Finite-Length MMSE Tomlinson–Harashima Precoding for Frequency Selective Vector Channels

Michael Joham, *Member, IEEE*, David A. Schmidt, *Student Member, IEEE*, Johannes Brehmer, *Student Member, IEEE*, and Wolfgang Utschick, *Senior Member, IEEE*

***Abstract*—We design Tomlinson–Harashima precoding for decentralized receivers and frequency-selective channels based on the minimum mean square error criterion, where the feedforward filter is restricted to have finite length. Contrary to most other publications on Tomlinson–Harashima precoding which rely on solutions for decision feedback equalization to find the corresponding precoding filters in a heuristic manner, we deduce the optimization for Tomlinson–Harashima precoding from the optimization for the linear minimum mean square error transmit filter. Thereby, we include the precoding order explicitly in the problem formulation and thus obtain the precoding filter solutions, together with the algorithms to compute the latency time, i.e., the time difference between application of the precoder at the transmitter and detection at the receiver and the precoding order from a single optimization.**

Since the algorithm for THP filter computation resulting from the optimization has a high computational complexity, we present an alternative algorithm to compute the Tomlinson–Harashima precoding filters based on a Cholesky factorization with symmetric permutation, resulting in an order of complexity that is the same as for the computation of the linear transmit filters. The simulations reveal that the latency time optimization can be omitted without performance degradation for most practical channel models, i.e., the latency time can be chosen to be the order of the feedforward filter.

***Index Terms*—Broadcast channel, Cholesky factorization, decentralized receivers, nonlinear transmit processing, Tomlinson–Harashima precoding, Wiener filtering.**

I. Introduction

FOR point-to-point *multiple input multiple output* (MIMO) systems, the standard solution is the usage of transmit and receive filters resulting from joint optimizations as in [1]–[3], which lead to a diagonalization of the channel into its eigenmodes. However, when considering the *multi-user multiple input single output* (MU-MISO) *broadcast* scenario, e.g., the downlink of a mobile communication system, several *decentralized* (*non-cooperative*) receivers are served by one centralized transmitter. Thus, the joint optimization of transmit and receive filters as in [1]–[3] is impossible, as signals of different receivers can only be processed separately. Instead, we have to use *transmit processing* (*precoding*, [4], [5]), where the receivers are restricted applying scalar weights only, which are jointly optimized with the filter applied at the transmitter. A system with precoding clearly outperforms a system with only *receive processing* (see, e.g., [6]), where the receivers have to equalize the received signals separately due to the decentralization in the broadcast setup. This is because the number of degrees of freedom available at each of the receivers is smaller than the number available at the transmitter.

A major problem for systems with precoding is the availability of the *channel state information* (CSI) at the transmitter. The CSI can be easily obtained from the channel estimation during reception in *time division duplex* systems [7], [8], since the channel can be assumed to be reciprocal. However, there is a time difference between estimation and application of the CSI leading to erroneous CSI. In *frequency division duplex* systems, the only way to obtain the CSI is feedback (e.g., [9]). Again, the CSI contains errors due to feedback delay and the quantization in the feedback channel. In either duplex case, a *robust design* of the precoding filter must be performed to take into account the CSI errors [10]–[13]. Note that we do not consider a robust design in this paper. Instead, we make the popular assumption of error-free CSI at the transmitter. However, the optimizations developed in this paper are the *necessary basis* for a robust design and can be extended to be robust as shown in [13] for example.

Linear precoding is an approach to deal with MU-MISO broadcast channels, where the data signals for the different receivers are linearly transformed to obtain the transmit signal (see [5] and references therein). Probably due to the simplicity of the concept, linear precoding based on the *zero-forcing criterion* (*transmit zero-forcing filter*—TxZF) has gained the most attention [14]–[18]. However, the TxZF is substantially outperformed by the *transmit Wiener filter* (TxWF) that minimizes the *mean square error* (MSE) under a transmit power constraint [5], [18]–[20] and finds a good trade-off between interference and noise at the receivers. Alternatively, the transmitter can also use a matched filter (e.g., [21]–[23]) that we do not consider in this paper, since it is interference limited (see, e.g., [5]). Another well-researched design criterion for linear precoding is the maximization of the minimum *signal-to-interference-plus-noise-ratio* (SINR) under a transmit power constraint (see, e.g., [24] and [25]). Although SINR maximization is of high importance, e.g., the possible data rate directly relates to the SINR, we do not employ the SINR criterion in this paper because it cannot be solved analytically (iterative algorithms of high complexity, as in [24] and [25] have to be utilized) and its application to nonlinear precoding schemes is

Manuscript received October 14, 2005; revised August 24, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Martin Haardt.

The authors are with the Associate Institute for Signal Processing, Technical University of Munich, 80290 Munich, Germany (e-mail: joham@tum.de; dschmidt@tum.de; brehmer@tum.de; utschick@tum.de).

Digital Object Identifier 10.1109/TSP.2007.893954

difficult [26] since the optimization of the precoding order is an open problem [27].

We will focus on nonlinear precoding in this paper due to the superior performance compared to linear precoders. One type of nonlinear precoding minimizes the *bit error probability* (BEP) [28], [29]. Unfortunately, analytical solutions exist only for special channel matrices [29]. Otherwise, the non-convex optimization has to be solved numerically [28]. Thus, we do not consider BEP minimization in this paper. Likewise, we do not consider *vector precoding* [30], [31] since it has prohibitive complexity due to the necessary closest point search in a lattice (non-polynomial in the number of receivers, e.g., [32] and [33]).

The nonlinear *Tomlinson–Harashima precoding* (THP) is based on modulo operators at the receivers and has a complexity comparable to that of linear precoders, as will be shown in this paper. Note that THP is strongly related to dirty paper coding [34]. In fact, THP is a suboptimal implementation of dirty paper coding (see e.g., [35]). THP was originally proposed in [36] and [37] to combat *intersymbol interference* (ISI) in single-user transmission over frequency selective channels (see also [38] and references therein). Gibbard *et al.* [39] employed THP for an asymmetric single-user transmission to simplify the mobile terminal. In [40], Ginis *et al.* applied *zero-forcing* THP (ZF-THP) to a *digital subscriber line* (DSL) system without optimizing the precoding order, and Fischer *et al.* [41], [42] performed spatial equalization (or multi-user separation) with ZF-THP in multi-user transmission over frequency flat channels, where the precoding order is optimized. In [43], Liu *et al.* designed ZF-THP for *code division multiple access* (CDMA) systems similar to [41]. Joham *et al.* introduced THP based on the *minimum mean square error* (MMSE) criterion for decentralized receivers over frequency flat channels [44] and frequency-selective channels [45], [46], where both the precoding order and the latency time are optimized. We will call this the THP type *Wiener filter THP* (WF-THP) in the sequel. In [47], Fischer *et al.* introduced ordered ZF-THP with an IIR feedforward filter for frequency selective channels. Choi *et al.* [48] proposed ZF-THP for multi-user systems with frequency-selective channels, where the precoding order is optimized but the obviously suboptimum choice of zero latency time is made. Degen *et al.* [49] designed WF-THP for frequency-selective channels in the frequency domain. Schubert *et al.* optimized WF-THP for frequency flat channels with different weights at the receivers in [50], where an iterative solution for the equal MSE optimization was developed, but an exhaustive search for the minimization of the total MSE had to be applied. For both WF-THP types, Schubert *et al.* could not give any precoding order optimization. In [51], Shao *et al.* derived WF-THP without precoding order in an intuitive way, and Kusume *et al.* [52] presented an algorithm to compute ordered WF-THP for flat fading channels whose complexity has the same order as linear precoding. Liu *et al.* [53] investigated precoding order algorithms for ZF-THP and designed ZF-THP with equal weights at the receivers as in [11] and [44]. THP with erroneous or partial CSI was considered in [11]–[13], [54], and [55].

Most of the existing literature introduces THP intuitively and bases the THP filters on a zero-forcing condition without any optimization. Thus, it is unclear in which sense the THP filters are optimal. Contrary to the filters, the precoding order is found with some optimization. However, it is not ensured that the optimization for the precoding order fits the underlying optimization for the filters. For THP with FIR filters, the need for an optimization of the latency time seems to be unknown.

Our approach to THP is different. We base the design on an optimization, namely, on MSE minimization. The speciality of our formulation is the inclusion of the precoding order and the latency time in the THP optimization. Therefore, we guarantee that the THP filters, the precoding order, and the latency time are designed aiming at the same goal. In previous work [45], [46], we have already published the resulting WF-THP filters, which are superior to the state-ofthe-art zero-forcing THP filters. However, the notation used in this paper is simpler compared to [45] and [46], and we propose a low complexity algorithm for WF-THP over frequency-selective channels.

The contributions of this paper are as follows.

1) We design WF-THP with optimized latency time and precoding order based on a single optimization, where the feedforward filter is restricted to be FIR.
2) We develop a low complexity algorithm to compute the WF-THP filters. The analysis of the algorithm reveals that the WF-THP filters with optimized latency time and precoding order can be computed with the same order of complexity as the linear TxWF.
3) From the structure of the WF-THP solution, it becomes clear that the optimum latency time cannot be smaller than the order of the feedforward filter. Additionally, we conjecture from the simulations that the latency time optimization can even be skipped in realistic scenarios.
4) The simulations reveal that WF-THP clearly outperforms ZF-THP and the linear precoding approaches.

Since we deduce the WF-THP optimization from the TxWF optimization, we review the TxWF in Section II. In Section III, we show how the WF-THP optimization has to be formulated and derive the WF-THP filters. Then, we show how the WF-THP filters can be computed efficiently by employing a symmetrically permuted Cholesky factorization (e.g., [56]) in Section IV. The simulation results can be found in Section V.

A. Notation

Vectors and matrices are denoted by lower case bold and capital bold letters, respectively. We use $\mathrm{E}[\bullet]$, "$*$," "$\otimes$," $\mathrm{Re}(\bullet)$, $\mathrm{tr}(\bullet)$, $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, $\|\bullet\|_2$, and $\|\bullet\|_{\mathrm{F}}$ for expectation, convolution, Kronecker product, real part of the argument, trace of a matrix, complex conjugation, transposition, conjugate transposition, Euclidian norm, and Frobenius norm, respectively. All random sequences are assumed to be zero-mean and stationary. The covariance matrix of the vector random process $\boldsymbol{x}[n]$ is denoted by $\boldsymbol{R}_{\boldsymbol{x}} = \mathrm{E}[\boldsymbol{x}[n]\boldsymbol{x}^{\mathrm{H}}[n]]$, whereas the variance of the scalar random process $y[n]$ is denoted by $\sigma_y^2 = \mathrm{E}[|y[n]|^2]$. The $N \times N$ identity matrix is $\mathbf{1}_N$, whose ith column is $\boldsymbol{e}_i$. We use $\mathbf{0}_{N\times M}$ and $\mathbf{0}_N$ for the $N \times M$ zero matrix and the N-dimensional zero vector, respectively. The block-diagonal matrix with the blocks $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$ on its diagonal is denoted by $\mathrm{blockdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N)$. The unit impulse is $\delta[n]$, which is one for $n = 0$ and zero otherwise.

We use the same definition for the derivative $\partial a(\boldsymbol{B})/\partial \boldsymbol{B}$ of a scalar $a(\boldsymbol{B})$ with respect to the $N \times M$ matrix $\boldsymbol{B}$ as in [57], i.e., each entry of the resulting $N \times M$ matrix is the derivative of the scalar $a(\boldsymbol{B})$ with respect to the respective entry of $\boldsymbol{B}$. Since the cost functions of the investigated optimizations are not analytic, we employ the following derivative (e.g., [58]):

$$\frac{\partial b(\boldsymbol{A})}{\partial \boldsymbol{A}} = \frac{1}{2}\left(\frac{\partial b(\boldsymbol{A})}{\partial \mathrm{Re}(\boldsymbol{A})} - \mathrm{j}\frac{\partial b(\boldsymbol{A})}{\partial \mathrm{Im}(\boldsymbol{A})}\right)$$

where $b(\boldsymbol{A}) \in \mathbb{C}$ and $\boldsymbol{A} \in \mathbb{C}^{N\times M}$.

II. Transmit Wiener Filter (TxWF)

The linear TxWF with the restriction to be FIR for frequency-selective channels was first derived in [59] and used when investigating a single-user system. The extension to multi-user CDMA systems is found in [60]. We use a different notation than in [59] and [60] in order to be compatible with the later-needed notation for THP.

A. System Model

In a system with linear precoding, the data signal $\boldsymbol{s}[n] \in \mathbb{A}^B$ is passed through the FIR precoding filter of order L:[1]

$$\boldsymbol{P}[n] = \sum_{\ell=0}^{L} \boldsymbol{P}_\ell \delta[n-\ell]$$

where $\boldsymbol{P}_\ell \in \mathbb{C}^{N_\mathrm{a}\times B}$, $\ell = 0,\ldots,L$. Here, $\mathbb{A}$ denotes the symbol alphabet, B is the number of receivers, and N_a is the number of transmit antennas. For notational brevity, we make the simplifying assumption that $\mathrm{E}[\boldsymbol{s}[n]\boldsymbol{s}^\mathrm{H}[n+\nu]] = \boldsymbol{R}_{\boldsymbol{s}}\delta[\nu] \in \mathbb{C}^{B\times B}$, i.e., the symbols are temporally uncorrelated. The resulting transmit signal

$$\boldsymbol{y}[n] = \sum_{\ell=0}^{L} \boldsymbol{P}_\ell \boldsymbol{s}[n-\ell] \in \mathbb{C}^{N_\mathrm{a}}$$

propagates over the channel of order Q

$$\boldsymbol{H}[n] = \sum_{q=0}^{Q} \boldsymbol{H}_q \delta[n-q]$$

where $\boldsymbol{H}_q \in \mathbb{C}^{B\times N_\mathrm{a}}$, $q = 0,\ldots,Q$. The channel output is perturbed by the zero-mean noise $\boldsymbol{\eta}[n]$ with covariance matrix $\boldsymbol{R}_{\boldsymbol{\eta}} = \mathrm{E}[\boldsymbol{\eta}[n]\boldsymbol{\eta}^\mathrm{H}[n]]$ and weighted with the scalar receive filter $g \in \mathbb{C}$ to get the estimate (cf. Fig. 1)

$$\begin{aligned}\hat{\boldsymbol{s}}[n] &= g\boldsymbol{H}[n] * \boldsymbol{P}[n] * \boldsymbol{s}[n] + g\boldsymbol{\eta}[n] \\ &= g\sum_{q=0}^{Q}\sum_{\ell=0}^{L} \boldsymbol{H}_q\boldsymbol{P}_\ell \boldsymbol{s}[n-q-\ell] + g\boldsymbol{\eta}[n] \in \mathbb{C}^B \end{aligned} \quad (1)$$

[1]We consider a fixed filter order L. When optimizing also with respect to L, we would end up with $L = \infty$, i.e., $\boldsymbol{P}[n]$ is IIR.

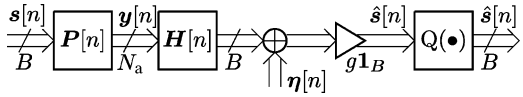

Fig. 1. System model for linear precoding.

which is then mapped to the symbol alphabet by the nearest neighbor quantizer $\mathrm{Q}(\bullet)$.

By defining

$$\begin{aligned}\boldsymbol{H} &= \sum_{q=0}^{Q}[\boldsymbol{0}_{L+1\times q}, \boldsymbol{1}_{L+1}, \boldsymbol{0}_{L+1\times Q-q}]^\mathrm{T} \otimes \boldsymbol{H}_q \quad \text{and} \\ \boldsymbol{P} &= \left[\boldsymbol{P}_0^\mathrm{T}, \ldots, \boldsymbol{P}_L^\mathrm{T}\right]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{a}(L+1)\times B} \end{aligned} \quad (2)$$

where $\boldsymbol{H}$ is a $B(Q+L+1)\times N_\mathrm{a}(L+1)$ block Toeplitz matrix, the coefficients of the convolution of $\boldsymbol{H}[n]$ and $\boldsymbol{P}[n]$ can be computed by the product $\boldsymbol{HP} \in \mathbb{C}^{B(Q+L+1)\times B}$, i.e., the ith $B\times B$ block element of $\boldsymbol{HP}$ is the ith coeffient $\sum_{q+\ell=i-1}\boldsymbol{H}_q\boldsymbol{P}_\ell$ of $\boldsymbol{H}[n] * \boldsymbol{P}[n]$. Thus, we can rewrite (1)

$$\hat{\boldsymbol{s}}[n] = g\sum_{i=0}^{Q+L} \boldsymbol{S}^{(i)}\boldsymbol{HP}\boldsymbol{s}[n-i] + g\boldsymbol{\eta}[n]. \quad (3)$$

Here, we introduced the selection matrix

$$\boldsymbol{S}^{(i)} = \boldsymbol{e}_{i+1}^\mathrm{T} \otimes \boldsymbol{1}_B \in \{0,1\}^{B\times B(Q+L+1)} \quad (4)$$

which gives the $Bi+1$th row up to the $B(i+1)$th row of a matrix when applied from the left. Therefore, $\boldsymbol{S}^{(i)}\boldsymbol{HP}$ is the ith coefficient of $\boldsymbol{H}[n] * \boldsymbol{P}[n]$. Note that $\boldsymbol{e}_i \in \{0,1\}^{Q+L+1}$.

B. Derivation of the TxWF

The linear TxWF is found by minimizing the MSE between the data signal $\boldsymbol{s}[n]$ and the estimate $\hat{\boldsymbol{s}}[n]$ under a total transmit power constraint (see [5]):[2]

$$\begin{aligned}\left\{\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{lin}}, g_{\mathrm{WF}}^{\mathrm{lin}}, \nu_{\mathrm{WF}}^{\mathrm{lin}}\right\} = \underset{\{\boldsymbol{P},g,\nu\}}{\mathrm{argmin}}\, &\mathrm{E}\left[\|\boldsymbol{s}[n-\nu] - \hat{\boldsymbol{s}}[n]\|_2^2\right] \\ &\text{s.t.}: \mathrm{E}\left[\|\boldsymbol{y}[n]\|_2^2\right] = E_{\mathrm{tr}}. \end{aligned} \quad (5)$$

Here, we introduced the latency time $\nu \in \{0,\ldots,Q+L\}$, i.e., the data signal $\boldsymbol{s}[n]$ is estimated ν time steps after it has been applied to the input of the precoding filter $\boldsymbol{P}[n]$. With the made assumptions, the MSE can be written as [see (3)]

$$\begin{aligned}\varepsilon(\boldsymbol{P}, g, \nu) &= \mathrm{E}\left[\|\boldsymbol{s}[n-\nu] - \hat{\boldsymbol{s}}[n]\|_2^2\right] \\ &= \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{s}}) - \mathrm{Re}\left(\mathrm{tr}\left(g\boldsymbol{S}^{(\nu)}\boldsymbol{HPR}_{\boldsymbol{s}}\right)\right) + |g|^2\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}}) \\ &\quad + |g|^2(\boldsymbol{HPR}_{\boldsymbol{s}}\boldsymbol{P}^\mathrm{H}\boldsymbol{H}^\mathrm{H}) \end{aligned}$$

[2]We could also apply an inequality for the transmit power constraint, that is, $\mathrm{E}[\|\boldsymbol{y}[n]\|_2^2] \le E_{\mathrm{tr}}$, as in [11] and [17]. However, the transmit power constraint would always be active, i.e., the constraint is an equality, because the MSE can be decreased by using more transmit power. To simplify the derivation of the TxWF and WF-THP, we use an equality in the transmit power constraint.

where we used $\sum_{i=0}^{Q+L} \boldsymbol{S}^{(i),\mathrm{T}}\boldsymbol{S}^{(i)} = \mathbf{1}_{B(Q+L+1)}$ for the last summand. For the transmit power constraint, we get

$$\mathrm{E}\left[||\boldsymbol{y}[n]||_2^2\right] = \mathrm{tr}(\boldsymbol{P}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{P}^{\mathrm{H}}).$$

With the last two results, the optimization (5) can be solved with the method of Lagrangian multipliers. We set the derivatives of the Lagrangian function with respect to the filter $\boldsymbol{P}$ and g to zero and incorporate the transmit power constraint, which leads to [see [5]]

$$\begin{aligned}\boldsymbol{P}(\nu) &= \frac{1}{g(\nu)}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{A}_{\mathrm{WF}}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{a}}(L+1)\times B} \quad \text{and} \\ g(\nu) &= \sqrt{\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{A}_{\mathrm{WF}}^{-2}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{S}^{(\nu)}\right)/E_{\mathrm{tr}}} \end{aligned} \tag{6}$$

with

$$\boldsymbol{A}_{\mathrm{WF}} = \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \xi\mathbf{1}_{B(Q+L+1)} \quad \text{and} \quad \xi = \frac{\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}})}{E_{\mathrm{tr}}} \tag{7}$$

where we used the *matrix inversion lemma* (see, e.g. [61]) and restrict $g(\nu)$ to be positive real. This result for the precoder $\boldsymbol{P}(\nu)$ and the estimator $g(\nu)$ is an optimizer of (5) for a given latency time ν. Since the latency time is discrete valued, we have to try all $Q+L+1$ possible values for ν and choose the one with minimum MSE $\varepsilon(\boldsymbol{P}(\nu), g(\nu), \nu)$ (cf. [59]):[3]

$$\nu_{\mathrm{WF}}^{\mathrm{lin}} = \underset{\nu\in\{0,\ldots,Q+L\}}{\mathrm{argmin}} \ \mathrm{tr}\left(\boldsymbol{S}^{(\nu)}\boldsymbol{A}_{\mathrm{WF}}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{R}_{\boldsymbol{s}}\right).$$

The resulting $\nu_{\mathrm{WF}}^{\mathrm{lin}}$ has to be plugged into $\boldsymbol{P}(\nu)$ and $g(\nu)$ to get the optimum filters $\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{lin}}$ and $g_{\mathrm{WF}}^{\mathrm{lin}}$, respectively.

C. Transmit Zero-Forcing Filter (TxZF)

The optimization for the TxZF follows from (5) by incorporating the constraint $\mathrm{E}[\hat{\boldsymbol{s}}[n]|\boldsymbol{s}[n], \boldsymbol{s}[n-1], \ldots] = \boldsymbol{s}[n-\nu]$, or equivalently, $g\boldsymbol{H}\boldsymbol{P} = \boldsymbol{S}^{(\nu),\mathrm{T}}$. Following similar steps as for the TxWF, it can be shown that the TxZF can be obtained from the TxWF solution by replacing $\boldsymbol{A}_{\mathrm{WF}}$ with $\boldsymbol{A}_{\mathrm{ZF}} = \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}$. Note that the TxZF only exists for $N_{\mathrm{a}}(L+1) \geq B(Q+L+1)$.

III. Wiener Filter Tomlinson Harashima Precoding (WF-THP)

THP for decentralized receivers based on the MMSE criterion was first presented in [45], where frequency flat and frequency-selective channels were considered. In [46], the WF-THP optimization of [45] was extended such that not only the FIR filter coefficients but also the latency time and the precoding order evolve from one optimization. We will basically follow the approach of [46] but employ a simpler notation, which allows a straightforward algorithmic solution in the next section.

[3]Since $\boldsymbol{A}_{\mathrm{WF}}$ has no special structure [contrary to $\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O})}$ in (23)], any value for ν between 0 and $Q+L$ can be optimal.

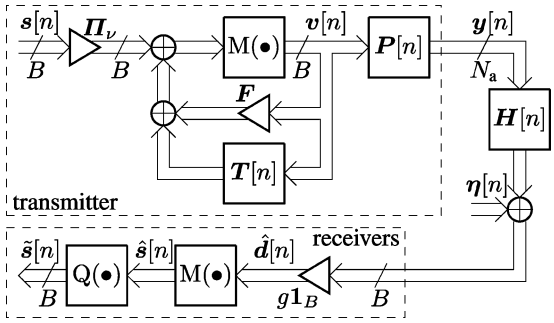

Fig. 2. System model for Tomlinson–Harashima precoding.

A. Preliminaries and Principle of THP

When employing THP, the receivers have to be equipped with the modulo operators

$$\mathrm{Mod}(x) = x - \tau\left\lfloor\frac{\mathrm{Re}(x)}{\tau}+\frac{1}{2}\right\rfloor - \mathrm{j}\tau\left\lfloor\frac{\mathrm{Im}(x)}{\tau}+\frac{1}{2}\right\rfloor \in \mathbb{V}$$

where $x \in \mathbb{C}$, τ is the modulo constant, $\lfloor\bullet\rfloor$ denotes the floor operation which gives the largest integer smaller than or equal to the argument, and

$$\mathbb{V} = \{x + \mathrm{j}y | x, y \in [-\tau/2, \tau/2)\}$$

is the fundamental Voronoi region of the lattice corresponding to $\mathrm{Mod}(\bullet)$. The modulo operators in Fig. 2 are defined element-wise, i.e.,

$$\mathrm{M}(\boldsymbol{x}) = [\mathrm{Mod}(x_1), \ldots, \mathrm{Mod}(x_B)]^{\mathrm{T}} \in \mathbb{V}^B.$$

Here, x_i denotes the ith entry of $\boldsymbol{x} \in \mathbb{C}^B$. Note that the modulo operation can also be expressed as

$$\mathrm{M}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{a} \in \mathbb{V}^B \tag{8}$$

with the auxiliary vector $\boldsymbol{a} \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$. We can interpret the modulo operator $\mathrm{M}(\bullet)$ as a device that chooses the auxiliary vector $\boldsymbol{a}$ from the lattice $\tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$ such that $\mathrm{M}(\bullet) \in \mathbb{V}^B$. Therefore, we have $\mathrm{M}(\boldsymbol{x}) = \boldsymbol{x}$, if $\boldsymbol{x} \in \mathbb{V}^B$. In the sequel, we make the standard assumption that the modulo constant τ is sufficiently large (e.g., $\tau = 2\sqrt{2}$ and $\tau = 8/\sqrt{10}$ for quadrature phase shift keying (QPSK) and 16QAM symbols with unit variance, respectively) such that $\mathbb{A} \subset \mathbb{V}$ is fulfilled. Thus, $\boldsymbol{s}[n] \in \mathbb{V}^B$ and $\boldsymbol{s}[n]$ is not changed when the modulo operator is applied. Note that $\mathbb{A} \subset \mathbb{V}$ is crucial, because we assume that the receivers apply the same nearest-neighbor quantizer $\mathrm{Q}(\bullet)$ as for linear precoding. Without $\mathbb{A} \subset \mathbb{V}$, the quantizer must be redesigned, because all elements of $\mathbb{A}$ are mapped into $\mathbb{V}$, i.e., the constellation $\mathbb{A}'$ seen by the receiver is different from $\mathbb{A}$. This restriction on $\mathbb{A}$, however, does not mean that the modulo operators are inactive and can be dropped, since the input of the modulo operator at the transmitter is the sum of $\boldsymbol{s}[n]$ and the outputs of the two feedback filters $\boldsymbol{F}$ and $\boldsymbol{T}[n]$.

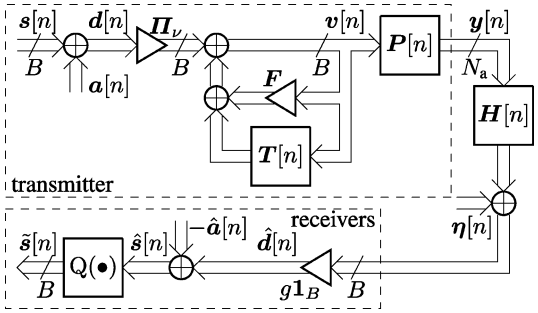

Fig. 3. System model for Tomlinson–Harashima precoding with auxiliary signals.

The modulo operators at the receivers introduce additional degrees of freedom, since any vector $\boldsymbol{a}' \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$ can be added to the modulo input without changing the output, that is, $\mathrm{M}(\boldsymbol{x} + \boldsymbol{a}') = \mathrm{M}(\boldsymbol{x})$. Consequently, the transmitter has the freedom to generate the signals $\boldsymbol{x}[n] + \boldsymbol{a}[n]$ at the receivers instead of $\boldsymbol{x}[n]$, i.e., the transmitter can choose the fundamental regions of the modulo operators at the receivers by the choice of some signal $\boldsymbol{a}[n] \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$. These degrees of freedom are optimally utilized by vector precoding (see, e.g., [30] and [31]). However, vector precoding is very complex due to the necessary closest point search in a lattice. To avoid this computationally costly search (exponential in the number of receivers; see, e.g., [32] and [33]), the heuristic which limits the amplitude of the signal $\boldsymbol{v}[n]$ is employed for THP, i.e., the sum of the permuted data signal $\boldsymbol{\Pi}_\nu \boldsymbol{s}[n]$ and the outputs of the spatial feedback filter $\boldsymbol{F}$ and the temporal feedback filter $\boldsymbol{T}[n]$ are passed through the modulo operator $\mathrm{M}(\bullet)$ (see Fig. 2).[4] By limiting the amplitude of $\boldsymbol{v}[n]$, the power of the transmit signal $\boldsymbol{y}[n]$ is also limited indirectly for a fixed feedforward filter $\boldsymbol{P}[n]$. As the *zero-forcing* vector precoding minimizes the transmit power for a fixed transformation by the choice of the signal $\boldsymbol{a}[n]$ (see [30]), the heuristic employed for THP is reasonable for ZF-THP. Although transmit power minimization is not optimum for *MMSE* vector precoding (see [31]), most publications on regularized vector precoding [30], [51], [62] also heuristically minimize the transmit power for a fixed transformation, leading to clearly better results than zero-forcing vector precoding. Therefore, the heuristic which limits the amplitude of the signal $\boldsymbol{v}[n]$ by the modulo operator $\mathrm{M}(\bullet)$ is also reasonable for WF-THP.

With the representation (8) of the modulo operator, we get Fig. 3 from Fig. 2 by introducing the auxiliary signals $\boldsymbol{\Pi}_\nu \boldsymbol{a}[n]$ at the transmitter and $-\hat{\boldsymbol{a}}[n]$ at the receivers, where the summation of the auxiliary signal $\boldsymbol{\Pi}_\nu \boldsymbol{a}[n]$ at the transmitter has been moved to the front of the permutation matrix $\boldsymbol{\Pi}_\nu$. Note that Figs. 2 and 3 are fully equivalent, as long as the auxiliary signals in Fig. 3 are chosen according to (8). Based on Figs. 2 and 3, the following remarks on THP are possible.

1) The modulo operation at the transmitter automatically computes the desired value $\boldsymbol{a}[n] \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$ for $\hat{\boldsymbol{a}}[n] \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$ at the receivers following the strategy to limit the amplitude of the signal $\boldsymbol{v}[n] \in \mathbb{V}^B$. Thereby, the *virtual* desired signal $\boldsymbol{d}[n]$ for the input $\hat{\boldsymbol{d}}[n]$ of the modulo operators at the receivers is created by adding the signal $\boldsymbol{a}[n]$ to the data signal $\boldsymbol{s}[n]$ (see Fig. 3). Note that $\boldsymbol{d}[n]$ does not exist in Fig. 2.
2) Not only the transmit signal $\boldsymbol{y}[n]$ but also the desired part of the input $\hat{\boldsymbol{d}}[n]$ of the modulo operators at the receivers linearly depend on the virtual desired signal $\boldsymbol{d}[n]$ (see Fig. 3). Hence, the system between $\boldsymbol{d}[n]$ and $\hat{\boldsymbol{d}}[n]$ has similar properties as the system with linear precoding in Fig. 1. We can infer that the optimization techniques for linear precoding are also applicable to the system in Fig. 3 when we use $\hat{\boldsymbol{d}}[n]$ as the estimate and $\boldsymbol{d}[n]$ as the desired signal in the optimization. Note, however, that the system between $\boldsymbol{s}[n]$ and $\hat{\boldsymbol{s}}[n]$ is nonlinear due to the addition of $\boldsymbol{a}[n]$ and $-\hat{\boldsymbol{a}}[n]$.
3) The estimates $\hat{\boldsymbol{s}}[n]$ at the receivers are the same as the data signal $\boldsymbol{s}[n]$ if the inputs $\hat{\boldsymbol{d}}[n]$ of the modulo operators at the receivers are the same as the virtual desired signal $\boldsymbol{d}[n] = \boldsymbol{s}[n] + \boldsymbol{a}[n]$, since $\hat{\boldsymbol{a}}[n] = \boldsymbol{a}[n]$, in this case, is due to the assumption that $\boldsymbol{s}[n] \in \mathbb{V}^B$. We can conclude that $\hat{\boldsymbol{d}}[n]$ must be as similar to $\boldsymbol{d}[n]$ as possible. This observation motivates us to employ the MMSE criterion when designing THP filters. Additionally, Erez *et al.* showed in [63] that the Shannon capacity for AWGN channels can be reached by a system with a modulo receiver if the received signal is scaled with an MMSE weight prior to the modulo operation (see also [64]). In [63], an intuitive explanation can additionally be found as to why such an *inflated lattice* decoder outperforms a lattice decoder without the MMSE weight.
4) To ensure the realizability of the feedback loop with $\boldsymbol{F}$ and $\boldsymbol{T}[n]$, the filter $\boldsymbol{T}[n]$ has to be strictly causal, i.e., $\boldsymbol{T}[n] = \boldsymbol{0}_{B\times B}$ for $n \leq 0$, and $\boldsymbol{F} \in \mathbb{C}^{B\times B}$ is constrained to be lower triangular with zero main diagonal.[5] Otherwise, we would end up with a non-causal feedback and/or a delay-free loop.
5) Usually, it is stated that the feedforward filter $\boldsymbol{P}[n]$ is necessary to form a minimum phase channel, i.e., the interference of preceding symbols is canceled by $\boldsymbol{P}[n]$, and the feedback filters $\boldsymbol{F}$ and $\boldsymbol{T}[n]$ remove the remaining interference. At this point, we can infer this allocation of responsibilities neither from Fig. 2 nor from Fig. 3. However, we will see in the following that the filters resulting from the WF-THP optimization show these properties (see Section III-D).
6) For ZF-THP, it can be shown (see, e.g., [4, Th. 3.1]) that the modulo operator at the transmitter in combination with the feedback filters $\boldsymbol{T}[n]$ and $\boldsymbol{F}$ lead to a decorrelation of the entries of $\boldsymbol{v}[n]$, that is, $\mathrm{E}[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+\nu]] = \sigma_{\boldsymbol{v}}^2 \boldsymbol{1}_B \delta[\nu]$. Additionally, the entries of $\boldsymbol{v}[n]$ are uniformly distributed over $\mathbb{V}$. Thus, $\sigma_{\boldsymbol{v}}^2 = \tau^2/6$. Although this result is only applicable to WF-THP for medium to high SNR, we make

[4]Note that the separation of the feedback operation into $\boldsymbol{F}$ and $\boldsymbol{T}[n]$ is only done for notational simplicity, because $\boldsymbol{F}$ has a special structure, i.e., it is strictly lower triangular, and $\boldsymbol{T}[n]$ only has to be strictly causal. As we will see later, $\boldsymbol{T}[n]$ suppresses the interference of the post-cursors, and $\boldsymbol{F}$ combats the interference due to the other scalar entries of $\boldsymbol{d}[n]$

[5]The choice of $\boldsymbol{F}$ to be lower triangular with zero main diagonal is made for notational simplicity. Other structures for $\boldsymbol{F}$ are possible as well, e.g., upper triangular with zero main diagonal.

the assumption that $\mathrm{E}[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+\nu]] = \sigma_{\boldsymbol{v}}^2 \mathbf{1}_B \delta[\nu]$ with $\sigma_{\boldsymbol{v}}^2 = \tau^2/6$ holds (see Section III-D). Note that the known statistics of $\boldsymbol{v}[n]$ and the unknown statistics of $\boldsymbol{d}[n]$ are key differences of THP compared to linear precoding.

7) For the argumentation that the receiver can recover the data signal $\boldsymbol{s}[n]$, the special structure of the transmitter, i.e., the partitioning into the permutation matrix $\boldsymbol{\Pi}_\nu$ and the filters $\boldsymbol{P}[n]$, $\boldsymbol{F}$, and $\boldsymbol{T}[n]$, is irrelevant. We could compute the signal $\boldsymbol{a}[n]$ directly by an exhaustive search and transform the sum of $\boldsymbol{a}[n]$ and the data signal $\boldsymbol{s}[n]$ by one linear filter to get the transmit signal $\boldsymbol{y}[n]$. However, the exhaustive search is prohibitively complex and is avoided by the successive computation of the elements of $\boldsymbol{a}[n]$ with the THP feedback loop consisting of $\boldsymbol{F}$, $\boldsymbol{T}[n]$, and the element-wise modulo operator $\mathrm{M}(\bullet)$. The structure of the filters $\boldsymbol{F}$ and $\boldsymbol{T}[n]$, together with the element-wise definition of the modulo operator $\mathrm{M}(\bullet)$, cause the successive computation of the signal $\boldsymbol{v}[n]$, that is, the ith entry $v_i[n]$ is computed based on the $i-1$ preceding entries $v_1[n], \ldots, v_{i-1}[n]$.
8) Since the signal $\boldsymbol{v}[n]$ is found successively, the order of computing the entries of $\boldsymbol{v}[n]$ has an influence on the performance of the precoder. The importance of the precoding order can also be understood with Fig. 3. As the system between $\boldsymbol{d}[n]$ and $\hat{\boldsymbol{d}}[n]$ is linear, the combination of $\boldsymbol{\Pi}_\nu$, $\boldsymbol{F}$, $\boldsymbol{T}[n]$, and $\boldsymbol{P}[n]$ is a linear TxWF (not FIR due to the feedback loop). As the statistics of $\boldsymbol{d}[n]$ depend on the feedback filters $\boldsymbol{F}$ and $\boldsymbol{T}[n]$ (cf. (10); the statistics of $\boldsymbol{v}[n]$ are given), the linear TxWF and, thus, its performance, depends on the choice of $\boldsymbol{F}$ and $\boldsymbol{T}[n]$, which depend on $\boldsymbol{\Pi}_\nu$. The reordering operation is expressed by the permutation matrix $\boldsymbol{\Pi}_\nu$.
9) Besides the advantage of additional degrees of freedom for the transmitter, the modulo operators at the receivers have the disadvantage that the system performance is deteriorated by the additionally allowed constellation points at the receivers. These additional points are introduced by the modulo operator $\mathrm{M}(\bullet)$, since two different received signals can lead to the same modulo output. Especially, the new neighbors for the outer symbols of the constellation set induce additional errors. Thus, this effect is more pronounced for small constellations (e.g., QPSK) because most of the symbols are outer symbols.

Following the above remarks, we will design the THP filters by minimizing the MSE between the virtual desired signal $\boldsymbol{d}[n]$ and the estimates $\hat{\boldsymbol{d}}[n]$ since the auxiliary signals $\boldsymbol{a}[n]$ and $\hat{\boldsymbol{a}}[n]$ are generated automatically by the modulo operators. In this optimization, we have to constrain the structure of the feedback filters to ensure realizability.

B. System Model

As we know the statistical properties of the modulo output $\boldsymbol{v}[n]$ at the transmitter, i.e., $\mathrm{E}[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+\nu]] = \sigma_{\boldsymbol{v}}^2 \mathbf{1}_B \delta[n]$, we will derive expressions for the virtual desired signal $\boldsymbol{d}[n]$ and the estimates $\hat{\boldsymbol{d}}[n]$ based on $\boldsymbol{v}[n]$. From Fig. 3, we can see that

$$\boldsymbol{v}[n] = \boldsymbol{\Pi}_\nu \boldsymbol{d}[n] + \boldsymbol{F}\boldsymbol{v}[n] + \boldsymbol{T}[n] * \boldsymbol{v}[n] \in \mathbb{V}^B$$

where the spatial feedback filter $\boldsymbol{F}$ has to be lower triangular with zero main diagonal and the temporal feedback filter must be strictly causal:

$$\boldsymbol{T}[n] = \sum_{i=1}^{N_T} \boldsymbol{T}_i \delta[n-i]$$

with $\boldsymbol{T}_i \in \mathbb{C}^{B\times B}$ and some filter order N_T to be determined later. The B-tuple

$$\mathcal{O} = (b_1, \ldots, b_B) \quad \text{with} \quad \{b_1, \ldots, b_B\} = \{1, \ldots, B\}$$

contains the indices of the reordering, i.e., the ith entry $v_i[n]$ of $\boldsymbol{v}[n]$ corresponds to the b_ith entry $s_{b_i}[n]$ of the data signal $\boldsymbol{s}[n]$. In the sequel, $\mathcal{O}$ will be called the precoding order. The permutation matrix representing the reordering is defined as

$$\boldsymbol{\Pi}_\nu = \sum_{i=1}^{B} \boldsymbol{e}_i \boldsymbol{e}_{b_i}^{\mathrm{T}} \in \{0,1\}^{B\times B} \tag{9}$$

with $\boldsymbol{e}_i, \boldsymbol{e}_{b_i} \in \{0,1\}^B$ and $\boldsymbol{\Pi}_\nu^{-1} = \boldsymbol{\Pi}_\nu^{\mathrm{T}}$. Consequently, we get for the virtual desired signal

$$\boldsymbol{d}[n] = \boldsymbol{\Pi}_\nu^{\mathrm{T}}(\mathbf{1}_B - \boldsymbol{F})\boldsymbol{v}[n] - \boldsymbol{\Pi}_\nu^{\mathrm{T}} \sum_{i=1}^{N_T} \boldsymbol{T}_i \boldsymbol{v}[n-i]. \tag{10}$$

Comparing Fig. 3 with Fig. 1, we observe that the estimate $\hat{\boldsymbol{d}}[n]$ has the same dependence on $\boldsymbol{v}[n]$ in Fig. 3 as $\hat{\boldsymbol{s}}[n]$ has on $\boldsymbol{s}[n]$ in Fig. 1. Hence, we can reuse (3):

$$\hat{\boldsymbol{d}}[n] = g \sum_{i=0}^{Q+L} \boldsymbol{S}^{(i)} \boldsymbol{H}\boldsymbol{P}\boldsymbol{v}[n-i] + g\boldsymbol{\eta}[n].$$

The channel matrix $\boldsymbol{H}$ and the feedforward filter matrix $\boldsymbol{P}$ can be found in (2). The selection matrix $\boldsymbol{S}^{(i)}$ is defined in (4). For the development of the efficient algorithm to compute the WF-THP filters together with the precoding order and the latency time in the next section, it is helpful to define the permuted $B(Q+L+1) \times N_{\mathrm{a}}(L+1)$ block Toeplitz channel matrix

$$\boldsymbol{C} = \mathrm{blockdiag}(\boldsymbol{\Pi}_0, \ldots, \boldsymbol{\Pi}_{Q+L})\boldsymbol{H} \tag{11}$$

where we have introduced the $B \times B$ permutation matrices $\boldsymbol{\Pi}_0, \ldots, \boldsymbol{\Pi}_{\nu-1}, \boldsymbol{\Pi}_{\nu+1}, \ldots, \boldsymbol{\Pi}_{Q+L}$. Note that only $\boldsymbol{\Pi}_\nu$ defined in (9) is used by the precoder. The other permutation matrices $\boldsymbol{\Pi}_i$, $i \neq \nu$ can be chosen arbitrarily and are only introduced to simplify the developed algorithms in Section IV. By above multiplication with the block diagonal permutation matrix $\mathrm{blockdiag}(\boldsymbol{\Pi}_0, \ldots, \boldsymbol{\Pi}_{Q+L})$, the $B \times N_{\mathrm{a}}(L+1)$ block rows of $\boldsymbol{H}$ are permuted differently, e.g., the $i+1$th block row is permuted with $\boldsymbol{\Pi}_i$. With $\boldsymbol{\Pi}_i^{\mathrm{T}} = \boldsymbol{\Pi}_i^{-1}$, the estimate reads as

$$\begin{aligned}\hat{\boldsymbol{d}}[n] &= g \sum_{i=0}^{Q+L} \boldsymbol{\Pi}_i^{\mathrm{T}} \boldsymbol{\Pi}_i \boldsymbol{S}^{(i)} \boldsymbol{H}\boldsymbol{P}\boldsymbol{v}[n-i] + g\boldsymbol{\eta}[n] \\ \hat{\boldsymbol{d}}[n] &= g \sum_{i=0}^{Q+L} \boldsymbol{\Pi}_i^{\mathrm{T}} \boldsymbol{S}^{(i)} \boldsymbol{C}\boldsymbol{P}\boldsymbol{v}[n-i] + g\boldsymbol{\eta}[n]\end{aligned} \tag{12}$$

since $\boldsymbol{\Pi}_i \boldsymbol{S}^{(i)} = \boldsymbol{S}^{(i)} \mathrm{blockdiag}(\boldsymbol{\Pi}_0, \ldots, \boldsymbol{\Pi}_{Q+L})$.

C. Derivation of the WF-THP Filters

Similar to the TxWF, the WF-THP filters are found by minimizing the MSE between the signals $\boldsymbol{d}[n-\nu]$ and $\hat{\boldsymbol{d}}[n]$ under the transmit power constraint. Additionally, we have to constrain the structure of the feedback filter $\boldsymbol{F}$:

$$\begin{aligned}&\{\boldsymbol{P}_{\mathrm{WF}}, \boldsymbol{F}_{\mathrm{WF}}, \boldsymbol{T}_{\mathrm{WF},1}, \ldots, \boldsymbol{T}_{\mathrm{WF},N_T}, g_{\mathrm{WF}}, \nu_{\mathrm{WF}}, \mathcal{O}_{\mathrm{WF}}\} \\ &\quad = \underset{\{\boldsymbol{P},\boldsymbol{F},\boldsymbol{T}_1,\ldots,\boldsymbol{T}_{N_T},g,\nu,\mathcal{O}\}}{\operatorname{argmin}} \mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu]-\hat{\boldsymbol{d}}[n]\right\|_2^2\right] \\ &\text{s.t.:} \quad \boldsymbol{F} \in \mathbb{L}^{B\times B} \quad \text{and} \quad \mathrm{E}\left[\|\boldsymbol{y}[n]\|_2^2\right] = E_{\mathrm{tr}} \end{aligned} \tag{13}$$

with the set $\mathbb{L}^{B\times B}$ of the $B\times B$ complex lower triangular matrices whose main diagonal is zero. Here, we introduced the latency time $\nu \in \{0,\ldots,Q+L\}$. With (10) and (12), the MSE (i.e., the cost function) can be rewritten as

$$\begin{aligned}&\mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu]-\hat{\boldsymbol{d}}[n]\right\|_2^2\right] \\ &\quad = \sigma_{\boldsymbol{v}}^2 \mathrm{tr}(\boldsymbol{1}_B + \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) + \sigma_{\boldsymbol{v}}^2 \sum_{i=1}^{N_T} \mathrm{tr}\left(\boldsymbol{T}_i \boldsymbol{T}_i^{\mathrm{H}}\right) + |g|^2 \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}}) \\ &\qquad + \sigma_{\boldsymbol{v}}^2 |g|^2 \mathrm{tr}(\boldsymbol{C}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}^{\mathrm{H}}) \\ &\qquad - 2\sigma_{\boldsymbol{v}}^2 \mathrm{Re}\left(\mathrm{tr}\left(g\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}(\boldsymbol{1}_B-\boldsymbol{F})^{\mathrm{H}}\right)\right) \\ &\qquad + 2\sigma_{\boldsymbol{v}}^2 \sum_{i=1}^{N_T} \mathrm{Re}\left(\mathrm{tr}\left(g\boldsymbol{\Pi}_{\nu+i}^{\mathrm{T}}\boldsymbol{S}^{(\nu+i)}\boldsymbol{C}\boldsymbol{P}\boldsymbol{T}_i^{\mathrm{H}}\boldsymbol{\Pi}_{\nu}\right)\right) \end{aligned} \tag{14}$$

where we made use of $\sum_{i=0}^{Q+L} \boldsymbol{S}^{(i),\mathrm{T}}\boldsymbol{S}^{(i)} = \boldsymbol{1}_{B(Q+L+1)}$ and incorporated the fact that $\boldsymbol{v}[n]$ is temporally and spatially uncorrelated, i.e., $\mathrm{E}[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+\nu]] = \sigma_{\boldsymbol{v}}^2 \boldsymbol{1}_B \delta[\nu]$, and that $\boldsymbol{F}$ has a zero main diagonal. By defining the selection matrix

$$\boldsymbol{S}_i = [\boldsymbol{1}_i, \boldsymbol{0}_{i\times B-i}] \in \{0,1\}^{i\times B} \tag{15}$$

which gives the first i rows of a matrix with B rows, when applied from the left, we can reformulate the constraint on $\boldsymbol{F}$ to be an element of $\mathbb{L}^{B\times B}$:

$$\boldsymbol{S}_i \boldsymbol{F} \boldsymbol{e}_i = \boldsymbol{0}_i, \quad i = 1,\ldots,B \tag{16}$$

that is, we constrain the structure of $\boldsymbol{F}$ column-wise by setting the first i elements of the ith column to zero. Since $\boldsymbol{y}[n] = \boldsymbol{P}[n] * \boldsymbol{v}[n]$ (see Fig. 3) and $\sum_{i=0}^{L} \mathrm{tr}(\boldsymbol{P}_i \boldsymbol{P}_i^{\mathrm{H}}) = \mathrm{tr}(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}})$, we get, for the transmit power constraint

$$\mathrm{E}\left[\|\boldsymbol{y}[n]\|_2^2\right] = \sigma_{\boldsymbol{v}}^2 \mathrm{tr}(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}) = E_{\mathrm{tr}}. \tag{17}$$

The solution to the WF-THP optimization (13) can be obtained by incorporating (14), (16), and (17) into the optimization (13) and using the method of Lagrangian multipliers, as shown in Appendix I. The optimum latency time and precoding order follow from

$$\{\nu_{\mathrm{WF}}, \mathcal{O}_{\mathrm{WF}}\} = \underset{\{\nu,\mathcal{O}\}}{\arg\min} \sum_{i=1}^{B} \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{S}^{(\nu)} \boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1} \boldsymbol{S}^{(\nu),\mathrm{T}} \boldsymbol{e}_i \tag{18}$$

which are needed to compute the WF-THP filters:

$$\boldsymbol{P}_{\mathrm{WF}} = \frac{1}{g_{\mathrm{WF}}} \sum_{i=1}^{B} \boldsymbol{C}^{\mathrm{H}} \boldsymbol{A}_{\mathrm{WF},\nu_{\mathrm{WF}},i}^{(\mathcal{O}_{\mathrm{WF}}),-1} \boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}} \tag{19}$$

$$\boldsymbol{F}_{\mathrm{WF}} = -g_{\mathrm{WF}} \sum_{i=1}^{B} \left(\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i\right) \boldsymbol{S}^{(\nu_{\mathrm{WF}})} \boldsymbol{C}\boldsymbol{P}_{\mathrm{WF}} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}} \tag{20}$$

$$\boldsymbol{T}_{\mathrm{WF},i} = -g_{\mathrm{WF}} \boldsymbol{\Pi}_{\nu_{\mathrm{WF}}} \boldsymbol{\Pi}_{\nu_{\mathrm{WF}}+i}^{\mathrm{T}} \boldsymbol{S}^{(\nu_{\mathrm{WF}}+i)} \boldsymbol{C}\boldsymbol{P}_{\mathrm{WF}} \tag{21}$$

for $i = 1,\ldots,Q+L-\nu_{\mathrm{WF}}$ and $\boldsymbol{T}_{\mathrm{WF},i} = \boldsymbol{0}_{B\times B}$ otherwise. The permutation matrix $\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}$ can be obtained by inserting $\mathcal{O}_{\mathrm{WF}}$ into (9). The WF-THP weight at the receivers is

$$g_{\mathrm{WF}} = \sqrt{\frac{\sigma_{\boldsymbol{v}}^2}{E_{\mathrm{tr}}} \sum_{i=1}^{B} \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{S}^{(\nu_{\mathrm{WF}})} \boldsymbol{C}\boldsymbol{C}^{\mathrm{H}} \boldsymbol{A}_{\mathrm{WF},\nu_{\mathrm{WF}},i}^{(\mathcal{O}_{\mathrm{WF}}),-2} \boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}} \boldsymbol{e}_i}. \tag{22}$$

For notational brevity, we have introduced

$$\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O})} = \boldsymbol{S}_{\nu,i}^{\mathrm{T}} \boldsymbol{S}_{\nu,i} \boldsymbol{C}\boldsymbol{C}^{\mathrm{H}} \boldsymbol{S}_{\nu,i}^{\mathrm{T}} \boldsymbol{S}_{\nu,i} + \xi \boldsymbol{1}_{B(Q+L+1)} \tag{23}$$

with the $B\nu+i \times B(Q+L+1)$ selection matrix

$$\boldsymbol{S}_{\nu,i} = \left[\boldsymbol{1}_{B\nu+i}, \boldsymbol{0}_{B\nu+i\times B(Q+L-\nu+1)-i}\right] \tag{24}$$

which gives the first $B\nu+i$ rows of a matrix with $B(Q+L+1)$ rows, when applied from the left.

D. Discussion of the WF-THP Solution

Based on the WF-THP solution in (18)–(22) [see also (28)–(30)], we make following remarks.

1) The feedforward filter $\boldsymbol{P}_{\mathrm{WF}}$ in (19) depends only on the first $B(\nu_{\mathrm{WF}}+1)$ rows of the channel matrix $\boldsymbol{C}$.[6] From (12), we see that these rows of $\boldsymbol{C}$ together with $\boldsymbol{P}_{\mathrm{WF}}$ form the coefficients for $\boldsymbol{v}[n],\ldots,\boldsymbol{v}[n-\nu_{\mathrm{WF}}]$, and hence, $\boldsymbol{d}[n],\ldots,\boldsymbol{d}[n-\nu_{\mathrm{WF}}]$ in the estimate $\hat{\boldsymbol{d}}[n]$. As $\boldsymbol{d}[n-\nu_{\mathrm{WF}}]$ is the desired part, the feedforward filter $\boldsymbol{P}_{\mathrm{WF}}[n]$ combats the signal portions due to the preceding symbols, viz. $\boldsymbol{d}[n],\ldots,\boldsymbol{d}[n-\nu_{\mathrm{WF}}+1]$. Similarly, $\boldsymbol{P}_{\mathrm{WF}}[n]$ combats the interference caused by $d_{b_{i+1}}[n-\nu_{\mathrm{WF}}],\ldots,d_{b_B}[n-\nu_{\mathrm{WF}}]$ in the estimate $\hat{d}_{b_i}[n]$ of $d_{b_i}[n-\nu_{\mathrm{WF}}]$, as the ith column of $\boldsymbol{P}_{\mathrm{WF}}$ only depends on the first $B\nu+i$ rows of $\boldsymbol{C}$.
2) Dividing the matrix $\boldsymbol{C}^{\mathrm{H}}$ into $\mathbb{C}^{N_{\mathrm{a}}\times B}$ blocks, we see that the $L+1-i$ lower block elements of the ith block column for $i = 1,\ldots,L$ are zero. As a consequence, the coefficients $\boldsymbol{P}_{\mathrm{WF},\nu+1},\ldots,\boldsymbol{P}_{\mathrm{WF},L}$ of $\boldsymbol{P}_{\mathrm{WF}}[n]$ resulting from the feedforward filter solution (19) for a latency time $\nu < L$ are zero. This corresponds to a reduction of the feedforward filter order L to $\nu < L$. It is intuitively clear that the properties of a feedforward filter are deteriorated by decreasing the filter order, i.e., the MSE is increased. In other words, the MSE for $\nu = L$ is always smaller than or equal to the MSE for any $\nu < L$. Hence, the latency time must always be larger than or equal to the filter order, that is, $\nu_{\mathrm{WF}} \in \{L,\ldots,Q+L\}$. For a strict proof of this statement, see [65].

[6] Multiplying with $\boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}}\boldsymbol{e}_i$ from the right picks out the $B\nu_{\mathrm{WF}}+i$th column. Only the first $B\nu_{\mathrm{WF}}+i$ elements of the $B\nu_{\mathrm{WF}}+i$th column of $\boldsymbol{A}_{\mathrm{WF},\nu_{\mathrm{WF}},i}^{(\mathcal{O}_{\mathrm{WF}}),-1}$ are nonzero for $i = 1,\ldots,B$.

3) The product $g_{\mathrm{WF}}\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}+i}\boldsymbol{S}^{(\nu_{\mathrm{WF}}+i)}\boldsymbol{C}\boldsymbol{P}_{\mathrm{WF}}$ in (21) multiplied with $\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}^{\mathrm{T}}$ gives the $\nu_{\mathrm{WF}}+i$th coefficient of $g_{\mathrm{WF}}\boldsymbol{H}[n]*\boldsymbol{P}_{\mathrm{WF}}[n]$, i.e., the weight for the signal portion in the estimate $\hat{\boldsymbol{d}}[n]$ due to the symbol $\boldsymbol{d}[n-\nu_{\mathrm{WF}}-i]$. Thus, the temporal feedback filter $\boldsymbol{T}_{\mathrm{WF}}[n]$ subtracts the interference caused by the succeeding symbols.
4) The spatial feedback filter $\boldsymbol{F}_{\mathrm{WF}}$ in (20) is constructed from the ν_{WF}th coefficient of $g_{\mathrm{WF}}\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}\boldsymbol{H}[n]*\boldsymbol{P}_{\mathrm{WF}}[n]$, where the ith column of this coefficient is projected by $\mathbf{1}_B-\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i$, which sets the first i elements to zero. The resulting lower triangular $\boldsymbol{F}_{\mathrm{WF}}$ combats the interference caused by $d_{b_1}[n-\nu_{\mathrm{WF}}],\ldots,d_{b_{i-1}}[n-\nu_{\mathrm{WF}}]$ in the estimate $\hat{d}_{b_i}[n]$ for $d_{b_i}[n-\nu_{\mathrm{WF}}]$.
5) With decreasing SNR, the ratio ξ [see (7)] increases. Therefore, the identity matrix in the definition of $\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O})}$ in (23) becomes dominant for low SNR. We can conclude that $\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\rightarrow\xi^{-1}\mathbf{1}_{B(Q+L+1)}$, and hence, $\boldsymbol{F}_{\mathrm{WF}},\boldsymbol{T}_{\mathrm{WF},1},\ldots,\boldsymbol{T}_{\mathrm{WF},Q+L-\nu_{\mathrm{WF}}}\rightarrow\mathbf{0}_{B\times B}$ for very low SNR, i.e., the feedback is switched off. Without feedback, however, the statistical assumptions for the signal $\boldsymbol{v}[n]$ are no longer valid because $\boldsymbol{v}[n]=\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}\boldsymbol{s}[n]$. As the variance $\sigma_{\boldsymbol{v}}^2=\tau^2/6$ assumed for the design is usually larger than the variance of the data symbols $\boldsymbol{s}[n]$, the wrong assumption for the covariance matrix of $\boldsymbol{v}[n]$ leads to a feedforward filter $\boldsymbol{P}_{\mathrm{WF}}$, which does not use all of the available transmit power E_{tr} for low SNR.
6) The first step to compute the WF-THP solution, namely, the latency time and precoding order optimization (18), is the most expensive step. For all combinations of the integer $\nu\in\{L,\ldots,Q+L\}$ and the B-tuple $\mathcal{O}$, we have to invert the $B(Q+L+1)\times B(Q+L+1)$ matrices $\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O})}$, $i=1,\ldots,B$. Thus, the complexity of this step is $\mathrm{O}(B!B^3Q(Q+L)^3)$. In the next subsection, we show how this complexity can be reduced dramatically.

E. Successive Precoding Order Computation

To avoid the high complexity of $\mathrm{O}(B!B^3Q(Q+L)^3)$ for the optimum latency time and precoding order computation (18), the standard suboptimum approach is the successive computation of the precoding order $\mathcal{O}$ (like V-BLAST for DFE, e.g., [66]). To this end, let us rewrite the cost of (18) by employing the identity $\boldsymbol{\Pi}_j^{\mathrm{T}}\boldsymbol{S}^{(j)}=\boldsymbol{S}^{(j)}\mathrm{blockdiag}(\boldsymbol{\Pi}_0^{\mathrm{T}},\ldots,\boldsymbol{\Pi}_{Q+L}^{\mathrm{T}})$:

$$\begin{aligned}&\sum_{i=1}^{B}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\\&\quad=\sum_{i=1}^{B}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{\Pi}_\nu\boldsymbol{\Pi}_\nu^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{\Pi}_\nu\boldsymbol{\Pi}_\nu^{\mathrm{T}}\boldsymbol{e}_i\\&\quad=\sum_{i=1}^{B}\boldsymbol{e}_{b_i}^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{\Gamma}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_{b_i}\end{aligned}\tag{25}$$

with the $B(Q+L+1)\times B(Q+L+1)$ matrix [see also (2), (9), and (11)]

$$\boldsymbol{\Gamma}_{\mathrm{WF},\nu,i}^{(\mathcal{O})}=\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})}+\xi\mathbf{1}_{B(Q+L+1)}.$$

The projector $\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})}\in\{0,1\}^{B(Q+L+1)\times B(Q+L+1)}$ is defined as [see (38)]

$$\begin{aligned}\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})}&=\mathrm{blockdiag}\left(\boldsymbol{\Pi}_0^{\mathrm{T}},\ldots\right)\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\mathrm{blockdiag}(\boldsymbol{\Pi}_0,\ldots)\\&=\sum_{j=0}^{\nu}\boldsymbol{S}^{(j),\mathrm{T}}\boldsymbol{S}^{(j)}-\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{\Pi}_\nu^{\mathrm{T}}\left(\mathbf{1}_B-\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i\right)\boldsymbol{\Pi}_\nu\boldsymbol{S}^{(\nu)}\\&=\sum_{j=0}^{\nu}\boldsymbol{S}^{(j),\mathrm{T}}\boldsymbol{S}^{(j)}-\begin{cases}\mathbf{0}_{B(Q+L+1)\times B(Q+L+1)}, & i=B\\ \boldsymbol{S}^{(\nu),\mathrm{T}}\sum_{j=i+1}^{B}\boldsymbol{e}_{b_j}\boldsymbol{e}_{b_j}^{\mathrm{T}}\boldsymbol{S}^{(\nu)}, & \text{else.}\end{cases}\end{aligned}$$

We observe that $\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})}$ only depends on the indices $b_{i+1},\ldots,b_B$ of the precoding order $\mathcal{O}$. Thus, the ith summand of the cost (25) only depends on the indices $b_i,\ldots,b_B$. This property of (25) motivates us to compute the indices successively: First, compute b_B by minimizing the Bth summand ($i=B$) of (25). Second, compute the ith index b_i of the precoding order $\mathcal{O}$ by minimizing the ith summand of (25) for fixed $b_{i+1},\ldots,b_B$:

$$\begin{aligned}\mathcal{O}'_{\mathrm{WF}}&=(b_1,\ldots,b_B)\\b_i&=\operatorname*{argmin}_{b\in\mathbb{O}_i}\boldsymbol{e}_b^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{\Gamma}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_b\\&\qquad i=B,\ldots,1\end{aligned}\tag{26}$$

where $\mathbb{O}_i=\{1,\ldots,B\}\setminus\{b_{i+1}\ldots,b_B\}$ is the set of allowed values for b_i. Note that $\mathbb{O}_B=\{1,\ldots,B\}$.

The precoding order must be optimized with (26) for every possible $\nu\in\{L,\ldots,Q+L\}$ to find the latency time ν'_{WF} minimizing the sum (25). Thus, the complexity of the latency time and precoding order optimization is $\mathrm{O}(B^4Q(Q+L)^3)$.

F. Zero-Forcing Tomlinson–Harashima Precoding (ZF-THP)

Similar to the respective linear precoders (see Section II), the ZF-THP optimization can be obtained from the WF-THP optimization (13) by including the zero-forcing constraint $\mathrm{E}[\hat{\boldsymbol{d}}[n]|\boldsymbol{d}[n],\boldsymbol{d}[n-1],\ldots]=\boldsymbol{d}[n-\nu]$. The steps to solve the ZF-THP optimization are similar to the ones used in Appendix I. By replacing $\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O})}$ with

$$\boldsymbol{A}_{\mathrm{ZF},\nu,i}^{(\mathcal{O})}=\begin{bmatrix}\boldsymbol{S}_{\nu,i}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{X}\end{bmatrix}$$

we get the ZF-THP solution from the WF-THP solution in (18)–(22), where $\boldsymbol{X}^{B(Q+L-\nu+1)-i\times B(Q+L-\nu+1)-i}$ can be any invertible matrix, e.g., $\boldsymbol{X}=\mathbf{1}_{B(Q+L-\nu+1)-i}$.

IV. Efficient Precoding Order and Latency Time Computation

As highlighted in the previous section, the first step in finding the WF-THP filters is the optimization of the latency time and the precoding order. For fixed latency time, the precoding order can be found via (26). Consequently, the optimum latency time can be obtained by performing (26) for all possibly optimum values $L,\ldots,Q+L$ for the latency time.

We will show how the order of complexity in solving the precoding order and latency time optimization can be reduced by employing a *Cholesky factorization with symmetrical permutation*. First, the precoding order optimization (26) is reformulated

to be able to incorporate it into the computation of the Cholesky factorization. Second, we develop an algorithm where the latency time optimization is also a byproduct of the Cholesky factorization. Interestingly, the resulting order of complexity for WF-THP is the same as for the linear TxWF.

A. Optimized Latency Time

We start with making the assumption that the following Cholesky factorization (e.g., [56]) is known:

$$\begin{aligned}\left(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}+\xi\mathbf{1}_{B(Q+L+1)}\right)^{-1}&\\ =\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{L}&\\ =\boldsymbol{\Pi}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}+\xi\mathbf{1}_{B(Q+L+1)}\right)^{-1}\boldsymbol{\Pi}^{\mathrm{T}}&\end{aligned}\tag{27}$$

where $\boldsymbol{L}\in\mathbb{C}^{B(Q+L+1)\times B(Q+L+1)}$ is unit lower triangular,[7] $\boldsymbol{D}\in\mathbb{R}^{B(Q+L+1)\times B(Q+L+1)}$ is a non-negative real diagonal matrix, and the $B(Q+L+1)\times B(Q+L+1)$ permutation matrix

$$\boldsymbol{\Pi}=\mathrm{blockdiag}(\boldsymbol{\Pi}_0,\ldots,\boldsymbol{\Pi}_{Q+L})$$

is block diagonal [cf. (11)]. From the last line of (27), we see that $\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{L}$ is the Cholesky factorization with symmetric permutation of $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}+\xi\mathbf{1}_{B(Q+L+1)})^{-1}$ [56, p. 147].

The Cholesky factorization (27) can be used to rewrite the WF-THP filter expressions in (19)–(21):

$$\boldsymbol{P}_{\mathrm{WF}}=\frac{1}{g_{\mathrm{WF}}}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}}\tag{28}$$

$$\boldsymbol{F}_{\mathrm{WF}}=\mathbf{1}_B-\boldsymbol{S}^{(\nu_{\mathrm{WF}})}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}}\tag{29}$$

$$\boldsymbol{T}_{\mathrm{WF},i}=-\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}+i}^{\mathrm{T}}\boldsymbol{S}^{(\nu_{\mathrm{WF}}+i)}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}}\tag{30}$$

with $i=1,\ldots,Q+L-\nu_{\mathrm{WF}}$. The alternative expression for g_{WF} follows from (28) and the transmit power constraint (17). For a detailed derivation of above expressions, see Appendix II.

We see that $\boldsymbol{P}_{\mathrm{WF}}$ only depends on the $\nu_{\mathrm{WF}}+1$th $B\times B$ block row of $\boldsymbol{L}$ and $\boldsymbol{F}_{\mathrm{WF}}$ on the $\nu_{\mathrm{WF}}+1$th $B\times B$ block diagonal element of the lower triangular $\boldsymbol{L}$, since (29) is equivalent to $\boldsymbol{F}_{\mathrm{WF}}=\mathbf{1}_B-(\boldsymbol{S}^{(\nu_{\mathrm{WF}})}\boldsymbol{L}\boldsymbol{S}^{(\nu_{\mathrm{WF}}),\mathrm{T}})^{-1}$. Due to the inversion of $\boldsymbol{L}$ in (30), we could conclude that the complete Cholesky factorization (27) is necessary to compute the WF-THP filters. However, the coefficients of the temporal feedback filter can also be computed via (21). Consequently, we do not need the complete Cholesky factorization (27) to compute the WF-THP filters. Instead, the $B\nu_{\mathrm{WF}}+1$th up to the $B(\nu_{\mathrm{WF}}+1)$th row of $\boldsymbol{L}$ and the $\nu_{\mathrm{WF}}+1$th $B\times B$ block diagonal of $\boldsymbol{D}$ are sufficient to find the filters with (21), (28), and (29).

More importantly, when inserting (27) into the cost function of the precoding order optimization (26), we obtain a simple rule of how the Cholesky factorization (27) has to be computed to obtain an optimal $\boldsymbol{\Pi}_{\nu_{\mathrm{WF}}}$, i.e., a precoding order $\mathcal{O}'_{\mathrm{WF}}$ optimal with respect to (26). Let $f_i(b)=\boldsymbol{e}_b^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{\Gamma}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_b$ denote the cost of (26). Due to (25) and (44), we get

$$f_i(b)=\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i=\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{B}_{\nu,i}^{(\mathcal{O})}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i.$$

[7] That is, $\boldsymbol{L}$ is lower triangular with unit main diagonal [56, p. 91].

Note that $\boldsymbol{B}_{\nu,i}^{(\mathcal{O})}$ depends on the index b. Inserting (42) and (27) into above cost $f_i(b)$ of (26) yields

$$\begin{aligned}f_i(b)&=\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\left(\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1}\boldsymbol{D}^{-1}\boldsymbol{L}^{\mathrm{H},-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1}\boldsymbol{S}_{\nu,i}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\\ &=\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{D}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\end{aligned}\tag{31}$$

where we employed (46) and (47) for the second line. We observe that the cost to be minimized for finding b_i, $i=B,\ldots,1$ is the $B\nu+i$th diagonal entry of $\boldsymbol{D}$. Therefore, when the $\nu+1$th diagonal block $\boldsymbol{D}_{\nu+1}\in\mathbb{R}^{B\times B}$ of $\boldsymbol{D}$ is computed according to (27), the following procedure must be used: For fixed permutation of the last $B-i$ rows/columns of the $\nu+1$th block row/column of $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}+\xi\mathbf{1}_{B(Q+L+1)})^{-1}$ permute the first i rows/columns of the $\nu+1$th block row/column such that the resulting ith diagonal entry of $\boldsymbol{D}_{\nu+1}$ is minimized. Fortunately, this minimization can easily be included in the Cholesky factorization algorithm.

The result for the cost in (31) is also very important for an efficient implementation of the latency time optimization. Summing (31) up for $i=1,\ldots,B$ gives the cost to be minimized by the latency time (cf. (25)):

$$\nu'_{\mathrm{WF}}=\underset{\nu\in\{L,\ldots,Q+L\}}{\mathrm{argmin}}\ \mathrm{tr}\left(\boldsymbol{S}^{(\nu)}\boldsymbol{D}\boldsymbol{S}^{(\nu),\mathrm{T}}\right).\tag{32}$$

Interestingly, the cost for some latency time ν only depends on the $\nu+1$th $B\times B$ diagonal block of $\boldsymbol{D}$. Thus, above optimization for the latency time can be incorporated into the Cholesky factorization (27) as follows: First, compute the ith block diagonal entry $\boldsymbol{D}_i\in\mathbb{R}^{B\times B}$, $i=Q+L+1,\ldots,L+1$, of $\boldsymbol{D}$ such that the resulting permutation matrix $\boldsymbol{\Pi}_{i-1}$ represents the optimal precoding order for $\nu=i-1$. Second, choose the optimum latency time ν'_{WF} according to (32).

The resulting algorithm to compute the WF-THP filters with optimum precoding order $\mathcal{O}'_{\mathrm{WF}}$ and latency time ν'_{WF} can be found in Table I. In order to find the latency time with minimum MSE, we initialize the minimum cost $\varepsilon_{\min}$ found so far with infinity in line 3. The precoding order optimization for the latency time ν is performed in line 9. Lines 13–15 are the core of the plain Cholesky factorization. In line 17, the latency time optimization can be found. Since the WF-THP filters only depend on parts of the Cholesky factorization, the factorization is not computed completely (see line 4). Therefore, the first ν'_{WF} diagonal blocks of the projector in line 21 are chosen to be identity matrices.

The incomplete Cholesky factorization performed by the algorithm in Table I has less complexity than the most costly line 1: the inversion of the positive definite Hermitian matrix $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}+\xi\mathbf{1}_{B(Q+L+1)}$. This inverse is also needed for the linear TxWF filter [cf. (6) and (7)]. We can conclude that the order of complexity to compute the WF-THP filters is $\mathrm{O}(B^3(Q+L)^3)$, i.e., it is the same as for the linear TxWF.

B. Fixed Latency Time

As will be demonstrated by simulation in the next section (see Fig. 6), the latency time can be set to a fixed value ν_{fix} for most scenarios without performance loss. For a fixed latency

TABLE I
WF-THP FILTER COMPUTATION WITH OPTIMIZED PRECODING ORDER AND LATENCY TIME

$\boldsymbol{\Theta} \leftarrow (\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \xi \mathbf{1}_{B(Q+L+1)})^{-1}$
$\boldsymbol{D} \leftarrow \mathbf{0}_{B(Q+L+1)\times B(Q+L+1)}$
$\varepsilon_{\min} \leftarrow \infty$
for $\nu = Q+L, \ldots, L$:
  $\boldsymbol{\Pi}_{\nu} \leftarrow \mathbf{1}_B$
  $\varepsilon \leftarrow 0$
  for $i = B, \ldots, 1$:
    $k \leftarrow B\nu + i$
    $b_i \leftarrow \underset{b\in\{1,\ldots,i\}}{\operatorname{argmin}} \boldsymbol{\Theta}(B\nu+b, B\nu+b)$
    $\boldsymbol{\Pi}_{\nu} \leftarrow \boldsymbol{\Pi}_{\nu}$ with rows b_i and i exchanged
    $\boldsymbol{\Pi} \leftarrow \mathbf{1}_{B(Q+L+1)}$ with rows $B\nu+b_i$ and k exch.
    $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Pi}\boldsymbol{\Theta}\boldsymbol{\Pi}^{\mathrm{T}}$
    $\boldsymbol{D}(k,k) \leftarrow \boldsymbol{\Theta}(k,k)$
    $\boldsymbol{\Theta}(k, 1:k) \leftarrow \boldsymbol{\Theta}(k, 1:k)/\boldsymbol{D}(k,k)$
    $\boldsymbol{\Theta}(1:k-1, 1:k-1) \leftarrow \boldsymbol{\Theta}(1:k-1, 1:k-1)$
      $-\boldsymbol{D}(k,k)(\boldsymbol{\Theta}(k, 1:k-1))^{\mathrm{H}}\boldsymbol{\Theta}(k, 1:k-1)$
    $\varepsilon \leftarrow \varepsilon + \boldsymbol{D}(k,k)$
  if $\varepsilon < \varepsilon_{\min}$:
    $\nu'_{\mathrm{WF}} \leftarrow \nu$
    $\varepsilon_{\min} \leftarrow \varepsilon$
$\boldsymbol{L} \leftarrow$ lower triangular part of $\boldsymbol{\Theta}$
$\boldsymbol{C} \leftarrow \mathrm{blockdiag}(\mathbf{1}_B, \ldots, \mathbf{1}_B, \boldsymbol{\Pi}_{\nu'_{\mathrm{WF}}}, \ldots, \boldsymbol{\Pi}_{Q+L})\boldsymbol{H}$
$\boldsymbol{P} \leftarrow \boldsymbol{C}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{S}^{(\nu'_{\mathrm{WF}}),\mathrm{T}}$
$g_{\mathrm{WF}} \leftarrow \sqrt{\sigma_{\boldsymbol{v}}^2 \|\boldsymbol{P}\|_{\mathrm{F}}^2 / E_{\mathrm{tr}}}$
$\boldsymbol{P}_{\mathrm{WF}} \leftarrow \boldsymbol{P}/g_{\mathrm{WF}}$
$\boldsymbol{F}_{\mathrm{WF}} \leftarrow \mathbf{1}_B - (\boldsymbol{S}^{(\nu'_{\mathrm{WF}})}\boldsymbol{L}\boldsymbol{S}^{(\nu'_{\mathrm{WF}}),\mathrm{T}})^{-1}$
for $i = 1, \ldots, Q+L-\nu'_{\mathrm{WF}}$:
  $\boldsymbol{T}_i \leftarrow -\boldsymbol{\Pi}_{\nu'_{\mathrm{WF}}}\boldsymbol{\Pi}^{\mathrm{T}}_{\nu'_{\mathrm{WF}}+i}\boldsymbol{S}^{(\nu'_{\mathrm{WF}}+i)}\boldsymbol{C}\boldsymbol{P}$

time ν_{fix}, the computation of the WF-THP filters can be further simplified by employing the following Cholesky factorization with symmetric permutation:

$$\boldsymbol{\Pi}_{\mathrm{fix}}\left(\boldsymbol{S}_{\nu_{\mathrm{fix}},B}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \xi\mathbf{1}_{B(Q+L+1)}\right)\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B}\right)^{-1} \times \boldsymbol{\Pi}^{\mathrm{T}}_{\mathrm{fix}} = \boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda} \quad (33)$$

where $\boldsymbol{\Lambda} \in \mathbb{C}^{B(\nu_{\mathrm{fix}}+1)\times B(\nu_{\mathrm{fix}}+1)}$ is unit lower triangular, and $\boldsymbol{\Delta} \in \mathbb{R}^{B(\nu_{\mathrm{fix}}+1)\times B(\nu_{\mathrm{fix}}+1)}$ is a diagonal matrix. The projector $\boldsymbol{\Pi}_{\mathrm{fix}}$ is defined as

$$\boldsymbol{\Pi}_{\mathrm{fix}} = \mathrm{blockdiag}\left(\mathbf{1}_B, \ldots, \mathbf{1}_B, \boldsymbol{\Pi}_{\nu_{\mathrm{fix}}}\right)$$

that is we simply set $\boldsymbol{\Pi}_i = \mathbf{1}_B$ for $i = 0, \ldots, Q+L$ and $i \neq \nu_{\mathrm{fix}}$. By incorporating (33) into (19) and (20), and with steps similar to the ones found in Appendix II, we find, for the WF-THP feedforward and spatial feedback filter

$$\boldsymbol{P}_{\mathrm{WF}} = \frac{1}{g_{\mathrm{WF}}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B}\boldsymbol{\Pi}^{\mathrm{T}}_{\mathrm{fix}}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Delta}\boldsymbol{S}_{\nu_{\mathrm{fix}},B}\boldsymbol{S}^{(\nu_{\mathrm{fix}}),\mathrm{T}} \quad (34)$$

$$\boldsymbol{F}_{\mathrm{WF}} = \mathbf{1}_B - \left(\boldsymbol{S}^{(\nu_{\mathrm{fix}})}\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B}\boldsymbol{\Lambda}\boldsymbol{S}_{\nu_{\mathrm{fix}},B}\boldsymbol{S}^{(\nu_{\mathrm{fix}}),\mathrm{T}}\right)^{-1} \quad (35)$$

respectively. Note that only the last B rows of $\boldsymbol{\Lambda}$ and the last $B\times B$ diagonal block of $\boldsymbol{\Delta}$ are needed to compute the WF-THP filters with the above expressions, since the coefficients of the temporal feedback filter can be found with (21). Therefore, we do not have to perform the complete Cholesky factorization

TABLE II
WF-THP FILTER COMPUTATION WITH OPTIMIZED PRECODING ORDER AND FIXED LATENCY TIME

$\boldsymbol{\Theta} \leftarrow (\boldsymbol{S}_{\nu_{\mathrm{fix}},B}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \xi\mathbf{1}_{B(Q+L+1)})\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B})^{-1}$
$\boldsymbol{\Delta} \leftarrow \mathbf{0}_{B(\nu_{\mathrm{fix}}+1)\times B(\nu_{\mathrm{fix}}+1)}$
$\boldsymbol{\Pi}_{\nu_{\mathrm{fix}}} \leftarrow \mathbf{1}_B$
for $i = B, \ldots, 1$:
  $k \leftarrow B\nu_{\mathrm{fix}} + i$
  $b_i \leftarrow \underset{b\in\{1,\ldots,i\}}{\operatorname{argmin}} \boldsymbol{\Theta}(B\nu_{\mathrm{fix}}+b, B\nu_{\mathrm{fix}}+b)$
  $\boldsymbol{\Pi}_{\nu_{\mathrm{fix}}} \leftarrow \boldsymbol{\Pi}_{\nu_{\mathrm{fix}}}$ with rows b_i and i exchanged
  $\boldsymbol{\Pi} \leftarrow \mathbf{1}_{B(\nu_{\mathrm{fix}}+1)}$ with rows $B\nu_{\mathrm{fix}}+b_i$ and k exchanged
  $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Pi}\boldsymbol{\Theta}\boldsymbol{\Pi}^{\mathrm{T}}$
  $\boldsymbol{\Delta}(k,k) \leftarrow \boldsymbol{\Theta}(k,k)$
  $\boldsymbol{\Theta}(k, 1:k) \leftarrow \boldsymbol{\Theta}(k, 1:k)/\boldsymbol{\Delta}(k,k)$
  $\boldsymbol{\Theta}(1:k-1, 1:k-1) \leftarrow \boldsymbol{\Theta}(1:k-1, 1:k-1)$
    $-\boldsymbol{\Delta}(k,k)(\boldsymbol{\Theta}(k, 1:k-1))^{\mathrm{H}}\boldsymbol{\Theta}(k, 1:k-1)$
$\boldsymbol{\Lambda} \leftarrow$ lower triangular part of $\boldsymbol{\Theta}$
$\boldsymbol{C} \leftarrow \mathrm{blockdiag}(\mathbf{1}_B, \ldots, \mathbf{1}_B, \boldsymbol{\Pi}_{\nu_{\mathrm{fix}}}, \mathbf{1}_B, \ldots, \mathbf{1}_B)\boldsymbol{H}$
$\boldsymbol{P} \leftarrow \boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Delta}\boldsymbol{S}_{\nu_{\mathrm{fix}},B}\boldsymbol{S}^{(\nu_{\mathrm{fix}}),\mathrm{T}}$
$g_{\mathrm{WF}} \leftarrow \sqrt{\sigma_{\boldsymbol{v}}^2 \|\boldsymbol{P}\|_{\mathrm{F}}^2 / E_{\mathrm{tr}}}$
$\boldsymbol{P}_{\mathrm{WF}} \leftarrow \boldsymbol{P}/g_{\mathrm{WF}}$
$\boldsymbol{F}_{\mathrm{WF}} \leftarrow \mathbf{1}_B - (\boldsymbol{S}^{(\nu_{\mathrm{fix}})}\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B}\boldsymbol{\Lambda}\boldsymbol{S}_{\nu_{\mathrm{fix}},B}\boldsymbol{S}^{(\nu_{\mathrm{fix}}),\mathrm{T}})^{-1}$
for $i = 1, \ldots, Q+L-\nu_{\mathrm{fix}}$:
  $\boldsymbol{T}_{\mathrm{WF},i} \leftarrow -\boldsymbol{\Pi}_{\nu_{\mathrm{fix}}}\boldsymbol{S}^{(\nu_{\mathrm{fix}}+i)}\boldsymbol{H}\boldsymbol{P}$

(33). Inserting (33) into the cost of the precoding order optimization (26) yields

$$\boldsymbol{e}_b^{\mathrm{T}}\boldsymbol{S}^{(\nu_{\mathrm{fix}})}\boldsymbol{\Gamma}^{(\mathcal{O}),-1}_{\mathrm{WF},\nu_{\mathrm{fix}},i}\boldsymbol{S}^{(\nu_{\mathrm{fix}}),\mathrm{T}}\boldsymbol{e}_b = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu_{\mathrm{fix}})}\boldsymbol{S}^{\mathrm{T}}_{\nu_{\mathrm{fix}},B}\boldsymbol{\Delta}\boldsymbol{S}_{\nu_{\mathrm{fix}},B}\boldsymbol{S}^{(\nu_{\mathrm{fix}}),\mathrm{T}}\boldsymbol{e}_i. \quad (36)$$

From this result, we see that the permutation of the Cholesky factorization (33) has to minimize the diagonal entries of the lower right $B \times B$ diagonal block of $\boldsymbol{\Delta}$.

The algorithm to compute the WF-THP filters with optimized precoding order but fixed latency time can be found in Table II. Again, the matrix inversion in line 1 is the most complex operation, and the order of complexity is $\mathrm{O}(B^3\nu_{\mathrm{fix}}^3)$.

V. SIMULATION RESULTS

We begin by investigating the performance of THP for frequency flat channels ($Q = 0$), as in this case, the optimum precoders for decentralized modulo receivers are known. As noted in Section III-A, the combination of the feedback loop and the modulo operator at the transmitter can be interpreted as a suboptimum procedure for choosing the perturbation vectors $\boldsymbol{a}[n]$, which would be optimally found through an extremely costly closest point search in a lattice for each precoded symbol. This is realized by the vector precoders [30], [31], which require exponential complexity in the number of users B. We set $\nu_{\mathrm{fix}} = L = 0$, since only spatial equalization is necessary. In Fig. 4, the performance of THP is compared to that of the WF and ZF vector precoders (WF-VP and ZF-VP) and the linear precoders (lin. WF and lin. ZF) for an i i.d. channel model and 16QAM symbols, where we also assumed perfect *channel state information* at the transmitter. First of all, it can be seen that nonlinear precoding is far superior to linear precoding. Also, the WF solutions clearly outperform the respective ZF solutions. Remarkably, the performance of WF-THP is close to that of the optimum WF-VP. For a *bit error rate* (BER) of 10^{-2},

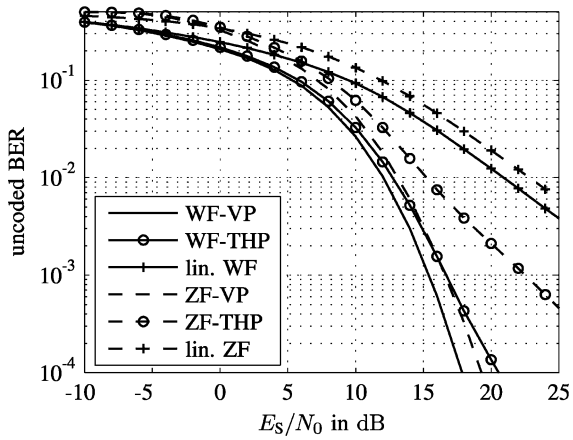

Fig. 4. Tomlinson–Harashima precoding in comparison with vector precoding and linear precoding, i.i.d. Frequency flat channel, $B = 4$ users, $N_a = 4$ transmit antennas, 16QAM.

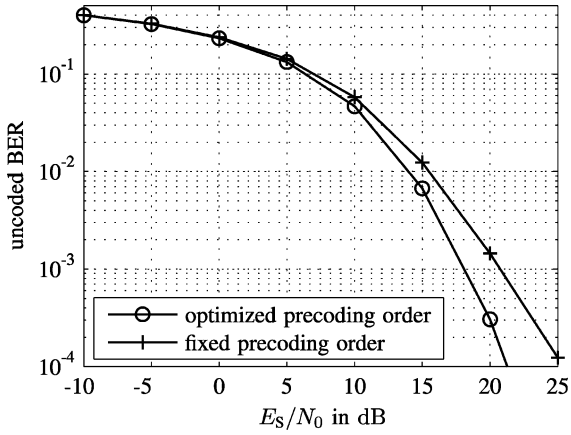

Fig. 5. Effect of precoding order optimization, "Pedestrian A" power delay profile ($Q = 3$), 10° Laplacian angular spread, $B = 3$ users, $N_a = 4$ transmit antennas, filter order $L = 4$, 16QAM.

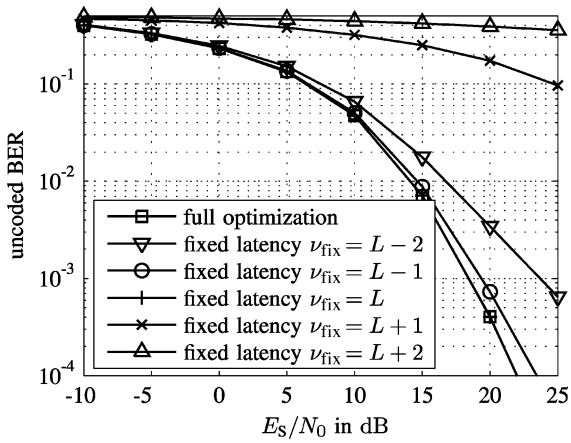

Fig. 6. Effect of fixed latency time, "Pedestrian A" power delay profile ($Q = 3$), 10° Laplacian angular spread, $B = 3$ users, $N_a = 4$ transmit antennas, filter order $L = 2$, 16QAM.

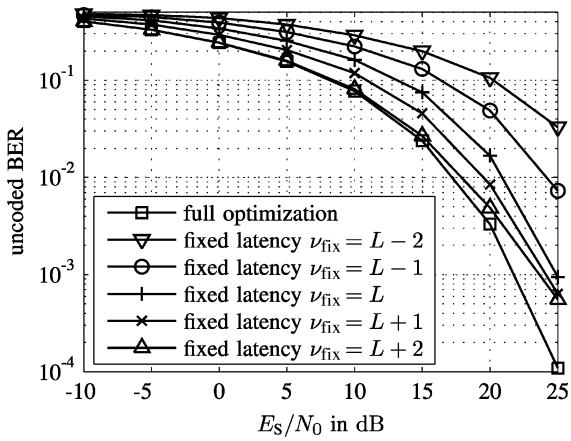

Fig. 7. Effect of fixed latency time, exponentially increasing power delay profile (3 dB per tap, $Q = 5$), 10° Laplacian angular spread, $B = 3$ users, $N_a = 4$ transmit antennas, filter order $L = 3$, 16QAM.

the difference between WF-THP and WF-VP is below 0.7 dB, and WF-THP even outperforms ZF-VP at a complexity comparable to that of the linear filters. For higher *signal-to-noise ratios* (SNRs), however, the slope of the WF-THP graph decreases noticably. This can be explained by the fact that with spatial THP ($L = 0$), the symbol of the first user in the precoding order is always precoded linearly, as it is not superimposed by any feedback and, therefore, is always left unaltered by the modulo operator. For high SNRs, the performance of this first user dominates the BER, leading to the observed decrease in the slope of the THP graphs.

Next, we proceed to frequency-selective channels and investigate the gain that can be achieved by optimizing the precoding order with WF-THP, which has been done for frequency flat channels in [52] and [53] and for frequency selective channels with ZF-THP in [47]. For the results in Fig. 5, we employed the standardization "Pedestrian A" power delay profile [67], which is strongly decreasing. The transmit antennas were assumed to be arranged in a uniform linear array with $\lambda/2$ spacing, and the receivers were located at random angles around the transmitter with 10° Laplacian angular spread. We assumed no temporal correlations. Here, the gain is considerable; at a BER of 10^{-2}, we can measure about 1.5 dB. For other power delay profiles, such as the more spread-out "Vehicular A" scenario, the gain turns out to be lower. In general, the more average energy the first path of the channel has, the more important the spatial feedback component becomes, which benefits from the optimized precoding order.

As was noted in the previous section, setting the latency time to a fixed value can be done without performance degradation in most scenarios, which is demonstrated in Fig. 6, where the performance of different fixed latency times (see Table II) is compared to that of full optimization (Table I). Further simulations with different system parameters and channel models indicate that as long as the power delay profile of the channel model is decaying, constant, or even "U-shaped," fixed latency time $\nu_{fix} = L$ performs nearly exactly as well as full optimization, regardless of all other system parameters. Even for increasing power delay profiles, the filter order L can always be chosen large enough so that latency optimization can be omitted without penalty by setting $\nu_{fix} = L$. Degradation was only observed for "short" feedforward filters in combination with exponentially increasing power delay profiles. One such scenario is shown in Fig. 7. Note that increasing power delay profiles do not usually occur in wireless communications.

VI. Conclusion

We derived closed-form expressions for FIR THP filters based on the MMSE criterion. Inserting the Cholesky factorization with symmetric permutation of the regularized channel Gram into the resulting filter expressions enabled us to find an algorithm for the WF-THP filter computation with optimum precoding order and latency time, where the complexity is comparable to linear precoding filters. We also presented a second algorithm for fixed latency time with further reduced complexity. The simulation results showed that WF-THP leads to near-optimum results and that the optimization of the precoding order is crucial for the performance of WF-THP. We also observed that the latency time optimization can be omitted in realistic scenarios and that the latency time can be chosen to be simply the feedforward filter order.

Appendix I
Solution of the WF-THP Optimization

Following the WF-THP optimization (13), the MSE has to be minimized with respect to the feedforward filter $\boldsymbol{P}$, the spatial feedback filter $\boldsymbol{F}$, the temporal feedback filter coefficients $\boldsymbol{T}_i$, $i = 1, \ldots, N_T$, the receivers' weight g, the latency time ν, and the precoding order $\mathcal{O}$. Contrary to the other quantities, the latency time $\nu \in \{0, \ldots, Q+L\}$ and the precoding order $\mathcal{O}$ can only have discrete values. Thus, we have to use following optimum strategy.

1) Find expressions for the WF-THP filters depending on the latency time and the precoding order. As the WF-THP filters have to fulfill the constraints of (13), we employ the Lagrangian function for this step.
2) Plug the solutions for the WF-THP filters into the cost function of (13). Try all possible values for the latency time ν and the precoding order $\mathcal{O}$, and choose the ones with minimum cost.

To find the WF-THP filters depending on the latency time and the precoding order, we form the Lagrangian function for (13) by employing (16) and (17)

$$L(\boldsymbol{P}, \ldots) = \mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu] - \hat{\boldsymbol{d}}[n]\right\|_2^2\right] - \sum_{i=1}^{B} 2\mathrm{Re}\left(\boldsymbol{\mu}_i^{\mathrm{T}} \boldsymbol{S}_i \boldsymbol{F} \boldsymbol{e}_i\right) - \lambda\left(\sigma_{\boldsymbol{v}}^2 \mathrm{tr}(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}) - E_{\mathrm{tr}}\right)$$

where $\lambda \in \mathbb{R}$ and $\boldsymbol{\mu}_i \in \mathbb{C}^i$, $i = 1, \ldots, B$ are the Lagrangian multipliers. The MSE $\mathrm{E}[\|\boldsymbol{d}[n-\nu] - \hat{\boldsymbol{d}}[n]\|_2^2]$ can be found in (14). Setting the derivatives of the Lagrangian function $L(\boldsymbol{P}, \ldots)$ with respect to the THP filters to zero yields the necessary KKT conditions (e.g., [68], [69]), which we will use to find the WF-THP filters.

The derivative $\partial L(\boldsymbol{P}, \ldots)/\partial \boldsymbol{T}_i$ with respect to the ith coefficient of the temporal feedback filter directly leads to (21)

$$\boldsymbol{T}_i = \begin{cases} -g \boldsymbol{\Pi}_\nu \boldsymbol{\Pi}_{\nu+i}^{\mathrm{T}} \boldsymbol{S}^{(\nu+i)} \boldsymbol{C}\boldsymbol{P}, & i = 1, \ldots, Q+L-\nu \\ \boldsymbol{0}_{B\times B}, & \text{otherwise.} \end{cases}$$

Thereby, we see that the order N_T of the temporal feedback filter is $Q+L-\nu$, i.e., N_T depends on the latency time ν. From the derivative of the Lagrangian function $L(\boldsymbol{P}, \ldots)$ with respect to the spatial feedback filter $\boldsymbol{F}$, we get

$$\boldsymbol{F} = \frac{1}{\sigma_{\boldsymbol{v}}^2} \sum_{i=1}^{B} \boldsymbol{S}_i^{\mathrm{T}} \boldsymbol{\mu}_i^* \boldsymbol{e}_i^{\mathrm{T}} - g \boldsymbol{S}^{(\nu)} \boldsymbol{C}\boldsymbol{P}.$$

Combining this result with the constraint (16) on the structure of $\boldsymbol{F}$ yields, for the Lagrangian multiplier

$$\boldsymbol{\mu}_i^* = \sigma_{\boldsymbol{v}}^2 g \boldsymbol{S}_i \boldsymbol{S}^{(\nu)} \boldsymbol{C}\boldsymbol{P}\boldsymbol{e}_i.$$

Since $\boldsymbol{1}_B = \sum_{i=1}^{B} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}$, we have found (20)

$$\boldsymbol{F} = -g \sum_{i=1}^{B} \left(\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}} \boldsymbol{S}_i\right) \boldsymbol{S}^{(\nu)} \boldsymbol{C}\boldsymbol{P}\boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}.$$

With above results for $\boldsymbol{T}_i$ and $\boldsymbol{F}$, the derivative with respect to the feedforward filter $\boldsymbol{P}$ can be written as

$$\begin{aligned} \frac{\partial L(\boldsymbol{P}, \ldots)}{\partial \boldsymbol{P}} &= -\sigma_{\boldsymbol{v}}^2 g \boldsymbol{C}^{\mathrm{T}} \boldsymbol{S}^{(\nu),\mathrm{T}} \\ &\quad \times \left(\boldsymbol{1}_B + g^* \sum_{i=1}^{B} \left(\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}} \boldsymbol{S}_i\right) \boldsymbol{S}^{(\nu)} \boldsymbol{C}^* \boldsymbol{P}^* \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}\right) \\ &\quad - \sigma_{\boldsymbol{v}}^2 |g|^2 \sum_{i=1}^{Q+L-\nu} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{S}^{(\nu+i),\mathrm{T}} \boldsymbol{S}^{(\nu+i)} \boldsymbol{C}^* \boldsymbol{P}^* \\ &\quad + \sigma_{\boldsymbol{v}}^2 |g|^2 \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C}^* \boldsymbol{P}^* - \sigma_{\boldsymbol{v}}^2 \lambda \boldsymbol{P}^* \end{aligned} \tag{37}$$

which must be zero, that is, $\partial L(\boldsymbol{P}, \ldots)/\partial \boldsymbol{P} = \boldsymbol{0}_{N_{\mathrm{a}}(L+1)\times B}$. When multiplying $\partial L(\boldsymbol{P}, \ldots)/\partial \boldsymbol{P}$ with $\boldsymbol{P}^{\mathrm{T}}$ and applying the trace operator, we obtain an expression that helps to find the Lagrangian multiplier $\lambda \in \mathbb{R}$

$$\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{T}} \frac{\partial L(\boldsymbol{P}, \ldots)}{\partial \boldsymbol{P}}\right) = -|g|^2 \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}}) - \lambda \sigma_{\boldsymbol{v}}^2 \mathrm{tr}(\boldsymbol{P}^{\mathrm{T}} \boldsymbol{P}^*) = 0$$

where the first term $-|g|^2 \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}})$ follows from the derivative of the Lagrangian function $L(\boldsymbol{P}, \ldots)$ with respect to the weight g applied by the receivers. Due to the transmit power constraint (17), we get [see also (7)]

$$\lambda = -|g|^2 \xi.$$

With this result, the ith column of the complex conjugate of the derivative (37) reads

$$\begin{aligned} \left(\frac{\partial L(\boldsymbol{P}, \ldots)}{\partial \boldsymbol{P}}\right)^* \boldsymbol{e}_i &= +\sigma_{\boldsymbol{v}}^2 |g|^2 \boldsymbol{C}^{\mathrm{H}} \boldsymbol{S}_{\nu,i}^{\mathrm{T}} \boldsymbol{S}_{\nu,i} \boldsymbol{C}\boldsymbol{P}\boldsymbol{e}_i \\ &\quad - \sigma_{\boldsymbol{v}}^2 g^* \boldsymbol{C}^{\mathrm{H}} \boldsymbol{S}^{(\nu),\mathrm{T}} \boldsymbol{e}_i + \sigma_{\boldsymbol{v}}^2 |g|^2 \xi \boldsymbol{P}\boldsymbol{e}_i \\ &= \boldsymbol{0}_{N_{\mathrm{a}}(L+1)}. \end{aligned}$$

Here, we made the substitution

$$\boldsymbol{S}_{\nu,i}^{\mathrm{T}} \boldsymbol{S}_{\nu,i} = \boldsymbol{1}_{B(Q+L+1)} - \sum_{j=1}^{Q+L-\nu} \boldsymbol{S}^{(\nu+j),\mathrm{T}} \boldsymbol{S}^{(\nu+j)} - \boldsymbol{S}^{(\nu),\mathrm{T}} \left(\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}} \boldsymbol{S}_i\right) \boldsymbol{S}^{(\nu)} \tag{38}$$

with the selection matrix $\boldsymbol{S}_{\nu,i} \in \{0,1\}^{B\nu+i\times B(Q+L+1)}$ defined in (24). The above equation enables us to compute the ith column of the feedforward filter $\boldsymbol{P}$. Collecting the B columns in one matrix gives the following expression for the feedforward filter:

$$\boldsymbol{P} = \frac{1}{g}\sum_{i=1}^{B}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{C} + \xi\mathbf{1}_{N_{\mathrm{a}}(L+1)}\right)^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}. \tag{39}$$

Finally, the receivers' weight is found by inserting the result for $\boldsymbol{P}$ into the transmit power constraint (17):

$$|g|^2 = \frac{1}{E_{\mathrm{tr}}}\sum_{i=1}^{B}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{C} \times\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{C} + \xi\mathbf{1}_{N_{\mathrm{a}}(L+1)}\right)^{-2}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i.$$

We see that only the amplitude of the weight $g \in \mathbb{C}$ is determined. As any value for the phase of g is compensated by the feedforward filter $\boldsymbol{P}$, we choose $g \in \mathbb{R}_+$, i.e.,g is simply found by taking the square root of above equation.

Up to now, we have found expressions for the WF-THP filters $\boldsymbol{P}, \boldsymbol{F}, \boldsymbol{T}_i$, $i = 1,\ldots,Q+L-\nu$, and g for some latency time ν and precoding order $\mathcal{O}$. Since the previous steps hold for any choice of ν and $\mathcal{O}$, optimality with respect to (13) still holds.

Before we proceed with solving (13), we plug the results for the WF-THP filters into the cost function (14) of (13). With the expressions (20) and (21) for the feedback filters, we get

$$\begin{aligned}
&\mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu]-\hat{\boldsymbol{d}}[n]\right\|_2^2\right]\\
&\quad= \sigma_{\boldsymbol{v}}^2 B - 2\sigma_{\boldsymbol{v}}^2\mathrm{Re}\Big(\mathrm{tr}\Big(g\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}\Big)\Big) + |g|^2\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}})\\
&\qquad+\sigma_{\boldsymbol{v}}^2|g|^2\mathrm{tr}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{P})\\
&\qquad-\sigma_{\boldsymbol{v}}^2\sum_{j=1}^{Q+L-\nu}|g|^2\mathrm{tr}\Big(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}^{(\nu+j),\mathrm{T}}\boldsymbol{S}^{(\nu+j)}\boldsymbol{C}\boldsymbol{P}\Big)\\
&\qquad-\sigma_{\boldsymbol{v}}^2\sum_{i=1}^{B}|g|^2\mathrm{tr}\Big(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}^{(\nu),\mathrm{T}}\Big(\mathbf{1}_B-\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i\Big)\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\Big)\\
&\quad= \sigma_{\boldsymbol{v}}^2 B - 2\sigma_{\boldsymbol{v}}^2\mathrm{Re}\Big(\mathrm{tr}\Big(g\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}\Big)\Big) + |g|^2\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}}) + \sigma_{\boldsymbol{v}}^2\sum_{i=1}^{B}|g|^2\\
&\qquad\times\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{C}\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\right)
\end{aligned}$$

where we incorporated (38) for the second equality. Note that $\sigma_{\boldsymbol{v}}^2\mathrm{tr}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})/E_{\mathrm{tr}} = 1$ due to the transmit power constraint (17). Therefore, we have with (7)

$$|g|^2\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}}) = |g|^2\sigma_{\boldsymbol{v}}^2\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\xi\mathbf{1}_{N_{\mathrm{a}}(L+1)}\boldsymbol{P}\right)$$

and with the obvious relation $\mathbf{1}_B = \sum_{i=1}^{B}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}$, the MSE reads as

$$\begin{aligned}
\mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu]-\hat{\boldsymbol{d}}[n]\right\|_2^2\right] &= \sigma_{\boldsymbol{v}}^2 B - 2\sigma_{\boldsymbol{v}}^2\mathrm{Re}\left(\mathrm{tr}\left(g\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}\right)\right)\\
&+\sigma_{\boldsymbol{v}}^2\sum_{i=1}^{B}|g|^2\mathrm{tr}\left(\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}^{\mathrm{H}}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{C}+\xi\mathbf{1}_{N_{\mathrm{a}}(L+1)}\right)\boldsymbol{P}\right).
\end{aligned}$$

From (39) for the feedforward filter $\boldsymbol{P}$, we see that the last summand of above MSE expression is simply $\sigma_{\boldsymbol{v}}^2\mathrm{tr}(g\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}) \in \mathbb{R}$. Thus

$$\mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu]-\hat{\boldsymbol{d}}[n]\right\|_2^2\right] = \sigma_{\boldsymbol{v}}^2 B - \sigma_{\boldsymbol{v}}^2\mathrm{tr}\left(g\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{P}\right). \tag{40}$$

Note that

$$\boldsymbol{S}^{(\nu)} = \boldsymbol{S}^{(\nu)}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}$$

since $\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}$ sets the last $B(Q+L-\nu+1)-i$ rows to zero but leaves the $B\nu+1$th up to the $B(\nu+1)$th rows unchanged, which are picked out by $\boldsymbol{S}^{(\nu)}$. Therefore, the matrix inversion lemma (e.g.,[61]) can be used to rewrite the expressions for the feedforward filter $\boldsymbol{P}$ and the weight g applied by the receivers:[8]

$$\begin{aligned}
\boldsymbol{P} &= \frac{1}{g}\sum_{i=1}^{B}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}_{\nu,i}^{(\mathcal{O})}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}} \quad\text{and}\\
&= \sqrt{\frac{\sigma_{\boldsymbol{v}}^2}{E_{\mathrm{tr}}}\sum_{i=1}^{B}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}_{\nu,i}^{(\mathcal{O}),2}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i}
\end{aligned} \tag{41}$$

where we introduced the $B(Q+L+1)\times B(Q+L+1)$ block matrix

$$\begin{aligned}
\boldsymbol{B}_{\nu,i}^{(\mathcal{O})} &= \boldsymbol{S}_{\nu,i}^{\mathrm{T}}\left(\boldsymbol{S}_{\nu,i}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}+\xi\mathbf{1}_{B\nu+i}\right)^{-1}\boldsymbol{S}_{\nu,i}\\
&= \begin{bmatrix}\left(\boldsymbol{S}_{\nu,i}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}+\xi\mathbf{1}_{B\nu+i}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}
\end{aligned} \tag{42}$$

whose kth power is denoted by $\boldsymbol{B}_{\nu,i}^{(\mathcal{O}),k}$. Note that only the upper left $B\nu+i\times B\nu+i$ block of $\boldsymbol{B}_{\nu,i}^{(\mathcal{O})}$ is nonzero, and from (see (4) and [57])

$$\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i = \boldsymbol{e}_{\nu+1}\otimes\boldsymbol{e}_i = \boldsymbol{e}_{B\nu+i} \in \{0,1\}^{B(Q+L+1)} \tag{43}$$

we see that multiplying with $\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i$ from the right gives the $B\nu+i$th column of a matrix with $B(Q+L+1)$ columns. Consequently, $\boldsymbol{B}_{\nu,i}^{(\mathcal{O}),k}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i$ is the right-most nonzero column of $\boldsymbol{B}_{\nu,i}^{(\mathcal{O}),k}$. Since this column of $\boldsymbol{B}_{\nu,i}^{(\mathcal{O}),k}$ is not influenced by the lower right block, and the two antidiagonal blocks are zero, we can make following substitution:

$$\boldsymbol{B}_{\nu,i}^{(\mathcal{O}),k}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i = \boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-k}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i \tag{44}$$

with the $B(Q+L+1)\times B(Q+L+1)$ matrix

$$\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O})} = \begin{bmatrix}\boldsymbol{S}_{\nu,i}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}+\xi\mathbf{1}_{B\nu+i} & \mathbf{0}\\ \mathbf{0} & \xi\mathbf{1}\end{bmatrix}$$

which is defined more concisely in (23). The filter solutions resulting from above substitution can be found in (19) and (22). Additionally, we can plug (19) into the MSE expression (40) and find

$$\mathrm{E}\left[\left\|\boldsymbol{d}[n-\nu]-\hat{\boldsymbol{d}}[n]\right\|_2^2\right] = \xi\sigma_{\boldsymbol{v}}^2\sum_{i=1}^{B}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\boldsymbol{A}_{\mathrm{WF},\nu,i}^{(\mathcal{O}),-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i.$$

[8]In particular, we use $(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}+\alpha\mathbf{1}_N)^{-1}\boldsymbol{X}^{\mathrm{H}} = \boldsymbol{X}^{\mathrm{H}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}+\alpha\mathbf{1}_M)^{-1}$, which holds for any $\boldsymbol{X} \in \mathbb{C}^{M\times N}$.

This MSE expression only depends on the latency time ν and the precoding order $\mathcal{O}$. Thus, the optimum ν and $\mathcal{O}$ with respect to the WF-THP optimization (13) minimize the above MSE, as expressed by (18), where the constants ξ and $\sigma_{\boldsymbol{v}}^2$ have been dropped.

APPENDIX II
INCORPORATING THE CHOLESKY FACTORIZATION

In the following, we will use the expression for the feedforward filter $\boldsymbol{P}$ in (41), which is fully equivalent to (19). Inserting the Cholesky factorization (27) into (41), we obtain

$$\boldsymbol{P} = \frac{1}{g}\sum_{i=1}^{B} \boldsymbol{C}^{\mathrm{H}} \boldsymbol{S}_{\nu,i}^{\mathrm{T}} \left(\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1}\boldsymbol{D}^{-1}\boldsymbol{L}^{\mathrm{H},-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1} \times \boldsymbol{S}_{\nu,i}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}. \quad (45)$$

Note that the inverse of a unit upper (lower) triangular matrix is unit upper (lower) triangular [56, p. 91]. Thus, $\boldsymbol{L}^{-1}$ is unit lower triangular. Remember that the selection matrix $\boldsymbol{S}_{\nu,i}$ gives the first $B\nu+i$ rows of a matrix when applied from the left [see (24)]. Since $\boldsymbol{L}^{-1}$ is lower triangular, the upper right $B\nu+i \times B(Q+L-\nu+1)-i$ block of $\boldsymbol{L}^{-1}$ is zero. Hence, following equality holds:

$$\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1} = \boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i} \in \mathbb{C}^{B\nu+i\times B(Q+L+1)}$$

which helps to rewrite the inverse found in the expression for the feedforward filter $\boldsymbol{P}$:

$$\begin{aligned}\left(\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1}\boldsymbol{D}^{-1}\boldsymbol{L}^{\mathrm{H},-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1} &= \left(\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{\mathrm{H},-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{S}_{\nu,i}\boldsymbol{D}^{-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1} \\ &= \boldsymbol{S}_{\nu,i}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{D}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{L}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}. \quad (46)\end{aligned}$$

For the last line, we have used the property of triangular matrices[9] that the inverse of the upper left block is the respective upper left block of the inverse of the triangular matrix.[10] As $\boldsymbol{L}$ is unit lower triangular by definition, the $B\nu+i$th column of $\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{L}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}$ is simply $\boldsymbol{e}_{B\nu+i} \in \{0,1\}^{B(Q+L+1)}$, and since $\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i$ picks out the $B\nu+i$th column of a matrix [see (43)], we have

$$\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{L}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i = \boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i \quad (47)$$

and

$$\left(\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{-1}\boldsymbol{D}^{-1}\boldsymbol{L}^{\mathrm{H},-1}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\right)^{-1}\boldsymbol{S}_{\nu,i}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i = \boldsymbol{S}_{\nu,i}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{D}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i.$$

Before substituting this result into (45) for $\boldsymbol{P}$, remember that $\boldsymbol{D}$ is diagonal and $\boldsymbol{L}^{\mathrm{H}}$ is upper triangular. Consequently, the last $B(Q+L-\nu+1)-i$ elements of the $B\nu+i$th column of both $\boldsymbol{D}$ and $\boldsymbol{L}^{\mathrm{H}}$ are zero, and both occurrences of the projector $\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}$ can be dropped in (45):

$$\begin{aligned}\boldsymbol{P} &= \frac{1}{g}\sum_{i=1}^{B}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{S}_{\nu,i}^{\mathrm{T}}\boldsymbol{S}_{\nu,i}\boldsymbol{D}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}} \\ &= \frac{1}{g}\sum_{i=1}^{B}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}.\end{aligned}$$

Due to $\sum_{i=1}^{B}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}} = \boldsymbol{1}_B$, the last line is equivalent to (28).

To obtain (29), we plug (28) into (20):

$$\boldsymbol{F} = -\sum_{i=1}^{B}\left(\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i\right)\boldsymbol{S}^{(\nu)}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}.$$

Remember that the projector $\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i$ sets the first i elements of a B-dimensional vector to zero. As only the first i elements of the $B\nu+i$th column of the matrix $\boldsymbol{S}^{(\nu)}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{D}$ are different from zero ($\boldsymbol{L}^{\mathrm{H}}$ is upper triangular), we can replace $\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}$ by $\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}} + \xi\boldsymbol{1}_{B(Q+L+1)} = \boldsymbol{L}^{-1}\boldsymbol{D}^{-1}\boldsymbol{L}^{\mathrm{H},-1}$ and get

$$\boldsymbol{F} = -\sum_{i=1}^{B}\left(\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i\right)\boldsymbol{S}^{(\nu)}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}.$$

Since $\boldsymbol{L}^{-1}$ is unit lower triangular, its $\nu+1$th diagonal $B \times B$ block $\boldsymbol{S}^{(\nu)}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}}$ is also unit lower triangular. Thus, the projection of the ith column $\boldsymbol{S}^{(\nu)}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{e}_i$ with $\boldsymbol{1}_B - \boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i$ only sets the ith element, i.e., the ith diagonal element of $-\boldsymbol{S}^{(\nu)}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}}$ to zero. The diagonal of $-\boldsymbol{S}^{(\nu)}\boldsymbol{L}^{-1}\boldsymbol{S}^{(\nu),\mathrm{T}}$ can also be set to zero by adding an identity matrix, as is done in (29).

To get the expression (30) of the temporal feedback filter coefficients depending on the Cholesky factorization (27), similar steps as for the spatial feedback filter (29) are necessary.

[9]A diagonal matrix is a special case of a triangular matrix.

[10]This can be seen from the matrix inversion lemma for partitioned matrices (e.g., [61]) when including the condition that at least one of the off-diagonal blocks is zero:

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}\\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{D}^{-1}\end{bmatrix}$$

for $\boldsymbol{B} = 0$ and/or $\boldsymbol{C} = 0$.

ACKNOWLEDGMENT

The authors would like to thank F. Dietrich and K. Kusume for the discussions on latency time optimization and Cholesky factorization with symmetric permutation, respectively.

REFERENCES

[1] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Tech. J.*, vol. 64, no. 6, pp. 1147–1159, Jul.–Aug. 1985.

[2] J. Yang and S. Roy, "On joint transmitter and receiver optimization for Multiple-Input-Multiple-Output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[3] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[4] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.

[5] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[6] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[7] R. Esmailzadeh, M. Nakagawa, and E. A. Sourour, "Time-division duplex CDMA communications," *IEEE Pers. Commun.*, vol. 4, no. 2, pp. 51–56, Apr. 1997.

[8] M. Haardt, A. Klein, R. Koehn, S. Oestreich, M. Purat, V. Sommer, and T. Ulrich, "The TD-CDMA based UTRA TDD mode," *IEEE J. Select. Areas Commun.*, vol. 18, no. 8, pp. 1375–1385, Aug. 2000.

[9] P. M. Castro and L. Castedo, "Adaptive vector quantization for precoding using blind channel prediction in frequency selective MIMO mobile channels," in *Proc. ITG WSA 2005*, Apr. 2005.

[10] F. Rey, M. Lamarca, and G. Vázquez, "Transmit filter optimization based on partial CSI knowledge for wireless applications," in *Proc. ICC*, May 2003, vol. 4, pp. 2567–2571.

[11] R. Hunger, F. Dietrich, M. Joham, and W. Utschick, "Robust transmit zero-forcing filters," in *Proc. ITG Workshop Smart Antennas*, Mar. 2004, pp. 130–137.

[12] A. P. Liavas, "Tomlinson-Harashima precoding with partial channel knowledge," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 5–9, Jan. 2005.

[13] F. A. Dietrich and W. Utschick, "Robust Tomlinson-Harashima precoding," in *Proc. PIMRC*, Sep. 2005, vol. 1, pp. 136–140.

[14] B. R. Vojčić and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, Oct. 1998.

[15] M. Brandt-Pearce and A. Dharap, "Transmitter-based multiuser interference rejection for the down-link of a wireless CDMA system in a multipath environment," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 407–417, Mar. 2000.

[16] M. Meurer, P. W. Baier, T. Weber, Y. Lu, and A. Papathanassiou, "Joint transmission: Advantageous downlink concept for CDMA mobile radio systems using time division duplexing," *Electron. Lett.*, vol. 36, no. 10, pp. 900–901, May 2000.

[17] A. N. Barreto and G. Fettweis, "Joint signal precoding in the downlink of spread spectrum systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 511–518, May 2003.

[18] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[19] H. R. Karimi, M. Sandell, and J. Salz, "Comparison between transmitter and receiver array processing to achieve interference nulling and diversity," in *Proc. PIMRC'99*, Sep. 1999, vol. 3, pp. 997–1001.

[20] R. L. Choi and R. D. Murch, "New transmit schemes and simplified receiver for MIMO wireless communication systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1217–1230, Nov. 2003.

[21] R. E. McIntosh and S. E. El-Khamy, "Optimum pulse transmission through a plasma medium," *IEEE Trans. Antennas Propagat.*, vol. AP-18, no. 5, pp. 666–671, Sep. 1970.

[22] R. Esmailzadeh and M. Nakagawa, "Pre-RAKE diversity combination for direct sequence spread spectrum mobile communications systems," *IEICE Trans. Commun.*, vol. E76-B, no. 8, pp. 1008–1015, Aug. 1993.

[23] R. L. Choi, K. B. Letaief, and R. D. Murch, "MISO CDMA transmission with simplified receiver for wireless communication handsets," *IEEE Trans. Commun.*, vol. 49, no. 5, pp. 888–898, May 2001.

[24] 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.

[25] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear MIMO precoders for fixed receivers," in *Proc. ICASSP 2004*, May 2004, vol. IV, pp. IV-481–IV-484.

[26] M. Schubert and H. Boche, "Joint 'Dirty Paper' pre-coding and downlink beamforming," in *Proc. ISSSTA 2002*, Sep. 2002, vol. 2, pp. 536–540.

[27] ——, "User ordering and power allocation for optimal multiantenna precoding/decoding," in *Proc. ITG WSA 2004*, Mar. 2004, pp. 174–181.

[28] R. Irmer, R. Habendorf, W. Rave, and G. Fettweis, "Nonlinear multiuser transmission using multiple antennas for TDD-CDMA," in *Proc. WPMC*, Oct. 2003, vol. 3, pp. 251–255.

[29] T. Weber and M. Meurer, "Optimum joint transmission: Potentials and dualities," in *Proc. WPMC*, Oct. 2003, vol. 1, pp. 79–83.

[30] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication—Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 195–202, Mar. 2005.

[31] D. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding," in *Proc. PIMRC*, Sep. 2005, vol. 1, pp. 107–111.

[32] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[33] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[34] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[35] C. B. Peel, "On "dirty-paper coding"," *IEEE Signal Processing Mag.*, vol. 20, no. 3, pp. 112–113, May 2003.

[36] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5/6, pp. 138–139, Mar. 1971.

[37] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, Aug. 1972.

[38] G. D. Forney and M. V. Eyuboğlu, "Combined equalization and coding using precoding," *IEEE Commun. Mag.*, vol. 29, no. 12, pp. 25–34, Dec. 1991.

[39] M. R. Gibbard and A. B. Sesay, "Asymmetric signal processing for indoor wireless LAN's," *IEEE Trans. Veh. Technol.*, vol. 48, no. 6, pp. 2053–2064, Nov. 1999.

[40] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2000, vol. 2, pp. 1627–1631.

[41] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "MIMO precoding for decentralized receivers," in *Proc. ISIT 2002*, Jun./Jul. 2002, p. 496.

[42] 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.

[43] J. Liu and A. Duel-Hallen, "Tomlinson-Harashima transmitter precoding for synchronous multiuser communications," in *Proc. CISS*, Mar. 2003.

[44] M. Joham and W. Utschick, "Ordered spatial Tomlinson Harashima precoding," in *Smart Antennas—State-ofthe-Art*, ser. EURASIP Book Series on Signal Processing and Communications, T. Kaiser, A. Bourdoux, H. Boche, J. R. Fonollosa, J. B. Andersen, and W. Utschick, Eds. : EURASIP, Hindawi, 2006, vol. 3, ch. III, pp. 401–422.

[45] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. ITG SCC*, Jan. 2004, pp. 387–394.

[46] M. Joham, J. Brehmer, A. Voulgarelis, and W. Utschick, "Multiuser spatio-temporal Tomlinson-Harashima precoding for frequency selective vector channels," in *Proc. ITG Workshop Smart Antennas*, Mar. 2004, pp. 208–215.

[47] R. F. H. Fischer, C. Stierstorfer, and J. B. Huber, "Precoding for point-to-multipoint transmission over MIMO ISI channels," in *Proc. Zurich Seminar Commun.*, Feb. 2004, pp. 208–211.

[48] L. Choi and R. D. Murch, "A pre-BLAST-DFE technique for the downlink of frequency-selective fading MIMO channels," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 737–743, May 2004.

[49] C. Degen and L. Brühl, "Linear and successive predistortion in the frequency domain: Performance evaluation in SDMA systems," in *Proc. WCNC*, Mar. 2005.

[50] M. Schubert and S. Shi, "MMSE transmit optimization with interference pre-compensation," in *Proc. VTC Spring*, May 2005.

[51] X. Shao, J. Yuan, and P. Rapajic, "Precoder design for MIMO broadcast channels," in *Proc. ICC*, May 2005.

[52] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Efficient Tomlinson-Harashima precoding for spatial multiplexing on flat MIMO channel," in *Proc. ICC*, May 2005, vol. 3, pp. 2021–2025.

[53] J. Liu and W. A. Krzymień, "A novel nonlinear precoding algorithm for the downlink of multiple antenna multi-user systems," in *Proc. VTC Spring*, May 2005.

[54] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," in *Proc. Int. Zurich Seminar on Broadband Commun.*, Feb. 2002, pp. 7-1–7-6.

[55] 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.

[56] G. Golub and C. V. Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[57] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.

[58] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[59] M. Joham, W. Utschick, and J. A. Nossek, "Latency time optimization for FIR and block transmit filters," in *Proc. ISSPA*, Jul. 2003, vol. 1, pp. 273–276.

[60] M. Joham, R. Irmer, S. Berger, G. Fettweis, and W. Utschick, "Linear precoding approaches for the TDD DS-CDMA downlink," in *Proc. WPMC*, Oct. 2003, vol. 3, pp. 323–327.
[61] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
[62] R. Habendorf, R. Irmer, W. Rave, and G. Fettweis, "Nonlinear multiuser precoding for non-connected decision regions," in *Proc. SPAWC*, Jun. 2005.
[63] U. Erez and R. Zamir, "Achieving $(1/2)\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.
[64] G. D. Forney, Jr., "On the role of MMSE estimation in approaching the informationtheoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. Allerton Conf.*, Oct. 2003.
[65] D. Schmidt, M. Joham, F. Dietrich, K. Kusume, and W. Utschick, "Complexity reduction for MMSE multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. ITG WSA*, Apr. 2005.
[66] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, Sep. 1998, pp. 295–300.
[67] *"Universal Mobile Telecommunications System (UMTS): Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03 Version 3.2.0),"* ETSI, Apr. 1998.
[68] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.
[69] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Michael Joham (S'99–M'05) was born in Kufstein, Austria, in 1974. He received the Dipl.-Ing. and Dr.-Ing. degrees (both *summa cum laude*) in electrical engineering from the Technische Universität München (TUM), Munich, Germany, in 1999 and 2004, respectively.

He was with the Institute of Circuit Theory and Signal Processing at the TUM from 1999 to 2004. Since 2004, he has been with the Associate Institute for Signal Processing at the TUM, where he is currently a senior researcher. In the summers of 1998 and 2000, he visited Purdue University, West Lafayette, IN. His main research interests are estimation theory, reduced-rank processing, and precoding in mobile communications.

Dr. Joham received the VDE Preis for his diploma thesis in 1999 and the Texas-Instruments-Preis for his dissertation in 2004.

David A. Schmidt (S'06) was born in Saarbrücken, Germany, in 1981. He studied electrical engineering at the Munich University of Technology (TUM), Munich, Germany, from 2000 until 2005 and, in the fall of 2003, at Clemson University, Clemson, SC. He received the Dipl.-Ing. degree (summa cum laude) in the summer of 2005 and is currently a Research Assistant at the Associate Institute for Signal Processing, TUM, where he is also working towards his doctorate degree.

His research interests include linear and nonlinear precoding techniques, as well as iterative equalization.

Johannes Brehmer (S'04) was born in Hannover, Germany, in 1978. He received the Dipl.-Ing. in electrical engineering from the Munich University of Technology (TUM), Munich, Germany, in 2003. He was an IMCC exchange student at the University of Texas at Austin, Austin, TX, during 2001–2002. Since 2004, he has been working towards the Dr.-Ing. degree at the Associate Institute for Signal Processing, TUM.

His research interests include cross-layer optimization and signal processing for communications.

Mr. Brehmer received the Werner von Siemens Excellence Award 2003 for his diploma thesis. He held a scholarship from the German National Academic Foundation from 1998 to 2003.

Wolfgang Utschick (M'97–SM'06) completed several industrial education programs before he received the diploma and doctoral degrees, both with honors, in electrical engineering from Munich University of Technology (TUM), Munich, Germany, in 1993 and 1998. During this period, he held a scholarship of the Bavarian Ministry of Education for exceptional students and a scholarship of the Siemens AG.

In 1993, he became a part-time lecturer at a Technical School for Industrial Education. From 1998 to 2002, he was co-director of the Signal Processing Group at the Institute of Circuit Theory and Signal Processing at the TUM. Since 2000, he has been instrumental in the 3rd Generation Partnership Project as an academic consultant in the field of multi-element antenna wireless communication systems. In October 2002, he was appointed Professor at the TUM in the Department of EI, where he is head of the Associate Institute for Signal Processing. His research interests are in signal processing, communications, and applied mathematics in information technology.

Dr. Utschick is Senior Member of the VDE/ITG. He currently an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.