QUANTIZED FEEDBACK AND FEEDBACK REDUCTION FOR PRECODED SPATIAL MULTIPLEXING MIMO SYSTEMS

Geert Leus and Claude Simon

Delft University of Technology, Fac. EEMCS, Mekelweg 4, 2628CD Delft, The Netherlands

ABSTRACT

In this paper, we discuss different options for quantized feedback and feedback reduction for a precoded spatial-multiplexing multiple-input multiple-output (MIMO) system in a time-varying channel. The novel contributions of this paper are a quantized feedback strategy based on the bit-error-rate (BER) of a linear receiver, and two new feedback reduction strategies for a time-varying channel. Both these feedback reduction schemes exploit the time correlation of the MIMO channel. They basically can be viewed as an optimized generalization of existing feedback reduction strategies.

1. INTRODUCTION

In the last few years, spatial multiplexing emerged as a promising scheme to fulfill the data rate requirements of future wireless services. A technique to make spatial multiplexing more robust to rank deficient channels and to allow for simpler receiver architectures is linear precoding [1].

The optimal precoder is calculated as a function of the channel state information (CSI). However, since CSI is in general only available at the receiver, it needs to be fed back to the transmitter. Since the precoder is generally restricted to be unitary it is beneficial [2] to feed back the quantized precoder, instead of the quantized channel. To feed back precoder information over a bandlimited feedback link, we require a codebook consisting of a set of quantized precoders. Such a codebook is generally designed using some distortion measure [6, 8, 10, 11, 12, 13]. One novel contribution of this paper is the use of the exact bit-error-rate (BER) of a linear receiver as a distortion measure. Further, a selection measure is required that maps the channel to a specific quantized precoder. This selection measure could be similar to the distortion measure, but this does not have to be the case. For instance, when the BER of a linear receiver is the criterion of interest, it is always better to use this as a selection measure, independent of the codebook design [10].

The feedback requirements can be further reduced by exploiting the temporal correlation of the channel. In [3] and [4], a first-order Markov chain is introduced to model the feedback of a set of quantized beamformers that is designed following one of the above approaches. Based on this Markov model, no feedback is sent for the quantized beamformer with the highest transition probability (always the previous quantized beamformer), an equi-length bitword is sent for the quantized beamformers with the second highest transition probability, and no feedback is sent for the remaining quantized beamformers, if there are any. When there are remaining quantized beamformers, as in [4], such a feedback method is clearly lossy, and thus a feedback reduction can only be obtained at the price of a reduced performance. When there are no remaining quantized beamformers, as in [3], the feedback method is lossless, but the feedback rate is not optimally reduced. In [5], another type of feedback reduction is proposed. This method was proposed for time-invariant MIMO-OFDM systems, but it can easily be translated for the current set-up. In that case, a set of quantized beamformers is designed based on maximizing the minimum distance between the beamformers. Further, only a few quantized beamformers that are closest to the previous quantized beamformer are considered and the best one is used. This approach does not require any statistical knowledge about the channel, but when the Doppler spread is large, the method could lose track. To solve this problem, [5] proposes a trellis-based extension at the cost of latency. This partially solves the problem, but we believe it still has tracking problems when the Doppler spread is too large.

In this paper, we will show a few optimized methods for reducing the feedback rate. We will basically focus on two methods. One that optimally adapts the bitwords depending on the previously fed back information, called the *adaptive bitword approach*, and one that adapts the precoders depending on the previously fed back information, called the *adaptive precoder approach*. The previous feedback reduction methods [3],[4],[5] can be considered as special (suboptimal) cases of the proposed approaches.

Notation: Vectors are designated with lowercase boldface letters, and matrices with uppercase boldface letters. $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^{-1}$ denote the transpose, the complex conjugate transpose, and the inverse of the matrix $\mathbf{A}$, respectively. In addition, $\text{abs}(\mathbf{A})$ represents the element-wise absolute value of the matrix $\mathbf{A}$, $\text{tr}(\mathbf{A})$ its trace, $\det(\mathbf{A})$ its determinant, and $\text{diag}(\mathbf{A})$ a diagonal matrix obtained by removing the off-diagonal elements of $\mathbf{A}$. Further, $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$ denote the two-norm and Frobenius norm of $\mathbf{A}$, respectively. Finally, $E(\cdot)$ represents expectation, and $P(\cdot)$ probability.

2. SYSTEM MODEL

We consider a narrowband linearly precoded spatial multiplexing MIMO system, with N_T transmit and N_R receive antennas, transmitting $N_S \leq \min(N_T, N_R)$ symbol streams. For a particular time instant, the input-output relation can then be written as

$$\mathbf{y} = \mathbf{HFs} + \boldsymbol{\nu}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{N_R \times 1}$ is the received vector, $\boldsymbol{\nu} \in \mathbb{C}^{N_R \times 1}$ is the additive noise vector, $\mathbf{s} \in \mathbb{C}^{N_S \times 1}$ is the data symbol vector, $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is the channel matrix, and $\mathbf{F} \in \mathbb{C}^{N_T \times N_S}$ is the linear precoder matrix. We assume that the elements of $\boldsymbol{\nu}$ are i.i.d. and complex Gaussian distributed with zero mean and variance 1, that the elements of $\mathbf{s}$ are i.i.d. and uniformly distributed over a finite alphabet $\mathcal{A}$ with zero mean and variance 1, and that the elements of $\mathbf{H}$ are i.i.d. and complex Gaussian distributed with zero mean and variance P (note that the transmit power is embedded in $\mathbf{H}$). Further, in order to reduce the feedback [6], $\mathbf{F}$ is limited to be unitary, i.e., $\mathbf{F} \in \mathcal{U}_{N_T \times N_S}$, with $\mathcal{U}_{n \times m}$ denoting the set of unitary $n \times m$ matrices. Hence, the received signal-to-noise ratio (SNR) per transmit antenna is given by P.

The singular value decomposition (SVD) of $\mathbf{H}$ will be denoted as $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ belong to $\mathcal{U}_{N_R \times N_R}$

This research was supported in part by NWO-STW under the VIDI program (DTC.6577).

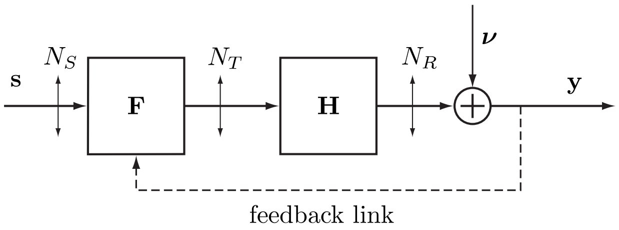

Fig. 1. System model.

and $\mathcal{U}_{N_T \times N_T}$, respectively, and $\mathbf{\Sigma}$ is a diagonal $N_R \times N_T$ matrix with the diagonal starting in the top left corner. Let us also define $\bar{\mathbf{U}} = [\mathbf{U}]_{:,1:N_S} \in \mathcal{U}_{N_R \times N_S}$, $\bar{\mathbf{V}} = [\mathbf{V}]_{:,1:N_S} \in \mathcal{U}_{N_T \times N_S}$, and $\bar{\mathbf{\Sigma}} = [\mathbf{\Sigma}]_{1:N_S,1:N_S}$, where we use Matlab notation in the subscript to select the appropriate rows and columns.

We consider a block-wise transmission structure, where every block consists of a frame of length T_f (in seconds). We assume that the receiver can perfectly estimate the channel state information (CSI) at the beginning of each block, and can feed back some precoder information to the transmitter over a bandlimited yet delay- and error-free feedback link. We will make a distinction between a non-dedicated and a dedicated feedback link. In case of a non-dedicated feedback link, the feedback information has to be instantaneously decodable. In other words, the bitwords that are sent back should be prefix-free (PF), i.e., a bitword can not contain any other bitword as a prefix. This is not the case for a dedicated feedback link, where non-prefix-free (NPF) bitwords can be used. In this context, note that an empty bitword also counts as an NPF bitword.

3. QUANTIZED FEEDBACK

Since the feedback link is bandlimited we need to quantize the precoder information. This requires a codebook consisting of a set of quantized precoders and a set of related channel regions. Further, we need a selection procedure that maps the channel to a specific quantized precoder. Finally, a different bitword is assigned to every quantized precoder, which will be sent over the non-dedicated or dedicated feedback link to inform the transmitter (see Figure 1).

3.1. Codebook Design

Most existing codebook design algorithms, construct the quantized precoders and related channel regions, $\{\mathbf{F}_i, \mathcal{R}_i\}_{i=1}^K$, so that an expected distortion measure between the channel and the quantized precoder is minimized. More specifically, if we assume that the quantized precoder $\mathbf{F}_i$ is picked whenever $\mathbf{H} \in \mathcal{R}_i$, then we wish to minimize

$$\sum_{i=1}^{K} E[D(\mathbf{H}, \mathbf{F}_i)|\mathbf{H} \in \mathcal{R}_i] P(\mathbf{H} \in \mathcal{R}_i), \tag{2}$$

over $\{\mathbf{F}_i, \mathcal{R}_i\}_{i=1}^K$, under the constraints that $\mathbf{F}_i \in \mathcal{U}_{N_T \times N_S}$, $\mathcal{R}_i \subset \mathbb{C}^{N_R \times N_T}$, $\bigcup_i \mathcal{R}_i = \mathbb{C}^{N_R \times N_T}$, and $\mathcal{R}_i \cap \mathcal{R}_{i'} = \varnothing, \forall i \neq i'$, where $D(\mathbf{H}, \mathbf{F})$ represents a distortion measure between the channel $\mathbf{H}$ and the precoder $\mathbf{F}$. Minimizing (2) can be done by the Lloyd algorithm [7].

Note that although the codebook is designed based on a certain selection procedure, the actual selection procedure could be different. More specifically, for a given channel matrix $\mathbf{H}$, we pick the quantized precoder as

$$Q(\mathbf{H}) = \arg\min_{\mathbf{F} \in \{\mathbf{F}_i\}_{i=0}^{K-1}} S(\mathbf{H}, \mathbf{F}), \tag{3}$$

precoders	bitwords non-dedicated	bitwords dedicated
$\mathbf{F}_1$	$w_1 = 00$	$w_1 = /$
$\mathbf{F}_2$	$w_2 = 01$	$w_2 = 0$
$\mathbf{F}_3$	$w_3 = 10$	$w_3 = 1$
$\mathbf{F}_4$	$w_4 = 11$	$w_4 = 00$

Table 1. Example of a 4-entry ($K = 4$) codebook for a non-dedicated and dedicated feedback link.

where $S(\mathbf{H}, \mathbf{F})$ is the selection measure between the channel $\mathbf{H}$ and the precoder $\mathbf{F}$. As explained earlier, the selection measure $S(\mathbf{H}, \mathbf{F})$ can be the same as the distortion measure $D(\mathbf{H}, \mathbf{F})$, in which case $Q(\mathbf{H}) = \mathbf{F}_i$ whenever $\mathbf{H} \in \mathcal{R}_i$, but this does not have to be the case. In the following subsections, we will discuss in more detail possible distortion and selection measures.

Assuming that the probabilities $\{P(Q(\mathbf{H}) = \mathbf{F}_i)\}_{i=1}^K$ are generally all close to $1/K$, there are basically two strategies to assign bitwords $\{w_i\}_{i=1}^K$ to the quantized precoders $\{\mathbf{F}_i\}_{i=1}^K$. Which strategy is chosen depends on the type of feedback channel. For a non-dedicated feedback link, we take K equi-length PF bitwords, whereas for a dedicated feedback link we take K increasing-length NPF bitwords. An example of a 4-entry ($K = 4$) codebook is presented in Table 1 for a non-dedicated and dedicated feedback link. Clearly, the average feedback rate for the non-dedicated feedback link, which is about $\lceil \log_2(K) \rceil$, is larger than the average feedback rate for the dedicated feedback link, which is about $1/K \sum_{i=1}^K \lfloor \log_2 i \rfloor$.

3.2. Distortion Measures

In [6] and [8], a number of performance measures has been transformed into subspace distances between $\bar{\mathbf{V}}$ and $\mathbf{F}_i$, where $\bar{\mathbf{V}}$ is the optimal precoder for that performance measure. For instance, the minimum received distance between two noiseless received vectors [related to the performance of the maximum likelihood (ML) receiver and called the ML performance measure], the minimum singular value (MSV) of the product of the channel and the precoder (called the MSV performance measure), and the trace of the mean square error (MSE) of the linear minimum mean squared error (LMMSE) receiver (called the trace-MSE performance measure) can all be transformed into the projection two-norm distance between $\bar{\mathbf{V}}$ and $\mathbf{F}_i$ [6]:

$$d_{p2}(\bar{\mathbf{V}}, \mathbf{F}_i) = \|\bar{\mathbf{V}}\bar{\mathbf{V}}^H - \mathbf{F}_i\mathbf{F}_i^H\|_2.$$

Further, the determinant of the MSE of the LMMSE receiver (called the det-MSE performance measure) and the capacity can be transformed into the Fubini-Study distance between $\bar{\mathbf{V}}$ and $\mathbf{F}_i$ [6]:

$$d_{FS}(\bar{\mathbf{V}}, \mathbf{F}_i) = \arccos|\det(\bar{\mathbf{V}}^H\mathbf{F}_i)|.$$

Finally, the Frobenius norm of the product of the channel and the precoder [related to the performance of an orthogonal space-time block code (OSTBC) used on top of the precoder and called the OSTBC performance measure] can be transformed into the chordal distance between $\bar{\mathbf{V}}$ and $\mathbf{F}_i$ [8]:

$$d_c(\mathbf{H}, \mathbf{F}_i) = \frac{1}{\sqrt{2}}\|\bar{\mathbf{V}}\bar{\mathbf{V}}^H - \mathbf{F}_i\mathbf{F}_i^H\|_F.$$

The squares of all these distances could be used as a distortion measure in (2), i.e., $D_{p2} = d_{p2}^2$, $D_{FS} = d_{FS}^2$, and $D_c = d_c^2$. Actually, [6] and [8] do not minimize (2) using these distortion measures but they try to maximize the minimum subspace distance between the quantized precoders $\{\mathbf{F}_i\}_{i=1}^K$. This results in a subspace packing problem on a Grassmann manifold and algorithms presented in [9] are adopted. In [10], however, (2) is minimized using the above distortion measures, leading to a slightly improved performance. The reason why squared subspace distances are generally used as a distortion measure in (2) is because the centroid computation of the Lloyd algorithm can then be carried out in closed form.

In [11], an alternative distortion measure is adopted, which is related to the capacity loss introduced by quantization. In contrast to the squared subspace distances mentioned earlier, this distortion measure depends on the SNR (through $\bar{\mathbf{\Sigma}}$):

$$D_{cl}(\mathbf{H}, \mathbf{F}_i) = \text{tr}\left(\mathbf{\Lambda} - \mathbf{\Lambda}\bar{\mathbf{V}}^H\mathbf{F}_i\mathbf{F}_i^H\bar{\mathbf{V}}\right), \quad (4)$$

where $\mathbf{\Lambda} = (\mathbf{I}_{N_S} + \bar{\mathbf{\Sigma}}^2)^{-1}\bar{\mathbf{\Sigma}}^2$. Note the close resemblance to the squared chordal distance:

$$\begin{aligned} D_c(\mathbf{H}, \mathbf{F}_i) &= \frac{1}{2}\|\bar{\mathbf{V}}\bar{\mathbf{V}}^H - \mathbf{F}_i\mathbf{F}_i^H\|_F^2 \\ &= \text{tr}(\mathbf{I}_{N_S} - \bar{\mathbf{V}}^H\mathbf{F}_i\mathbf{F}_i^H\bar{\mathbf{V}}), \end{aligned} \quad (5)$$

which is independent of the SNR. Although (4) is not strictly speaking a squared subspace distance, the centroid can still be computed in closed form [11].

A common problem to all the above distortion measures is that all precoder matrices in the same subspace have the same distortion. This is not a problem if the related performance measures are considered, because they are invariant to a left multiplication of the precoder with a unitary matrix. However, if the BER of a linear receiver is considered, for instance, this becomes a problem because the BER of a linear receiver is not invariant to a left multiplication of the precoder with a unitary matrix. It has been shown in [10] that for some SNR regions $\mathbf{F} = \bar{\mathbf{V}}$ is optimal, whereas for other SNR regions $\mathbf{F} = \bar{\mathbf{V}}\mathbf{M}$ is optimal, with $\mathbf{M} \in \mathcal{U}_{N_S \times N_S}$ having constant modulus entries, e.g., the Hadamard or the DFT matrix. Further, even though there are also SNR regions for which the optimal unitary precoder is not known, choosing either one of the precoders $\mathbf{F} = \bar{\mathbf{V}}$ or $\mathbf{F} = \bar{\mathbf{V}}\mathbf{M}$ is a good strategy. This means we actually need a quantization procedure that quantizes $\bar{\mathbf{V}}$ by retaining the order of the singular vectors. Depending on the linear receiver and the SNR region, we can then either use the quantized precoder $\mathbf{F}_i$ or $\mathbf{F}_i\mathbf{M}$ at the transmitter. In [12], for instance, the squared Frobenius norm distance between $\bar{\mathbf{V}}$ and $\mathbf{F}_i$ was considered:

$$D_F(\mathbf{H}, \mathbf{F}_i) = \|\bar{\mathbf{V}} - \mathbf{F}_i\|_F^2 = 2\text{tr}[\mathbf{I}_{N_S} - \Re(\bar{\mathbf{V}}^H\mathbf{F}_i)]. \quad (6)$$

This approach was modified in [13] to take the phase ambiguity of the singular vectors into account, labeled the squared modified Frobenius norm distance:

$$\begin{aligned} D_{mF}(\mathbf{H}, \mathbf{F}_i) &= \min_{\mathbf{\Theta} \in \mathcal{D}_{N_S}} \|\bar{\mathbf{V}}\mathbf{\Theta} - \mathbf{F}_i\|_F^2 \\ &= \|\bar{\mathbf{V}}\text{diag}(\bar{\mathbf{V}}^H\mathbf{F}_i)\text{diag}^{-1}(\text{abs}(\bar{\mathbf{V}}^H\mathbf{F}_i)) - \mathbf{F}_i\|_F^2 \\ &= 2\text{tr}[\mathbf{I}_{N_S} - \text{abs}(\bar{\mathbf{V}}^H\mathbf{F}_i)], \end{aligned} \quad (7)$$

with $\mathcal{D}_n$ denoting the set of diagonal unitary $n \times n$ matrices. Note that these two distortion measures are again independent of the SNR. Also observe the difference with the squared chordal distance of (5). Through the use of the real value or the absolute value of $\bar{\mathbf{V}}^H\mathbf{F}_i$ instead of the product $\bar{\mathbf{V}}^H\mathbf{F}_i\mathbf{F}_i^H\bar{\mathbf{V}}$, we truly encode $\bar{\mathbf{V}}$ and not only its subspace. The problem with the two above distortion measures is that the centroid computation required for the Lloyd algorithm can not be carried out in closed form as with the previous distortion measures. Therefore, we apply a brute-force centroid computation by exhaustively searching for the best center, i.e., the channel which has the minimal average distortion within a region. This trick actually allows us to use all kinds of distortion measures, including the BER of a linear receiver, for which exact expressions exist [14]. Hence, we can consider a BER distortion function that is given by

$$D_{ber}(\mathbf{H}, \mathbf{F}_i) = BER(\mathbf{H}, \mathbf{F}_i),$$

where $BER(\mathbf{H}, \mathbf{F}_i)$ is the average BER of a linear receiver for a channel $\mathbf{H}$ and a precoder $\mathbf{F}_i$. Note that this BER distortion function depends again on the SNR.

3.3. Selection Measures

In the previous section, many distortion measures were derived based on specific performance measures. Usually these performance measures are used as a selection measure in the selection procedure. In other words, for the squared projection two-norm distance D_{p2}, we will use the ML, MSV, or trace-MSE performance measure in the selection procedure, for the squared Fubini-Study distance D_{FS}, we will use the det-MSE or capacity performance measure in the selection procedure, and for the squared chordal distance D_c, we will use the OSTBC performance measure in the selection procedure. For the capacity loss distortion measure D_{cl}, the squared Frobenius norm distance D_F, the modified squared Frobenius distance D_{mF}, and the BER distortion measure D_{ber}, we will generally use the distortion measure itself to guide our precoder selection, i.e., $S(\mathbf{H}, \mathbf{F}_i) = D(\mathbf{H}, \mathbf{F}_i)$.

The above approach seems to be the most natural. However, in [10], the BER of the adopted linear receiver is always used as a selection measure, independent of the distortion measure used to construct the codebook. When the BER of a linear receiver is the criterion of interest, this approach is of course always the best. Note that in case we want to apply this approach when D_F or D_{mF} are used as a distortion measure, we should first decide on whether to use $\mathbf{F}_i$ or $\mathbf{F}_i\mathbf{M}$ at the transmitter, a decision which will depend on the linear receiver and the SNR region [10].

4. FEEDBACK REDUCTION

There are basically two ways to reduce the feedback requirements in a time-varying channel. You can stick to the original codebook and adapt the bitwords depending on the previously fed back information. We call this approach the *adaptive bitword approach*. Alternatively, we could also adapt the precoders depending on the previously fed back information. We call this approach the *adaptive precoder approach*. We next show how to optimize both approaches. For simplicity reasons, we only take the most recent feedback instant into account for which the channel matrix was given by $\mathbf{H}'$. Remember that the current channel matrix is denoted by $\mathbf{H}$.

4.1. Adaptive Bitword Approach

For this approach, we adopt a codebook $\{\mathbf{F}_i, \mathcal{R}_i\}_{i=1}^K$ that was designed as discussed in Section 3. However, we construct the bitwords that are assigned to the quantized precoders in an adaptive fashion depending on the previous quantized precoder. If for instance $Q(\mathbf{H}') = \mathbf{F}_j$, we assign to the quantized precoders $\{\mathbf{F}_i\}_{i=1}^K$ the bitwords $\{w_{i,j}\}_{i=1}^K$ that are found by minimizing

$$\sum_{i=1}^{K} l(w_{i,j})P(Q(\mathbf{H}) = \mathbf{F}_i | Q(\mathbf{H}') = \mathbf{F}_j), \quad (8)$$

where $l(w)$ represents the length of the bitword w and $P(Q(\mathbf{H}) = \mathbf{F}_i | Q(\mathbf{H}') = \mathbf{F}_j)$ stands for the transition probability that the current quantized precoder is $\mathbf{F}_i$ given that the previous quantized precoder is $\mathbf{F}_j$. Hence, this approach is based on a first-order Markov chain for the quantized precoders, as done in [3] and [4].

The solution of the above problem depends on the type of feedback link. For a non-dedicated feedback link, the solution is given by K PF bitwords that are computed using the Huffman algorithm based on the transition probabilities [15], whereas for a dedicated feedback link the solution is given by K increasing-length NPF bitwords with decreasing order of transition probability. An example of an 8-entry ($K = 8$) codebook when $Q(\mathbf{H}') = \mathbf{F}_1$ is presented in Table 2 for a non-dedicated and dedicated feedback link. Note that as the Doppler frequency increases and hence the correlation between the channel matrices at the feedback instances decreases, the transition probabilities converge to $1/K$ and the average feedback rate will be the same as in Section 3.

We remark that the schemes proposed in [3] and [4] can be viewed as suboptimal solutions of the proposed approach in case of a dedicated feedback link (note that these papers only consider beamforming, i.e., $N_S = 1$). These schemes adopt an empty bitword for the quantized precoder with the highest transition probability (always the previous quantized precoder), K_a equi-length PF bitwords for the K_a quantized precoders with the second highest transition probabilities, and $K - K_a - 1$ empty bitwords for the remaining quantized precoders, if there are any. See again Table 2 for an example with $K = 8$ and $K_a = 4$. When the transmitter receives an empty bitword, it assumes that the previous quantized precoder should be used. When $K_a < K - 1$ [4], such a feedback method is clearly lossy (see also the example). Hence, for the same feedback rate, it has a worse performance than the proposed adaptive bitword approach. However, it has the flexibility to realize a smaller average feedback rate at the cost of an additional performance loss. When $K_a = K - 1$ [3], the feedback method is lossless, but for the same performance, it has a higher feedback rate than the proposed adaptive bitword approach, since the latter is optimized in this sense.

4.2. Adaptive Precoder Approach

In the adaptive precoder approach, we design a codebook in an adaptive fashion depending on the previous quantized precoder. Let us start from a set of K fixed channel regions $\{\mathcal{R}_i\}_{i=1}^K$, which could for instance correspond to the channel regions of a codebook that was designed as discussed in Section 3. Let us further relate to each channel region $\mathcal{R}_j$ the codebook $\{\mathbf{F}_{i,j}, \mathcal{R}_{i,j}\}_{i=1}^{K_a}$. If we now assume that, given $\mathbf{H}' \in \mathcal{R}_j$, the quantized precoder $\mathbf{F}_{i,j}$ is picked whenever $\mathbf{H} \in \mathcal{R}_{i,j}$, then we basically wish to minimize

$$\sum_{i=1}^{K_a} E[D(\mathbf{H}, \mathbf{F}_{i,j})|\mathbf{H} \in \mathcal{R}_{i,j}] P(\mathbf{H} \in \mathcal{R}_{i,j}|\mathbf{H}' \in \mathcal{R}_j) \quad (9)$$

over $\{\mathbf{F}_{i,j}, \mathcal{R}_{i,j}\}_{i=1}^{K_a}$, under the constraints that $\mathbf{F}_{i,j} \in \mathcal{U}_{N_T \times N_S}$, $\mathcal{R}_{i,j} \subset \mathbb{C}^{N_R \times N_T}$, $\bigcup_i \mathcal{R}_{i,j} = \mathbb{C}^{N_R \times N_T}$, and $\mathcal{R}_{i,j} \cap \mathcal{R}_{i',j} = \varnothing$, $\forall i \neq i'$. Minimizing (9) can again be done by the Lloyd algorithm [7].

The problem now is that the transmitter does not know in which channel region $\mathcal{R}_j$ the previous channel matrix $\mathbf{H}'$ resides solely based on the previous quantized precoder $Q(\mathbf{H}')$. Hence, we assume that the transmitter estimates the previous channel matrix as

$$\hat{\mathbf{H}}' = [Q(\mathbf{H}'), \mathbf{0}_{N_T \times (N_R - N_S)}]^H,$$

and checks to which channel region $\mathcal{R}_j$ it belongs. Under this assumption, the selection procedure is given by

$$Q(\mathbf{H}) = \arg\min_{\mathbf{F} \in \{\mathbf{F}_{i,j}\}_{i=1}^{K_a} | \hat{\mathbf{H}}' \in \mathcal{R}_j} S(\mathbf{H}, \mathbf{F}). \quad (10)$$

See Table 3 for an example of such an adaptive precoder scheme with $K = 8$ and $K_a = 4$.

The recursive feedback reduction scheme proposed in [5] can actually be viewed as a suboptimal solution of the above adaptive precoder approach. Although this method was originally proposed for MIMO-OFDM systems, it can easily be translated for the current set-up. The authors start from a set of K fixed quantized precoders $\{\mathbf{F}_i\}_{i=1}^K$, which correspond to the quantized precoders of a codebook that was designed using subspace packing based on the chordal distance. Further, they relate to each quantized precoder $\mathbf{F}_j$ the set of K_a quantized precoders $\{\mathbf{F}_{i,j}\}_{i=1}^{K_a}$ from the set $\{\mathbf{F}_i\}_{i=1}^K$ that have the smallest chordal distance to $\mathbf{F}_j$. The selection procedure is then given by

$$Q(\mathbf{H}) = \arg\min_{\mathbf{F} \in \{\mathbf{F}_{i,j}\}_{i=1}^{K_a} | Q(\mathbf{H}') = \mathbf{F}_j} S(\mathbf{H}, \mathbf{F}). \quad (11)$$

Hence, compared to the proposed adaptive precoder approach, the starting point is a set of K fixed quantized precoders $\{\mathbf{F}_i\}_{i=1}^K$, instead of a set of K fixed channel regions $\{\mathcal{R}_i\}_{i=1}^K$. Further, the set $\{\mathbf{F}_{i,j}\}_{i=0}^{K_a-1}$ is related to the previous precoder $\mathbf{F}_j$, instead of to the previous channel region $\mathcal{R}_j$, and it is a subset of these K fixed quantized precoders $\{\mathbf{F}_i\}_{i=1}^K$, instead of the result of an optimization procedure. The advantage of this approach is that no statistical knowledge about the channel is required. However, when the Doppler spread is too large, the method could lose track, because every set of K_a quantized precoders $\{\mathbf{F}_{i,j}\}_{i=1}^{K_a}$ only covers a small region around $\mathbf{F}_j$. To solve this problem, [5] proposes a trellis-based extension at the cost of latency. This partially solves the problem, but we believe it still has tracking problems when the Doppler spread is too large. The proposed adaptive precoder approach does not suffer from this problem because every set of K_a quantized precoders $\{\mathbf{F}_{i,j}\}_{i=1}^{K_a}$ is designed optimally depending on the Doppler spread.

As before, we may assume that the probabilities $\{P(Q(\mathbf{H}) = \mathbf{F}_{i,j})\}_{i=1}^{K_a}$ are generally all close to $1/K_a$. Hence, there are again two strategies to assign bitwords $\{w_i\}_{i=1}^{K_a}$ to the quantized precoders $\{\mathbf{F}_{i,j}\}_{i=1}^{K_a}$. For a non-dedicated feedback link, we take K_a equi-length PF bitwords, whereas for a dedicated feedback link we take K_a increasing-length NPF bitwords. Hence, for the different cases presented in Table 3, we assign bitwords as shown in Table 1 with "$\mathbf{F}_i$" replaced by "case i".

5. SIMULATIONS

In this section, we illustrate the proposed ideas with a couple of simulation results. We consider a MIMO system with $N_T = 4$ transmit and $N_R = 4$ receive antennas, over which we transmit $N_S = 2$ symbol streams that are QPSK modulated. First, let us take a look at the codebook that is designed using the actual BER of a linear receiver as a distortion measure, and let us compare this to a codebook that is designed using the squared chordal distance. For both codebooks, we use the BER of the linear receiver as a selection measure. Figures 2 and 3 plot the obtained BER results for a linear zero forcing receiver using codebooks with $K = 4$ and $K = 16$, respectively (2 bits/frame and 4 bits/frame, respectively). It turns out that the actual codebook design does not matter much if the BER is used as a selection measure. A similar observation was made in [10].

Let us next take a look at the adaptive precoder approach. For simplicity, we assume that the distortion measure and the selection measure are both given by the squared modified Frobenius norm distance. The performance measure we consider is the BER of the linear MMSE receiver. Since for this setup $\bar{\mathbf{V}}\mathbf{M}$ instead of $\bar{\mathbf{V}}$ is always optimal [10], we use the precoder $\mathbf{F}_{i,j}\mathbf{M}$ instead of $\mathbf{F}_{i,j}$ at the transmitter. The channel taps are modeled using Jakes' model [16]. We fix the Doppler spread f_D to 30 Hz and change the frame length T_f (note that we could equivalently fix the frame length T_f and change the Doppler spread f_D). For simplicity, we assume that the channel is constant within a frame. Figure 4 gives a comparison between the constant codebook with $K = 64$ (6 bits/frame for a non-dedicated link), the constant codebook with $K = 4$ (2 bits/frame for a non-dedicated link), and the adaptive precoder approach with $K = 64$ and

precoders	transition probability	bitwords non-dedicated	bitwords dedicated	method of [4]
$\mathbf{F}_1$	0.25	$w_{1,1} = 01$	$w_{1,1} = /$	$w_{1,1} = /$
$\mathbf{F}_2$	0.20	$w_{2,1} = 11$	$w_{2,1} = 0$	$w_{2,1} = 00$
$\mathbf{F}_3$	0.18	$w_{3,1} = 000$	$w_{3,1} = 1$	$w_{3,1} = 01$
$\mathbf{F}_4$	0.16	$w_{4,1} = 001$	$w_{4,1} = 00$	$w_{4,1} = 10$
$\mathbf{F}_5$	0.10	$w_{5,1} = 101$	$w_{5,1} = 01$	$w_{5,1} = 11$
$\mathbf{F}_6$	0.08	$w_{6,1} = 1000$	$w_{6,1} = 10$	$w_{6,1} = /$
$\mathbf{F}_7$	0.02	$w_{7,1} = 10010$	$w_{7,1} = 11$	$w_{7,1} = /$
$\mathbf{F}_8$	0.01	$w_{8,1} = 10011$	$w_{8,1} = 000$	$w_{8,1} = /$

Table 2. Example of an 8-entry ($K = 8$) codebook when $Q(\mathbf{H}') = \mathbf{F}_1$ for a non-dedicated and dedicated feedback link.

	$\hat{\mathbf{H}}' \in \mathcal{R}_1$	$\hat{\mathbf{H}}' \in \mathcal{R}_2$	$\hat{\mathbf{H}}' \in \mathcal{R}_3$	$\hat{\mathbf{H}}' \in \mathcal{R}_4$	$\hat{\mathbf{H}}' \in \mathcal{R}_5$	$\hat{\mathbf{H}}' \in \mathcal{R}_6$	$\hat{\mathbf{H}}' \in \mathcal{R}_7$	$\hat{\mathbf{H}}' \in \mathcal{R}_8$
case 1	$\mathbf{F}_{1,1}$	$\mathbf{F}_{1,2}$	$\mathbf{F}_{1,3}$	$\mathbf{F}_{1,4}$	$\mathbf{F}_{1,5}$	$\mathbf{F}_{1,6}$	$\mathbf{F}_{1,7}$	$\mathbf{F}_{1,8}$
case 2	$\mathbf{F}_{2,1}$	$\mathbf{F}_{2,2}$	$\mathbf{F}_{2,3}$	$\mathbf{F}_{2,4}$	$\mathbf{F}_{2,5}$	$\mathbf{F}_{2,6}$	$\mathbf{F}_{2,7}$	$\mathbf{F}_{2,8}$
case 3	$\mathbf{F}_{3,1}$	$\mathbf{F}_{3,2}$	$\mathbf{F}_{3,3}$	$\mathbf{F}_{3,4}$	$\mathbf{F}_{3,5}$	$\mathbf{F}_{3,6}$	$\mathbf{F}_{3,7}$	$\mathbf{F}_{3,8}$
case 4	$\mathbf{F}_{4,1}$	$\mathbf{F}_{4,2}$	$\mathbf{F}_{4,3}$	$\mathbf{F}_{4,4}$	$\mathbf{F}_{4,5}$	$\mathbf{F}_{4,6}$	$\mathbf{F}_{4,7}$	$\mathbf{F}_{4,8}$

Table 3. Example of the adaptive precoder scheme with $K = 8$ and $K_a = 4$.

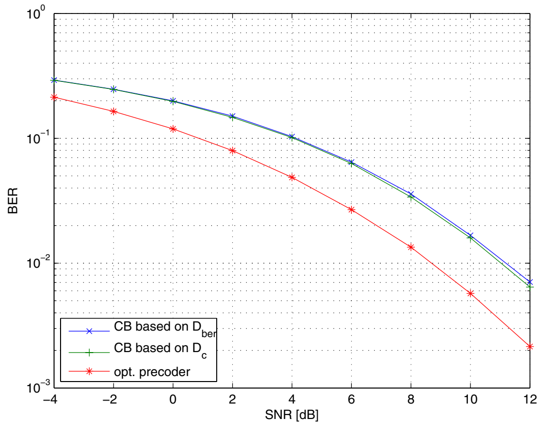

Fig. 2. Performance results for the codebooks with $K = 4$ (2 bits/frame)

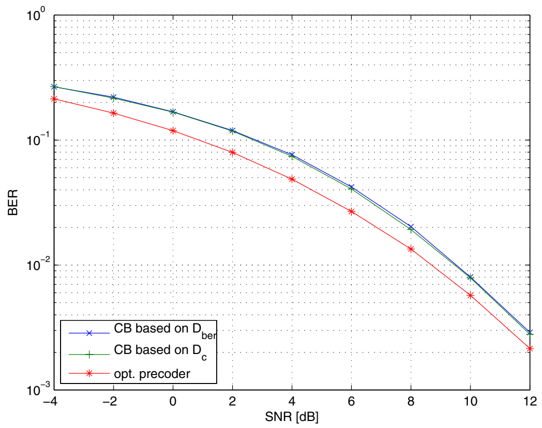

Fig. 3. Performance results for the codebooks with $K = 16$ (4 bits/frame)

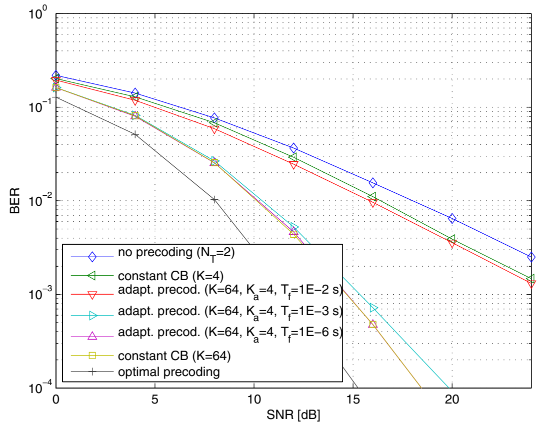

Fig. 4. Performance comparison of the adaptive precoder approach with two constant codebooks.

$K_a = 4$ (2 bits/frame for a non-dedicated link). The performance of the adaptive precoder approach with 2 bits/frame feedback has a performance that approaches the performance of the constant codebook with 2 bits/frame feedback for a large T_f and approaches the performance of the constant codebook with 6 bits/frame feedback for a small T_f.

Finally, let us compare the adaptive precoder approach with the adaptive bitword approach. The same setup as above is considered. We fix the SNR to 10 dB and we assume the feedback link is non-dedicated. Figure 5 shows respectively the average feedback rate and the BER of the linear MMSE receiver as a function of the frame length T_f. For the adaptive bitword approach we consider codebooks with K =2, 4, 8, and 16, whereas for the adaptive precoder approach we take $K = 64$ and K_a =2, 4, 8, and 16. For the adaptive bitword approach, the BER is constant and the average feedback rate increases with an increasing Doppler spread. On the other hand, for the adaptive precoder approach, the average feedback rate is constant and the BER increases with an increasing Doppler spread. Hence, the question basically is how their average feedback rates (BERs) compare for the same BER (average feedback rate). To answer this question, let us take a look at a few examples. We see that the adaptive bitword approach with $K = 8$ has the same average feedback rate as the adaptive precoder approach with $K = 64$ and

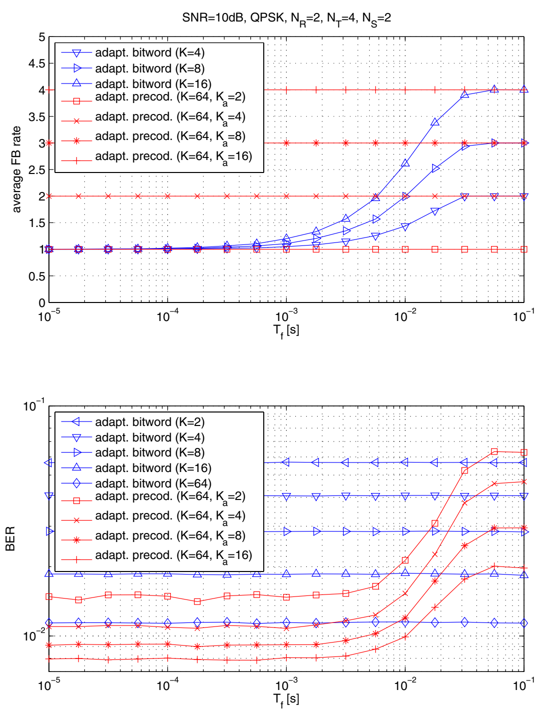

Fig. 5. Performance comparison of the adaptive precoder approach with the adaptive bitword approach.

$K_a = 4$ at $T_f \approx 0.01$. However, at this frame length, the first has a worse BER as the latter. Similarly, we see that the adaptive bitword approach with $K = 8$ has the same BER as the adaptive precoder approach with $K = 64$ and $K_a = 4$ at $T_f \approx 0.02$. But at this frame length, the first has a higher average feedback rate as the latter. Other examples show the same behavior. Hence, we can conclude that for this particular set-up, the adaptive bitword approach is worse than the adaptive precoder approach.

6. CONCLUSIONS

We have given an overview of different quantized feedback and feedback reduction schemes for a precoded spatial-multiplexing MIMO system in a time-varying channel. The novel contributions of this paper are a quantized feedback strategy based on the BER of a linear receiver, and two new feedback reduction strategies for a time-varying channel. Both these feedback reduction schemes exploit the time correlation of the MIMO channel. They basically can be viewed as an optimized generalization of existing feedback reduction strategies.

7. REFERENCES

[1] Anna Scaglione, Petre Stoica, Sergio Barbarossa, Georgios B. Giannakis, and Hemanth Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[2] David J. Love and Robert W. Heath, Jr., "Limited feedback precoding for spatial multiplexing systems using linear receivers," in *Proc. Military Communications Conference (MILCOM'03)*, Oct. 2003, vol. 1, pp. 627 – 632.

[3] Kaibin Huang, Bishwarup Mondal, Robert W. Heath, Jr., and Jeffrey G. Adams, "Effect of feedback delay on multi-antenna limited feedback for temporally-correlated channels," in *Proc. Global Telecommunications Conf. (GLOBECOM'06)*, Dec. 2006.

[4] Kaibin Huang, Bishwarup Mondal, Robert W. Heath, Jr., and Jeffrey G. Adams, "Multi-antenna limited feedback for temporally-correlated channels: Feedback compression," in *Proc. Global Telecommunications Conf. (GLOBECOM'06)*, Dec. 2006.

[5] Shengli Zhou and Peter Willett, "Recursive and trellis-based feedback reduction for MIMO-OFDM with transmit beamforming," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM'05)*, St. Louis, MO, USA,, Nov. 2005, vol. 6.

[6] David J. Love and Robert W. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[7] Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compressing*, Kluwer Academic Publishers, 1995.

[8] David J. Love and Robert W. Heath, Jr., "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 64–73, Jan. 2005.

[9] Bertrand M. Hochwald, Thomas L. Marzetta, Thomas J. Richardson, Wim Sweldens, and Rüdiger Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.

[10] Shengli Zhou and Baosheng Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1653–1665, May 2006.

[11] June Chul Roh and Bhaskar D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Processing*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[12] Nadia Khaled, Bishwarup Mondal, Robert W. Heath, Jr., Geert Leus, and Frederik Petré, "Quantized multi-mode precoding for spatial multiplexing MIMO-OFDM system," in *Proc. IEEE 62nd Vehicular Technology Conference (VTC'05)*, Sept. 2005, vol. 2, pp. 867–871.

[13] Geert Leus, Claude Simon, and Nadia Khaled, "Spatial multiplexing with linear precoding in time-varying channels with limited feedback," in *Proc. 14th European Signal Processing Conference (EUSIPCO'06)*, Florence, Italy, Sept. 2006.

[14] Kyongkuk Cho and Dongweon Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, July 2002.

[15] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[16] William C. Jakes, Jr., *Microwave Mobile Communications*, John Wiley & Sons, 1974.