MIMO Channel Estimation with Dimension Reduction

Matthias Stege, Peter Zillmann, Gerhard Fettweis

Mobile Communications Systems Chair, Dresden University of Technology , Germany

{zillmann, stege}@ifn.et.tu-dresden.de

Abstract

The number of channel estimates that have to be estimated in Multiple-Input Multiple-Output (MIMO) system is in general much larger, than in a single antenna communication scheme. This leads to lower signal to noise ratios (SNR) of the channel estimates if a constant pilot power independent from the number of transmit antennas is assumed. The use of long-term spatial channel characteristics can improve channel estimation for MIMO wireless systems. Separating the signal and the noise subspace followed by a dimension reduction can significantly reduce additive noise on channel estimates. This leads to improved channel estimation, especially for MIMO systems with high numbers of antennas, and to lower pilot power requirements.[1]

Keywords

MIMO, Channel Estimation, Dimension Reduction, MIMO-Systems, BLAST.

1. Introduction

Wireless systems with multiple transmit and receive antennas are an important part of discussions on future wireless communication systems, in 3GPP standards and for use in WLAN systems. In a rich scattering environment, these systems offer large capacity gains, as shown by information theory [1]. The MIMO channel estimation system has to provide the receiver with accurate information of all subchannels to ensure a reliable suppression of self-interference within a successive interference cancellation (SIC) receiver.

The need for a sophisticated channel estimation for multiple-input multiple-output systems is determined by the fact, that the more transmit antennas are applied in a communication system, the more channel coefficients need to be estimated. Considering a constant pilot power which is independent from the number of TX-antennas, the power of pilot signal emitted from a specific antenna reduces with increasing number of antennas. That means that less pilot power is available for the estimation of a channel coefficient leading to a worse signal to noise ratio for the channel estimates. That highlights a general problem of MIMO-techniques. They are designed to exploit more capacity within a communication system. The MIMO capacity gain can only be reached, if reliable estimates of the channel coefficients are available at the receiver.

[1]This work was financed by the German Federal Ministry of Education, Science, Research and Technology (BMBF) and Ericsson as part of the COMCAR-project. The authors thank U. Wachsmann and J. Thielecke for the fruitful discussions.

Noise reduction for the channel estimates can be achieved by temporal averaging. This requires the channel to be constant over the averaging time, otherwise a bias would be introduced. In case of a communications system with multiple transmit antennas orthogonal pilot sequences are transmitted from the different antennas. Thus, a correlation over $M^{\mathrm{T}x}$ symbols of these sequences is necessary to resolve the interference between them. The resulting signal to noise ratio of the channel estimates after this correlation is the same as for a single-input single-output system, but remark, that only one estimate is available for $M^{\mathrm{T}x}$ transmitted pilot symbols [2]. This means, that there are less pilot symbols available for averaging in a given coherence interval. This reduces the potential gain, that could be achieved by temporal averaging.

In [3] and [4] it has been shown, that knowledge of long-term spatial channel characteristics can improve channel estimation. This approach can be extended to the problem of MIMO-channel estimation. In general, a noise reduction on the channel estimates can be achieved by separating the signal- from the noise-subspace. The signal-dependent Karhunen-Loève-Transform (KLT) is applied to determine the signal subspace, and a dimension reduction algorithm is used to reduce additive noise power corrupting the channel estimates. The improvement for the estimates and for a V-BLAST MIMO system are evaluated.

2. Signal Model

Flat fading channel characteristics are assumed throughout the paper. Furthermore, a discrete-time representations of the signals are used. For a system with M^{Rx} receive antennas, the receive signal vector $\mathbf{y}(k)$ has M^{Rx} components and is defined as

$$\mathbf{y}(k) = \mathbf{r}(k) + \mathbf{n}(k) = \mathbf{H}(k)\mathbf{s}(k) + \mathbf{n}(k), \tag{1}$$

where the white Gaussian noise vector process $\mathbf{n}(k)$ accounts for thermal noise and interferences:

$$\mathbf{n}(k) = [n_1(k) \ \dots \ n_{M^{Rx}}(k)]^T. \tag{2}$$

The components of $\mathbf{n}(k)$ are zero mean, complex random variables with variance σ_n^2. Thus, this random process is sometimes also called temporally and spatially white.

The transmitted signal vector $\mathbf{s}(k)$ of M^{Tx} complex symbols, where M^{Tx} is the number of transmit antennas, may be represented by

$$\mathbf{s}(k) = [s_1(k) \ \dots \ s_{M^{Tx}}(k)]^T. \tag{3}$$

Consequently, $\mathbf{H}(k)$ is a $M^{Rx} \times M^{Tx}$ matrix of complex channel coefficients $h_{i,j}(k)$ with $1 \leq i \leq M^{Tx}$ and $1 \leq j \leq M^{Rx}$,

0-7803-7442-8/02/$17.00 © 2002 IEEE

Parameter	Microcell	Picocell
distance BS → MS [m]	300	10
cluster diameter [m]	10	30
average AOA [deg]	30	0
AOA angular spread (RMS) [deg]	≈ 1	≈ 110
AOA angular spread (max-min) [deg]	≈ 9	≈ 350
average AOD [deg]	53	90
AOD angular spread (RMS) [deg]	≈ 2	≈ 85
AOD angular spread (max-min) [deg]	≈ 14	≈ 350

Table 1: Parameters for MIMO scenarios

reading

$$\mathbf{H}(k) = \begin{pmatrix} h_{1,1}(k) & \dots & h_{M^{Tx},1}(k) \\ \vdots & \ddots & \vdots \\ h_{1,M^{Rx}}(k) & \dots & h_{M^{Tx},M^{Rx}}(k) \end{pmatrix}. \quad (4)$$

The channel is assumed to be passive and normalized, resulting in

$$E\{ \mid h_{i,j}(k) \mid^2 \} = 1. \quad (5)$$

The actual transmission coefficients $h_{i,j}(k)$ result from a superposition of different wavefronts reflected by a set of scatterers within the scenario. Two different scenarios are investigated in this paper. The Picocell scenario demonstrates a typical MIMO scenario with low spatial correlation. Transmit and receive antenna arrays lie within one set of narrow scatterers. Angular spreads at transmitter and receiver are large. The Microcell scenario serves as a comparison. The scattering cluster is small and and some distance away from the antennas. Therefore, angular spreads at both transmitter and receiver are also small. This leads to highly correlated channel coefficients. The Microcell scenario can be considered as a typical beamforming scenario due to that high correlation. Table 1 summarizes important parameters of the scenarios. For a detailed description of this MIMO channel model, the reader is referred to [5, 6].

Finally, the problem of channel estimation shall be described. The output of an estimator operating on orthogonal pilot sequences from the transmit antennas may be expressed by

$$\hat{\mathbf{H}}(k) = \mathbf{H}(k) + \mathbf{N}(k), \quad (6)$$

where the estimation error $\mathbf{N}(k)$ results from noise and interferences and consists of zero mean, complex Gaussian random variables with variance σ_n^2. This is due to the fact that M^{Tx} pilot symbols have to be correlated to resolve interference between them.

3. KLT and Dimension Reduction

Signal transforms are frequently used in the field of digital signal processing. Transformed signals can yield better insight into certain properties. Well known orthonormal transforms are the Discrete Cosine Transform (DCT) and the Discrete Fourier Transform (DFT) [7]. Both of these transforms have a *fixed* set of orthonormal basis vectors, which means, that the basis vectors are independent of the actual signal to be processed.

In contrast *signal-dependent* transforms, such as the Karhunen-Loève Transform (KLT) are known. It's basis vectors are matched to the statistics of the input signal to deliver a set of uncorrelated coefficients in the transform domain. The KLT is the optimum least squares decorrelating transform and has been successfully applied to problems in the fields of speech processing and pattern recognition [8]. It can be shown [7], that it concentrates a maximum amount of signal energy into a small number of orthogonal Eigenspaces. But when only a few signal dimensions contain a significant amount of signal energy, these dimensions are sufficient for an approximate reconstruction of the signal. Additive uncorrelated noise, in contrast, results in energy of equal size σ_n^2 in all dimensions of the Eigenspace. Therefore, in the KLT domain, it is possible to separate the signal space from the evenly distributed noise by discarding coefficients, that contain mainly noise power.

For the problem of channel estimation, the input vector process to be investigated is a vector form of $\hat{\mathbf{H}}(k)$,

$$\begin{aligned} \hat{\mathbf{h}}(k) &= \operatorname{vec}(\hat{\mathbf{H}}(k)) = \operatorname{vec}(\mathbf{H}(k)) + \operatorname{vec}(\mathbf{N}(k)), \quad (7) \\ &= \mathbf{h}(k) + \hat{\mathbf{n}}(k). \quad (8) \end{aligned}$$

The operation $\operatorname{vec}(\mathbf{H}(k))$ defines a vector of length $M^{\mathrm{Tx}} \cdot M^{\mathrm{Rx}}$ obtained by stacking the columns of $\mathbf{H}(k)$. KLT basis vectors are obtained from the input signal's estimated covariance matrix

$$\mathbf{R}_{\hat{\mathbf{h}}} = \mathrm{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} \quad (9)$$

via Eigenvalue Decomposition (EVD) [9]. An EVD of the hermitian matrix $\mathbf{R}_{\hat{\mathbf{h}}}$ can be written as

$$\mathbf{R}_{\hat{\mathbf{h}}} = \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^H, \quad (10)$$

where $\mathbf{U}$ is the eigenvector matrix of $\mathbf{R}_{\hat{\mathbf{h}}}$ and $\mathbf{\Lambda}^2$ is a diagonal matrix containing the eigenvalues of $\mathbf{R}_{\hat{\mathbf{h}}}$. Essentially, the KLT changes the basis vectors of the random vector process to the eigenvectors in $\mathbf{U}$ and an eigenvalue represents the projection of signal- and noise energy onto the corresponding vector. With knowledge on the structure of the process $\hat{\mathbf{h}}$, $\mathbf{\Lambda}^2$ can be written as

$$\begin{aligned} \mathbf{\Lambda}^2 &= diag\left[\lambda_1^2, \dots, \lambda_N^2\right] \\ &= diag\left[\sigma_1^2 + \sigma_n^2, \dots, \sigma_N^2 + \sigma_n^2\right], \quad (11) \end{aligned}$$

where σ_i^2 denotes the signal energy and σ_n^2 the noise variance. N is the number of eigenvalues and thus the dimension of $\hat{\mathbf{h}}(k)$. Now, when the signal part of a certain eigenvalue is smaller than the noise part, it is beneficial to discard this component for reconstruction, because the overall mean squared error (MSE) of the estimation will be smaller then. Dimension reduction is done by using the reduced transform matrix $\mathbf{U}_L$, consisting of the first $L \leq N$ eigenvectors. This can be viewed as a kind of filtering by a filter with an ideal lowpass characteristic. The bidirectional transform and dimension reduction can be done in one step, since

$$\tilde{\mathbf{h}} = \mathbf{U}_L\mathbf{U}_L^H\hat{\mathbf{h}} = \mathbf{\Phi}_L\hat{\mathbf{h}}, \quad (12)$$

where $\tilde{\mathbf{h}}$ denotes the filtered estimate of $\hat{\mathbf{h}}$. Projector matrix $\mathbf{\Phi}_L$ is idempotent and therefore holds [10]:

$$\mathbf{\Phi}_L = \mathbf{\Phi}_L^H = \mathbf{\Phi}_L^2. \quad (13)$$

Such a transform results in two types of errors: First, the signal energy within the eigenvalues $\lambda_{L+1}^2 \ldots \lambda_N^2$ is discarded, leading to a *Bias*. The remaining noise within the reduced signal space still corrupts the signal. The optimum MMSE criterion for selecting the number of dimensions L_{opt} is therefore known as *Bias-Variance-Tradeoff* [11]. It minimizes the sum of both errors after the transform and has the form

$$\begin{aligned} L_{opt} &= \arg\min_L \left[\sum_{i=L+1}^{N} (\lambda_i^2 - \sigma_n^2) + L\sigma_n^2 \right] \\ &= \arg\min_L \left[\sum_{i=L+1}^{N} \lambda_i^2 + (2L - N)\sigma_n^2 \right], \quad (14) \end{aligned}$$

which follows from the eigenvalue structure shown in (11). Remark, that the noisy eigenvalues λ_i^2 must be known to solve (14). They are obtained from the EVD of $\mathbf{R}_{\hat{\mathbf{h}}}$. Additionally knowledge on the noise variance σ_n^2 is required. This can be estimated from known pilots signals.

This section is concluded by a closer observation of the channel coefficients' covariance matrix. Note that $\mathbf{R}_{\hat{\mathbf{h}}}$ is the long-term *spatial* covariance matrix, averaged over the time duration for which the spatial characteristics of the scenario remain approximately constant. This time is usually much longer than the coherence time of the channels, describing the short-term channel variations. In practice, $\mathbf{R}_{\hat{\mathbf{h}}}$ could simply be estimated via block-averaging over an averaging length B:

$$\mathbf{R}_{\hat{\mathbf{h}}} = \frac{1}{B-1} \sum_{k=0}^{B-1} \hat{\mathbf{h}}(k) \big(\hat{\mathbf{h}}(k)\big)^H. \quad (15)$$

It is important to choose B sufficiently large to get an accurate estimate. However, large block-sizes contribute to latency and processing power requirements. Other, more sophisticated methods are possible but beyond the scope of this work.

4. Properties of the Channel Estimates After the Transformation

After applying a KLT, there are still some error components on the channel estimate $\tilde{\mathbf{h}}$. The mean squared error (MSE) $\varepsilon_{\tilde{\mathbf{h}}}^2$ follows:

$$\begin{aligned} \varepsilon_{\tilde{\mathbf{h}}}^2 &= ||\mathbf{h}(k) - \tilde{\mathbf{h}}(k)||_F^2 \quad (16) \\ &= ||\mathbf{h}(k) - \mathbf{\Phi}_L \hat{\mathbf{h}}(k)||_F^2 \quad (17) \end{aligned}$$

($||\quad||_F^2$ defines the Frobenius norm [9].) Using (8) the MSE can be expresses in terms of a *bias* $\mathbf{b}(k)$ and an *additional noise* $\tilde{\mathbf{n}}(k)$:

$$\begin{aligned} \varepsilon_{\tilde{\mathbf{h}}}^2 &= ||(\mathbf{I} - \mathbf{\Phi}_L)\mathbf{h}(k) + \mathbf{\Phi}_L \hat{\mathbf{n}}(k)||_F^2 \quad (18) \\ &= ||\mathbf{b}(k) + \tilde{\mathbf{n}}(k)||_F^2 \quad (19) \end{aligned}$$

The properties of the remaining noise are investigated first. Since $L \leq N$ uncorrelated noise processes are transformed back onto N processes, it follows that these noise processes must be mutually correlated in general. Let the remaining noise vector can be denoted by

$$\tilde{\mathbf{n}}(k) = \mathbf{\Phi}_L \hat{\mathbf{n}}(k), \quad (20)$$

where $\hat{\mathbf{n}}(k) = \mathrm{vec}(\mathbf{N}(k))$.

With (13) the noise covariance matrix can then be written as

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{n}}} &= E\{\mathbf{\Phi}_L \hat{\mathbf{n}}(k) \big(\mathbf{\Phi}_L \hat{\mathbf{n}}(k)\big)^H\} \\ &= \sigma_n^2 \mathbf{\Phi}_L \mathbf{I}_N \mathbf{\Phi}_L^H \\ &= \sigma_n^2 \mathbf{\Phi}_L. \quad (21) \end{aligned}$$

Remaining noise processes are in general spatially mutually correlated, but they are still temporally white, because every process is a weighted sum of white input processes.

Next, the bias $\mathbf{b}(k)$ shall be analyzed. It is an approximation error resulting from discarded signal energy and can be expressed by

$$\mathbf{b}(k) = \big(\mathbf{I} - \mathbf{\Phi}_L\big)\mathbf{h}(k), \quad (22)$$

and this notation reveals that $\mathbf{b}(k)$ cannot considered to be temporally white. Being a weighted sum of the channel coefficients itself, it has the same spectral characteristics as the coefficients. The correlation matrix of the subchannels after the transformation can be obtained in a similar way as (21), resulting in

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \mathbf{\Phi}_L \mathbf{R}_{\mathbf{h}} \mathbf{\Phi}_L^H, \quad (23)$$

with $\mathbf{R}_{\mathbf{h}}$ being the covariance matrix of the channel coefficients vector $\mathbf{h}(k)$ (without noise). This again leads to some spatially correlated error component. The temporal characteristics are the same as the fading characteristic of the channel coefficients.

5. Channel Estimation Improvement

With the framework developed so far it is possible to investigate the ability of the system to improve channel estimation. Eigenvalue distributions of typical scenarios were simulated. Figure 1 shows characteristic Eigenvalue distributions for Picocell and Microcell scenarios with $M^{Tx} = M^{Rx} = 4$ and 8. Eigenvalues are normalized to a sum of 1, modeling fixed transmit power for different numbers of transmit antennas. Only nonzero eigenvalues are shown. Since 30 scatterers are considered within the cluster, at most 30 eigenvalues greater than zero can occur. The Picocell case is a typical MIMO scenario for systems aiming at exploiting the higher capacity of a rich scattering environment. Due to the small correlation between subchannels, a higher number of eigenvalues contains significant amounts of signal energy. Compare this to the Microcell scenario. A large amount of signal energy is contained in the first eigenvalue, and the curve is steeply decreasing. This is a beneficial situation for dimension reduction. However, the Microcell scenario is not well suited for MIMO systems like V-BLAST, because the low MIMO-capacity in these channel scenarios [6]. It is a typical beamforming scenario and shall serves only for comparison here. The other extreme case, where the channel coefficients were completely uncorrelated, all eigenvalues would have the same value. No improvement is possible then with the KLT.

These results were now used to evaluate the possible noise reduction of the proposed channel estimator. Criterion (14) was computed for different values of σ_n^2, and the resulting overall MSE compared to the input MSE of $N\sigma_n^2$ delivers the effective noise reduction . Figure 2 summarizes the results. It shows the noise reduction versus the pilot SNR before the transformation. In general, Microcell scenarios offer more possibilities for improvement because of the signal energy being concentrated in very few dimensions. This leads to a small value of L_{opt}, and

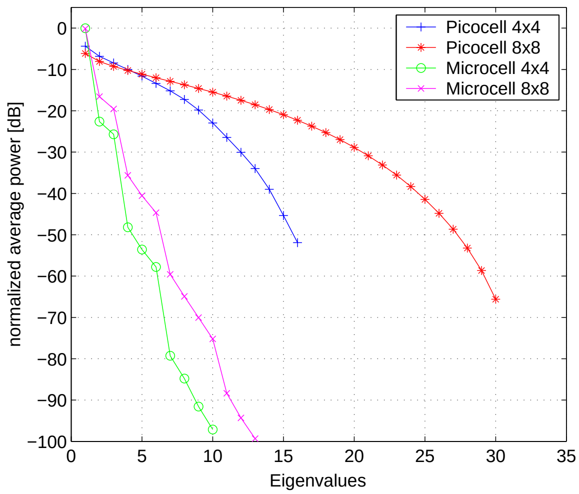

Figure 1: Average nonzero eigenvalues for certain scenarios

most of the noise power can be discarded. Picocell shows less improvement, especially for small numbers of antennas, but in the case of $M^{Tx} = M^{Rx} = 8$, significant improvements are possible. The algorithm becomes more efficient, as higher numbers of antennas are available. Another important property of the system is, that higher gains are obtained for low input SNR. That is extremely useful, since it helps reducing the pilot power to a lowest possible level to avoid interference by to other parallel transmitted signals.

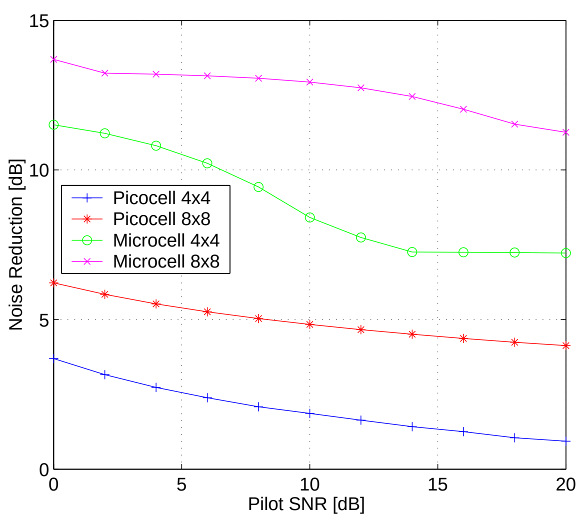

Figure 2: Channel estimation improvement for certain scenarios

6. MIMO Simulation Results

Figure 3 shows the principle of the MIMO-channel estimation with dimension reduction. Simulations were conducted to evaluate the influence of real channel estimation on the performance of a V-BLAST MIMO system. V-BLAST is a, layered scheme aiming at MIMO capacity gains. Independent data streams are transmitted over the different antennas. The receiver uses *Serial Interference Canceling (SIC)* and detects the most reliable layer first. Remodulated, already detected layers are subtracted from the input signal to improve detection of the remaining layers. For a detailed description of the system, the reader is referred to [12].

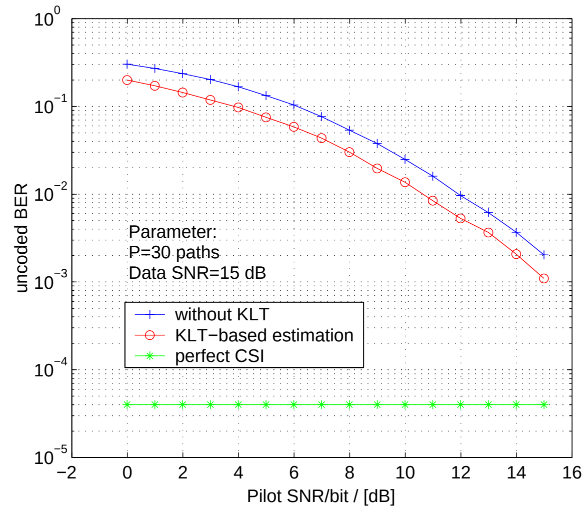

Figure 4: Performance of a V-BLAST, Picocell, 4×4 system vs. pilot SNR

A pilot structure similar to the 3GPP-standard was assumed. That means, that a continuous pilot signal is available over a pilot channel (CPICH) along with the data channel. The overall pilot power is constant independent from the number transmit antennas. The data layer SNR is held constant at a value of 15 dB. Bit error rates are shown versus different pilot SNR for ideal channel knowledge, conventional channel estimation and KLT-based estimation. Thus, possible pilot power savings can be seen directly. The V-BLAST system was only simulated for Picocell scenarios, since it requires a rich scattering environment in order to achieve capacity gains.

Figure 4 displays the results for the 4×4 case. About 1.5 - 2 dB pilot power reduction can be achieved through dimension reduction. Equivalently, when the same pilot level is used as for conventional systems, performance improves significantly. Figure 5 reveals the potential of the proposed system: In an 8×8 system, 4.5 dB improvement are gained at a BER of 10^{-2}, and still more than 4 dB at a BER of 10^{-3}. As already shown in Figure 2, more improvement is reached for high numbers of antennas. This is a very important property, because with conventional systems, channel estimation gets worse for increasing number of transmit antennas. The reason for that is the smaller amount of pilot power for each antenna. The pilot SNR therefore gets worse, and conventional channel estimation systems offer a degraded performance. The dimension reduction system reverses this behavior and ensures more reliable channel estimation even for large antenna arrays at the transmitter.

7. Summary

A novel approach for estimating channel coefficients for MIMO wireless systems has been presented in this paper. The signal-

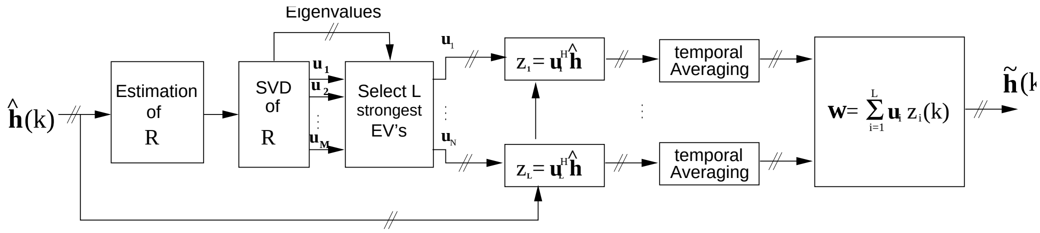

Figure 3: Overall Principle of the KLT-based MIMO Channel Estimation

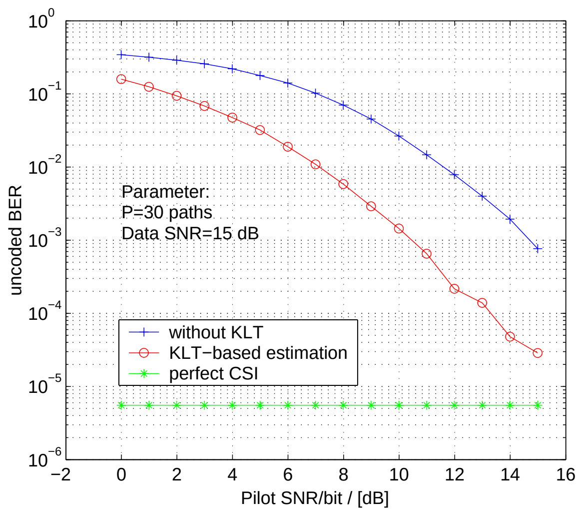

Figure 5: Performance of a V-BLAST, Picocell, 8×8 system vs. pilot SNR

dependent decorrelating KLT enables the system to reduce the number of coefficients in the transform domain used for estimation, provided the noise is uncorrelated and there is some subchannel correlation in the MIMO scenario.

Simulation results in section 6 showed the performance of a V-BLAST reference system. Pilot power can be significantly reduced by the concept presented here. The system is beneficial especially for high numbers of antennas and is therefore able to overcome the effect of splitting pilot power between the antennas. It has to be mentioned, that only an uncoded V-BLAST system was discussed here, similar to the approach in [13]. For a complete view of the the effect of channel estimation errors coding has to be considered as well.

8. References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications 6*, 1998.

[2] M. Stege, M. Bronzel, and G. Fettweis, "On the performance of space-time blockcodes," in *IEEE Proc. Vehicular Technology Conference*, Rhodes, Greece, 6.-9. May 2001, vol. 3, pp. 2282–2286.

[3] Matthias Stege, Tom Ruprich, Marcus Bronzel, and Gerhard P. Fettweis, "Channel estimation using long-term spatial channel characteristics," in *Proceedings of WPMC*, 2001.

[4] Monica Nicoli and Umberto Spagnolini, "Reduced rank channel estimation and rank order selection for CDMA systems," in *Proceedings ICC*, 2001.

[5] M. Stege, J. Jelitto, M. Bronzel, and G. Fettweis, "A multiple input - multiple output channel model for simulation of Tx- and Rx-diversity wireless systems," in *IEEE Proc. Vehicular Technology Conference*, 2000.

[6] M. Stege, M. Bronzel, and G. Fettweis, "MIMO-capacities for realistic COST 259 scenarios," in *Proc. of the International Zurich Seminar*, 2002.

[7] Richard A. Haddad and Thomas W. Parsons, *Digital Signal Processing - Theory, Applications and Hardware*, Computer Science Press, 1991.

[8] G. Ruske, *Automatische Spracherkennung*, R. Oldenbourg Verlag München Wien, 1988.

[9] G. H. Golub and C. F. van Loan, *Matrix Computations*, The John Hopkins University Press, 1996.

[10] Louis L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1990.

[11] J. Jelitto, M. Bronzel, and G. Fettweis, "From Smart Antennas to Reduced Dimension Space-Time Processing," *FREQUENZ - Journal of Telecommunications*, vol. 55, no. 5-6, pp. 165–70, May/June 2001.

[12] 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 *Proceedings of ISSSE*, 1998.

[13] S. Bäro, G. Bauch, A. Pavlic, and A. Semmler, "Improving BLAST performance using space-time block codes and turbo decoding," in *Globecom 2000*, 2000, vol. 2, pp. 1067–1071.