Joint Optimization of Pilot Assisted Channel Estimation and Equalization applied to Space-Time Decision Feedback Equalization

Frank A. Dietrich, Michael Joham, and Wolfgang Utschick
Institute for Circuit Theory and Signal Processing, Munich University of Technology,
Arcisstrasse 21, 80290 Munich, Germany, E-Mail: {Dietrich, Joham, Utschick}@nws.ei.tum.de

***Abstract*— Traditionally, channel estimation and equalization are optimized separately and independently: The channel estimate is simply plugged into the equalizer as if it had no errors. We propose a new method for joint pilot symbol assisted channel estimation and equalization and apply it to the design of the space-time decision feedback equalizer. The explicit solution of this joint approach is obtained with the same order of complexity as a separate design and results in a significant performance improvement. Furthermore, we discuss its relation to robust optimization and regularization techniques for equalizer design.**

***Index Terms*— decision feedback equalization, channel estimation, multiuser detection, adaptive antennas**

I. Introduction

A wireless communication link with a frequency selective channel, multiple transmit (no cooperation among transmitters), and multiple receive antennas is considered. Pilot symbols are available, e.g. time multiplexed, for channel estimation. The standard approach to optimize a receiver for equalization of a frequency selective MIMO channel with coherent detection is: 1) *Estimation of the channel parameters* with an *a priori* chosen method (e.g. maximum likelihood, linear minimum mean square error (LMMSE), correlator) from the received pilot sequence. 2) *Optimization of the equalizer* assuming that the channel parameters are perfectly known. 3) Application of the channel estimates to the equalizer design as if they were known without errors.

This approach of *separate optimization* of pilot symbol assisted channel estimation [1] and channel equalization, has the advantage that standard methods [2], [3] can be applied and combined. This advantage is obtained at the expense that synergies of joint optimization cannot be exploited resulting in a performance loss. Different channel estimation methods have to be investigated for a given equalization paradigm to find out the best combination. Finally, the assumption of perfect channel knowledge made for equalizer design is not justified.

As a consequence a *joint optimization* of pilot assisted channel estimation and equalization should be performed.[1] Some previous approaches to joint channel estimation and equalization are not based on the availability of a pilot sequence, but operate in a blind fashion (see e.g. a summary in [4]). They have a significantly increased complexity compared to the corresponding method of equalization based on perfect channel knowledge. Others solve the problem based on pilots but iteratively, e.g., [5], [6]. The channel estimate obtained from a pilot sequence can also be improved iteratively, e.g., to track the time-variant channel [7], [8]. Robust optimization of preequalization at the transmitter taking into account channel estimation errors is considered in [9], [10].

[1]This is true for a finite SNR and a finite number of pilot symbols. Asymptotically both optimization tasks will be separated.

Our novel approach to joint optimization follows the *Bayesian paradigm* minimizing the average cost function, where the expectation is taken w.r.t. the unknown channel parameters conditioned on the received pilot sequence (Sec. III). This general paradigm may be applied to equalizers, whose optimization problem can be translated into a cost function with constraints for which all parameters are known. In contrast to, e.g., [5], [6] it is *not* an iterative approach. As an example, we present the joint optimization of *space-time decision feedback equalization* (ST-DFE) for frequency selective MIMO channels (Sec. IV) [11], [12], [13]. Incorporating a latency time as well as spatial ordering the solution for the ST-DFE based on the minimum mean square error (MSE) is briefly stated in Sec. IV-A, which serves as a basis for applying the novel joint optimization (Sec. IV-B). The relation of this approach to regularization and stochastic programming is sketched in Sec. IV-C. The solution of the joint optimization has the *same order of complexity* as the standard design with separate optimization (Sec. IV-D), but leads to a *significant performance improvement*, as can be seen from the simulations in Sec. V. In the Appendix a brief derivation of the solution for the standard optimization approach of the ST-DFE is given, which is stated in Sec. IV-A. Also notation used in Sec. IV-A is introduced here (see [14] for more details).

Notation: Random vectors and matrices are denoted by lower and upper case sans serif bold letters, whereas the respective realizations or deterministic variables are italic. The operators $*$, $\mathrm{E}[\bullet]$, $\mathrm{E}_{\mathbf{a}}[\bullet]$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, and $\mathrm{tr}(\bullet)$ stand for convolution, a general expectation, expectation with respect to $\mathbf{a}$, transpose, Hermitian transpose, and trace of a matrix, respectively. $\boldsymbol{e}_i$ is the i-th column of an $N \times N$ identity matrix $\boldsymbol{I}_N$ and $\mathbf{0}_{M\times N}$ the $M \times N$ matrix of zeros.

II. System Model and Assumptions

Data symbols $\boldsymbol{s}_{\mathrm{d}}[n] \in \mathbb{B}^K$ from a set of QAM symbols $\mathbb{B}$ are transmitted from K antennas via a frequency selective

0-7803-8938-7/05/$20.00 (C) 2005 IEEE

MIMO channel $\boldsymbol{H}[n] = \sum_{\ell=0}^{L} \boldsymbol{H}_\ell \delta[n-\ell] \in \mathbb{C}^{M \times K}$ of order L (Fig. 1). The received signal of M antennas is

$$\mathbf{y}_\mathrm{d}[n] = \boldsymbol{H}[n] * \mathbf{s}_\mathrm{d}[n] + \mathbf{n}_\mathrm{d}[n] \in \mathbb{C}^M, \tag{1}$$

where $\mathbf{n}_\mathrm{d}[n] \sim \mathcal{N}_\mathrm{c}(\mathbf{0}, \boldsymbol{C}_\mathbf{n})$ is additive temporally white complex Gaussian noise with covariance matrix $\boldsymbol{C}_\mathbf{n} \in \mathbb{C}^{M \times M}$.

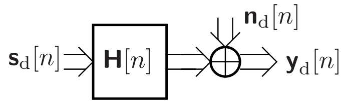

Fig. 1. Frequency selective MIMO channel.

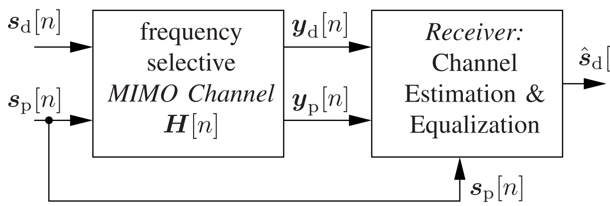

Fig. 2. Channel and Receiver Model

Pilot channel: N_p pilot symbols $\boldsymbol{s}_\mathrm{p}[n] \in \mathbb{B}^K$ with $n \in \{-L, \cdots, N_\mathrm{p}-1\}$ are time multiplexed including L guard symbols. The received pilot sequence $\mathbf{y}_\mathrm{p}[n]$ is defined similar to (1) and can be rewritten as

$$\mathbf{y}_\mathrm{p} = \boldsymbol{S}_\mathrm{p}\mathbf{h} + \mathbf{n}_\mathrm{p} \in \mathbb{C}^{MN_\mathrm{p}}, \tag{2}$$

where $\boldsymbol{S}_\mathrm{p} = \boldsymbol{S}_\mathrm{p}'^{\mathrm{T}} \otimes \boldsymbol{I}_M \in \mathbb{C}^{MN_\mathrm{p} \times (L+1)KM}$ is block Toeplitz, $\boldsymbol{S}_\mathrm{p}' \in \mathbb{C}^{K(L+1) \times N_\mathrm{p}}$ has block Toeplitz structure with first block row $[\boldsymbol{s}_\mathrm{p}[0], \cdots, \boldsymbol{s}_\mathrm{p}[N_\mathrm{p}-1]]$ and first column $[\boldsymbol{s}_\mathrm{p}[0]^\mathrm{T}, \cdots, \boldsymbol{s}_\mathrm{p}[-L]^\mathrm{T}]^\mathrm{T}$, $\mathbf{y}_\mathrm{p} = [\mathbf{y}_\mathrm{p}[0]^\mathrm{T}, \cdots, \mathbf{y}_\mathrm{p}[N_\mathrm{p}-1]^\mathrm{T}]^\mathrm{T}$, and $\mathbf{n}_\mathrm{p} = [\mathbf{n}_\mathrm{p}[0]^\mathrm{T}, \cdots, \mathbf{n}_\mathrm{p}[N_\mathrm{p}-1]^\mathrm{T}]^\mathrm{T}$. All channel coefficients are summarized in $\mathbf{h} = \mathrm{vec}([\mathbf{H}_0, \cdots, \mathbf{H}_L])$. The additive noise $\mathbf{n}_\mathrm{p}[n]$ is distributed as $\mathbf{n}_\mathrm{d}[n]$. The channel is assumed to be i.i.d. block fading and Gaussian $\mathbf{h} \sim \mathcal{N}_\mathrm{c}(\boldsymbol{\mu}_\mathbf{h}, \boldsymbol{C}_\mathbf{h})$ with mean $\boldsymbol{\mu}_\mathbf{h} = \mathrm{E}[\mathbf{h}]$ and covariance $\boldsymbol{C}_\mathbf{h} = \mathrm{E}[(\mathbf{h}-\boldsymbol{\mu}_\mathbf{h})(\mathbf{h}-\boldsymbol{\mu}_\mathbf{h})^\mathrm{H}]$.

III. Joint Optimization of Channel Estimation and Equalization

Consider the cost function $C_\mathrm{S}(\mathcal{P}, \boldsymbol{h}, \mathcal{M}_\mathrm{S})$, e.g. the mean square error (MSE), to design an equalizer: It depends on the independent parameters $\mathcal{P} \in \mathbb{P}$, where $\mathbb{P}$ also includes constraints independent of $\boldsymbol{h}$, a set of model parameters $\mathcal{M}_\mathrm{S} \in \mathbb{M}_\mathrm{S}$ assumed to be perfectly known, and the channel parameters $\boldsymbol{h}$, which are unknown and have to be estimated from $\boldsymbol{y}_\mathrm{p}$ (Eqn. 2). The *standard approach* uses the estimated parameters $\hat{\boldsymbol{h}}$ as if they were error-free to obtain the optimum equalizer coefficients $\mathcal{P}_\mathrm{S}$, as depicted in Fig. 3,

$$\mathcal{P}_\mathrm{S} = \operatorname*{argmin}_{\mathcal{P} \in \mathbb{P}} C_\mathrm{S}(\mathcal{P}, \hat{\boldsymbol{h}}, \mathcal{M}_\mathrm{S}). \tag{3}$$

To estimate the channel the receiver relies on $\boldsymbol{s}_\mathrm{p}[n]$ and $\boldsymbol{y}_\mathrm{p}$ (Fig. 2). Given these the channel can be modeled as a random variable from the point of view of the receiver, e.g. $\mathbf{h} \sim \mathcal{N}_\mathrm{c}(\boldsymbol{\mu}_\mathbf{h}, \boldsymbol{C}_\mathbf{h})$. Thus, the cost function $C_\mathrm{S}(\mathcal{P}, \mathbf{h}, \mathcal{M}_\mathrm{S})$ is now a random variable, too.

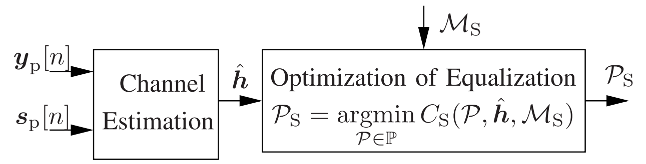

Fig. 3. Standard receiver design: Separate optimization of pilot symbol assisted channel estimation and equalization.

For a *joint optimization* the receiver exploits its knowledge about the channel through $\boldsymbol{y}_\mathrm{p}$ and aims at minimizing the mean cost, which is the conditional mean estimate of the cost, given the received pilot sequence

$$C_\mathrm{J}(\mathcal{P}, \boldsymbol{y}_\mathrm{p}, \mathcal{M}_\mathrm{J}) = \mathrm{E}_\mathbf{h}[C_\mathrm{S}(\mathcal{P}, \mathbf{h}, \mathcal{M}_\mathrm{S})|\boldsymbol{y}_\mathrm{p}]. \tag{4}$$

The set of model parameters assumed known in the joint optimization $\mathcal{M}_\mathrm{J} \in \mathbb{M}_\mathrm{J} = \mathbb{M}_\mathrm{S} \cup \{\mathrm{p}(\boldsymbol{h}|\boldsymbol{y}_\mathrm{p})\}$ is now extended by the conditional probability density function $\mathrm{p}(\boldsymbol{h}|\boldsymbol{y}_\mathrm{p})$ or, alternatively, its parameters $\{\mathrm{p}(\boldsymbol{h}|\boldsymbol{y}_\mathrm{p})\} \equiv \{\boldsymbol{\mu}_{\mathbf{h}|\boldsymbol{y}_\mathrm{p}}, \boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}_\mathrm{p}}\}$ in case of Gaussian distributed channel parameters (see Sec. IV-B) [3]. The joint optimization problem is now given as

$$\mathcal{P}_\mathrm{J} = \operatorname*{argmin}_{\mathcal{P} \in \mathbb{P}} C_\mathrm{J}(\mathcal{P}, \boldsymbol{y}_\mathrm{p}, \mathcal{M}_\mathrm{J}) \tag{5}$$

and does *not* depend on the channel coefficients $\boldsymbol{h}$, but on the received pilot symbols $\boldsymbol{y}_\mathrm{p}$. Thus, it directly chooses the optimum equalizer parameters $\mathcal{P}$ without explicit estimation of the channel (Fig. 4).

This *approach is feasible* if the standard optimization problem (3) can be transformed such that the constraints are either deterministic, i.e., depend on perfectly known parameters, or can be approximated by including them in the cost function (e.g. zero forcing constraints [10]).

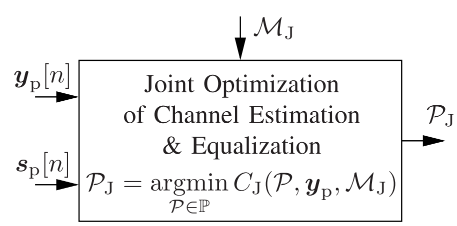

Fig. 4. Joint optimization of receiver: Joint pilot symbol assisted channel estimation and equalization.

IV. Application to Space-Time Decision Feedback Equalization (ST-DFE)

The FIR *space-time decision feedback equalizer* (ST-DFE) [11] depicted in Fig. 5 consists of the matrix FIR feedforward filter $\boldsymbol{G}[n] = \sum_{f=0}^{F} \boldsymbol{G}_f \delta[n-f] \in \mathbb{C}^{K \times M}$ of order F, the temporal feedback filter $\boldsymbol{B}[n] = \sum_{i=1}^{F_\mathrm{B}} \boldsymbol{B}_i \delta[n-i] \in \mathbb{C}^{K \times K}$ of order $F_\mathrm{B} = F + L - \nu$, and the spatial feedback filter $\boldsymbol{D} \in \mathbb{C}_\mathrm{lower}^{K \times K}$, where $\mathbb{C}_\mathrm{lower}^{K \times K}$ denotes the set of lower triangular $K \times$

K matrices with zero main diagonal to ensure *spatial causality* of $\boldsymbol{D}$. We collect the corresponding FIR filter coefficients in

$$\boldsymbol{G} = [\boldsymbol{G}_0, \cdots, \boldsymbol{G}_F] \in \mathbb{C}^{K \times M(F+1)} \quad \text{and} \tag{6}$$

$$\boldsymbol{B} = [\boldsymbol{B}_1, \cdots, \boldsymbol{B}_{F_{\mathrm{B}}}] \in \mathbb{C}^{K \times K F_{\mathrm{B}}}. \tag{7}$$

The $K \times K$ *permutation* matrix $\boldsymbol{P}^{(\mathcal{O})} = \sum_{i=1}^{K} \boldsymbol{e}_i \boldsymbol{e}_{b_i}^{\mathrm{T}}$ describes the spatial ordering $\mathcal{O} \in \mathbb{O}$, where the index b_i of the i-th detected symbol $\boldsymbol{e}_{b_i}^{\mathrm{T}} \mathbf{s}_{\mathrm{d}}[n-\nu]$ is the i-th element of the K-tuple $\mathcal{O}$ and ν is the latency time. The set of all possible orderings is $\mathbb{O} = \{[b_1, \ldots, b_K] \,|\, b_i \in \{1, \ldots, K\} \backslash \{b_1, \ldots, b_{i-1}\}\}$. The symbolwise nearest neighbor mapping from the complex plane to the finite symbol constellation $\mathbb{B}$ is denoted by $\mathcal{Q}(\bullet)$.

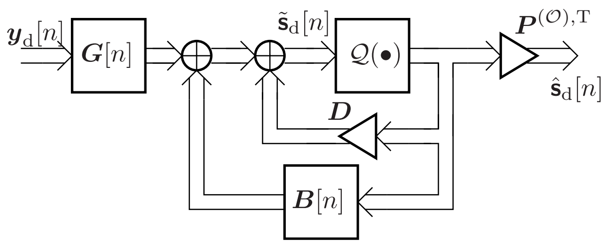

Fig. 5. Structure and parameters of FIR ST-DFE [11].

A. Standard Approach

In the sequel, we consider a ST-DFE minimizing the *mean square error* (MSE)

$$\sigma^2_{\varepsilon,\mathrm{S}}(\mathcal{P}, \boldsymbol{h}, \mathcal{M}_{\mathrm{S}}) = \mathrm{E}[\|\mathbf{s}_{\mathrm{d}}[n-\nu] - \boldsymbol{P}^{(\mathcal{O}),\mathrm{T}} \tilde{\mathbf{s}}_{\mathrm{d}}[n]\|_2^2]. \tag{8}$$

under the common assumption of correctly detected symbols, i.e., $\hat{\mathbf{s}}_{\mathrm{d}}[n] = \mathbf{s}_{\mathrm{d}}[n]$. As the standard approach is well known in the literature [11], [12], [13], we only give a short derivation of the solution in the Appendix. For understanding the main features of our new approach with a joint optimization in the next section, only the structure of the standard solution below is of importance.

The resulting *optimization problem* is

$$\mathcal{P}_{\mathrm{S}} = \operatorname*{argmin}_{\mathcal{P} \in \mathbb{P}} \sigma^2_{\varepsilon,\mathrm{S}}(\mathcal{P}, \boldsymbol{h}, \mathcal{M}_{\mathrm{S}}) \tag{9}$$

with the set of independent variables to be optimized

$$\mathbb{P} = \Big\{ \boldsymbol{G} \in \mathbb{C}^{K \times M(F+1)}, \boldsymbol{B} \in \mathbb{C}^{K \times K F_{\mathrm{B}}}, \boldsymbol{D} \in \mathbb{C}^{K \times K}_{\mathrm{lower}}, \nu \in \{0, \cdots, F+L\}, \mathcal{O} \in \mathbb{O} \Big\} \tag{10}$$

and the set $\mathbb{M}_{\mathrm{S}} = \{\boldsymbol{C}_{\mathbf{n}}, \sigma_{\mathrm{s}}^2\}$ of parameters assumed known. The data symbols are zero mean with $\boldsymbol{C}_{\bar{\mathbf{s}}_{\mathrm{d}}} = \sigma_{\mathrm{s}}^2 \boldsymbol{I}_{K(L+F+1)}$.

The solution for the *filter coefficients* is sketched in the Appendix, where also the necessary definitions are given, and reads as

$$\begin{aligned}
\boldsymbol{G}_{\mathrm{S}} &= \sum_{k=1}^{K} \boldsymbol{e}_k (\boldsymbol{e}_{\nu_{\mathrm{S}}+1}^{\mathrm{T}} \otimes \boldsymbol{e}_{b_{\mathrm{S},k}}^{\mathrm{T}}) \boldsymbol{H}^{\mathrm{H}} (\boldsymbol{H} \boldsymbol{\Pi}_k^{(\mathcal{O}_{\mathrm{S}},\nu_{\mathrm{S}})} \boldsymbol{H}^{\mathrm{H}} + \sigma_{\mathrm{s}}^{-2} \boldsymbol{C}_{\bar{\mathbf{n}}})^{-1}, \\
\boldsymbol{B}_{\mathrm{S}} &= -\boldsymbol{G}_{\mathrm{S}} \boldsymbol{H} (\boldsymbol{I}_{F+L+1} \otimes \boldsymbol{P}^{(\mathcal{O}_{\mathrm{S}}),\mathrm{T}}) \boldsymbol{S}^{(\nu_{\mathrm{S}}),\mathrm{T}}, \quad \text{and} \\
\boldsymbol{D}_{\mathrm{S}} &= \sum_{k=1}^{K} \boldsymbol{e}_k \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{G}_{\mathrm{S}} \boldsymbol{H} (\boldsymbol{e}_{\nu_{\mathrm{S}}+1} \otimes \boldsymbol{I}_K) \boldsymbol{P}^{(\mathcal{O}_{\mathrm{S}}),\mathrm{T}} (\boldsymbol{S}_k \boldsymbol{S}_k^{\mathrm{T}} - \boldsymbol{I}_K).
\end{aligned} \tag{11}$$

Besides some notational details, we make the following observations: From the solution in (11) we see that the temporal feedback filter $\boldsymbol{B}_{\mathrm{S}}$ cancels intersymbol interference from previously detected symbols (determined by latency time ν_{S}), the spatial feedback filter $\boldsymbol{D}_{\mathrm{S}}$ cancels multiple access or interstream interference from already detected symbols (determined by spatial ordering). In essence the feedforward filter $\boldsymbol{G}_{\mathrm{S}}$ has the same structure as any MMSE solution—besides the notation required for describing spatial ordering and latency time—and suppresses interference from symbols, which have not been detected yet. For high SNR it converges to the zero forcing solution.

A suboptimum but efficient solution [13], [12] for the latency time ν_{S} *and* the ordering $\mathcal{O}_{\mathrm{S}}$ is given in Eqn. (33) of the Appendix.

B. Joint Optimization

Based on this brief review of the main ST-DFE results we can proceed with our new method for direct pilot assisted equalizer design from Eqn. (5).

As discussed in Sec. III the channel parameters $\boldsymbol{h}$ in (9) have to be estimated and are only known with errors. The standard optimization does not account for this effect, which could be done applying robust optimization techniques (see Sec. IV-C). But explicit channel estimation is not necessary with joint optimization as will be seen in the sequel.

Following (4) the new cost function of ST-DFE is

$$\sigma^2_{\varepsilon,\mathrm{J}}(\mathcal{P}, \boldsymbol{y}_{\mathrm{p}}, \mathcal{M}_{\mathrm{J}}) = \mathrm{E}_{\mathbf{h}}[\sigma^2_{\varepsilon,\mathrm{S}}(\mathcal{P}, \mathbf{h}, \mathcal{M}_{\mathrm{S}}) | \boldsymbol{y}_{\mathrm{p}}] \tag{12}$$

resulting in the joint optimization problem

$$\mathcal{P}_{\mathrm{J}} = \operatorname*{argmin}_{\mathcal{P} \in \mathbb{P}} \sigma^2_{\varepsilon,\mathrm{J}}(\mathcal{P}, \boldsymbol{y}_{\mathrm{p}}, \mathcal{M}_{\mathrm{J}}) \tag{13}$$

with the set of model parameters

$$\mathbb{M}_{\mathrm{J}} = \mathbb{M}_{\mathrm{S}} \cup \{\boldsymbol{\mu}_{\mathbf{h}}, \boldsymbol{C}_{\mathbf{h}}, \boldsymbol{S}_{\mathrm{p}}\}. \tag{14}$$

Compared to (8) - due to linearity of the expectation - the following *substitutions for the Gram* of the *channel matrix* $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}$ and the channel matrix have to be made to obtain (12):

$$\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} \longleftarrow \mathrm{E}[\mathbf{H}\mathbf{H}^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{p}}] \quad \text{and} \tag{15}$$

$$\boldsymbol{H} \longleftarrow \hat{\boldsymbol{H}} = \mathrm{E}[\mathbf{H} | \boldsymbol{y}_{\mathrm{p}}]. \tag{16}$$

The first expression is the conditional mean estimator of the Gram matrix based on $\boldsymbol{y}_{\mathrm{p}}$. With selection matrix $\boldsymbol{\Sigma}_\ell$ (29) and (28) it can be written as

$$\begin{aligned}
\mathrm{E}[\mathbf{H}\mathbf{H}^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{p}}] &= \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L} \boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'}^{\mathrm{T}} \otimes \mathrm{E}[\mathbf{H}_\ell \mathbf{H}_{\ell'}^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{p}}] \\
\mathrm{E}[\mathbf{H}_\ell \mathbf{H}_{\ell'}^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{p}}] &= \sum_{k=1}^{K} \mathrm{E}[\mathbf{h}_{\ell,k} \mathbf{h}_{\ell',k}^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{p}}],
\end{aligned} \tag{17}$$

where $\boldsymbol{h}_{\ell,k}$ is the k-th column of $\boldsymbol{H}_\ell$. Thus, it can be computed as a sum of subblock matrices of the conditional correlation matrix $\mathrm{E}[\mathbf{h}\mathbf{h}^{\mathrm{H}} | \boldsymbol{y}_{\mathrm{p}}]$ of $\boldsymbol{h}$. To simplify the conditional expectation we assume that $\mathbf{h}$ and $\mathbf{y}_\mathrm{p}$ are *jointly complex Gaussian* distributed [3] and obtain

$$\mathrm{E}[\mathbf{h}\mathbf{h}^\mathrm{H}|\boldsymbol{y}_\mathrm{p}] = \hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^\mathrm{H} + \boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}_\mathrm{p}}. \quad (18)$$

The conditional covariance matrix is given as

$$\boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}_\mathrm{p}} = \boldsymbol{C}_\mathbf{h} - \boldsymbol{W}\boldsymbol{S}_\mathrm{p}\boldsymbol{C}_\mathbf{h} \quad (19)$$

and the conditional mean of $\mathbf{h}$, which is the LMMSE (conditional mean) channel estimator,

$$\hat{\boldsymbol{h}} = \mathrm{E}[\mathbf{h}|\boldsymbol{y}_\mathrm{p}] = \boldsymbol{\mu}_\mathbf{h} + \boldsymbol{W}(\boldsymbol{y}_\mathrm{p} - \boldsymbol{S}_\mathrm{p}\boldsymbol{\mu}_\mathbf{h}) \quad (20)$$
$$\boldsymbol{W} = \boldsymbol{C}_\mathbf{h}\boldsymbol{S}_\mathrm{p}^\mathrm{H}(\boldsymbol{S}_\mathrm{p}\boldsymbol{C}_\mathbf{h}\boldsymbol{S}_\mathrm{p}^\mathrm{H} + \boldsymbol{C}_{\mathbf{n}_\mathrm{p}})^{-1} \quad \text{with} \quad \boldsymbol{C}_{\mathbf{n}_\mathrm{p}} = \mathrm{E}[\mathbf{n}_\mathrm{p}\mathbf{n}_\mathrm{p}^\mathrm{H}].$$

From (17) and (18) we observe that the conditional mean of the Gram (15) can be expressed as the sum of the Gram of the LMMSE channel estimate $\hat{\boldsymbol{H}}$ and the conditional covariance matrix

$$\mathrm{E}[\mathbf{H}\mathbf{H}^\mathrm{H}|\boldsymbol{y}_\mathrm{p}] = \hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\mathrm{H} + \boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_\mathrm{p}}. \quad (21)$$

With the definition of the conditional covariance matrix of $\mathbf{H}$ and the fact that the estimation error and $\mathbf{y}_\mathrm{p}$ are uncorrelated for LMMSE estimation [3] we have

$$\boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_\mathrm{p}} = \mathrm{E}[(\mathbf{H} - \mathrm{E}[\mathbf{H}|\boldsymbol{y}_\mathrm{p}])(\mathbf{H} - \mathrm{E}[\mathbf{H}|\boldsymbol{y}_\mathrm{p}])^\mathrm{H}], \quad (22)$$

which can be interpreted as the *covariance matrix of the estimation error* with the mean square error on the diagonal.

The conditional mean estimator $\hat{\boldsymbol{H}}$ of $\mathbf{H}$ in (16) is derived similarly using (20) and considering the block Toeplitz structure (cf. 28) of $\hat{\boldsymbol{H}}$: $\mathrm{E}[\mathbf{H}|\boldsymbol{y}_\mathrm{p}] = \sum_{\ell=0}^{L} \boldsymbol{\Sigma}_\ell \otimes \mathrm{E}[\mathbf{H}_\ell|\boldsymbol{y}_\mathrm{p}]$.

Thus, the *solution of the joint optimization problem* (13) is given by (11) and (33) with the following substitutions

$$\begin{aligned} \boldsymbol{C}_{\bar{\mathbf{n}}} &\longleftarrow \boldsymbol{C}_{\bar{\mathbf{n}}} + \sigma_\mathrm{s}^2\boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_\mathrm{p}} \\ \boldsymbol{H} &\longleftarrow \hat{\boldsymbol{H}} = \mathrm{E}[\mathbf{H}|\boldsymbol{y}_\mathrm{p}]. \end{aligned} \quad (23)$$

This is a consequence from applying (15) and (16) with the preceding results to (8).

It shows two *structural differences* compared to the standard solution (11) and (33): 1) It does not depend on the channel parameters $\boldsymbol{h}$, but on the received pilot sequence $\boldsymbol{y}_\mathrm{p}$. 2) All inverses include an additional (structured) loading matrix $\boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_\mathrm{p}}$. Obviously, the residual uncertainty about the channel coefficients is represented as an additional "noise source".

C. Interpretation

The approach in (5) and (13) can be viewed as a *pilot assisted equalizer design* without explicit estimation of the channel. From (23) we conclude that the solution of the joint optimization of channel estimation and ST-DFE (cf. Eqn. 12) results in LMMSE channel estimation together with a loading (23) in the inverse for computing the forward filter $\boldsymbol{G}$ in (11) and in the latency time/ordering optimization (33), that represents the correlations in the estimation error. This observation allows further conclusions w.r.t. robust optimization as described below.

The cost function (12) can be rewritten as

$$\sigma_{\varepsilon,\mathrm{J}}^2(\mathcal{P}, \boldsymbol{y}_\mathrm{p}, \mathcal{M}_\mathrm{J}) = \sigma_{\varepsilon,\mathrm{S}}^2(\mathcal{P}, \hat{\boldsymbol{h}}, \mathcal{M}_\mathrm{S}) + \sigma_\mathrm{s}^2\mathrm{tr}(\boldsymbol{G}\boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_\mathrm{p}}\boldsymbol{G}^\mathrm{H}),$$

where the second term represents a Tikhonov regularization [15] of the feedforward filter $\boldsymbol{G}$ with the regularization parameter given by the model $\mathcal{M}_\mathrm{J}$.

Our approach of joint optimization is equivalent to a *robust optimization* based on the paradigm of *static stochastic programming* [16], if an LMMSE channel estimator (20) is chosen. (Additional approaches for robust optimization are discussed in [17].) It models the channel as a random variable centered at the channel estimate and a stochastic estimation error described by its first and second order statistics:

$$\mathbf{H} = \hat{\boldsymbol{H}} + \mathbf{E} \quad \text{with} \quad \boldsymbol{C}_\mathbf{E} = \mathrm{E}[\mathbf{E}\mathbf{E}^\mathrm{H}]. \quad (24)$$

We assumed $\mathrm{E}[\mathbf{E}] = \mathbf{0}$ for simplicity. In this case $\boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_\mathrm{p}}$ is equal to the covariance $\boldsymbol{C}_\mathbf{E}$ of the channel estimation error $\mathbf{E}$ (compare to [9]). The robust cost function for $\mathbf{h} = \hat{\boldsymbol{h}} + \mathbf{e}$ is

$$\begin{aligned} \sigma_{\varepsilon,\mathrm{R}}^2(\mathcal{P}, \hat{\boldsymbol{h}}, \mathcal{M}_\mathrm{R}) &= \mathrm{E}_\mathbf{e}[\sigma_{\varepsilon,\mathrm{S}}^2(\mathcal{P}, \hat{\boldsymbol{h}} + \mathbf{e}, \mathcal{M}_\mathrm{S})] \\ &= \sigma_{\varepsilon,\mathrm{S}}^2(\mathcal{P}, \hat{\boldsymbol{h}}, \mathcal{M}_\mathrm{S}) + \sigma_\mathrm{s}^2\mathrm{tr}(\boldsymbol{G}\boldsymbol{C}_\mathbf{E}\boldsymbol{G}^\mathrm{H}). \end{aligned} \quad (25)$$

Note, that for robust optimization no assumption on the distribution of the error is needed—only knowledge about its first and second order statistics is required—, whereas for the joint optimization in Sec. IV-B we assumed a joint Gaussian distribution of $\mathbf{h}$ and $\mathbf{y}_\mathrm{p}$.

Joint optimization can be viewed as a technique to improve the interface between channel estimator and equalizer communicating the statistical properties of the estimation error. Contrary to worst case robust optimization (min-max) [17] this approach easily incorporates the structure of the estimation error into the design with little additional complexity.

D. Complexity

At first glance, one might predict that the *computational complexity* of solving two problems jointly increases compared to a separate solution of both, as the problem statement is more "complicated" and involves more parameters. In contrast to this intuition, the solution of the new approach (13) is obtained with the *same order* of complexity as a separate optimization with an MMSE ST-DFE (9) and an LMMSE channel estimator (20): the joint solution can be derived by simply substituting (23) into the standard solution (11) and (33) as shown in Sec. IV-B.

V. Performance

Monte Carlo simulations for the mean uncoded *bit error rate* (BER) are shown for the following parameters: 16-QAM modulation, $M = 8$ receive antennas in an uniform linear array with half wavelength interelement spacing, $K = 6$ transmitters, channel order $L = 3$, filter order $F = 8$, $N_\mathrm{p} = 30$ QPSK pilot symbols. The BER is averaged over 1000 blocks each with 300 symbols for every transmitter. The channels between the K transmitters are uncorrelated and angles of arrival at the receiver are Laplace distributed (angular spread of $10°$) [18]. The power delay profile is exponential with rate of decay $1 \mu s$ (symbol rate $1.28\,\mathrm{MHz}$). An LMMSE channel estimator (20) is used for the standard design of the ST-DFE, whereas the joint optimization uses the solution from (23). Latency time optimization and spatial ordering is performed as described above in (33).

Two scenarios are considered (Fig. 6): 1) The channel remains constant during reception of the pilot sequence and data (solid lines). The *joint optimization gains* 2 dB compared to a separate design, which loses about 4.7 dB at an BER of 10^{-2} compared to perfect channel knowledge. 2) *A delay* of 300 symbols between pilot sequence and application of the ST-DFE to equalize the channel is assumed, which is a rough model for equalizing the channel at the end of the data block (dashed lines). In this case the channel has a maximum Doppler frequency of $5 \cdot 10^{-5}$ (Jakes Doppler spectrum) normalized to the symbol period, i.e., temporally correlated. Now, the BER of standard design saturates due to large errors in channel knowledge (Fig. 6). A *joint design* tremendously *decreases the error floor*.

The improved performance is due to the exchange of information about the size of the estimation errors and their structure described by the estimation error covariance matrix $\boldsymbol{C}_{\mathbf{H}|\boldsymbol{y}_{\mathrm{p}}}$ in (21).

As mentioned before, the joint optimization converges to a separate optimization for large number N_{p} of pilot symbols. To investigate the convergence behavior, Fig. 7 shows the BER performance in scenario 1 for $N_{\mathrm{p}} \in \{10, 50, 100\}$. For $N_{\mathrm{p}} = 100$ the gain of the joint optimization over the standard optimization is already small. Below $N_{\mathrm{p}} = 50$ the estimation error is large enough that the equalizer design gains from knowledge of its size and structure. This is the interesting region of operation for joint optimization. Of course, for $N_{\mathrm{p}} = 10$ the BER performance is not acceptable anymore due to the small number of pilots; by joint optimization we gain significantly but not sufficiently in this case.

Comparing the case of $N_{\mathrm{p}} = 100$ and $N_{\mathrm{p}} = 50$ for moderate SNR we observe a similar performance for the joint design with $N_{\mathrm{p}} = 50$ and the standard design for $N_{\mathrm{p}} = 100$: With joint design the pilot symbols are used more efficiently, which results in an SNR gain or may also be used to reduce the number of pilot symbols.

VI. Conclusions

A paradigm for joint optimization of pilot symbol assisted channel estimation and equalization was presented. As an example, it is applied to MMSE optimization of a space-time decision feedback equalizer. But it is also applicable to a wide range of linear and non-linear equalizer designs, such as zero-forcing, Wiener filter, and other DFE approaches. The solution shows that the joint optimization is equivalent to LMMSE channel estimation and a structured loading of the inverse in the filter design with the estimation error covariance matrix. It can be interpreted as an improved interface between channel estimator and equalizer communicating the statistical properties of the estimation error. Large performance gains are obtained with the same order of computational complexity as for the standard design.

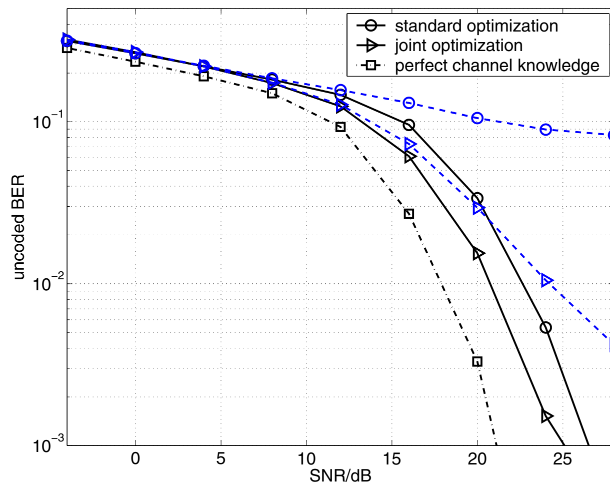

Fig. 6. Comparison of joint optimization of channel estimation and ST-DFE with a separate optimization (*solid lines:* constant channel—scenario 1; *dashed lines:* time delay between pilot and equalization—scenario 2).

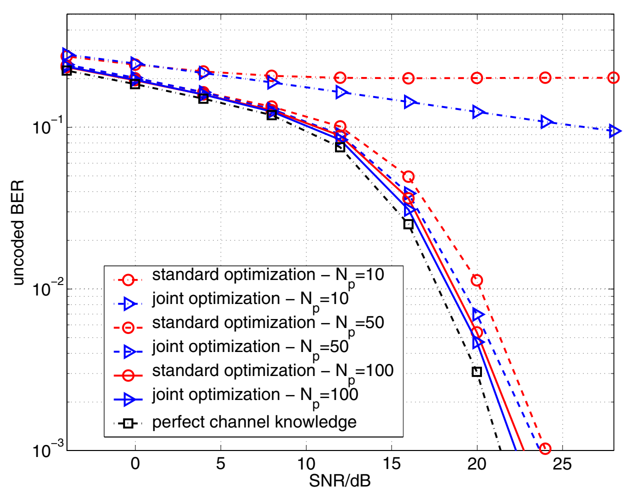

Fig. 7. Joint and separate optimization for different number of pilot symbols $N_{\mathrm{p}} \in \{10, 50, 100\}$ (scenario 1). With increasing number of pilot symbols N_{p} the joint design converges to a separate design.

Appendix

The following sketch of the derivation for the ST-DFE standard approach (Sec. IV-A) is based on [11], [12], [13]. See [14] for a detailed explanation with the same notation.

With the standard assumption of correctly detected symbols, that is $\mathcal{Q}(\tilde{\mathbf{s}}_{\mathrm{d}}[n]) = \boldsymbol{P}^{(\mathcal{O})}\mathbf{s}_{\mathrm{d}}[n-\nu]$, we get for the estimate

$$\tilde{\mathbf{s}}_{\mathrm{d}}[n] = \boldsymbol{D}\,\boldsymbol{P}^{(\mathcal{O})}\mathbf{s}_{\mathrm{d}}[n-\nu] + \boldsymbol{B}[n] * \boldsymbol{P}^{(\mathcal{O})}\mathbf{s}_{\mathrm{d}}[n-\nu] + \boldsymbol{G}[n] * \boldsymbol{H}[n] * \mathbf{s}_{\mathrm{d}}[n] + \boldsymbol{G}[n] * \mathbf{n}_{\mathrm{d}}[n] \in \mathbb{C}^{K} \quad (26)$$

or in matrix-vector notation with $\boldsymbol{e}_{\nu+1} \in \{0,1\}^{F+L+1}$:

$$\tilde{\mathbf{s}}_{\mathrm{d}}[n] = (\boldsymbol{e}_{\nu+1}^{\mathrm{T}} \otimes \boldsymbol{D} + \boldsymbol{B}\boldsymbol{S}^{(\nu)})(\boldsymbol{I}_{F+L+1} \otimes \boldsymbol{P}^{(\mathcal{O})})\bar{\mathbf{s}}_{\mathrm{d}}[n] + \boldsymbol{G}\boldsymbol{H}\bar{\mathbf{s}}_{\mathrm{d}}[n] + \boldsymbol{G}\bar{\mathbf{n}}_{\mathrm{d}}[n] \in \mathbb{C}^{K}. \quad (27)$$

Here, we introduced the block Toeplitz channel matrix

$$\boldsymbol{H} = \sum_{\ell=0}^{L} \boldsymbol{\Sigma}_\ell \otimes \boldsymbol{H}_\ell \in \mathbb{C}^{M(F+1)\times K(F+L+1)} \tag{28}$$

with selection matrix

$$\boldsymbol{\Sigma}_\ell = [\mathbf{0}_{F+1\times\ell}, \boldsymbol{I}_{F+1}, \mathbf{0}_{F+1\times L-\ell}], \tag{29}$$

the data vector

$$\bar{\mathbf{s}}_\mathrm{d}[n] = [\mathbf{s}_\mathrm{d}^\mathrm{T}[n], \ldots, \mathbf{s}_\mathrm{d}^\mathrm{T}[n-L-F]]^\mathrm{T},$$

and the noise vector

$$\bar{\mathbf{n}}_\mathrm{d}[n] = [\mathbf{n}_\mathrm{d}^\mathrm{T}[n], \ldots, \mathbf{n}_\mathrm{d}^\mathrm{T}[n-F]]^\mathrm{T}.$$

The matrix $\boldsymbol{S}^{(\nu)} = [\mathbf{0}_{KF_\mathrm{B}\times K(\nu+1)}, \boldsymbol{I}_{KF_\mathrm{B}}]$ selects the last KF_B elements of the $K(F+L+1)$-dimensional symbol vector $\bar{\mathbf{s}}_\mathrm{d}[n]$ after permutation available for temporal feedback.

We consider a ST-DFE minimizing the *mean square error* (MSE), which reads as (cf. Eqn. 8)

$$\begin{aligned}\sigma_{\varepsilon,\mathrm{S}}^2(\mathcal{P}, \boldsymbol{h}, \mathcal{M}_\mathrm{S}) &= \sigma_\mathrm{s}^2 K - 2\sigma_\mathrm{s}^2 \mathrm{Re}(\mathrm{tr}(\boldsymbol{P}^{(\mathcal{O}),\mathrm{T}} \boldsymbol{G}\boldsymbol{H}(\boldsymbol{e}_{\nu+1}\otimes \boldsymbol{I}_K)) \\ &+ 2\sigma_\mathrm{s}^2 \mathrm{Re}(\mathrm{tr}((\boldsymbol{e}_{\nu+1}^\mathrm{T}\otimes \boldsymbol{D} + \boldsymbol{S}^{(\nu)}\boldsymbol{B})(\boldsymbol{I}_{F+L+1}\otimes \boldsymbol{P}^{(\mathcal{O})})\boldsymbol{H}^\mathrm{H}\boldsymbol{G}^\mathrm{H})) \\ &+ \sigma_\mathrm{s}^2 \mathrm{tr}(\boldsymbol{D}\boldsymbol{D}^\mathrm{H} + \boldsymbol{B}\boldsymbol{B}^\mathrm{H}) + \mathrm{tr}(\boldsymbol{G}\,(\sigma_\mathrm{s}^2 \boldsymbol{H}\boldsymbol{H}^\mathrm{H} + \boldsymbol{C}_{\bar{\mathbf{n}}})\,\boldsymbol{G}^\mathrm{H}),\end{aligned}$$

since $\mathbf{s}_\mathrm{d}[n-\nu] = (\boldsymbol{e}_{\nu+1}^\mathrm{T} \otimes \boldsymbol{I}_K)\bar{\mathbf{s}}_\mathrm{d}[n]$ and $\boldsymbol{D} \in \mathbb{C}_\mathrm{lower}^{K\times K}$.

The resulting *optimization problem* is (9). For solving (9) we split up the constraint $\boldsymbol{D} \in \mathbb{C}_\mathrm{lower}^{K\times K}$ into constraints on the rows of $\boldsymbol{D}$:

$$\boldsymbol{e}_k^\mathrm{T} \boldsymbol{D}\boldsymbol{S}_k = \mathbf{0}_{K-k+1}^\mathrm{T}, k = 1, \ldots, K, \tag{30}$$

where $\boldsymbol{S}_k = [\mathbf{0}_{K-k+1\times k-1}, \boldsymbol{I}_{K-k+1}]^\mathrm{T}$ selects columns with zero elements in k-th row of $\boldsymbol{D}$. With this reformulation of the constraint, the Lagrange multiplier method is applied to solve the optimization problem. Some additional properties of the selection matrices are needed as detailed in [14]. In the solution (11) the projector $\boldsymbol{\Pi}_k^{(\mathcal{O}_\mathrm{S},\nu_\mathrm{S})}$ is introduced. It sets the last KF_B and the $k-1$ columns of $\boldsymbol{H}$ to zero, which correspond to the previously detected symbols:

$$\begin{aligned}\boldsymbol{\Pi}_k^{(\mathcal{O}_\mathrm{S},\nu_\mathrm{S})} &= \boldsymbol{\Pi}^{(\nu_\mathrm{S})} - \sum_{i=1}^{k-1} (\boldsymbol{e}_{\nu_\mathrm{S}+1} \otimes \boldsymbol{e}_{b_{\mathrm{S},i}})(\boldsymbol{e}_{\nu_\mathrm{S}+1}^\mathrm{T} \otimes \boldsymbol{e}_{b_{\mathrm{S},i}}^\mathrm{T}) \quad \text{with} \\ \boldsymbol{\Pi}^{(\nu_\mathrm{S})} &= \boldsymbol{I}_{K(F+L+1)} - \boldsymbol{S}^{(\nu_\mathrm{S}),\mathrm{T}}\boldsymbol{S}^{(\nu_\mathrm{S})}.\end{aligned} \tag{31}$$

Note that $\boldsymbol{\Pi}_i^{(\mathcal{O}_\mathrm{S},\nu_\mathrm{S})}$ only depends on the indices $b_{\mathrm{S},1}, \ldots, b_{\mathrm{S},i-1}$ of the previously detected symbols. Plugging the standard ST-DFE solution (11) into (8), yields for the MSE:

$$\sigma_{\varepsilon,\mathrm{S}}^2(\mathcal{P}_\mathrm{S}, \boldsymbol{h}, \mathcal{M}_\mathrm{S}) = \sigma_\mathrm{s}^2(K - \mathrm{tr}(\boldsymbol{P}^{(\mathcal{O}_\mathrm{S}),\mathrm{T}}\boldsymbol{G}_\mathrm{S}\boldsymbol{H}(\boldsymbol{e}_{\nu_\mathrm{S}+1}\otimes \boldsymbol{I}_K))). \tag{32}$$

Obviously, minimizing this MSE w.r.t. the latency time ν_S *and* the ordering $\mathcal{O}_\mathrm{S}$ has prohibitive complexity, since the inverse in (11) has to be computed $K!(F+L+1)$ times. Therefore, we apply following *suboptimum strategy* instead:

$$\begin{aligned}\nu_\mathrm{S} &= \underset{\nu\in\{0,\cdots,L+F\}}{\mathrm{argmax}}\ \mathrm{tr}((\boldsymbol{e}_{\nu+1}^\mathrm{T}\otimes \boldsymbol{I}_K)\boldsymbol{H}^\mathrm{H}\boldsymbol{A}^{(\nu),-1}\boldsymbol{H}(\boldsymbol{e}_{\nu+1}\otimes \boldsymbol{I}_K)) \\ \mathcal{O}_\mathrm{S} &= [b_{\mathrm{S},1}, \cdots, b_{\mathrm{S},K}] \quad \text{with} \\ b_{\mathrm{S},k} &= \underset{b\in\mathbb{O}_k}{\mathrm{argmax}}\ (\boldsymbol{e}_{\nu_{\mathrm{S}+1}}\otimes \boldsymbol{e}_b)\boldsymbol{H}^\mathrm{H}\boldsymbol{A}_k^{(\mathcal{O}_\mathrm{S},\nu_\mathrm{S}),-1}\boldsymbol{H}(\boldsymbol{e}_{\nu_{\mathrm{S}+1}}^\mathrm{T}\otimes \boldsymbol{e}_b^\mathrm{T}), \\ \boldsymbol{A}^{(\nu)} &= \boldsymbol{A}_1^{(\mathcal{O},\nu)}, \quad \text{and} \quad \boldsymbol{A}_k^{(\mathcal{O},\nu)} = \boldsymbol{H}\boldsymbol{\Pi}_k^{(\mathcal{O},\nu)}\boldsymbol{H}^\mathrm{H} + \sigma_\mathrm{s}^{-2}\boldsymbol{C}_{\bar{\mathbf{n}}}.\end{aligned} \tag{33}$$

where $\mathbb{O}_k = \{1, \cdots, K\} \setminus \{b_1, \cdots, b_{k-1}\}$. Thus, we choose the latency time ν_S under the assumption of an inactive spatial feedback filter (as in [11]), i.e., $\boldsymbol{D} = \mathbf{0}_{K\times K}$, and the ordering $\mathcal{O}_\mathrm{S}$ is found successively by minimizing the MSE of the k-th symbol $\boldsymbol{e}_{b_k}^\mathrm{T}\mathbf{s}_\mathrm{d}[n-\nu]$ under the assumption that the ordering of the previous symbols is fixed (compare to [13], [12]).

References

[1] J. K. Cavers. An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels. *IEEE Trans. Veh. Technol.*, 40(4):686–693, Nov. 1991.

[2] John G. Proakis. *Digital Communications*. McGraw-Hill, 3rd edition, 1995.

[3] Steven M. Kay. *Fundamentals of Statistical Signal Processing - Estimation Theory*. PTR Prentice Hall, 1st edition, 1993.

[4] H. Meyr, M. Moeneclaey, and S. A. Fechtel. *Digital Communication Receivers*. John Wiley, 1998.

[5] L. M. Davis, I. B. Collings, and P. Hoeher. Joint MAP Equalization and Channel Estimation for Frequency Selective and Frequency-Flat Fast-Fadign Channels. *IEEE Trans. Commun.*, pages 2106–2114, Dec. 2001.

[6] M. Tuechler and M. Mecking. Equalization for non-ideal channel knowledge. In *Proc. Conf. on Information Sciences and Systems (CISS)*, March 2003.

[7] N. Sellami, S. Lasaulce, and I. Fijalkow. Iterative Channel Estimation for Coded DS-CDMA Systems over Frequency Selective Channels. In *4th IEEE Workshop on Signal Processing Advances in Wireless Communications*, 2003.

[8] R. Otnes and M. Tuechler. On iterative estimation, equalization and decoding. In *Proc. Intern. Conf. on Communications*, May 2003.

[9] F. A. Dietrich, R. Hunger, M. Joham, and W. Utschick. Robust Transmit Wiener Filter for Time Division Duplex Systems. In *Proceedings of the International Symposium on Signal Processing and Information Technology (ISSPIT)*, Darmstadt, Germany, Dec. 2003.

[10] R. Hunger, F. A. Dietrich, M. Joham, and W. Utschick. Robust Transmit Zero-Forcing Filters. In *Proceedings of the ITG Workshop on Smart Antennas*, Munich, Germany, March 2004.

[11] N. Al-Dhahir and A. H. Sayed. The Finite-Length Multi-Input Multi-Output MMSE-DFE. *IEEE Transactions on Signal Processing*, 48(10):2921–2936, October 2000.

[12] A. Lozano and C. Papadias. Layered Space-Time Receivers for Frequency-Selective Wireless Channels. *IEEE Transactions on Communications*, 50(1):65–73, Jan. 2002.

[13] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela. V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel. In *Proc. ISSSE 98*, pages 295–300, September 1998.

[14] Michael Joham. *Optimization of Linear and Nonlinear Transmit Signal Processing*. Ph. D. Dissertation, Technische Universität München, Shaker Verlag, Aachen, Germany, 2004.

[15] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of Ill-Posed Problems*. V. H. Winston & Sons, 1977.

[16] András Prékopa. *Stochastic Programming*. Kluwer, 1995.

[17] J. Dahl, L. Vandenberghe, and B. H. Fleury. Robust least-squares estimators based on semidefinite programming. In *Proc. 36rd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002.

[18] P. E. Mogensen, K. I. Pedersen, B. Fleury, F. Frederiksen, P. Leth-Espensen, K. Olesen, and S. L. Larsen. TSUNAMI II: 2D-Channel Characteristics. Technical report, Aalborg University, Dec. 1997.