IEEE 6th CAS Symp. on Emerging Technologies: Mobile and Wireless Comm.
Shanghai, China, May 31~June 2, 2004

Transmit Processing in MIMO Wireless Systems

Josef A. Nossek, Michael Joham, and Wolfgang Utschick
Institute for Circuit Theory and Signal Processing,
Munich University of Technology, 80333 Munich, Germany
Phone: +49 (89) 289-28501, Email: nossek@nws.ei.tum.de

***Abstract*— By restricting the receive filter to be scalar, we derive the optimizations for linear transmit processing from the respective joint optimizations of transmit and receive filters. We can identify three filter types similar to receive processing: the *transmit matched filter* (TxMF), the *transmit zero-forcing filter* (TxZF), and the *transmit Wiener filter* (TxWF). The TxWF has similar convergence properties as the receive Wiener filter, i.e. it converges to the matched filter and the zero-forcing filter for low and high signal to noise ratio, respectively. Additionally, the *mean square error* (MSE) of the TxWF is a lower bound for the MSEs of the TxMF and the TxZF.**

The optimizations for the linear transmit filters are extended to obtain the respective *Tomlinson-Harashima precoding* (THP) solutions. The new formulation of the THP optimizations directly leads to an algorithm to compute the precoding order. We also discuss the reasons why the linear transmit filters are outperformed by the respective THP schemes.

I. Introduction

The most general approach to *multiple input multiple output* (MIMO) communication systems is the application of jointly optimized transmit and receive filters (see e.g. [1], [2], [3], [4], [5], [6], [7], [8]). Since the *joint optimization of transmit and receive filters* is usually based on the assumption that the transmitters but also the receivers *cooperate*, it can neither be applied to the *uplink* (non-cooperating transmitters) nor to the *downlink* (non-cooperating receivers) of a multi-user wireless communication system.

In a system with *receive processing* (e.g. [9]), the transmitter utilizes an *a priori* defined signal processing and the receiver has to equalize the channel. Obviously, this approach leads to simple transmitters and the transmitters can be *non-cooperative*. Thus, receive processing is well suited for the *uplink* of wireless MIMO systems. However, the receivers have to be cooperative for optimum receive processing. Therefore, receive processing is not appropriate for the downlink.

To end up with *simple non-cooperating receivers*, *transmit processing* has to be employed for the downlink. The receive filter is *a priori* defined and known to the transmitter, whereas the *transmit filter* is applied to equalize the channel.

Contrary to receive processing, the optimizations for transmit processing are not well established. We show in Section III that the well known optimizations for receive processing evolve from the respective joint optimizations of transmit and receive filters (see Section II) by restricting the transmit filter to be an identity matrix weighted by a scalar. Motivated by this result, we obtain the optimizations for transmit processing from the respective joint optimizations of transmit and receive filters by restricting the receive filter to be an identity matrix weighted by a scalar in Section IV. In Section V, we compare linear transmit and receive processing.

A popular nonlinear transmit processing scheme is *Tomlinson Harashima precoding* (THP) originally proposed for dispersive *single input single output* (SISO) systems in [10], [11], but can also be applied to MIMO systems [12], [13]. THP is similar to *decision feedback equalization* (DFE), but contrary to DFE, THP feeds back the *already transmitted symbols* to reduce the interference caused by these symbols at the *receivers*. Although THP is based on the application of nonlinear modulo operators at the receivers and the transmitter, similar optimizations as for the linear transmit filters can be used to find the THP filters. In Section VI, we show how the optimizations of Section IV for linear transmit processing have to be extended to get the respective optimizations for THP, where we incorporate the precoding order into the system model. Consequently, we end up with an algorithm to find the precoding order directly resulting from the THP optimization. In Section VII, we compare the linear transmit filters to THP.

Note that we assume that the channel and the statistics of the received noise are perfectly known to the receiver and the transmitter. Generally, these parameters have to be signaled from the receiver to the transmitter, but in *time division duplex* systems, the transmitter can reuse the channel estimate obtained during the reception in the other link to design the transmit filter. Since the channel is time varying, the performance deteriorates [14], [15] and the applicability of transmit processing is limited. A *robust design* (see [16]) reduces the negative effect of outdated channel knowledge and therefore enhances the applicability of transmit processing.

Notation: Vectors and matrices are denoted by lower case bold and capital bold letters, respectively. We use $\mathrm{E}[\bullet]$, $\mathrm{tr}(\bullet)$, $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, and $(\bullet)^{\mathrm{H}}$ for expectation, trace of a matrix, complex conjugation, transposition, and conjugate transposition, respectively. All random processes are assumed to be zero-mean and stationary. The covariance matrix of the vector process $\boldsymbol{x}$ is denoted by $\boldsymbol{R}_{\boldsymbol{x}} = \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}]$ and the variance of the scalar process y is $\sigma_y^2 = \mathrm{E}[|y|^2]$. The K-dimensional zero vector is $\mathbf{0}_K$, the $K \times L$ zero matrix is $\mathbf{0}_{K\times L}$, and the $L \times L$ identity matrix is $\mathbf{1}_L$, whose ℓ-th column is $\boldsymbol{e}_\ell$.

II. Joint Optimization of Transmit and Receive Filters

As depicted in Fig. 1, the data signal $\boldsymbol{s} \in \mathbb{C}^B$ is passed through the transmit filter $\boldsymbol{P} \in \mathbb{C}^{M\times B}$ to form the transmit signal $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{s} \in \mathbb{C}^M$, where M denotes the number of

0-7803-7938-1/04 /$20.00

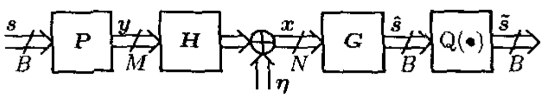

Fig. 1. MIMO System with Linear Transmit and Receive Filters

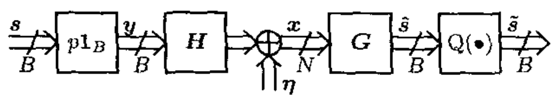

Fig. 2. MIMO System with Linear Receive Filter

scalar channel inputs. After propagation over the channel $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ and perturbation by the noise $\boldsymbol{\eta} \in \mathbb{C}^{N}$, the resulting received signal $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y} + \boldsymbol{\eta} \in \mathbb{C}^{N}$ is transformed by the receive filter $\boldsymbol{G} \in \mathbb{C}^{B\times N}$ to get the estimate

$$\hat{\boldsymbol{s}} = \boldsymbol{G}\boldsymbol{H}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{G}\boldsymbol{\eta} \in \mathbb{C}^{B}. \tag{1}$$

Here, N denotes the number of scalar channel outputs and in Fig. 1, $Q(\bullet)$ denotes quantization.

Although also other joint optimizations have been proposed (e. g. in [8]), we concentrate on the following three criteria, because they are a generalization of the most often used criteria for receive processing.

The *eigenprecoder* [17], [18] was originally proposed for CDMA systems, but can also be derived for the more general system of Fig. 1 by maximizing the desired signal portion in the estimate under a mean transmit energy constraint:

$$\{\boldsymbol{P}_{\mathrm{MF}}^{\mathrm{joint}}, \boldsymbol{G}_{\mathrm{MF}}^{\mathrm{joint}}\} = \underset{\{\boldsymbol{P},\boldsymbol{G}\}}{\operatorname{argmax}} \frac{\left|\mathrm{E}\left[\boldsymbol{s}^{\mathrm{H}}\hat{\boldsymbol{s}}\right]\right|^2}{\mathrm{E}\left[\|\boldsymbol{s}\|_2^2\right]\mathrm{E}\left[\|\boldsymbol{G}\boldsymbol{\eta}\|_2^2\right]} \quad \text{s.t.: } \mathrm{E}\left[\|\boldsymbol{y}\|_2^2\right] = E_{\mathrm{tr}}. \tag{2}$$

The *joint zero-forcing optimization* is based on the minimization of the MSE together with a zero-forcing constraint and a transmit energy constraint (e. g. [19], [20], [21]):

$$\{\boldsymbol{P}_{\mathrm{ZF}}^{\mathrm{joint}}, \boldsymbol{G}_{\mathrm{ZF}}^{\mathrm{joint}}\} = \underset{\{\boldsymbol{P},\boldsymbol{G}\}}{\operatorname{argmin}} \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_2^2\right] \quad \text{s.t.: } \hat{\boldsymbol{s}}|_{\boldsymbol{\eta}=\boldsymbol{0}_N} = \boldsymbol{s} \text{ and } \mathrm{E}\left[\|\boldsymbol{y}\|_2^2\right] = E_{\mathrm{tr}}. \tag{3}$$

We call the minimization of the MSE with a mean transmit energy constraint the *joint Wiener optimization* (e. g. [5], [8]):

$$\{\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{joint}}, \boldsymbol{G}_{\mathrm{WF}}^{\mathrm{joint}}\} = \underset{\{\boldsymbol{P},\boldsymbol{G}\}}{\operatorname{argmin}} \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_2^2\right] \quad \text{s.t.: } \mathrm{E}\left[\|\boldsymbol{y}\|_2^2\right] = E_{\mathrm{tr}}. \tag{4}$$

Note that the solution for the joint Wiener optimization (4) converges to the eigenprecoder (2) and the solution of the joint zero-forcing optimization (3) for low and high *signal to noise ratio* (SNR), respectively.

III. Linear Receive Processing

By restricting the transmit filter to be an identity matrix weighted by a scalar, that is $\boldsymbol{P} = p\mathbf{1}_B$ with $p \in \mathbb{C}$, we obtain Fig. 2 from Fig. 1. The resulting system in Fig. 2 is appropriate for receive processing, since the channel $\boldsymbol{H} \in \mathbb{C}^{N\times B}$ can only be equalized by the receive filter $\boldsymbol{G} \in \mathbb{C}^{B\times N}$. The estimate can be written as (cf. Eq. 1)

$$\hat{\boldsymbol{s}} = p\boldsymbol{G}\boldsymbol{H}\boldsymbol{s} + \boldsymbol{G}\boldsymbol{\eta} \in \mathbb{C}^{B}. \tag{5}$$

With the same restriction on the transmit filter to be equal to $p\mathbf{1}_B$, we obtain the well known optimizations for receive processing from (2), (3), and (4).

The optimization for the *receive matched filter* (RxMF) follows from the eigenprecoder optimization (2):

$$\{p_{\mathrm{MF}}, \boldsymbol{G}_{\mathrm{MF}}\} = \underset{\{p,\boldsymbol{G}\}}{\operatorname{argmax}} \frac{\left|\mathrm{tr}\left(p\boldsymbol{H}\boldsymbol{G}\boldsymbol{R}_{\boldsymbol{s}}\right)\right|^2}{\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{s}}\right)\mathrm{tr}\left(\boldsymbol{G}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{G}^{\mathrm{H}}\right)} \quad \text{s.t.: } |p|^2\,\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{s}}\right) = E_{\mathrm{tr}}. \tag{6}$$

A possible solution of above RxMF optimization is:

$$p_{\mathrm{MF}} = \sqrt{\frac{E_{\mathrm{tr}}}{\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{s}}\right)}}, \quad \boldsymbol{G}_{\mathrm{MF}} = p_{\mathrm{MF}}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{7}$$

By setting $\boldsymbol{P} = p\mathbf{1}_B$ in the joint zero-forcing optimization (3), we obtain the *receive zero-forcing filter* (RxZF) optimization:

$$\{p_{\mathrm{ZF}}, \boldsymbol{G}_{\mathrm{ZF}}\} = \underset{\{p,\boldsymbol{G}\}}{\operatorname{argmin}}\, \mathrm{tr}\left(\boldsymbol{G}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{G}^{\mathrm{H}}\right) \quad \text{s.t.: } p\boldsymbol{H}\boldsymbol{G} = \mathbf{1}_B \text{ and } |p|^2\,\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{s}}\right) = E_{\mathrm{tr}}. \tag{8}$$

With the restriction $p \in \mathbb{R}_+$, the RxZF of (8) is unique:

$$p_{\mathrm{ZF}} = p_{\mathrm{MF}}, \quad \boldsymbol{G}_{\mathrm{ZF}} = p_{\mathrm{ZF}}^{-1}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{9}$$

The *receive Wiener filter* (RxWF) for the system in Fig. 2 can be found with (cf. Eq. 4):

$$\{p_{\mathrm{WF}}, \boldsymbol{G}_{\mathrm{WF}}\} = \underset{\{p,\boldsymbol{G}\}}{\operatorname{argmin}} \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_2^2\right] \quad \text{s.t.: } |p|^2\,\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{s}}\right) = E_{\mathrm{tr}}. \tag{10}$$

The resulting RxWF for $p \in \mathbb{R}_+$ can be expressed as

$$p_{\mathrm{WF}} = p_{\mathrm{MF}}, \; \boldsymbol{G}_{\mathrm{WF}} = p_{\mathrm{WF}}\left(\boldsymbol{R}_{\boldsymbol{s}}^{-1} + p_{\mathrm{WF}}^2\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{11}$$

Obviously, $\boldsymbol{G}_{\mathrm{WF}} \rightarrow \boldsymbol{G}_{\mathrm{MF}}$ for low SNR and $\boldsymbol{G}_{\mathrm{WF}} \rightarrow \boldsymbol{G}_{\mathrm{ZF}}$ for high SNR. Note that the optimizations for receive processing are usually formulated without the transmit energy constraint and the scalar weight p at the transmitter (e. g. [9]). Since $p = \sqrt{E_{\mathrm{tr}}/\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{s}})}$ is only necessary to fulfill the transmit energy constraint and the receiver cannot distinguish between p and $\boldsymbol{H}$, the receiver assumes $E_{\mathrm{tr}} = \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{s}})$ ($p' = 1$) and bases the receive filter on the total channel $\boldsymbol{H}' = p\boldsymbol{H}$.

IV. Linear Transmit Processing

When restricting the receive filter in Fig. 1 to be an identity matrix weighted by a scalar, i. e. $\boldsymbol{G} = g\mathbf{1}_B$ with $g \in \mathbb{C}$, we end up with the system for transmit processing in Fig. 3. The estimate reads as (cf. Eq. 1)

$$\hat{\boldsymbol{s}} = g\boldsymbol{H}\boldsymbol{P}\boldsymbol{s} + g\boldsymbol{\eta} \in \mathbb{C}^{B}. \tag{12}$$

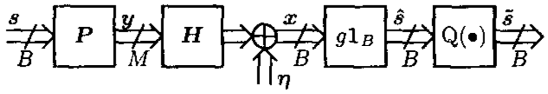

Fig. 3. MIMO System with Linear Transmit Filter

The channel $\boldsymbol{H} \in \mathbb{C}^{B\times M}$ has to be equalized by the transmit filter $\boldsymbol{P} \in \mathbb{C}^{M\times B}$. Similar to receive processing in the previous section, the optimizations for transmit processing evolve from (2), (3), and (4) with the restriction that $\boldsymbol{G} = g\mathbf{1}_B$.

The *transmit matched filter* (TxMF) was first proposed by McIntosh et al. [22] and became especially popular in CDMA applications as *prerake* (see e. g. [23], [24], [25]). The TxMF maximizes the received signal power [22]:

$$\{\boldsymbol{P}_{\mathrm{MF}}, g_{\mathrm{MF}}\} = \underset{\{\boldsymbol{P},g\}}{\operatorname{argmax}} \frac{|\mathrm{tr}\,(g\boldsymbol{H}\boldsymbol{P}\boldsymbol{R}_s)|^2}{\mathrm{tr}\,(\boldsymbol{R}_s)\,\mathrm{tr}\left(|g|^2\boldsymbol{R}_\eta\right)} \quad \text{s. t.: } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{R}_s\boldsymbol{P}^{\mathrm{H}}\right) = E_{\mathrm{tr}}. \tag{13}$$

The TxMF

$$\boldsymbol{P}_{\mathrm{MF}} = \sqrt{\frac{E_{\mathrm{tr}}}{\mathrm{tr}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_s\boldsymbol{H}\right)}}\boldsymbol{H}^{\mathrm{H}} \quad \text{and} \quad g_{\mathrm{MF}} \in \mathbb{C} \tag{14}$$

is thus advantageous for systems with low transmit energy or high received noise power. Since the TxMF does not take into account interference, its performance is poor for interference limited scenarios.

This property of the TxMF motivates the *transmit zero-forcing filter* (TxZF) which completely suppresses the interference at the receive filter output (cf. Eq. 3):

$$\{\boldsymbol{P}_{\mathrm{ZF}}, g_{\mathrm{ZF}}\} = \underset{\{\boldsymbol{P},g\}}{\operatorname{argmin}}\, |g|^2\,\mathrm{tr}\,(\boldsymbol{R}_\eta) \quad \text{s. t.: } g\boldsymbol{H}\boldsymbol{P} = \mathbf{1}_B \text{ and } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{R}_s\boldsymbol{P}^{\mathrm{H}}\right) = E_{\mathrm{tr}}. \tag{15}$$

With $g_{\mathrm{ZF}} \in \mathbb{R}_+$, we find for the TxZF:

$$\boldsymbol{P}_{\mathrm{ZF}} = g_{\mathrm{ZF}}^{-1}\boldsymbol{H}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1} \quad \text{and} \quad g_{\mathrm{ZF}} = \sqrt{\frac{\mathrm{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\boldsymbol{R}_s\right)}{E_{\mathrm{tr}}}}. \tag{16}$$

Most publications on the TxZF (e. g. [26], [27], [28], [29], [30], [14]) focus on CDMA systems, but the TxZF can also be applied to MIMO systems [31]. Due to the interference suppression, the TxZF outperforms the TxMF in interference limited scenarios, but the TxZF is worse than the TxMF, if the transmit energy is low or the received noise power is high.

The *transmit Wiener filter* (TxWF) minimizes the MSE under the average transmit energy constraint (cf. Eq. 4):

$$\{\boldsymbol{P}_{\mathrm{WF}}, g_{\mathrm{WF}}\} = \underset{\{\boldsymbol{P},g\}}{\operatorname{argmin}}\, \mathrm{E}\left[\|\boldsymbol{s}-\hat{\boldsymbol{s}}\|_2^2\right] \quad \text{s. t.: } \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{R}_s\boldsymbol{P}^{\mathrm{H}}\right) = E_{\mathrm{tr}}. \tag{17}$$

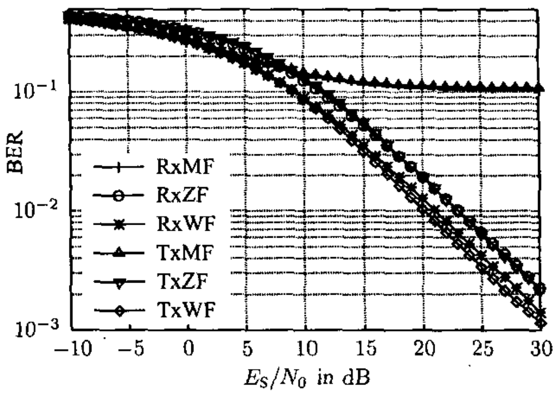

Fig. 4. QPSK Transmission over MIMO Channel with Two Transmitting and Two Receiving Antenna Elements: Uncoded BER vs. SNR for Linear Receive and Transmit Processing

For $g_{\mathrm{WF}} \in \mathbb{R}_+$, the TxWF can be written as

$$\boldsymbol{P}_{\mathrm{WF}} = g_{\mathrm{WF}}^{-1}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi_{\mathrm{WF}}\mathbf{1}_M\right)^{-1}\boldsymbol{H}^{\mathrm{H}} \quad \text{and} \quad g_{\mathrm{WF}} = \sqrt{\frac{\mathrm{tr}\left(\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi_{\mathrm{WF}}\mathbf{1}_M\right)^{-2}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_s\boldsymbol{H}\right)}{E_{\mathrm{tr}}}}. \tag{18}$$

where $\xi_{\mathrm{WF}} = \mathrm{tr}(\boldsymbol{R}_\eta)/E_{\mathrm{tr}}$. Karimi et al. [31] introduced this filter in an intuitive way, whereas Choi et al. [32] and Joham et al. [33] published above optimization to obtain the TxWF. Interestingly, the TxWF converges to the TxMF for low SNR and to the TxZF for high SNR:

$$\boldsymbol{P}_{\mathrm{WF}} \to \boldsymbol{P}_{\mathrm{MF}} \quad \text{for} \quad E_{\mathrm{tr}}/\mathrm{E}\left[\|\boldsymbol{\eta}\|_2^2\right] \to 0 \quad \text{and}$$
$$\boldsymbol{P}_{\mathrm{WF}} \to \boldsymbol{P}_{\mathrm{ZF}} \quad \text{for} \quad E_{\mathrm{tr}}/\mathrm{E}\left[\|\boldsymbol{\eta}\|_2^2\right] \to \infty.$$

Note that the TxWF $\boldsymbol{P}_{\mathrm{WF}}$ depends on the average received noise power $\mathrm{E}[\|\boldsymbol{\eta}\|_2^2]$ contrary to the TxMF and the TxZF. Since the transmitter cannot estimate this parameter, the receiver has to signal the noise power to the transmitter. Fortunately, the TxWF is very robust against wrong values for $\mathrm{E}[\|\boldsymbol{\eta}\|_2^2]$ (see e. g. [33]). Consequently, the transmitter only needs a rough estimate of this value.

V. Comparison of Linear Transmit Filters

Obviously, the MSEs of the TxMF and the TxZF are always lower bounded by the MSE of the TxWF, as the TxWF minimizes the MSE. Moreover, it can be shown analytically that the MSE of the TxZF is higher than the MSE of the TxMF for low SNR, but the MSE of the TxZF is smaller for high SNR (see [34]). We have also shown that the MSEs of the transmit filters are equal to the MSEs of the respective receive filters, if the symbols $\boldsymbol{s}$ and the noise $\boldsymbol{\eta}$ are white [34].

The same relations are true for the BERs of the transmit filters, but cannot be shown analytically. To illustrate this statement, we present the uncoded BER results for the

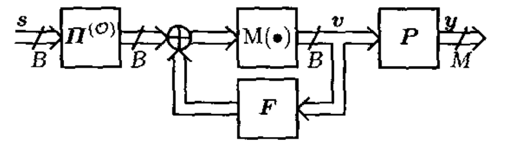

Fig. 5. Tomlinson-Harashima Precoder with Modulo Operator M(•) and Permutation Matrix $\boldsymbol{\Pi}$

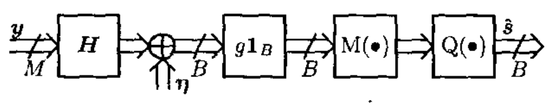

Fig. 6. Channel and Receiver with Quantizer Q(•)

discussed transmit filters and the respective receive filters applied to a MIMO system with two transmit and two receive antenna elements in Fig. 4. We see that the TxWF is superior compared to the TxMF and the TxZF. The BER of the TxMF is smaller than the BER of the TxZF for low SNR, but the BER of the TxMF saturates for high SNR due to the remaining interference.

VI. Tomlinson-Harashima Precoding

The interference suppressing linear transmit filters, i. e. the TxZF and the TxWF (see Section IV), depend on the pseudo inverse of the channel $\boldsymbol{H}$. Therefore, the power of the received signal can be small, if the channel is ill-conditioned, because the transmit energy is fixed. To overcome this difficulty, a feedback loop with the feedback filter $\boldsymbol{F} \in \mathbb{C}^{B\times B}$ is introduced (see Fig. 5). Inside the loop, the nonlinear element-wise modulo operator

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

is applied to ensure that the amplitude of the output $\boldsymbol{v} \in \mathbb{C}^B$ is limited. The modulo constant τ is chosen according to the modulation alphabet (see e. g. [35]). To counteract this modulo operator at the transmitter, the receiver also has to apply a modulo operator M(•) together with the quantizer Q(•) in a THP system (see Fig. 6). In Fig. 5, we also see that the data signal $\boldsymbol{s} \in \mathbb{C}^B$ is transformed by a permutation matrix

$$\boldsymbol{\Pi}^{(\mathcal{O})} = \sum_{i=1}^{B} \boldsymbol{e}_i \boldsymbol{e}_{b_i}^{\mathrm{T}} \in \{0,1\}^{B\times B} \quad \text{with}$$
$$b_i \neq b_j \text{ for } i \neq j \quad \text{and} \quad b_i \in \{1,\ldots,B\} \text{ for } i = 1,\ldots,B,$$

before it is passed through the precoder, i. e. the b_i-th scalar data stream is precoded i-th. This reordering represents a degree of freedom and is part of the optimization to further optimize the cost function. For brevity, we collect the B indices defining $\boldsymbol{\Pi}^{(\mathcal{O})}$ in the B-tupel $\mathcal{O} = (b_1, \ldots, b_B)$.

Note that the modulo operator M(•) can be expressed as the sum of the input and an auxiliary signal. When including this linear representation of M(•) into Figs. 5 and 6, we end up with Figs. 7 and 8. The resulting signals $\boldsymbol{d}$ and $\hat{\boldsymbol{d}}$ are used

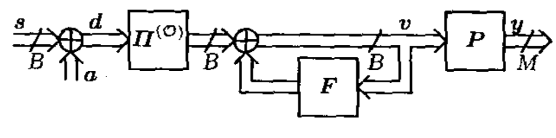

Fig. 7. Linear Representation of Modulo Operation at Transmitter

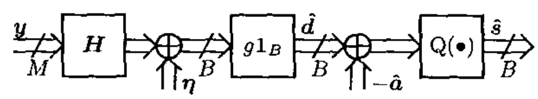

Fig. 8. Linear Representation of Modulo Operation at Receiver

for the THP optimizations instead of $\boldsymbol{s}$ and $\hat{\boldsymbol{s}}$, because the auxiliary signals $\boldsymbol{a}$ and $-\hat{\boldsymbol{a}}$ are added automatically by the modulo operator M(•). The estimate depending on the modulo output $\boldsymbol{v}$ reads as

$$\hat{\boldsymbol{d}} = g\boldsymbol{H}\boldsymbol{P}\boldsymbol{v} + g\boldsymbol{v} \in \mathbb{C}^B, \tag{19}$$

whereas we get for the desired value for the estimate:

$$\boldsymbol{d} = \boldsymbol{\Pi}^{(\mathcal{O}),\mathrm{T}} (\mathbf{1}_B - \boldsymbol{F})\, \boldsymbol{v} \in \mathbb{C}^B. \tag{20}$$

Since the modulo operator M(•) is applied element-wise, the feedback filter $\boldsymbol{F}$ has to be *lower triangular with zero main diagonal*. Thereby, delay-free loops are avoided ensuring realizability, because only already precoded data streams are fed back by $\boldsymbol{F}$. Including this constraint, we can rewrite the TxWF optimization (17) to obtain the *Wiener THP* (WF-THP) optimization [36], [37]:

$$\{\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{THP}}, \boldsymbol{F}_{\mathrm{WF}}^{\mathrm{THP}}, g_{\mathrm{WF}}^{\mathrm{THP}}, \mathcal{O}_{\mathrm{WF}}^{\mathrm{THP}}\} = \underset{\{\boldsymbol{P},\boldsymbol{F},\beta,\mathcal{O}\}}{\operatorname{argmin}}\ \mathrm{E}\left[\|\boldsymbol{d} - \hat{\boldsymbol{d}}\|_2^2\right] \tag{21}$$
$$\text{s.t.: } \mathrm{E}\left[\|\boldsymbol{y}\|_2^2\right] = E_{\mathrm{tr}} \ \text{ and } \ \boldsymbol{F}: \text{ lower triang., zero diagonal.}$$

Under the common assumption of uncorrelated modulo outputs $\boldsymbol{v}$, that is $\boldsymbol{R}_{\boldsymbol{v}} = \mathrm{diag}(\sigma_{v_1}^2, \ldots, \sigma_{v_B}^2)$, the WF-THP filters depending on the ordering $\mathcal{O}$ can be written as

$$\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{THP}} = \frac{1}{g_{\mathrm{WF}}^{\mathrm{THP}}} \sum_{i=1}^{B} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}_i \left(\boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}_i + \xi_{\mathrm{WF}} \mathbf{1}_B\right)^{-1} \boldsymbol{e}_{b_i} \boldsymbol{e}_i^{\mathrm{T}},$$
$$\boldsymbol{F}_{\mathrm{WF}}^{\mathrm{THP}} = g_{\mathrm{WF}}^{\mathrm{THP}} \sum_{i=1}^{B} \left(\boldsymbol{S}_i^{\mathrm{T}} \boldsymbol{S}_i - \mathbf{1}_B\right) \boldsymbol{\Pi}^{(\mathcal{O})} \boldsymbol{H} \boldsymbol{P}_{\mathrm{WF}}^{\mathrm{THP}} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}, \tag{22}$$
$$g_{\mathrm{WF}}^{\mathrm{THP}} = \sqrt{\frac{\sum_{i=1}^{B} \sigma_{v_i}^2 \boldsymbol{e}_{b_i}^{\mathrm{T}} (\boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}_i + \xi_{\mathrm{WF}} \mathbf{1}_B)^{-2} \boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}_i \boldsymbol{e}_{b_i}}{E_{\mathrm{tr}}}},$$

where we introduced $\xi_{\mathrm{WF}} = \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{\eta}})/E_{\mathrm{tr}}$. The selection matrix $\boldsymbol{S}_i = [\mathbf{1}_i, \mathbf{0}_{i\times B-i}]$ cuts out the first i elements of a B-dimensional column vector, and the projection matrix $\boldsymbol{\Pi}_i = \mathbf{1}_B - \sum_{j=i+1}^{B} \boldsymbol{e}_{b_j} \boldsymbol{e}_{b_j}^{\mathrm{T}} \in \{0,1\}^{B\times B}$ sets the b_{i+1}-th to the b_B-th row of the channel matrix $\boldsymbol{H}$ to zero. Therefore, the i-th column of the feedforward filter $\boldsymbol{P}_{\mathrm{WF}}^{\mathrm{THP}} \in \mathbb{C}^{M\times B}$ only depends on rows of the channel matrix $\boldsymbol{H}$ corresponding to data streams which are precoded later.

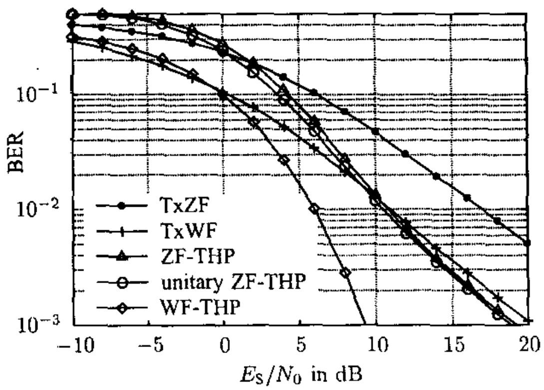

Fig. 9. QPSK Transmission over MIMO Channel with Four Transmitting and Four Receiving Antenna Elements: Uncoded BER vs. SNR for Linear and Nonlinear Transmit Processing

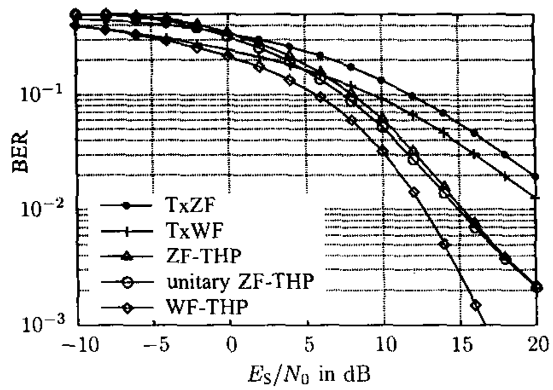

Fig. 10. 16QAM Transmission over MIMO Channel with Four Transmitting and Four Receiving Antenna Elements: Uncoded BER vs. SNR for Linear and Nonlinear Transmit Processing

Plugging the result of (22) into the cost function of (21), we find for the optimum WF-THP ordering:

$$\mathcal{O}'_{\mathrm{WF}} = \underset{\mathcal{O}}{\operatorname{argmin}} \sum_{i=1}^{B} \sigma_{v_i}^2 e_{b_i}^{\mathrm{T}} \left(\boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}_i + \xi_{\mathrm{WF}} \mathbf{1}_B \right)^{-1} e_{b_i}. \tag{23}$$

To avoid the high complexity $\mathrm{O}(B!B^3)$ of this optimization, we suggest following suboptimum approach instead:

$$b_{\mathrm{WF},i}^{\mathrm{THP}} = \underset{b \in \mathbb{O}_i}{\operatorname{argmin}}\, e_b^{\mathrm{T}} \left(\boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}_i + \xi_{\mathrm{WF}} \mathbf{1}_B \right)^{-1} e_b, \tag{24}$$

for $i = B, \ldots, 1$, i.e. instead of minimizing the whole sum, each summand is minimized separately for fixed succeeding indices. Here, $\mathbb{O}_i = \{1, \ldots, B\} \backslash \{b_{\mathrm{WF},i+1}^{\mathrm{THP}}, \ldots, b_{\mathrm{WF},B}^{\mathrm{THP}}\}$ is the set of possible indices for the i-th precoded data stream.

The *zero-forcing THP* (ZF-THP) optimization can be found by including the zero-forcing constraint $\hat{\boldsymbol{d}}|_{\eta=\mathbf{0}_B} = \boldsymbol{d}$ into the WF-THP optimization (21). The WF-THP filters in (22) converge to the ZF-THP filters for the limit $E_{\mathrm{tr}}/\operatorname{tr}(\boldsymbol{R}_\eta) \to \infty$ ($\xi_{\mathrm{WF}} \to 0$). Note that we get for ZF-THP [36]:

$$\boldsymbol{\Pi}_{\mathrm{ZF}}^{\mathrm{THP,T}} \boldsymbol{H} \boldsymbol{P}_{\mathrm{ZF}}^{\mathrm{THP}} = \mathbf{1}_B - \boldsymbol{F}_{\mathrm{ZF}}^{\mathrm{THP}},$$

where the ZF-THP feedback filter $\boldsymbol{F}_{\mathrm{ZF}}^{\mathrm{THP}}$ is a lower triangular matrix with zero main diagonal. We can conclude that the feedforward filter $\boldsymbol{P}_{\mathrm{ZF}}^{\mathrm{THP}}$ removes the interference of data streams precoded later, whereas the lower triangular feedback filter $\boldsymbol{F}_{\mathrm{ZF}}^{\mathrm{THP}}$ suppresses the interference caused by already precoded data streams. Since the THP feedforward filter $\boldsymbol{P}_{\mathrm{ZF}}^{\mathrm{THP}}$ has to suppress less interference than the TxZF $\boldsymbol{P}_{\mathrm{ZF}}$ (cf. Eq. 16), more degrees of freedom are left to maximize the received signal power. Hence, THP outperforms the respective linear transmit filter.

VII. Comparison of Linear and Nonlinear Transmit Processing

In Fig. 9, we compare the linear TxZF and TxWF with the nonlinear ZF-THP and WF-THP for a system with four transmit and four receive antenna elements and QPSK transmission. Additionally, we include the uncoded BER results for *unitary ZF-THP* [38], a variant of ZF-THP with unitary feedforward filter and a weighting with a diagonal matrix at the receiver different from a weighted identity matrix. We observe that the THP approaches clearly outperform the respective linear transmit filters for high SNR as expected, but are worse for low SNR due to the additional allowed constellation points introduced by the modulo operation at the receiver. Note that the BERs of the two ZF-THP types have the same slope as the linear transmit filters for high SNR, since the last column of $\boldsymbol{P}_{\mathrm{ZF}}^{\mathrm{THP}}$ is the weighted b_B-th column of the linear TxZF $\boldsymbol{P}_{\mathrm{ZF}}$. Thus, the diversity order of the data stream precoded last is the same as the diversity order of the TxZF data streams. As the lowest diversity order is dominant, the ZF-THP approaches have the same diversity order as the linear transmit filters. Not unexpected, the unitary ZF-THP outperforms the ZF-THP approach discussed in this paper, because the diagonal weighting at the receiver offers more degrees of freedom.

For 16QAM (see Fig. 10), the negative effect of the modulo operation at the receiver can be neglected. Therefore, the THP approaches are superior in the whole depicted SNR range. Note that WF-THP needs about 2 dB less SNR than ZF-THP for a uncoded BER of 10 %.

In Fig. 11, we show the uncoded BER results for QPSK, four transmit and three receive antenna elements. Obviously, all approaches have an improved performance compared to Fig. 9. However, the advantage of the linear transmit filters for low SNR is more pronounced and the unitary ZF-THP is worse than the ZF-THP with scalar weight at the receiver. We can conclude that the intuitively designed unitary ZF-THP is suboptimum.

References

[1] R. J. Pilc, "The Optimum Linear Modulator for a Gaussian Source Used with a Gaussian Channel," *Bell System Technical Journal*, vol. 48, no. 9, pp. 3075–3089, November 1969.

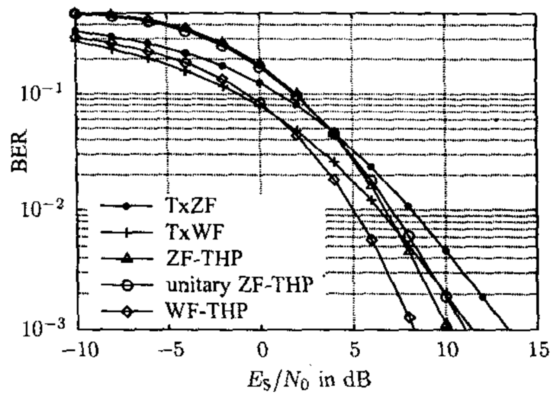

Fig. 11. QPSK Transmission over MIMO Channel with Four Transmitting and Three Receiving Antenna Elements: Uncoded BER vs. SNR for Linear and Nonlinear Transmit Processing

[2] J. Salz, "Digital Transmission Over Cross-Coupled Linear Channels," *AT&T Technical Journal*, vol. 64, no. 6, pp. 1147–1159, July–August 1985.

[3] H. S. Malvar and D. H. Staelin, "Optimal Pre- and Postfilters for Multichannel Signal Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 287–289, February 1988.

[4] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of Near- and Far-End Crosstalk by Linear Pre- and Post-Filtering," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 614–629, April 1992.

[5] J. Yang and S. Roy, "On Joint Transmitter and Receiver Optimization for Multiple-Input-Multiple-Output (MIMO) Transmission Systems," *IEEE Transactions on Communications*, vol. 42, no. 12, pp. 3221–3231, December 1994.

[6] W. M. Jang, B. R. Vojčić, and R. L. Pickholtz, "Joint Transmitter-Receiver Optimization in Synchronous Multiuser Communications over Multipath Channels," *IEEE Transactions on Communications*, vol. 46, no. 2, pp. 269–278, February 1998.

[7] H. Sampath, P. Stoica, and A. Paulraj, "Generalized Linear Precoder and Decoder Design for MIMO Channels Using the Weighted MMSE Criterion," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2198–2206, December 2001.

[8] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal Designs for Space-Time Linear Precoders and Decoders," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[9] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] F. Dietrich, R. Hunger, M. Joham, and W. Utschick, "Linear Precoding over Time-Varying Channels in TDD Systems," in *Proc. ICASSP 2003*, vol. V, April 2003, pp. 117–120.

[16] F. A. Dietrich, R. Hunger, M. Joham, and W. Utschick, "Robust Transmit Wiener Filter for Time Division Duplex Systems," in *Proc. ISSPIT 2003*, December 2003.

[17] J. Wang, M. Zhao, S. Zhou, and Y. Yao, "A Novel Multipath Transmission Diversity Scheme in TDD-CDMA Systems," *IEICE Transactions on Communications*, vol. E82-B, no. 10, pp. 1706–1709, October 1999.

[18] 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.

[19] J. W. Smith, "The Joint Optimization of Transmitted Signal and Receiving Filter for Data Transmission Systems," *Bell System Technical Journal*, vol. 44, no. 10, pp. 2363–2392, December 1965.

[20] G. D. Golden, J. E. Mazo, and J. Salz, "Transmitter Design for Data Transmission in the Presence of a Data-Like Interferer," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 837–850, February/March/April 1995.

[21] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant Filterbank Precoders and Equalizers Part I: Unification and Optimal Designs," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[22] R. E. McIntosh and S. E. El-Khamy, "Optimum Pulse Transmission Through a Plasma Medium," *IEEE Transactions on Antennas and Propagation*, vol. AP-18, no. 5, pp. 666–671, September 1970.

[23] 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.

[24] R. Esmailzadeh, E. Sourour, and M. Nakagawa, "Prerake Diversity Combining in Time-Division Duplex CDMA Mobile Communications," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 3, pp. 795–801, May 1999.

[25] R. L. Choi, K. B. Letaief, and R. D. Murch, "MISO CDMA Transmission with Simplified Receiver for Wireless Communication Handsets," *IEEE Transactions on Communications*, vol. 49, no. 5, May 2001.

[26] H. Liu and G. Xu, "Multiuser Blind Channel Estimation and Spatial Channel Pre-Equalization," in *Proc. ICASSP'95*, vol. 3, May 1995, pp. 1756–1759.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," 2004, *Accepted for publication in IEEE Transactions on Signal Processing.*

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

[36] 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.

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

[38] 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.