Iterative MMSE Transmit and Receive Filter Design for Frequency Selective MU-MISO Systems

Ralf Bendlin, Michael Joham, and Josef A. Nossek
Institute for Circuit Theory and Signal Processing
Munich University of Technology, 80333 Munich, Germany
Email: Joham@tum.de

Yih-Fang Huang
Department of Electrical Engineering
University of Notre Dame, Notre Dame, IN 46556 U.S.A.
Email: Huang@nd.edu

Abstract— We consider a *multi-user multiple-input single-output* (MU-MISO) scenario, where the decentralized users are served by a centralized transmitter with multiple channel inputs via frequency selective vector channels. For this broadcast setup, we jointly design the transmitter and receivers based on a minimum sum *mean square error* (MSE) criterion. Contrary to previous work in this field, we do not restrict the receivers to be scalar weights but employ *finite impulse response* (FIR) receive filters. Since the sum MSE minimization has no closed-form solution neither for linear precoding nor for *Tomlinson-Harashima precoding* (THP), we propose to use an alternating optimization and prove the convergence of the resulting iterative algorithm. The simulations show that the obtained linear and nonlinear precoding solutions with FIR receivers clearly outperform the state-of-the-art precoders with scalar receivers.

I. Introduction

Precoding is inevitable for the broadcast setup [1], i.e., one centralized transmitter serves several decentralized receivers, e.g., the downlink of a cellular system with a *base station* (BS) with multiple antennas transmitting to several single antenna *mobile stations* (MS). Clearly, precoding is superior to receive processing in such MU-MISO systems, since the decentralized receivers have fewer degrees of freedom than the transmitter.

The data are transformed linearly before transmission in the case of linear precoding. Compared to matched filter [2] and *zero-forcing* (ZF) schemes [3]–[7], *minimum MSE* (MMSE) based linear precoders show an excellent performance [8]–[11]. For THP, the receivers are equipped with modulo operators giving the transmitter the additional degrees of freedom to add a perturbation signal by means of a modulo feedback loop to reduce transmit power (e.g., [12]). As for linear precoding, MMSE-THP [11]–[14] clearly outperforms ZF-THP [14]–[17]. Thus, we focus on linear and non-linear precoding based on the MMSE criterion in this paper.

Most publications on precoding considered block processing so far, i.e., either precoding for flat channels [3], [4], [10], [11], [13], [16] or precoding for frequency selective channels but with block filters which process the data of a whole block [2], [5]–[7], [17]. Nevertheless, FIR precoders were proposed in [8], [12] and *infinite impulse response* (IIR) THP was considered in [14]. FIR precoders are preferable compared to block processing for frequency selective channels due to the substantially reduced complexity. Because of the practical disadvantages of IIR filters, viz., possible instability due to

finite precision and additional delay due to the necessity to use a non-causal implementation, we only consider FIR precoders.

Although the conventional design of precoding assuming scalar receivers is appropriate for flat channels, enhancing the receivers improves the performance for frequency selective channels. In [2], the matched filter at the transmitter was supported by a matched filter at the receiver. The resulting *signal-to-noise ratio* (SNR) gain is advantageous in single-user scenarios. In [6], a ZF equalizer supplemented the ZF precoder which leads to a substantial transmit power decrease.

We also enhance the receivers in this paper; instead of simple scalar weights, the equalizers are FIR filters. These FIR equalizers are jointly optimized with the precoder minimizing the MSE. Unfortunately, the resulting optimization has no closed-form solution. Hence, we propose to employ following alternating optimization (similar to [11] for flat fading channels) to find the MMSE filters. First, the precoder is updated while keeping the receivers fixed. Second, the receivers are updated while keeping the transmitter fixed. These two steps are repeated until convergence which we prove in this paper. Moreover, we give a ZF solution in closed-form which is an excellent initial value for our iterative algorithm.

We assume perfect knowledge of the *channel state information* (CSI). Clearly, this is an unrealistic assumption, but our problem statement and solution can easily be extended to erroneous CSI such as in [18], [19]. In this paper, we focus on the problem formulation/solution and highlight the performance improvement by a joint MMSE precoder/equalizer design.

After introducing the system model in Section II, we develop and discuss the iterative joint MMSE precoder/equalizer solution in Section III. The simulation results can be found in Section IV and conclusions are drawn in Section V.

Notation: Vectors and matrices are respectively denoted by bold lower and upper case letters. $[\boldsymbol{A}]_{m,n}$ is the m-th element in the n-th column of the matrix $\boldsymbol{A}$. $\mathrm{E}[\bullet]$, $\| \bullet \|_2$, $*$, $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, $\otimes$, $\mathrm{tr}(\bullet)$, and $(\bullet)^+$ denote expectation, Euclidean norm of a vector, convolution, complex conjugation, transposition, conjugate transposition, Kronecker product, trace of a matrix, and Moore-Penrose pseudo inverse, respectively. $\mathrm{diag}(\bullet)$ returns a diagonal matrix with the arguments on the diagonal. $\delta[n]$ is the unit impulse. $\boldsymbol{0}_M$, $\boldsymbol{0}_{M \times N}$, and $\boldsymbol{1}_M$ represent the M-dimensional zero vector, $M \times N$ zero matrix, and $M \times M$ identity matrix, respectively. $\boldsymbol{e}_i \in \mathbb{B}^M$ is the i-th column of $\boldsymbol{1}_M$, where M is given by the context and $\mathbb{B} = \{0, 1\}$.

II. SYSTEM MODEL

Fig. 1 shows the MU-MISO block diagram of the downlink (FIR filters are square boxes, pure spatial filters are triangles) comprising a BS equipped with N_a antennas and B single-antenna MSs. The channel's impulse response of order[1] Q is

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

where $\boldsymbol{H}_q \in \mathbb{C}^{B \times N_a}$ and the k-th row of $\boldsymbol{H}[n]$ is the FIR vector channel from the BS to the k-th MS. Furthermore, the received signal is perturbed by temporally white and spatially uncorrelated Gaussian noise, i.e., $\boldsymbol{\eta}[n] \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{\eta}})$ with $\boldsymbol{R}_{\boldsymbol{\eta}} = \mathrm{diag}(\sigma_{\eta_1}^2, \ldots, \sigma_{\eta_B}^2)$ and $\mathrm{E}[\boldsymbol{\eta}[n]\boldsymbol{\eta}^{\mathrm{H}}[n+\nu]] = \boldsymbol{R}_{\boldsymbol{\eta}}\delta[\nu]$.

For THP, *modulo operators* M($\bullet$) are deployed at the receivers to increase the degrees of freedom for the transmitter, since they map any element of a coset of the corresponding lattice to its representative in the *fundamental Voronoi region* (see e.g., [12], [20]). The transmitter chooses the element of the coset by means of the modulo feedback loop comprising the spatial feedback filter $\boldsymbol{F}$ (lower triangular with zero main diagonal to avoid a delay-free loop) and the temporal feedback filter $\boldsymbol{T}[n]$ (strictly causal to ensure realizability) of order N_{T}

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

with $\boldsymbol{T}_i \in \mathbb{C}^{B \times B}$. Since the output $\boldsymbol{v}[n]$ of the feedback loop is computed successively, the MSE of THP can be further minimized by reordering the symbols. Correspondingly, the data vector $\boldsymbol{s}[n]$ is first processed by the permutation matrix

$$\boldsymbol{\Pi}_\nu = \sum_{i=1}^{B} \boldsymbol{e}_i \boldsymbol{e}_{b_i}^{\mathrm{T}} \in \mathbb{B}^{B \times B}, \tag{2}$$

where $\boldsymbol{e}_i, \boldsymbol{e}_{b_i} \in \mathbb{B}^B$ and the B-tupel $\mathcal{O} = \{b_1, \ldots, b_B\}$ with $b_i \in \{1, \ldots, B\}$ is the precoding order. In other words, the i-th entry of $\boldsymbol{v}[n]$ corresponds to the b_i-th entry of $\boldsymbol{s}[n]$. The signal $\boldsymbol{v}[n]$ is passed through the feedforward filter of order L

$$\boldsymbol{P}[n] = \sum_{\ell=0}^{L} \boldsymbol{P}_\ell \delta[n-\ell] \quad \text{with} \quad \boldsymbol{P}_\ell \in \mathbb{C}^{N_a \times B} \tag{3}$$

to get the transmit signal $\boldsymbol{y}[n] \in \mathbb{C}^{N_a}$. Note (e.g., [15]) that $\mathrm{E}[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+\nu]] = \sigma_v^2 \boldsymbol{1}_B \delta[\nu]$ due to the modulo operator.

The quantizers Q($\bullet$) at the receivers map the estimate $\hat{\boldsymbol{s}}[n] \in \mathbb{C}^B$ to the modulation alphabet, i.e., $\tilde{\boldsymbol{s}}[n], \boldsymbol{s}[n] \in \mathbb{A}^B$. The common weight β of all receivers is only introduced for notational simplicity, as we will see in Section III.

The impulse responses of the FIR equalizers of order F at the receivers are comprised in

$$\boldsymbol{G}[n] = \sum_{f=0}^{F} \boldsymbol{G}_f \delta[n-f].$$

Note that the taps $\boldsymbol{G}_f \in \mathbb{C}^{B \times B}$, $f = 0, \ldots, F$, are diagonal, since the receivers cannot cooperate.
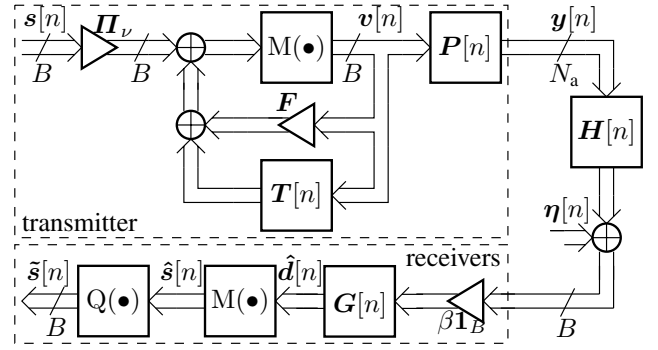

Fig. 1. System model for THP with FIR equalizer $\boldsymbol{G}[n]$.
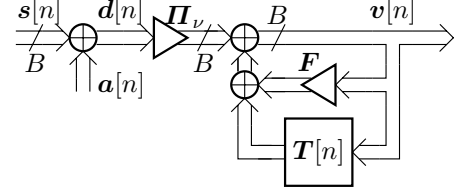

Fig. 2. Linear representation of the modulo operator.

Since any vector $\boldsymbol{a}' \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$ can be added to the modulo input without changing its output, i.e., M($\boldsymbol{x} + \boldsymbol{a}'$) = M($\boldsymbol{x}$), the transmitter can choose a desired value for this shift of the received signal (e.g., [12]). This shift of the received signal is removed by the modulo operators at the receivers and its desired value is created by the modulo operator in the feedback loop at the transmitter. With this understanding, we can interpret the linear representation in Fig. 2, where the signal $\boldsymbol{a}[n] \in \tau\mathbb{Z}^B + \mathrm{j}\tau\mathbb{Z}^B$ added by the transmitter's modulo operator has been moved to the front of the feedback loop. The data signal $\boldsymbol{s}[n]$ is the desired value of the modulo outputs at the receivers. As the signal portion $\boldsymbol{a}[n]$ is removed by the receivers' modulo operators, the signal $\boldsymbol{d}[n]$ is the desired signal for the *inputs* of the receivers' modulo operators.

From Fig. 2, we have that

$$\boldsymbol{d}[n] = \boldsymbol{\Pi}_\nu^{\mathrm{T}}(\boldsymbol{1}_B - \boldsymbol{F})\boldsymbol{v}[n] - \boldsymbol{\Pi}_\nu^{\mathrm{T}} \sum_{i=0}^{N_{\mathrm{T}}} \boldsymbol{T}_i \boldsymbol{v}[n-i]. \tag{4}$$

The estimate for $\boldsymbol{d}[n]$, i.e., the input of the receivers' modulo operators, can be written as (see Fig. 1)

$$\hat{\boldsymbol{d}}[n] = \beta \boldsymbol{G}[n] * \boldsymbol{H}[n] * \boldsymbol{P}[n] * \boldsymbol{v}[n] + \beta \boldsymbol{G}[n] * \boldsymbol{\eta}[n].$$

Alternatively, the estimate is given by

$$\hat{\boldsymbol{d}}[n] = \beta \sum_{i=0}^{F+Q+L} \boldsymbol{S}^{(i)} \boldsymbol{GHP}\boldsymbol{v}[n-i] + \beta \sum_{i=0}^{F} \tilde{\boldsymbol{S}}^{(i)} \tilde{\boldsymbol{G}} \boldsymbol{\eta}[n-i], \tag{5}$$

where we introduced the selection matrices

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

---

[1]Note that $Q_k = Q, \forall k$ for notational simplicity.

and put the coefficients of the filters into the matrices

$$\boldsymbol{P} = \left[\boldsymbol{P}_0^{\mathrm{T}}, \ldots, \boldsymbol{P}_L^{\mathrm{T}} \right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{a}}(L+1) \times B},$$

$$\boldsymbol{H} = \sum_{q=0}^{Q} [\boldsymbol{0}_{L+1 \times q}, \boldsymbol{1}_{L+1}, \boldsymbol{0}_{L+1 \times Q-q}]^{\mathrm{T}} \otimes \boldsymbol{H}_q,$$

$$\boldsymbol{G} = \sum_{f=0}^{F} [\boldsymbol{0}_{Q+L+1 \times f}, \boldsymbol{1}_{Q+L+1}, \boldsymbol{0}_{Q+L+1 \times F-f}]^{\mathrm{T}} \otimes \boldsymbol{G}_f, \text{ and}$$

$$\tilde{\boldsymbol{G}} = \left[\boldsymbol{G}_0^{\mathrm{T}}, \ldots, \boldsymbol{G}_F^{\mathrm{T}} \right]^{\mathrm{T}} \in \mathbb{C}^{B(F+1) \times B}.$$

The block Toeplitz convolution matrices for the channel and the non-cooperative equalizers are $\boldsymbol{H} \in \mathbb{C}^{B(Q+L+1) \times N_{\mathrm{a}}(L+1)}$ and $\boldsymbol{G} \in \mathbb{C}^{B(F+Q+L+1) \times B(Q+L+1)}$, respectively. Note that $\boldsymbol{S}^{(i)} \boldsymbol{G} \boldsymbol{H} \boldsymbol{P}$ is the tap of $\boldsymbol{G}[n] * \boldsymbol{H}[n] * \boldsymbol{P}[n]$ with delay i.

For linear precoding (no modulo operators, no feedback loop: $\mathrm{M}(\boldsymbol{x}) = \boldsymbol{x}$, $\boldsymbol{\Pi}_\nu = \boldsymbol{1}_B$, $\boldsymbol{T}[n] = \boldsymbol{0}\delta[n]$, and $\boldsymbol{F} = \boldsymbol{0}$), (5) is also applicable, but with $\boldsymbol{v}[n] = \boldsymbol{s}[n]$ and $\hat{\boldsymbol{d}}[n] = \hat{\boldsymbol{s}}[n]$.

III. JOINT TRANSMIT AND RECEIVE FILTER DESIGN

A. Problem Formulation

We try to find the precoding filters $\boldsymbol{P}[n]$, $\boldsymbol{F}$, $\boldsymbol{T}[n]$, the precoding order $\mathcal{O}$ (represented by $\boldsymbol{\Pi}_\nu$), the equalizers $\boldsymbol{G}[n]$, and the latency time ν (time difference between application of precoder and decision at receivers) minimizing the MSE[2]

$$\varepsilon(\boldsymbol{P}, \boldsymbol{F}, \boldsymbol{T}_1, \ldots, \boldsymbol{T}_{N_{\mathrm{T}}}, \beta, \nu, \mathcal{O}, \boldsymbol{G}) = \mathrm{E}\left[\|\boldsymbol{d}[n-\nu] - \hat{\boldsymbol{d}}[n]\|_2^2 \right] \tag{6}$$

under the total transmit power constraint

$$\mathrm{E}[\|\boldsymbol{y}[n]\|_2^2] = \sigma_{\boldsymbol{v}}^2 \, \mathrm{tr}(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}) = E_{\mathrm{tr}}. \tag{7}$$

Besides minimizing the MSE subject to above constraint, we must ensure the structural properties of $\boldsymbol{F}$ (strictly lower triangular) and $\boldsymbol{G}$ (matrix with diagonal blocks).

Unfortunately, the MSE $\varepsilon(\ldots)$ is non-convex in $\boldsymbol{P}$, $\boldsymbol{F}$, $\boldsymbol{T}_i, i = 1, \ldots, N_{\mathrm{T}}$, β, $\boldsymbol{G}$ and no closed-form solution can be found (this is also true for linear precoding). Therefore, we must resort to iterative techniques. Since the solution of the optimization for fixed $\boldsymbol{G}$ can be solved as in [12] (see Subsection III-B) and $\boldsymbol{G}$ can easily be found (see Subsection III-C and e.g., [21]), when fixing the other variables, we propose to apply alternating optimization to find the precoder and equalizer filters, as depicted in Fig. 3. Based on an initial value $\boldsymbol{G}^{(0)}[n]$ for the equalizers, the transmitter and the receivers are updated alternately. During a transmitter update, the feedforward filter $\boldsymbol{P}[n]$, the feedback filters $\boldsymbol{T}[n]$ and $\boldsymbol{F}$, the latency time ν, and the precoding order $\mathcal{O}$ are recomputed based on the fixed receivers $\boldsymbol{G}[n]$. Additionally, a new value for the common weight β of the receivers is computed which is necessary to have a closed-form solution for the precoding filters [12]. In a receiver update, $\boldsymbol{G}[n]$ is recomputed while keeping all the other variables fixed. The iteration is terminated after the maximum number of updates i_{id} or when the relative MSE decrease is below the threshold c_{th}.

[2]For linear precoding: $\varepsilon(\boldsymbol{P}, \beta, \nu, \boldsymbol{G}) = \mathrm{E}[\|\boldsymbol{s}[n-\nu] - \hat{\boldsymbol{s}}[n]\|_2^2]$.
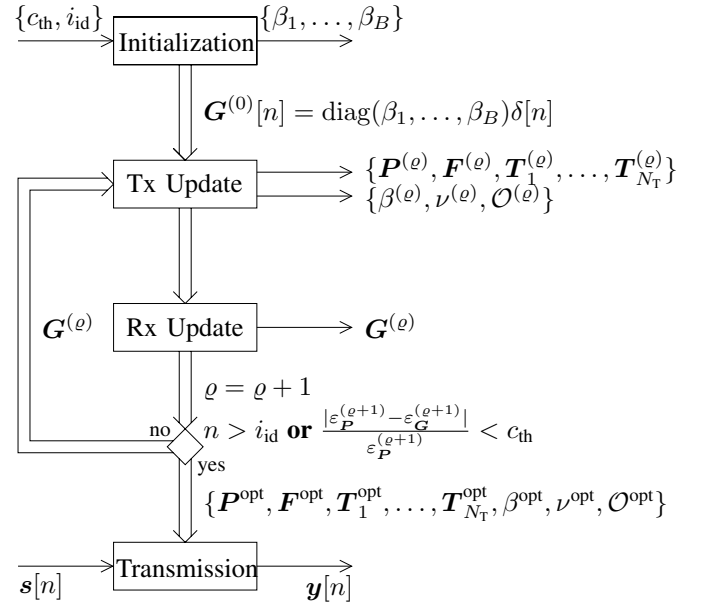

Fig. 3. Schematic of the alternating optimization for the joint transmit and receive filter computation.

Note that the iteration is completely performed at the transmitter. The receivers are only updated virtually. Once the optimum transmit filters and parameters are determined, the transmitter can transmit pilot symbols which enable the receivers to design the equalizers based on the MMSE criterion without the explicit knowledge of the precoding filters.

B. Update of Transmitter

When recomputing the precoding filters, we form an equivalent channel $\boldsymbol{H}_{\mathrm{equ}} = \boldsymbol{G}[n] * \boldsymbol{H}[n]$ with the equivalent noise $\boldsymbol{\eta}_{\mathrm{equ}} = \boldsymbol{G}[n] * \boldsymbol{\eta}[n]$. Then, the MMSE optimization of the precoding filters is performed as in [12] based on $\boldsymbol{H}_{\mathrm{equ}}$ and $\boldsymbol{\eta}_{\mathrm{equ}}$. We do not show the derivation of the THP filters as in [12] due to space limitations. Instead, we refer to [12] for the development of the efficient algorithm to compute the precoding filters based on a symmetrically permuted Cholesky factorization shown in Table I. The matrix $\boldsymbol{H}_{\mathrm{equ}} \in \mathbb{C}^{B(F+Q+L+1) \times N_{\mathrm{a}}(L+1)}$ is defined similarly to $\boldsymbol{H}$ but contains the coefficients of $\boldsymbol{H}_{\mathrm{equ}}[n]$ instead of $\boldsymbol{H}[n]$. The constant $\xi = \mathrm{tr}(\tilde{\boldsymbol{G}}\boldsymbol{R}_{\boldsymbol{\eta}}\tilde{\boldsymbol{G}}^{\mathrm{H}})/E_{\mathrm{tr}}$ is the reciprocal SNR.

For linear precoding we get via Lagrangian multipliers

$$\boldsymbol{P}_{\mathrm{WF}} = \frac{1}{\beta_{\mathrm{WF}}} \boldsymbol{H}_{\mathrm{equ}}^{\mathrm{H}} \left(\boldsymbol{H}_{\mathrm{equ}} \boldsymbol{H}_{\mathrm{equ}}^{\mathrm{H}} + \xi \boldsymbol{1}_{B(F+Q+L+1)} \right)^{-1} \boldsymbol{S}^{(\nu_{\mathrm{WF}})}.$$

The scalar β_{WF} follows from $\sigma_s^2 \, \mathrm{tr}(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}) = E_{\mathrm{tr}}$. By plugging the filter solution into the MSE, ν_{WF} can be found by minimizing the resulting expression (see [12]).

C. Update of Receivers

For the update of the receivers, we need a new system model. The input of the modulo operator of the k-th receiver can be expressed as

$$\hat{d}_k[n] = \beta g_k[n] * \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{H}[n] * \boldsymbol{P}[n] * \boldsymbol{v}[n] + \beta g_k[n] * \eta_k[n].$$

TABLE I
ALGORITHM FOR TRANSMITTER UPDATE.

```
1:   Θ ← (H_equ H_equ^H + ξ1_B(F+Q+L+1))^{-1}
     D ← 0_B(F+Q+L+1)×B(F+Q+L+1)
3:   ε_min ← ∞
4:   for ν = F + Q + L, ..., L:
         Π_ν ← 1_B
         ε ← 0
         for i = B, ..., 1:
             k ← Bν + i
9:           b_i ← argmin  Θ(Bν + b, Bν + b)
                    b∈{1,...,i}
             Π_ν ← Π_ν with rows b_i and i exchanged
             Π ← 1_B(F+Q+L+1) with rows Bν + b_i and k exch.
             Θ ← ΠΘΠ^T
13:          D(k,k) ← Θ(k,k)
             Θ(k,1:k) ← Θ(k,1:k)/D(k,k)
15:          Θ(1:k-1,1:k-1) ← Θ(1:k-1,1:k-1)
                  −D(k,k)(Θ(k,1:k-1))^H Θ(k,1:k-1)
             ε ← ε + D(k,k)
17:      if ε < ε_min:
             ν_WF ← ν
             ε_min ← ε
     L ← lower triangular part of Θ
21:  C ← blockdiag(1_B, ..., 1_B, Π_ν_WF, ..., Π_F+Q+L) H_equ
     P ← C^H L^H D S^(ν_WF),T
     β_WF ← √(σ_v^2 ||P||_F^2 / E_tr)
     P_WF ← P/β_WF
     F_WF ← 1_B − (S^(ν_WF) L S^(ν_WF),T)^{-1}
     for i = 1, ..., F + Q + L − ν_WF:
         T_i ← −Π_ν_WF Π_ν_WF+i^T S^(ν_WF+i) CP
```

In matrix-vector notation, we have

$$\hat{d}_k[n] = \beta \boldsymbol{g}_k^T \boldsymbol{A}_k \boldsymbol{v}'[n] + \beta \boldsymbol{g}_k^T \boldsymbol{\eta}_k[n], \quad (8)$$

where the coefficients of $g_k[n] = \sum_{f=0}^F g_{k,f}\delta[n-f]$ are collected in $\boldsymbol{g}_k = [g_{k,0}, \ldots, g_{k,F}]^T \in \mathbb{C}^{F+1}$ and

$$\boldsymbol{A}_k = \sum_{q=0}^{Q+L} [\boldsymbol{0}_{F+1\times q}, \boldsymbol{1}_{F+1}, \boldsymbol{0}_{F+1\times Q+L-q}] \otimes \boldsymbol{a}_{k,q}^T$$

is the $F + 1 \times B(F + Q + L + 1)$ block Toeplitz convolution matrix corresponding to $\boldsymbol{a}_k^T[n] = \boldsymbol{e}_k^T \boldsymbol{H}[n] * \boldsymbol{P}[n]$, i.e., the combination of the k-th channel and the feedforward filter. Furthermore, $\boldsymbol{\eta}_k[n] = [\eta_k[n], \ldots, \eta_k[n-F]]^T$ and $\boldsymbol{v}'[n] = [\boldsymbol{v}^T[n], \ldots, \boldsymbol{v}^T[n-F-Q-L]]^T$, whose covariance matrix is $\sigma_v^2 \boldsymbol{1}_{B(F+Q+L+1)}$. The desired value for the modulo input $\hat{d}_k[n]$ of receiver k is

$$d_k[n-\nu] = \boldsymbol{e}_k^T \boldsymbol{\Pi}_\nu^T \left((\boldsymbol{1}_B - \boldsymbol{F}) \boldsymbol{S}^{(\nu)} - \sum_{i=1}^{N_T} \boldsymbol{T}_i \boldsymbol{S}^{(i+\nu)} \right) \boldsymbol{v}'[n].$$

The MMSE equalizers result from

$$\boldsymbol{g}_{WF,k}^T = \operatorname*{argmin}_{\boldsymbol{g}^T} E\left[|d_k[n-\nu] - \hat{d}_k[n]|^2 \right]. \quad (9)$$

Setting the derivative of the MSE cost function to zero leads to

$$\boldsymbol{g}_{WF,k}^T = \beta^{-1} \boldsymbol{x}_k^T \boldsymbol{A}_k^H \left(\boldsymbol{A}_k \boldsymbol{A}_k^H + \xi_k \boldsymbol{1}_{F+1} \right)^{-1} \quad (10)$$

with $\boldsymbol{x}_k^T = \boldsymbol{e}_k^T \boldsymbol{\Pi}_\nu^T ((\boldsymbol{1}_B - \boldsymbol{F})\boldsymbol{S}^{(\nu)} - \sum_{i=1}^{N_T} \boldsymbol{T}_i \boldsymbol{S}^{(i+\nu)})$ and $\xi_k = \sigma_{\eta_k}^2/\sigma_v^2$. Not surprisingly, (10) contains the term β^{-1}. We see

that the scalar β needed for the precoder update drops out after the equalizer update.

From (10), we see that the k-th receiver must have the means to estimate $\boldsymbol{x}_k^T \boldsymbol{A}_k^H$ to be able to design its equalizer after the transmitter has obtained the precoding filters from the iteration. Thus, the training symbols must be precoded accordingly.

Clearly, above expression (10) for the equalizer is also valid for linear precoding, but we have to perform following substitutions: $\sigma_v^2 \rightarrow \sigma_s^2$ and $\boldsymbol{x}_k \rightarrow \boldsymbol{S}^{(\nu),T} \boldsymbol{e}_k$.

D. Initialization of Algorithm

The most important property of a useful initialization is that none of the equalizers may be zero, i.e., $g_k[n] \neq 0\delta[n], \forall k$. A very simple initialization would be $\boldsymbol{G}^{(0)}[n] = \boldsymbol{1}_B\delta[n]$. However, we observed that a ZF initialization leads to a considerable reduction of necessary iterations. Via the closed-form ZF solution presented in the sequel, we find the weights $\beta_1, \ldots, \beta_B$ which are used to initialize the equalizers as $\boldsymbol{G}^{(0)}[n] = \boldsymbol{B}\delta[n] = \operatorname{diag}(\beta_1, \ldots, \beta_B)\delta[n]$. With $\beta = 1$ and $\boldsymbol{G}[n] = \boldsymbol{G}^{(0)}[n]$, we get for the estimate

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

Due to the ZF constraint, the MSE simplifies to

$$E\left[||\boldsymbol{d}[n-\nu] - \hat{\boldsymbol{d}}[n]||_2^2 \right] = \operatorname{tr}(\boldsymbol{B}\boldsymbol{R}_\eta \boldsymbol{B}^H) = \sum_{k=1}^B |\beta_k|^2 \sigma_{\eta_k}^2.$$

Consequently, the optimization problem can be stated as

$$\{\boldsymbol{P}_{ZF}, \boldsymbol{F}_{ZF}, \boldsymbol{T}_{ZF,1}, \ldots, \boldsymbol{T}_{ZF,N_T}, \boldsymbol{B}_{ZF}, \nu_{ZF}, \mathcal{O}_{ZF}\} =$$
$$\operatorname*{argmin}_{\{\boldsymbol{P},\boldsymbol{F},\boldsymbol{T}_1,\ldots,\boldsymbol{T}_{N_T},\boldsymbol{B},\nu,\mathcal{O}\}} \operatorname{tr}(\boldsymbol{B}\boldsymbol{R}_\eta \boldsymbol{B}^H)$$

subject to the transmit power constraint $E[||\boldsymbol{y}[n]||_2^2] = \sigma_v^2 \operatorname{tr}(\boldsymbol{P}\boldsymbol{P}^H) = E_{tr}$, a constraint on the structure of $\boldsymbol{F}$, i.e., $\boldsymbol{S}_i \boldsymbol{F} \boldsymbol{e}_i = \boldsymbol{0}_i$, $\forall i = 1, \ldots, B$, a constraint on the structure of $\boldsymbol{B}$, i.e., $\boldsymbol{B} = \sum_{k=1}^B \beta_k \boldsymbol{e}_k \boldsymbol{e}_k^T$, and the ZF condition

$$\boldsymbol{H}\boldsymbol{P} = \boldsymbol{S}^{(\nu),T}\boldsymbol{B}^{-1}\boldsymbol{\Pi}_\nu^T(\boldsymbol{1}_B - \boldsymbol{F}) - \sum_{i=1}^{N_T} \boldsymbol{S}^{(\nu+i),T}\boldsymbol{B}^{-1}\boldsymbol{\Pi}_\nu^T \boldsymbol{T}_i.$$

The resulting ZF transmit filter can be written as

$$\boldsymbol{F} = \boldsymbol{1}_B - \boldsymbol{\Pi}_\nu \boldsymbol{B}\boldsymbol{S}^{(\nu)}\boldsymbol{H}\boldsymbol{P},$$
$$\boldsymbol{T}_i = \boldsymbol{\Pi}_\nu \boldsymbol{B}\boldsymbol{S}^{(\nu+i)}\boldsymbol{H}\boldsymbol{P}, \quad \forall i = 1, \ldots, N_T = Q + L - \nu, \text{ and}$$
$$\boldsymbol{P} = \sum_{i=1}^B \beta_{b_i}^{-1} \left(\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})} \boldsymbol{S}^{(\nu)}\boldsymbol{H} \right)^+ \boldsymbol{e}_{b_i} \boldsymbol{e}_i^T$$

with $\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})} = \boldsymbol{\Pi}_\nu^T \boldsymbol{S}_i^T \boldsymbol{S}_i \boldsymbol{\Pi}_\nu$. The B different receiver weights read as

$$|\beta_{b_i}| = \sqrt[4]{\xi_{b_i}^{-1}\alpha_{\nu,i,b_i}^{(\mathcal{O})}} \sqrt{\frac{\sum_{m=1}^B \sigma_v^2 \sqrt{\xi_{b_m}\alpha_{\nu,i,b_m}^{(\mathcal{O})}}}{E_{tr}}}$$

with

$$\alpha_{\nu,i,b_i}^{(\mathcal{O})} = \left\| \left(\boldsymbol{\Pi}_{\nu,i}^{(\mathcal{O})} \boldsymbol{S}^{(\nu)} \boldsymbol{H} \right)^{+} \boldsymbol{e}_{b_i} \right\|_2^2,$$

$$b_i = \operatorname*{argmin}_{\ell \in \{1,\ldots,B\} \setminus \{b_{i+1},\ldots,b_B\}} \alpha_{\nu,i,\ell}^{(\mathcal{O})}, \quad \forall i = B, B-1,\ldots,1.$$

$$(11)$$

The precoding order optimization has to be performed for all possible latency times ν to find the optimal latency time. However, we set $\nu_{\text{fix}} = L$ following the arguments in [12].

E. Convergence Analysis

Let the MSE be denoted by $\varepsilon(\mathcal{P}, \mathcal{G})$, where $\mathcal{P}$ represents the variables recomputed in the transmitter update and $\mathcal{G}$ that recomputed in the receiver update. In the transmitter update at the ϱ-th iteration, the variables in $\mathcal{P}^{(\varrho)}$ are chosen to minimize the MSE for a given $\mathcal{G}^{(\varrho-1)}$. Thus, $\varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho-1)}) \leq \varepsilon(\mathcal{P}, \mathcal{G}^{(\varrho-1)})$ for any $\mathcal{P}$ and in particular,

$$\varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho-1)}) \leq \varepsilon(\mathcal{P}^{(\varrho-1)}, \mathcal{G}^{(\varrho-1)}).$$

Similarly, the variables in $\mathcal{G}^{(\varrho)}$ are chosen to minimize the MSE for fixed $\mathcal{P}^{(\varrho)}$. We can follow that $\varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho)}) \leq \varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G})$ for any $\mathcal{G}$ and therefore,

$$\varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho)}) \leq \varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho-1)}).$$

Combining the two results, we see that the MSE is monotonically decreasing during the iteration:

$$\varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho)}) \leq \varepsilon(\mathcal{P}^{(\varrho)}, \mathcal{G}^{(\varrho-1)}) \leq \varepsilon(\mathcal{P}^{(\varrho-1)}, \mathcal{G}^{(\varrho-1)}).$$

Since the MSE is lower bounded, i.e., it is non-negative, our proposed iterative algorithm converges.[3]

IV. SIMULATION RESULTS

We now want to analyze our proposed signal processing algorithm in terms of its BER over different SNR values by using Monte-Carlo computer simulations. The BERs as seen in the plots are obtained by averaging over 10000 different channel realizations. The channel taps $\boldsymbol{H}_q$ were determined by the 3GPP *spatial channel model* (SCM) for MIMO systems in urban micro-cell environments and the 3GPP 'Pedestrian A' power delay profile [23], so that $\sum_{q=0}^{Q} \mathrm{E}[\mathrm{tr}(\boldsymbol{H}_q \boldsymbol{H}_q^{\mathrm{H}})] = B N_{\text{a}}$. We "transmitted" 100 symbols per user over each channel within an SNR range of $[-10\text{dB}, 25\text{dB}]$. The signal-to-noise ratio is defined as $\text{SNR} = \frac{E_{\text{tr}}}{\text{tr}(\boldsymbol{R}_{\boldsymbol{\eta}})}$. The iterative scheme is abandoned when $|\varepsilon_{\boldsymbol{P}}^{(\varrho+1)} - \varepsilon_{\boldsymbol{G}}^{(\varrho+1)}|/\varepsilon_{\boldsymbol{P}}^{(\varrho+1)} < 0.001$, i.e., the *iteration depth* (ID) is variable. The remaining parameters are: $B = 4$ users, $N_{\text{a}} = 4$ transmit antennas, $Q + 1 = 4$ multipaths, filter order $L = F = 4$, and 16-QAM symbols with unit variance. Fig. 4 displays the simulation results in the most general case. Our proposed signal processing algorithm ("RxSP, ID=var.") is compared to the case of pure precoding,
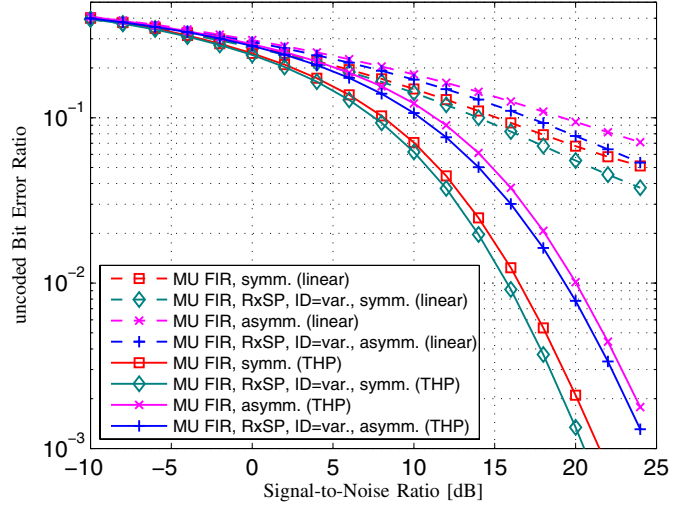

Fig. 4. Performance improvements for different multi-user (MU) scenarios in a frequency-selective (FIR) channel environment. The proposed iterative algorithm, both linear and nonlinear, shows considerable gains for symmetric and asymmetric channels.
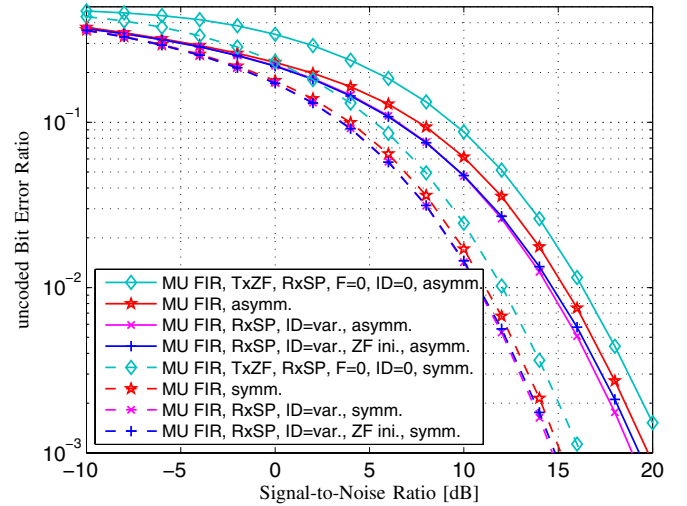

Fig. 5. BER over SNR performance for different initializations of THP. Note that $N_{\text{a}} = 7$ to ensure the existence of the ZF solution. The performance of the closed-form ZF solution, which is used for initialization, is also plotted.

i.e., $\boldsymbol{G}[n] = \mathbf{1}_B \delta[n]$ and $\varrho = 0$. Both linear and nonlinear filters at the transmitter are treated. We observe gains between 1dB and 2dB for the linear and nonlinear case (at a BER of 10^{-1} which we assume to be the target uncoded BER for coded transmission), respectively, both for symmetric and asymmetric channels.[4] In the latter, the multi-paths of two users are ten times stronger than those of the two others without violating $\sum_{q=0}^{Q} \mathrm{E}[\mathrm{tr}(\boldsymbol{H}_q \boldsymbol{H}_q^{\mathrm{H}})] = B N_{\text{a}}$. We want to point out, that in the nonlinear case, the choice of the initialization slightly changes the BER (see Fig. 5). A ZF initialization on the one hand decreases the number of necessary iterations by a

---

[3]At this point, it has not been proven yet, whether the algorithm converges to the global minimum or not. In the case of pure linear processing, the algorithm *should* converge to the global minimum following considerations similar to those in [22]. In the case of nonlinear precoding, a convergence to the global minimum cannot be guaranteed due to the *suboptimal* optimization of the precoding ordering in (11).

[4]This still holds true when $Q = 0$, i.e., the channel is frequency flat. This is in contrast to results in [11], where a spatially uncorrelated channel model was used and considerable gains were only observed for asymmetric channels.
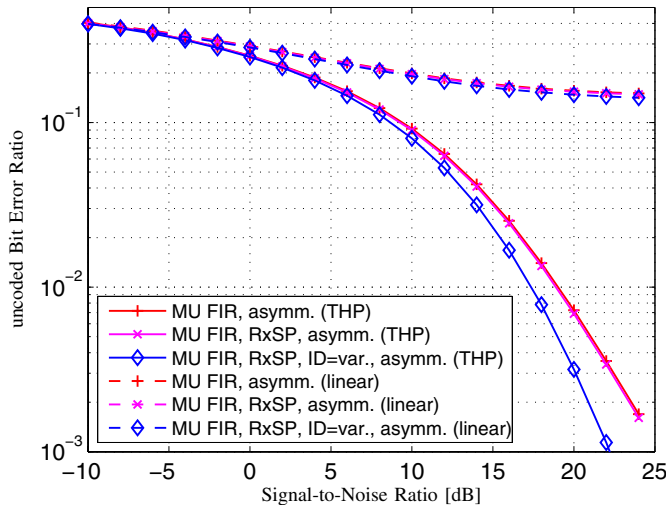
Fig. 6. The linear scheme is not capable to operate more users than transmit antennas available at the base station. For the nonlinear scheme, we observe a huge gain, although the system is overloaded. This is intercessional for iterative, nonlinear signal processing algorithms.

factor of 10 (not shown here), but on the other hand results in a slightly worse performance compared to an initialization with an identity matrix. For linear transmit filters, the iteration depth can be cut down by a factor of 3 without degradation. In Fig. 5, we also compare our proposed algorithm to zero-forcing THP without iteration, i.e., the initialization solution. Although we allowed for different receiver weights for different users, zero-forcing THP with scalar receivers is outperformed by all MMSE approaches over the whole SNR range.

Last but not least, we want to show the superiority of *nonlinear, iterative* algorithms in systems, that are overloaded (see Fig. 6). To this end, we transmit to more users than antennas available at the transmitter ($B = 5$, $N_a = 4$). Fig. 6 also shows the results for the case, where the receivers design their equalizers based on the given precoder after only one iteration ("RxSP" without "ID"). As we can see, the linear filters are not capable of operating such a scenario. Contrary, the nonlinear iterative approach does not collapse, it comes along with a huge gain, also compared to systems without enhanced receivers or only one iteration.

V. CONCLUSIONS

We presented an iterative algorithm to compute jointly optimized precoders and FIR equalizers for frequency selective MU-MISO channels. We mainly discussed the case of non-linear precoding (THP), since it is harder to solve, but also gave the solutions for the linear case. We proved the convergence of the iterative algorithm and demonstrated that considerable gains (up to 2dB) compared to state-of-the-art precoding approaches can be achieved by the proposed scheme. Furthermore, we suggested a ZF initialization, that dramatically cuts down the number of necessary iterations.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
[2] R. Irmer and G. Fettweis, "Combined Transmitter and Receiver Optimization for Multiple Antenna Frequency-Selective Channels," in *Proc. WPMC 2002*, October 2002, pp. 412–416.
[3] B. R. Vojcic and W. M. Jang, "Transmitter Precoding in Synchronous Multiuser Communications," *IEEE Transactions on Communications*, vol. 46, no. 10, pp. 1346–1355, October 1998.
[4] M. Brandt-Pearce and A. Dharap, "Transmitter-Based Multiuser Interference Rejection for the Down-Link of a Wireless CDMA System in a Multipath Environment," *IEEE Journal on Selected Areas of Communication*, vol. 18, no. 3, pp. 407–417, March 2000.
[5] M. Meurer, P. W. Baier, T. Weber, Y. Lu, and A. Papathanassiou, "Joint transmission: advantageous downlink concept for CDMA mobile radio systems using time division duplexing," *Electronics Letters*, vol. 36, no. 10, pp. 900–901, May 2000.
[6] M. Meurer, H. Tröger, T. Weber, and P. W. Baier, "Synthesis of joint detection and joint transmission in CDMA downlinks," *Electronics Letters*, vol. 37, no. 14, pp. 919–920, July 2001.
[7] A. Noll Barreto and G. Fettweis, "Joint Signal Precoding in the Downlink of Spread Spectrum Systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 3, pp. 511–518, May 2003.
[8] M. Joham, W. Utschick, and J. A. Nossek, "Latency Time Optimization for FIR and Block Transmit Filters," in *Proc. ISSPA 2003*, vol. 1, July 2003, pp. 273–276.
[9] L. Choi and R. D. Murch, "Transmit-Preprocessing Techniques With Simplified Receivers for the Downlink of MISO TDD-CDMA Systems," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 2, pp. 285–295, March 2004.
[10] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, August 2005.
[11] R. Hunger, W. Utschick, D. Schmidt, and M. Joham, "Alternating Optimization for MMSE Broadcast Precoding," in *Proc. ICASSP 2006*, vol. IV, May 2006, pp. IV–757–IV–760.
[12] M. Joham, D. A. Schmidt, J. Brehmer, and W. Utschick, "Finite-Length MMSE Tomlinson-Harashima Precoding for Frequency Selective Vector Channels," 2006, *Accepted for publication in IEEE Transactions on Signal Processing*.
[13] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Efficient Tomlinson-Harashima Precoding for Spatial Multiplexing on Flat MIMO Channel," in *Proc. ICC 2005*, vol. 3, May 2005, pp. 2021–2025.
[14] R. F. H. Fischer, "Sorted Spectral Factorization of Matrix Polynomials in MIMO Communications," *IEEE Transactions on Communications*, vol. 53, no. 6, pp. 945–951, June 2005.
[15] ——, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, 2002.
[16] 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.
[17] M. Meurer, T. Weber, and W. Qiu, "Transmit Nonlinear Zero Forcing: Energy Efficient Receiver Oriented Transmission in MIMO CDMA Mobile Radio Downlinks," in *Proc. ISSSTA 2004*, September 2004, pp. 260–269.
[18] A. P. Liavas, "Tomlinson-Harashima Precoding With Partial Channel Knowledge," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 5–9, January 2005.
[19] F. A. Dietrich and W. Utschick, "Robust Tomlinson-Harashima Precoding," in *Proc. PIMRC '05*, September 2005.
[20] D. Schmidt, M. Joham, and W. Utschick, "Minimum Mean Square Error Vector Precoding," in *Proc. PIMRC '05*, vol. 1, September 2005, pp. 107–111.
[21] L. L. Scharf, *Statistical Signal Processing*. Addison-Wesley, 1991.
[22] R. Hunger, W. Utschick, D. A. Schmidt, and M. Joham, "On linear and nonlinear precoders for decentralized receivers." 2006, *in preparation*.
[23] 3 GPP TSG-RAN-WG1, "Spacial channel model for Multiple Input Multiple Output (MIMO) simulations," 3GPP, Tech. Rep. TR 25.996, August 2003.