2003 4th IEEE Workshop on Signal Processing
Advances in Wireless Communications

LINEAR AND NON-LINEAR PRECODING/DECODING FOR MIMO SYSTEMS USING THE FADING CORRELATION AT THE TRANSMITTER

O. Simeone[(1)], *U. Spagnolini*[(1)] *and Y. Bar-Ness*[(2)]

[(1)]Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milano (Italy)
[(2)]CCSPR, New Jersey Institute of Technology, Newark (NJ, USA)

ABSTRACT

A transceiver structure for MIMO systems that comprises linear/non-linear precoding/decoding is optimized according to the MMSE criterion under the assumption that only long-term channel state information (or LT-CSI, i.e., fading channel and noise correlation matrices) is available at the transmitter. The structure generalizes different techniques known from the literature, such as BLAST, linear precoding and decoding and Tomlinson-Harashima precoding. Performance comparison of these known techniques and the proposed scheme, carried out by means of simulations, shows that relevant benefits can be obtained by taking advantage of LT-CSI at the transmitter.

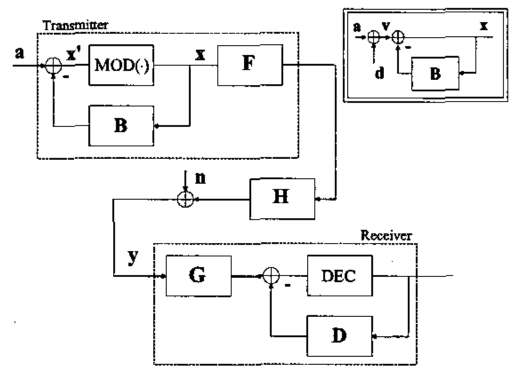

Fig. 1. Block diagram of the generalized transceiver.

1. INTRODUCTION

In order to achieve the high spectral efficiencies promised by the information theory over a radio link with multiple antennas at both the transmitter and the receiver (i.e., MIMO link), different approaches have been proposed (e.g., space-time codes and V-BLAST [1]). In this paper, we are concerned with the optimization of the transceiver structure shown in fig. 1 under the constraint that the channel state information (CSI) at the transmitter is limited to the second order statistics of channel fading and noise [2] [3] (long-term CSI or, in short, LT-CSI). This assumption is of crucial relevance for systems in which the fading channel is sufficiently fast-varying to make the condition of instantaneous CSI (I-CSI) at the transmitter not realistic. The LT-CSI can be acquired by the transmitter either directly from measurements of the opposite link or by feedback from the receiver. On the other hand, in the design of the receiver, the instantaneous realization of the channel matrix $\mathbf{H}$ (i.e., I-CSI) is assumed known (effects of channel estimation errors are studied by means of simulations). Linear and non-linear precoding and decoding are considered in the scheme of fig. 1. The structure reduces to known systems for specific constraints on the matrices $\{\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}\}$. For instance, imposing $\mathbf{B} = \mathbf{0}$ and $\mathbf{F} = \mathbf{I}$ the scheme reduces to a decision feedback equalizer (or equivalently to the V-BLAST receiver without optimal ordering [1]); for $\mathbf{B} = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$ we have the linear precoding-linear decoding structure (LP-LD), studied, e.g., in [4] [5], under the assumption of I-CSI at both the transmitter and the receiver; for $\mathbf{F} = \mathbf{I}$ and $\mathbf{D} = \mathbf{0}$ the Tomlinson-Harashima precoding (THP) structure proposed in [6] and studied for LT-CSI at the transmitter in [3] is obtained.

This work was partially supported by NSF under grant CCR-0085846.

The paper is organized as follows. The transceiver scheme depicted in fig. 1 is described in Sec. 2 along with the basic assumptions. The expressions for the optimal linear/non linear precoding/decoding operators $\{\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}\}$ are derived and discussed in the context of the existing literature on the subject in Sec. 3. Sec. 4 compares different structures obtained from the scheme of fig. 1 in terms of uncoded symbol error probability (SER) by means of simulation.

2. SIGNAL MODEL

We focus on a MIMO wireless link with an equal number of transmit and receive antennas N. The $N \times 1$ data vector $\mathbf{a}$ (the time dependence of all the variables is implied) is composed of complex symbols taken from the $M-$QAM constellation, i.e., each entry a_i ($i = 1, .., N$) belongs to the set $\mathcal{A} = \{a^I + ja^Q | a^I, a^Q \in \{\pm 1, \pm 3, .., \pm\sqrt{M} - 1\}\}$. The data vector is passed through the non-linear part of the precoder defined by the $N \times N$ strictly upper triangular matrix $\mathbf{B}$ (i.e., $B_{ii} = 0$). In order to stabilize the precoder, or equivalently to limit the dynamic range of the precoded sequence, a non-linear modulo-arithmetic operation (MOD) is introduced, as it is done in THP (see, e.g., [7]). This operation performs a periodic mapping (or modulo reduction) of its input $\mathbf{x}'$ on the square region of the complex plane that contains $\mathcal{A}$ and has side length $2\sqrt{M}$, i.e., $\mathcal{R} = \{x^I + jx^Q | x^I, x^Q \in (-\sqrt{M}, \sqrt{M}]\}$. In other words, $x_i = MOD(x_i') = x_i' + 2\sqrt{M}k_i$, where the real and imaginary parts of k_i are integers chosen to reduce $x_i \in \mathcal{R}$. Notice that there is only one k_i that satisfies this condition. It follows that the non-linear part of the precoder can be equivalently redrawn by deleting the block MOD and adding at the input an input-dependent vector $\mathbf{d}$ (see box in fig. 1) such that $d_i = 2k_i\sqrt{M}$. Therefore, the effective symbols input to the non-linear precoder are $\mathbf{v} = \mathbf{a} + \mathbf{d}$. After linear precoding with the $N \times N$ matrix $\mathbf{F}$ and propagation through the $N \times N$ radio channel $\mathbf{H}$, the received signal can be written as

$$\mathbf{y} = \mathbf{HFx} + \mathbf{n} \tag{1}$$

where the circularly symmetric Gaussian noise has correlation $E[\mathbf{nn}^H] = \mathbf{R}_n$ and $E[\mathbf{xx}^H] = \sigma_x^2\mathbf{I}$. Notice that the latter assumption, also made in [6] and [3], is not rigorously satisfied when $\mathbf{B} \neq \mathbf{0}$ since in this case $E[\mathbf{xx}^H]$ is a function of the unknown (i.e., design target) matrix $\mathbf{B}$. In optimizing the scheme of fig. 1, the power constraint $E[||\mathbf{Fx}||^2] = \sigma_x^2 \operatorname{tr}\{\mathbf{FF}^H\} = \sigma_x^2 N$ will be imposed. This condition is clearly satisfied when no linear precoding is employed ($\mathbf{F} = \mathbf{I}$).

The channel matrix $\mathbf{H}$ is assumed to be zero-mean (Rayleigh fading) circularly symmetric complex Gaussian distributed with a separable spatial correlation function [8]. The correlation between the channel gains H_{ij} and $H_{\ell m}$, i.e., $E[H_{ij}H_{\ell m}^*]$, is thus given by the product of the spatial correlation at the receiver $\gamma_R(i, \ell)$ and the spatial correlation at the transmitter $\gamma_T(j, m)$ so that

$$\mathbf{H} = \mathbf{R}_R^{H/2}\mathbf{H}_w\mathbf{R}_T^{1/2}, \tag{2}$$

where the correlation matrices $\mathbf{R}_R$ and $\mathbf{R}_T$ are defined as $R_{R,i\ell} = \gamma_R(i, \ell)$ and $R_{T,jm} = \gamma_T(j, m)$ and $\mathbf{H}_w$ is a matrix of independent identically distributed circularly symmetric complex Gaussian variables with unit power. The notation $(\cdot)^{1/2}$ defines the Cholesky factorization. We set the Frobenius norm of the channel matrix as $E[||\mathbf{H}||^2] = \operatorname{tr}\{\mathbf{R}_T\}\operatorname{tr}\{\mathbf{R}_R\} = N$ to account for the linear increase of the SNR as a function of the number of receive antennas. For simplicity, the numerical evaluation of the performance of the presented algorithms will be carried out for an AR(1) model of the spatial correlation: $\gamma_R(i, \ell) = 1/\sqrt{N}\rho_R^{|i-\ell|}$ and $\gamma_T(j, m) = 1/\sqrt{N}\rho_T^{|j-m|}$. For later use, we remark that the correlation function of the channel (that is related to the LT-CSI available at the transmitter, see next Section) is

$$E[\mathbf{H}^H\mathbf{H}] = \mathbf{R}_T \cdot \operatorname{tr}\{\mathbf{R}_R\} = \sqrt{N}\mathbf{R}_T. \tag{3}$$

At the receiver side, the signal is linearly processed by the $N \times N$ matrix $\mathbf{G}$, passed through the feedback loop defined by the $N \times N$ strictly upper triangular matrix $\mathbf{D}$ and modulo reduced into $\mathcal{R}$ if $\mathbf{B} \neq \mathbf{0}$ (not shown in fig. 1).

3. MMSE-BASED PRECODING AND DECODING

Here we optimize the general transceiver scheme of fig. 1 by minimizing the mean square error (MSE) between the variables at the input of the decision device and the effective data symbols $\mathbf{v} = \mathbf{a} + \mathbf{d}$ [6]. As previously discussed, we constraint the design of the operators at the transmitter side, i.e., of the matrices $\mathbf{F}$ and $\mathbf{B}$, to be based only on LT-CSI, represented by the second order statistics of channel and noise. In particular, the transmitter is given only the correlation matrix $E[\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}]$. On the other hand, the operators $\mathbf{G}$ and $\mathbf{D}$ at the receiver side are allowed to depend directly on the I-CSI, i.e., on the channel matrix $\mathbf{H}$. Furthermore, we will assume perfect error recovery at the output of the decision device as it is usually done in the literature on decision feedback equalization. To conclude the set of hypotheses, we recall that the transmitted power is constrained as $E[||\mathbf{Fx}||^2] = \sigma_x^2 \operatorname{tr}\{\mathbf{FF}^H\} = \sigma_x^2 N$.

We now proceed with the derivation of the optimum precoding and decoding matrices. Since the vector at the input of the decision device can be written as $\mathbf{Gy} - \mathbf{Dv}$ (recall the assumption of no error propagation made above), the design problem can be stated as

$$\{\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}\} = \arg \min_{\{\mathbf{B},\mathbf{F},\mathbf{G},\mathbf{D}\}} E[||\mathbf{Gy} - \mathbf{Dv} - \mathbf{v}||^2], \tag{4}$$

we will show below how we take into account the different types of CSI's at the transmitter and the receiver. From fig. 1 one can easily show that $\mathbf{v} = \mathbf{Cx}$ where $\mathbf{C} = \mathbf{I} + \mathbf{B}$, so that

$$MSE(\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}) = E[||\mathbf{Gy} - \mathbf{ECx}||^2] \tag{5}$$

with $\mathbf{E} = \mathbf{I} + \mathbf{D}$. The upper triangular (with unit diagonal) feedback matrices $\mathbf{C}$ and $\mathbf{E}$ (or equivalently $\mathbf{B}$ and $\mathbf{D}$) can not be independently identified using the MMSE criterion. In the following we will thus set $\mathbf{K} = \mathbf{EC}$ and restate (5) as $MSE(\mathbf{F}, \mathbf{G}, \mathbf{K}) = E[||\mathbf{Gy} - \mathbf{Kx}||^2]$. We remark that $\mathbf{K}$ is still an upper triangular matrix with unit diagonal.

From the standard theory of Wiener linear filtering, we get $\mathbf{G} = \mathbf{K}E[\mathbf{x}\mathbf{y}^H]E[\mathbf{y}\mathbf{y}^H]^{-1}$ and after algebraic manipulations

$$\begin{aligned}\mathbf{G} &= \mathbf{K}\mathbf{F}^H\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\mathbf{F}\mathbf{F}\tilde{\mathbf{H}}^H + 1/\sigma_x^2\mathbf{I})^{-1}\cdot\mathbf{R}_n^{-H/2} = \\ &= \mathbf{K}\mathbf{Q}^{-1}\mathbf{F}^H\tilde{\mathbf{H}}^H\cdot\mathbf{R}_n^{-H/2} \qquad (6)\end{aligned}$$

where

$$\mathbf{Q} = \mathbf{F}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{F} + 1/\sigma_x^2\mathbf{I} \qquad (7)$$

and $\tilde{\mathbf{H}} = \mathbf{R}_n^{-H/2}\mathbf{H}$. The result (6) states the the optimum linear filter at the front-end of the receiver performs the whitening of the received signal and then applies a linear operator that has the classical Wiener structure. Substituting (6) into the expression of $MSE(\mathbf{F}, \mathbf{G}, \mathbf{K})$ we obtain

$$MSE(\mathbf{F}, \mathbf{K}) = \mathrm{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}. \qquad (8)$$

In our framework, minimizing (8) w.r.t. $\mathbf{K}$ leads to different results depending on the way the matrix $\mathbf{K}$ is factorized into the transmitter ($\mathbf{C}$) and receiver ($\mathbf{E}$) part. Here, we consider two cases:

1) *non-linear decoding* ($\mathbf{C} = \mathbf{I} \Rightarrow \mathbf{K} = \mathbf{E}$): since the receiver has access to the I-CSI $\mathbf{H}$, the feedback matrix $\mathbf{K} = \mathbf{E}$ is obtained as

$$\mathbf{E} = \mathbf{V}\mathbf{Q}^{1/2}, \qquad (9)$$

where $\mathbf{V}$ is a diagonal matrix that scales to unity the elements on the main diagonal of $\mathbf{K}$. The non-linear decoding matrix $\mathbf{E}$ in (9) follows from the observation that (8) is minimized when $\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H$ is diagonal [6];

2) *non-linear precoding* [3] ($\mathbf{E} = \mathbf{I} \Rightarrow \mathbf{K} = \mathbf{C}$): since the transmitter is given only the LT-CSI $E[\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}] = E[\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}]$ we can not minimize (8). Instead, it is reasonable to consider $E[\mathrm{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}]$ as the loss function. It can be shown that $E[\mathrm{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}] \geq \mathrm{tr}\{\mathbf{K}\bar{\mathbf{Q}}^{-1}\mathbf{K}^H\}$ where

$$\bar{\mathbf{Q}} = \mathbf{F}^H E[\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}]\mathbf{F} + 1/\sigma_x^2\mathbf{I} \qquad (10)$$

(see Appendix). Therefore, similarly to the approach of [2] and [3], we minimize the lower bound $\overline{MSE}(\mathbf{F}, \mathbf{K}) = \mathrm{tr}\{\mathbf{K}\bar{\mathbf{Q}}^{-1}\mathbf{K}^H\}$ obtaining

$$\mathbf{C} = \mathbf{V}\bar{\mathbf{Q}}^{1/2}, \qquad (11)$$

where $\mathbf{V}$ is the scaling matrix as in (9).

After substitution of (9) if $\mathbf{C} = \mathbf{I}$ or (11) if $\mathbf{E} = \mathbf{I}$ according to the two cases discussed above, we should in principle minimize $\overline{MSE}(\mathbf{F}, \mathbf{K}) = \mathrm{tr}\{\mathbf{K}\bar{\mathbf{Q}}^{-1}\mathbf{K}^H\}$ with respect to the transmit precoding matrix $\mathbf{F}$. To make the problem analytically tractable and obtain a solution independent on $\mathbf{K}$, we minimize $\mathrm{tr}\{\bar{\mathbf{Q}}^{-1}\}$ instead. In other words, the matrix $\mathbf{F}$ is designed by assuming that neither non-linear precoding nor non-linear decoding is included in the transceiver ($\mathbf{K} = \mathbf{I}$). Nonetheless, simulation results show that the so obtained linear precoder performs satisfactorily even for $\mathbf{K} \neq \mathbf{I}$ (see Sec. 4). The precoder can now be obtained following the steps outlined in [4]. It is

$$\mathbf{F} = \mathbf{U}\Phi \qquad (12)$$

where $\mathbf{U}$ is obtained from the eigenvalue decomposition of the LT-CSI $E[\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}] = \mathbf{U}\Lambda\mathbf{U}^H$ and Φ is a diagonal matrix such that

$$|\Phi_{ii}|^2 = \left(\frac{N + \sum_{n=1}^{\bar{N}}\lambda_{nn}^{-1}}{\sigma_x^2\sum_{n=1}^{\bar{N}}\lambda_{nn}^{-1/2}}\lambda_{ii}^{-1/2} - \frac{1}{\lambda_{ii}\sigma_x^2}\right)^+, \qquad (13)$$

where $(x)^+ = \max(x, 0)$ and $\bar{N} \leq N$ is such that $|\Phi_{nn}|^2 > 0$ for $n \in [1, \bar{N}]$ and $|\Phi_{nn}|^2 = 0$ for all other n. Since $\bar{N}$ is a function of Φ_{ii}, (13) requires to be evaluated iteratively starting with $\bar{N} = N$ and decreasing $\bar{N}$ progressively by one (for details, refer to [4]).

Some remarks on the results of the optimization (6), (9)-(11) and (12) are in order. 1) In case we relax the assumption of LT-CSI at the transmitter, i.e., we allow the matrices $\mathbf{C}$ and $\mathbf{F}$ to depend on the instantaneous CSI, we get $\mathbf{K} = \mathbf{I}$ and $\mathbf{F}$ and $\mathbf{G}$ coincide with the results derived in [4]. In other words, if both sides of the link have access to the channel matrix $\mathbf{H}$, the setting that minimizes the MSE (4) results in linear precoding and decoding. 2) Setting $\mathbf{F} = \mathbf{I}$ and $\mathbf{E} = \mathbf{I}$ leads to the THP followed by a MMSE residual linear equalizer derived in [3].

4. SIMULATION RESULTS

The performance of the general precoder/decoder structure of fig. 1 is first evaluated in terms of uncoded SER for a 16-QAM constellation ($M = 16$), $N = 4$ antennas and the diversity model. We compare the performance of the general setting with $\mathbf{K} = \mathbf{C}$ or $\mathbf{K} = \mathbf{E}$, referred to as NP-LD (non-linear precoding, linear decoding) and LP-ND (linear precoding, non-linear decoding) respectively, with the following special cases: 1) $\mathbf{B} = \mathbf{0}$, $\mathbf{F} = \mathbf{I}$: MMSE-V-BLAST receiver (or MMSE-DFE); 2) $\mathbf{B} = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$: linear precoding-linear decoding (LP-LD); 3) $\mathbf{F} = \mathbf{I}$ and $\mathbf{D} = \mathbf{0}$: THP with MMSE residual equalization [3]. We further limit the analysis to the spatially white noise case, i.e., $\mathbf{R}_n = \sigma_n^2\mathbf{I}$. The signal to noise ratio is defined as $SNR = \sigma_x^2/\sigma_n^2$.

In fig. 2 the SER is plotted as a function of SNR for $\rho_T = 0.6$ (left) and $\rho_T = 0.9$ (right) ($\rho_R = 0.2$). Apart from the expected performance degradation due to the decreased spatial diversity, it can be seen that the benefits (if any) of precoding based on LT-CSI compared to the DFE

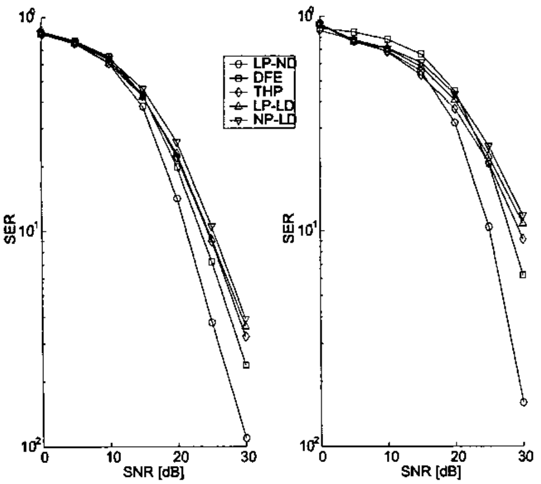

Fig. 2. SER vs. SNR for $\rho_T = 0.6$ (left) and $\rho_T = 0.9$ (right) ($\rho_R = 0.2$, $N = 4$)

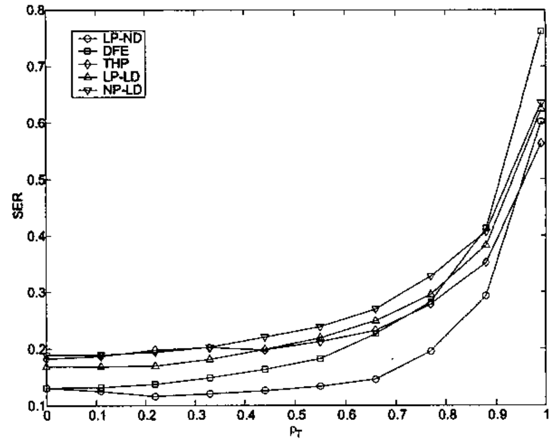

Fig. 3. Effect of spatial correlation at the transmitter side on the uncoded SER ($SNR = 20dB$, $N = 4$, $\rho_R = 0.2$).

receiver are more relevant for increasing values of ρ_T. This is intuitively clear since for $\rho_T = 0$ the LT-CSI $E[\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}] = \sqrt{N}/\sigma_n^2 \mathbf{R}_T = \sqrt{N}/\sigma_n^2 \mathbf{I}$ does not bring any side information that can be exploited by the transmitter to improve the performance of the link. In this case, it is $\mathbf{F} = \mathbf{I}$ and $\mathbf{B} = 0$ from (12) and (11) respectively. Furthermore, it can be concluded that the LP-ND gives the best performance in terms of uncoded SER. It is important to remark that in comparing the performance of the schemes of interest other considerations, apart from the SER, should be taken into account. For instance, it is well-known that non-linear precoding at the transmitter causes a relevant increase of the dynamic range at the input of the decision device that can limit its feasibility [9].

To have a clearer understanding of the role of the spatial correlation at the transmitter side on the performance of different schemes, fig. 3 plots the uncoded SER as a function of ρ_T for $SNR = 20dB$, $N = 4$ and $\rho_R = 0.2$. In accordance with the previous discussion, all precoding schemes outperform the DFE for ρ_T large enough. Moreover, the LP-ND structure shows the lowest SER except for very high values of ρ_T, where it is slightly outperformed by the THP scheme.

We now want to assess the effects of an imperfect I-CSI at the receiver. To this end, we assume that for the design of $\mathbf{G}$ (6) and $\mathbf{D}$ (9) only a noisy version of the channel matrix is available. A conventional LS estimate of the channel is carried out at the receiver. The training sequences from all transmit antennas are assumed to be mutually orthogonal. Therefore, the estimate is unbiased and the estimated channel gains (conditioned on $\mathbf{H}$) are i.i.d. variables with variance $1/SNR \cdot 1/N_t$, where N_t is the length of the training sequence [10]. The SER is plotted as a function of SNR in fig. 4 for $\rho_T = 0.6$ (left) and $\rho_T = 0.9$ (right) respectively ($\rho_R = 0.2$, $N_t = N = 4$). The same considerations discussed for perfect I-CSI apply also to the case in which the channel estimation error is taken into account, except for the performance degradation due to the imperfect I-CSI. In particular, LP-ND still gives the best performance uniformly with respect to the SNR.

In summary, precoding with long term channel state information appears to be advantageous in dense multipath channels with relatively large correlation at the transmitter ($\rho_T \geq 0.2$). Moreover, the experimental results show that the most promising scheme is LP-ND, also considering the practical limitations of non-linear precoding [9]. The preferred scheme essentially adds a linear precoder to a modified BLAST receiver, where the feedforward filter $\mathbf{G}$ and the feedback filter $\mathbf{D}$ are designed by taking into account the precoder $\mathbf{F}$ according to (6) and (9).

5. CONCLUSION

A transceiver structure for frequency non-selective MIMO channels that includes linear/non-linear precoding/decoding has been studied under the assumption that the state information available at the transmitter is limited to the second order statistics of channel and noise. Simulations have shown that relevant benefits can be obtained by exploiting the long term channel state information at the transmitter in dense multipath channels with relatively large correlation at the transmitter side. Moreover, the preferred scheme essentially adds a linear precoder to a modified BLAST receiver.

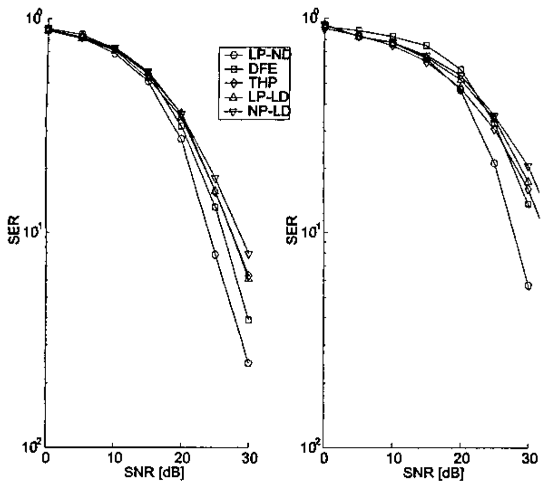

Fig. 4. SER vs. SNR for $\rho_T = 0.6$ (left) and $\rho_T = 0.9$ (right) in case of imperfect I-CSI at the receiver ($\rho_R = 0.2$, $N = N_t = 4$)

6. REFERENCES

[1] G. D. Golden, C. J. Foschini, R. A. Valenzuela, P. W. Wolnianski, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics letters*, vol. 35, no. 1, pp. 14-16, Jan. 1999.

[2] M. T. Ivrlac, T. P. Kurpjuhn, C. Brunner, W. Utschick, "Efficient use of fading correlations in MIMO systems," VTC 2001, vol.4, pp. 2763-2767, 2001.

[3] R. F. H. Fischer, C. Windpassinger, A. Lampe, J. B. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," International Zurich Seminar on Broadband Communications, pp. 7_1-7_6, 2002.

[4] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1987-2006, July 1999.

[5] A. Scaglione, G. B. Giannakis, S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: unification and optimal design," *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp. 1051-1064, May 2002.

[6] R. F. H. Fischer, C. Windpassinger, A. Lampe, J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," Proc. of 4 ITG Conference on Source and Channel Coding, pp. 139-147, January 2002.

[7] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774-780, Aug. 1972.

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

[9] R. F. H. Fischer, R. Tzschoppe, J. B. Huber, "Signal shaping for peak-power and dynamics reduction in transmission schemes employing precoding," *IEEE Trans. Comm.*, vol. 50, no. 5, pp. 736-741, May 2002.

[10] B. Hassibi, B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," Technical Memorandum, Bell Laboratories, Lucent Technologies, April 2000.

[11] R. A. Horn, C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1996.

7. APPENDIX

The inequality $E[\mathrm{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}] \geq \mathrm{tr}\{\mathbf{K}E[\mathbf{Q}]^{-1}\mathbf{K}^H\} = \mathrm{tr}\{\mathbf{K}\bar{\mathbf{Q}}^{-1}\mathbf{K}^H\}$ directly follows from the Jensen's inequality once the function $\mathrm{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}$ is proved to be convex in the positive definite matrix $\mathbf{Q}$. To show the convexity of the function of interest, it is sufficient to demonstrate that $\mathbf{k}^H(\lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2)^{-1}\mathbf{k} \leq \lambda\mathbf{k}^H\mathbf{Q}_1^{-1}\mathbf{k} + (1-\lambda)\mathbf{k}^H\mathbf{Q}_2^{-1}\mathbf{k}$, where $0 \leq \lambda \leq 1$ and $\mathbf{k}$ is any vector. The aforementioned condition can be stated as $(\lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2)^{-1} \leq \lambda\mathbf{Q}_1^{-1} + (1-\lambda)\mathbf{Q}_2^{-1}$ or equivalently as (th. 7.7.3 of [11])

$$\varrho((\lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2)(\lambda\mathbf{Q}_1^{-1} + (1-\lambda)\mathbf{Q}_2^{-1})) \geq 1, \tag{14}$$

where $\varrho(\cdot)$ denotes the spectral radius. After simple manipulations, we obtain

$$\varrho(\mathbf{Q}_1\mathbf{Q}_2^{-1} + \mathbf{Q}_2\mathbf{Q}_1^{-1}) \geq 2. \tag{15}$$

Since $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are hermitian matrices (in particular, they are positive definite), we can find a nonsingular matrix $\mathbf{Z}$ such that $\mathbf{Q}_1 = \mathbf{Z}\mathbf{Z}^H$ and $\mathbf{Q}_2 = \mathbf{Z}\Theta\mathbf{Z}^H$, where $\Theta = \mathrm{diag}([\Theta_1 \cdots \Theta_N])$ is diagonal and Θ_i is real and positive (th. 7.6.3 and 7.6.5 of [11]). It follows that

$$\mathbf{Q}_1\mathbf{Q}_2^{-1} + \mathbf{Q}_2\mathbf{Q}_1^{-1} = \mathbf{Z}(\Theta + \Theta^{-1})\mathbf{Z}^{-1}, \tag{16}$$

which implies that (15) becomes

$$\Theta_i + 1/\Theta_i \geq 2 \quad \forall i = 1, 2, .., N, \tag{17}$$

that is clearly satisfied $\forall\Theta_i > 0$.