14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

DESIGN OF ROBUST LINEAR DISPERSION CODES BASED ON IMPERFECT CSI FOR ML RECEIVERS

Svante Bergman and Björn Ottersten

Signal Processing, School of Electrical Engineering,
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, Sweden
phone: + (46) 8 790 84 31, fax: + (46) 8 790 72 60, email: svante.bergman@ee.kth.se
web: www.ee.kth.se

ABSTRACT

This paper concerns the design of codes for multiple-input multiple-output communication systems. The transmission scheme utilizes imperfect channel state information (CSI) in the design, assuming that maximum-likelihood detection is employed at the receiver. It is argued that channel diagonalizing codes are not robust to imperfections in the CSI. A robust non-diagonalizing code with good minimum distance separation between received codewords is proposed. The design is very suitable for systems operating at high data rates since the complexity scales nicely with the number of antennas. Numerical results show that the proposed code outperforms a state-of-the-art diagonalizing precoder.

1. INTRODUCTION

From information theory it is known that the maximum achievable data rate of multiple-input multiple-output (MIMO) communication systems can be increased if the transmitter has access to channel state information (CSI) [1]. Transmitter-side channel state information (TX-CSI) may be acquired from the receiver via feedback, or estimated from the reverse link for time division duplex systems. By exploiting TX-CSI, the transmitted data can be adapted to the spatial characteristics of the channel for improved system performance.

Optimization of the mutual information using perfect TX-CSI results in diagonalization of the channel and independent transmission over separate single-input single-output (SISO) spatial channels. The power allocated to each channel is given by the spatial water-filling solution [1], and the codebooks are Gaussian. In most practical applications symbols are however limited to finite sized constellation sets, and capacity-optimal precoding is not necessarily optimal in terms of maximum uncoded data rate or minimum uncoded bit error rate (BER). Some recent work on these systems employing linear transmitters and receivers has shown that diagonalization is not always optimal in terms of minimum BER [2, 3, 4].

Due to system limitations such as delay or capacity constraints in the feedback link, it is reasonable to assume that the quality of the TX-CSI is worse than the CSI available at the receiver. In the literature the TX-CSI imperfections are typically modeled using a Bayesian channel model [5, 6, 7]. Another approach is to model the TX-CSI as one out of a finite number of quantized channel states [8], which perhaps is more realistic than the Bayesian model if the capacity of the feedback link is the bottleneck. When the TX-CSI is imperfect, it is no longer possible to completely separate the spatial channels and crosstalk is inevitable. This makes the problem of minimizing the error rate very hard and the topic has recently received much attention. In [5], the mean squared error (MSE) was minimized assuming a linear minimum MSE (MMSE) receiver structure. For maximum-likelihood (ML) decoding receivers, [6, 9] designed weighted orthogonal space-time block codes (OSTBC) based on TX-CSI. OSTBC have been proven very efficient in terms of reducing the block error rate at low data rates. In practice however, the error rate can usually be allowed to be quite high due to efficient outer codes and packet-layer retransmissions. This motivates the use of linear dispersion codes (LDC, see [10]) where multiple data streams are transmitted in parallel, allowing high data rates at the cost of reduced diversity gain.

In our earlier work [11, 12], the focus was on bit and power loading (BPL) algorithms assuming the optimal (and nonlinear) ML detector is employed at the receiver. The spatial channels were decorrelated from the transmitter side (essentially diagonalizing the channel), and the bit and power of the data streams were adapted to minimize the union bound on the symbol error rate. This approach works well for systems with relatively few data streams and low data rates, but for large sized MIMO systems the complexity of optimizing the power for all bit load candidates becomes unfeasible. Furthermore, as will be argued below, diagonalization of the channel may cause problems when imperfections in the channel estimate are introduced. Fortunately blind transmission (i.e. transmission disregarding the TX-CSI) has a performance surprisingly close to BPL at high data rates, which along with the results in [3, 4, 7], has inspired us to search for non-diagonalizing codes.

In this paper, we identify a robustness-complication that affects diagonalizing precoders when the TX-CSI is perturbed. A class of precoders that is less sensitive to imperfections in the channel estimate is proposed. Within this class, it is shown that a precoder based on the discrete Fourier transform (DFT) maximizes a certain upper bound on the minimum distance. A lower bound on the minimum distance is also derived and is utilized to show that the DFT based precoder is the optimal robust precoder for certain channels and outperforms the only robust diagonalizing precoder for all channels. Numerical results show that for systems operating at high data rates the proposed LDC outperforms the optimized BPL from [11, 12]. The complexity of the design procedure is very low, which is clearly an advantage when the number of data streams is large.

The paper is organized as follows. In Section 2, the system model and channel state information model are described. Section 3 specifies the union bound of the ML-detection block error rate. Robustness against imperfections in the TX-CSI is discussed in Section 4, a robust LDC is proposed and motivated analytically. In Section 5, numerical examples show that the proposed codes outperform a state-of-the-art diagonalizing code. Conclusions are drawn in Section 6.

The set of complex valued N by M matrices is denoted

$\mathbb{C}^{N\times M}$, and similarly for real valued matrices $\mathbb{R}^{N\times M}$. The set of integers is denoted $\mathbb{Z}$. The Frobenius norm is denoted $\|\cdot\|_F$. Expectation with respect to $\mathbf{H}$ is denoted $\mathrm{E}_{\mathbf{H}}$, while E is the expectation with respect to all random variables. The vectorization operator on matrices is denoted $\mathrm{vec}(\cdot)$, and $\otimes$ is the Kronecker product of matrices. The complex Gaussian distribution is denoted $CN(\cdot,\cdot)$.

2. SYSTEM MODEL

Consider a MIMO communication system with N transmitting and M receiving antennas over a flat fading channel. Similar to [10], it is assumed that the channel matrix is constant during at least L channel uses. Using complex notation, the transmitted signal block is $\mathbf{C}\in\mathbb{C}^{N\times L}$, the channel matrix is $\mathbf{H}\in\mathbb{C}^{M\times N}$, and the additive noise block is $\mathbf{V}\in\mathbb{C}^{M\times L}$. The noise is assumed to be zero mean, white, complex Gaussian distributed with unit variance. The received signal block $\mathbf{Y}\in\mathbb{C}^{M\times L}$ is then modeled as

$$\mathbf{Y}=\mathbf{HC}+\mathbf{V}. \tag{1}$$

Without loss of generality the system is normalized so that $\mathrm{E}\left[\|\mathbf{H}\|_F^2\right]=MN$. Given these normalizations, the signal to noise ratio (SNR) is defined as the average transmitted power $P=L^{-1}\,\mathrm{E}\left[\|\mathbf{C}\|_F^2\right]$.

2.1 Channel state information model

Partial or imperfect CSI implies some amount of uncertainty about the channel matrix, $\mathbf{H}$. Assume that $\mathbf{H}$ given the CSI, $\boldsymbol{\zeta}$, is distributed as

$$\mathrm{vec}(\mathbf{H}|\boldsymbol{\zeta})\sim CN\big(\mathrm{vec}(\hat{\mathbf{H}}),\boldsymbol{\Sigma}_{\mathbf{H}}\big), \tag{2}$$

$$\boldsymbol{\Sigma}_{\mathbf{H}}\equiv\mathrm{E}_{\mathbf{H}|\boldsymbol{\zeta}}\left[\mathrm{vec}(\mathbf{H}-\hat{\mathbf{H}})\big(\mathrm{vec}(\mathbf{H}-\hat{\mathbf{H}})\big)^{\mathrm{H}}\right], \tag{3}$$

where the CSI consists of $\hat{\mathbf{H}}$ and $\boldsymbol{\Sigma}_{\mathbf{H}}$. As an example, the transmitter may have an imperfect channel estimate $\hat{\mathbf{H}}$, that is corrupted by some estimation error, $\tilde{\mathbf{H}}=\mathbf{H}-\hat{\mathbf{H}}$, which is spatially correlated as $\mathrm{E}\left[\mathrm{vec}(\tilde{\mathbf{H}})\mathrm{vec}(\tilde{\mathbf{H}})^{\mathrm{H}}\right]=\boldsymbol{\Sigma}_{\mathbf{H}}$. For simplicity, throughout this work the estimation error will be assumed to be uncorrelated, i.e. $\boldsymbol{\Sigma}_{\mathbf{H}}$ proportional to $\mathbf{I}$.

3. PERFORMANCE BOUND

The expectation of the probability that a transmitted codeword $\mathbf{C}_i$ is detected as another codeword $\mathbf{C}_j$ when performing ML-detection is denoted the block error rate, P_e. In general, the block error rate is difficult to evaluate analytically and we chose to approximate it with the union bound

$$\bar{P}_e=\mathrm{E}_{\mathbf{H}}\Big[\sum_{\substack{i,j\\ j\neq i}}p(\mathbf{C}_i)\,\mathrm{Pr}\left(\mathbf{C}_i\rightarrow\mathbf{C}_j\right)\Big]\geq P_e.$$

The bound is fairly tight for high SNR. Assume codewords are uniformly drawn from a finite sized codebook $\mathcal{C}$ so that the probability mass function is constant, $p(\mathbf{C}_i)=|\mathcal{C}|^{-1}$, where $|\cdot|$ denotes the cardinality of a set. Introducing the Q-function for the pairwise error probability [13] yields

$$\bar{P}_e=|\mathcal{C}|\;\mathrm{E}_{\mathbf{H}}\left[\sum_{\substack{i,j\\ j\neq i}}p(\mathbf{C}_i)p(\mathbf{C}_j)Q\bigg(\frac{\|\mathbf{H}(\mathbf{C}_i-\mathbf{C}_j)\|_F}{\sqrt{2}}\bigg)\right]. \tag{4}$$

From (4) we see that the double sum essentially corresponds to expectation over independent identically distributed random variables $\mathbf{C}_i$ and $\mathbf{C}_j$ that are uniformly distributed in $\mathcal{C}$. The union bound depends on $\mathbf{C}_i$, $\mathbf{C}_j$, and $\mathbf{H}$ only through

$$Z(\mathbf{H})\equiv\frac{1}{2}\|\mathbf{H}(\mathbf{C}_i-\mathbf{C}_j)\|_F^2,$$

which will be referred as the separation profile. The case $i=j$ is not included in (4) so we define a function $g(x)$ to cancel out these terms

$$g(x)=\begin{cases}0, & x\leq 0\\ 1, & x>0\end{cases}.$$

Using the definition of the separation profile and $g(x)$, the union bound (4) is reformulated as

$$\bar{P}_e=|\mathcal{C}|\;\mathrm{E}_{\mathbf{H}}\left[\mathrm{E}_{Z|\mathbf{H}}\left[Q\big(\sqrt{Z(\mathbf{H})}\big)\,g\big(Z(\mathbf{H})\big)\right]\right]. \tag{5}$$

The performance bound is the base for the following discussion on LDC design.

4. LINEAR DISPSERSION CODE DESIGN

Assume an LDC with quadrature amplitude modulated (QAM) or pulse amplitude modulated (PAM) symbols stacked in a vector $\mathbf{x}$, where the codewords are written on the following form

$$\mathrm{vec}(\mathbf{C})=\mathbf{Fx}.$$

The dispersion matrix $\mathbf{F}$ is a data-independent precoding matrix, whereas the data is mapped on the vector $\mathbf{x}$, assumed to have discrete-random elements that are statistically independent, zero mean and have variance one. The number of data streams is typically an integer multiple, K, of the block length, L, hence $\mathbf{F}\in\mathbb{C}^{NL\times KL}$ (where $NL\geq KL$) defines how the symbols are spread in space (and time if $L>1$). Since the symbols are QAM or PAM, the difference between two codewords is a vector on the complex integer lattice as follows

$$\frac{1}{\sqrt{2}}\mathrm{vec}(\mathbf{C}_i-\mathbf{C}_j)=\frac{1}{\sqrt{2}}\mathbf{F}(\mathbf{x}_i-\mathbf{x}_j)=\mathbf{F}\boldsymbol{\Sigma}\,\mathbf{z},$$

where $\mathbf{z}\in\mathcal{I}_{KL}\equiv\{\mathbb{Z}^{KL}+j\mathbb{Z}^{KL}\}\setminus\{0^{KL}\}$ (the set of complex integer vectors excluding the origin), and the diagonal matrix $\boldsymbol{\Sigma}$ scales $\mathbf{z}$ to ensure unit variance of $\mathbf{x}$. From these assumptions the separation profile can be written on a quadratic form

$$Z(\mathbf{H})=\mathbf{z}^{\mathrm{H}}\,\boldsymbol{\Sigma}\,\mathbf{F}^{\mathrm{H}}\mathbf{Q}_{\mathbf{H}}\mathbf{F}\,\boldsymbol{\Sigma}\,\mathbf{z}$$

where $\mathbf{Q}_{\mathbf{H}}=\mathbf{I}\otimes\mathbf{H}^{\mathrm{H}}\mathbf{H}$. The goal is to find the optimal bit load (i.e. element constellations of $\mathbf{z}$, with its corresponding normalization matrix $\boldsymbol{\Sigma}$), and precoder $\mathbf{F}$ that maximizes (5). Unfortunately this problem is very difficult. Instead, a suboptimal scheme is proposed by first analyzing the problem assuming perfect TX-CSI resulting in a minimum distance performance measure. Then, introducing CSI imperfections, a constraint on the precoder is imposed to maintain robustness. Finally, a precoder that is robust, while exhibiting good minimum distance properties is proposed.

Due to the steep (logarithmic) decent of the $Q(\cdot)$ function, a good strategy is to maximize the separation profile (minimum distance), which if $\mathbf{Q}_{\mathbf{H}}$ is deterministic can be formulated as

$$\max_{\mathrm{Tr}\{\mathbf{F}^{\mathrm{H}}\mathbf{F}\}\leq PL}\;\min_{\mathbf{z}\in\mathcal{I}_{KL}}\;\mathbf{z}^{\mathrm{H}}\,\boldsymbol{\Sigma}\,\mathbf{F}^{\mathrm{H}}\mathbf{Q}_{\mathbf{H}}\mathbf{F}\,\boldsymbol{\Sigma}\,\mathbf{z}. \tag{6}$$

This optimization must be repeated for all possible bit load constellations (and implicitly $\boldsymbol{\Sigma}$), with the maximizing bit load then selected for transmission. Clearly, this is not a convex problem and no analytical solution has been found in the general case. Furthermore, if $\mathbf{Q}_{\mathbf{H}}$ is stochastic the problem becomes ever harder. By imposing certain TX-CSI dependent structures on $\mathbf{F}$ the problem can be simplified, which allows us to draw some conclusions as shown in the following subsections.

4.1 Diagonalizing precoder

If the matrix $\mathbf{Q_H}$ is perfectly known to the transmitter, a common precoding strategy is to let $\mathbf{F}$ diagonalize the channel so that the crosstalk is eliminated [1, 14]. Assume the singular value decomposition of $\mathbf{Q_H}$ is $\mathbf{U_H \Lambda_H U_H^H}$. Let $\tilde{\mathbf{\Lambda}}_\mathbf{H} \in \mathbb{C}^{KL\times KL}$ be the diagonal matrix with the KL strongest singular values, and let $\tilde{\mathbf{U}}_\mathbf{H} \in \mathbb{C}^{NL\times KL}$ be the corresponding eigenvectors from $\mathbf{U_H}$. A diagonalizing precoder must be on the form $\mathbf{F} = \tilde{\mathbf{U}}_\mathbf{H}\mathbf{\Lambda}$, where $\mathbf{\Lambda} \in \mathbb{C}^{KL\times KL}$ is diagonal and positive definite. Within this class of precoders, problem (6) has an analytical solution (see Appendix A)

$$\mathbf{F} = \alpha\, \tilde{\mathbf{U}}_\mathbf{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}^{-1/2}\mathbf{\Sigma}^{-1}.$$

The scalar coefficient α ensures that the power constraint on the system is fulfilled. By redistributing the bit load on the spatial channels, $\mathbf{\Sigma}$ can be used to compensate and improve the power efficiency for potentially ill-conditioned channel matrices. Unfortunately, this diagonalizing precoder can show weak performance when the CSI is imperfect. To see this, introduce a small perturbation on the $\mathbf{Q_H}$ matrix, $\hat{\mathbf{Q}}_\mathbf{H} = \mathbf{Q_H} + \mathbf{\Delta}$, where $\mathbf{\Delta}$ is hermitian but not necessarily positive semi-definite (PSD). The perturbed separation profile becomes

$$Z(\mathbf{H}) = \alpha^2\, \mathbf{z}^\mathrm{H}\, (\mathbf{I} + \tilde{\mathbf{\Lambda}}_\mathbf{H}^{-1/2}\tilde{\mathbf{U}}_\mathbf{H}^\mathrm{H}\mathbf{\Delta}\tilde{\mathbf{U}}_\mathbf{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}^{-1/2})\, \mathbf{z}. \tag{7}$$

The smallest eigenvalue of $\tilde{\mathbf{\Lambda}}_\mathbf{H}$ will magnify the second term and potentially destroy the minimum distance separation. Thus, unless the channel eigenvalues are roughly equal, this precoder structure is not robust to imperfections in the channel estimate.

4.2 Non-diagonalizing precoder

In order to define a precoder that is robust while still providing good minimum distance separation we propose to use a precoder where $\mathbf{F\Sigma}$ is unitary, so that the second term in (7) is not magnified by large eigenvalues. More specifically the following structure is chosen

$$\mathbf{F} = \beta\, \tilde{\mathbf{U}}_\mathbf{H}\mathbf{U}\mathbf{\Sigma}^{-1}, \quad \beta^2 = \frac{PL}{\mathrm{Tr}\{\mathbf{\Sigma}^{-2}\}},$$

where $\mathbf{U}$ is a unitary matrix that should be selected to maximize the minimum distance, and the bit load should be as evenly distributed as possible to maximize the power efficiency. Ideally, for certain bit rates $\mathbf{\Sigma}$ can be proportional to the identity matrix, making the precoder completely unitary. In [8], unitary precoders were proposed for peak-power limited, quantized feedback precoders, and although the scenario herein is somewhat different this further motivates using this precoder for perturbed channels.

The optimal $\mathbf{U}$ depends on the singular values $\tilde{\mathbf{\Lambda}}_\mathbf{H}$ and is difficult to derive in the general case. However, the following upper bound on the minimum distance (disregarding the constant β^2)

$$\min_{i=1,\ldots,KL}[\mathbf{U}^\mathrm{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{U}]_{i,i} \geq \min_{\mathbf{z}\in\mathcal{I}_K} \mathbf{z}^\mathrm{H}\mathbf{U}^\mathrm{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{U}\mathbf{z},$$

is maximized if $\mathbf{U}$ is a DFT matrix, $\mathbf{U}_\mathrm{DFT}$. This suggests a DFT-based precoder will have good minimum distance properties. While the upper bound does not guarantee large minimum distance, for $\mathbf{U} = \mathbf{U}_\mathrm{DFT}$ a lower bound can be derived (see Appendix B),

$$\min_{\mathbf{z}\in\mathcal{I}_K} \mathbf{z}^\mathrm{H}\mathbf{U}_\mathrm{DFT}^\mathrm{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{U}_\mathrm{DFT}\mathbf{z} \geq \min_{n\in\{1,\ldots,KL\}} \frac{KL}{n^2}\sum_{k=1}^{n}\boldsymbol{\lambda}_k,$$

where $\boldsymbol{\lambda}_k$ contains the diagonal elements of $\tilde{\mathbf{\Lambda}}_\mathbf{H}$ sorted in increasing order. Interestingly, the lower bound coincides with the upper bound if the singular values are roughly equal, making the DFT precoder optimal. In the other extreme when the smallest singular value is significantly lower than the others, the lower bound can be simplified to

$$\min_{\mathbf{z}\in\mathcal{I}_K} \mathbf{z}^\mathrm{H}\mathbf{U}_\mathrm{DFT}^\mathrm{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{U}_\mathrm{DFT}\mathbf{z} \geq K\min_k[\tilde{\mathbf{\Lambda}}_\mathbf{H}]_{k,k},$$

which is a factor K times larger than what can be guaranteed with an arbitrary unitary matrix. In any case the DFT matrix always outperforms $\mathbf{U} = \mathbf{I}$ in terms of minimum distance.

To conclude, our proposed precoder is on the form

$$\mathbf{F} = \beta\, \tilde{\mathbf{U}}_\mathbf{H}\mathbf{U}_\mathrm{DFT}\mathbf{\Sigma}^{-1}, \quad \beta^2 = \frac{PL}{\mathrm{Tr}\{\mathbf{\Sigma}^{-2}\}},$$

and shares similarities with the precoder that was proposed for linear receivers in [7], although our problem and the motivations are somewhat different due to the non-linear ML-detection applied herein.

5. NUMERICAL EXAMPLES

This section presents simulation results that demonstrate the gains attained by using the non-diagonalizing precoder as opposed to diagonalizing precoders. The channel is modeled as a Rayleigh, MIMO block fading channel with no correlation between the matrix elements, i.e. the a-priori distribution of the channel matrix is $\mathrm{vec}(\mathbf{H}) \sim \mathit{CN}(\mathbf{0}, \mathbf{I})$. At all times, the channel matrix is assumed to be perfectly known at the receiver while the TX-CSI is imperfect. The distribution of the channel given the TX-CSI, $\mathbf{H}|\boldsymbol{\zeta}$, is

$$\mathrm{vec}(\mathbf{H}|\boldsymbol{\zeta}) \sim \mathit{CN}\big(\mathrm{vec}(\hat{\mathbf{H}}), (1-\sigma^2)\mathbf{I}\big),$$

where $\hat{\mathbf{H}}$ is an imperfect channel estimate. It is reasonable to assume that the estimation error is uncorrelated with the estimate, making the a-priori distribution of the channel estimate

$$\mathrm{vec}(\hat{\mathbf{H}}) \sim \mathit{CN}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{8}$$

The parameter $\sigma^2 \in [0, 1]$, is a measure of the quality of the channel estimate, $\sigma^2 = 0$ implies no instant channel knowledge (although the a-priori statistics are known), and $\sigma^2 = 1$ is equivalent to perfect channel knowledge.

Although the proposed LDC can be designed for arbitrary block lengths, this example assumes block length $L = 1$ for simplicity. The constellations used for the spatial channels are rectangular QAM, and the receiver employs ML detection implemented using the sphere-decoding technique [15].

5.1 The benchmarking schemes

The proposed robust precoding scheme from Section 4.2 will be compared with the diagonalizing precoder from Section 4.1 as well as two benchmarking transmission schemes, blind transmission and optimized BPL.

-*Blind transmission:* The first benchmarking scheme is the well known V-BLAST scheme [16], where the bits are evenly distributed among the spatial channels and no linear precoding is performed. This scheme does not consider the TX-CSI at all, and should intuitively perform the worst.

-*Optimal diagonal BPL transmission:* The second benchmarking scheme was presented in [12], in which the linear precoder has the following form

$$\mathbf{F} = \mathbf{U_H}\mathbf{\Lambda}\mathbf{\Sigma}^{-1},$$

14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

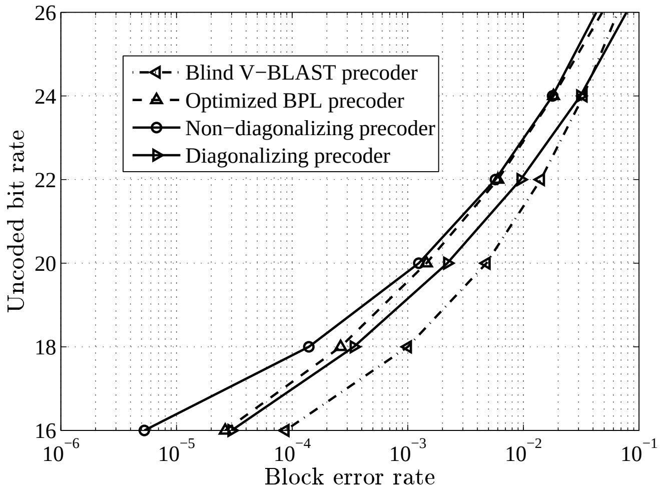

Figure 1: Monte Carlo simulated block error rate for a Rician 8×8 MIMO channel at different data rates. Four precoding strategies are compared. The SNR is set to 16 dB and the TX-CSI quality is $\sigma^2 = 0.8$.

where $\mathbf{U_H}$ contains the unitary eigenvectors of $\mathrm{E}_{\mathbf{H}|\zeta}[\mathbf{H}^{\mathrm{H}}\mathbf{H}]$, $\mathbf{\Lambda}$ is a diagonal matrix that defines the power allocation to the spatial channels, and $\mathbf{\Sigma}^{-1}$ is a diagonal matrix that scales the symbols to the complex integer lattice (with an offset $j\,0.5 + 0.5$). The power load, $\mathbf{\Lambda}$, and bit load are optimized to minimize the union bound of the block error rate given the TX-CSI. This optimization is complicated and not discussed in this paper, but this scheme is essentially optimal given the constraints of the code structure and disregarding some approximate steps in the design procedure [12]. It is important to point out that this is *not* the suboptimal diagonalizing scheme from Section 4.1 which assumed perfect CSI and only considered maximizing the minimum distance (not minimizing the union bound).

5.2 Simulation procedure and results

In this numerical example, 15 channel estimates were drawn independently according to (8). For each channel estimate LDCs where designed as described above (and optimized for the BPL code) at design data rates from 16 to 26 bits per channel use. The block error rate of the codes where then evaluated using Monte Carlo simulations of the joint ML-detection. Finally the block error rates where averaged over the channel estimates to approximate the long term error rate at each uncoded bit rate level.

Figure 1 shows the average performance as a function of the uncoded data rate. The MIMO system has eight transmitting and eight receiving antennas, the SNR is set to 16 dB, and the quality of the TX-CSI is $\sigma^2 = 0.8$. From the plot we can conclude that the robust precoder outperforms both the precoder from Section 4.1 and the optimized BPL precoder. This is a great improvement considering the design complexity is significantly lower for the robust precoder (at least in the latter case). No costly union bound evaluations and no iterative optimization of the spatial bit and power load are required what so ever.

6. CONCLUSIONS

Diagonalizing codes for MIMO systems using ML receivers were observed to suffer from problems with robustness when the TX-CSI is imperfect. A non-diagonalizing precoding technique that has been proven effective for linear receivers, was proposed to improve robustness. It was shown that the code also has good minimum distance separation between received codewords. The code is very suitable for systems operating at high data rates due to its low design complexity. Numerical results show that the precoder outperforms a state-of-the-art diagonalizing precoder.

A. DIAGONALIZING PRECODER: MAXIMAL MINIMUM DISTANCE

Assuming perfect TX-CSI, we derive the maximum minimum-distance precoder for the class of diagonalizing precoders. The singular value decomposition of $\mathbf{Q_H}$ is $\mathbf{U_H \Lambda_H U_H^{\mathrm{H}}}$. Let $\tilde{\mathbf{U}}_\mathbf{H} \in \mathbb{C}^{NL \times KL}$ be the columns of $\mathbf{U_H}$ that corresponds to the KL strongest singular values. Assume the precoder has the following form $\mathbf{F} = \tilde{\mathbf{U}}_\mathbf{H}\mathbf{\Lambda}^{1/2}$. Then the optimization problem (6) is

$$\max_{\mathrm{Tr}\{\mathbf{\Lambda}\} \le PL}\ \min_{\mathbf{z} \in \mathcal{I}_{KL}} \mathbf{z}^{\mathrm{H}}\,\mathbf{\Sigma}^2\mathbf{\Lambda}\tilde{\mathbf{\Lambda}}_\mathbf{H}\,\mathbf{z} = \max_{\substack{\mathbf{1}^{\mathrm{T}}\boldsymbol{\lambda} \le PL \\ \mathbf{1}t \le \tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{\Sigma}^2\boldsymbol{\lambda}}} t,$$

where $\boldsymbol{\lambda}$ is the vector of diagonal elements of the diagonal matrix $\mathbf{\Lambda}$, and $\mathbf{1}$ is the unit vector. Since $\mathbf{\Sigma}^2\tilde{\mathbf{\Lambda}}_\mathbf{H}$ is diagonal and strictly positive, the second inequality is equivalent with

$$(\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{\Sigma}^2)^{-1}\mathbf{1}t \le \boldsymbol{\lambda} \quad \Rightarrow \quad \mathbf{1}^{\mathrm{T}}(\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{\Sigma}^2)^{-1}\mathbf{1}t \le \mathbf{1}^{\mathrm{T}}\boldsymbol{\lambda} \le PL \quad \Rightarrow$$

$$t \le \frac{PL}{\mathbf{1}^{\mathrm{T}}(\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{\Sigma}^2)^{-1}\mathbf{1}} \equiv t_{\max}$$

This bound is attained when $\boldsymbol{\lambda} = t_{\max}\,(\tilde{\mathbf{\Lambda}}_\mathbf{H}\mathbf{\Sigma}^2)^{-1}\mathbf{1}$, and consequently the maximizing minimum distance precoder is

$$\mathbf{F} = \sqrt{t_{\max}}\,\tilde{\mathbf{U}}_\mathbf{H}\tilde{\mathbf{\Lambda}}_\mathbf{H}^{-1/2}\mathbf{\Sigma}^{-1}.$$

It is important that this type of precoders use bit loading (that indirectly affects $\mathbf{\Sigma}$) to maximize $t_{\max}$. In this case this is quite a simple numerical task though.

B. MINIMUM DISTANCE BOUND FOR THE NON-DIAGONALIZING CODE

Here, a lower bound on the maximum minimum distance is derived for the non-diagonal DFT precoders. For simplicity, the proof will be shown for block length $L = 1$. The extension to general $L > 1$ is straight forward.

A unitary DFT matrix, $\mathbf{U}_{\mathrm{DFT}} \in \mathbb{C}^{K \times K}$, satisfy $\mathbf{U}_{\mathrm{DFT}}^{\mathrm{H}}\mathbf{U}_{\mathrm{DFT}} = \mathbf{I}$, and $|[\mathbf{U}_{\mathrm{DFT}}]_{i,j}|^2 = \frac{1}{K}$ for all i, j from 1 to K. Given a complex valued vector, $\mathbf{z}$, we define the real valued vector

$$\mathbf{d} \equiv \mathrm{diag}\{\mathbf{U}_{\mathrm{DFT}}\mathbf{z}\mathbf{z}^{\mathrm{H}}\mathbf{U}_{\mathrm{DFT}}^{\mathrm{H}}\}.$$

The following rules are easily established

$$\mathbf{d}_i = |[\mathbf{U}_{\mathrm{DFT}}\mathbf{z}]_i|^2 \ge 0,$$
$$\mathbf{1}^{\mathrm{T}}\mathbf{d} = \mathrm{Tr}\{\mathbf{U}_{\mathrm{DFT}}\mathbf{z}\mathbf{z}^{\mathrm{H}}\mathbf{U}_{\mathrm{DFT}}^{\mathrm{H}}\} = \mathbf{z}^{\mathrm{H}}\mathbf{z},$$

where $\mathbf{1}$ is the unit vector of length K. Now restrict $\mathbf{z}$ to lie in the complex integer lattice (minus the origin),

$$\mathbf{z} \in \mathcal{I}_K \equiv \{\mathbb{Z}^K + j\mathbb{Z}^K\} \setminus \{0^K\},$$

hence either $|\mathbf{z}_i| = 0$ or $|\mathbf{z}_i| \ge 1$ for all i. As a consequence of this

$$\sum_{k=1}^{K} |\mathbf{z}_k| \le \mathbf{z}^{\mathrm{H}}\mathbf{z}. \tag{9}$$

Using the triangle inequality, (9), and the properties of the DFT matrix we have

$$\begin{aligned}\mathbf{d}_i &= |[\mathbf{U}_{\mathrm{DFT}}\mathbf{z}]_i|^2 = \Big|\sum_{k=0}^{K}[\mathbf{U}_{\mathrm{DFT}}]_{i,k}\mathbf{z}_k\Big|^2 \\ &\leq \Big(\sum_{k=0}^{K}|[\mathbf{U}_{\mathrm{DFT}}]_{i,k}||\mathbf{z}_k|\Big)^2 \leq \frac{(\mathbf{z}^{\mathrm{H}}\mathbf{z})^2}{K} \equiv d_{\max}\end{aligned}$$

Let $\boldsymbol{\lambda} \in \mathbb{R}_+^K$ be vector ordered such that $0 \leq \boldsymbol{\lambda}_1 \leq \boldsymbol{\lambda}_2 \leq ... \leq \boldsymbol{\lambda}_K$. Since each $\mathbf{d}_k \in [0, d_{\max}]$ and they sum up to $\mathbf{z}^{\mathrm{H}}\mathbf{z}$ we know that

$$\tilde{\mathbf{d}} \equiv [\underbrace{d_{\max}, d_{\max}, ..., d_{\max}}_{n \text{ elements}}, \mathbf{z}^{\mathrm{H}}\mathbf{z} - n\, d_{\max}, 0, 0, ..., 0]^{\mathrm{T}}$$

where $n \equiv \mathrm{floor}(K/\mathbf{z}^{\mathrm{H}}\mathbf{z})$ gives a lower bound to $\mathbf{d}^{\mathrm{T}}\boldsymbol{\lambda}$, i.e.

$$\mathbf{d}^{\mathrm{T}}\boldsymbol{\lambda} \geq \tilde{\mathbf{d}}^{\mathrm{T}}\boldsymbol{\lambda} = \frac{(\mathbf{z}^{\mathrm{H}}\mathbf{z})^2}{K}\Big(\Big(\frac{K}{\mathbf{z}^{\mathrm{H}}\mathbf{z}} - n\Big)\boldsymbol{\lambda}_{n+1} + \sum_{k=1}^{n}\boldsymbol{\lambda}_k\Big).$$

This bound contains $\mathbf{z}$ which makes it a little tedious to work with. It is however possible to remove this dependency by lower bounding the lower bound as follows. Define

$$r = K/\mathbf{z}^{\mathrm{H}}\mathbf{z} - n \in [0, 1).$$

If $\mathbf{z}^{\mathrm{H}}\mathbf{z} \geq K$, then $n = 0$ and the lower bound becomes $\mathbf{d}^{\mathrm{T}}\boldsymbol{\lambda} \geq \boldsymbol{\lambda}_1\mathbf{z}^{\mathrm{H}}\mathbf{z} \geq \boldsymbol{\lambda}_1 K$. With a similar approach, we will now analyze the bound for all possible $n = 0, 1, ..., K$. Define a helping function $f(\cdot)$ as

$$\frac{\tilde{\mathbf{d}}^{\mathrm{T}}\boldsymbol{\lambda}}{K} = \frac{1}{(n+r)^2}\Big(r\boldsymbol{\lambda}_{n+1} + \sum_{k=1}^{n}\boldsymbol{\lambda}_k\Big) \equiv f(r)\boldsymbol{\lambda}_{n+1}.$$

We wish to find a lower bound for all $r \in [0, 1]$, hence we should minimize the function

$$f(r) = \frac{r+\alpha}{(r+n)^2}, \quad 0 \leq \alpha < n.$$

Differentiation tell us that $f(r)$ has one optimum on $r \in (-\infty, \infty)$ that is a maximum. Hence, the minimum in $r \in [0, 1]$ must occur on the boundaries. In fact, if $r = 0$ both points will be considered when n ranges from 1 to K since $r = 0$ implies $r = 1$ when n is one smaller. Assuming $\boldsymbol{\lambda}$ is the sorted diagonal elements of $\tilde{\boldsymbol{\Lambda}}_{\mathbf{H}}$, we have proven that the minimum distance can be lower bounded as

$$\min_{\mathbf{z}\in\mathcal{I}_K} \mathbf{z}^{\mathrm{H}}\mathbf{U}_{\mathrm{DFT}}^{\mathrm{H}}\tilde{\boldsymbol{\Lambda}}_{\mathbf{H}}\mathbf{U}_{\mathrm{DFT}}\mathbf{z} \geq \min_{n\in\{1,...,K\}} \frac{K}{n^2}\sum_{k=1}^{n}\boldsymbol{\lambda}_k.$$

It may be hard to interpret the bound intuitively. However, if the elements of $\boldsymbol{\lambda}$ are roughly equal, then the minimizing $n = K$, while the minimizing n equals 1 if the $\boldsymbol{\lambda}_1$ is much smaller than the other $\boldsymbol{\lambda}_k$'s.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Technical Memorandum, Bell Laboratories (Published in European Transactions on Telecommunications, Vol. 10, No.6, pp. 585-595, Nov/Dec 1999)*, 1995.

[2] D. P. Palomar and S. Barbarossa, "Designing MIMO communication systems: Constellation choice and linear transciever design," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3804–3818, October 2005.

[3] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, September 2003.

[4] Y. Ding, T. N. Davidson, and K. M. Wong, "On improving the BER performance of rate-adaptive block transceivers, with applications to DMT," in *Proceedings IEEE Global Telecommunications Conference*, vol. 3, December 2003, pp. 1654–1658.

[5] E. A. Jorswieck, A. Sezgin, H. Boche, and E. Costa, "Optimal transmit strategies in MIMO Ricean channels with MMSE receiver," in *Vehicular Technology Conference*, vol. 5, September 2004, pp. 3787–3791.

[6] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, pp. 611–627, March 2002.

[7] M. Kiessling and J. Speidel, "Statistical prefilter design for MIMO ZF and MMSE receivers based on majorization theory," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, May 2004, pp. 313–316.

[8] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, August 2005.

[9] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding on channel mean feedback," *IEEE Transactions on Information Theory*, vol. 50, pp. 2599–2613, October 2002.

[10] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[11] S. Bergman, C. Martin, and B. Ottersten, "Bit and power loading for spatial multiplexing using partial channel state information," in *Proceedings ITG Workshop on Smart Antennas, Technische Universität Munich*, March 2004.

[12] S. Bergman and B. Ottersten, "Adaptive spatial bit loading using imperfect channel state information," in *Proceedings of International Workshop on Optical and Electronic Device Technology for Access Networks, Aalborg, Denmark*, September 2005.

[13] M. K. Simon and M.-S. Alouini, "A unified approach to the performance analysis of digital communications over generalized fading channels," *Proceedings of the IEEE*, vol. 86, no. 9, pp. 1860–1877, September 1998.

[14] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 357–366, March 1998.

[15] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389–2402, October 2003.

[16] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-blast: An architecture for realizing very high data rates over the rich-scattering wireless channel," *URSI International Symposium on Signals, Systems, and Electronics.*