COMPLEXITY REDUCTION FOR MMSE MULTIUSER SPATIO-TEMPORAL TOMLINSON-HARASHIMA PRECODING

David Schmidt, Michael Joham, Frank A. Dietrich,
Katsutoshi Kusume, and Wolfgang Utschick

9th International ITG/IEEE Workshop on Smart Antennas (WSA 2005)
Duisburg, Germany

April 4th–5th, 2005

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

COMPLEXITY REDUCTION FOR MMSE MULTIUSER SPATIO-TEMPORAL TOMLINSON-HARASHIMA PRECODING

David Schmidt∗*, Michael Joham*∗*, Frank A. Dietrich*∗*, Katsutoshi Kusume*†*, and Wolfgang Utschick*∗

∗Munich University of Technology, 80333 Munich, Germany, Email: Michael.Joham@nws.ei.tum.de
†DoCoMo Communications Laboratories Europe GmbH, 80687 Munich, Germany

ABSTRACT

We derive the *minimum mean square error* solution to *Tomlinson-Harashima precoding* in frequency selective multiuser scenarios with a centralized multi-antenna transmitter. This solution depends on the ordering of the precoded symbols as well as the delay between transmission and detection of a symbol (latency time). We present an algorithm for jointly optimizing these two parameters in a computational complexity that grows with the third power—instead of the fourth power—of the system parameters. In the course of this complexity reduction, we prove that latency times less than the order of the FIR feedforward filter do not need to be considered. We then use simulations to show that optimization of the latency time can be omitted altogether without performance degradation for most practical channel models.

1. INTRODUCTION

Most wireless communications scenarios feature a *base station* (BS) with high computational power and *mobile stations* (MSs) which should be inexpensive and consume as little power as possible, mandating that the computationally complex channel equalization is performed at the BS in the downlink as well as the uplink [1]. Pre-equalization for the downlink can be performed linearly (e. g. [2, 3, 4, 5, 6, 7]) or nonlinearly. In this work we focus on *Tomlinson-Harashima precoding* (THP), a nonlinear extension of linear precoding schemes, which is computationally much less expensive than more general nonlinear techniques, such as those proposed in [8, 9, 10]. THP requires full channel state information; in [11], the influence of imperfect channel state information was investigated and robust THP introduced.

THP was originally proposed to combat intersymbol interference in frequency selective channels [12, 13] by adding a feedback filter at the transmitter that cancels the temporal interference of already transmitted symbols, and a nonlinear modulo operation at both transmitter and receiver, in order to limit the signal amplitude. However, THP has also been applied to the problem of multiuser separation in frequency flat *multiple input multiple output* (MIMO) channels, e. g. in [14, 15, 16, 17]. Here the symbols for the different users are precoded in a certain order; the spatial interference of an already precoded symbol is removed by the feedback loop for all successively precoded symbols. Obviously, the performance of spatial THP depends on the order in which the symbols are precoded.

In [17, 18], these approaches were combined and *spatio-temporal Tomlinson-Harashima precoding* (ST-THP) was derived. In contrast to [19], where an IIR feedforward filter was obtained, all filters were assumed to be FIR. The use of adaptive IIR filters entails several implementation issues, such as quantization limit cycles due to the finite word length of digital signal processors, and is therefore not applied in this paper either. The order in which the symbols are precoded affects the performance of ST-THP, as does the latency time, i. e. the delay between transmission and detection of a symbol. It was noted in [17, 18] that jointly optimizing latency time and ordering is computationally infeasible. Instead, a suboptimum method was proposed, which first finds the best latency time for the assumption of inactive spatial feedback filter and then optimizes the ordering.

In this paper, we will present several approaches to overcome the prohibitive complexity of joint latency time and ordering optimization. In Sections 2 and 3 we will introduce the system model and derive the *Wiener Filter* (WF) solution for ST-THP, respectively. In the next two sections, we will present ways of reducing the complexity order of ordering optimization and latency time optimization. Finally, in Section 6, we will use simulations to show that in realistic scenarios latency time optimization can be omitted altogether, without a degradation of performance.

1.1. Notation

Throughout the paper, we will denote vectors and matrices by lower case bold and upper case bold letters, respectively. We use $\mathrm{E}[\bullet]$, '$*$', '$\otimes$', $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, $\mathrm{tr}(\bullet)$, and $\mathrm{Re}(\bullet)$ for expectation, convolution, the Kronecker product, transposition, conjugate transposition, the trace of a matrix, and the real part, respectively. 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 $\mathbf{1}_N$. $\delta[n]$ is the unit impulse function, which evaluates to one for $n = 0$ and to zero otherwise. All random sequences are assumed to be zero-mean and stationary. When describing algorithms, we use the notation $\boldsymbol{A}(i, j)$ for the element in row i and column j of matrix $\boldsymbol{A}$ and $\boldsymbol{A}(k : l, m : n)$ for the block consisting of rows k to l of the columns m to n of matrix $\boldsymbol{A}$.

2. SYSTEM MODEL

We consider a system with B non-cooperative single antenna receivers, N_a antennas at the transmitter, and an FIR-channel of order Q with the impulse response

$$\boldsymbol{H}[n] = \sum_{q=0}^{Q} \boldsymbol{H}_q \delta[n-q] \in \mathbb{C}^{B \times N_\mathrm{a}},$$

as depicted in Fig. 1.

The B data streams are collected in the vector signal

$$\boldsymbol{s}[n] = [s_1[n], \ldots, s_B[n]]^\mathrm{T} \in \mathbb{C}^B,$$

which is reordered by the permutation matrix

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

where the vector $\boldsymbol{e}_i$ is the i-th column of the identity matrix $\mathbf{1}_B$. Here we introduced the B-tupel $\mathcal{O} = (b_1, b_2, \ldots, b_B)$, with $b_i \in \{1, \ldots, B\} \setminus \{b_1, \ldots, b_{i-1}\}$. The data symbol for the b_1-th user is precoded first, the data symbol for the b_B-th user last. Note that $\boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}} \boldsymbol{\Pi}^{(\mathcal{O})} = \mathbf{1}_B$.

The precoder, which processes the permuted data signal, consists of a spatial feedback filter, a temporal feedback filter, and a feedforward filter. The FIR feedforward filter of order L has the impulse response

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

In order to be realizable without zero-delay feedback loops, the spatial feedback filter $\boldsymbol{F} \in \mathbb{C}^{B \times B}$ must have lower triangular structure with zero main diagonal. The FIR temporal feedback filter

$$\boldsymbol{T}[n] = \sum_{j=1}^{Q+L-\nu} \boldsymbol{T}_j \delta[n-j] \in \mathbb{C}^{B \times B} \qquad (1)$$

is strictly causal and needs $Q + L - \nu$ coefficients to remove temporal interference following the detection of the symbol at the receivers. Here, ν is the latency time, i. e. the delay between transmission and detection of a symbol.

In order to be able to analyze the system, we replace the modulo operators $\mathrm{M}(\bullet)$ with the summation of auxiliary signals $\boldsymbol{a}[n]$ and $-\tilde{\boldsymbol{a}}[n]$, as can be seen in Fig. 2. For a detailed description of the modulo operator as well as its linear representation, see [17, 18].

While the statistics of the signal $\boldsymbol{d}[n]$ are unknown, we can assume that the modulo operator outputs are temporally and spatially uncorrelated, i. e.

$$\mathrm{E}\left[\boldsymbol{v}[n]\boldsymbol{v}^\mathrm{H}[n+k]\right] = \sigma_{\boldsymbol{v}}^2 \mathbf{1}_B \delta[k], \qquad (2)$$

(cf. [20, Theorem 3.1]).

The additive noise at the receivers has the spatial covariance matrix $\mathrm{E}\left[\boldsymbol{\eta}[n]\boldsymbol{\eta}^\mathrm{H}[n]\right] = \boldsymbol{R}_{\boldsymbol{\eta}}$.

3. MMSE FILTER SOLUTION

Using (1) we can express the desired signal as

$$\boldsymbol{d}[n] = \boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}} \left((\mathbf{1}_B - \boldsymbol{F}) \boldsymbol{v}[n] - \sum_{j=1}^{Q+L-\nu} \boldsymbol{T}_j \boldsymbol{v}[n-j] \right). \qquad (3)$$

The estimate of the desired signal at the receivers calculates to

$$\begin{aligned} \tilde{\boldsymbol{d}}[n] &= \beta^{-1} \boldsymbol{H}[n] * \boldsymbol{P}[n] * \boldsymbol{v}[n] + \beta^{-1} \boldsymbol{\eta}[n] \\ &= \beta^{-1} \boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}} \sum_{j=0}^{Q+L} \boldsymbol{S}^{(j)} \tilde{\boldsymbol{H}}^{(\mathcal{O})} \boldsymbol{P} \boldsymbol{v}[n-j] + \beta^{-1} \boldsymbol{\eta}[n], \end{aligned} \qquad (4)$$

where we defined

$$\begin{aligned} \tilde{\boldsymbol{H}}^{(\mathcal{O})} &= \sum_{q=0}^{Q} \boldsymbol{S}_{(q,L+1,Q)}^\mathrm{T} \otimes (\boldsymbol{\Pi}^{(\mathcal{O})} \boldsymbol{H}_q), \\ \boldsymbol{P} &= \left[\boldsymbol{P}_0^\mathrm{T}, \ldots, \boldsymbol{P}_L^\mathrm{T}\right]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{a}(L+1) \times B}, \\ \boldsymbol{S}_{(q,L+1,Q)} &= [\mathbf{0}_{L+1 \times q}, \mathbf{1}_{L+1}, \mathbf{0}_{L+1 \times Q-q}] \\ &\qquad \in \{0,1\}^{L+1 \times Q+L+1}, \\ \boldsymbol{S}^{(j)} &= \boldsymbol{e}_{j+1}^\mathrm{T} \otimes \mathbf{1}_B \in \{0,1\}^{B \times B(Q+L+1)}, \end{aligned}$$

and $\boldsymbol{e}_{j+1}$ is the $j+1$-th column of the identity matrix $\mathbf{1}_{Q+L+1}$. $\tilde{\boldsymbol{H}}^{(\mathcal{O})}$ is a $B(Q+L+1) \times N_\mathrm{a}(L+1)$ matrix with block Toeplitz structure, which contains the permuted coefficients of the channel impulse response. Note that even though the permutation is not necessary at this point, this definition of the channel matrix will prove useful later on. $\boldsymbol{P}$ is obtained by stacking the coefficients of $\boldsymbol{P}[n]$. The selection matrix $\boldsymbol{S}^{(j)}$, when multiplied from the left, returns the $Bj+1$-th to $B(j+1)$-th rows of a matrix with $B(Q+L+1)$ rows.

Consequently, we can write the *mean square error* (MSE) of the estimate of the desired signal $\tilde{\boldsymbol{d}}[n]$ with regard

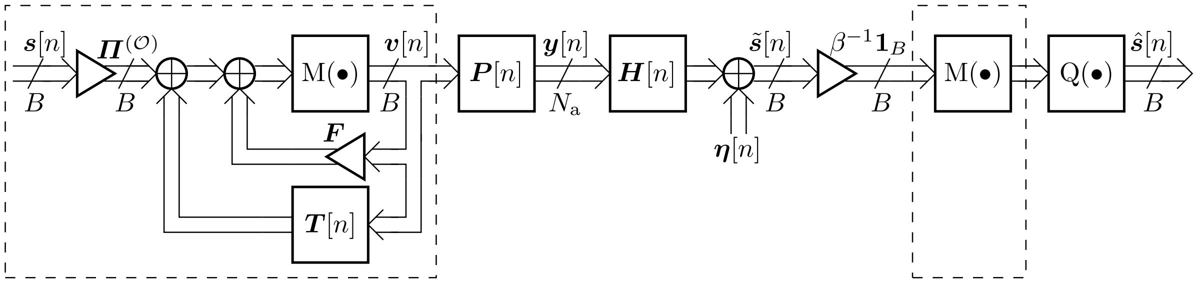


Fig. 1. THP Transmission over Time Dispersive MIMO Channels

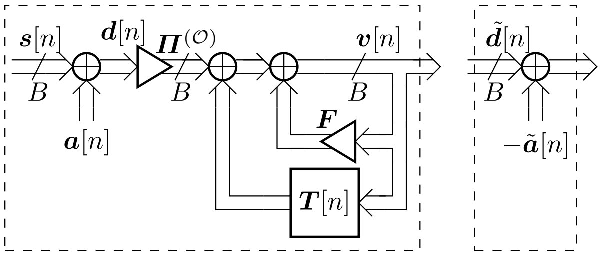


Fig. 2. Linear Representation of the Modulo Operators

to the delayed desired signal $\boldsymbol{d}[n-\nu]$ for a given ordering $\mathcal{O}$ and latency time ν as

$$\begin{aligned}\varepsilon^{(\mathcal{O})}(\nu) &= \mathrm{E}\left[\|\boldsymbol{d}[n-\nu]-\tilde{\boldsymbol{d}}[n]\|_2^2\right]\\ &= \sigma_{\boldsymbol{v}}^2\bigg(B+\mathrm{tr}\big(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big)+\sum_{j=1}^{Q+L-\nu}\mathrm{tr}\big(\boldsymbol{T}_j\boldsymbol{T}_j^{\mathrm{H}}\big)\\ &\quad -2\beta^{-1}\,\mathrm{Re}\big(\mathrm{tr}\big(\boldsymbol{S}^{(\nu)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\boldsymbol{P}(\mathbf{1}_B-\boldsymbol{F})^{\mathrm{H}}\big)\big)\\ &\quad +2\beta^{-1}\sum_{j=1}^{Q+L-\nu}\mathrm{Re}\big(\mathrm{tr}\big(\boldsymbol{S}^{(\nu+j)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\boldsymbol{P}\boldsymbol{T}_j^{\mathrm{H}}\big)\big)\\ &\quad +\beta^{-2}\,\mathrm{tr}\big(\tilde{\boldsymbol{H}}^{(\mathcal{O})}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\big)\bigg)+\beta^{-2}\,\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}}),\end{aligned}$$

where we used (2), (3), and (4).

Due to (2), the average transmit power evaluates to

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

The *minimum mean square error* (MMSE) filter solution minimizes the MSE for a given average transmit power E_{tr}. The lower triangular, zero main diagonal structure of the spatial feedback filter $\boldsymbol{F}$ is an additional constraint in our optimization:

$$\begin{aligned}\{\boldsymbol{P}_{\mathrm{WF}},\boldsymbol{F}_{\mathrm{WF}},\boldsymbol{T}_{\mathrm{WF},1},\ldots,\beta_{\mathrm{WF}}\} &= \underset{\{\boldsymbol{P},\boldsymbol{F},\boldsymbol{T}_1,\ldots,\beta\}}{\arg\min}\ \varepsilon^{(\mathcal{O})}(\nu)\\ \text{s.t.:}\quad \sigma_{\boldsymbol{v}}^2\,\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right)=E_{\mathrm{tr}} &\quad\text{and}\\ \boldsymbol{S}_i\boldsymbol{F}\boldsymbol{e}_i=\mathbf{0}_i, &\quad i=1,\ldots,B,\end{aligned}\tag{5}$$

with

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

Note that the selection matrix $\boldsymbol{S}_i$, when multiplied from the left, returns the first i out of B rows of a matrix or vector.

The solution to (5) can be found with the method of Lagrangian multipliers and, after applying the matrix inversion lemma (e. g. [21, Section 2.9]), reads as

$$\begin{aligned}\boldsymbol{P}_{\mathrm{WF}} &= \beta_{\mathrm{WF}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\sum_{i=1}^{B}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}\boldsymbol{C}_{(\nu,i)}^{(\mathcal{O}),-1}\boldsymbol{S}_{(\nu,i)}\boldsymbol{e}_{(\nu,i)}\boldsymbol{e}_i^{\mathrm{T}},\\ \boldsymbol{F}_{\mathrm{WF}} &= -\beta_{\mathrm{WF}}^{-1}\sum_{i=1}^{B}\sum_{k=i+1}^{B}\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{S}^{(\nu)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\boldsymbol{P}_{\mathrm{WF}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}},\ \text{and}\\ \boldsymbol{T}_{\mathrm{WF},j} &= -\beta_{\mathrm{WF}}^{-1}\boldsymbol{S}^{(\nu+j)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\boldsymbol{P}_{\mathrm{WF}},\end{aligned}\tag{6}$$

with

$$\begin{aligned}\boldsymbol{C}_{(\nu,i)}^{(\mathcal{O})} &= \boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}+\xi_{\mathrm{WF}}\mathbf{1}_{B\nu+i}\\ &\qquad\in\mathbb{C}^{B\nu+i\times B\nu+i},\\ \boldsymbol{S}_{(\nu,i)} &= \left[\mathbf{1}_{B\nu+i},\mathbf{0}_{B\nu+i\times B(Q+L+1-\nu)-i}\right]\\ &\qquad\in\{0,1\}^{B\nu+i\times B(Q+L+1)},\\ \boldsymbol{e}_{(\nu,i)} &= \boldsymbol{e}_{\nu+1}\otimes\boldsymbol{e}_i\in\{0,1\}^{B(Q+L+1)},\quad\text{and}\\ \xi_{\mathrm{WF}} &= \frac{\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}})}{E_{\mathrm{tr}}}.\end{aligned}\tag{7}$$

The selection matrix $\boldsymbol{S}_{(\nu,i)}$, when multiplied from the left, selects the first $B\nu+i$ out of $B(Q+L+1)$ rows of a matrix or vector. The gain factor β_{WF} is chosen so that the transmit power constraint is fulfilled. We can see that computing the filter solution according to (6) involves inverting the B matrices $\boldsymbol{C}_{(\nu,i)}^{(\mathcal{O})}$ with $i=1,\ldots,B$.

When the MMSE filter solution is employed, the MSE simplifies to

$$\varepsilon^{(\mathcal{O})}(\nu)=\sigma_{\boldsymbol{v}}^2\xi_{\mathrm{WF}}\sum_{i=1}^{B}\boldsymbol{e}_{(\nu,i)}^{\mathrm{T}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}\boldsymbol{C}_{(\nu,i)}^{(\mathcal{O}),-1}\boldsymbol{S}_{(\nu,i)}\boldsymbol{e}_{(\nu,i)}.\tag{8}$$

Note that each summand is the bottom right element of the inverse of $\boldsymbol{C}_{(\nu,i)}^{(\mathcal{O})}$.

Table 1. Standard Computation of Optimum Permutation Matrix $\boldsymbol{\Pi}^{(\mathcal{O})}$

$\boldsymbol{\Pi}^{(\mathcal{O})} \leftarrow \mathbf{1}_B$
$\varepsilon^{(\mathcal{O})}(\nu) \leftarrow 0$
for $i = B, \ldots, 1$:
$\quad \boldsymbol{G} \leftarrow$ bottom right $i \times i$-block of $\boldsymbol{C}^{(\mathcal{O}),-1}_{(\nu,i)}$
$\quad q \leftarrow \underset{q' \in \{1,\ldots,i\}}{\arg\min}\ \boldsymbol{G}(q',q')$
$\quad \boldsymbol{\Pi}_i \leftarrow \mathbf{1}_B$ with rows q and i exchanged
$\quad \boldsymbol{\Pi}^{(\mathcal{O})} \leftarrow \boldsymbol{\Pi}_i \boldsymbol{\Pi}^{(\mathcal{O})}$
$\quad \varepsilon^{(\mathcal{O})}(\nu) \leftarrow \varepsilon^{(\mathcal{O})}(\nu) + \sigma_{\boldsymbol{v}}^2 \xi_{\mathrm{WF}} \boldsymbol{G}(q,q)$

4. EFFICIENT ORDERING COMPUTATION

Computation of the optimum ordering $\mathcal{O}$ for a given latency time ν would require trying out all $B!$ possibilities for $\mathcal{O}$ and choosing the one with the lowest MSE. Since this quickly becomes too complex with an increasing number of users, the standard suboptimum approach is to successively minimize the summands of $\varepsilon^{(\mathcal{O})}(\nu)$, starting with the contribution of the data stream precoded last, i. e. the one with index b_B, since it does not depend on the ordering of the previously precoded symbols. We will refer to the result of this procedure as 'optimum' for the remainder of the paper. The algorithm involves inverting the matrices $\boldsymbol{C}^{(\mathcal{O})}_{(\nu,i)}$ and symmetrically permuting the result, so that the lowest diagonal element is in the bottom right position, for $i = B, \ldots, 1$ (cf. Table 1). The resulting complexity is $\mathrm{O}\left(B^4\nu^3\right)$, or $\mathrm{O}\left(B^4(Q+L)^4\right)$ in combination with trying out all possible latency times $\nu \in \{0, \ldots, Q+L\}$.

Now let us assume a given ordering $\mathcal{O}$ and latency time ν. Furthermore, assume that the Cholesky factorization of

$$\boldsymbol{C}^{(\mathcal{O}),-1}_{(\nu,B)} = \boldsymbol{L}^{\mathrm{H}} \boldsymbol{D} \boldsymbol{L} \tag{9}$$

is known, where $\boldsymbol{L}$ is lower triangular with unit main diagonal and $\boldsymbol{D}$ is diagonal with real-valued, positive entries. Note that we decomposed into the product of an upper triangular and a lower triangular matrix, instead of a lower and an upper triangular matrix.

The MSE in (8) then simplifies to

$$\varepsilon^{(\mathcal{O})}(\nu) = \sigma_{\boldsymbol{v}}^2 \xi_{\mathrm{WF}} \sum_{i=1}^{B} d_{B\nu+i,B\nu+i},$$

where $d_{B\nu+i,B\nu+i}$ is the $B\nu+i$-th diagonal element of $\boldsymbol{D}$.[1]

[1] This result is obtained by plugging (9) into (8) and making use of the properties of triangular matrices to simplify the result, among them the fact that inverting an upper left square block of a triangular matrix is equivalent to taking the upper left block of the inverse, as well as the fact that the inverse of a triangular matrix with unit main diagonal is itself triangular with unit main diagonal.

Table 2. Efficient Computation of Optimum Permutation Matrix $\boldsymbol{\Pi}^{(\mathcal{O})}$

$\boldsymbol{\Pi}^{(\mathcal{O})} \leftarrow \mathbf{1}_B$
$\boldsymbol{G} \leftarrow$ bottom right $B \times B$-block of $\boldsymbol{C}^{(\mathcal{O}),-1}_{(\nu,B)}$ for $\boldsymbol{\Pi}^{(\mathcal{O})} = \mathbf{1}_B$
$\boldsymbol{D} \leftarrow \mathbf{0}_{B\times B}$
for $i = B, \ldots, 1$:
$\quad q \leftarrow \underset{q' \in \{1,\ldots,i\}}{\arg\min}\ \boldsymbol{G}(q',q')$
$\quad \boldsymbol{\Pi}_i \leftarrow \mathbf{1}_B$ with rows q and i exchanged
$\quad \boldsymbol{\Pi}^{(\mathcal{O})} \leftarrow \boldsymbol{\Pi}_i \boldsymbol{\Pi}^{(\mathcal{O})}$
$\quad \boldsymbol{G} \leftarrow \boldsymbol{\Pi}_i \boldsymbol{G} \boldsymbol{\Pi}_i^{\mathrm{T}}$
$\quad \boldsymbol{D}(i,i) \leftarrow \boldsymbol{G}(i,i)$
$\quad \boldsymbol{G}(i,1:i) \leftarrow \boldsymbol{G}(i,1:i)/\boldsymbol{D}(i,i)$
$\quad \boldsymbol{G}(1:i-1,1:i-1) \leftarrow \boldsymbol{G}(1:i-1,1:i-1)-$
$\qquad -\boldsymbol{G}(i,1:i-1)^{\mathrm{H}} \boldsymbol{G}(i,1:i-1) \boldsymbol{D}(i,i)$
$\varepsilon^{(\mathcal{O})}(\nu) \leftarrow \sigma_{\boldsymbol{v}}^2 \xi_{\mathrm{WF}}\, \mathrm{tr}(\boldsymbol{D})$

The idea of the proposed efficient ordering algorithm, which was introduced for spatial THP in [22], is to incorporate the successive minimization of the MSE contributions into the Cholesky factorization algorithm (cf. [23]). Note that since the first factor is to be upper triangular, the factorization must begin with the bottom right element and continue upward and to the left. When an element of $\boldsymbol{D}$ is to be computed, we now insert a symmetric permutation, such that the respective element is minimized. The complete algorithm is shown in Table 2. It takes the matrix $\boldsymbol{C}^{(\mathcal{O})}_{(\nu,B)}$ with no permutation as input and returns the same 'optimum' ordering yielded by the standard procedure in Table 1.

For determining the ordering, only the bottom right $B \times B$ block of the matrix $\boldsymbol{C}^{(\mathcal{O}),-1}_{(\nu,B)}$ is relevant. However, if the Cholesky factorization is computed for the complete matrix, the filter solutions can be significantly simplified as well, in particular

$$\boldsymbol{P}_{\mathrm{WF}} = \beta_{\mathrm{WF}} \tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}} \boldsymbol{S}^{\mathrm{T}}_{(\nu,B)} \boldsymbol{L}^{\mathrm{H}} \boldsymbol{D} \boldsymbol{S}_{(\nu,B)} \boldsymbol{S}^{(\nu),\mathrm{T}} \quad \text{and}$$

$$\boldsymbol{F}_{\mathrm{WF}} = -\left(\boldsymbol{S}^{(\nu)} \boldsymbol{S}^{\mathrm{T}}_{(\nu,B)} \boldsymbol{L} \boldsymbol{S}_{(\nu,B)} \boldsymbol{S}^{(\nu),\mathrm{T}}\right)^{-1} + \mathbf{1}_B,$$

i. e. only one further inversion of a $B \times B$ triangular matrix is necessary, resulting in a complexity for finding the best ordering and computing the filters for a given latency time ν of $\mathrm{O}\left(B^3\nu^3\right)$. The fact that the lower triangular part of the matrix $\boldsymbol{G}$ returned by the algorithm in Table 2 is the bottom right block of $\boldsymbol{L}$ in (9) can be used to save several operations when factorizing the complete matrix.

4.1. Simulation Results

In order to show the effect of optimized ordering on the performance of ST-THP, we carried out numerical simulations in different scenarios. For the results in Fig. 3 and Fig. 4, we assumed $N_\mathrm{a} = 4$ transmit antennas, configured as a uniform linear array with $\lambda/2$ spacing, and $B = 3$ receivers, located at random angles around the transmitter, with Laplacian angular spread of $10°$. We assumed no temporal correlations, i. e. uncorrelated angles of departure for the different taps. We used ‘Pedestrian A’ and ‘Vehicular A’ power delay profiles (cf. [24]), normalized so that $\sum_{q=0}^{Q} \mathrm{E}\left[\|\boldsymbol{H}_q\|_\mathrm{F}^2\right] = N_\mathrm{a}B$, where $\|\bullet\|_\mathrm{F}$ denotes the Frobenius norm of a matrix. Furthermore, the transmitter had perfect channel state information and employed a feedforward filter of order $L = 4$. 100 16QAM symbols were transmitted to each receiver per channel realization, with optimized ordering (cf. Table 2) and with no ordering, i. e. $\boldsymbol{\Pi}^{(\mathcal{O})} = \mathbf{1}_B$. The bit error rate was obtained by averaging over 10000 channel realizations.

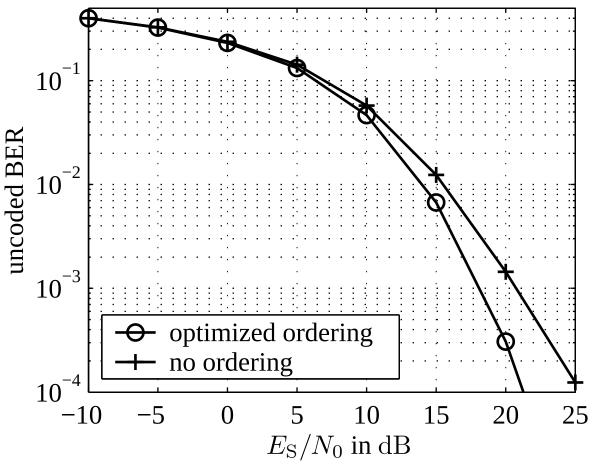


Fig. 3. Channel Order $Q = 3$, ‘Pedestrian A’ Power Delay Profile

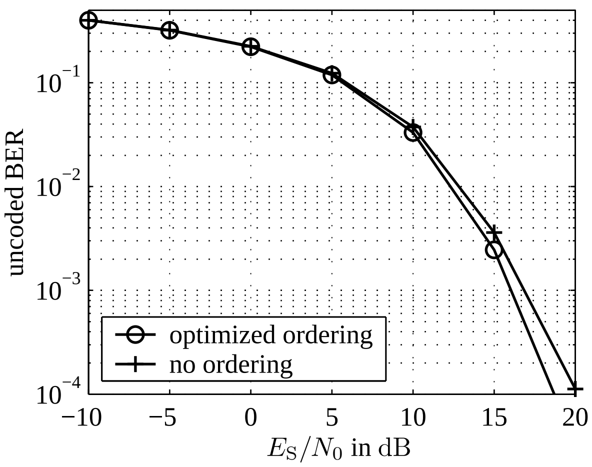


Fig. 4. Channel Order $Q = 5$, ‘Vehicular A’ Power Delay Profile

In Fig. 3 we can measure a gain of the ordering optimization of about $1.5\mathrm{dB}$ at an uncoded bit error rate of 10^{-2}, and about $2.5\mathrm{dB}$ at 10^{-3}. In Fig. 4 the gain is significantly lower. Further simulations have led us to conclude that the performance gain that can be achieved by employing the optimum ordering depends highly on the power delay profile of the channel model. The more average energy the first path of the channel has, the more important the spatial feedback component becomes, which benefits from optimum ordering.

5. EFFICIENT LATENCY TIME OPTIMIZATION

So far, we were able to reduce the complexity order of computing the optimum ordering; however, in order to determine the best combination of latency time and ordering, the MSE with optimum ordering must be calculated for all possible latency times $0 \leq \nu \leq Q + L$.

In this section, we will introduce two approaches for reducing the computational effort of this process. First, we will prove that not all latency times must be considered, as certain ones cannot outperform other latency times in terms of MSE. Second, we will show how the block matrix inversion properties can be exploited, making it unnecessary to perform a complete matrix inversion for each new latency time.

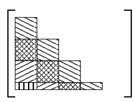
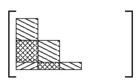
Theorem 1 *For latency times less than or equal to the length of the FIR feedforward filter L and for any given ordering $\mathcal{O}$, the MSE is non-increasing in ν, i. e.*

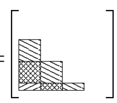
$$\varepsilon^{(\mathcal{O})}(\nu) \leq \varepsilon^{(\mathcal{O})}(\nu - 1), \quad 1 \leq \nu \leq L.$$

Proof. In the following, we always assume that $1 \leq \nu \leq L$. We recall that $\tilde{\boldsymbol{H}}^{(\mathcal{O})}$ has block Toeplitz structure and that $\boldsymbol{S}_{(\nu,i)}$, when multiplied from the left, selects the first $B\nu + i$ out of $B(Q + L + 1)$ rows of a matrix. We furthermore introduce the selection matrix

$$\tilde{\boldsymbol{S}} = [\mathbf{0}_{N_\mathrm{a}L \times N_\mathrm{a}}, \mathbf{1}_{N_\mathrm{a}L}] \in \{0,1\}^{N_\mathrm{a}L \times N_\mathrm{a}(L+1)},$$

which, when transposed and multiplied from the right, cuts off the first N_a out of $N_\mathrm{a}(L + 1)$ columns of a matrix. The structure of $\tilde{\boldsymbol{H}}^{(\mathcal{O})}$ with different selection matrices applied to it can be illustrated as follows:

$\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})} =$ , $\boldsymbol{S}_{(\nu-1,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})} =$ ,

and $\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{S}}^\mathrm{T} =$ .

Obviously, due to the special structure of the channel matrix, the following holds true:

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

and furthermore

$$\left(\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{S}}^{\mathrm{T}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}} + \xi_{\mathrm{WF}}\mathbf{1}_{B\nu+i}\right)^{-1} = \\ = \begin{bmatrix} \xi_{\mathrm{WF}}^{-1}\mathbf{1}_B & \mathbf{0}_{B\times B(\nu-1)+i} \\ \mathbf{0}_{B(\nu-1)+i\times B} & \boldsymbol{C}_{(\nu-1,i)}^{(\mathcal{O}),-1} \end{bmatrix}.$$

Thus, we can express the MSE for latency time $\nu - 1$ as

$$\varepsilon^{(\mathcal{O})}(\nu-1) = \\ = \sigma_{\boldsymbol{v}}^2\xi_{\mathrm{WF}}\sum_{i=1}^{B}\boldsymbol{e}_{(\nu-1,i)}^{\mathrm{T}}\boldsymbol{S}_{(\nu-1,i)}^{\mathrm{T}}\boldsymbol{C}_{(\nu-1,i)}^{(\mathcal{O}),-1}\boldsymbol{S}_{(\nu-1,i)}\boldsymbol{e}_{(\nu-1,i)} \\ = \sigma_{\boldsymbol{v}}^2\xi_{\mathrm{WF}}\sum_{i=1}^{B}\boldsymbol{e}_{(\nu,i)}^{\mathrm{T}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}\Big(\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{S}}^{\mathrm{T}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}} \\ + \xi_{\mathrm{WF}}\mathbf{1}_{B\nu+i}\Big)^{-1}\boldsymbol{S}_{(\nu,i)}\boldsymbol{e}_{(\nu,i)}.$$

Let

$$\mathrm{S}\left(\begin{bmatrix} \boldsymbol{A}' & \boldsymbol{b} \\ \boldsymbol{b}^{\mathrm{H}} & c \end{bmatrix}\right) = c - \boldsymbol{b}^{\mathrm{H}}\boldsymbol{A}'^{,-1}\boldsymbol{b}$$

denote the scalar Schur complement of a Hermitian matrix. It is known (e. g. [21, Section 2.9]) that the bottom right element of the inverse of a Hermitian matrix is the inverse of its scalar Schur complement. We can therefore write

$$\varepsilon^{(\mathcal{O})}(\nu) = \sigma_{\boldsymbol{v}}^2\xi_{\mathrm{WF}} \\ \cdot\sum_{i=1}^{B}\mathrm{S}^{-1}\left(\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}} + \xi_{\mathrm{WF}}\mathbf{1}_{B\nu+i}\right),$$

and

$$\varepsilon^{(\mathcal{O})}(\nu-1) = \sigma_{\boldsymbol{v}}^2\xi_{\mathrm{WF}} \\ \cdot\sum_{i=1}^{B}\mathrm{S}^{-1}\left(\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{S}}^{\mathrm{T}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}} + \xi_{\mathrm{WF}}\mathbf{1}_{B\nu+i}\right).$$

It can easily be seen that the following inequality holds true for every vector $\boldsymbol{x} \in \mathbb{C}^{B\nu+i}$:

$$\|\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}\boldsymbol{x}\|_2^2 \geq \|\tilde{\boldsymbol{S}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}\boldsymbol{x}\|_2^2.$$

Therefore the matrix

$$\boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}} - \boldsymbol{S}_{(\nu,i)}\tilde{\boldsymbol{H}}^{(\mathcal{O})}\tilde{\boldsymbol{S}}^{\mathrm{T}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{H}}^{(\mathcal{O}),\mathrm{H}}\boldsymbol{S}_{(\nu,i)}^{\mathrm{T}}$$

is positive semidefinite.

It has been shown (e. g. [25]) that if the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{A} - \boldsymbol{B}$ are positive semidefinite, then $\mathrm{S}(\boldsymbol{A}) \geq \mathrm{S}(\boldsymbol{B})$. Thus, every summand of $\varepsilon^{(\mathcal{O})}(\nu)$ is less than or equal to the according summand of $\varepsilon^{(\mathcal{O})}(\nu - 1)$. ■

Remark: For the special case of a frequency flat channel the matrix $\tilde{\boldsymbol{H}}^{(\mathcal{O})}$ is block diagonal, and the MSE becomes independent of the latency time. In this case equality holds in Theorem 1.

We can conclude from Theorem 1 that we only need to consider latency times $L \leq \nu \leq Q + L$, since lower latency times cannot perform better. Nonetheless, computing the best ordering as described in Section 4 for each latency time would require $Q + 1$ inversions of $\boldsymbol{C}_{(\nu,B)}^{(\mathcal{O})}$. Here we can make use of the relationship between $\boldsymbol{C}_{(\nu,B)}^{(\mathcal{O})}$ and $\boldsymbol{C}_{(\nu-1,B)}^{(\mathcal{O})}$ discussed in the following.

As can be seen from (7), $\boldsymbol{C}_{(\nu-1,B)}^{(\mathcal{O})}$ is the upper left block of $\boldsymbol{C}_{(\nu,B)}^{(\mathcal{O})}$. According to e. g. [21, Section 2.9], when the inverse of the upper left block of a matrix is known, the inverse of the complete matrix can be computed with several matrix multiplications and the inversion of a matrix the size of the bottom right block:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{H}} & \boldsymbol{C} \end{bmatrix}^{-1} = \\ = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}_{\boldsymbol{A}}^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}_{\boldsymbol{A}}^{-1} \\ -\boldsymbol{S}_{\boldsymbol{A}}^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{-1} & \boldsymbol{S}_{\boldsymbol{A}}^{-1} \end{bmatrix}, \tag{10}$$

where

$$\boldsymbol{S}_{\boldsymbol{A}} = \boldsymbol{C} - \boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{B}.$$

When jointly optimizing latency time and ordering, we therefore only have to invert the full $B(\nu + 1) \times B(\nu + 1)$ matrix for $\nu = L$. For all subsequent latency times, only a $B \times B$ matrix must be inverted. The algorithm for joint optimization can be seen in Table 3, its complexity is $\mathrm{O}(B^3(Q + L)^3)$.

Special attention however must be paid to the problem of numerical error propagation. Simulations have indicated that the successive inversion of a matrix of increasing size using (10) is only suitable for a small number of iterations.

6. FIXED LATENCY TIME

Simulations have shown that for most channel models the algorithm in Table 3 returns $\nu_{\mathrm{WF}} = L$ in the overwhelming majority of channel realizations. We therefore investigated the effect of using a fixed latency time through extensive simulations.

The simulation results indicate that as long as the power delay profile of the channel model is decaying, or even constant or 'U-shaped', ST-THP with fixed latency time $\nu = L$

Table 3. Joint Optimization of Latency Time and Ordering

$\nu \leftarrow L$
compute $\boldsymbol{C}_{(\nu,B)}^{(\mathcal{O}),-1}$ for $\boldsymbol{\Pi}^{(\mathcal{O})} = \mathbf{1}_B$
compute $\boldsymbol{\Pi}^{(\mathcal{O})}$ and $\varepsilon^{(\mathcal{O})}(\nu)$ using Table 2
$\boldsymbol{\Pi}^{(\mathcal{O}_{\mathrm{WF}})} \leftarrow \boldsymbol{\Pi}^{(\mathcal{O})}$
$\varepsilon_{\min} \leftarrow \varepsilon^{(\mathcal{O})}(\nu)$
$\nu_{\mathrm{WF}} \leftarrow L$
for $\nu = L+1, \ldots, L+Q$:
    compute $\boldsymbol{C}_{(\nu,B)}^{(\mathcal{O}),-1}$ for $\boldsymbol{\Pi}^{(\mathcal{O})} = \mathbf{1}_B$ using $\boldsymbol{C}_{(\nu-1,B)}^{(\mathcal{O}),-1}$ and (10)
    compute $\boldsymbol{\Pi}^{(\mathcal{O})}$ and $\varepsilon^{(\mathcal{O})}(\nu)$ using Table 2
    if $\varepsilon^{(\mathcal{O})}(\nu) < \varepsilon_{\min}$:
        $\boldsymbol{\Pi}^{(\mathcal{O}_{\mathrm{WF}})} \leftarrow \boldsymbol{\Pi}^{(\mathcal{O})}$
        $\varepsilon_{\min} \leftarrow \varepsilon^{(\mathcal{O})}(\nu)$
        $\nu_{\mathrm{WF}} \leftarrow \nu$

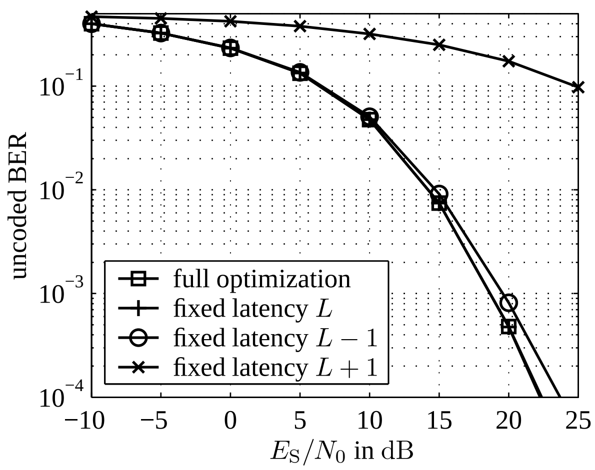


Fig. 5. Channel Order $Q = 3$, 'Pedestrian A' Power Delay Profile, Filter Order $L = 2$

performs nearly exactly as well as ST-THP with joint latency time and ordering optimization, regardless of all other system parameters. This can be seen in Fig. 5 and Fig. 6, where we used the same channel model and simulation parameters as in Section 4.1, unless noted otherwise in the caption. For channel models with increasing power delay profiles however, we found that setting the latency time to a fixed value can indeed result in significant degradation of the performance (cf. Fig. 7). In such scenarios the order of the feedforward filter L plays a crucial role: the larger L is chosen, the smaller the performance penalty of omitting the latency time optimization becomes. In our simulations, when L was larger than Q, the difference between fully optimized ST-THP and ST-THP with fixed latency time $\nu = L$ became unnoticable, regardless of the power delay profile.

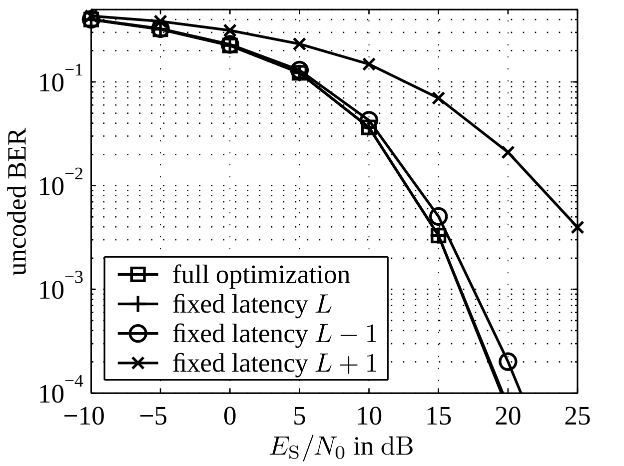


Fig. 6. Channel Order $Q = 5$, 'Vehicular A' Power Delay Profile, Filter Order $L = 3$

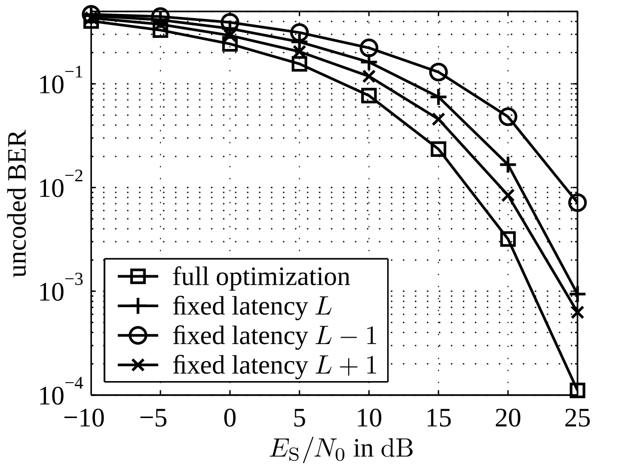


Fig. 7. Channel Order $Q = 5$, Exponentially Increasing Power Delay Profile (3dB per Tap), Filter Order $L = 3$

7. SUMMARY

In this paper, we derived the MMSE filter solutions for ST-THP in a simple notation that allows us to investigate the issues of ordering and latency time. We presented a procedure for jointly optimizing latency time and ordering in $\mathrm{O}(B^3(Q+L)^3)$ floating point operations. This was achieved on the one hand by using an efficient ordering algorithm based on the Cholesky factorization, on the other hand by utilizing the inversion properties of partitioned matrices for the latency time. It was also shown that the Cholesky factorization can be employed to reduce the complexity order of computing the filter solutions.

Furthermore, we proved that for THP with FIR feedforward filter, latency times smaller than the order of the feedforward filter do not need to be considered. This result of course is also significant for single input single output (SISO) systems.

Finally, in Section 6 we showed that for practical channel models with decaying power delay profiles, latency time optimization may be omitted altogether without a penalty in performance. Even for increasing power delay profiles, we can avoid the need for latency time optimization by *a priori* choosing a sufficiently large order of the feedforward filter.

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

[3] R. Esmailzadeh and M. Nakagawa, "Pre-RAKE Diversity Combination for Direct Sequence Spread Spectrum Mobile Communications Systems," *IEICE Transactions on Communications*, vol. E76-B, no. 8, pp. 1008–1015, August 1993.

[4] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *Accepted for Publication in IEEE Transactions on Signal Processing*.

[5] 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*, November 2002, vol. 1, pp. 429–433.

[6] 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*, September 2000, vol. 1, pp. 1–5.

[7] R. Irmer, A. Noll Barreto, and G. Fettweis, "Transmitter Precoding for Spread-Spectrum Signals in Frequency-Selective Fading Channels," in *Proc. 3G Wireless*, May 2001, pp. 939–944.

[8] 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 Transactions on Communications*, vol. 53, no. 1, pp. 195–202, January 2005.

[9] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-Reduction-Aided Broadcast Precoding," *IEEE Transactions on Communications*, vol. 52, no. 12, pp. 2057–2060, December 2004.

[10] S. Shi and M. Schubert, "Precoding and Power Loading for Multi-Antenna Broadcast Channels," in *Proc. 38th Annual Conference on Information Sciences and Systems (CISS)*, March 2004.

[11] R. Hunger, F. Dietrich, M. Joham, and W. Utschick, "Robust Transmit Zero-Forcing Filters," in *Proc. ITG Workshop on Smart Antennas*, March 2004.

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

[13] 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.

[14] 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*, October 2000, vol. 2, pp. 1627–1632.

[15] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "MIMO Precoding for Decentralized Receivers," in *Proc. ISIT 2002*, June/July 2002, p. 496.

[16] J. Liu and A. Duel-Hallen, "Tomlinson-Harashima Transmitter Precoding for Synchronous Multiuser Communications," in *Proc. CISS'03*, March 2003.

[17] 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.

[18] M. Joham, J. Brehmer, A. Voulgarelis, and W. Utschick, "Multiuser Spatio-Temporal Tomlinson-Harashima Precoding for Frequency Selective Vector Channels," in *Proc. ITG Workshop on Smart Antennas*, March 2004.

[19] R. F. H. Fischer, C. Stierstorfer, and J. B. Huber, "Precoding for Point-to-Multipoint Transmission over MIMO ISI Channels," in *International Zurich Seminar (IZS '04)*, February 2004, pp. 208–211.

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

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

[22] K. Kusume, M. Joham, and W. Utschick, "Efficient Tomlinson-Harashima Precoding for Spatial Multiplexing on Flat MIMO Channel," *Accepted for Publication at ICC 2005*.

[23] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[24] ETSI, "Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)," April 1998.

[25] C.-K. Li and R. Mathias, "Extremal Characterizations of the Schur Complement and Resulting Inequalities," *SIAM Review*, vol. 42, no. 2, pp. 233–246, 2000.