derive an explicit expression for $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}]$ assuming the receiver model from (8), we define the complex Gaussian random variable $z_k = \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k$ and consider the kth row of $\boldsymbol{G}\boldsymbol{H}$. Using the properties of the conditional expectation[7][47], we obtain

$$\mathrm{E}_{\boldsymbol{h}}[g_k\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}] = \mathrm{E}_{\boldsymbol{h}}\left[\frac{(\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k)^*}{|\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{q}_k|}\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}\right] = \mathrm{E}_{\boldsymbol{h},z_k}\left[\frac{z_k^*}{|z_k|}\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}\right] \\ = \mathrm{E}_{\boldsymbol{h}}\left[\frac{z_k^*}{|z_k|}\mathrm{E}_{\boldsymbol{h}}\left[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}},z_k\right]|\boldsymbol{y}_{\mathrm{T}}\right]. \quad (50)$$

[7]For two random variables z and h we can write $\mathrm{E}_{z,h}[\mathrm{f}(z)h] = \mathrm{E}_z[\mathrm{f}(z)\mathrm{E}_h[h|z]]$.

As $\mathrm{p}_{\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}},z_k}(\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}},z_k)$ is complex Gaussian, the CM estimator $\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}},z_k]$ is equivalent to the LMMSE estimator [42]

$$\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}},z_k] = \\ = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}] + \boldsymbol{c}_{\boldsymbol{h}_k z_k|\boldsymbol{y}_{\mathrm{T}}}c_{z_k|\boldsymbol{y}_{\mathrm{T}}}^{-1}(z_k - \mathrm{E}_{z_k}[z_k|\boldsymbol{y}_{\mathrm{T}}]) \quad (51)$$

with covariances

$$\begin{aligned}\boldsymbol{c}_{\boldsymbol{h}_k z_k|\boldsymbol{y}_{\mathrm{T}}} &= \mathrm{E}_{\boldsymbol{h},z_k}[(\boldsymbol{h}_k - \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}])(z_k - \mathrm{E}_{z_k}[z_k|\boldsymbol{y}_{\mathrm{T}}])^*|\boldsymbol{y}_{\mathrm{T}}] \\ &= \boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}}\boldsymbol{q}_k^* \\ c_{z_k|\boldsymbol{y}_{\mathrm{T}}} &= \mathrm{E}_{z_k}[|z_k - \mathrm{E}_{z_k}[z_k|\boldsymbol{y}_{\mathrm{T}}]|^2|\boldsymbol{y}_{\mathrm{T}}] = \boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}}^*\boldsymbol{q}_k. \quad (52)\end{aligned}$$

The LMMSE estimator (51) should be read as the estimator of $\boldsymbol{h}_k$ given z_k under the "a priori distribution" $\mathrm{p}_{\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}}(\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}})$ of $\boldsymbol{h}_k$. The conditional first-order moment in (51) is

$$\mu_{z_k|\boldsymbol{y}_{\mathrm{T}}} = \mathrm{E}_{z_k}[z_k|\boldsymbol{y}_{\mathrm{T}}] = \hat{\boldsymbol{h}}_k^{\mathrm{T}}\boldsymbol{q}_k \quad (53)$$

and the LMMSE estimate $\hat{\boldsymbol{h}}_k = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}]$ of $\boldsymbol{h}_k$ given $\boldsymbol{y}_{\mathrm{T}}$ from (10). Applying (51) to (50) yields

$$\mathrm{E}_{\boldsymbol{h}}[g_k\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}] = \mathrm{E}_{\boldsymbol{h}}[g_k|\boldsymbol{y}_{\mathrm{T}}]\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}] + \\ + \boldsymbol{c}_{\boldsymbol{h}_k z_k|\boldsymbol{y}_{\mathrm{T}}}c_{z_k|\boldsymbol{y}_{\mathrm{T}}}^{-1}(\mathrm{E}_{z_k}[|z_k|\boldsymbol{y}_{\mathrm{T}}] - \mathrm{E}_{\boldsymbol{h}}[g_k|\boldsymbol{y}_{\mathrm{T}}]\mathrm{E}_{z_k}[z_k|\boldsymbol{y}_{\mathrm{T}}]). \quad (54)$$

With [48] the remaining terms, i.e., the CM estimate of the receivers' processing g_k and of the magnitude $|z_k|$, are

$$\hat{g}_k = \mathrm{E}_{\boldsymbol{h}}[g_k|\boldsymbol{y}_{\mathrm{T}}] = \frac{\sqrt{\pi}}{2}\frac{|\mu_{z_k|\boldsymbol{y}_{\mathrm{T}}}|}{c_{z_k|\boldsymbol{y}_{\mathrm{T}}}^{1/2}}\frac{\mu_{z_k|\boldsymbol{y}_{\mathrm{T}}}^*}{|\mu_{z_k|\boldsymbol{y}_{\mathrm{T}}}|}\,{}_1F_1\left(\frac{1}{2},2,-\frac{|\mu_{z_k|\boldsymbol{y}_{\mathrm{T}}}|^2}{c_{z_k|\boldsymbol{y}_{\mathrm{T}}}}\right) \quad (55)$$

$$\mathrm{E}_{z_k}[|z_k|\,|\boldsymbol{y}_{\mathrm{T}}] = \frac{\sqrt{\pi}}{2}c_{z_k|\boldsymbol{y}_{\mathrm{T}}}^{1/2}\,{}_1F_1\left(-\frac{1}{2},1,-\frac{|\mu_{z_k|\boldsymbol{y}_{\mathrm{T}}}|^2}{c_{z_k|\boldsymbol{y}_{\mathrm{T}}}}\right) \quad (56)$$

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

Summarizing the derivation, we obtain

$$\mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}] = \hat{\boldsymbol{G}}\hat{\boldsymbol{H}} + \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}}, \hat{\boldsymbol{G}} = \mathrm{E}_{\boldsymbol{h}}[\boldsymbol{G}|\boldsymbol{y}_{\mathrm{T}}] = \operatorname{diag}[\hat{g}_k]_{k=1}^K, \quad (57)$$

from (54). The kth row of $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}}$ is

$$\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}_{\mathrm{T}}} = \boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_{\mathrm{T}}}^*\,c_{z_k|\boldsymbol{y}_{\mathrm{T}}}^{-1}\left(\mathrm{E}_{z_k}[|z_k|\,|\boldsymbol{y}_{\mathrm{T}}] - \mu_{z_k|\boldsymbol{y}_{\mathrm{T}}}\hat{g}_k\right). \quad (58)$$

ACKNOWLEDGMENT

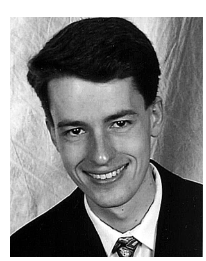
The authors would like to thank M. Joham for many helpful discussions on Tomlinson–Harashima precoding.

REFERENCES

[1] F. A. Dietrich and W. Utschick, "Robust Tomlinson–Harashima precoding," in *Proc. 16th IEEE Symp. Personal, Indoor Mobile Radio Communications*, Berlin, Germany, 2005, pp. 136–140.

[2] F. A. Dietrich, P. Breun, and W. Utschick, "Tomlinson-Harashima precoding: A continuous transition from complete to statistical channel knowledge," in *Proc. IEEE Global Telecommunications (IEEE GLOBECOM) Conf*, Saint Louis, MO, 2005, vol. 4, pp. 2379–2384.

[3] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. Signals, Systems,Computers*, Oct. 2000, vol. 2, pp. 1627–1631.
[4] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.
[5] M. Bengtsson and B. Ottersten, "Optimum and suboptimum transmit beamforming," in *Handbook of Antennas in Wireless Communications*. Boca Raton, FL: CRC Press, 2001, ch. 18.
[6] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.
[7] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983.
[8] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," in *Proc. Int. Symp. Information Theory Its Applications*, Nov. 2000, pp. 681–684.
[9] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE Global Telecommunications (IEEE GLOBECOM) Conf*, Nov. 2001, vol. 2, pp. 1344–1348.
[10] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5/6, pp. 138–139, Mar. 1971.
[11] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780, Aug. 1972.
[12] R. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.
[13] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, Jul. 2004.
[14] G. D. Forney and M. V. Eyuboglu, "Combined equalization and coding using precoding," *IEEE Commun. Mag.*, vol. 29, no. 12, pp. 25–34, Dec. 1991.
[15] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "MIMO precoding for decentralized receivers," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2002, p. 496.
[16] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio–temporal Tomlinson–Harashima precoding," in *Proc. ITG Conf. Source Channel Coding (ITG SCC'04)*, Jan. 2004, pp. 387–394.
[17] 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. New York: EURASIP, Hindawi Publishing Corporation, 2005, vol. 3.
[18] M. Joham, *Optimization of Linear and Nonlinear Transmit Signal Processing*. Aachen, Germany: Shaker Verlag, 2004.
[19] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Efficient Tomlinson-Harashima precoding for spatial multiplexing on flat MIMO channel," in *Proc. Int. Conf. Communications (ICC)*, May 2005, vol. 3, pp. 2021–2025.
[20] ——, "Cholesky factorization with symmetric permutation applied to detecting and precoding spatially multiplexed data streams," *IEEE Trans. Signal Process.*, 2006, accepted for publication.
[21] A. Prékopa, *Stochastic Programming*. Dordrecht, Germany: Kluwer, 1995.
[22] A. L. Swindlehurst and M. Viberg, "Bayesian approaches for robust array signal processing," in *Statistical Methods in Control & Signal Processing*. New York: Marcel Dekker, 1997.
[23] K.Öhrn, A. Ahlén, and M. Sternad, "A probabilistic approach to multivariable robust filtering and open-loop control," *IEEE Trans. Autom. Control*, vol. 40, pp. 405–718, Mar. 1995.
[24] F. Rey, M. Lamarca, and G. Vazquez, "Transmit filter optimization based on partial CSI knowledge for wireless applications," in *Proc. IEEE Int. Conf. Communications*, May 2003, vol. 4, pp. 2567–2571.
[25] P. Xia, S. Zhou, and G. B. Giannakis, "Adaptive MIMO OFDM based on partial channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 202–213, Jan. 2004.
[26] 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. Int. Conf. Communications*, Seoul, South Korea, May 2005, vol. 4, pp. 2162–2167.
[27] F. A. Dietrich, R. Hunger, M. Joham, and W. Utschick, "Robust transmit wiener filter for time division duplex systems," in *Proc. Int. Symp. Signal Processing Information Technology (ISSPIT)*, Darmstadt, Germany, Dec. 2003, pp. 415–418.
[28] R. Hunger, F. A. Dietrich, M. Joham, and W. Utschick, "Robust transmit zero-forcing filters," in *Proc. ITG Workshop on Smart Antennas*, Munich, Germany, Mar. 2004, pp. 130–137.
[29] A. P. Liavas, "Tomlinson–Harashima precoding with partial channel knowledge," *IEEE Trans. Commun.*, vol. 53, pp. 5–9, Jan. 2005.
[30] 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 Broadband Communications*, Feb. 2002, pp. 7-1–7-6.
[31] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Linear and nonlinear preequalization equalization for MIMO systems with long-term channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 373–378, Mar. 2004.
[32] M. Joham, K. Kusume, W. Utschick, and J. A. Nossek, "Transmit matched filter and transmit wiener filter for the downlink of FDD DS-CDMA systems," in *Proc. Personal, Indoor Mobile Radio Communication (PIMRC) Conf.*, Sep. 2002, vol. 5, pp. 2312–2316.
[33] B. Zerlin, M. Joham, W. Utschick, and J. A. Nossek, "Covariance based linear precoding," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 190–199, Jan. 2006.
[34] B. Zerlin, M. Joham, W. Utschick, and J. A. Nossek, "Covariance based linear precoding for full rank channels," in *Proc. Int. ITG Workshop on Smart Antennas*, Duisburg, Germany, Apr. 2005.
[35] F. A. Dietrich, W. Utschick, and P. Breun, "Linear precoding based on a stochastic MSE criterion," presented at the Proc. 13th Eur. Signal Processing Conf., Antalya, Turkey, Sep. 2005.
[36] M. Bengtsson and B. Ottersten, "Uplink and downlink beamforming for fading channels," in *Proc. IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communications*, Feb. 1999, pp. 350–353.
[37] M. Joham, J. Brehmer, A. Voulgarelis, and W. Utschick, "Multiuser spatio–temporal Tomlinson–Harashima precoding for frequency selective vector channels," in *Proc. ITG Workshop on Smart Antennas*, Munich, Germany, Mar. 2004, pp. 208–215.
[38] P. Monogioudis, K. Conner, D. Das, S. Gollamudi, J. A. C. Lee, A. L. Moustakas, S. Nagaraj, A. M. Rao, R. A. Soni, and Y. Yuan, "Intelligent antenna solutions for UMTS: Algorithms and simulation results," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 28–39, Oct. 2004.
[39] K. I. Pedersen, P. E. Mogensen, and J. Ramiro-Moreno, "Application and performance of downlink beamforming techniques in UMTS," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 134–143, Oct. 2003.
[40] R. Hunger, M. Joham, and W. Utschick, "Extension of linear and nonlinear transmit filters for decentralized receivers," in *Proc. Eur Wireless 2005*, Apr. 2005, vol. 1, pp. 40–46.
[41] M. Schubert and S. Shi, "MMSE transmit optimization with interference pre-compensation," in *Proc. Vehicular Technology Conf. (VTC) Spring*, May 2005, vol. 2, pp. 845–849.
[42] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*, 1st ed. Englewood Cliffs, NJ: PTR Prentice-Hall, 1993.
[43] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[44] D. A. Schmidt, M. Joham, F. A. Dietrich, K. Kusume, and W. Utschick, "Complexity reduction for MMSE multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. Int. ITG Workshop on Smart Antennas*, Duisburg, Germany, Apr. 2005.
[45] F. A. Dietrich, F. Hoffmann, and W. Utschick, "Conditional mean estimator for the Gramian matrix of complex Gaussian random variables," in *Proc. Int. Conf. Acoustic, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, vol. 3, pp. 1137–1140.
[46] Technical Specifications, 3GPP TS 25.221, Version 4.3.0 3GPP, 2001 [Online]. Available: http://www.3gpp.org/
[47] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[48] K. S. Miller, *Complex Stochastic Processes*, 1st ed. Reading, MA: Addison-Wesley, 1974.



Frank A. Dietrich (S'99) was born in Fürstenfeldbruck, Germany, in 1976. He received the Diploma degree (*summa cum laude*) in electrical engineering from the Munich University of Technology (TUM), Munich, Germany, in 2001. Funded by a Fulbright scholarship and a grant from the American European Engineering Exchange program he studied computer and systems engineering at the Rensselaer Polytechnic Institute, Troy, NY, from 1998 to 1999. He is currently working towards the Ph.D. degree at the Associate Institute for Signal Processing, TUM.

During summer 1999, he was involved in array processing research at Siemens Corporate Research, Princeton, NJ. Currently, he is working as a Research Assistant at the Munich University of Technology. His main research interests are in parameter identification and robust space–time signal processing with the focus on wireless communication systems and adaptive antennas.

Peter Breun was born in Erlangen, Germany, in 1980. He received the B.Sc. and Dipl.-Ing. degrees (both *summa cum laude*) in electrical engineering from the Munich University of Technology (TUM), Munich, Germany, in 2003 and 2005, respectively. He is currently working towards the Ph.D. degree at TUM.

Since 2005, he has been with the Associate Institute for Signal Processing, TUM, where he is working as a Research Assistant. His research interests include multiantenna communication systems, parameter estimation, and statistical signal processing in wireless communications.

Mr. Breun received the VDE Preis for his diploma thesis in 2005.

Wolfgang Utschick (S'93–M'97–SM'06) was born on May 6, 1964. He completed several industrial education programs before he received the diploma and doctoral degrees (both with honors) in electrical engineering from the Munich University of Technology (TUM), Munich, Germany, in 1993 and 1998, respectively.

In 1993, he became a part-time Lecturer at a technical school for industrial education. From 1998 to 2002, he was head of the Signal Processing Group at the Institute of Circuit Theory and Signal Processing at TUM. He was a guest Researcher at ETH Zurich, Switzerland, in 2000, and since 2000 he has been instrumental in the 3rd Generation Partnership Project as an academic consultant in the field of multielement antenna wireless communication systems. In October 2002, he was appointed Professor at the TUM in the Department of Electrical Engineering, where he is head of the Associate Institute for Signal Processing. He completed a biennial program for professional academic teaching and gives national and international undergraduate and graduate courses in the field of signal processing and communication systems. His research interests are statistical array signal processing, spectral theory, multiuser communication systems, multicriterion optimization theory, and kernel methods for approximation and detection.

Dr. Utschick is senior member of the VDE/ITG. He is an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. During his studies, he was awarded a scholarship from the Bavarian Ministry of Education for exceptional students and a scholarship from the Siemens AG.