MIMO Channel Estimation in Correlated Fading Environments

Mario Kiessling[1,2], Joachim Speidel[1], Yejian Chen[1]
[1]Institute of Telecommunications, University of Stuttgart, Germany
[2]Siemens AG, Information and Communication Mobile, Ulm, Germany
mario.kiessling@gmx.net, speidel@inue.uni-stuttgart.de

***Abstract*-Pilot aided channel estimation is considered for wireless MIMO systems in presence of fading correlation. Assuming a linear minimum mean squared error based channel estimator, we study statistical shaping of the training sequence via a matrix prefilter depending on the correlation properties of the channel. The new scheme is a general concept that can readily be applied to smart antenna multiple-input single-output systems. Furthermore, in a quasi-stationary environment with limited mobility, it requires only a long-term update interval of the prefilter, thus admitting its use in FDD systems with low-rate feedback link for the correlation matrices. Simulation results of a MIMO system with maximum eigenmode transmission including realistic channel estimation demonstrate the effectiveness of the proposed scheme with a significant SNR gain.**

I. Introduction

Provided that both receiver and transmitter of a MIMO system are aware of the long-term stable correlation properties of the channel (e.g. via low-rate feedback etc.), this knowledge should be exploited to improve the overall system performance. In this paper we study channel estimation (CE) aspects of a MIMO system in the presence of fading correlation. With a growing number of TX antennas, the length of the training sequence needs also to be increased, leaving less space for data transmission and thus impacting the overall spectral efficiency of the MIMO system [1]. Hence, CE is a critical part of the system, motivating research of more advanced pilot symbol assisted CE schemes [2][3] that minimize training overhead. We mention that on the other hand there are approaches that try to avoid channel estimation (at the expense of a performance penalty) by introducing differential modulation [4].

Standard orthogonal training sequences are the starting point for our work. On the basis of a well-known simplified flat fading model for spatially correlated wireless MIMO channels (an extension to more general channel models is straightforward) we derive statistical transmit prefiltering schemes for the MIMO training sequence. Alternatively, the prefiltering scheme can be conceived as an optimal training sequence design. At the RX we assume linear minimum mean squared error (MMSE) CE [5]. By minimizing the resulting overall mean squared error (MSE) of the channel estimator, we derive the optimum linear transmit prefilter.

The effectiveness of the proposed scheme is demonstrated via Monte-Carlo simulations of the CE MSE and on the other hand via bit error rate (BER) simulations comprising CE effects for transmission on the maximum eigenmode of the MIMO channel. Application of the concept to smart antenna multiple-input single-output (MISO) systems with an antenna array at the base station only and a single antenna at the mobile station is possible.

II. Signal And Channel Model

In the remainder of the paper, bold lowercase letters denote column vectors, bold uppercase letters describe matrices, by $\mathbf{I}_s$ we denote an identity matrix of size s×s, vec($\mathbf{X}$) stacks the columns of matrix $\mathbf{X}$ in a column vector, diag($\mathbf{x}$)=diag($\mathbf{x}^T$) returns a diagonal matrix with the elements of $\mathbf{x}$ on the diagonal, $\otimes$ is the Kronecker product, $\mathbf{X}^*$ means complex conjugate, $\mathbf{X}^T$ means transpose, and $\mathbf{X}^H$ means Hermitian (conjugate transpose).

We consider the transmission of a training sequence over a flat fading MIMO link in Fig. 1

$$\boldsymbol{Y} = \boldsymbol{HFS} + \boldsymbol{N}, \tag{1}$$

where $\mathbf{S}$ is a M_{TX}×N_t training sequence of length N_t symbols and M_{TX} is the number of TX antennas. We presume orthogonal training sequences. A possible choice for a training sequence fulfilling this criterion could be a standard DFT matrix with elements

$$[\boldsymbol{S}]_{k,l} = e^{-j2\pi\frac{(k-1)(l-1)}{N_t}}, \quad \begin{cases} 1 \le k \le M_{TX} \\ 1 \le l \le N_t \end{cases} \tag{2}$$

with the orthogonality property

$$\boldsymbol{SS}^{H} = \boldsymbol{S}^{H}\boldsymbol{S} = N_t \cdot \boldsymbol{I}. \tag{3}$$

$\mathbf{F}$ is a M_{TX}×M_{TX} linear matrix transmit prefilter. We mention that the product $\mathbf{FS}$ could also be interpreted as a new training sequence $\tilde{\boldsymbol{S}}$, however, due to the invertibility of $\mathbf{S}$, both formulations are mathematically equivalent.

$\mathbf{H}$ is the M_{RX}×M_{TX} MIMO channel matrix with correlated Rayleigh fading elements, $\mathbf{N}$ is the M_{RX}×N_t noise matrix with

$$\boldsymbol{N} = \left[\tilde{\boldsymbol{n}}_1 \;\ldots\; \tilde{\boldsymbol{n}}_{N_t}\right], \tag{4}$$

and covariance matrix of the noise column vectors

$$\mathrm{E}\left[\tilde{\boldsymbol{n}}_i\tilde{\boldsymbol{n}}_i^{H}\right] = \boldsymbol{R}_{\tilde{n}\tilde{n}}, \quad 1 \le i \le N_t. \tag{5}$$

Matrix **Y** is the noisy $M_{RX}\times N_t$ receive sequence (see Fig. 1). By M_{RX} we denote the number of RX antennas. By appropriate processing of the received training sequence **Y,** the MIMO receiver is capable of producing a channel estimate $\hat{\boldsymbol{H}}$ in Fig. 1.

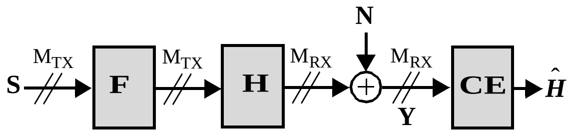

Fig. 1: System model

Using a widely accepted simplified channel model (e.g. [6][7]), the correlated MIMO channel can be described by the matrix product

$$\boldsymbol{H} = \boldsymbol{A}^H \boldsymbol{H}_w \boldsymbol{B}, \tag{6}$$

where $\mathbf{H}_w$ is a $M_{RX}\times M_{TX}$ matrix of complex i.i.d. Gaussian variables of unity variance and

$$\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{R}_{RX} \qquad \boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{R}_{TX}, \tag{7}$$

where $\mathbf{R}_{RX}$ and $\mathbf{R}_{TX}$ is the long-term stable (normalized) receive and transmit correlation matrix, respectively, with

$$\mathrm{tr}(\boldsymbol{R}_{RX}) = M_{RX} \qquad \mathrm{tr}(\boldsymbol{R}_{TX}) = M_{TX} \tag{8}$$

in order to normalize the channel gain.

III. Linear MMSE MIMO Channel Estimation

In order to derive the MMSE MIMO channel estimator, we rewrite (1) in vector form in order to be able to apply standard results from estimation theory

$$\underbrace{\mathrm{vec}(\boldsymbol{Y})}_{\boldsymbol{y}} = \underbrace{((\boldsymbol{F}\boldsymbol{S})^{\mathrm{T}} \otimes \boldsymbol{I})}_{\boldsymbol{X}} \cdot \underbrace{\mathrm{vec}(\boldsymbol{H})}_{\boldsymbol{h}} + \underbrace{\mathrm{vec}(\boldsymbol{N})}_{\boldsymbol{n}}, \tag{9}$$

where we have used [8]

$$\mathrm{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A}) \cdot \mathrm{vec}(\boldsymbol{B}). \tag{10}$$

Denoting the covariance matrix of **h** by $\mathbf{R}_{hh}$, while the covariance matrix of **n** is

$$\boldsymbol{R}_{\boldsymbol{nn}} = \mathrm{E}[\boldsymbol{n}\boldsymbol{n}^H] = \boldsymbol{I} \otimes \boldsymbol{R}_{\tilde{n}\tilde{n}}, \tag{11}$$

the linear MMSE estimator of **h** is given by the well-known equation [9]

$$\hat{\boldsymbol{h}} = (\boldsymbol{R}_{hh}^{-1} + \boldsymbol{X}^H \boldsymbol{R}_{nn}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{H}} \boldsymbol{R}_{\tilde{n}\tilde{n}}^{-1} \cdot \boldsymbol{y}. \tag{12}$$

We deploy the overall MSE ε [9]

$$\begin{aligned} \varepsilon &= \mathrm{tr}(\boldsymbol{R}_{hh}^{-1} + \boldsymbol{X}^H \boldsymbol{R}_{nn}^{-1} \boldsymbol{X})^{-1} \\ &= \mathrm{tr}(\boldsymbol{R}_{hh}^{-1} + (\boldsymbol{F}^* \boldsymbol{F}^{\mathrm{T}} \otimes \boldsymbol{R}_{\tilde{n}\tilde{n}}^{-1}))^{-1} \end{aligned} \tag{13}$$

as a measure of the quality of the MIMO channel estimator. Note that for the derivation of (13) we have used (3) and the properties of the Kronecker product [8]

$$\begin{aligned} (\boldsymbol{A} \otimes \boldsymbol{B}) \cdot (\boldsymbol{C} \otimes \boldsymbol{D}) &= \boldsymbol{A}\boldsymbol{B} \otimes \boldsymbol{C}\boldsymbol{D} \\ (\boldsymbol{A} \otimes \boldsymbol{B})^{-1} &= \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1} \end{aligned}. \tag{14}$$

For the specific channel model in (6) it can be shown with (10) that the channel covariance reads

$$\boldsymbol{R}_{hh} = E[\boldsymbol{h}\boldsymbol{h}^H] = \boldsymbol{R}_{TX}^* \otimes \boldsymbol{R}_{RX} \tag{15}$$

and with (14) the MSE is given by

$$\varepsilon = \mathrm{tr}\,((\boldsymbol{R}_{TX}^*)^{-1} \otimes \boldsymbol{R}_{RX}^{-1} + N_t(\boldsymbol{F}^* \boldsymbol{F}^{\mathrm{T}} \otimes \boldsymbol{R}_{\tilde{n}\tilde{n}}^{-1}))^{-1}. \tag{16}$$

After introducing the eigenvalue decompositions (EVD) with matrices $\Lambda_{RX} = \mathrm{diag}(\lambda_{\mathrm{R},1}, \ldots, \lambda_{\mathrm{R},M_{RX}})$ and $\Lambda_{TX} = \mathrm{diag}(\lambda_{T,1}, \ldots, \lambda_{\mathrm{T},M_{TX}})$, which contain the sorted (descending) eigenvalues

$$\boldsymbol{R}_{RX} = \boldsymbol{V}_{RX}\Lambda_{RX}\boldsymbol{V}_{RX}^{\mathrm{H}} \qquad \boldsymbol{R}_{TX}^* = \boldsymbol{V}_{TX}\Lambda_{TX}\boldsymbol{V}_{TX}^{\mathrm{H}}, \tag{17}$$

we find from (16) after simple manipulations

$$\varepsilon = \mathrm{tr}\,(\Lambda_{TX}^{-1} \otimes \Lambda_{RX}^{-1} + N_t(\boldsymbol{V}_{TX}^{\mathrm{H}} \boldsymbol{F}^* \boldsymbol{F}^{\mathrm{T}} \boldsymbol{V}_{TX} \otimes \boldsymbol{V}_{RX}^{\mathrm{H}} \boldsymbol{R}_{\tilde{n}\tilde{n}}^{-1} \boldsymbol{V}_{RX}))^{-1}. \tag{18}$$

Without loss of generality ($\mathbf{V}_{TX}$ is unitary and thus invertible) we decompose **F** with general matrix Φ_f

$$\boldsymbol{F}^* = \boldsymbol{V}_{TX}\Phi_{\mathrm{f}} \tag{19}$$

and introduce for brevity the $M_{RX}\times M_{RX}$ matrix

$$\boldsymbol{R} = \boldsymbol{V}_{RX}^{\mathrm{H}} \boldsymbol{R}_{\tilde{n}\tilde{n}}^{-1} \boldsymbol{V}_{RX}. \tag{20}$$

Now (18) reads

$$\varepsilon = \mathrm{tr}\,(\Lambda_{TX}^{-1} \otimes \Lambda_{RX}^{-1} + N_t(\Phi_{\mathrm{f}}\Phi_{\mathrm{f}}^{\mathrm{H}} \otimes \boldsymbol{R}))^{-1}, \tag{21}$$

whereas we emphasize again that (21) is valid for the specific channel model in (6) with its Kronecker covariance structure (15).

IV. Training Sequence Prefilter Design

A. Structure of the prefilter

For the design of the optimum prefilter in the sense that it minimizes the MSE in (32), we have to to solve the constrained optimization problem under a power constraint ρ

$$\begin{aligned} &\underbrace{\min}_{\Phi_{\mathrm{f}}} \mathrm{tr}\,(\Lambda_{TX}^{-1} \otimes \Lambda_{RX}^{-1} + N_t(\Phi_{\mathrm{f}}\Phi_{\mathrm{f}}^{\mathrm{H}} \otimes \boldsymbol{R}))^{-1} \\ &\text{s.t. } \mathrm{tr}(\Phi_{\mathrm{f}}\Phi_{\mathrm{f}}^{\mathrm{H}}) = \rho \end{aligned}, \tag{22}$$

where 's.t.' stands for subject to and we normalize the transmit power to $\rho = M_{TX}$. For solving (22), we state the following

Definition 1. [10**,** 7.7.1]
Let **A**, **B**$\in M_n$ (where M_n is the set of complex n×n matrices) be Hermitian matrices. We write A≥B, if the matrix **A-B** is positive semidefinite. Equivalently, A>B means **A-B** is positive definite.

Theorem 1. [10, 7.7.8]
Let **A**$\in M_n$ be positive definite and S⊂{1,2,…,n} be an index set. Then $[\boldsymbol{A}^{-1}]_S \geq ([\boldsymbol{A}]_S)^{-1}$, where the left-hand side is the

principal submatrix of $\mathbf{A}^{-1}$ determined by deletion of the rows and columns indicated by S, while the right-hand side is the inverse of the corresponding submatrix of $\mathbf{A}$. The equality is only valid for diagonal block matrices $\mathbf{A}$.

Corollary 1. [10, 7.1.5]
The trace, the determinant, and all principal minors of a positive definite matrix are positive.

First we note from (21) that

$$\Lambda_{TX}^{-1} \otimes \Lambda_{RX}^{-1} + N_t(\Phi_f \Phi_f^H \otimes \boldsymbol{R}) > 0 \,. \tag{23}$$

By Theorem 1 in combination with Corollary 1 it can readily be checked that for minimum MSE in (22), the matrix in (23) has to be block-diagonal, which can only be fulfilled with diagonal $\Phi_f \Phi_f^H$. Without loss of generality we can therefore assume that the optimal

$$\Phi_f = \operatorname{diag}(\phi_1, \ldots, \phi_{M_{TX}}) \tag{24}$$

is a real diagonal power allocation (PA) matrix that assign transmit power to the different long-term eigenmodes of the channel. To this end, note that the constrained optimization problem (22) is a function of $\Phi_f \Phi_f^H$ alone. Due to the block-diagonal structure, with the general formula

$$[\operatorname{diag}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)]^{-1} = \operatorname{diag}(\boldsymbol{X}_1^{-1}, \ldots, \boldsymbol{X}_N^{-1}) \tag{25}$$

the MSE in (21) and (22) can be rewritten as

$$\varepsilon = \sum_{k=1}^{M_{TX}} \operatorname{tr}(\lambda_{T,k}^{-1} \cdot \Lambda_{RX}^{-1} + N_t \phi_k^2 \cdot \boldsymbol{R})^{-1} \,. \tag{26}$$

In order to further simplify matters and to gain a better insight into the problem, in the following we focus on the AWGN case with $\mathbf{R}=1/N_0 \cdot \mathbf{I}$ [see (20)], such that (26) reduces to

$$\begin{aligned} \varepsilon &= \sum_{k=1}^{M_{TX}} \operatorname{tr}\left(\lambda_{T,k}^{-1} \cdot \Lambda_{RX}^{-1} + \frac{N_t}{N_0}\phi_k^2 \cdot \boldsymbol{I}\right)^{-1} \\ &= \sum_{k=1}^{M_{TX}} \sum_{l=1}^{M_{RX}} \frac{\lambda_{T,k} \cdot \lambda_{R,l}}{\left(1 + \frac{N_t}{N_0}\phi_k^2 \cdot \lambda_{T,k} \cdot \lambda_{R,l}\right)} , \end{aligned} \tag{27}$$

which is the basis for the following derivations. For high SNR ($N_0 \rightarrow 0$), we get from (27)

$$\varepsilon \approx \sum_{k=1}^{M_{TX}} \sum_{l=1}^{M_{RX}} \frac{\lambda_{T,k} \cdot \lambda_{R,l}}{\frac{N_t}{N_0}\phi_k^2 \cdot \lambda_{T,k} \cdot \lambda_{R,l}} = \frac{N_0 \cdot M_{RX}}{N_t} \cdot \sum_{k=1}^{M_{TX}} \frac{1}{\phi_k^2} \,. \tag{28}$$

Obviously, the channel estimation error is independent of the correlation properties of the channel at high SNR.

B. Design of Φ_f for both RX and TX correlation

Applying the method of Lagrange multipliers (with multiplier μ) for minimizing the MSE in (27) under a power contraint leads to the condition

$$-\lambda_{T,l}^2 \cdot \frac{N_t}{N_0} \cdot \sum_{k=1}^{M_{RX}} \frac{\lambda_{R,k}^2}{1 + \frac{N_t}{N_0} \cdot \lambda_{T,l}\lambda_{R,k}\phi_{f,l}^2} + \mu = 0 \,. \tag{29}$$

for all $1 \le l \le M_{TX}$. A closed form solution for the general case of arbitrary array sizes is not possible (the equation in (29) essentially reduces to finding the roots of a polynomial) and one has to resort to numerical optimization methods. However, one can find solutions for the low and high SNR regions.

For high SNR ($N_0 \rightarrow 0$), we get from (28) by Lagrange optimization or majorization theory [11] the optimal $\Phi_f = \rho/M_{TX} \cdot \mathbf{I}$, resulting in $\mathbf{F} = \rho/M_{TX} \cdot \mathbf{I}$, i.e. there is essentially no prefilter. We conclude that standard orthogonal training sequences are optimal in the high SNR region.

On the other hand, for low SNR ($N_0 \rightarrow \infty$), we get from (27) with the series expansion for small x

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-1) \cdot x^k \tag{30}$$

and (8) the MSE approximation

$$\varepsilon \approx M_{RX} \cdot M_{TX} - \frac{N_t}{N_0} \cdot \sum_{k=1}^{M_{TX}} \phi_k^2 \cdot \lambda_{T,k}^2 \sum_{l=1}^{M_{RX}} \lambda_{R,l}^2 \,. \tag{31}$$

The second term should be maximized for minimizing the MSE. To this end, it is obvious that we should pour all power on the strongest long-term eigenmode of the channel, i.e. use all power for the ϕ_k corresponding to the maximum λ_k. Obviously, standard training sequences are clearly suboptimum in the low SNR region. Interestingly, an optimum training sequence design for this region consists of a one-dimensional training sequence that is transmitted on the strongest long-term eigenmode of the channel.

C. Design of Φ_f for TX correlation only

Without RX correlation, from (29) we can compute a closed-form equation for Φ_f, namely

$$\Phi_f = \left[\mu^{-1/2} \cdot \left(\frac{N_t}{N_0}\right)^{-1/2} \cdot \boldsymbol{I} - \left(\frac{N_t}{N_0}\right)^{-1} \Lambda_{TX}^{-1}\right]_+^{1/2} , \tag{32}$$

The Lagrange multiplier can be determined via the power constraint and reads

$$\mu^{-1/2} = \frac{\left(\frac{N_t}{N_0}\right)^{-1} \cdot \operatorname{tr}(\Lambda_{TX}^{-1}) + \rho}{\left(\frac{N_t}{N_0}\right)^{-1/2} \cdot \operatorname{tr}(\boldsymbol{I}_{M_{TX}})} \,. \tag{33}$$

In (32) the '+' sign indicates that all diagonal elements have to be greater or equal to 0. However, from (32) this is equivalent to

$$\mu^{-1/2} \ge \left(\frac{N_t}{N_0}\right)^{-1/2} \cdot \frac{1}{\lambda_{T,M_{TX}}} , \tag{34}$$

An iterative procedure can be used to assure (34), where the last diagonal element of Φ_f is subsequently set to 0 and the matrices involved in the calculation of (33) and (34) are reduced by one element. Inserting (33) in (34) and solving for N_0 one can determine the switching point, where the transmitter switches from 1 to 2 active eigenmodes used for transmission of the training sequence

$$N_0\big|_{\text{switch},1,2} = \frac{\rho N_t}{\frac{1}{\lambda_{T,2}} - \frac{1}{\lambda_{T,1}}}. \tag{35}$$

For longer training sequences, the switching point obviously shifts to lower SNR values. Furthermore, if a vector $\left[\lambda_{T,1}\ \lambda_{T,2}\right]^T$ is majorized by a vector $\left[\tilde{\lambda}_{T,1}\ \tilde{\lambda}_{T,2}\right]^T$ (higher correlation), then the switching point moves to higher SNR values. Equation (35) can be generalized for the switching point between L and L+1 eigenmodes

$$N_0\big|_{\text{switch},L,L+1} = \frac{\rho N_t}{\frac{L}{\lambda_{T,L+1}} - \sum_{l=1}^{L}\frac{1}{\lambda_{T,l}}}. \tag{36}$$

Higher channel correlation and shorter training sequences shift the switching points to higher SNR, such that in these cases prefiltering becomes more effective.

D. Design of Φ_f for RX correlation only

On the other hand, if the fading at the transmit antenna array is completely uncorrelated, i.e. $\Lambda_{TX}=\mathbf{I}$, and there is only correlation present at the receiver, (29) reads again for all $1 \le l \le M_{TX}$

$$\frac{N_t}{N_0} \cdot \sum_{k=1}^{M_{RX}} \frac{\lambda_{\mathrm{RX},k}^2}{1 + \frac{N_t}{N_0} \cdot \lambda_{\mathrm{RX},k}\phi_{\mathrm{f},l}^2} + \mu = 0. \tag{37}$$

This can only be fulfilled if all $\phi_{f,l}$ agree in size, i.e. for this special case we find in matrix notation $\Phi_f=\rho/M_{TX}\cdot\mathbf{I}$, implying that the training sequence is left unchanged. This result agrees with intuition. When there is no transmit correlation present, there are no prominent directions and the transmitter equally distributes power.

V. Structure Of MMSE Channel Estimator

In an AWGN environment and with the optimal prefilter structure according to (19), it can be shown from (12) that the channel estimator is given by

$$\hat{\boldsymbol{h}} = (\boldsymbol{V}_{TX} \otimes \boldsymbol{V}_{RX})(N_0 \cdot \Lambda_{TX}^{-1} \otimes \Lambda_{RX}^{-1} + N_t(\Phi_f\Phi_f^H \otimes \boldsymbol{I}))^{-1}(\Phi_f\boldsymbol{S}^* \otimes \boldsymbol{V}_{RX}^H) \cdot \boldsymbol{y}. \tag{38}$$

Obviously, with the optimum training sequence design the receiver just has to invert a diagonal matrix, resulting in simple scalar operations and thus low complexity. In particular, for vanishing RX correlation we find

$$\hat{\boldsymbol{h}} = (\boldsymbol{V}_{TX}(N_0 \cdot \Lambda_{TX}^{-1} + N_t\Phi_f\Phi_f^H)^{-1}\Phi_f\boldsymbol{S}^* \otimes \boldsymbol{I}) \cdot \boldsymbol{y}. \tag{39}$$

Using (10), it can be seen that (39) is equivalent to the matrix equation

$$\hat{\boldsymbol{H}} = \boldsymbol{Y} \cdot \boldsymbol{S}^H\Phi_f \cdot (N_0 \cdot \Lambda_{TX}^{-1} + N_t\Phi_f\Phi_f^H)^{-1}\boldsymbol{V}_{TX}^T. \tag{40}$$

Note that estimated channel coefficients for one particular RX antenna are a function of the received training sequence at that antenna only. As there is no correlation between the RX antenna elements, no information can be acquired from the signal at other RX antennas. Again, the diagonalization in the eigenmode domain becomes obvious in (40). In the low SNR region it was shown above that only the strongest eigenmode is used for training sequence transmission. In this special case, (40) reduces to

$$\hat{\boldsymbol{H}} = \sqrt{\rho} \cdot \frac{\lambda_{T,1}}{N_0 + \rho N_t \lambda_{T,1}} \cdot \boldsymbol{Y} \cdot \boldsymbol{s}_1 \cdot \boldsymbol{v}_1^T, \tag{41}$$

where we have defined

$$\boldsymbol{S}^H = \begin{bmatrix}\boldsymbol{s}_1 & \dots & \boldsymbol{s}_{M_{TX}}\end{bmatrix} \qquad \boldsymbol{V}_{TX}^T = \begin{bmatrix}\boldsymbol{v}_1^T \\ \cdot \\ \boldsymbol{v}_{M_{TX}}^T\end{bmatrix}. \tag{42}$$

Note that the channel estimator in (41) thus reduces to a rank 1 matrix.

On the other hand, consider the case of RX correlation only, where (38) reduces to

$$\hat{\boldsymbol{h}} = (\boldsymbol{S}^* \otimes (N_0\boldsymbol{R}_{RX}^{-1} + N_t\boldsymbol{I})^{-1}) \cdot \boldsymbol{y}. \tag{43}$$

Applying (10), we have the matrix equivalent

$$\hat{\boldsymbol{H}} = (N_0\boldsymbol{R}_{RX}^{-1} + N_t\boldsymbol{I})^{-1} \cdot \boldsymbol{Y} \cdot \boldsymbol{S}^H. \tag{44}$$

VI. MSE And BER Simulations

We study the effects of statistically shaping the training sequence according to the correlation properties of the channel with a prefilter **F** designed according to (19) and matrix Φ_f according to (32), i.e. we are focusing on the case of TX correlation only. To this end, we investigate the overall MSE of the channel estimator given in (18) or (21), respectively.

In Fig. 2 we have plotted the resulting MSE of the channel estimator for a 4×4 system with uncorrelated fading for reference and with TX correlation according to a single main direction of departure, Laplacian power distribution and an angular spread (AS) of 2 degrees (strong correlation). For the correlated case, we show results with and without prefiltering of a DFT training sequence. The antenna array at the transmitter has an antenna element spacing of 0.5 wavelengths. We emphasize that this scenario corresponds to a heavily correlated channel suited for maximum eigenmode transmission (beamforming), with a TX correlation matrix given by

$$\mathrm{abs}(\boldsymbol{R}_{TX}) \approx \begin{bmatrix} 1 & 0.99 & 0.98 & 0.95 \\ 0.99 & 1 & 0.99 & 0.98 \\ 0.98 & 0.99 & 1 & 0.99 \\ 0.95 & 0.98 & 0.99 & 1 \end{bmatrix}. \tag{45}$$

With a length N_t=4 training sequence one can observe a significant reduction of the MSE with prefiltering over a wide range in the low SNR region with a maximum gain factor of about 3 in the range of -10 to 10 dB. The insights gained from the MSE simulations are the key for effectively deploying the proposed CE scheme. Obviously, it can only improve the system performance, if the operating point of the system agrees with the range of MSE improvements.

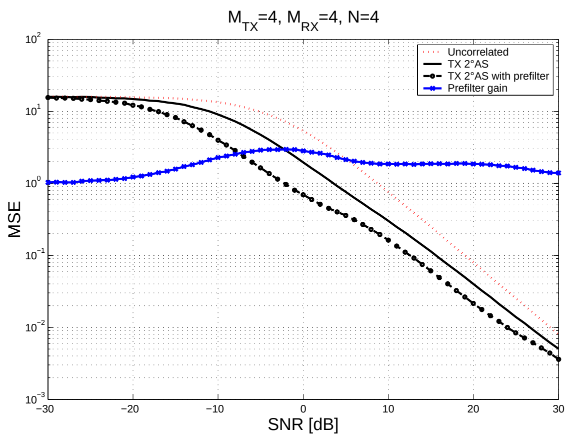

Fig. 2: MSE and gain with TX AS 2°, M_{TX}=4, M_{RX}=4, N_t=4

For the same scenario the squared power allocation coefficients are depicted in (3), where the switching points according to (35) and (36) clearly emerge.

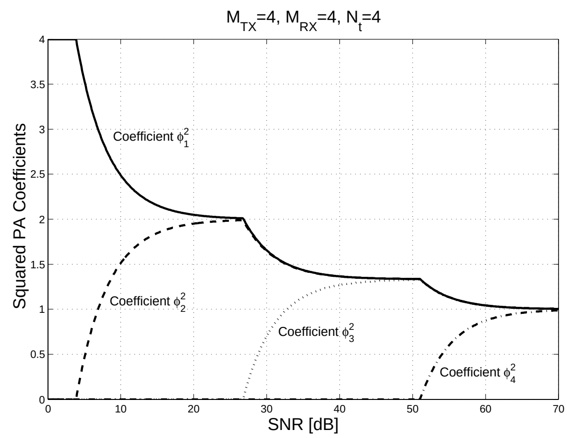

Fig. 3: PA coefficients with TX AS 2°, M_{TX}=4, M_{RX}=4, N_t=4

As an example, we consider the downlink of a MIMO system with maximum eigenmode transmission. In order to separate the effects of CE at RX and TX, we presume ideal CE at the TX, while RX CE is based on the novel CE scheme. In Fig. 4 we have plotted BER simulation results for a 4×4 system with QPSK modulation and N_t={8,16}. The performance improvement due to enhanced CE is obvious in this strongly semi-correlated scenario with again an AS of 2 degrees at the TX. As expected, in the given SNR range the gain is higher with a shorter training sequence.

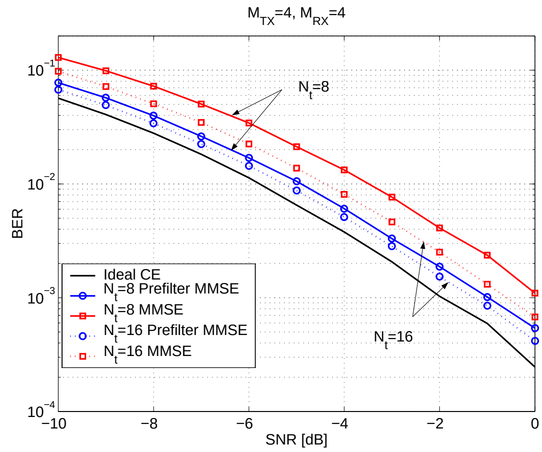

Fig. 4: BER of maximum eigenmode transmission, M_{TX}=4, M_{RX}=4

References

[1] Marzetta T. L., “BLAST training: estimating channel characteristics for high capacity space-time wireless”, Annual Allerton Conference on Communication, Control, and Computing, Sept. 1999

[2] Cavers J. K., “An analysis of pilot symbol assisted modulation for Rayleigh fading channels”, *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686-693, Nov. 1991

[3] Qinfang Sun, Cox D.C., Huang, H.C., Lozano A., “Estimation of continuous flat fading MIMO channels”, IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 549 -553, Oct. 2002

[4] Hochwald B.M., Marzetta T.L., “Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading”, *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543 -564, March 2000

[5] Baltersee J., Fock G., Meyr H., “Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving”, *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 12, Dec. 2001

[6] D. Shiu, G. J. Foschini, M. J. Gans, J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems", *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502-513, March 2000

[7] Chuah C. -N., Tse D. N. C., Kahn J. M., Valenzuela R. A., “Capacity Scaling in MIMO wireless systems under correlated fading”, *IEEE Transactions on Information Theory*, vol. 48, no. 3, March 2002

[8] Lütkepohl H., *Handbook of matrices*, John Wiley&Sons, 1996

[9] Kay S. M., *Fundamentals of statistical signal processing volume 1 (estimation theory)*, Prentice Hall, 1993

[10] Horn R. A., Johnson C. R., *Matrix analysis*, Cambridge University Press, 1985

[11] Marshall A. W., Olkin I., *Inequalities: theory of majorization and its applications*, Academic Press, 1979

[12] Kiessling M., Speidel J., “Analytical performance of MIMO zero-forcing receivers in correlated Rayleigh fading environments”, SPAWC, June 2003, unpublished

[13] Kiessling M., Viering I., Reinhardt M., Speidel J., "Short-term and long-term diagonalization of correlated MIMO channels with adaptive modulation," PIMRC, Sept. 2002

[14] Kiessling M., Speidel J., Geng N., “Performance analysis of MIMO maximum likelihood receivers with channel correlation, colored gaussian noise, and linear prefiltering”, ICC, May 2003, in press