LINEAR PRECODING BASED ON A STOCHASTIC MSE CRITERION

Frank A. Dietrich, Wolfgang Utschick, and Peter Breun

Institute for Circuit Theory and Signal Processing, Munich University of Technology
Arcisstrasse 21, 80290, Munich, Germany
phone: + (49) 89-289 28511, fax: + (49) 89-289 28504, email: {Dietrich, Utschick, Breun}@nws.ei.tum.de
web: www.nws.ei.tum.de

ABSTRACT

The performance of algorithms for preprocessing the signal at the transmitter in mobile communications is severely limited by the amount of available channel state information (CSI). The channel can be modeled as random variable conditioned on a delayed and noisy observation of the channel realization. We introduce novel criteria for linear precoding based on an MMSE criterion, which is stochastic due to the transmitter's channel model. Moreover, employing an estimation theoretic perspective, a new model for the receivers' processing is incorporated into the optimization. We show that the transmitter's partial CSI can is exploited efficiently by this design method and leads to combined linear precoding and channel estimation, which contains the categories of complete and statistical CSI as asymptotic cases.

1. INTRODUCTION

Precoding or preequalization schemes may be designed based on complete, partial, statistical, or no channel state information (CSI). We define *partial CSI* in terms of a conditional probability density function (PDF) assuming an indirect (delayed and noisy) observation of the current channel's realization. This observation can be obtained based on training symbols in the uplink in a time division duplex (TDD) system [1] or from the feedback of the receivers in a closed loop system [2]. Thus, from the transmitter's perspective the channel is modeled as a random variable with a conditional PDF given the noisy observation. *Complete* or *statistical CSI* are viewed as asymptotic cases of partial CSI: in the former the channel realization is known completely and in the latter only knowledge about the channel's unconditioned PDF is available.

The slot structure in TDD systems, the limited feedback, and the Doppler spread of the channel lead to a delayed CSI, which causes significant performance degradation of signal processing at the transmitter. Thus, *robust algorithms* are required to exploit knowledge about size and structure of the errors in CSI given by the conditional PDF. In this article the problem is treated, how different categories of CSI can be incorporated into a clean optimization problem for a linear preequalizer.

In [3] the MMSE solution from the *uplink* assuming statistical CSI and a phase correction at the receiver was applied for downlink beamforming. An MMSE design for preequalization *in the downlink*, i.e., without using an analogy to the uplink as in [3], was presented in [4, 5] assuming complete CSI and without modeling the receivers' channel knowledge explicitly. Assuming a (channel) matched filter at the receiver this approach was extended to statistical CSI in [6] with the detour over a "power equivalent model" based on the channel covariance matrix.

Our contributions are: 1) The MMSE design of [4, 5] is extended to partial and statistical CSI. 2) Criteria for a combined optimization of channel estimation and precoding are proposed and solved. 3) The "detour" over the power equivalent model [6] is avoided with the new "direct" approaches. 4) The receivers' degree of CSI and processing capability is included into the optimization of the transmitter to avoid a conservative design.

We consider a TDD system with multiple transmit antennas and non-cooperative receivers (downlink). Although our results are derived for a frequency-flat fading channel to be able to focus on the main ideas, they are directly applicable to frequency selective channels (space-time filtering) and CDMA systems following [6]. Precoding is optimized based on a *mean square error* (MSE) criterion with average power constraint [4, 5], which is now *stochastic*, i.e., a random variable, due to the transmitter's partial channel knowledge.

As the transmitter knows the precoding/beamforming, which it uses for transmitting the training sequence(s), it has all information about the receivers' degree of CSI. We exploit this information to model the receivers' signal processing based on their amount of CSI. Assuming an amplitude and phase modulation, coherent detection is required at every receiver, which includes the correction of the channel phase. Therefore, we describe the receivers' dependency on the current channel realization in terms of a phase correction (Sec. 2). As this simple processing is part of every coherent receiver's processing chain, it does not pose any constraint on receiver design. Thus, the receiver can still be optimized independent of the precoding. Previously, the receivers have been modeled to have no CSI or only an automatic gain control when optimizing beamforming and precoding. This is not a realistic model for a wireless communication system with coherent demodulation and results in a performance loss in case of partial CSI at the transmitter.

In Sec. 3 different *principal strategies* for dealing with stochastic systems are introduced in the framework of Bayesian estimation, where the focus is on a *combined optimization of channel estimation and precoding*. It is shown how different assumptions made by the transmitter about the *receiver's CSI* and processing capabilities influence the solution of the new optimization problems, which can be given explicitly. Connections to robust optimization [1, 7] and regularization are emphasized. In Sec. 4 and 5 it is shown analytically and with simulations—w.r.t. uncoded bit error rate (BER)—that preequalization based on a conditional mean (CM) estimate of the cost function leads to a consistent solution, which converges nicely to the case of complete CSI (e.g. for stationary users with an infinite number of training symbols) and statistical CSI (e.g. for very high Doppler frequencies). Viewing our problem in the context of robust optimization we show, that the receivers' processing capabilities have to be considered by the transmitter to avoid unnecessarily conservative robust preequalizers resulting in a bad performance. Modeling the receivers' CSI, i.e., their dependency on the current channel realization, together with a CM estimate of the resulting MSE cost function introduces a new methodology in the design of precoding or downlink beamforming. It is more robust to partial CSI and allows to model the degree of CSI at the receivers at the transmitter.

Notation: Random vectors and matrices are denoted by lower and upper case sans serif bold letters (e.g. $\boldsymbol{\mathsf{b}}$, $\boldsymbol{\mathsf{B}}$), whereas the realizations or deterministic variables are, e.g., $\boldsymbol{b}$, $\boldsymbol{B}$. The operators $\mathrm{E}[\bullet]$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, and $\mathrm{tr}(\bullet)$ stand for expectation, transpose, Hermitian transpose, and trace of a matrix, respectively. $\otimes$ and $\delta_{k,k'}$ denote the Kronecker product and function, $\mathrm{vec}(\boldsymbol{B})$ stacks the columns of $\boldsymbol{B}$ in a vector. $\boldsymbol{e}_i$ is the i-th column of an $N \times N$ identity matrix $\boldsymbol{I}_N$

2. SYSTEM MODEL

Downlink Data Channel: Data symbols $\boldsymbol{\mathsf{s}}[n] \in \mathbb{B}^K$ ($\mathrm{E}[\boldsymbol{\mathsf{s}}[n]\boldsymbol{\mathsf{s}}[n]^{\mathrm{H}}] = \boldsymbol{I}_K$, modulation alphabet $\mathbb{B}$) are preequalized/precoded with $\boldsymbol{P} \in \mathbb{C}^{M \times K}$ and transmitted using M antennas over the channel $\boldsymbol{H}_q \in \mathbb{C}^{K \times M}$ to K receivers in the *downlink* (time slot q). The channel is assumed constant during one time slot ("block-fading"). The (non-cooperative) receivers are modeled as $\boldsymbol{G} = \text{diag}[g_k]_{k=1}^K \in \mathbb{C}^{K \times K}$, which may be a function of the channel $\boldsymbol{G}(\boldsymbol{H}_q)$. Including white additive complex Gaussian noise $\boldsymbol{n}[n] \sim \mathcal{N}_c(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_K)$ the estimate of the signal $\boldsymbol{s}[n]$ reads as (Fig. 1)

$$\hat{\boldsymbol{s}}[n] = \boldsymbol{G}\boldsymbol{H}_q\boldsymbol{P}\boldsymbol{s}[n] + \boldsymbol{G}\boldsymbol{n}[n] \in \mathbb{C}^K. \tag{1}$$

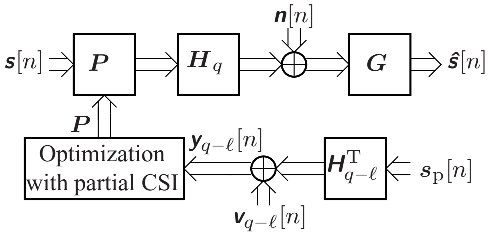

Figure 1: Flat fading system model of up- and downlink.

Downlink Training Channel: There are two main concepts for training symbol based channel estimation in the downlink: Providing K receiver specific (dedicated) training sequences (in some systems restricted to M sequences) or transmitting one common training sequence to the receivers. For the former case a different transmit filter $\boldsymbol{d}_k$ may be used for every training sequence, e.g., $\boldsymbol{d}_k$ is the k-th column of $\boldsymbol{P}$. In the latter case it is transmitted with the same filter $\boldsymbol{d}_k = \boldsymbol{d}$, e.g., over the first antenna $\boldsymbol{d}_k = \boldsymbol{e}_1$. Thus, we assume that receiver k knows $\boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{d}_k$ ($\boldsymbol{h}_{k,q}$ is the k-th column of $\boldsymbol{H}_q^{\mathrm{T}}$) and corrects the phase based on this CSI:

$$\boldsymbol{G} = \text{diag}\left[g_k\right]_{k=1}^K = \text{diag}\left[\left(\boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{d}_k\right)^* / \left|\boldsymbol{h}_{k,q}^{\mathrm{T}}\boldsymbol{d}_k\right|\right]_{k=1}^K. \tag{2}$$

As mentioned above, this model of the receivers' dependencies on the current channel realization is novel in optimization of transmit processing. It captures the most important aspect of a coherent receiver. On the other hand it is still simple enough, that it is included in any—even low-complexity—receiver design without any need for standardization.

Uplink Training Channel: In a TDD system the *channel for designing* $\boldsymbol{P}$ can be estimated from N training symbols (per receiver) $\boldsymbol{s}_{\mathrm{p}}[n] \in \mathbb{C}^K$ ($n \in \{1, \ldots, N\}$) in an *uplink* slot. We assume alternating up-/downlink slots and a delay of 3 slots (due to processing the training sequence) to the first uplink slot available with a training sequence. The receive training signal is (Fig. 1)

$$\boldsymbol{y}_q[n] = \boldsymbol{H}_q^{\mathrm{T}}\boldsymbol{s}_{\mathrm{p}}[n] + \boldsymbol{v}_q[n] \in \mathbb{C}^M,\ n \in \{1, \ldots, N\} \tag{3}$$

with additive white noise $\boldsymbol{v}_q[n] \sim \mathcal{N}_c(\mathbf{0}, \sigma_v^2 \boldsymbol{I}_M)$. Collecting all N training symbols $\boldsymbol{s}_{\mathrm{p}}[n]$ in one matrix $\boldsymbol{S}'_{\mathrm{p}} \in \mathbb{C}^{K \times N}$ we obtain

$$\boldsymbol{Y}_q = \boldsymbol{H}_q^{\mathrm{T}}\boldsymbol{S}'_{\mathrm{p}} + \boldsymbol{V}_q \in \mathbb{C}^{M \times N} \tag{4}$$

$$\bar{\boldsymbol{y}}_q = \text{vec}[\boldsymbol{Y}_q] = (\boldsymbol{S}_{\mathrm{p}}'^{\mathrm{T}} \otimes \boldsymbol{I}_M)\boldsymbol{h}_q + \bar{\boldsymbol{v}}_q \in \mathbb{C}^{MN}, \tag{5}$$

where $\boldsymbol{h}_q = \text{vec}[\boldsymbol{H}_q^{\mathrm{T}}]$. Training signals from Q previous uplink slots[1] are available for estimating/predicting the channel realization at time q. Thus, the total observation is

$$\boldsymbol{y}_q = \boldsymbol{S}\boldsymbol{h}_{\mathrm{T},q} + \boldsymbol{v}_q \in \mathbb{C}^{MNQ} \tag{6}$$

with $\boldsymbol{h}_{\mathrm{T},q} = [\boldsymbol{h}_{q-3}^{\mathrm{T}}, \boldsymbol{h}_{q-5}^{\mathrm{T}}, \ldots, \boldsymbol{h}_{q-(2Q+1)}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{QMK}$, $\boldsymbol{y}_q = [\bar{\boldsymbol{y}}_{q-3}^{\mathrm{T}}, \bar{\boldsymbol{y}}_{q-5}^{\mathrm{T}}, \ldots, \bar{\boldsymbol{y}}_{q-(2Q+1)}^{\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{S} = \boldsymbol{I}_Q \otimes \boldsymbol{S}_{\mathrm{p}}'^{\mathrm{T}} \otimes \boldsymbol{I}_M$.

The channel coefficients $\boldsymbol{h}_q = \text{vec}[\boldsymbol{H}_q^{\mathrm{T}}]$ are modeled as a stationary zero mean complex Gaussian random vector with covariance matrix $\boldsymbol{C}_{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h}_q\boldsymbol{h}_q^{\mathrm{H}}]$, which is block diagonal assuming $\mathrm{E}[\boldsymbol{h}_{k,q}\boldsymbol{h}_{k',q}^{\mathrm{H}}] = \boldsymbol{C}_{\boldsymbol{h}_k}\delta_{k,k'}$. $\boldsymbol{h}_{k,q} \sim \mathcal{N}_c(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{h}_k})$ is the k-th column of $\boldsymbol{H}_q^{\mathrm{T}}$. For simplicity, we assume identical autocorrelation $r[i]$ (normalized to $r[0] = 1$) for all elements of $\boldsymbol{h}_q$ and a time-difference of i slots, i.e., $\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}} = \boldsymbol{C}_{\mathrm{T}} \otimes \boldsymbol{C}_{\boldsymbol{h}}$. $\boldsymbol{C}_{\mathrm{T}}$ is Toeplitz with first column $[r[0], r[2], \ldots, r[2Q-2]]^{\mathrm{T}}$.

Throughout the article we assume that first and second order channel and noise statistics are given. For convenience, we omit the slot index q in $\boldsymbol{H}_q$, $\boldsymbol{h}_q$, $\boldsymbol{y}_q$, and $\boldsymbol{h}_{\mathrm{T},q}$ in the sequel and write: $\boldsymbol{H}$, $\boldsymbol{h}$, $\boldsymbol{y}$, and $\boldsymbol{h}_{\mathrm{T}}$.

3. STOCHASTIC MSE CRITERION: PARTIAL CSI AT TRANSMITTER

The filter $\boldsymbol{P}$ is optimized minimizing the modified MSE [5]

$$\begin{aligned}\sigma_{\boldsymbol{\varepsilon}}^2(\boldsymbol{P}, \beta, \boldsymbol{H}) &= \mathrm{E}[\|\boldsymbol{s}[n] - \beta^{-1}\hat{\boldsymbol{s}}[n]\|_2^2] = K + \beta^{-2}\sigma_n^2\mathrm{tr}[\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}] \\ &+ \beta^{-2}\mathrm{tr}[\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}\boldsymbol{P}] - 2\beta^{-1}\,\mathrm{Re}[\mathrm{tr}[\boldsymbol{G}\boldsymbol{H}\boldsymbol{P}]]\end{aligned} \tag{7}$$

subject to an average transmit power constraint:

$$\{\boldsymbol{P}, \beta\} = \operatorname*{argmin}_{\boldsymbol{P}', \beta'} \sigma_{\boldsymbol{\varepsilon}}^2(\boldsymbol{P}', \beta', \boldsymbol{H}) \ \text{ s.t. } \mathrm{E}[\|\boldsymbol{P}\boldsymbol{s}\|_2^2] \leq P_{\mathrm{Tx}}. \tag{8}$$

Rescaling of the estimate $\hat{\boldsymbol{s}}[n]$ by β^{-1} provides the necessary degree of freedom to incorporate the power constraint and may be interpreted, e.g., as a model for the automatic gain control at the receivers [5].

Complete knowledge about the realization $\boldsymbol{H}$ is never available at the transmitter, but obtained via the observations in $\boldsymbol{y}$. Thus, the channel $\boldsymbol{h}$ is considered as a random variable by the transmitter with conditional (complex) Gaussian PDF $\mathrm{p}_{\boldsymbol{H}|\boldsymbol{y}}(\boldsymbol{H}|\boldsymbol{y})$ with mean $\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}} = \mathrm{E}[\boldsymbol{h}|\boldsymbol{y}]$ and covariance $\boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}}$:

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{y}} &= \hat{\boldsymbol{h}} = \boldsymbol{W}\boldsymbol{y}, \quad \boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}}\boldsymbol{S}^{\mathrm{H}} + \sigma_v^2\boldsymbol{I}_{MNQ})^{-1} \\ \boldsymbol{C}_{\boldsymbol{h}|\boldsymbol{y}} &= \boldsymbol{C}_{\boldsymbol{h}} - \boldsymbol{W}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{T}}\boldsymbol{h}},\end{aligned} \tag{9}$$

where $\boldsymbol{C}_{\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}} = \mathrm{E}[\boldsymbol{h}\boldsymbol{h}_{\mathrm{T}}^{\mathrm{H}}] = [r[3], r[5], \ldots, r[2Q+1]] \otimes \boldsymbol{C}_{\boldsymbol{h}}$ and $\boldsymbol{W}$ is equivalent to the LMMSE estimator [9]. Due to the partial knowledge about $\boldsymbol{H}$ via the conditional PDF the cost function (7) and the argument of the optimization (8) are random variables, too.

The solution of (8) depends on $\boldsymbol{H}$. Thus, it is a random variable itself. With $\alpha = \sigma_n^2\mathrm{tr}(\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G})/P_{\mathrm{Tx}}$ and β chosen to satisfy the constraint with equality it is given by

$$\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{H}) = \beta(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H} + \alpha\boldsymbol{I}_M)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}. \tag{10}$$

Certainly, assuming no CSI at the receiver [6, 1] is not justified in most systems. The strategy for transmitting the training sequence in the downlink determines the receivers' channel knowledge. Their processing is modeled by a phase correction with $\boldsymbol{G}$ (2). Although $\boldsymbol{G}$ is fixed at the receivers for one channel realization, from the transmitter's perspective $\boldsymbol{G}$ is now also a random variable described by $\boldsymbol{d}_k$ and $\mathrm{p}_{\boldsymbol{H}|\boldsymbol{y}}(\boldsymbol{H}|\boldsymbol{y})$.

There are *three principal approaches* how to deal with the *stochastic MSE criterion* in (8), which are presented in the next subsections.

[1] In closed loop systems $\boldsymbol{y}_q$ can be substituted by estimates $\hat{\boldsymbol{h}}_{\mathrm{T},q}$ with appropriate Gaussian statistical model in the sequel [2, 8].

3.1 Traditional Approach

The channel realization $\boldsymbol{H}$ is estimated, e.g., $\hat{\boldsymbol{H}} = \mathrm{E}[\boldsymbol{H}|\boldsymbol{y}]$, and simply plugged into the optimization (8) problem based on $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G} = \boldsymbol{I}_K$, (10), and (2), which yields [5]

$$\boldsymbol{P}_{\mathrm{T}} = \beta_{\mathrm{T}}(\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}} + \alpha \boldsymbol{I}_M)^{-1}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}_{\mathrm{T}}^{\mathrm{H}} \tag{11}$$

with $\alpha = K\sigma_{\boldsymbol{n}}^2/P_{\mathrm{Tx}}$ and β_{T} chosen to satisfy the constraint with equality. The estimate $\hat{\boldsymbol{G}}_{\mathrm{T}}$ for the receivers' processing is obtained applying $\hat{\boldsymbol{h}}$ from (9) to (2). The estimate $\hat{\boldsymbol{H}}$ of the channel is used as *if it was the true one*. The numerical complexity is in the order of $O(K^3)$ after applying the matrix-inversion lemma [9].

3.2 Conditional Mean Estimate of the Cost Function

A conditional mean (CM) estimate of the MSE (7), i.e., the expected cost given past observations of the channel in $\boldsymbol{y}$, is minimized

$$\min_{\boldsymbol{P}',\beta'} \mathrm{E}_{\boldsymbol{H}}\left[\sigma_{\boldsymbol{\varepsilon}}^2(\boldsymbol{P}',\beta',\boldsymbol{H}_q)|\boldsymbol{y}\right] \text{ s.t. } \mathrm{E}[\|\boldsymbol{P}\boldsymbol{s}\|_2^2] \leq P_{\mathrm{Tx}}. \tag{12}$$

It can be viewed as a *combined optimization of channel estimation and precoding*, as the optimization problem only depends on the observation $\boldsymbol{y}$ and not on the channel $\boldsymbol{H}$ or its estimate $\hat{\boldsymbol{H}}$.

The CM estimate of the MSE (7) is

$$\begin{aligned}\sigma_{\boldsymbol{\varepsilon}}^2(\boldsymbol{P},\beta,\boldsymbol{H}) &= K + \beta^{-2}\sigma_{\boldsymbol{n}}^2\mathrm{tr}[\mathrm{E}[\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}|\boldsymbol{y}]] \\ &+ \beta^{-2}\mathrm{tr}[\boldsymbol{P}^{\mathrm{H}}\mathrm{E}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}]\boldsymbol{P}] - 2\beta^{-1}\,\mathrm{Re}[\mathrm{tr}[\mathrm{E}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}]\boldsymbol{P}]]\end{aligned} \tag{13}$$

The CM estimate of the MSE (13) can be simplified with $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G} = \boldsymbol{I}_K$, i.e., $\mathrm{E}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}] = \mathrm{E}[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}|\boldsymbol{y}]$ the CM estimate of the channel Gramian $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ is

$$\mathrm{E}_{\boldsymbol{H}}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}|\boldsymbol{y}\right] = \hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}},\ \hat{\boldsymbol{H}} = \mathrm{E}[\boldsymbol{H}|\boldsymbol{y}]. \tag{14}$$

The conditional covariance can be computed using (9) $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}} = \mathrm{E}[(\boldsymbol{H}-\hat{\boldsymbol{H}})^{\mathrm{H}}(\boldsymbol{H}-\hat{\boldsymbol{H}})|\boldsymbol{y}] = \sum_{k=1}^{K}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}}^*$ and is identical to the covariance matrix of the estimation error $\mathrm{E}[(\boldsymbol{H}-\hat{\boldsymbol{H}})^{\mathrm{H}}(\boldsymbol{H}-\hat{\boldsymbol{H}})]$ due to the orthogonality property of the LMMSE estimator *and* the jointly (complex) Gaussian distribution of $\boldsymbol{y}$ and $\boldsymbol{h}$ [9].

The CM estimate of the effective channel $\boldsymbol{G}\boldsymbol{H}$ reads

$$\mathrm{E}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}] = \hat{\boldsymbol{G}}\hat{\boldsymbol{H}} + \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}},\ \hat{\boldsymbol{G}} = \mathrm{E}[\boldsymbol{G}|\boldsymbol{y}] = \mathrm{diag}\,[\hat{g}_k]_{k=1}^K,$$

where the kth row of $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}} \in \mathbb{C}^{K\times M}$ is given by

$$\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}} = \boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}}^*\,c_{\mathsf{x}_k|\boldsymbol{y}}^{-1}\left(\mathrm{E}[|\mathsf{x}_k|\ |\boldsymbol{y}] - \mu_{\mathsf{x}_k|\boldsymbol{y}}\hat{g}_k\right). \tag{15}$$

$\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}}$ accounts for the CM estimation of the *product* $\boldsymbol{G}\boldsymbol{H}$, as $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{H}}$ are separate CM estimates. The channel estimated by the k-th receiver is $\mathsf{x}_k = \boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{h}_k$ with first and second order moments

$$\begin{aligned}\mu_{\mathsf{x}_k|\boldsymbol{y}} &= \mathrm{E}[\mathsf{x}_k|\boldsymbol{y}] = \boldsymbol{d}_k^{\mathrm{T}}\hat{\boldsymbol{h}}_k, \\ c_{\mathsf{x}_k|\boldsymbol{y}} &= \mathrm{E}[|\mathsf{x}_k - \mu_{\mathsf{x}_k|\boldsymbol{y}}|^2|\boldsymbol{y}] = \boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}}\boldsymbol{d}_k^*.\end{aligned} \tag{16}$$

With [10] the remaining terms, i.e., the CM estimate of the receiver g_k and of the magnitude $|\mathsf{x}_k|$ are

$$\begin{aligned}\hat{g}_k &= \mathrm{E}[\mathsf{g}_k|\boldsymbol{y}] \\ &= \frac{\sqrt{\pi}}{2}\frac{|\mu_{\mathsf{x}_k|\boldsymbol{y}}|}{c_{\mathsf{x}_k|\boldsymbol{y}}^{1/2}}\frac{\mu_{\mathsf{x}_k|\boldsymbol{y}}^*}{|\mu_{\mathsf{x}_k|\boldsymbol{y}}|}\,{}_1F_1\left(\frac{1}{2},2,-\frac{|\mu_{\mathsf{x}_k|\boldsymbol{y}}|^2}{c_{\mathsf{x}_k|\boldsymbol{y}}}\right),\end{aligned} \tag{17}$$

$$\mathrm{E}[|\mathsf{x}_k|\ |\boldsymbol{y}] = \frac{\sqrt{\pi}}{2}c_{\mathsf{x}_k|\boldsymbol{y}}^{1/2}\,{}_1F_1\left(-\frac{1}{2},1,-\frac{|\mu_{\mathsf{x}_k|\boldsymbol{y}}|^2}{c_{\mathsf{x}_k|\boldsymbol{y}}}\right), \tag{18}$$

where ${}_1F_1(\alpha,\beta,z)$ is the confluent hypergeometric function.

With β_{P} chosen to satisfy the constraint with equality, the solution of (12) based on the Lagrange function is

$$\boldsymbol{P}_{\mathrm{P}} = \beta_{\mathrm{P}}\left(\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}} + \alpha\boldsymbol{I}_M\right)^{-1}\left(\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}} + \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}}^{\mathrm{H}}\right). \tag{19}$$

Interpretation: The combined optimization is identical to CM channel estimation of $\boldsymbol{H}$ with $\boldsymbol{W}$ (9) and of $\boldsymbol{G}$ (17), whose estimation error size and structure are described by $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}}$ and $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}}$. The structured loading with $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}}$ in the inverse shows a close relation to *robust optimization* and Tikhonov regularization based on a stochastic error model and the paradigm of (static) stochastic programming [7, 1]. Thus, knowledge about size and structure of the (random) estimation error is considered in the design of $\boldsymbol{P}$. Comparing (19) with (11) we can interpret $\hat{\boldsymbol{G}}\hat{\boldsymbol{H}} + \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}}$ as the channel model of the transmitter, which is used instead of $\hat{\boldsymbol{G}}_{\mathrm{T}}\hat{\boldsymbol{H}}$ and is strongly influenced by the choice of $\boldsymbol{d}_k$ above.

The additional complexity compared to the traditional approach (11) with LMMSE channel estimation[2] is small and dominated by the computation of $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}}$ (9) with $O(M^3K^2NQ)$.

3.3 Conditional Mean Estimate of the Argument

The best estimator for $\boldsymbol{P}(\boldsymbol{H})$ (10), i.e., for the solution of (8), in the mean square sense given $\boldsymbol{y}$ can be derived from

$$\min_{\boldsymbol{P}_{\mathrm{CM}}} \mathrm{E}_{\boldsymbol{H}}[\|\boldsymbol{P}(\boldsymbol{H}) - \boldsymbol{P}_{\mathrm{CM}}\|_{\mathrm{F}}^2|\boldsymbol{y}] \text{ s.t. } \|\boldsymbol{P}_{\mathrm{CM}}\|_{\mathrm{F}}^2 = P_{\mathrm{Tx}}.$$

It is the CM estimate of $\boldsymbol{P}(\boldsymbol{H})$

$$\boldsymbol{P}_{\mathrm{CM}} = \beta_{\mathrm{CM}}\mathrm{E}_{\boldsymbol{H}}[\boldsymbol{P}(\boldsymbol{H})|\boldsymbol{y}], \tag{20}$$

β_{CM} chosen to satisfy the constraint. An explicit solution is not possible, but can be obtained via Monte Carlo simulations of the CM, which is very expensive.[3]

4. ASYMPTOTIC CASES OF CSI

Further insights in the solution (19) for partial CSI can be obtained when considering the situation of complete and statistical CSI at the transmitter. We show that a continuous transition between the different categories of CSI is possible.

4.1 Complete CSI at Transmitter

For complete CSI, i.e., $\hat{\boldsymbol{H}} = \boldsymbol{H}$, the error covariance matrix $\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}}$ of the channel estimate is zero and the matrix $\mathrm{E}[\boldsymbol{G}\boldsymbol{H}|\boldsymbol{y}]$ is equivalent to the effective channel $\boldsymbol{G}\boldsymbol{H}$:

$$\boldsymbol{C}_{\boldsymbol{H}^{\mathrm{H}}|\boldsymbol{y}} \to \boldsymbol{0}_{M\times M},\ \boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}} \to \boldsymbol{0}_{K\times M},\ \hat{\boldsymbol{H}} \to \boldsymbol{H},\ \hat{\boldsymbol{G}} \to \boldsymbol{G}.$$

Convergence is achieved for $\sigma_{\boldsymbol{v}}^2 \to 0$ and $r[i] \to 1\,\forall i$. The MSE (7) is no random variable, as the conditional PDF $\mathrm{p}_{\boldsymbol{H}|\boldsymbol{y}}(\boldsymbol{H}|\boldsymbol{y}) = \delta(\boldsymbol{h} - \boldsymbol{W}\boldsymbol{y})$ is a Dirac distribution centered at $\boldsymbol{W}\boldsymbol{y}$. Thus, all approaches of Sec. 3 converge to

$$\boldsymbol{P}_{\mathrm{C}} = \beta_{\mathrm{C}}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \alpha\boldsymbol{I}_M\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}} \tag{21}$$

with α and β_{C} as above.

---

[2] The computational complexity of the LMMSE estimator—computation of $\boldsymbol{W}$ in (9)—is $O(M^3N^3Q^3 + K^2M^3Q^2N)$ if no symmetry and structure of the linear system of equations is exploited.

[3] A approach in two stages is also introduced in [11] for statistical CSI using the dominant eigenvector of the precoder's correlation matrix.

4.2 Statistical CSI at Transmitter

The situation is more involved in the case of statistical CSI, i.e., only knowledge about the PDF $\mathrm{p}_{\boldsymbol{H}}(\boldsymbol{H})$, where the observation $\boldsymbol{y}$ is independent of the current channel $\boldsymbol{H}$ due to high Doppler frequency (temporally uncorrelated channel). If no information is available about the current channel, e.g., as $\sigma_v^2 \rightarrow$ or $r[i] \rightarrow 0$ for $i > 0$, the CM channel estimate (9) is $\hat{\boldsymbol{H}} = \boldsymbol{0}_{K \times M}$.

The traditional approach (11) yields $\boldsymbol{P}_\mathrm{T} \rightarrow \boldsymbol{0}_{M \times K}$ for $\hat{\boldsymbol{H}} = \boldsymbol{0}_{K \times M}$, which can be derived from the Karush-Kuhn-Tucker conditions. Obviously, not transmitting at all is not an adequate solution for the case of statistical CSI.

Furthermore, it has to be noted that $\|\hat{\boldsymbol{H}}^\mathrm{H}\hat{\boldsymbol{H}}\|_\mathrm{F} > 0$ numerically: For example, the traditional solution will not reach its limit in practice but converge to $\boldsymbol{P}_\mathrm{T} \rightarrow \beta_\mathrm{T}\hat{\boldsymbol{H}}^\mathrm{H}$ as $\|\hat{\boldsymbol{H}}^\mathrm{H}\hat{\boldsymbol{H}}\|_\mathrm{F} \ll \|\alpha \boldsymbol{I}_M\|_\mathrm{F}$. This is a matched filter on the channel. Applying an LMMSE estimator together with the traditional approach, ensures that the rows of $\hat{\boldsymbol{H}}$ are part of the subspace in case $\boldsymbol{C}_{\boldsymbol{h}_k}$ is of low rank. This yields a BER smaller than 0.5 as seen in Fig. 3.

As the PDF $\mathrm{p}_{\boldsymbol{H}|\boldsymbol{y}}(\boldsymbol{H}|\boldsymbol{y}) \rightarrow \mathrm{p}_{\boldsymbol{H}}(\boldsymbol{H})$ and $\boldsymbol{U}_{\boldsymbol{H}|\boldsymbol{y}} \rightarrow \boldsymbol{U}_{\boldsymbol{H}}$ the solution (19) for partial CSI at the transmitter and some CSI at the receiver converges to

$$\boldsymbol{P}_\mathrm{P} \rightarrow \boldsymbol{P}_\mathrm{S} = \beta_\mathrm{S}\left(\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} + \alpha \boldsymbol{I}_M\right)^{-1}\boldsymbol{U}_{\boldsymbol{H}}^\mathrm{H} \tag{22}$$

with the channel's second order statistics $\boldsymbol{C}_{\boldsymbol{H}^\mathrm{H}} = \mathrm{E}[\boldsymbol{H}^\mathrm{H}\boldsymbol{H}] = \sum_{k=1}^{K} \boldsymbol{C}_{\boldsymbol{h}_k}^*$ and $\boldsymbol{e}_k^\mathrm{T}\boldsymbol{U}_{\boldsymbol{H}} = \sqrt{\pi}\boldsymbol{d}_k^\mathrm{T}\boldsymbol{C}_{\boldsymbol{h}_k}^\mathrm{T}/(\boldsymbol{d}_k^\mathrm{H}\boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{d}_k)^{1/2}/2$. This converges is a consequence of the receiver model (2). $\boldsymbol{P}_\mathrm{S}$ shows a close relation to traditional beamforming solutions [3] with a rank one model for each receiver's channel. Note, that $\boldsymbol{U}_{\boldsymbol{H}}$ is determined by the system's concept for transmitting the downlink training sequence with spatial filters $\boldsymbol{d}_k$ (Sec. 2).

5. PERFORMANCE EVALUATION

Simulation parameters: QPSK data symbols, $M = 4$ transmit antennas in a uniform linear array (half wavelength spacing), and $K = 3$ receivers are used. All complex Gaussian channel coefficients have the same Jakes power spectrum with maximum Doppler frequency f_d (normalized to the slot period). The azimuth directions of the receivers' channels are Laplace distributed with mean $[-15°, 0°, 15°]$ and standard deviation $3°$. Walsh-Hadamard sequences of length $N = 32$ are used for training in the uplink and the received training sequences from $Q = 5$ previous uplink slots are considered for prediction. The receivers (2) perform a phase correction with $\boldsymbol{d}_k$ equal to the principal eigenvector of the correlation matrices $\boldsymbol{R}_{\boldsymbol{h}_k|\boldsymbol{y}} = \hat{\boldsymbol{h}}_k\hat{\boldsymbol{h}}_k^\mathrm{H} + \boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}}$. For computing $\boldsymbol{P}_\mathrm{CM}$ in (20) we choose $\boldsymbol{G}$ as in (2) and use 100 random realizations to compute the CM estimate via Monte Carlo simulations.

Results: Fig. 2 shows the uncoded BER vs. SNR: $\boldsymbol{P}_\mathrm{P}$ (19) based on the CM estimate of the cost results in a significantly lower BER floor than the traditional design $\boldsymbol{P}_\mathrm{T}$ (11). Having CSI at the receiver is clearly better than traditional design $\boldsymbol{P}_\mathrm{T}$ without Rx-processing $\boldsymbol{G} = \boldsymbol{I}_K$ (cf. Fig. 3). With increasing Doppler frequency f_d our systematic approach $\boldsymbol{P}_\mathrm{P}$ (19) results in a perfect transition in BER from complete CSI $\boldsymbol{P}_\mathrm{C}$ to statistical CSI $\boldsymbol{P}_\mathrm{S}$ (cf. Fig. 3). The CM estimate $\boldsymbol{P}_\mathrm{CM}$ (20) of the argument of (8), which requires a high computational complexity, performs worse than $\boldsymbol{P}_\mathrm{P}$: Thus, estimating the MSE cost function is the correct paradigm for dealing with partial CSI. At low SNR and low Doppler frequency all approaches perform similarly. For medium SNR, where interference is the performance limiting factor, and medium Doppler frequency we already have a large performance advantage—in this simple scenario—when modeling the receiver more accurately and performing a CM estimate of the MSE.

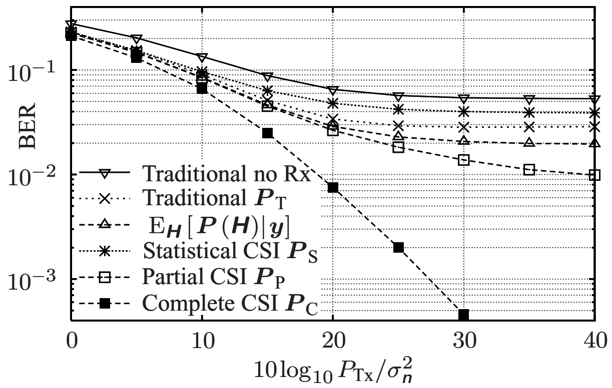

Figure 2: Uncoded BER vs. SNR at Doppler freq. $f_\mathrm{d} = 0.11$.

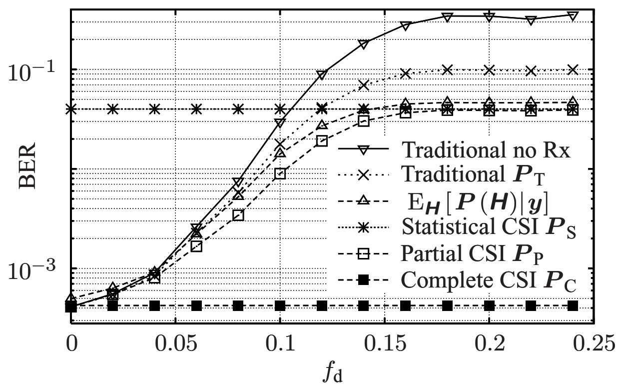

Figure 3: Uncoded BER vs. f_d at $10\log_{10}(P_\mathrm{Tx}/\sigma_n^2) = 30\,\mathrm{dB}$.

REFERENCES

[1] F. A. Dietrich, R. Hunger, M. Joham, and W. Utschick. Robust Transmit Wiener Filter for Time Division Duplex Systems. In *Proc. of ISSPIT*, Germany, Dec. 2003.

[2] F. Rey, M. Lamarca, and G. Vazquez. A Joint Transmitter-Receiver Design in MIMO systems Robust to Channel Uncertainty for W-LAN Applications. In *Proceedings of the IST Mobile & Wireless Telecommunications Summit*, pages 314–318, June 2002.

[3] M. Bengtsson and B. Ottersten. Uplink and Downlink Beamforming for Fading Channels. In *Proc. of IEEE Signal Proc. Workshop on Signal Proc. Advances in Wireless Communications*, pages 350–353, Feb. 1999.

[4] 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*, volume 1, pages 9–13, Sept. 2002.

[5] M. Joham, W. Utschick, and J. A. Nossek. Linear Transmit Processing in MIMO Communications Systems. *IEEE Trans. on Signal Proc.*, Aug. 2005.

[6] M. Joham, K. Kusume, W. Utschick, and J. A. Nossek. Transmit Matched Filter and Transmit Wiener Filter for the Downlink of FDD DS-CDMA Systems. In *Proc. PIMRC 2002*, volume 5, pages 2312–2316, Sept. 2002.

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

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

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

[10] K. S. Miller. *Complex Stochastic Processes*. Addison-Wesley, 1st edition, 1974.

[11] M. T. Ivrlac, Ruly Lai U Choi, Ross Murch, and Josef A. Nossek. Effective use of long-term transmit channel state information in multi-user MIMO communication systems. In *Proc. VTC Fall*, pages 409–413, Orlando, Florida, Oct. 2003.