Multiuser Spatio-Temporal Tomlinson-Harashima Precoding for Frequency Selective Vector Channels

Michael Joham, Johannes Brehmer, Andreas Voulgarelis, and Wolfgang Utschick

8th ITG Workshop on Smart Antennas (WSA 2004)
Munich, Germany
pp. 208–215, March 18th–19th, 2004

Munich University of Technology
Institute for Circuit Theory and Signal Processing
http://www.nws.ei.tum.de

Multiuser Spatio-Temporal Tomlinson-Harashima Precoding for Frequency Selective Vector Channels

Michael Joham, Johannes Brehmer, Andreas Voulgarelis, and Wolfgang Utschick
Institute for Circuit Theory and Signal Processing, Munich University of Technology, 80333 Munich, Germany
Phone: +49 (89) 289-28510, Email: Michael.Joham@nws.ei.tum.de

***Abstract*— We derive and compare different approaches to *Tomlinson-Harashima precoding* (THP) in frequency-selective multi-user scenarios with a centralized multi-antenna transmitter and single-antenna receivers. Starting with spatio-temporal THP, we discuss how to simplify this relatively complex approach to systems that perform either spatial THP or temporal THP only. For each approach, we give a MMSE design rule for computing the feedforward and the feedback filters as well as the optimum latency time and the optimum order in which to encode the multiple data streams. Since the zero-forcing variants are special cases of the MMSE THP types, we only explain how zero-forcing THP can be obtain from MMSE THP. We compare the three THP approaches in terms of complexity and, using simulations, in terms of BER performance. While both simplifications to spatio-temporal THP can significantly reduce complexity, simulation results show that in frequency-selective scenarios, the temporal THP component is of particular importance with respect to BER performance.**

I. Introduction

Current and next-generation wireless communication systems such as cellular networks and *wireless local area networks* (WLANs) show a highly asymmetric structure in terms of complexity: A high-complexity *base station* (BS) serves a number of low-complexity *mobile stations* (MSs). This heterogeneous structure is due to the fact that the design of the MSs is tightly constrained in terms of cost and power dissipation, while these constraints are less stringent at the BS. Evidently, asymmetric systems require signal processing algorithms that mirror this asymmetry [1]. In the downlink of asymmetric systems, i. e. transmission from the BS to the MSs, transmit processing represents the appropriate strategy for achieving high-performance wireless transmission while preserving the low complexity of the receivers by processing the signal at the transmitter, that is prior to transmission (precoding or pre-equalization).

Tomlinson-Harashima precoding is a nonlinear precoding scheme that extends the well-known linear precoding schemes *transmit zero-forcing filter* (TxZF, e. g. [2], [3], [4], [5], [6], [7], [8]) and *transmit Wiener filter* (TxWF, e. g. [4], [9], [10]) by adding a feedback filter at the transmitter and a modulo operation at both transmitter and receiver. Similar to *decision feedback equalization* (DFE, [11], [12], [13], [14], [15]) at the receiver, where already *detected symbols* are fed back to reduce interference, THP at the transmitter feeds back already *precoded symbols* to remove interference at the receiver.

Originally, THP was proposed to combat *intersymbol interference* (ISI) in single-user transmission over frequency-selective channels (see [16], [17] and also [18]). Generalizing this concept to the multi-user case leads to what we denote as *temporal THP* (T-THP), i. e. THP is used to combat interference from symbols sent at an earlier time instant.

In [19], [20], THP was used to perform spatial equalization (or multi-user separation) in multi-user transmission over frequency-flat channels. This we call *spatial THP* (S-THP), i. e. THP is used to combat interference from symbols sent at the same time instant (but designated for different users) as the symbol under consideration.

In [21], we combined both temporal THP and spatial THP into joint *spatio-temporal THP* (ST-THP) to end up with a general THP approach for multi-user transmission over frequency-selective channels. Motivated by the advantages of FIR filter structures with respect to real-world implementations, we generally constrain the feedforward filter in a THP system to be FIR, in contrast to [22]. Due to the fact that a FIR transmit filter is used, the performance of ST-THP depends on the chosen latency time. A second parameter that influences system performance is the order in which the data streams designated for the different users are encoded. As discussed in [21], for ST-THP, the joint optimization of latency time and order is computationally infeasible. Therefore, a sub-optimum approach has to be chosen. Still, even the sub-optimum approach presented in [21] induces a high computational complexity.

In this work, we introduce a new optimization for WF-ST-THP which also includes the optimziation of the latency time and the ordering. We also investigate how simplifications of the ST-THP procedure can reduce computational complexity and how these simplifications affect system performance. More precisely, we compare ST-THP with systems that perform either S-THP or T-THP only. In terms of complexity, it turns out that, under certain assumptions, S-THP can greatly simplify the computation of the latency time, whereas the computation of the order is not affected. In contrast to S-THP, in T-THP the computation of the optimum latency time is equivalent to the sub-optimum approach presented in [21] for ST-THP. Still, complexity is significantly reduced, too, as the order does no longer have an impact on the system performance – thus, it is not necessary to compute an optimum order.

We use simulations to compare ST-THP with S-THP and T-THP in terms of performance. Simulation results show that for the scenarios under consideration, system performance is

significantly degraded if ST-THP is simplified into S-THP. On the other hand, T-THP can provide most of the performance of the more complex ST-THP.

After introducing the system model and reviewing the principle of THP in Section II, we employ an new optimization which includes the latency time and the ordering to compute the WF-ST-THP filters together with the appropriate ordering and latency time in Section III. In Sections IV and V, we discuss and compare special cases of the WF-ST-THP, respectively. The simulation results are presented in Section VI.

A. Notation

Vectors and matrices are denoted by lower case bold and capital bold letters, respectively. We use $\mathrm{E}[\bullet]$, '$*$', '$\otimes$', $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, and $(\bullet)^{\mathrm{H}}$ for expectation, convolution, the Kronecker product, complex conjugation, transposition, and conjugate transposition, respectively. The pseudo inverse is denoted by $(\bullet)^+$. All random processes are assumed to be zero-mean and stationary. The variance of the scalar process $y[n]$ is denoted by $\sigma_y^2 = \mathrm{E}\big[|y[n]|^2\big]$. The $N \times M$ zero matrix is $\mathbf{0}_{N\times M}$, the M-dimensional zero vector is $\mathbf{0}_M$, and the $N \times N$ identity matrix is $\mathbf{1}_N$, whose n-th column is $\boldsymbol{e}_n$.

II. System Model

We investigate a system with N_{a} antenna elements at the transmitter and K single antenna receivers as depicted in Fig. 1. The K scalar data streams are collected in the vector data signal

$$\boldsymbol{s}[n] = \big[s_1[n], \ldots, s_K[n]\big]^{\mathrm{T}} \in \mathbb{C}^K,$$

which is first transformed by the permutation matrix

$$\boldsymbol{\Pi}^{(\mathcal{O})} = \sum_{i=1}^{K} \boldsymbol{e}_i \boldsymbol{e}_{b_i}^{\mathrm{T}} \in \{0,1\}^{K\times K},$$

where $\boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}}\boldsymbol{\Pi}^{(\mathcal{O})} = \mathbf{1}_K$, $\boldsymbol{e}_i \in \{0,1\}^K$, and we introduced the ordering

$$\mathcal{O} = \big[b_1, \ldots, b_K\big].$$

The i-th element $b_i, i = 1, \ldots, K$, of the ordering $\mathcal{O}$ is the index of the i-th precoded scalar data stream $s_{b_i}[n]$. Note that $b_i \in \{1, \ldots, K\}\backslash\{b_1, \ldots, b_{i-1}\}$. We chose the above definition for the permutation matrix $\boldsymbol{\Pi}^{(\mathcal{O})}$, since the resulting notation is simple. Note that the permutation matrix $\boldsymbol{\Pi}^{(\mathcal{O})}$ leads to a sorted output $\boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{s}[n]$, whose first entry is precoded first and whose last entry is precoded last

The permuted data signal $\boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{s}[n]$ is passed through the precoder consisting of the matrix FIR feedforward filter

$$\boldsymbol{P}[n] = \sum_{\ell=0}^{L} \boldsymbol{P}_\ell \delta[n-\ell] \in \mathbb{C}^{N_{\mathsf{a}}\times K},$$

and the feedback filters

$$\boldsymbol{F} \in \mathbb{C}^{K\times K} \quad \text{and} \quad \boldsymbol{T}[n] = \sum_{i=1}^{L+Q-\nu} \boldsymbol{T}_i \delta[n-i] \in \mathbb{C}^{K\times K}$$

for spatial and temporal THP, respectively. The temporal feedback filter $\boldsymbol{T}[n]$ has to be strictly *temporally causal*, that is all coefficients for delays smaller than 1 have to be zero. Similarly, the spatial feedback filter $\boldsymbol{F}$ has to be strictly *spatially causal*, i. e. $\boldsymbol{F}$ has to be lower triangular with zero main diagonal due to the sorting introduced by $\boldsymbol{\Pi}^{(\mathcal{O})}$. This causality constraint on the spatial feedback filter $\boldsymbol{F}$ can be split into K constraints on the columns of $\boldsymbol{F}$:

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

where $\boldsymbol{e}_i \in \{0,1\}^K$ and the selection matrix

$$\boldsymbol{S}_i = \big[\mathbf{1}_i, \mathbf{0}_{i\times K-i}\big] \in \mathbb{C}^{i\times K}$$

cuts out the first i elements of a K-dimensional vector. The modulo operation

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

is applied element-wise, e. g. for $\boldsymbol{x} = [x_1, \ldots, x_N]^{\mathrm{T}} \in \mathbb{C}^N$:

$$\mathrm{M}(\boldsymbol{x}) = \big[\mathrm{M}(x_1), \ldots, \mathrm{M}(x_N)\big]^{\mathrm{T}} \in \mathbb{M}^N,$$

and limits the output amplitude of the feedback loop, since the output of $\mathrm{M}(\boldsymbol{x})$ has entries with real and imaginary parts whose absolute values are smaller than $\tau/2$. Equivalently, the output of $\mathrm{M}(x)$ is element of the set

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

The floor operator $\lfloor\bullet\rfloor$ gives the integer number smaller than or equal to the argument. The constant τ depends on the employed modulation alphabet, e. g. $\tau = 2\sqrt{2}$ for QPSK and $\tau = 8/\sqrt{10}$ for 16QAM under the assumption that the symbols have unit variance, that is $\sigma_{s_k}^2 = 1, k = 1, \ldots, K$. We observe that the output of the modulo operator $\mathrm{M}(\bullet)$ is simply the sum of the argument and a signal whose entries have real and imaginary parts which are integer multiples of the constant τ. Hence, we can replace the modulo operators at the transmitter and the receivers by the summation with the auxiliary signals $\boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{a}[n]$ and $-\hat{a}_1[n], \ldots, -\hat{a}_K[n]$, respectively, whose real and imaginary parts are integer multiples of the constant τ. When moving $\boldsymbol{\Pi}^{(\mathcal{O})}\boldsymbol{a}[n]$ to the front of the feedback loop, we obtain Fig. 2. We utilize $\boldsymbol{d}[n] \in \mathbb{C}^K$ as the desired signal and $\hat{\boldsymbol{d}}[n] = [\hat{d}_1[n], \ldots, \hat{d}_K[n]]^{\mathrm{T}} \in \mathbb{C}^K$ as the estimate in the following optimizations, because the auxiliary signals $\boldsymbol{a}[n]$ and $-\hat{a}_1[n], \ldots, -\hat{a}_K[n]$ are automatically generated by the modulo operators to ensure that $\boldsymbol{v}[n] \in \mathbb{M}^K$ and $\tilde{s}_1[n], \ldots, \tilde{s}_K[n] \in \mathbb{M}$. Fig. 2 also helps to understand the basic principle of THP. When the filters are designed following a *zero-forcing* criterion and when the noise is zero, the combination of the precoding filters $\boldsymbol{P}[n]$, $\boldsymbol{F}$, and $\boldsymbol{T}[n]$ together with the scalar weight β^{-1} invert the channels $\boldsymbol{h}_1[n], \ldots, \boldsymbol{h}_K[n]$. Hence, $\hat{a}_i[n] = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{a}[n]$ and $\tilde{s}_i[n] = s_i[n]$ in this case, if the symbols $s_i[n] \in \mathbb{M}, i = 1, \ldots, K$.

From Fig. 2, we see that the desired signal $\boldsymbol{d}[n]$ can be expressed as a function of the modulo output $\boldsymbol{v}[n]$:

$$\boldsymbol{d}[n] = \boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}} \left(\boldsymbol{v}[n] - \boldsymbol{F}\boldsymbol{v}[n] - \boldsymbol{T}[n] * \boldsymbol{v}[n]\right).$$

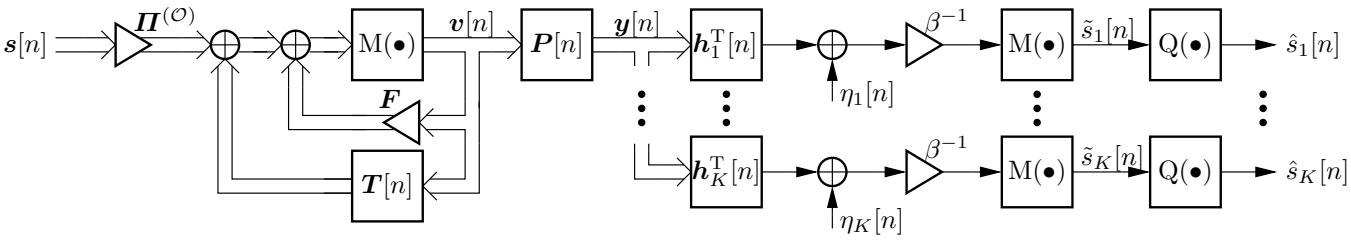

Fig. 1. Transmission over Time Dispersive Vector Channels to K Receivers Applying Modulo Operator $\mathrm{M}(\bullet)$ and Quantizer $\mathrm{Q}(\bullet)$

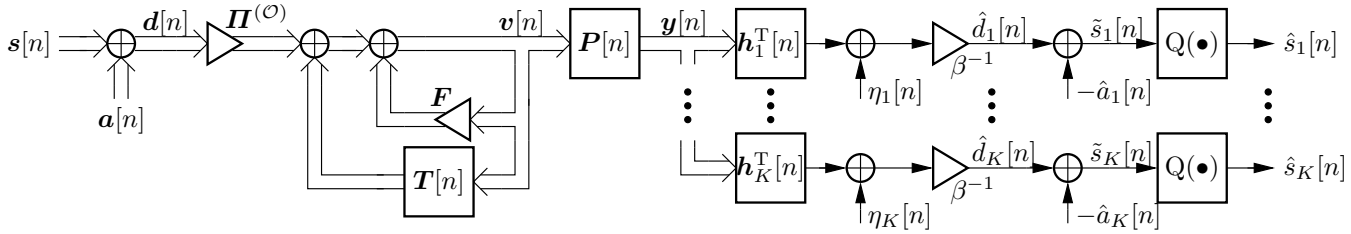

Fig. 2. Linear Representation of the Modulo Operators for the Transmission over Time Dispersive Vector Channels

Obviously, the convolution of $\boldsymbol{v}[n]$ with $\boldsymbol{T}[n]$ can be decomposed in the following way:

$$\boldsymbol{T}[n] * \boldsymbol{v}[n] = \sum_{\ell=1}^{L+Q-\nu} \sum_{i=1}^{K} \boldsymbol{T}_\ell \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{v}[n-\ell],$$

where $\boldsymbol{e}_i \in \{0,1\}^K$. With the definition of the vector containing delayed versions of the i-th modulo output, the coefficients of the temporal feedback filter $\boldsymbol{T}[n]$ applied to the i-th modulo output, and the selection matrix cutting out the last $L+Q-\nu$ elements of an $L+Q+1$-dimensional vector

$$\begin{aligned}
\boldsymbol{v}_i[n] &= \left[\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{v}[n], \ldots, \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{v}[n-L-Q]\right]^{\mathrm{T}} \in \mathbb{M}^{L+Q+1}, \\
\bar{\boldsymbol{T}}_i &= \left[\boldsymbol{T}_1\boldsymbol{e}_i, \ldots, \boldsymbol{T}_{L+Q-\nu}\boldsymbol{e}_i\right] \in \mathbb{C}^{K\times L+Q-\nu}, \quad \text{and} \\
\boldsymbol{S}^{(\nu)} &= \left[\boldsymbol{0}_{L+Q-\nu\times\nu+1}, \boldsymbol{1}_{L+Q-\nu}\right] \in \{0,1\}^{L+Q-\nu\times L+Q+1},
\end{aligned}$$

respectively, we can rewrite above result for the b_k-th desired signal $d_{b_k}[n] = \boldsymbol{e}_{b_k}^{\mathrm{T}}\boldsymbol{d}[n], k = 1,\ldots,K$:

$$d_{b_k}[n-\nu] = \boldsymbol{e}_{\nu+1}^{\mathrm{T}}\boldsymbol{v}_k[n] - \boldsymbol{e}_k^{\mathrm{T}} \sum_{i=1}^{K} \left(\boldsymbol{F}\boldsymbol{e}_i\boldsymbol{e}_{\nu+1}^{\mathrm{T}} + \bar{\boldsymbol{T}}_i\boldsymbol{S}^{(\nu)}\right)\boldsymbol{v}_i[n], \tag{2}$$

since $\boldsymbol{e}_{b_k}^{\mathrm{T}}\boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}} = \boldsymbol{e}_k^{\mathrm{T}}$. Here, $\boldsymbol{e}_{\nu+1} \in \{0,1\}^{L+Q+1}$ and $\boldsymbol{e}_{\nu+1}^{\mathrm{T}}\boldsymbol{v}_k[n] = \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{v}[n-\nu]$. Note that the delayed $d_{b_k}[n-\nu]$ is the desired signal. In other words, we allow a latency time $\nu \in \{0,\ldots,L+Q\}$. Thus, a symbol is estimated at the receiver ν symbol times after it has been applied to the input of the precoder at the transmitter.

The output $\boldsymbol{v}[n]$ of the modulo operator at the transmitter is passed through the feedforward filter $\boldsymbol{P}[n]$, propagates over the vector FIR channel of order Q

$$\boldsymbol{h}_{b_k}[n] = \sum_{q=0}^{Q} \boldsymbol{h}_{b_k,q}\delta[n-q] \in \mathbb{C}^{N_{\mathrm{a}}},$$

is perturbed by the complex noise $\eta_{b_k}[n] \in \mathbb{C}$, and is weighted with the scalar $\beta^{-1} \in \mathbb{R}_+$ to form the estimate at the b_k-th receiver:

$$\begin{aligned}
\hat{d}_{b_k}[n] &= \beta^{-1}\boldsymbol{h}_{b_k}^{\mathrm{T}}[n] * \boldsymbol{P}[n] * \boldsymbol{v}[n] + \beta^{-1}\eta_{b_k}[n], \\
&= \beta^{-1}\sum_{\ell=0}^{L}\sum_{q=0}^{Q}\sum_{i=1}^{K} \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}_\ell^{\mathrm{T}}\boldsymbol{h}_{b_k,q}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{v}[n-\ell-q] + \beta^{-1}\eta_{b_k}[n].
\end{aligned}$$

Here, $\boldsymbol{e}_i \in \{0,1\}^K$ and $b_k \in \{1,\ldots,K\}\backslash\{b_1,\ldots,b_{k-1}\}$ is the k-th element of the ordering $\mathcal{O}$. Alternatively, we can write:

$$\hat{d}_{b_k}[n] = \beta^{-1}\sum_{i=1}^{K} \boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{H}_{b_k}\boldsymbol{v}_i[n] + \beta^{-1}\eta_{b_k}[n], \tag{3}$$

where we introduced the vector with the feedforward filter coefficients for the i-th precoded data stream and the block Toeplitz convolution matrix of the b_k-th channel:

$$\boldsymbol{p}_i = \left[\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}_0^{\mathrm{T}}, \ldots, \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}_L^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{a}}(L+1)} \quad \text{and}$$

$$\boldsymbol{H}_{b_k} = \sum_{q=0}^{Q} \boldsymbol{S}_{(q,L+1,Q)} \otimes \boldsymbol{h}_{b_k,q} \in \mathbb{C}^{N_{\mathrm{a}}(L+1)\times L+Q+1},$$

respectively. The selection matrix employed for the b_k-th channel matrix $\boldsymbol{H}_{b_k}$ is defined as

$$\boldsymbol{S}_{(q,M,N)} = \left[\boldsymbol{0}_{M\times q}, \boldsymbol{1}_M, \boldsymbol{0}_{M\times N-q}\right] \in \{0,1\}^{M\times M+N}.$$

Note that the desired signal for the estimate $\hat{d}_{b_k}[n]$ in (3) is the delayed signal $d_{b_k}[n-\nu]$ (see Eq. 2).

The transmit power can be expressed as

$$\mathrm{E}\left[\|\boldsymbol{y}[n]\|_2^2\right] = \sum_{i=1}^{K} \sigma_{\boldsymbol{v}}^2 \boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{p}_i, \tag{4}$$

because we make the assumption that the outputs $\boldsymbol{v}[n]$ of the modulo operator at the transmitter are uncorrelated, i. e. $\mathrm{E}[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+k]] = \sigma_{\boldsymbol{v}}^2\delta[k]\boldsymbol{1}_K$. Under the popular assumption that the entries of $\boldsymbol{v}[n]$ are uniformly distributed over $\mathbb{M}$, we get $\sigma_{\boldsymbol{v}}^2 = \tau^2/6$.

III. Multiuser Wiener Spatio-Temporal THP (WF-ST-THP)

The linear *transmit Wiener filter* (TxWF) minimizes the *mean square error* (MSE) together with a transmit power constraint (see e. g. [9]). We use a similar optimization for *Wiener spatio-temporal THP* (WF-ST-THP). As was noted in Section II, the spatial feedback filter $\boldsymbol{F}$ has to be spatially causal:

$$\begin{aligned}\left\{\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{st}}, \boldsymbol{F}_{\mathrm{WF}}^{\mathrm{st}}, \boldsymbol{T}_{\mathrm{WF}}^{\mathrm{st}}, \beta_{\mathrm{WF}}^{\mathrm{st}}, \mathcal{O}_{\mathrm{WF}}^{\mathrm{st}}, \nu_{\mathrm{WF}}^{\mathrm{st}}\right\} &= \underset{\{\boldsymbol{P},\boldsymbol{F},\boldsymbol{T},\beta,\mathcal{O},\nu\}}{\operatorname{argmin}} \sum_{k=1}^{K} \sigma_{\varepsilon_k}^2 \\ \text{s. t.:}\quad \mathrm{E}\Big[\|\boldsymbol{y}[n]\|_2^2\Big] &= E_{\mathrm{tr}} \quad \text{and} \\ \boldsymbol{F}:&\ \text{lower triangular, zero main diagonal.}\end{aligned} \tag{5}$$

For brevity, we combine the feedforward filter and temporal feedback filter coefficients in

$$\boldsymbol{P} = \big[\boldsymbol{p}_1, \dots, \boldsymbol{p}_K\big] \quad \text{and} \quad \boldsymbol{T} = \big[\bar{\boldsymbol{T}}_1, \dots, \bar{\boldsymbol{T}}_K\big],$$

respectively, and the MSE for the b_k-th data stream reads as

$$\sigma_{\varepsilon_k}^2 = \mathrm{E}\Big[\Big|d_{b_k}[n-\nu] - \hat{d}_{b_k}[n]\Big|^2\Big]. \tag{6}$$

Plugging (2) and (3) into the MSE definition, yields:

$$\begin{aligned}\sigma_{\varepsilon_k}^2 &= \beta^{-2}\sigma_{\eta_{b_k}}^2 + \sigma_{\boldsymbol{v}}^2\left(1 + \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{e}_k + \sum_{i=1}^{K}\boldsymbol{e}_k^{\mathrm{T}}\bar{\boldsymbol{T}}_i\bar{\boldsymbol{T}}_i^{\mathrm{H}}\boldsymbol{e}_k\right) \\ &+ 2\beta^{-1}\sigma_{\boldsymbol{v}}^2\sum_{i=1}^{K}\mathrm{Re}\left(\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{H}_{b_k}\left(\boldsymbol{e}_{\nu+1}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{F}^{\mathrm{H}} + \boldsymbol{S}^{(\nu),\mathrm{T}}\bar{\boldsymbol{T}}_i^{\mathrm{H}}\right)\boldsymbol{e}_k\right) \\ &- 2\beta^{-1}\sigma_{\boldsymbol{v}}^2\,\mathrm{Re}\left(\boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{H}_{b_k}\boldsymbol{e}_{\nu+1}\right) + \beta^{-2}\sigma_{\boldsymbol{v}}^2\sum_{i=1}^{K}\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{H}_{b_k}\boldsymbol{H}_{b_k}^{\mathrm{H}}\boldsymbol{p}_i^*,\end{aligned}$$

since the diagonal elements of the spatial feedback filter $\boldsymbol{F}$ are zero and the modulo outputs are uncorrelated, that is $\mathrm{E}[\boldsymbol{v}_i[n]\boldsymbol{v}_k^{\mathrm{H}}[n]] = \sigma_{\boldsymbol{v}}^2\delta[i-k]\mathbf{1}_{L+Q+1}$. The transmit power constraint can be rewritten with (4) and the causality constraint on $\boldsymbol{F}$ can be found in (1). Above WF-ST-THP optimization in (5) can be solved with the Lagrangian function

$$\begin{aligned}L(\boldsymbol{P},\dots) = &\sum_{k=1}^{K}\sigma_{\varepsilon_k}^2 - \rho\left(\sum_{i=1}^{K}\sigma_{\boldsymbol{v}}^2\boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{p}_i - E_{\mathrm{tr}}\right) \\ &- \sum_{i=1}^{K}2\,\mathrm{Re}\left(\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{S}_i\boldsymbol{F}\boldsymbol{e}_i\right),\end{aligned}$$

where the Lagrangian multipliers are $\boldsymbol{\mu}_i \in \mathbb{C}^i, i = 1,\dots,K$, and $\rho \in \mathbb{R}$. The derivatives of the Lagrangian function $L(\boldsymbol{P},\dots)$ with respect to the i-th temporal feedback filter and the spatial feedback filter result in ($\boldsymbol{F}_{\mathrm{WF}}^{\mathrm{st}} \in \mathbb{C}^{K\times K}$):

$$\bar{\boldsymbol{T}}_{\mathrm{WF},i}^{\mathrm{st}} = -\beta_{\mathrm{WF}}^{\mathrm{st},-1}\sum_{k=1}^{K}\boldsymbol{e}_k\boldsymbol{p}_{\mathrm{WF},i}^{\mathrm{st},\mathrm{T}}\boldsymbol{H}_{b_k}\boldsymbol{S}^{(\nu),\mathrm{T}} \in \mathbb{C}^{K\times L+Q-\nu}, \tag{7}$$

$$\boldsymbol{F}_{\mathrm{WF}}^{\mathrm{st}} = \beta_{\mathrm{WF}}^{\mathrm{st},-1}\sum_{i=1}^{K}\sum_{k=1}^{K}\left(\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i - \mathbf{1}_K\right)\boldsymbol{e}_k\boldsymbol{p}_{\mathrm{WF},i}^{\mathrm{st},\mathrm{T}}\boldsymbol{H}_{b_k}\boldsymbol{e}_{\nu+1}\boldsymbol{e}_i^{\mathrm{T}}, \tag{8}$$

respectively. For the spatial feedback filter expression, we used the causality condition for $\boldsymbol{F}$ in (1). Multiplying the derivative of the Lagrangian function $L(\boldsymbol{P},\dots)$ with respect to $\boldsymbol{p}_i$ with $\boldsymbol{p}_i^{\mathrm{T}}$ from the left, summing up over $i = 1,\dots,K$, and plugging this result into the derivative of $L(\boldsymbol{P},\dots)$ with respect to β leads to:

$$\rho = -\beta^{-2}\,\frac{\sum_{k=1}^{K}\sigma_{\eta_{b_k}}^2}{E_{\mathrm{tr}}} = -\beta^{-2}\xi_{\mathrm{WF}},$$

where we also utilized the transmit power constraint. When including this result, (7), and (8) into the derivative of $L(\boldsymbol{P},\dots)$ with respect to the feedforward filter $\boldsymbol{p}_i$, we obtain the WF-ST-THP feedforward filter:

$$\boldsymbol{p}_{\mathrm{WF},i}^{\mathrm{st},\mathrm{T}} = \beta_{\mathrm{WF}}^{\mathrm{st}}\boldsymbol{e}_{\nu+1}^{\mathrm{T}}\boldsymbol{H}_{b_i}^{\mathrm{H}}\left(\sum_{k=1}^{K}\boldsymbol{H}_{b_k}\boldsymbol{\Pi}_{k,i}^{(\nu)}\boldsymbol{H}_{b_k}^{\mathrm{H}} + \xi_{\mathrm{WF}}\mathbf{1}_{N_{\mathrm{a}}(L+1)}\right)^{-1}. \tag{9}$$

The introduced projector $\boldsymbol{\Pi}_{k,i}^{(\nu)} \in \{0,1\}^{L+Q+1\times L+Q+1}$ can be written as

$$\boldsymbol{\Pi}_{k,i}^{(\nu)} = \mathbf{1}_{L+Q+1} + \left(\|\boldsymbol{S}_i\boldsymbol{e}_k\|_2 - 1\right)\boldsymbol{e}_{\nu+1}\boldsymbol{e}_{\nu+1}^{\mathrm{T}} - \boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{S}^{(\nu)}. \tag{10}$$

Consequently, the last $L+Q-\nu$ columns of $\boldsymbol{H}_{b_k}$ and the last rows of $\boldsymbol{H}_{b_k}^{\mathrm{H}}$ are always set to zero, whereas the $\nu+1$-th column of $\boldsymbol{H}_{b_k}$ (the $\nu+1$-th row of $\boldsymbol{H}_{b_k}^{\mathrm{H}}$) is only set to zero, if $k > i$, that is the $\nu+1$-th column is only changed for channels belonging to data streams which are precoded later. Note that $\boldsymbol{\Pi}_{k,K}^{(\nu)} = \mathbf{1}_{L+Q+1} - \boldsymbol{S}^{(\nu),\mathrm{T}}\boldsymbol{S}^{(\nu)}$, i. e. only the last $L+Q-\nu$ columns of $\boldsymbol{H}_{b_k}, k = 1,\dots,K$, are set to zero in this case.

In (7), (8), and (9), we can find the WF-ST-THP filters for a fixed latency time ν and a fixed ordering $\mathcal{O}$. To obtain the optimum ν and $\mathcal{O}$, we plug (7), (8), and (9) into the MSE for the b_i-th data stream:

$$\sigma_{\varepsilon_i}^2 = \sigma_{\boldsymbol{v}}^2 - \beta_{\mathrm{WF}}^{\mathrm{st},-1}\sigma_{\boldsymbol{v}}^2\boldsymbol{p}_{\mathrm{WF},i}^{\mathrm{st},\mathrm{T}}\boldsymbol{H}_{b_i}\boldsymbol{e}_{\nu+1}.$$

After applying the *matrix inversion lemma* (e. g. [23]) to $\boldsymbol{p}_{\mathrm{WF},i}^{\mathrm{st},\mathrm{T}}$ and summing up over $i = 1,\dots,K$, we get the total MSE:

$$\sigma_{\varepsilon}^2 = \sum_{i=1}^{K}\sigma_{\varepsilon_i}^2 = \xi_{\mathrm{WF}}\sigma_{\boldsymbol{v}}^2\sum_{i=1}^{K}\boldsymbol{e}_{b_i}^{(\nu),\mathrm{T}}\boldsymbol{C}_i^{(\mathcal{O},\nu),-1}\boldsymbol{e}_{b_i}^{(\nu)}, \tag{11}$$

where we use the abbreviation

$$\boldsymbol{C}_i^{(\mathcal{O},\nu)} = \boldsymbol{\Pi}_i^{(\mathcal{O},\nu)}\check{\boldsymbol{H}}^{\mathrm{H}}\check{\boldsymbol{H}}\boldsymbol{\Pi}_i^{(\mathcal{O},\nu)} + \xi_{\mathrm{WF}}\mathbf{1}_{K(L+Q+1)},$$

with

$$\begin{aligned}\boldsymbol{\Pi}_i^{(\mathcal{O},\nu)} &= \sum_{k=1}^{K}\left(\boldsymbol{e}_{b_k}\boldsymbol{e}_{b_k}^{\mathrm{T}}\right)\otimes\boldsymbol{\Pi}_{k,i}^{(\nu)}, \\ \check{\boldsymbol{H}} &= \big[\boldsymbol{H}_1,\dots,\boldsymbol{H}_K\big] \in \mathbb{C}^{N_{\mathrm{a}}(L+1)\times K(L+Q+1)}, \quad \text{and} \\ \boldsymbol{e}_{b_i}^{(\nu)} &= \boldsymbol{e}_{b_i}\otimes\boldsymbol{e}_{\nu+1} \in \{0,1\}^{K(L+Q+1)}.\end{aligned} \tag{12}$$

Here, $\boldsymbol{e}_{b_i} \in \{0,1\}^K$ and $\boldsymbol{e}_{\nu+1} \in \{0,1\}^{L+Q+1}$. The optimum choice for the latency time ν and the ordering $\mathcal{O}$ can be obtained by trying all $K!(L+Q+1)$ possible combinations of ν and $\mathcal{O}$ and using the one which minimizes the total MSE in (11). Since this approach is too complex, we suggest to employ

1:	$\nu \leftarrow \underset{\nu'}{\operatorname{argmin}} \sum_{i=1}^{K} \boldsymbol{e}_i^{(\nu'),\mathrm{T}} \boldsymbol{C}_K^{(\mathcal{O},\nu'),-1} \boldsymbol{e}_i^{(\nu')}$
	$\mathbb{O} \leftarrow \{1,\ldots,K\}$
	$\boldsymbol{G} \leftarrow \check{\boldsymbol{H}} \boldsymbol{\Pi}_K^{(\mathcal{O},\nu)}$
4:	for $i = K,\ldots,1$:
5:	$\quad \boldsymbol{C} \leftarrow (\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G} + \xi_{\mathrm{WF}} \mathbf{1}_{K(L+Q+1)})^{-1}$
6:	$\quad b_i \leftarrow \underset{b\in\mathbb{O}}{\operatorname{argmin}}\, \boldsymbol{e}_b^{(\nu),\mathrm{T}} \boldsymbol{C} \boldsymbol{e}_b^{(\nu)}$
7:	$\quad \boldsymbol{p}_i^{\mathrm{T}} \leftarrow \boldsymbol{e}_{b_i}^{(\nu),\mathrm{T}} \boldsymbol{C}\boldsymbol{G}^{\mathrm{H}}$
	$\quad \mathbb{O} \leftarrow \mathbb{O} \backslash \{b_i\}$
9:	$\quad \boldsymbol{G} \leftarrow \boldsymbol{G}(\mathbf{1}_{K(L+Q+1)} - \boldsymbol{e}_{b_i}^{(\nu)} \boldsymbol{e}_{b_i}^{(\nu),\mathrm{T}})$
	$\chi \leftarrow \sigma_{\boldsymbol{v}}^2 \sum_{i=1}^{K} \lVert \boldsymbol{p}_i \rVert_2^2$
	$\beta \leftarrow \sqrt{E_{\mathrm{tr}}/\chi}$
	for $i = 0,\ldots,L$:
	$\quad \boldsymbol{P}_i = \beta[\mathbf{0}_{N_{\mathrm{a}}\times N_{\mathrm{a}} i}, \mathbf{1}_{N_{\mathrm{a}}}, \mathbf{0}_{N_{\mathrm{a}}\times N_{\mathrm{a}}(L-i)}][\boldsymbol{p}_1,\ldots,\boldsymbol{p}_K]$
	for $i = 1,\ldots,K$:
15:	$\quad \boldsymbol{f}_i \leftarrow (\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i - \mathbf{1}_K) \sum_{j=1}^{K} \boldsymbol{e}_j \boldsymbol{p}_i^{\mathrm{T}} \check{\boldsymbol{H}} \boldsymbol{e}_{b_j}^{(\nu)}$
16:	$\quad \bar{\boldsymbol{T}}_i \leftarrow -\sum_{j=1}^{K} \boldsymbol{e}_j \boldsymbol{p}_i^{\mathrm{T}} \check{\boldsymbol{H}} (\boldsymbol{e}_{b_j} \otimes \mathbf{1}_{L+Q+1}) \boldsymbol{S}^{(\nu),\mathrm{T}}$
	for $i = 1,\ldots,L+Q-\nu$:
18:	$\quad \boldsymbol{T}_i = [\bar{\boldsymbol{T}}_1 \boldsymbol{e}_i,\ldots,\bar{\boldsymbol{T}}_K \boldsymbol{e}_i]$

TABLE I

FILTER COMPUTATION FOR WF-ST-THP

1:	for $i = 1,\ldots,K$:
2:	$\quad v_i[n] \leftarrow \mathrm{M}(s_{b_i}[n] + \boldsymbol{e}_i^{\mathrm{T}} \sum_{j=1}^{i-1} \boldsymbol{f}_j v_j[n])$
3:	$\quad v_i[n] \leftarrow \mathrm{M}(v_i[n] + \boldsymbol{e}_i^{\mathrm{T}} \sum_{j=1}^{L+Q-\nu} \tilde{\boldsymbol{T}}_j \boldsymbol{v}[n-j])$
	$\boldsymbol{v}[n] = [v_1[n],\ldots,v_K[n]]^{\mathrm{T}}$
	$\boldsymbol{y}[n] = \sum_{\ell=0}^{L} \boldsymbol{P}_\ell \boldsymbol{v}[n-\ell]$

TABLE II

ORDERED ST-THP

following suboptimum approach which decouples the latency time optimization from the ordering optimization. First, we optimize the latency time ν under the assumption that the spatial feedback filter $\boldsymbol{F}$ is inactive. Thus, the projector in the matrix $\boldsymbol{C}_i^{(\mathcal{O},\nu)}$ simplifies to

$$\boldsymbol{\Pi}_K^{(\mathcal{O},\nu)} = \mathbf{1}_K \otimes \left(\mathbf{1}_{L+Q+1} - \boldsymbol{S}^{(\nu),\mathrm{T}} \boldsymbol{S}^{(\nu)}\right),$$

that is $\boldsymbol{C}_i^{(\mathcal{O},\nu)}$ is independent of the ordering $\mathcal{O}$ and the latency time $\nu_{\mathrm{WF}}^{\mathrm{st}}$ can be found by minimizing the sum of diagonal elements of $\boldsymbol{C}_K^{(\mathcal{O},\nu),-1}$. Note that we still have to compute the inverse of $\boldsymbol{C}_K^{(\mathcal{O},\nu)}$ for every value of the latency time $\nu \in \{0,\ldots,L+Q+1\}$, since $\boldsymbol{C}_K^{(\mathcal{O},\nu)}$ depends on ν. Second, the ordering $\mathcal{O}$ is optimized for the fixed latency time $\nu_{\mathrm{WF}}^{\mathrm{st}}$ and instead of minimizing the whole sum of the total MSE in (11), we minimize each summand under the assumption that the following elements of the ordering $\mathcal{O}_{\mathrm{WF}}^{\mathrm{st}}$ are fixed:

$$b_{\mathrm{WF},i}^{\mathrm{st}} = \underset{b\in\{1,\ldots,K\}\backslash\{b_{\mathrm{WF},i+1}^{\mathrm{st}},\ldots,b_{\mathrm{WF},K}^{\mathrm{st}}\}}{\operatorname{argmin}} \boldsymbol{e}_b^{(\nu),\mathrm{T}} \boldsymbol{C}_i^{(\mathcal{O}_{\mathrm{WF}}^{\mathrm{st}},\nu_{\mathrm{WF}}^{\mathrm{st}}),-1} \boldsymbol{e}_b^{(\nu)},$$

where $i = K,\ldots,1$. Note that the matrix $\boldsymbol{C}_i^{(\mathcal{O}_{\mathrm{WF}}^{\mathrm{st}},\nu_{\mathrm{WF}}^{\mathrm{st}})}$ only depends on $b_{\mathrm{WF},i+1}^{\mathrm{st}},\ldots,b_{\mathrm{WF},K}^{\mathrm{st}}$ and is therefore independent of $b_{\mathrm{WF},1}^{\mathrm{st}},\ldots,b_{\mathrm{WF},i}^{\mathrm{st}}$.

The resulting algorithm for the computation of the WF-ST-THP filters is given as pseudo code in Table I. The procedure

1:	$\nu \leftarrow \underset{\nu'}{\operatorname{argmin}} \sum_{i=1}^{K} \lVert (\boldsymbol{\Pi}_K^{(\mathcal{O},\nu)} \check{\boldsymbol{H}}^{\mathrm{T}})^{+} \boldsymbol{e}_i^{(\nu')} \rVert_2^2$
5:	$\quad \boldsymbol{P} \leftarrow \boldsymbol{G}^{+}$
6:	$\quad b_i \leftarrow \underset{b\in\mathbb{O}}{\operatorname{argmin}} \lVert \boldsymbol{P}^{\mathrm{T}} \boldsymbol{e}_b^{(\nu)} \rVert_2^2$
7:	$\quad \boldsymbol{p}_i^{\mathrm{T}} \leftarrow \boldsymbol{e}_{b_i}^{(\nu),\mathrm{T}} \boldsymbol{P}$

TABLE III

FILTER COMPUTATION FOR ZF-ST-THP

to compute the transmit signal $\boldsymbol{y}[n]$ of ST-THP can be found in Table II.

IV. SPECIAL CASES OF WF-ST-THP

A. Zero-Forcing Spatio-Temporal THP (WF-ST-THP)

The WF-ST-THP optimization in (5) can be used to find the *zero-forcing spatio-temporal THP* (ZF-ST-THP), when following zero-forcing conditions are included:

$$\hat{d}_{b_k}[n]\Big|_{\eta_{b_k}[n]=0} = d_{b_k}[n], \quad k = 1,\ldots,K.$$

We do not show the derivation of the ZF-ST-THP filters, but it is intuitively clear that they can be obtained from their WF-ST-THP counterparts, when we apply the limit $\xi_{\mathrm{WF}} \to 0$ (see e. g. [21]). For brevity, we only give the solution for the ZF-ST-THP feedforward filter:

$$\boldsymbol{p}_{\mathrm{ZF},i}^{\mathrm{st},\mathrm{T}} = \beta_{\mathrm{ZF}}^{\mathrm{st}} \boldsymbol{e}_{b_i}^{(\nu),\mathrm{T}} \left(\check{\boldsymbol{H}} \boldsymbol{\Pi}_i^{(\mathcal{O},\nu)}\right)^{+}.$$

The ZF-ST-THP algorithm is identical to the WF-ST-THP algorithm in Table I except for the lines 1, 5, 6, and 7, which can be found in Table III.

B. Wiener Spatial THP (WF-S-THP)

When adding the constraint

$$\boldsymbol{T}[n] = \mathbf{0}_{K\times K}\delta[n]$$

to the WF-ST-THP optimization in (5), we end up with the *Wiener spatial THP* (WF-S-THP), since the temporal feedback filter is switched off. The WF-ST-THP results for the feedforward filter in (9) and for the spatial feedback filter in (8) are also valid for WF-S-THP, if we redefine the projector (cf. Eq. 10)

$$\boldsymbol{\Pi}_{k,i}^{(\nu)} = \mathbf{1}_{L+Q+1} + (\lVert \boldsymbol{S}_i \boldsymbol{e}_k \rVert_2 - 1)\, \boldsymbol{e}_{\nu+1} \boldsymbol{e}_{\nu+1}^{\mathrm{T}},$$

which is part of the projector $\boldsymbol{\Pi}_i^{(\mathcal{O},\nu)}$ in (12). After skipping lines 16–18 in Table I and taking into account the redefinition of the projector $\boldsymbol{\Pi}_i^{(\mathcal{O},\nu)}$, we end up with the WF-S-THP algorithm to compute the filter coefficients. In Table II, we must drop line 3. Again, *zero-forcing spatial THP* (ZF-S-THP) can be found by applying the limit $\xi_{\mathrm{WF}} \to 0$.

We see that WF-S-THP has significantly lower complexity than WF-ST-THP, because the latency time optimization is dramatically simplified. This simplification is due to the sub-optimality of the latency time optimization, since we assume that the spatial feedback filter is inactive. Hence,

$$\boldsymbol{C}_i^{(\mathcal{O},\nu)} = \check{\boldsymbol{H}}^{\mathrm{H}} \check{\boldsymbol{H}} + \xi_{\mathrm{WF}} \mathbf{1}_{K(L+Q+1)}$$

4:	$\boldsymbol{C} \leftarrow (\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G} + \xi_{\mathrm{WF}}\mathbf{1}_{K(L+Q+1)})^{-1}$
	for $i = 1, \ldots, K$:
	$\boldsymbol{p}_i^{\mathrm{T}} \leftarrow \boldsymbol{e}_i^{(\nu),\mathrm{T}}\boldsymbol{C}\boldsymbol{G}^{\mathrm{H}}$

TABLE IV
FILTER COMPUTATION FOR WF-T-THP

is independent of ν and the WF-S-THP latency time optimization is the same as the TxWF latency time optimization. However, we have to perform an optimization of the ordering for WF-S-THP.

C. Wiener Temporal THP (WF-T-THP)

By restricting the spatial feedback filter to be zero, that is

$$\boldsymbol{F} = \mathbf{0}_{K\times K},$$

the WF-ST-THP optimization in (5) leads to the *Wiener temporal THP* (WF-T-THP). The solutions for the feedback filter in (7) and for the feedforward filter in (9) can also be used for WF-T-THP, if we plug in $\boldsymbol{\Pi}_{K,K}^{(\nu)}$ instead of $\boldsymbol{\Pi}_{k,i}^{(\nu)}$ (see Eq. 10) and set (cf. Eq. 12):

$$\boldsymbol{\Pi}_i^{(\mathcal{O},\nu)} = \mathbf{1}_K \otimes \boldsymbol{\Pi}_{K,K}^{(\nu)},$$

i. e. the filters are independent of the ordering $\mathcal{O}$. Table I is also valid for WF-T-THP, if we drop line 15 and replace the lines 4–9 by the lines of Table IV. Moreover, we have to skip the second line in Table II and replace $v_i[n]$ by $s_{b_i}[n]$ in the third line of Table II. Note that the limit $\xi_{\mathrm{WF}} \to 0$ yields *zero-forcing temporal THP* (ZF-T-THP).

Since we do not have to optimize the ordering for WF-T-THP, the complexity is significantly lower than for WF-ST-THP, but the optimization of the latency time is still necessary.

D. Wiener Transmit Filter (TxWF)

Obviously, the linear TxWF can be found, when none of the feedback filters is active, that is we have to add the constraints

$$\boldsymbol{F} = \mathbf{0}_{K\times K} \quad \text{and} \quad \boldsymbol{T}[n] = \mathbf{0}_{K\times K}\delta[n]$$

to the WF-ST-THP optimization to find the TxWF (cf. [9]).

The TxWF has lower complexity than all THP approaches, as no ordering optimization is necessary and the latency time optimization is very simple.

V. COMPARISON OF SPECIAL CASES

To see the differences of the ST-THP variants, we concentrate on the zero-forcing types, since they are more illustrative. The conclusions also apply to the Wiener types, the main difference is that most zero entries are replaced by entries which only converge to zero for high SNR. To understand the zero-forcing approaches, we investigate the coefficients of the combination of the scalar weighting β^{-1} at the receivers, the channels which are collected in the sorted matrix filter $\boldsymbol{H}[n] = [\boldsymbol{h}_{b_1}[n], \ldots, \boldsymbol{h}_{b_K}[n]]^{\mathrm{T}} \in \mathbb{C}^{K\times N_{\mathrm{a}}}$, and the feedforward filter $\boldsymbol{P}[n]$:

$$\boldsymbol{A}[n] = \beta^{-1}\boldsymbol{H}[n] * \boldsymbol{P}[n] = \sum_{i=0}^{L+Q} \boldsymbol{A}_i\delta[n-i].$$

When putting the $L+Q+1$ coefficients of $\boldsymbol{A}[n]$ into one matrix, we get:

$$\begin{aligned}
\text{TxZF:}\quad & \big[\mathbf{0}_{K\times K}, \ldots, \mathbf{0}_{K\times K}, \mathbf{1}_K, \mathbf{0}_{K\times K}, \ldots, \mathbf{0}_{K\times K}\big],\\
\text{ZF-S-THP:}\quad & \big[\mathbf{0}_{K\times K}, \ldots, \mathbf{0}_{K\times K}, \boldsymbol{L}, \mathbf{0}_{K\times K}, \ldots, \mathbf{0}_{K\times K}\big],\\
\text{ZF-T-THP:}\quad & \big[\mathbf{0}_{K\times K}, \ldots, \mathbf{0}_{K\times K}, \mathbf{1}_K, \boldsymbol{A}_{\nu+1}, \ldots, \boldsymbol{A}_{L+Q}\big],\\
\text{ZF-ST-THP:}\quad & \big[\mathbf{0}_{K\times K}, \ldots, \mathbf{0}_{K\times K}, \boldsymbol{L}, \boldsymbol{A}_{\nu+1}, \ldots, \boldsymbol{A}_{L+Q}\big].
\end{aligned}$$

The matrix $\boldsymbol{L}$ is lower triangular with ones on the main diagonal, whereas $\boldsymbol{A}_{\nu+1}, \ldots, \boldsymbol{A}_{L+Q}$ can have any structure. We can see that $\boldsymbol{A}_i = \mathbf{0}_{K\times K}$ for $i = 0, \ldots, \nu-1$, i. e. the interference due to the *preceding* symbols is suppressed by the feedforward filter $\boldsymbol{P}[n]$ for all approaches. Since $\boldsymbol{A}_i = \mathbf{0}_{K\times K}$ for $i > \nu+1$, the feedforward filter $\boldsymbol{P}[n]$ also suppresses the interference of the *succeeding* symbols for the TxZF and ZF-S-THP, but this interference is removed by the feedback filter $\boldsymbol{T}[n]$ for ZF-T-THP and ZF-ST-THP. Similarly, the *multi-user interference* (MUI) is suppressed by the feedforward filter $\boldsymbol{P}[n]$ for the TxZF and ZF-T-THP, whereas the MUI is partly removed by the feedback filter $\boldsymbol{F}$ for ZF-S-THP and ZF-ST-THP.

We can conclude that the feedforward filter $\boldsymbol{P}[n]$ has to suppress less interference, when more feedback is included in the precoder. As a consequence, $\boldsymbol{P}[n]$ can reach a lower MSE or BER for increasing feedback.

VI. SIMULATION RESULTS

We investigate a system with $N_{\mathrm{a}} = 4$ antenna elements at the transmitter and $K = 3$ receivers. The channel *power delay profile* is Vehicular A (Case II, [24]) with $Q+1 = 6$ Rayleigh fading paths and the CSI is known perfectly to the BS. The feedforward filters have order $L = 15$ for the linear transmit filters (TxZF and TxWF) and S-THP to ensure the existence of the zero-forcing variants. Contrary, the feedforward filters have order $L = 7$ for T-THP and ST-THP to end up with a fair comparison (T-THP and ST-THP include the temporal feedback filter $\boldsymbol{T}[n]$). The uncoded BER results are the mean of 30000 channel realizations with 100 symbols per user and channel realization. The SNR is defined as the ratio of E_{S} over N_0, that is the mean transmitted energy per symbol and user over the noise spectral density per receiver.

In Fig. 3, we observe that ZF-S-THP outperforms the linear TxZF only for very high SNR although ZF-S-THP needs an ordering algorithm of high complexity. As THP does not suffer from error propagation like DFE, we must conclude that the modulo operation at the receiver deteriorates the results in this case because of the infinite continuation of the QPSK constellation. Contrary, the ZF-T-THP and ZF-ST-THP clearly beat the linear TxZF although the order of the feedforward filter for ZF-T-THP and ZF-ST-THP is half of the order of

the TxZF. Note that ZF-ST-THP only has little additional performance gain compared to ZF-T-THP.

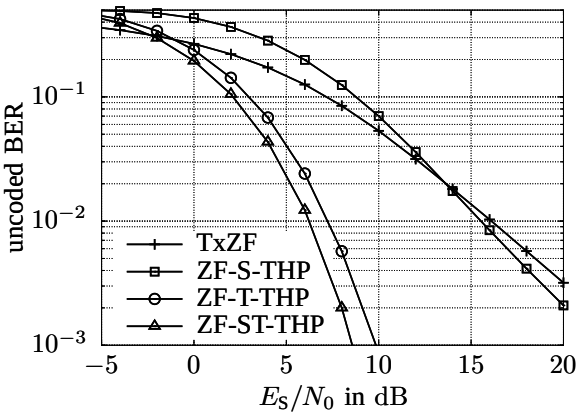

Fig. 3. Comparison of Zero-Forcing Precoders for QPSK Transmission

When employing the Wiener types of precoders for QPSK transmission, we obtain the results depicted in Fig. 4. We see that the best zero-forcing precoder ZF-ST-THP (cf. Fig. 3) is worse than WF-T-THP but nevertheless has a comparable performance as the Wiener precoder types. Surprisingly, the TxWF leads to the smallest uncoded BER below an SNR of about $2\,\mathrm{dB}$. This disadvantage of the THP transmitters is again due to the infinite extension of the constellation set caused by the modulo operations. Note that this introduction of additional neighbors especially deteriorates the BER results for QPSK transmission. However, we also have to take into account that the WF-T-THP and WF-ST-THP have a feedforward filter with the half order of the TxWF filter order. Moreover, we

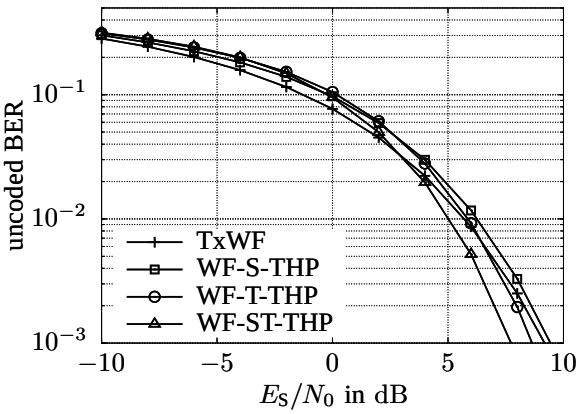

Fig. 4. Comparison of Wiener Precoders for QPSK Transmission

investigate the correctness of our assumption concerning the power of the precoded symbols $v_k[n]$. Recall that we assumed $\sigma_{\boldsymbol{v}}^2 = \tau^2/6$ Accordingly, we define the *transmit power loss* (TPL) as follows:

$$\mathrm{TPL} = -10\log_{10}\left(\frac{6}{KN\tau^2}\sum_{n=0}^{N-1}\|\boldsymbol{v}[n]\|_2^2\right)\,\mathrm{dB},$$

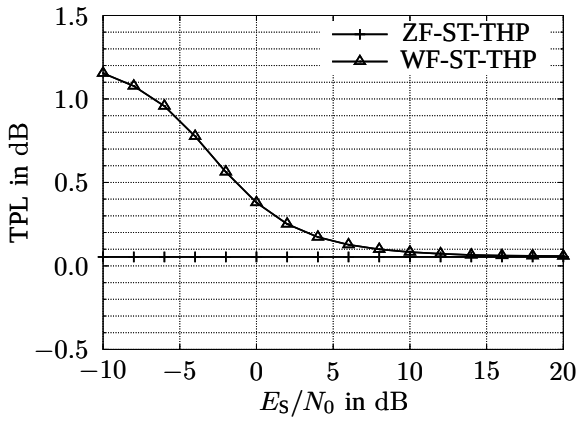

Fig. 5. QPSK Transmission, Exponential Power Delay Profile

where $N = 100$ denotes the number of transmitted symbols of one block, $K = 3$ is the number of receivers, and $\tau = 2\sqrt{2}$ for QPSK. Fig. 5 shows the transmit power loss for ZF-ST-THP and WF-ST-THP. We can conclude that for ZF-THP, $\sigma_{\boldsymbol{v}}^2 = \tau^2/6$ represents a good approximation of the precoder output symbol power, since the power loss is around $0.25\,\mathrm{dB}$. For WF-ST-THP, the transmit power loss is much more significant, especially for low SNR values. It can be shown that for low SNR values, the average power of the feedback terms is small compared to the average data symbol power $\sigma_{s_k}^2$. As a result, for decreasing SNR the assumption of a uniform distribution of the modulo outputs over $\mathbb{M}$ is rendered more and more imprecise. In the limit SNR$\rightarrow 0$, the data symbols pass the precoder unchanged. We can conclude that if QPSK modulation is employed, the performance of Wiener THP is clearly degraded by the transmit power loss in low SNR regions.

In Fig. 6, we present the results for the zero-forcing precoders, when 16QAM is transmitted. Like for QPSK transmission (cf. Fig. 3), we observe a dramatic improvement of the uncoded BER performance, when switching from the TxZF or ZF-S-THP to ZF-T-THP or ZF-ST-THP. Moreover, ZF-T-THP is very close to ZF-ST-THP, whereas ZF-S-THP cannot reach the ZF-ST-THP performance.

The uncoded BER results for 16QAM transmission with Wiener precoders are shown in Fig. 7. Interestingly, the linear TxWF is outperformed by all Wiener THP approaches for the whole depicted SNR range, because the additional neighbors introduced by the modulo operations do not have such an impact as for QPSK (cf. Fig. 4). Like for ZF-THP, WF-T-THP is close to WF-ST-THP which has much higher complexity due to the ordering algorithm. Thus, T-THP is a good alternative for ST-THP, since both lead to comparable results, but T-THP has significantly lower complexity. Note that for 16QAM, the ZF-T-THP and ZF-ST-THP nearly lead to the same uncoded BER performances as the WF-T-THP and WF-ST-THP, respectively.

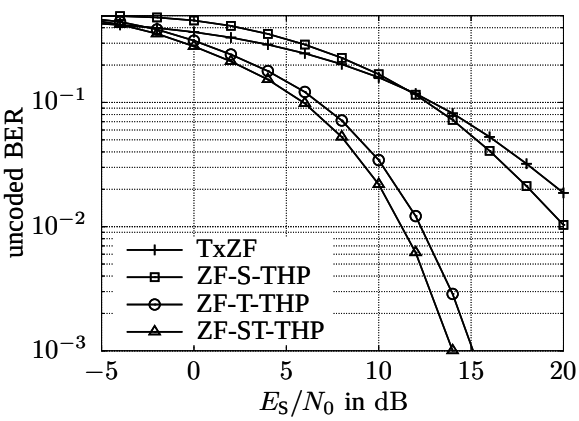

Fig. 6. Comparison of Zero-Forcing Precoders for 16QAM Transmission

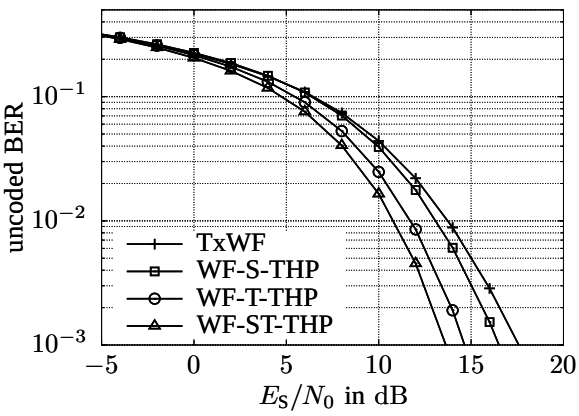

Fig. 7. Comparison of Wiener Precoders for 16QAM Transmission

VII. Conclusions

We have presented a WF-ST-THP optimization, which does not only optimize the filter coefficients but also the latency time and the ordering. Since the computation of the optimum latency time and the optimum ordering has prohibitive complexity, we proposed a suboptimum approach. We have discussed special cases of WF-ST-THP, namely the ZF-ST-THP, WF-S-THP, and WF-T-THP. The simulation results revealed that most of the performance gain of WF-ST-THP is due to the temporal THP component. Therefore, WF-T-THP is a good low complexity alternative for WF-ST-THP.

References

[1] M. R. Gibbard and A. B. Sesay, "Asymmetric Signal Processing for Indoor Wireless LAN's," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 6, pp. 2053–2064, November 1999.

[2] G. Montalbano and D. T. M. Slock, "Spatio-Temporal Array Processing for Matched Filter Bound Optimization in SDMA Downlink Transmission," in *Proc. ISSSE'98*, September 1998, pp. 416–421.

[3] B. R. Vojčić and W. M. Jang, "Transmitter Precoding in Synchronous Multiuser Communications," *IEEE Transactions on Communications*, vol. 46, no. 10, pp. 1346–1355, October 1998.

[4] 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*, vol. 3, September 1999, pp. 997–1001.

[5] P. W. Baier, M. Meurer, T. Weber, and H. Tröger, "Joint Transmission (JT), an alternative rationale for the downlink of Time Division CDMA using multi-element transmit antennas," in *Proc. ISSSTA 2000*, vol. 1, September 2000, pp. 1–5.

[6] M. Joham and W. Utschick, "Downlink Processing for Mitigation of Intracell Interference in DS-CDMA Systems," in *Proc. ISSSTA 2000*, vol. 1, September 2000, pp. 15–19.

[7] F. Kowalewski and P. Mangold, "Joint Predistortion and Transmit Diversity," in *Proc. Globecom '00*, vol. 1, November 2000, pp. 245–249.

[8] A. Noll Barreto and G. Fettweis, "Capacity Increase in the Downlink of Spread Spectrum Systems through Joint Signal Precoding," in *Proc. ICC 2001*, vol. 4, June 2001, pp. 1142–1146.

[9] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems," in *Proc. ISSSTA 2002*, vol. 1, September 2002, pp. 9–13.

[10] R. L. Choi and R. D. Murch, "Transmit MMSE Pre-Rake Pre-processing with Simplified Receivers for the Downlink of MISO TDD-CDMA Systems," in *Proc. Globecom 2002*, vol. 1, November 2002, pp. 429–433.

[11] M. E. Austin, "Decision-feedback equalization for digital communication over dispersive channels," M.I.T. Lincoln Laboratory, Technical report 437, August 1967.

[12] P. Monsen, "Feedback Equalization for Fading Dispersive Channels," *IEEE Transactions on Information Theory*, vol. IT-17, no. 1, pp. 56–64, January 1971.

[13] 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 98*, September 1998, pp. 295–300.

[14] N. Al-Dhahir and A. H. Sayed, "The Finite-Length Multi-Input Multi-Output MMSE-DFE," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2921–2936, October 2000.

[15] A. Lozano and C. Papadias, "Layered Space-Time Receivers for Frequency-Selective Wireless Channels," *IEEE Transactions on Communications*, vol. 50, no. 1, pp. 65–73, Jan. 2002.

[16] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Letters*, vol. 7, no. 5/6, pp. 138–139, March 1971.

[17] H. Harashima and H. Miyakawa, "Matched-Transmission Technique for Channels With Intersymbol Interference," *IEEE Transactions on Communications*, vol. 20, no. 4, pp. 774–780, August 1972.

[18] G. D. Forney and M. V. Eyuboğlu, "Combined Equalization and Coding Using Precoding," *IEEE Communications Magazine*, vol. 29, no. 12, pp. 25–34, December 1991.

[19] G. Ginis and J. M. Cioffi, "A Multi-user Precoding Scheme achieving Crosstalk Cancellation with Application to DSL Systems," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, vol. 2, October 2000, pp. 1627–1631.

[20] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-Time Transmission using Tomlinson-Harashima Precoding," in *Proc. 4th ITG Conference on Source and Channel Coding*, January 2002, pp. 139–147.

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

[22] R. F. H. Fischer, C. Stierstorfer, and J. B. Huber, "Precoding for Point-to-Multipoint Transmission over MIMO ISI Channels," February 2004, *Accepted for presentation at 2004 Zurich Seminar on Communications*.

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

[24] 3GPP-3GPP2 Spatial Channel Ad-hoc group, "Spatial Channel Model Text Describtion," 3rd Generation Partnership Project (3GPP), www.3gpp.org, SCM 134, April 2003.