Feedback Compression for Correlated Broadcast Channels

Claude Simon*, Ruben de Francisco†, Dirk T.M. Slock†, and Geert Leus*
*TU Delft, Fac. EEMCS, Mekelweg 4, 2628 CD Delft, The Netherlands
†Eurécom Institute, 2229, Route des Crêtes, 06560 Valbonne Sophia-Antipolis, France

***Abstract*—In this paper we apply Predictive Vector Quantization (PVQ) to quantize time-correlated broadcast channels. PVQ exploits the time-correlation of the channel to reduce the quantization error of the channels, and thus to improve the sum rate of the system. PVQ predicts the actual channel based on a number of previous channels, and then quantizes the difference between the prediction and the true channel. In this paper we show how the corresponding codebooks can be designed, and we present a prediction strategy. The performance of PVQ for a broadcast system is depicted through numerical simulations.**

***Index Terms*—Broadcast channels, vector quantization, time-varying channels, multiuser channels, MIMO systems.**

I. Introduction

Space division multiple access (SDMA) has emerged in the last years as an attractive transmission scheme for multiple-input multiple-output (MIMO) broadcast channels [1], [2]. It has been shown to outperform time division multiple access (TDMA) [3]. The optimal SDMA scheme for the Gaussian MIMO broadcast channel is dirty-paper coding (DPC) [4], [5], i.e., the rate region of DPC corresponds to the capacity region of the channel. Unfortunately, DPC has a high computational complexity, and is thus difficult to implement. However, zero-forcing (ZF) beamforming has been lately shown [6] to reach asymptotically the same performance as DPC for a high number of users. Most existing SDMA schemes assume channel knowledge at the transmitter side. However, in general, channel state information (CSI) knowledge is only available at the receiver side, and must be fed back to the transmitter. The feedback link is generally assumed to be bandwidth limited, meaning that only a limited number of bits can be fed back to the transmitter. The CSI must thus be quantized before it can be fed back to the transmitter.

A low-complexity SDMA scheme that works with limited feedback is opportunistic SDMA (OSDMA) [7]. OSDMA is an extension of opportunistic beamforming [8] to multiple users. It uses a random set of orthonormal beamforming vectors at the base station with N_T antennas to simultaneously transmit independent data streams to the N_T users with the highest signal-to-noise ratio (SINR). Several extensions of OSDMA have been proposed lately [9], [10] to incorporate larger sets of beamforming vectors, and thus, to improve the performance for scenarios with a lower number of users. Even though the OSDMA algorithms have a good performance for i.i.d. channels, there exists, to the best of the authors' knowledge, no extension of these algorithms to exploit time-correlated channels.

In this paper we present a scheme that uses Predictive Vector Quantization (PVQ) [11] to exploit the correlation between successive channel realizations in order to improve the quantization, and thus to improve the sum rate of the system. Further, our scheme does not make any assumptions on the scheduling function and on the transmission strategy, which allows for a high flexibility.

Notation: We use capital boldface letters to denote matrices, and small boldface letters to denote vectors. $E(\cdot)$ denotes expectation, $|\mathcal{A}|$ the cardinality of a set $\mathcal{A}$, and $\|\mathbf{a}\|$ the l^2-norm of a vector $\mathbf{a}$. $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, and $\otimes$ represents the Kronecker product.

II. System Model

We assume the downlink of a flat-fading multiuser system, where the base station is equipped with N_T antennas, serving K single-antenna users. Given a set $\mathcal{S}$ of N_T users scheduled for transmission, the corresponding data model using linear beamforming at time instant n is

$$y_k[n] = \sum_{i \in \mathcal{S}} \mathbf{h}_k^H[n]\mathbf{w}_i[n]s_i[n] + n_k[n] \quad (1)$$

where $y_k \in \mathbb{C}$ is the received symbol of user k, $\mathbf{h}_k^H \in \mathbb{C}^{1\times N_T}$ the channel vector of user k, $\mathbf{w}_i \in \mathbb{C}^{N_T \times 1}$ the beamforming vector for user i, and s_i the data symbol transmitted to user i. The noise $n_k \in \mathbb{C}$ is i.i.d., and zero mean circularly symmetric complex Gaussian distributed with variance N_0.

Although the proposed methods work for more general channel models, we assume for simplicity that the different channel vectors are i.i.d., and that the channel correlation is separable in space and time:

$$\mathbf{R}_m = E(\mathbf{h}_k[n]\mathbf{h}_k^H[n-m]) = \mathbf{R}\rho_m \quad (2)$$

where $\mathbf{R}$ is the space-correlation matrix, and ρ_m is the time-correlation function. In this work, we will mainly concentrate on the time-correlation. Exploitation of the space-correlation for limited feedback in broadcast channels has been studied in [12].

The research of the authors at TU Delft was supported in part by NWO-STW under the VIDI program (DTC.6577)
The research of the authors at Eurecom was supported in part by the Eurecom Institute, and by the national RNRT project OPUS.

The data is transmitted in a block-wise fashion. We assume a data-rate limited feedback link that can feed back B bits at the beginning of each block. Further, the feedback is assumed to be instantaneous and error-free.

We assume that the receivers have achieved perfect CSI through the use of pilots. The users then quantize the full channel, or just the channel direction if the feedback of scalars is permitted, to an element of a codebook $\mathcal{C}$, and feed back the according index to the base station. The base station then decides, based on the received feedback, which set of users to serve, and their corresponding beamforming vector.

The performance of the vector quantization (VQ) step can be improved by taking the time correlation of the channel into account. Vector quantizers with memory allow to quantize the actual channel more efficiently, i.e., the quantization error of VQ with memory is smaller than the quantization error of VQ without memory for the same amount of feedback. Even though there exists a large number of VQs with memory [13], we focus in this paper solely on predictive VQ (PVQ) since its simplicity makes it a good candidate for practical systems. It allows to exploit the correlation of the channel by considering a variable number of previous channels, without an exponential increase of the storage requirements for the codebooks as is the case for finite-state vector quantizers.

A. Linear Beamforming

The most common linear beamforming schemes are transmit matched filtering and zero-forcing (ZF) beamforming. Transmit matched filtering uses the normalized channel vector as beamforming vector. A scheme with a better performance is ZF beamforming. It provides a good tradeoff between the high-complexity schemes with good performance, e.g., DPC, and schemes like matched filtering.

The different ZF beamforming vectors are calculated based on the concatenated matrix $\hat{\mathbf{H}}$. The rows of $\hat{\mathbf{H}}$ consist of all the quantized channels $\hat{\mathbf{h}}_i^H$ of the users from the set $\mathcal{S}$. The ZF beamforming vectors are then the normalized columns of the pseudo-inverse of $\hat{\mathbf{H}}$.

B. User Selection

The optimal set of active users scheduled for transmission, denoted as $\mathcal{S}^*$, is selected to maximize the sum rate by an extensive search over all possible combinations of users

$$\mathcal{S}^* = \arg\max_{\mathcal{S}} \sum_{k\in\mathcal{S}} \log_2(1+\mathrm{SINR}_k) \tag{3}$$

where the signal-to-interference-and-noise ratio (SINR) is calculated as

$$\mathrm{SINR}_k = \frac{|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2}{\sum_{i\in\mathcal{S}, i\neq k} |\hat{\mathbf{h}}_k^H \mathbf{w}_i|^2 + N_0} \tag{4}$$

and $\hat{\mathbf{h}}_k$ is the quantized CSI known to the transmitter.

III. Predictive Vector Quantization

This section gives an overview of PVQ and its application to channel quantization of broadcast channels. For simplicity reasons, we omit the user index here.

PVQ starts by estimating the actual channel $\mathbf{h}[n]$ based on the m previously quantized channels $\hat{\mathbf{h}}[n-i], i = 1 \ldots m$, at both the base station and the users, resulting in

$$\tilde{\mathbf{h}}[n] = P(\hat{\mathbf{h}}[n-1], \hat{\mathbf{h}}[n-2], \ldots, \hat{\mathbf{h}}[n-m]) \tag{5}$$

where $P(\cdot)$ denotes the prediction function. The users, who have full CSI knowledge, then calculate the true error $\mathbf{e}[n]$ between the estimated channel $\tilde{\mathbf{h}}[n]$ and the true channel $\mathbf{h}[n]$:

$$\mathbf{e}[n] = \mathbf{h}[n] - \tilde{\mathbf{h}}[n] \tag{6}$$

The error is quantized by finding the entry in the quantization codebook $\mathcal{C}$ with the smallest Euclidean distance to the true error

$$\mathbf{e}_Q[n] = \arg\min_{\mathbf{c}\in\mathcal{C}} \|\mathbf{e}[n] - \mathbf{c}\|^2. \tag{7}$$

The quantized error $\mathbf{e}_Q[n]$ is fed back to the base station, and the quantized channel at time instant n is then computed as

$$\hat{\mathbf{h}}[n] = \tilde{\mathbf{h}}[n] + \mathbf{e}_Q[n]. \tag{8}$$

The challenge of PVQ is to design the codebook and the prediction function.

A. Codebook Design

A popular approach to design a codebook for PVQ is the open-loop approach [13]. It does not have an iterative nature, and it relies on the assumption that the quantized channels are a good approximation of the real channels. The codebook design assumes that the prediction function is known, and it uses regular VQ without memory on a training set $\mathcal{T}$, where the different elements of the training set $\mathcal{T}$ are the ideal prediction errors calculated as

$$\mathbf{e}_{\text{ideal}}[n] = \mathbf{h}[n] - P(\mathbf{h}[n-1], \mathbf{h}[n-2], \ldots, \mathbf{h}[n-m]). \tag{9}$$

The application of a memoryless VQ is possible since the prediction step in (9) removes, in the ideal case, the time correlation between the channels at different time instances.

Note that the ideal prediction error $\mathbf{e}_{\text{ideal}}[n]$ differs from the true error $\mathbf{e}[n]$ in (6). The true error is calculated as a function of the previously quantized channels, and thus depends on the quantization codebook. Using the ideal prediction error to design the codebooks removes this dependence, hence the name open-loop approach. Iterative designs, i.e., closed-loop approaches [11], only provide a minor gain.

The most common algorithm to design codebooks is the generalized Lloyd algorithm (GLA) [14]. It is a descent algorithm [13], i.e., it reduces the average distortion of the codebook with every iteration. However, the GLA is not guaranteed to find the global optimal codebook for non-convex distortion functions [15], since it may get trapped in a local minimum.

A more robust approach to find good codebooks is a Monte-Carlo based codebook design [12]. This approach generates random codebooks, estimates their performance through Monte-Carlo simulations, and finally keeps the codebook with the best performance. Even though this approach works well for small codebooks, it becomes computational expensive for larger codebooks.

The optimal design aims at finding a codebook that maximizes the overal sum rate of the system [12]. However, this design objective is computationally complex, and it depends on all the components of the system, e.g., the number of users, the selected beamforming strategy, the selection function.

To reduce the computational complexity, we focus instead on codebooks which minimize the average Euclidean distance between the ideal prediction error, and the quantized prediction error

$$\mathcal{C}^* = \arg\min_{\mathcal{C}} E(\|\mathbf{e}_{\text{ideal}}[n] - \mathbf{e}_{\text{ideal},Q}[n]\|^2) \tag{10}$$

with

$$\mathbf{e}_{\text{ideal},Q}[n] = \arg\min_{\mathbf{c}\in\mathcal{C}} \|\mathbf{e}_{\text{ideal}}[n] - \mathbf{c}\|^2. \tag{11}$$

B. Prediction Function

The other crucial part in designing the PVQ is the prediction function. A common technique for PVQ [13] is vector linear prediction [16].

Based on the previous m known channel vectors we want to predict the actual vector $\mathbf{h}[n]$ using the coefficient matrices $\mathbf{A}_j$:

$$\widetilde{\mathbf{h}}[n] = -\sum_{j=1}^{m} \mathbf{A}_j \mathbf{h}[n-j] \tag{12}$$

The goal is to minimize the average mean square prediction error. Using the orthogonality principle, the coefficient matrices can be derived from

$$\mathbf{R}_{0j} = -\sum_{\mu=1}^{m} \mathbf{A}_\mu \mathbf{R}_{\mu j} \quad j = 1, \dots, m \tag{13}$$

where $\mathbf{R}_{ij}$ is the channel correlation matrix

$$\mathbf{R}_{ij} = E(\mathbf{h}[n-i]\mathbf{h}[n-j]^H). \tag{14}$$

Stacking (13) in matrix form as

$$\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1m} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \dots & \mathbf{R}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{m1} & \mathbf{R}_{m2} & \dots & \mathbf{R}_{mm} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1^H \\ \mathbf{A}_2^H \\ \vdots \\ \mathbf{A}_m^H \end{bmatrix} = - \begin{bmatrix} \mathbf{R}_{10} \\ \mathbf{R}_{20} \\ \vdots \\ \mathbf{R}_{m0} \end{bmatrix} \tag{15}$$

the coefficient matrices $\mathbf{A}_j$ can now be found through simple matrix inversion.

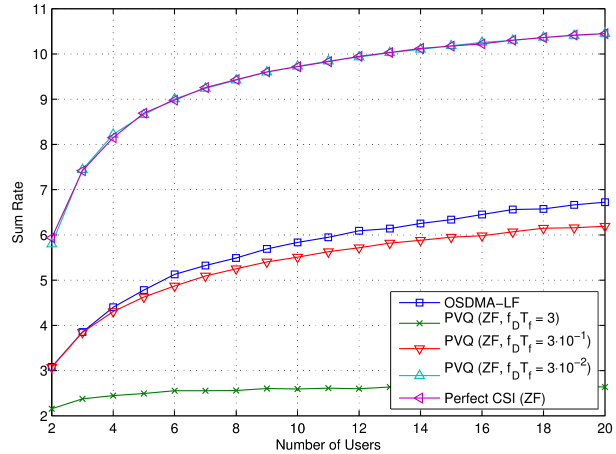

Fig. 1. The sum rate for different number of users. ($N_T = 2$, and SNR = 10 dB)

For the channel model presented in Section II, we have that $\mathbf{R}_{ij} = \mathbf{R}_{j-i} = \mathbf{R}\rho_{j-i}$. In that case, (15) becomes

$$\begin{bmatrix} \rho_0 & \rho_{-1} & \dots & \rho_{-m} \\ \rho_1 & \rho_0 & \dots & \rho_{-m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_m & \rho_{m-1} & \dots & \rho_0 \end{bmatrix} \otimes \mathbf{R} \begin{bmatrix} \mathbf{A}_1^H \\ \mathbf{A}_2^H \\ \vdots \\ \mathbf{A}_m^H \end{bmatrix} = - \begin{bmatrix} \rho_{-1} \\ \rho_{-2} \\ \vdots \\ \rho_{-m} \end{bmatrix} \otimes \mathbf{R}. \tag{16}$$

If $\mathbf{R}$ is assumed diagonal, it is clear that this equation can be solved for every channel entry separately.

IV. Simulations

Fig. 1 depicts the sum rate of PVQ with ZF beamforming, and of OSDMA-LF [10]. We assume a base station with $N_T = 2$ antennas, K users with SNR = 10 dB, and a data rate limited feedback link ($B = 3$ bits). The channel is modeled through (2) with $\mathbf{R} = \mathbf{I}_{N_T}$ and $\rho_m = J_0(2\pi f_D T_f m)$ where J_0 is the Bessel function of zeroth-order, f_D the Doppler spread, and T_f the frame length (Jakes' model [17]). Thus, we simply have to simulate different products $f_D T_f$. The algorithm predicts the actual channel based on the last $m = 3$ channels using polynomial extrapolation of order $p = m - 1$. The initial channels are assumed to be known perfectly, which can be approximated by starting the algorithm with a high-resolution memoryless VQ. In order to make a fair comparison to OSDMA-LF possible, we enforce the feedback limitation, i.e., no scalar SINR feedback is allowed. Thus, we also have to quantize the SINR feedback of the OSDMA-LF scheme. The SINR codebook is generated with the GLA using the mean squared error as distortion function. We simulate all the possible bit-distributions between SINR quantization and beamforming indexing, and finally choose the distribution which results in the highest SINR [18]. We see in Fig. 1 how the performance of PVQ with ZF improves for higher $f_D T_f$ values, i.e., for scenarios with a higher time correlation between the channels. Simulations depicting the performance

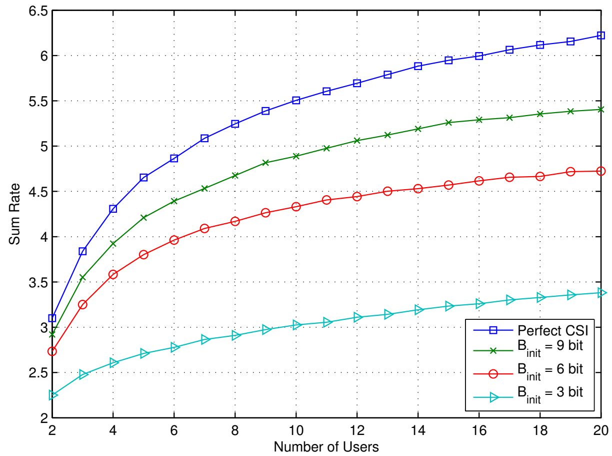

Fig. 2. Influence of the size of the initialization codebook on the sum rate. ($N_T = 2$, $T_f = 10^{-2}$ s, $f_D = 30$ Hz, $B = 3$ bits, and SNR = 10 dB)

of CSI quantization for spatially correlated channels can be found in [12].

Fig. 2 shows the influence of the initial quantization of the first m channels on the average sum rate. The plot compares the scenario where perfect CSI of the first m channels is available to scenarios where the first m channels have been quantized. We see how the sum rate increases after the first frame for larger codebooks. However, the importance of the quantization of the first m channels degrades over time, and all the schemes would converge to the same sum rate after a while.

V. Conclusions

We depicted through numerical simulations the benefits of using PVQ for time-correlated channels. PVQ uses a simple prediction step to remove the correlation between the channel to be quantized and the previous channels. This allows to improve the performance of the quantization step.

References

[1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[2] H. Bölcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen, Eds., *Space-Time Wireless Systems Space-Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge University Press, 2006.

[3] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[4] M. H. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[5] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, Sept. 2006.

[6] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[7] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[8] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[9] W. Choi, A. Forenza, J. G. Andrews, and R. W. Heath Jr., "Opportunistic space division multiple access with beam selection," accepted to IEEE Trans. on Comm.

[10] K. Huang, J. G. Andrews, and R. W. Heath Jr., "Orthogonal beamforming for SDMA downlink with limited feedback," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Apr. 2007, pp. 97–100.

[11] A. Gersho and V. Cuperman, "Vector quantization: A pattern-matching technique for speech coding," *IEEE Commun. Mag.*, vol. 21, no. 9, pp. 15–21, Dec. 1983.

[12] R. de Francisco, C. Simon, G. Leus, and D. T. Slock, "Codebook design for MIMO broadcast channels with quantized channel state information," submitted to ICC'08.

[13] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compressing*. Kluwer Academic Publishers, 1995.

[14] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[15] M. J. Sabin and R. M. Gray, "Global convergence and empirical consistency of the generalized Lloyd algorithm," *IEEE Trans. Inform. Theory*, vol. 32, no. 2, pp. 148–155, Mar. 1986.

[16] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

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

[18] M. Kountouris, R. de Francisco, D. Gesbert, D. T. Slock, and T. Sälzer, "Multiuser diversity - multiplexing tradeoff in MIMO broadcast channels with limited feedback," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Oct. 2006, pp. 364–368.