Linköping University Post Print

Combining Long-Term and Low-Rate Short-Term Channel State Information over Correlated MIMO Channels

Tùng T. Kim, Mats Bengtsson, Erik G. Larsson and Mikael Skoglund

N.B.: When citing this work, cite the original article.

Transactions Letters

Combining Long-Term and Low-Rate Short-Term Channel State Information over Correlated MIMO Channels

Tùng T. Kim, Mats Bengtsson, Erik G. Larsson, and Mikael Skoglund

Abstract—A simple structure to exploit both long-term and partial short-term channel state information at the transmitter (CSIT) over a family of correlated multiple-antenna channels is proposed. Partial short-term CSIT in the form of a weighting matrix is combined with a unitary transformation based on the long-term channel statistics. The heavily quantized feedback link is directly optimized to maximize the expected achievable rate under different power constraints, using vector quantization and convex optimization techniques on a sample channel distribution. Robustness against errors in the feedback link is also pursued with tools in channel optimized vector quantization. Simulations indicate the benefits of the proposed scheme.

Index Terms—MIMO systems, fading channels, information rates, feedback communication, adaptive systems.

I. INTRODUCTION

THE use of multiple antennas at the transmitter and the receiver is a well recognized technique to achieve high data rates in wireless communications. A multitude of different transmission techniques have been proposed in the literature, especially for the special cases of full channel state information at the transmitter (CSIT) and no CSIT, respectively. At least when using a small number of antennas, the throughput can be significantly improved if CSIT is available. However, in practice this either requires carefully calibrated radio chains and duplex times lower than the channel coherence time, if the channel reciprocity is exploited in time-division duplex systems, or that a significant bandwidth is allocated to feed back channel estimates from the receiver. This has led to a great deal of interest in low-rate feedback schemes, see for

T. T. Kim, M. Bengtsson, and M. Skoglund are with the ACCESS Linneaus Center, School of Electrical Engineering, Royal Institute of Technology, SE-10044 Stockholm, Sweden (e-mail: {tung.kim, mats.bengtsson, mikael.skoglund}@ee.kth.se).

E. Larsson was with Royal Institute of Technology. He is now with the Dept. of Electrical Engineering, Linköping University (e-mail: erik.larsson@isy.liu.se).

example [1]–[9], that can achieve a significant portion of the full-CSIT performance using only a few bits of feedback for each fading state.

The fading in wireless communications is generally governed by two components: A slowly-varying component caused by, e.g., shadowing, and a short-time variation caused by multipath-fading. Even if it is impossible to obtain accurate short-term CSIT, the long-term channel characteristics can often be estimated with good accuracy. For fixed long-term channel statistics, short-term feedback designs to maximize the ergodic capacity using vector quantization techniques are studied in [2]. The present work, by contrast, proposes a simple scheme that successfully combines both long-term and quantized short-term CSIT over a family of multiple-input multiple-output (MIMO) channels. The idea presented here is related to [4], [5] which propose to combine a codebook based on Grassmanian line packing with information from the channel covariance matrix. Our approach differs fundamentally in that it uses a mutual information criterion whereas [4], [5] minimize a bound on the CSIT error and is limited to beamforming.

Our proposed transmission scheme includes a unitary transformation influenced by the available knowledge of the channel statistics. Such a transformation can be motivated by the Karhunen-Loève transformation in vector quantization [10], and also by its optimality in the absence of short-term feedback [11]–[13]. The short-term CSIT is exploited in the form of a weighting matrix, which is designed using a modified version of the Lloyd algorithm. Unlike in [2] where approximations are required, leading to possible divergence, we show that a major step in the design procedure can be cast as a variation of the determinant maximization problem [14], which can be solved efficiently. In contrast to [4]–[6], our approach will sometimes lead to spatial multiplexing solutions. Simulation results confirm the benefits of the proposed scheme. The results also indicate that temporal power control yields little extra gain over a system that only allocates power over spatial modes. Finally, a robust design with respect to errors in the feedback link is proposed under the framework of channel optimized vector quantization.

II. SYSTEM MODEL

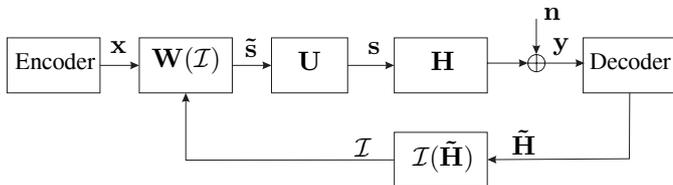Consider the discrete-time complex-baseband equivalent model of a MIMO communication system with N_t transmit

Fig. 1. System model.

antennas and N_r receive antennas. The received signal at time instant k of *block* l can be written as

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

where $\mathbf{H}_l(k)$ denotes the channel matrix and $\mathbf{s}_l(k)$ is the transmitted vector. The components of the temporally and spatially white noise $\mathbf{n}_l(k)$ are complex Gaussian with zero mean and unit variance. A block consists of N consecutive channel uses, during which the vector $\text{vec}[\mathbf{H}_l(k)]$ is assumed to be independent and identically distributed (i.i.d.) zero-mean complex Gaussian with covariance matrix $\mathbf{R}_l$, i.e., $\text{vec}[\mathbf{H}_l(k)] \sim \mathcal{CN}(0, \mathbf{R}_l)$. Herein $\text{vec}[\cdot]$ denotes vectorization. The channel covariance matrix $\mathbf{R}_l$, however, changes independently from one block to the next according to some stationary and ergodic distribution. A transmitted codeword is assumed to span a *single* block. We study the system in the limit of a very large block length N. This models a communication system where a codeword is sufficiently long to capture the ergodicity of short-term changes, but still short enough to experience a single $\mathbf{R}_l$.

For readability, we will omit the block index l and the time index k whenever this is unambiguous. Since $\mathbf{R}$ is a slowly-varying parameter, we assume that $\mathbf{R}$ is perfectly known at both sides of the link. Such information may be obtained from collected uplink measurements or using a low-rate feedback channel [15]. We further assume that $\mathbf{H}$ is fully known at the receiver. For a system with fixed long-term channel statistics, a transmitter using an "i.i.d. Gaussian" codebook and a *weighting matrix* which depends *only* on short-term feedback information is optimal in a capacity sense under certain assumptions [2], [16]. However, over a family of channels, this would require infinitely many quantization codebooks, one for each realization of $\mathbf{R}$. We therefore propose a simple alternative, illustrated in Fig. 1.

The transmitter first weighs the symbols $\mathbf{x}$, taken from a "Gaussian codebook," with $\text{E}[\mathbf{xx}^H] = \mathbf{I}_{N_t}$, by $\mathcal{W}(\mathcal{I})$, producing $\tilde{\mathbf{s}}$. The notation $[\cdot]^H$ denotes conjugate and transpose. Herein $\mathcal{W}$ is a mapping from a feedback index $\mathcal{I}$ to a finite set of weighting matrices. Such an index is obtained via a noiseless, zero-delay dedicated feedback link.[1] The weighted signals $\tilde{\mathbf{s}}$ are then linearly transformed by a unitary matrix influenced only by long-term channel statistics $\mathbf{U} \equiv \mathbf{U}(\mathbf{R})$. To produce the feedback index, the receiver employs an *index mapping* from the current *effective* channel realization $\tilde{\mathbf{H}} = \mathbf{HU}$ to an integer $\mathcal{I} \equiv \mathcal{I}(\tilde{\mathbf{H}})$. We assume that $\mathcal{I}$ takes a value in the set $\{1, \ldots, K\}$ where K is a constant

positive integer. In other words, we consider a resolution-constrained quantizer. For convenience, let $\mathbf{W}_i \triangleq \mathcal{W}(i)$, and $\mathbf{Q}_i \triangleq \mathbf{W}_i \mathbf{W}_i^H$, $i = 1, \ldots, K$. The system model (1) can then be written in the form

$$\mathbf{y} = \mathbf{HUW}_i\mathbf{x} + \mathbf{n}. \tag{2}$$

Conditioned on a feedback index $\mathcal{I} = i$, the average transmit power is

$$\text{E}\,\text{tr}(\mathbf{ss}^H) = \text{E}\,\text{tr}(\tilde{\mathbf{s}}\tilde{\mathbf{s}}^H) = \text{E}\,\text{tr}(\mathbf{W}_i\mathbf{xx}^H\mathbf{W}_i^{\,H}) = \text{tr}\,\mathbf{Q}_i,$$

where $\text{tr}\,\mathbf{X}$ denotes the trace of a matrix $\mathbf{X}$. We consider two different types of power constraints. A *short-term* power constraint requires that the transmit power does not exceed P for any feedback index:

$$\text{tr}\,\mathbf{Q}_{\mathcal{I}(\tilde{\mathbf{H}})} \le P, \quad \forall \tilde{\mathbf{H}}. \tag{3}$$

This models a system where temporal power control is not possible. Under the more relaxed *long-term* power constraint, the transmitter can vary the power over the transmission of infinitely many codewords so that

$$\text{E}_{\mathbf{R}}\,\text{E}_{\mathbf{H}}[\text{tr}\,\mathbf{Q}_{\mathcal{I}(\tilde{\mathbf{H}})}|\mathbf{R}] \le P. \tag{4}$$

Note that the distribution of $\tilde{\mathbf{H}}$ depends on the distribution of $\mathbf{R}$.

Let $I(\mathbf{R})$ denote the expected value of the mutual information between the transmitted and received signals, conditioned on $\mathbf{R}$ and for a fixed feedback scheme. We are interested in the design of a feedback scheme that maximizes the *expected rate* over infinitely many blocks, i.e.,

$$\max_{\mathcal{I}(\tilde{\mathbf{H}}),\{\mathbf{Q}_i\}} \text{E}_{\mathbf{R}}\,I(\mathbf{R}) \quad \text{s.t.} \quad \text{(3) or (4).} \tag{5}$$

The objective function in (5) can be interpreted as the achievable rate by coding over a family of information stable channels, where each member of the family is parameterized by a covariance matrix $\mathbf{R}$. In practice, the distribution of $\mathbf{R}$ has to be known beforehand. However, as will be shown in Section IV, our proposed design approach does not require the exact distribution, but only an empirical distribution of $\mathbf{R}$.

III. DECORRELATING LINEAR TRANSFORMATION

We propose to choose the unitary transformation $\mathbf{U}$ as the eigenvectors of the transmit side covariance matrix. As is shown below, this will decorrelate the channel coefficients before the quantization. We emphasize the simplicity and intuitive appeal of such a decorrelation, but do not claim its optimality, because unlike in [11]–[13], partial short-term CSIT is available in our model.

For simplicity, we begin with the case of a single receive antenna, where we use the notation $\mathbf{h} = \mathbf{H}^H$, and $\mathbf{R}^{Tx} = \mathbf{R}$. Thus the received signal can be written as $y = \mathbf{h}^H\mathbf{s} + n$ with $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{R}^{Tx})$. Introduce the eigendecomposition $\mathbf{U}^{Tx}\mathbf{D}^{Tx}(\mathbf{U}^{Tx})^H = \mathbf{R}^{Tx}$ with unitary $\mathbf{U}^{Tx}$ and diagonal $\mathbf{D}^{Tx}$. Now if we choose $\mathbf{U} = \mathbf{U}^{Tx}$, then $I(\mathbf{R}) = \text{E}_{\tilde{\mathbf{h}}|\mathbf{R}} \log(1 + \tilde{\mathbf{h}}^H\mathbf{Q}\tilde{\mathbf{h}})$, where $\tilde{\mathbf{h}} = (\mathbf{U}^{Tx})^H\mathbf{h}$ is a vector of decorrelated variables, i.e., $\tilde{\mathbf{h}} \sim \mathcal{CN}(0, \mathbf{D}^{Tx})$. This can be viewed as a property of the unitary-independent-unitary model [17]. Such

---

[1]Taking the noise in the feedback link into the design is also possible, as will be demonstrated in Section VI.

Karhunen-Loève transformations are commonly used in vector quantization [10].

The same ideas can be applied to a MIMO channel, but it is in general impossible to find a unitary precoding matrix $\mathbf{U}$ that fully decorrelates the channel coefficients. To proceed, we assume here that the second-order statistics of the channel follows the so-called Kronecker model [18], i.e., $\mathrm{vec}[\mathbf{H}] \sim \mathcal{CN}(0, \mathbf{R})$, where $\mathbf{R} = (\mathbf{R}^{\mathrm{Tx}})^{\mathrm{T}} \otimes \mathbf{R}^{\mathrm{Rx}}$. Herein $[\cdot]^{\mathrm{T}}$ denotes transpose and $\otimes$ denotes the Kronecker product. Introduce the eigendecomposition $\mathbf{U}^{\mathrm{Tx}}\mathbf{D}^{\mathrm{Tx}}(\mathbf{U}^{\mathrm{Tx}})^{\mathrm{H}} = \mathbf{R}^{\mathrm{Tx}}$ and let us choose $\mathbf{U} = \mathbf{U}^{\mathrm{Tx}}$. Recalling that $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{U}$, we then have

$$I(\mathbf{R}) = \mathrm{E}_{\tilde{\mathbf{H}}|\mathbf{R}} \log \det \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^{\mathrm{H}} \right),$$

with $\mathrm{vec}[\tilde{\mathbf{H}}] \sim \mathcal{CN}(0, \mathbf{D}^{\mathrm{Tx}} \otimes \mathbf{R}^{\mathrm{Rx}})$.

IV. FEEDBACK LINK DESIGN

A. Short-term Power Constraint

The feedback link is designed using a modified version of the Lloyd algorithm [2], [10]. However, instead of using an ad-hoc approximation that does not necessarily guarantee convergence as in [2][2], we herein exploit some results in determinant maximization [14].

We first discretize the problem (5) and consider

$$\max_{\mathcal{I}(\tilde{\mathbf{H}}),\{\mathbf{Q}_i\}} \frac{1}{|\mathcal{H}|} \sum_{\tilde{\mathbf{H}} \in \mathcal{H}} \log \det(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_{\mathcal{I}(\tilde{\mathbf{H}})}\tilde{\mathbf{H}}^{\mathrm{H}}) \quad \text{s.t.} \quad (3)$$

$$(6)$$

where $\mathcal{H}$ is a set of $|\mathcal{H}|$ samples drawn from the distribution of $\tilde{\mathbf{H}}$, which is used to approximate the continuous distribution of $\tilde{\mathbf{H}}$ [10]. The design procedure iteratively optimizes the index mapping $\mathcal{I}(\tilde{\mathbf{H}})$ and the weight codebook $\{\mathbf{Q}_i\}_{i=1}^K$. Since each optimization subproblem is solved *exactly*, the design guarantees convergence to a local optimum, but not necessarily to a global one. We summarize the two iteration steps as follows, where n indicates the iteration index.

First, given a set $\{\mathbf{Q}_i^{(n)}\}_{i=1}^K$ satisfying $\mathrm{tr}\,\mathbf{Q}_i^{(n)} \leq P, \forall i$, the optimal index mapping is given by

$$\mathcal{I}^{(n)}(\tilde{\mathbf{H}}) = \arg \max_i \log \det \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_i^{(n)}\tilde{\mathbf{H}}^{\mathrm{H}} \right).$$

Next, fix $\mathcal{I}^{(n)}(\tilde{\mathbf{H}})$ and define the quantization regions

$$\mathcal{H}_i^{(n)} \triangleq \{\tilde{\mathbf{H}} \in \mathcal{H} : \mathcal{I}^{(n)}(\tilde{\mathbf{H}}) = i\}.$$

The elements of the weight codebook can then be optimized individually:

$$\mathbf{Q}_i^{(n+1)} = \arg \max_{\mathbf{Q} \succeq 0} \frac{1}{|\mathcal{H}_i^{(n)}|} \sum_{\tilde{\mathbf{H}} \in \mathcal{H}_i^{(n)}} \log \det \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^{\mathrm{H}} \right)$$

$$\text{s.t.} \quad \mathrm{tr}\,\mathbf{Q} \leq P.$$

This convex problem is a slightly modified version of the standard determinant maximization problem. In our numerical examples, we have for simplicity solved this maximization using a direct generalization of the fixed-reduction algorithm in [14].

---

[2]We have observed numerically that the objective function does not always increase in every iteration when using [2].

B. Long-term Power Constraint

The technique outlined in Section IV-A can also be applied to the long-term power constraint case. The design however becomes more involved as we have to optimize the elements of the codebook $\{\mathbf{Q}_i\}_{i=1}^K$ jointly. Using a sample distribution $\mathcal{H}$, we can reformulate the problem as

$$\max_{\mathcal{I}(\tilde{\mathbf{H}}),\{\mathbf{Q}_i \succeq 0\}} \frac{1}{|\mathcal{H}|} \sum_{\tilde{\mathbf{H}} \in \mathcal{H}} \log \det \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_{\mathcal{I}(\tilde{\mathbf{H}})}\tilde{\mathbf{H}}^{\mathrm{H}} \right)$$

$$(7)$$

$$\text{s.t.} \quad \frac{1}{|\mathcal{H}|} \sum_{\tilde{\mathbf{H}} \in \mathcal{H}} \mathrm{tr}\,\mathbf{Q}_{\mathcal{I}(\tilde{\mathbf{H}})} \leq P.$$

We will iteratively solve the dual problem of (7). Given a fixed $\{\mathbf{Q}_i^{(n)}\}_{i=1}^K$ and a Lagrange multiplier associated with the power constraint $\lambda^{(n)}$, we assume that a constraint qualification holds so that the optimal index mapping solves

$$\max_{\mathcal{I}(\tilde{\mathbf{H}})} \left\{ \frac{1}{|\mathcal{H}|} \sum_{i=1}^K \sum_{\tilde{\mathbf{H}} \in \mathcal{R}_i} \log \det(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_i^{(n)}\tilde{\mathbf{H}}^{\mathrm{H}}) \right.$$

$$\left. - \lambda^{(n)} \sum_{i=1}^K \frac{|\mathcal{H}_i|}{|\mathcal{H}|} \mathrm{tr}\,\mathbf{Q}_i^{(n)} \right\}.$$

This can be rewritten as

$$\max_{\mathcal{I}(\tilde{\mathbf{H}})} \frac{1}{|\mathcal{H}|} \sum_{i=1}^K \sum_{\tilde{\mathbf{H}} \in \mathcal{H}_i} \left(\log \det(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_i^{(n)}\tilde{\mathbf{H}}^{\mathrm{H}}) \right.$$

$$\left. - \lambda^{(n)} \mathrm{tr}\,\mathbf{Q}_i^{(n)} \right).$$

Note that the maximization can also be seen as one performed over all possible ways of partitioning $\mathcal{H}$ into K subsets $\mathcal{H}_1, \ldots, \mathcal{H}_K$. The solution is readily given by

$$\mathcal{I}^{(n)}(\tilde{\mathbf{H}}) = \arg \max_i \left\{ \log \det \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_i^{(n)}\tilde{\mathbf{H}}^{\mathrm{H}} \right) \right.$$

$$\left. - \lambda^{(n)} \mathrm{tr}\,\mathbf{Q}_i^{(n)} \right\}.$$

In the initial step, we can select $\mathbf{Q}_i^{(0)}$ so that $\mathrm{tr}\,\mathbf{Q}_1^{(0)} = \cdots = \mathrm{tr}\,\mathbf{Q}_K^{(0)}$, to remove the dependence of $\mathcal{I}^{(0)}(\tilde{\mathbf{H}})$ on $\lambda^{(0)}$.

Next, given the quantization regions $\mathcal{H}_i^{(n)} \triangleq \{\tilde{\mathbf{H}} \in \mathcal{H} : \mathcal{I}^{(n)}(\tilde{\mathbf{H}}) = i\}$, the optimal weight codebook $\{\mathbf{Q}^{(n+1)}\}$ is the solution to

$$\max_{\{\mathbf{Q}_i \succeq 0\}} \frac{1}{|\mathcal{H}|} \sum_{i=1}^K \sum_{\tilde{\mathbf{H}} \in \mathcal{H}_i^{(n)}} \log \det \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_i\tilde{\mathbf{H}}^{\mathrm{H}} \right)$$

$$\text{s.t.} \quad \sum_{i=1}^K \frac{|\mathcal{H}_i^{(n)}|}{|\mathcal{H}|} \mathrm{tr}\,\mathbf{Q}_i \leq P.$$

We solve also this convex optimization using a barrier method with Newton steps. The optimal Lagrange multiplier can be shown to be

$$\lambda^{(n+1)} = \frac{\mathrm{tr}\sum_{\tilde{\mathbf{H}} \in \mathcal{H}_1} \mathbf{X}_1\mathbf{Q}_1}{|\mathcal{H}_1|\,\mathrm{tr}\,\mathbf{Q}_1} = \cdots = \frac{\mathrm{tr}\sum_{\tilde{\mathbf{H}} \in \mathcal{H}_K} \mathbf{X}_K\mathbf{Q}_K}{|\mathcal{H}_K|\,\mathrm{tr}\,\mathbf{Q}_K},$$

where $\mathbf{X}_i \equiv \mathbf{X}_i(\tilde{\mathbf{H}}) \triangleq \tilde{\mathbf{H}}^{\mathrm{H}} \left(\mathbf{I}_{N_\mathrm{r}} + \tilde{\mathbf{H}}\mathbf{Q}_i\tilde{\mathbf{H}}^{\mathrm{H}} \right)^{-1} \tilde{\mathbf{H}}.$
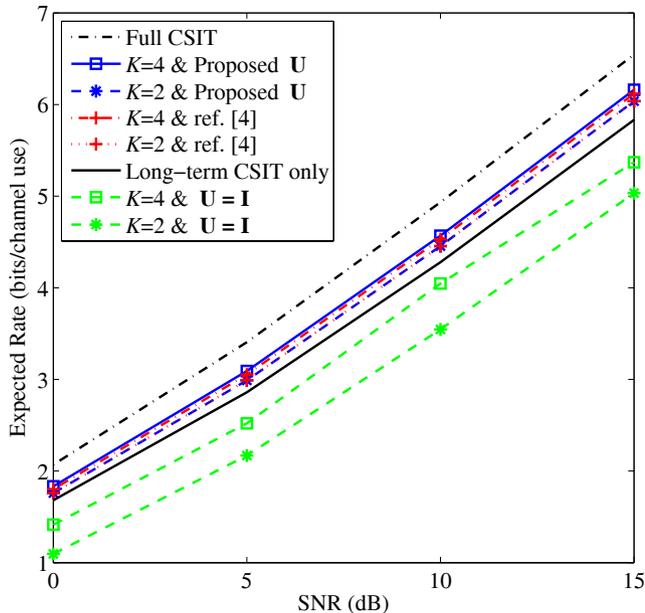
Fig. 2. Performance of the proposed scheme over a family of 4×1 channels. A short-term power constraint is assumed.

Fig. 3. Performance of the proposed scheme over a family of 4×2 channels. A short-term power constraint is assumed.

Fig. 4. Performance of the proposed scheme over a family of 2×1 channels under different power constraints.

V. NUMERICAL RESULTS

In this section we present simulation results for a specific class of correlated channels. For simplicity, the family of $\mathbf{R}^{\text{Tx}}$ is taken as Toeplitz matrices with $[1, \rho, \rho^2, \ldots, \rho^{N_t - 1}]$ as the top row, where ρ is a complex-valued random variable with phase uniformly distributed in $[0, 2\pi]$ and modulus distributed as $f(|\rho|) = C \exp\left(-\lambda(1 - |\rho|)\right)$ if $|\rho| < 1$ and $f(|\rho|) = 0$ otherwise, where $\lambda > 0$ and $C = \lambda/(1 - \exp(-\lambda))$ is a normalization factor. The goal is to study the general behaviors of the systems when the channels vary from highly correlated to fully uncorrelated, rather than to simulate a specific scenario. All the simulations are obtained with $\lambda = -\log(0.01)$, modeling a family of channels where $|\rho|$ is greater than 0.5 more than 90 percent of the time. A design for practical applications may rely on more realistic channel models. Since the noise variance is normalized to unity, we define the signal-to-noise ratio as SNR $\stackrel{\triangle}{=} P$. The feedback link is trained with 100 realizations of $\mathbf{R}$ and 1000 channel realizations for each $\mathbf{R}$, using 20 random starting points and 15 iterations in the Lloyd algorithm. Further increasing the number of channel realizations does not seem to change the performance.

The performance of the proposed scheme under a short-term power constraint over a family of 4×1 channels is plotted in Fig. 2. For comparison, we tried a system that only uses short-term feedback but ignores the correlation properties of the channel, using exactly the same vector quantization technique except that $\mathbf{U} = \mathbf{I}_{N_t}$. Clearly, the long-term information provides a consistent gain. For example, with $K = 2$ or one bit of feedback, the proposed approach provides nearly a 3 dB gain at an expected rate of 4 bits per channel use. We also plot the performance of the CSI quantization scheme proposed in [4], combined with beamforming. Interestingly enough, this scheme performs almost identically to our proposed scheme, even though it is not specifically optimized to maximize the
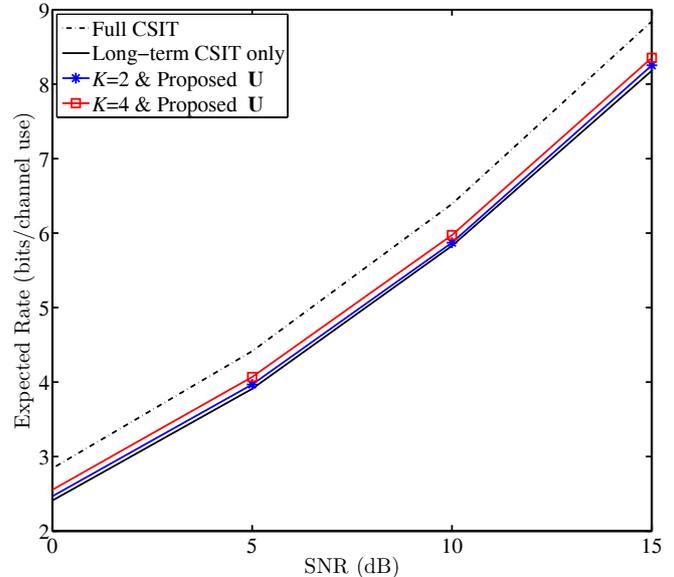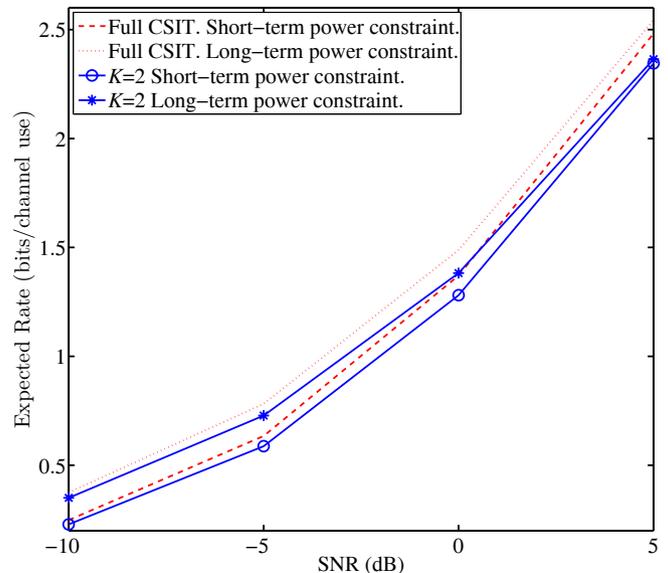
expected rate and always will correspond to $\mathbf{Q}$ matrices of rank 1.[3] On the other hand, [4] exploits some information also from the eigenvalues of $\mathbf{R}$, not only the eigenvectors. Finally, for comparison, the figure shows the performance of using full CSIT (i.e. beamforming) and of using only long-term CSIT [11]–[13], [19], where the transmit covariance matrix is optimized for each $\mathbf{R}$, i.e., for each fading block, using sample distributions. At an expected rate of 4 bits per channel use, combining 2 bits of short-term CSIT ($K = 4$) with long-term statistics yields a gain of roughly 1 dB over using only long-term statistics. A similar behavior, but with less pronounced gains, can be seen in the 4×2 MIMO case, as illustrated in

---

[3]For MISO channels, our scheme has always resulted in $\mathbf{Q}$ matrices of rank 1, but for MIMO most $\mathbf{Q}$ matrices were high rank.

Fig. 3. Other experiments (not reported in the figures) have shown that the resulting design is extremely insensitive to the SNR and also to the assumed channel statistical distribution. Also, we have tried to exploit knowledge of the eigenvalues of **R** to select between different code books, but the additional gain was extremely small.

Fig. 4 compares the performance of the proposed scheme under different power constraints over a family of 2×1 channels. The results indicate that temporal power control provides a negligible gain for moderate and high SNRs, consistent with some results under the assumption of perfect CSIT [20]. At low SNR, however, a long-term power-constrained system outperforms a short-term one by a wide margin. For instance, at SNR $= -5$ dB, a long-term power controlled system using one bit of feedback information even outperforms a perfect-CSIT system without temporal power control. However, the validity of the assumption about perfect channel knowledge at the receiver at so low SNR-values may be questioned.

VI. ROBUST DESIGN TO ERRONEOUS FEEDBACK LINKS

Noise and delays in the feedback link may lead to erroneous feedback, i.e., the feedback index received at the transmitter is not identical to the one sent by the receiver. In this section, we demonstrate how such defects in the feedback link can be explicitly taken into account in the system design. We exclusively focus on the short-term power constraint case. The design under a long-term power constraint problem can be handled under a similar principle, but does not necessarily give any additional insight into the system behavior. The key tool in our robust design is channel optimized vector quantization (COVQ) [7], [21].

To distinguish the feedback index from the one actually seen at the transmitter, let us denote $\mathcal{J}(\tilde{\mathbf{H}})$ as the index mapping used by the receiver that takes the erroneous feedback link into account. Thus, upon knowing $\tilde{\mathbf{H}}$, the receiver sends back $j = \mathcal{J}(\tilde{\mathbf{H}}) \in \{1, \ldots, K\}$, and the transmitter receives some index $i \in \{1, \ldots, K\}$, potentially different from j. We model the feedback link as a discrete-input discrete-output memoryless channel with transition probabilities $p(i|j)$. In practice, the values of the transition probabilities may need to be estimated based on e.g., the SNR of the feedback link. Note that even if the index is not correctly received, the effective channel matrix is still assumed to be perfectly tracked at the receiver. That is, the errors in the feedback link only affect the weighting matrix used at the transmitter.

The design problem can be reformulated as

$$\max_{\mathcal{J}(\tilde{\mathbf{H}}), \{\mathbf{Q}_i\}} \frac{1}{|\mathcal{H}|} \sum_{\tilde{\mathbf{H}}} \sum_{i=1}^{K} p(i|\mathcal{J}(\tilde{\mathbf{H}})) \log \det \left(\mathbf{I}_{N_r} + \tilde{\mathbf{H}} \mathbf{Q}_i \tilde{\mathbf{H}}^H \right) \tag{8}$$

where we again approximate the true distribution of $\tilde{\mathbf{H}}$ with a sample distribution. An iterative procedure, which essentially follows the methodologies in Section IV with some slight modifications, can be applied to the extended design problem (8). The iterative steps are described in the following, where we omit the iteration index n to improve readability.

Given the covariance matrices $\{\mathbf{Q}_i\}_{i=1}^{K}$, the optimal index
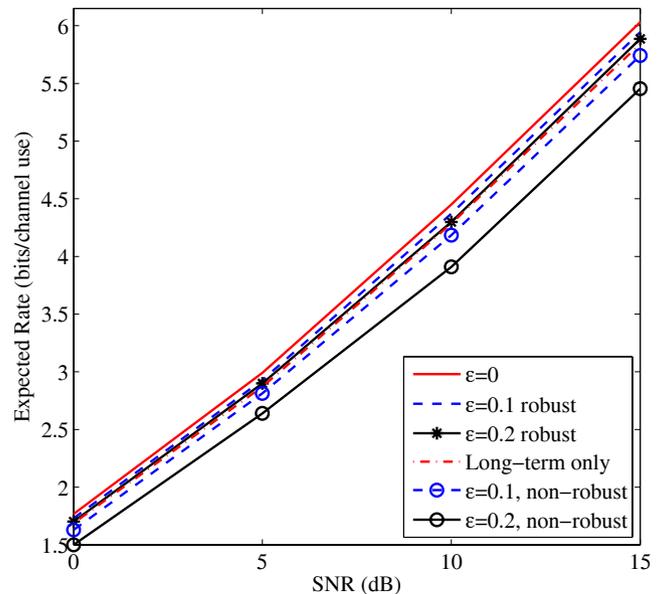

Fig. 5. Performance of the proposed schemes designed taking into account error probability in the feedback link over a family of 4×1 channels. The feedback resolution is $K = 2$.

mapping is given by

$$\mathcal{J}(\tilde{\mathbf{H}}) = \arg \max_{j \in \{1, \ldots, K\}} \sum_{i=1}^{K} p(i|j) \log \det \left(\mathbf{I}_{N_r} + \tilde{\mathbf{H}} \mathbf{Q}_i \tilde{\mathbf{H}}^H \right).$$

That is, the optimal index mapping also takes into account the possibilities of different outcomes of the random feedback link. Next, given the index mapping $\mathcal{J}(\tilde{\mathbf{H}})$, then for each value of i, the optimal transmit covariance matrix $\mathbf{Q}_i$ solves the following problem

$$\max_{\mathbf{Q} \succeq 0} \sum_{j=1}^{K} p(i|j) \sum_{\tilde{\mathbf{H}} \in \mathcal{H}_j} \log \det \left(\mathbf{I}_{N_r} + \tilde{\mathbf{H}} \mathbf{Q} \tilde{\mathbf{H}}^H \right)$$
$$\text{s.t.} \quad \text{tr} \, \mathbf{Q} \leq P,$$

where we define the quantization region j as $\mathcal{H}_j = \{\tilde{\mathbf{H}} \in \mathcal{H} : \mathcal{J}(\tilde{\mathbf{H}}) = j\}$. Clearly, introducing the weighting factors $p(i|j)$ does not change the concavity of the cost function; thus the optimization can be solved numerically for the global optimum.

We plot the performance of the robust design over a family of 4×1 channels (generated as described in Section V) in Fig. 5. In this example, the feedback link is modeled as a K-input K-output memoryless channel with transition probabilities $p(i|j) = 1 - \epsilon$ if $i = j$ and $p(i|j) = \epsilon/(K-1)$ if $i \neq j$, i.e., the error probability in the feedback link is ϵ and the errors are uniformly distributed over all possible erroneous outcomes. A finer error model on the *bit* level can also be used. As can be seen, the robust design successfully takes into account the errors in the feedback link and strictly improves the performance compared to that obtained with only long-term statistics, even if the error probability in the feedback link is relatively high (up to $\epsilon = 0.2$). This of course comes at the price of a higher complexity in the design. The curves marked by circles are the ones obtained by directly using an

error-free codebook over a noisy feedback link. Notice that in this case, not using a robust codebook even gives worse performance than relying on long-term statistics only. Such a sensitivity to error in the feedback link is somewhat reduced at higher values of the feedback level K (not plotted herein for readability).

VII. CONCLUSION

We have presented a simple transceiver structure that successfully combines short-term, fast feedback based on actual channel realizations and long-term, slowly-varying CSI containing the second-order statistics of the MIMO channel. While the proposed structure is not claimed to be optimal, we emphasize its simplicity and versatility, as well as its excellent performance. We have also studied an important extension from the basic setup, which allows the design to take into account errors in the feedback link.

REFERENCES

[1] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[2] V. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, pp. 62–70, Jan. 2004.

[3] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2562–2579, Oct. 2003.

[4] D. J. Love and R. W. Heath, Jr., "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 718–722, Mar. 2006.

[5] D. J. Love and R. W. Heath Jr, "Grassmannian beamforming on correlated MIMO channels," in *Proc. GLOBECOM*, vol. 1, pp. 106–110, 2004.

[6] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2735–2747, Oct. 2003.

[7] G. Jöngren and M. Skoglund, "Quantized feedback information in orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2473–2486, Oct. 2004.

[8] A. D. Dabbagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. Inform. Theory*, vol. 52, pp. 2190–2202, May 2006.

[9] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: a VQ-based approach," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1101–1112, Mar. 2006.

[10] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic, 1992.

[11] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[12] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[13] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sept. 2004.

[14] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 2, pp. 499–533, 1998.

[15] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed. CRC Press, Aug. 2001, ch. 18.

[16] M. Skoglund and G. Jöngren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 395–405, Apr. 2003.

[17] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[18] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[19] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.

[20] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[21] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 799–809, July 1990.