Design and Analysis of Transmit-Beamforming based on Limited-Rate Feedback

Pengfei Xia, *Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—This paper deals with design and performance analysis of transmit beamformers for multiple-input multiple-output (MIMO) systems based on bandwidth-limited information that is fed back from the receiver to the transmitter. By casting the design of transmit beamforming based on limited-rate feedback as an equivalent sphere vector quantization (SVQ) problem, multiantenna beamformed transmissions through independent and identically distributed (i.i.d.) Rayleigh fading channels are first considered. The rate-distortion function of the vector source is upper-bounded, and the operational rate-distortion performance achieved by the generalized Lloyd's algorithm is lower-bounded. Although different in nature, the two bounds yield asymptotically equivalent performance analysis results. The average signal-to-noise ratio (SNR) performance is also quantified. Finally, beamformer codebook designs are studied for correlated Rayleigh fading channels, and a low-complexity codebook design that achieves near-optimal performance is derived.**

***Index Terms*—Beamforming, generalized Lloyd algorithm, Grassmannian line packing, limited rate feedback, multiple-input multiple-output (MIMO), partial channel state information, rate-distortion function, vector quantization, Welch bound.**

I. INTRODUCTION

TRANSMIT beamforming can improve considerably the performance of multiple-input multiple-output (MIMO) systems [6], [17], [20]. Channel state information (CSI), however, has to be furnished to the transmitter in order to enable the beamforming operation. The question is what kind of CSI can be made practically available to the transmitter in a constantly changing wireless setting, especially for frequency-division duplex (FDD) systems where the downlink and uplink channels are not reciprocal. The various system imperfections that are present include channel estimation errors, feedback delay, as well as limited feedback bandwidth, and give rise to partial (imperfect) CSI at the transmitter (CSIT). Certain CSI imperfections, such as feedback delay and estimation errors, can be captured by a statistical channel mean information model, out of which the beamforming vector is derived (see [10], [20], [25], and [29] and the references therein).

Another important CSI imperfection in practice is the bandwidth constraint over the feedback link, which conveys to the transmitter only finite bits per fading block. For such cases, the transmitter and the receiver need to maintain a common beamformer codebook, i.e., a finite collection of beamforming vectors (codewords). For each received codeword index, the transmitter chooses the corresponding beamforming vector for data transmission. Codebook design criteria include maximizing the average signal-to-noise ratio (SNR) at the maximum ratio combining (MRC) output [16], [20], maximizing the average mutual information [15], or minimizing the outage probability [19]. In general, the codebook design can be viewed as a vector quantization (VQ) problem, and the generalized Lloyd algorithm can be employed to actually construct the codebook. Specifically for independent and identically distributed (i.i.d.) Rayleigh fading channels, it can be readily shown that designing beamformer codebooks reduces to a sphere vector quantization (SVQ) problem, where the codewords are vectors constrained on the unit hypersphere and the vector source input is uniformly distributed on the unit hypersphere. Different from the conventional Euclidean distance, special to this SVQ problem is the distortion metric, which is a projective distance from the source input vector to the codeword vector. Besides the SVQ approach, achieving the maximum Welch bound (or, equivalently, Grassmannian line packing) has also been pursued recently in an effort to obtain near-optimal codebook designs [16], [19], [27] for i.i.d. Rayleigh fading channels. Specifically, [27] provides an analytic codebook construction method, which leads to constant modulus transmissions as well as low-complexity quantization enabled by the fast Fourier transform (FFT) operation.

A main issue in source coding and quantization has been the optimal tradeoff between distortion and rate (number of bits used to describe the random source), "both in theory and in actual codes" [11]. The study of such optimal tradeoffs for the special SVQ problem arising in the design of transmit beamformers under limited-rate feedback constitutes the major goal of this paper. We are interested in theoretical performance limits of this SVQ problem, i.e., the rate-distortion function of the vector source input that specifies the ultimate limit on the minimum number of bits required to achieve a certain distortion level. Also to our interest is the operational rate-distortion performance, i.e., number of feedback bits versus average SNR

Manuscript received July 18, 2004; revised March 16, 2005. Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. This work was also supported by the NSF Grant no. 01-0516. Part of this work appeared in IEEE Vehicular Technology Conference 2004-Fall, Los Angeles, CA, September 26–29, 2004. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Steven L. Grant.

P. Xia was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with Samsung Electronics, San Jose, CA 95134 USA (e-mail: Pengfei.xia@samsung.com).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TSP.2006.871967

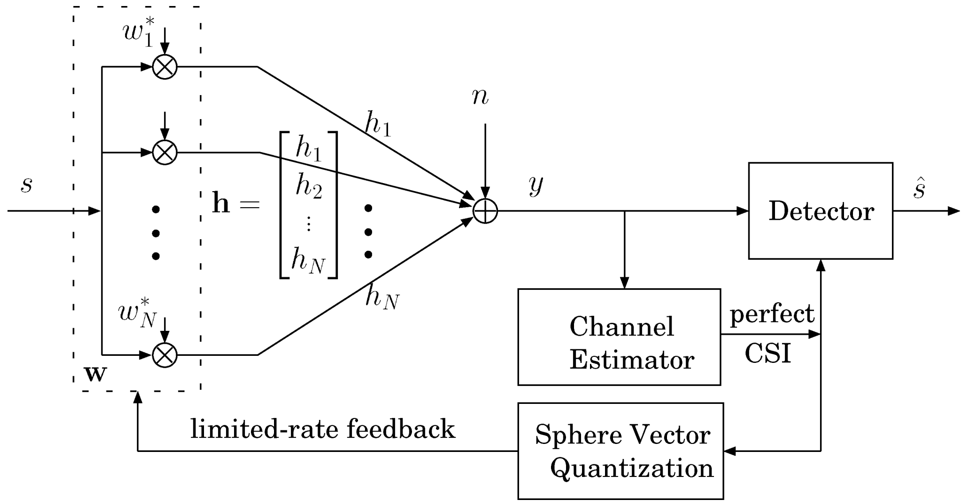

Fig. 1. System model.

degradation, achieved by the generalized Lloyd algorithm applied to our problem at hand. As usual, a certain performance gap exists between the theoretical performance and the operational performance, since the former generally requires infinitely long sequence-by-sequence quantization, while the latter is achieved via the simplest vector-by-vector quantization.[1] It is thus of interest to investigate the relationship between theoretical and operational performance.

Regarding performance analysis of the SVQ problem, the following are the contributions of this paper .

1) Relying on rate-distortion theory tools, we upper-bound the rate-distortion function of the vector source input, using as a distortion metric a suitable projective distance (Section III-B). We also explore the bound-achieving conditional distribution.
2) Using the bounding technique developed in [19] and [30], we lower-bound the operational rate-distortion performance, and thus the average SNR degradation, achieved by the generalized Lloyd algorithm (Section III-C). Numerical examples confirm that this lower bound is universally tight, implying that it also serves as a very close performance approximation.
3) We also quantify the average SNR performance achieved by the so-designed beamformer codebook over i.i.d. Rayleigh fading channels.

Investigating the SVQ problem related to transmit beamforming from the rate-distortion function perspective is novel and useful. On the other hand, the lower-bounding technique in Section III-C is similar to that used initially by [19] to lower-bound the outage probability for delay-sensitive applications, and later by [30] to lower-bound the symbol error probability for phase-shift keying (PSK)/quadrature amplitude modulation (QAM) constellations. We find that using SNR degradation as performance metric allows one to quantify the achieved SNR performance, and is thus attractive for most applications. Interestingly as we will see, although different in nature, the two bounds yield asymptotically equivalent performance analysis results. Finally, we study codebook designs for correlated Rayleigh fading channels. Based on a proper codebook designed for i.i.d. Rayleigh fading channels, we develop a beamformer codebook that achieves near-optimal performance in the correlated scenario, while enjoying very low design complexity. Throughout the paper, we adopt the following notational conventions.

Notation:: Bold upper and lower case letters denote matrices and column vectors, respectively; $(\cdot)^T$ and $(\cdot)^{\mathcal{H}}$ denote transpose and Hermitian transpose, respectively; $\mathbb{E}[\cdot]$ denotes the ensemble average; $\mathcal{C}^N$ denotes the complex N-dimensional space; Ω^N denotes the unit hypersphere in the complex N-dimensional space; $\mathcal{U}(\cdot)$ denotes the uniform distribution; and $\mathcal{CN}(\mathbf{a}, \mathbf{\Sigma})$ denotes a complex Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{\Sigma}$.

II. System Model and Sphere Vector Quantization Problems

We will deal with i.i.d. Rayleigh fading channels in this section and the next, and consider a wireless communication system with multiple (N) transmit antennas and a single receive antenna, as depicted in Fig. 1. Beamformer codebooks designed for such multiple-input single-output (MISO) systems can be directly applied to i.i.d. Rayleigh MIMO channels without loss of optimality [16]. Also, as we will show in Section III-E, performance analysis of MISO systems with optimal transmit beamforming can be used for MIMO systems with only minor modifications. In short, focusing on MISO systems incurs no loss of generality.

Our focus will be on narrowband block transmissions where the wireless channel is frequency nonselective. Let h_k denote the channel coefficient between the kth transmit and the single receive antenna, and $\mathbf{h} := [h_1, h_2, \ldots, h_N]^T$, where N is the number of transmit antennas. The channel coefficients h_k are uncorrelated and zero mean circularly complex Gaussian; i.e., $h_k \sim \mathcal{CN}(0, \sigma_h^2)$, $\forall k$. This assumption corresponds to a rich scattering environment and transmit antennas placed sufficiently far apart from each other. We further consider channel realizations that do not vary within a block but can change from block to block, which corresponds to the so-called block fading channel model that has been adopted by many

[1]Here, a vector is the smallest unit to perform sphere VQ under the special distortion metric.

practical wireless systems [24]. We also suppose that CSI is perfectly available at the receiver, which allows one to isolate the partial CSI effects due to the limited-rate feedback [20]. As in [15], [16], [19], and [20], the feedback channel is assumed to be error and delay free, but limited in bandwidth, thus conveying to the transmitter only B feedback bits per block. The error-free assumption can be well approximated through the use of sufficiently powerful error control codes over the feedback link, whereas the delay-free assumption is accurate when the processing and feedback delays are small relative to the channel's coherence time.

Corresponding to each B-tuple, the transmitter selects a unit-norm beamformer (steering vector) $\mathbf{w}$ and transmits through the multiple antennas the information symbol s with average energy E_s. The received symbol in noise is thus expressed as

$$y = \mathbf{w}^{\mathcal{H}}\mathbf{h}s + n \tag{1}$$

where n is additive white Gaussian noise (AWGN) with zero mean and variance $N_0/2$ per real and imaginary dimension. Without loss of generality, we set $E_s/N_0 = 1$ throughout the paper.

Both transmitter and receiver maintain a common codebook $\mathcal{W}$ comprising $M := 2^B$ beamforming vectors (codewords) $\mathcal{W} := \{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$. Following ideal channel estimation, the receiver chooses the best (in the maximum *instantaneous* SNR sense) beamforming vector

$$\begin{aligned}\mathbf{w}_* &= \arg\max_{\mathbf{w}\in\mathcal{W}} |\mathbf{w}^{\mathcal{H}}\mathbf{h}|^2 \\ &= \arg\min_{\mathbf{w}\in\mathcal{W}} \left(|\mathbf{h}|^2 - |\mathbf{w}^{\mathcal{H}}\mathbf{h}|^2\right)\end{aligned} \tag{2}$$

and feeds back the B-tuple of the corresponding codeword index. By matching the steering vector with a quantized channel codeword, transmit beamforming is implemented using only partial CSI obtained through the limited-rate feedback channel.

The design task now becomes finding an appropriate codebook $\mathcal{W}$ in order to optimize a certain objective. Maximizing the *average* receive SNR, $\mathbb{E}[|\mathbf{w}_*^{\mathcal{H}}\mathbf{h}|^2]$ has been widely accepted [16], [20] and will also be adopted in this paper. To facilitate performance analysis, we equivalently minimize

$$\mathbb{E}\left[|\mathbf{h}|^2 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{h}\right|^2\right] \tag{3}$$

which expresses the average SNR degradation due to quantization, while $\mathbb{E}[|\mathbf{h}|^2]$ is the achievable average SNR in presence of perfect CSIT. Our codebook design can thus be formally stated as

$$\begin{aligned}&\min_{\mathcal{W}} \mathbb{E}\left[|\mathbf{h}|^2 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{h}\right|^2\right] \\ &\text{subject to } \mathbf{w}_i^{\mathcal{H}}\mathbf{w}_i = 1, \qquad \forall i.\end{aligned} \tag{4}$$

Codebook design is straightforward when $M \leq N$, and the optimal codebook $\mathcal{W}$ can be chosen as any M orthogonal columns of an arbitrary $N \times N$ unitary matrix [16], [19]. In this paper, we focus on the more difficult $M > N$ case, which can be viewed as a VQ problem with

$$\begin{aligned}&\text{source input}: && \mathbf{h} \sim \mathcal{CN}\left(0, \sigma_h^2\mathbf{I}_N\right) \\ &\text{codebook}: && \mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_M\}, |\mathbf{w}_i| = 1, \quad \forall i \\ &\text{distortion metric}: && d_1(\mathbf{h}, \mathbf{w}_i) := |\mathbf{h}|^2 - \left|\mathbf{w}_i^{\mathcal{H}}\mathbf{h}\right|^2, \quad \forall \mathbf{h}, i.\end{aligned} \tag{5}$$

Notice that this distortion measure is valid for $N \geq 2$ only, since for $N = 1$, the distance is trivially zero. To cast the problem (5) in a more convenient form, consider the normalized version of the average SNR degradation in (3)

$$\frac{\mathbb{E}\left[|\mathbf{h}|^2 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{h}\right|^2\right]}{\mathbb{E}\left[|\mathbf{h}|^2\right]} = \mathbb{E}\left[1 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{g}\right|^2\right] \tag{6}$$

where $\mathbf{g} := \mathbf{h}/|\mathbf{h}|$ is the direction of $\mathbf{h}$, and the equality in (6) is due to the fact that with $\mathbf{h}$ complex i.i.d. Rayleigh distributed, $\mathbf{g}$ and $|\mathbf{h}|$ are independent [20]. Since $\mathbb{E}[|\mathbf{h}|^2] = N\sigma_h^2$ is fixed, one can equivalently maximize (6), instead of optimizing the average SNR degradation in (3). In accordance with this new objective function, the original VQ interpretation in (5) becomes

$$\begin{aligned}&\text{source input}: && \mathbf{g} \sim \mathcal{U}(\Omega^N) \\ &\text{codebook}: && \mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_M\}, |\mathbf{w}_i| = 1, \quad \forall i \\ &\text{distortion metric}: && d_1(\mathbf{g}, \mathbf{w}_i) = 1 - \left|\mathbf{w}_i^{\mathcal{H}}\mathbf{g}\right|^2, \quad \forall \mathbf{g}, i\end{aligned} \tag{7}$$

i.e., the source input is replaced by $\mathbf{g}$, while the distortion measure becomes $d_1(\mathbf{g}, \mathbf{w}_i)$. Notice that the codewords $\{\mathbf{w}_i\}$ are vectors constrained on the unit hypersphere Ω^N, while the source input $\mathbf{g}$ is uniformly distributed on Ω^N [20]. The problem formulations in (5) and (7) constitute what are known as SVQ problems. Unique to the SVQ problem under consideration is the distortion metric, which is a projective distance from the source input vector to the codeword vector, instead of the Euclidean distance adopted by conventional codebook designs in source coding.

III. Algorithm and Performance Analysis for i.i.d. Rayleigh Fading Channels

In this section, we first briefly outline the generalized Lloyd algorithm that solves our SVQ problem at hand. Subsequently, we derive an upper bound to the rate-distortion function of the vector source input (Section III-B), along with a lower bound to the operational rate-distortion performance achieved by the generalized Lloyd algorithm (Section III-C). We further identify a simple and valuable relationship between the theoretical performance and the operational performance and also quantify the achievable average SNR performance, which offers a useful guideline for practical system designs.

A. Generalized Lloyd Algorithm

Associated with every quantizer is a partition of the input space $\mathcal{C}^N$ into M regions $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_M$, with $\mathcal{A}_i$ denoting the neighborhood (or cluster region) of the codeword $\mathbf{w}_i$. Designing a quantizer amounts to finding a codebook and a partition rule that jointly minimize the overall average distortion measure, which is exactly the average SNR degradation in (3). Two necessary conditions prove to be essential for the quantizer design [8]. First, necessary to the optimality of the codebook, is the so-called *centroid condition*, which decrees that for each region, the optimal codeword should be chosen to minimize the distortion measure averaged over that region, or the local average distortion. Take the ith region $\mathcal{A}_i$, for example, whose local channel correlation matrix is $\mathbf{R}_i := \mathbb{E}[\mathbf{h}\mathbf{h}^{\mathcal{H}}|\mathbf{h} \in \mathcal{A}_i]$. The optimal beamforming vector $\mathbf{w}_i^{\mathrm{opt}}$ should maximize $\mathbf{w}_i^{\mathcal{H}}\mathbf{R}_i\mathbf{w}_i$, subject to the unit norm constraint. Therefore

$$\mathbf{w}_i^{\mathrm{opt}} = \arg \max_{\mathbf{w}_i^{\mathcal{H}}\mathbf{w}_i=1} \mathbf{w}_i^{\mathcal{H}}\mathbf{R}_i\mathbf{w}_i = \mathbf{u}_i \quad (8)$$

where $\mathbf{u}_i$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}_i$. In other words, for each region $\mathcal{A}_i$, the optimal beam points along the principal eigenvector of $\mathbf{R}_i$, which is also known as eigen beamforming [10], [20], [25], [29]. Second, necessary to the optimality of the channel space partition is the *nearest neighbor rule*

$$\mathbf{h} \in \mathcal{A}_i \quad \text{iff} \quad d_1(\mathbf{h}, \mathbf{w}_i) \le d_1(\mathbf{h}, \mathbf{w}_j), \quad \forall i, j = 1, \ldots, N \quad (9)$$

which dictates that all input vectors that are closer to the codeword $\mathbf{w}_i$ than to any other codeword be assigned to the neighborhood of $\mathbf{w}_i$ or region $\mathcal{A}_i$. The generalized Lloyd algorithm repeatedly examines the two necessary conditions to find the optimal codebook and the channel space partition.

Generalized Lloyd Algorithm [8]

s1) Initialize with any valid codebook.
s2) For the given codebook, use the *nearest neighbor rule* (9) to find the optimal regions.
s3) For the given regions, use the *centroid condition* (8) to determine the optimal codewords.
s4) Loop back to s2) until convergence.

Due to the centroid condition and the nearest neighbor rule, the overall average distortion monotonically decreases, or at least does not increase after each step. Normally it takes only five to eight iterations for the algorithm to converge. For a detailed discussion of practical issues about the algorithm, including global/local optimality properties of the codebook, and training-sequence-based implementation, the reader is referred to [8, Ch. 11]. We emphasize that the generalized Lloyd algorithm is applicable to arbitrary fading channels. In contrast, alternative design approaches to be discussed in Section III-D generally work for i.i.d. Rayleigh fading channels only. We will present a low-complexity codebook design method for correlated fading channels in Section IV.

B. Upper Bounding the Information Rate-Distortion Function

Describing any continuous-valued random source input $\mathbf{h}$ using only finite bits entails distortion. Given the source distribution and a proper distortion measure, the rate-distortion function of $\mathbf{h}$ specifies the ultimate limit on the minimum number of bits required to achieve a certain distortion level. Naturally, this function limit depends on the source input statistics and the distortion measure. For different distortion measures, the function limits are different even for the same source input. In this subsection, we derive an upper bound of this function limit.

Recall that the codeword $\mathbf{w}_i$ is constrained on the unit hypersphere Ω^N. Although this unit norm constraint is necessary to fix the transmit energy, it complicates the analysis of the rate-distortion function. Fortunately, as far as the rate-distortion performance is concerned, this constraint can be relaxed, provided that a proper distortion metric can be devised to be scale invariant. Specifically, for any codebook $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$, we can define a scaled version of it

$$\mathcal{V} := \{\mathbf{v}_1 := \alpha_1\mathbf{w}_1, \ldots, \mathbf{v}_M := \alpha_M\mathbf{w}_M\} \quad (10)$$

where $\alpha_1, \ldots, \alpha_M > 0$ are arbitrary positive real scalars. It is easy to show that the distortion metric $d_2(\mathbf{h}, \mathbf{v}_i) := |\mathbf{h}|^2 - |\mathbf{v}_i^{\mathcal{H}}\mathbf{h}|^2/|\mathbf{v}_i|^2$ is scale invariant

$$d_2(\mathbf{h}, \mathbf{v}_i) = d_1(\mathbf{h}, \mathbf{w}_i), \ \forall i, \quad (11)$$

for any codebook $\mathcal{W}$, source input realization $\mathbf{h}$, and arbitrary positive real numbers $\alpha_1, \ldots, \alpha_M$. Designing $\mathcal{W}$ under the unit-norm constraint is equivalent to designing $\mathcal{V}$ without any codeword constraint in the sense that they both provide identical rate-distortion performance.

In a nutshell, for the SVQ problem under consideration, we can safely remove the unit norm constraint and use $d_2(\mathbf{h}, \mathbf{v}_i)$ as the distortion metric describing the projective distance from $\mathbf{h}$ to the unconstrained codeword $\mathbf{v}_i$. The rate-distortion function under this distortion metric, is defined as [5], [7]

$$\begin{aligned} R(D) &:= \inf_{p(\mathbf{x}|\mathbf{h})} \mathcal{I}(\mathbf{h}; \mathbf{x}) \\ \text{s.t.} \quad & \mathbb{E}_{p(\mathbf{h})p(\mathbf{x}|\mathbf{h})}\left[d_2(\mathbf{h}, \mathbf{x})\right] \le D \end{aligned} \quad (12)$$

where $p(\mathbf{x}|\mathbf{h})$ is any conditional distribution of $\mathbf{x}$ given $\mathbf{h}$ over which the infimum is taken, $p(\mathbf{h})$ is the known input distribution (zero mean i.i.d. complex Gaussian here), and $\mathcal{I}(\mathbf{h}; \mathbf{x})$ is the mutual information between $\mathbf{h}$ and $\mathbf{x}$. Naturally, if Euclidean distance is used as the distortion measure, the corresponding rate-distortion function would be different from the so-defined $R(D)$ in (12).

Because it is difficult to find the optimal conditional distribution $p(\mathbf{x}|\mathbf{h})$ and the rate-distortion function, we next derive an upper bound of it. Such an upper bound, together with the bound-achieving test channels, will prove to be very useful, as they bear a close relationship with the operational rate-distortion performance in Section III-C.

To this end, we first construct a reverse test channel as in Fig. 2, where $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}, \sigma_\eta^2\mathbf{I}_N)$, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \sigma_x^2\mathbf{I}_N)$, $\boldsymbol{\eta}$ is uncorrelated with $\mathbf{x}$, and $\sigma_\eta^2 + \sigma_x^2 = \sigma_h^2$. For such a reverse test channel, the mutual information between $\mathbf{x}$ and $\mathbf{h}$ is given by

$$\begin{aligned} \mathcal{I}(\mathbf{h}, \mathbf{x}) &= \mathcal{H}(\mathbf{h}) - \mathcal{H}(\mathbf{h}|\mathbf{x}) = \mathcal{H}(\mathbf{h}) - \mathcal{H}(\boldsymbol{\eta}) \\ &= \sum_{i=1}^{N} \mathcal{H}(h_i) - \sum_{i=1}^{N} \mathcal{H}(\eta_i) \\ &= N \log_2\left(2\pi e \sigma_h^2\right)\big) - N \log_2\left(2\pi e \sigma_\eta^2\right)\big) \\ &= N \log_2\left(\frac{\sigma_h^2}{\sigma_\eta^2}\right) \end{aligned} \quad (13)$$

where $\mathcal{H}(\mathbf{h})$ is the differential entropy of the complex Gaussian random vector $\mathbf{h}$, the second equality is due to independence of

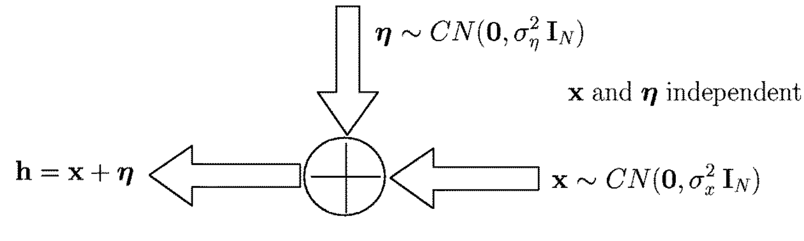

Fig. 2. Reverse test channel.

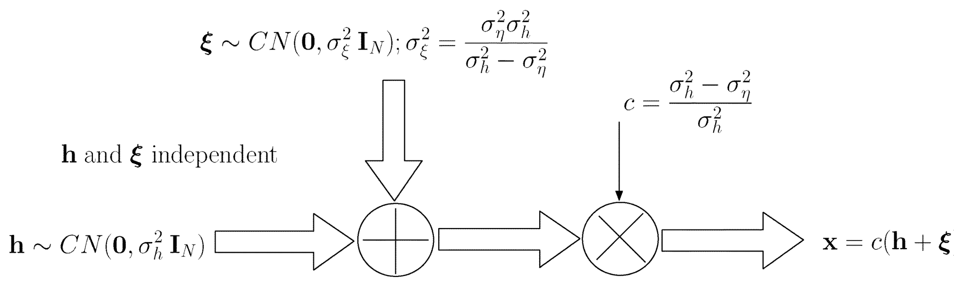

Fig. 3. Forward test channel.

$\mathbf{x}$ and $\boldsymbol{\eta}$, and the third equality is due to independence among the coefficients of $\mathbf{x}$ and $\boldsymbol{\eta}$.

We next evaluate the average distortion between $\mathbf{h}$ and $\mathbf{x}$ for this test channel. Notice that

$$\begin{aligned}\mathbb{E}\left[\frac{\mathbf{x}^{\mathcal{H}}\boldsymbol{\eta}\boldsymbol{\eta}^{\mathcal{H}}\mathbf{x}}{\mathbf{x}^{\mathcal{H}}\mathbf{x}}\right] &= \mathbb{E}\left[\text{trace}\left(\frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathcal{H}}\mathbf{x}\mathbf{x}^{\mathcal{H}}}{\mathbf{x}^{\mathcal{H}}\mathbf{x}}\right)\right]\\ &= \text{trace}\left(\mathbb{E}\left[\frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathcal{H}}\mathbf{x}\mathbf{x}^{\mathcal{H}}}{\mathbf{x}^{\mathcal{H}}\mathbf{x}}\right]\right)\\ &= \text{trace}\left(\mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathcal{H}}]\cdot\mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^{\mathcal{H}}}{\mathbf{x}^{\mathcal{H}}\mathbf{x}}\right]\right)\\ &= \text{trace}\left(\mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathcal{H}}]\cdot\frac{\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathcal{H}}]}{\mathbb{E}[\mathbf{x}^{\mathcal{H}}\mathbf{x}]}\right)\\ &= \text{trace}\left(\sigma_\eta^2\mathbf{I}_N\cdot\frac{1}{N}\mathbf{I}_N\right) = \sigma_\eta^2 \qquad (14)\end{aligned}$$

where $\text{trace}(\cdot)$ is the matrix trace operator, the second equality is due to the interchangeability of the trace and the expectation operators, the third equality is due to independence of $\boldsymbol{\eta}$ and $\mathbf{x}$, and the fourth equality is due to the independence of the vector length $|\mathbf{x}|$ and the vector direction $\mathbf{x}/|\mathbf{x}|$ for complex i.i.d. Gaussian vectors, i.e., $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathcal{H}}] = \mathbb{E}[|\mathbf{x}|^2]\cdot\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathcal{H}}/|\mathbf{x}|^2]$. Therefore, the average distortion is

$$\begin{aligned}\mathbb{E}\left[d_2(\mathbf{h},\mathbf{x})\right] &= \mathbb{E}\left[|\mathbf{h}|^2 - \frac{|\mathbf{x}^{\mathcal{H}}\mathbf{h}|^2}{|\mathbf{x}|^2}\right]\\ &= \mathbb{E}\left[|\boldsymbol{\eta}|^2 - \frac{|\mathbf{x}^{\mathcal{H}}\boldsymbol{\eta}|^2}{|\mathbf{x}|^2}\right]\\ &= (N-1)\sigma_\eta^2. \qquad (15)\end{aligned}$$

Letting $\sigma_\eta^2 = D/(N-1)$, the mutual information becomes

$$\mathcal{I}(\mathbf{h};\mathbf{x}) = N\log_2\left(\frac{N-1}{D}\sigma_h^2\right) \qquad (16)$$

while the average distortion is simply D.

Following the reverse test channel, we can construct a forward test channel, as shown in Fig. 3, where $\boldsymbol{\xi} \sim \mathcal{CN}(\mathbf{0}, \sigma_\xi^2\mathbf{I}_N)$ with $\sigma_\xi^2 = (\sigma_\eta^2\sigma_h^2)/(\sigma_\eta^2 + \sigma_h^2)$, $\boldsymbol{\xi}$ uncorrelated with $\mathbf{h}$, and the scaling constant $c := (\sigma_h^2 - \sigma_\eta^2)/\sigma_h^2$. It can be readily verified that (13), (15), and (16) also hold true for the forward test channel. These two test channels are actually equivalent, which can be seen from the facts that for both test channels, $\mathbf{h}$ and $\mathbf{x}$ are jointly Gaussian, and $\mathbb{E}[\mathbf{h}^{\mathcal{H}}\mathbf{h}]$ and $\mathbb{E}[\mathbf{x}^{\mathcal{H}}\mathbf{h}]$ are the same.

The reverse and the forward test channels demonstrate that a mutual information as shown in (16) is achievable subject to the average distortion constraint $\mathbb{E}[d_2(\mathbf{h},\mathbf{x})] \le D$ in (12). Equation (16) therefore serves as an upper-bound to the source input's rate-distortion function, when the aforementioned projective distance $d_2(\mathbf{h},\mathbf{x})$ is used as the distortion measure. Summarizing, we have Theorem 1.

Theorem 1: For a complex i.i.d. Gaussian source input $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2\mathbf{I}_N)$ with the distortion metric $d_2(\mathbf{h},\mathbf{x}) = |\mathbf{h}|^2 - |\mathbf{x}^{\mathcal{H}}\mathbf{h}|^2/|\mathbf{x}|^2$, an upper bound to the rate-distortion function of the source input $\mathbf{h}$ is given by

$$R_u(D) = \begin{cases} N\log_2\left(\frac{N-1}{D}\sigma_h^2\right), & 0 \le D \le (N-1)\sigma_h^2\\ 0, & D > (N-1)\sigma_h^2.\end{cases} \qquad (17)$$

Notice that when the required distortion $D > (N-1)\sigma_h^2$, no feedback bit is needed. Actually, it can be readily shown that in the worst beamforming case where the beam-steering vector is arbitrarily chosen, the average SNR is exactly σ_h^2, resulting in an average SNR degradation of $(N-1)\sigma_h^2$. Thus, the average SNR degradation can not exceed $(N-1)\sigma_h^2$.

An important dual to the rate-distortion function is the source input's distortion-rate function, defined as [11]

$$\begin{aligned}D(R) &:= \inf_{p(\mathbf{x}|\mathbf{h})} \mathbb{E}_{p(\mathbf{h})p(\mathbf{x}|\mathbf{h})}\left[d_2(\mathbf{h},\mathbf{x})\right]\\ &\text{s.t.} \quad \mathcal{I}(\mathbf{h};\mathbf{x}) \le R. \qquad (18)\end{aligned}$$

Due to the forward and the reverse test channels, an upper bound to the distortion-rate function can be similarly found as

$$D_u(R) = (N-1)\sigma_h^2 \cdot 2^{-\frac{2R}{N}}. \qquad (19)$$

For both bounds in (17) and (19), the bound-achieving conditional distribution $p(\mathbf{x}|\mathbf{h})$ is indeed specified by the forward test channel

$$p(\mathbf{x}|\mathbf{h}) \sim \mathcal{CN}\left(c\mathbf{h}, c^2\sigma_\xi^2\mathbf{I}_N\right) \qquad (20)$$

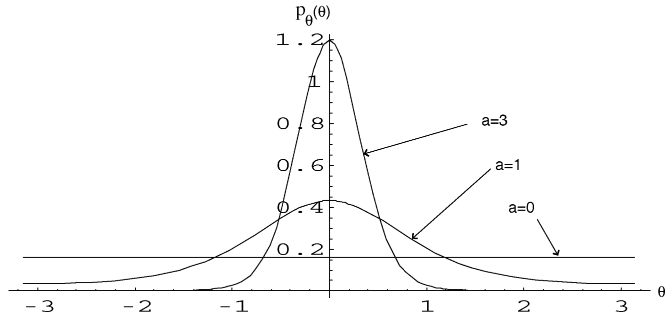

Fig. 4. Probability density function of the codeword direction θ.

which is complex normal distributed with mean proportional to $\mathbf{h}$ and covariance proportional to the identity matrix. Because the beamforming vector in practice is unit-norm constrained, we are particularly interested in the distribution of the codeword direction $\mathbf{x}/|\mathbf{x}|$. A tractable expression for the latter is not available in general. In the following, we give a simple example that provides useful insight to the distribution of the codeword direction.

Example: For illustration purpose, let us consider a two-dimensional real-valued example. Without loss of generality, let $\mathbf{h} := [h_1, 0]^T$, $a := |h_1|/(c\sigma_\xi)$, and $\mathbf{x} := [x_1, x_2]$. In this example, the direction of $\mathbf{x}$ is fully represented by the phase $\theta := \tan^{-1}(x_2/x_1)$. According to (20), the conditional distribution of the phase, or codeword direction, given channel realization $\mathbf{h} = [h_1, 0]^T$, is

$$p(\theta|\mathbf{h}) = \frac{e^{-a^2}}{2\pi} + \sqrt{\frac{1}{2\pi}} a \cos\theta e^{-\frac{a^2 \sin^2\theta}{2}} \cdot \frac{1 + \mathrm{Erf}\left(\frac{a\cos\theta}{\sqrt{2}}\right)}{2} \tag{21}$$

where $\mathrm{Erf}(\cdot)$ is the standard error function. Fig. 4 plots this conditional distribution for several values of a. The phase θ is uniformly distributed when $a = 0$; as a increases, the phase distribution becomes more and more concentrated around the given channel realization $\mathbf{h}$. In fact, when a is large enough, the conditional distribution of the codeword direction approaches a Gaussian density function with variance $1/a^2$

$$p(\theta|\mathbf{h}) \approx \frac{a}{\sqrt{2\pi}} e^{-\frac{a^2\theta^2}{2}}. \tag{22}$$

This can be justified because when a is large, the values of θ for which the density function (21) has significant values are getting very small, and thus $\sin(\theta) \approx \theta$ and $\cos(\theta) \approx 1$. The larger a is, the more clustered the conditional distribution is around the channel vector $\mathbf{h}$.

C. Quantifying the Operational Rate-Distortion Performance

It is well known that in order to achieve the rate-distortion function limit, quantization should be performed on a sequence-by-sequence basis, with the sequence length going to infinity [5], [7]. For the SVQ problem under consideration, the distortion measure between an input vector sequence and a codeword vector sequence can be defined as

$$d_2(\underline{\mathbf{h}}, \underline{\mathbf{w}}_i) := \frac{1}{L}\sum_{l=1}^{L} d_2\left(\mathbf{h}^{(l)}, \mathbf{w}_i^{(l)}\right) \tag{23}$$

where L is the sequence length, $\underline{\mathbf{h}} := \{\mathbf{h}^{(1)}, \dots, \mathbf{h}^{(L)}\}$ is an input sequence containing L input vectors, and $\underline{\mathbf{w}}_i := \{\mathbf{w}_i^{(1)}, \dots, \mathbf{w}_i^{(L)}\}$ is the ith codeword sequence containing L codeword vectors. The associated computational complexity and delay are prohibitively large, making the sequence-by-sequence quantization practically impossible.

On the other hand, the generalized Lloyd algorithm in Section III-A implements quantization on a vector-by-vector basis. Justifiably, a performance gap is always present between the rate-distortion function and the operational rate-distortion performance that is measured by the number of feedback bits versus average SNR degradation achieved by the generalized Lloyd algorithm. We next derive a lower bound to this operational rate-distortion performance. Interestingly, as we show next, a simple and important relationship emerges between the operational rate-distortion performance lower-bound and the rate-distortion function upper bound we derived in Section III-B.

Consider any codebook $\mathcal{W}$ and define the normalized SNR degradation

$$z := \frac{|\mathbf{h}|^2 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{h}\right|^2}{|\mathbf{h}|^2} = 1 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{g}\right|^2 \tag{24}$$

where z is a random variable taking values between 0 and 1. Let $F(z)$ denote the cumulative distribution function (CDF) of z. Even though $F(z)$ depends on the specific codebook, a codebook-independent upper bound is available [19], [30], as follows:

$$F(z) \le \tilde{F}(z) = \begin{cases} Mz^{N-1} & 0 \le z < M^{-\frac{1}{(N-1)}} \\ 1 & 1 > z \ge M^{-\frac{1}{(N-1)}}. \end{cases} \tag{25}$$

Let $R_o(D)$ denote the operational rate-distortion performance and $D_o(R)$ denote the dual operational distortion-rate performance. Based on (25)

$$\begin{aligned} D_o(R) &:= \mathbb{E}\left[|\mathbf{h}|^2 - \left|\mathbf{w}_*^{\mathcal{H}}\mathbf{h}\right|^2\right] \\ &= \int_0^1 z dF(z) \cdot \mathbb{E}\left[|\mathbf{h}|^2\right] \\ &\ge \int_0^1 z d\tilde{F}(z) \cdot \mathbb{E}\left[|\mathbf{h}|^2\right] \\ &= (N-1)\sigma_h^2 2^{-\frac{R}{N-1}} \end{aligned} \tag{26}$$

where the second equality follows from the independence of $|\mathbf{h}|$ and $\mathbf{g}$, and the inequality $\int_0^1 z dF(z) \ge \int_0^1 z d\tilde{F}(z)$ is due to the fact that $F(z) \le \tilde{F}(z)$, which can be verified using integration by parts. Thus

$$D_l(R) := (N-1)\sigma_h^2 2^{-\frac{R}{N-1}} \tag{27}$$

serves as a lower bound to the operational distortion-rate performance $D_o(R)$. Inverting (27), we obtain

$$R_l(D) := \begin{cases} (N-1)\log_2\left(\frac{N-1}{D}\sigma_h^2\right), & \text{if } 0 \le D \le (N-1)\sigma_h^2 \\ 0, & \text{if } D > (N-1)\sigma_h^2 \end{cases} \tag{28}$$

which is a lower bound to the operational rate-distortion performance $R_o(D)$. We emphasize that the lower bounds in (27) and

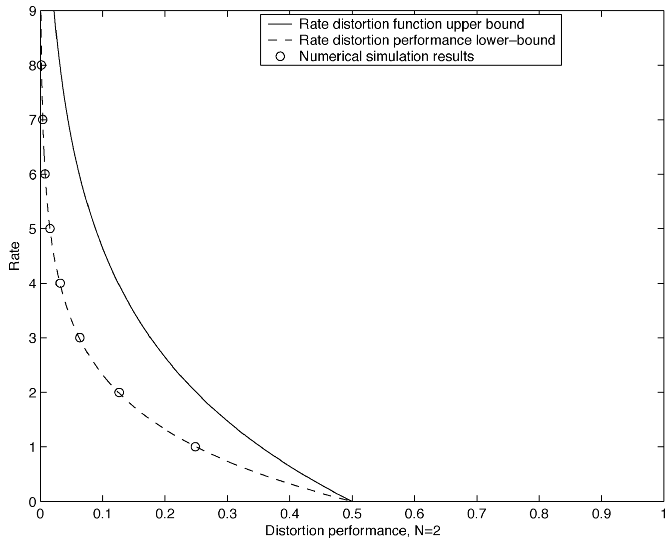

Fig. 5. Rate distortion performance, $N = 2$.

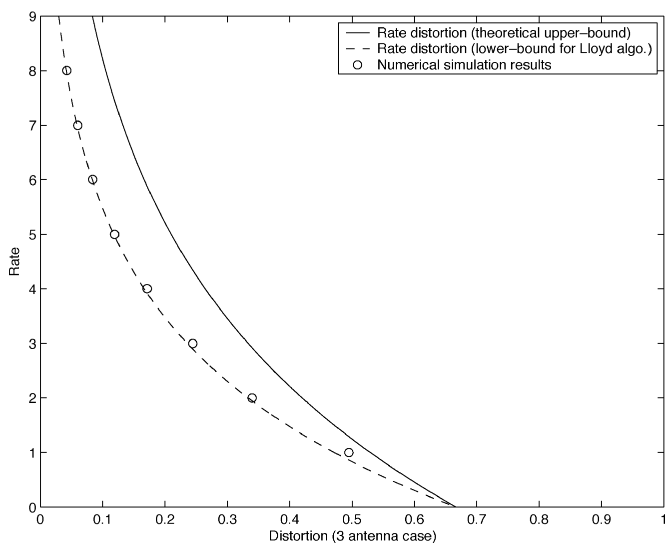

Fig. 6. Rate distortion performance, $N = 3$.

(28) are independent of the actual codebook design, and thus applicable to any beamformer codebook.

To investigate the tightness of the performance bounds in (27) and (28), we carry out Monte Carlo simulations for $N = 2$ up to 4 and plot the operational rate-distortion performance $R_o(D)$ in Figs. 5–7, respectively. The training sequence size of the generalized Lloyd algorithm was set to 100 000, the number of iterations to 20, and the channel covariance was $\sigma_h^2 = 1/N$. The performance lower-bound $R_l(D)$, and the source input's rate-distortion function upper bound $R_u(D)$ are also plotted for comparison. It is observed for all the numerical examples that with the codebook obtained from the generalized Lloyd algorithm, the lower bound $R_l(D)$ is sufficiently tight throughout the entire distortion range; and thus, the lower bound $R_l(D)$ indeed serves as a close approximation to $R_o(D)$.

Surprisingly, the operational rate-distortion performance lower bound (28) is nothing but a simple, scaled version of the source input's rate-distortion function upper bound in (17)

$$R_l(D) = \frac{(N-1)}{N} \cdot R_u(D). \tag{29}$$

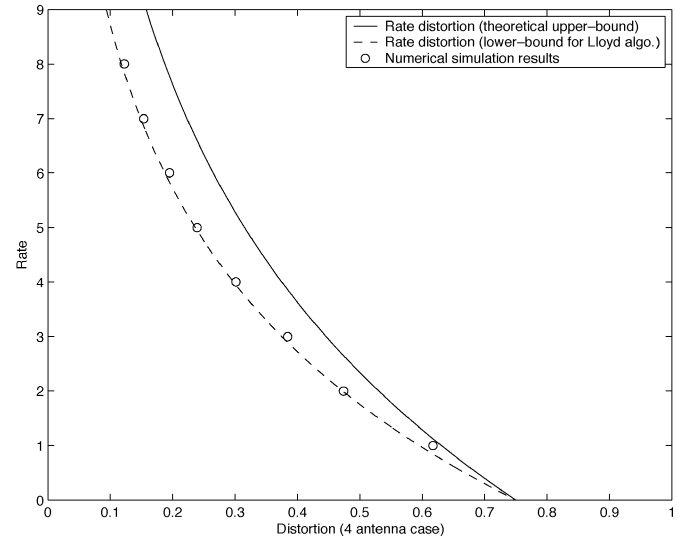

Fig. 7. Rate distortion performance, $N = 4$.

The scaling factor $(N-1)/N$ steadily approaches 1 as N increases. As we can see, the two performance analysis approaches, although very different in their nature, yield asymptotically equivalent results. To the best of our knowledge, such a simple relationship has not been identified before.

Summarizing, we have the following important result.

Proposition 1: For a complex i.i.d. Gaussian input $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I}_N)$ with distortion metric $d_2(\mathbf{h}, \mathbf{x}) = |\mathbf{h}|^2 - |\mathbf{x}^{\mathcal{H}}\mathbf{h}|^2/|\mathbf{x}|^2$, the operational rate-distortion performance achieved by the generalized Lloyd algorithm can be closely approximated by its lower bound in (28), which is a simple scaled version of the source's rate-distortion function upper bound in (17), as follows:

$$\begin{aligned} R_o(D) &\approx R_l(D) \\ &= \frac{N-1}{N} R_u(D) \\ &= (N-1)\log_2\left(\frac{N-1}{D}\sigma_h^2\right) \end{aligned} \tag{30}$$

when $0 \le D \le (N-1)\sigma_h^2$ and $R_o(D) = 0$ when $D > (N-1)\sigma_h^2$.

On the other hand, the dual operational distortion-rate performance achieved by the generalized Lloyd algorithm is also closely related to its lower bound (27) and the source's distortion-rate function (19) as follows:

$$D_o(R) \approx D_l(R) = D_u\left(\frac{N}{(N-1)}R\right) = (N-1)\sigma_h^2 2^{-\frac{R}{N-1}}. \tag{31}$$

More important, we are able to quantify the achieved average SNR by the so-designed beamformer codebook.

Proposition 2: For the considered N-dimensional transmit beamforming system operating over an i.i.d. Rayleigh fading link, with the size-2^B beamformer codebook designed by the generalized Lloyd algorithm, the achieved average SNR γ_o can be closely approximated by

$$\gamma_o(B) = N\sigma_h^2 - D_o(B) \approx N\sigma_h^2 - (N-1)\sigma_h^2 2^{-\frac{B}{N-1}}. \tag{32}$$

Proposition 2 provides a very important guideline for system engineering: Given a desired average SNR, it provides the number of feedback bits required to achieve it; conversely, given a fixed number of feedback bits, it quantifies the achievable average SNR.

Another important figure of merit is the average SNR improvement relative to the worst beamforming case, i.e., beamforming with an arbitrary beam-steering vector, which is simply given by

$$\kappa(B) := \gamma_o(B) - \gamma_o(0) = (N-1)\sigma_h^2 \left(1 - 2^{\frac{B}{N-1}}\right). \quad (33)$$

D. Alternate Codebook Design Approaches

As we have seen, to minimize the average SNR degradation in (3), the codebook design problem amounts to solving an SVQ problem, using the generalized Lloyd algorithm. Interestingly, the SVQ problem we dealt with is closely related to another interesting problem.

Consider a complex-valued codebook $\mathcal{W}$ of size $M \times N$ as in (7), where every codeword has unit norm. The maximum cross-correlation amplitude of such a codebook is defined as

$$I_{\max}(\mathcal{W}) := \max_{k \neq l} \left|\mathbf{w}_k^{\mathcal{H}} \mathbf{w}_l\right| \quad (34)$$

and Welch's lower bound on $I_{\max}$ is given by [26]

$$I_{\max}(\mathcal{W}) \geq \sqrt{\frac{(M-N)}{((M-1)N)}}. \quad (35)$$

As shown in [16] and [19], codebooks that achieve this maximum Welch bound[2] provide near-optimal solutions to the original SVQ problem for i.i.d. Rayleigh fading channels. We will call $\mathcal{W}$ a maximum Welch bound with equality (MWBE) codebook if it satisfies (35) as an equality [22].

Finding MWBE codebooks or achieving the maximum Welch bound, however, is not easy analytically or numerically [16], [22]. Analytical constructions are possible in certain cases [23], [27], while in other cases, one has to resort to numerical search algorithms in order to obtain near MWBE codebooks. In particular, a discrete Fourier transform (DFT)-based codebook search algorithm can be used [12]. As recently shown in [27], such a construction will also lead to desirable low-complexity quantization enabled by the FFT. In the following, we will not differentiate between codebooks designed using the generalized Lloyd algorithm, and those (near) MWBE codebooks designed as detailed in [12], [27], because both design approaches yield codebooks of similar performance (see, e.g., [16, Fig. 4] for an illustrating example). We will call these designs collectively *proper* codebooks for the SVQ problem under consideration. The maximum cross-correlation amplitude of such a proper codebook closely approaches, if not achieves, the maximum Welch bound (35).

E. From MISO to MIMO

Beamformer codebooks designed for MISO systems can be directly applied to i.i.d. Rayleigh MIMO channels without any loss of optimality [16]. To shed more light in the performance over i.i.d. Rayleigh fading MIMO channels, let $\mathbf{H}$ be the $N \times K$ channel matrix, with K being the number of receive antennas, and its entries i.i.d. Gaussian distributed as $h_{ij} \sim \mathcal{CN}(0, \sigma_h^2)$. Other assumptions are identical to the MISO setup. For the considered MIMO systems, an upper bound to the average SNR degradation is given by [16]

$$\mathbb{E}\left[\lambda \left(1 - \left|\mathbf{w}_*^{\mathcal{H}} \boldsymbol{\nu}\right|^2\right)\right] \quad (36)$$

where $\lambda, \boldsymbol{\nu}$ are the largest eigenvalue and the corresponding eigenvector of $\mathbf{H}^{\mathcal{H}}\mathbf{H}$. Because $\mathbf{H}^{\mathcal{H}}\mathbf{H}$ follows the Wishart distribution, λ and $\boldsymbol{\nu}$ are independent, with $\boldsymbol{\nu}$ uniformly distributed on the unit hypersphere Ω^N [28]. Furthermore, $\mathbb{E}[\lambda] = \sigma_s^2$ and can be calculated through integration [6, eq. (23)].

Comparing (36) with the objective function in (6), λ and $\boldsymbol{\nu}$ play exactly the roles of $|\mathbf{h}|^2$ and $\mathbf{g}$ in the MISO case. Therefore, it is clear that the performance analysis in Section III-B and Section III-C for i.i.d. Rayleigh MISO systems carries over to the i.i.d. Rayleigh MIMO case as well. Specifically, with σ_h^2 replaced by σ_s^2, (30) serves as an approximate upper bound to the average SNR degradation, and (32) serves as an approximate lower bound to the achievable average SNR.

IV. Codebook Designs for Correlated Rayleigh Fading Channels

So far, we have developed transmit beamformers for i.i.d. Rayleigh fading channels. In this section, we pursue transmit beamformer designs for correlated Rayleigh fading MISO channels, where $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ and $\mathbf{R}$ is the channel correlation matrix that depends on the relative geometry of the propagation environment; see also [9] for a detailed justification of this channel model. Matrix $\mathbf{R}$ is known both to the transmitter and to the receiver, and as before, the feedback link can support only B bits per block.

For such a correlated fading scenario, it is still desirable to maximize the average receive SNR. The problem though is different from that in (4), and its solution does not bear a close relationship with MWBE codebooks. Nevertheless, the problem can still be interpreted as a VQ problem, although no longer a spherical one. The generalized Lloyd algorithm can thus be employed to find the optimal codebook, called $\mathcal{W}_o$. The optimal solution, however, has to be reevaluated whenever the propagation geometry changes, which is rather complex and motivates an alternative approach.

To this end, we consider a proper (or near MWBE) codebook $\mathcal{W}_i = \{\mathbf{w}_{i1}, \ldots, \mathbf{w}_{iM}\}$ originally designed for i.i.d. Rayleigh fading channels, with a codebook matrix $\mathbf{W}_i = [\mathbf{w}_{i1} \ldots \mathbf{w}_{iM}]$. As shown in the Appendix, we have the result in Lemma 1.

Lemma 1: A proper (or near MWBE) codebook $\mathcal{W}_i$ satisfies $\mathbf{W}_i \mathbf{W}_i^{\mathcal{H}} \approx (M/N)\mathbf{I}_N$.

Using the proper codebook $\mathcal{W}_i$ and the factorization $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathcal{H}}$, we can first construct a transformed codebook

$$\{\mathbf{A}\mathbf{w}_{i1}, \ldots, \mathbf{A}\mathbf{w}_{iM}\} \quad (37)$$

[2]The problem of achieving the maximum Welch bound is also known as Grassmannian line packing in mathematics.

with codebook matrix $\mathbf{AW}_i$. According to Lemma 1, the second-order sample moment of this transformed codebook is

$$\left(\frac{1}{M}\right)\sum_{j=1}^{M}\mathbf{Aw}_{ij}\mathbf{w}_{ij}^{\mathcal{H}}\mathbf{A}^{\mathcal{H}}=\left(\frac{1}{M}\right)\mathbf{AW}_i\mathbf{W}_i^{\mathcal{H}}\mathbf{A}^{\mathcal{H}}\approx\left(\frac{1}{N}\right)\mathbf{R} \tag{38}$$

approximately matching the ensemble correlation of the underlying channel. This kind of moment matching property is also true for proper codebooks in i.i.d. Rayleigh fading channels

$$\left(\frac{1}{M}\right)\sum_{j=1}^{M}\mathbf{w}_{ij}\mathbf{w}_{ij}^{\mathcal{H}}\approx\left(\frac{1}{N}\right)\mathbf{I}_N \tag{39}$$

and is consistent with the principle behind successful quantization: construct a collection of codewords to represent the random source input as close as possible.

Since codewords in such a transformed codebook (37) violate the unit norm constraint on beamforming vectors, they have to be properly normalized. For this reason, we select our beamformer codebook for correlated Rayleigh fading channels as

$$\mathcal{W}_r=\left\{\frac{\mathbf{Aw}_{i1}}{\|\mathbf{Aw}_{i1}\|_2},\ldots,\frac{\mathbf{Aw}_{iM}}{\|\mathbf{Aw}_{iM}\|_2}\right\}. \tag{40}$$

For illustration purposes, we plot a two-dimensional real-valued codebook example in Fig. 8. The channel correlation matrix is randomly set as [0.445 −0.027; −0.027 0.052], and $M = 8$ is chosen. An optimal codebook $\mathcal{W}_o$ obtained using the generalized Lloyd algorithm is compared with the proposed codebook $\mathcal{W}_r$, and the proper codebook $\mathcal{W}_i$ based on which $\mathcal{W}_r$ is constructed. As we can see, the proposed design $\mathcal{W}_r$ comes very close to the optimal codebook $\mathcal{W}_o$, while $\mathcal{W}_i$ is considerably different from $\mathcal{W}_o$. We further compare the achieved average SNR performance by these different codebooks for complex-valued fading channels. The correlation matrix is set as $\mathrm{diag}(81\ 25\ 9\ 1)/\sqrt{116}$, and $N = 4$ is chosen. In Fig. 9, the achieved average SNR is plot versus the number of regions M. Clearly, the proposed codebook $\mathcal{W}_r$ achieves most of the optimal performance associated with $\mathcal{W}_o$, while the performance gap between $\mathcal{W}_i$ and $\mathcal{W}_o$ is significant.

As corroborated by these numerical studies, the proposed codebook achieves near optimal performance, while enjoying extremely low complexity. The only thing required is to maintain a proper codebook for i.i.d. Rayleigh fading channels. When the propagation geometry changes, a new beamformer codebook is constructed easily from the stored proper codebook and the channel correlation matrix.

V. Conclusion

We analyzed performance of transmit beamforming for i.i.d. Rayleigh MIMO systems under limited-rate feedback. We first established an upper bound to the rate-distortion function of the vector source input and next derived a lower bound to the operational rate-distortion performance achieved by the generalized Lloyd algorithm. Interestingly, these two performance analysis approaches, although very different in nature, yield asymptotically equivalent results. We further quantified the achievable average SNR, which provides an effective guideline for practical system designs. Quantifying the performance provides a concrete basis for future analytical study of more general transmission schemes under limited-rate feedback, such as joint adaptive modulation/beamforming. We also developed a low-complexity beamformer codebook design for correlated Rayleigh fading channels. Performance analysis of such beamformer codebooks constitutes an interesting direction for future research.

Transmit beamforming offers great application potential in wireless networks, and is currently being considered by the IEEE 802.11n standard. In this paper, we considered transmissions over flat-fading channels only. When the fading channel is frequency selective, it is possible to pursue related designs in the context of MIMO OFDM systems, where correlation is typically present across subcarriers. In such cases, it may be beneficial to combine the limited-rate beamforming scheme with adaptive modulation schemes. In addition, it may not be practical to feed back the same amount of bits for all subcarriers as the number of carriers is usually very large. Allocating a finite number of feedback bits to certain subcarriers depending

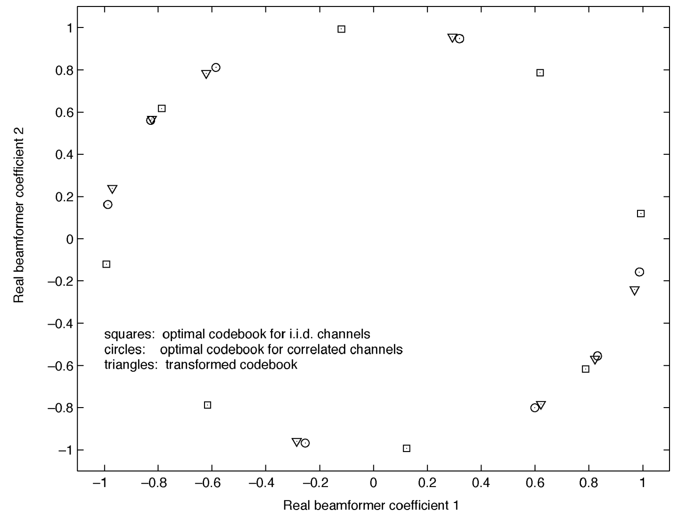

Fig. 8. Real-valued codebook example for correlated Rayleigh fading channels.

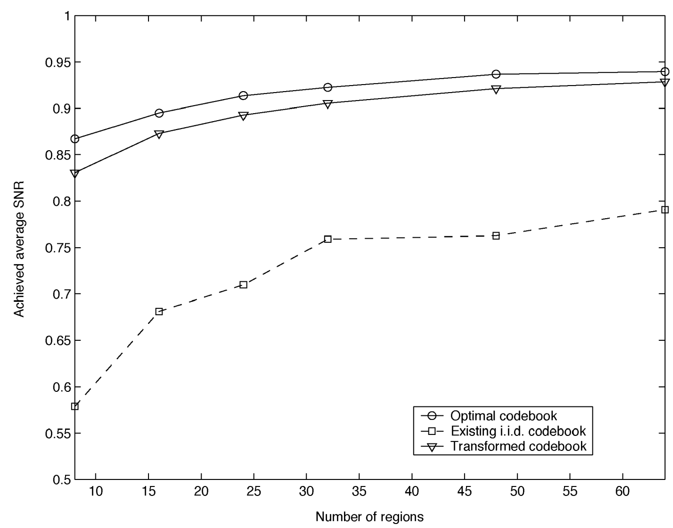

Fig. 9. Average SNR comparisons over correlated Rayleigh fading channels.

on the channel conditions is well motivated. In practical system designs, interference from adjacent cells or networks should also be taken into account. In such cases, it may be interesting to study limited-rate system designs for multiuser MIMO channels, which is a direction that warrants future research.

APPENDIX

Lemma 1: A proper (or near MWBE) codebook $\mathcal{W}_i$ satisfies $\mathbf{W}_i\mathbf{W}_i^{\mathcal{H}} \approx (M/N)\mathbf{I}_N$.

Proof: To see this, we introduce another Welch bound different than the bound in (35). Consider a complex-valued codebook $\mathcal{W}_i$ of size $M \times N$ as in (7), where every codeword is of unit norm. For such a codebook, the root-mean-square cross-correlation amplitude is defined as

$$I_{\rm rms}(\mathcal{W}) := \sqrt{\frac{1}{M(M-1)}\sum_{k=1}^{M}\sum_{\ell\neq k}^{M}\left|\mathbf{w}_{ik}^{\mathcal{H}}\mathbf{w}_{i\ell}\right|^2} \tag{41}$$

and the Welch lower bound on $I_{\rm rms}$ is given by

$$I_{\rm rms}(\mathcal{W}) \geq \sqrt{\frac{M-N}{(M-1)N}}. \tag{42}$$

If equality holds in (42), the codebook $\mathcal{W}_i$ is said to meet the Welch bound on $I_{\rm rms}$ with equality and will be referred to as a WBE codebook. It can be readily shown that [18], [21], [22] an MWBE codebook is also WBE and that a codebook $\mathcal{W}_i$ is WBE if and only if $\mathbf{W}_i\mathbf{W}_i^{\mathcal{H}} = (M/N)\mathbf{I}_N$. Combining these two facts and seeing that a proper codebook is approximately MWBE, we arrive at Lemma 1, which has also been verified by numerical results.

ACKNOWLEDGMENT

The authors would like to thank Dr. Y. Yao for helpful discussions on this and related topics.

REFERENCES

[1] S. Bhashyam, A. Sabharwal, and B. A. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 785–798, May 2002.
[2] R. Blum, "MIMO with limited feedback of channel state information," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 4, Hong Kong, China, Apr. 6–10, 2003, pp. 89–92.
[3] R. Blum and J. Winters, "On optimum MIMO with antenna selection," *IEEE Commun. Lett.*, vol. 6, no. 8, pp. 322–324, Aug. 2002.
[4] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139–159, 1996.
[5] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[6] P. A. Dighe, K. Mallik, and S. S. Jamuar, "Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 694–703, Apr. 2003.
[7] R. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.
[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.
[9] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna methods for multipath environments," in *Proc. IEEE GLOBECOM* , vol. 1, San Francisco, CA, Nov. 28–Dec. 2 1994, pp. 425–429.
[10] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
[11] R. M. Gray. Lecture notes on quantization and compression, fundamentals of quantization. [Online]. Available: http://www.stanford.edu/class/ee372/basics4.pdf
[12] B. M. Hochwald, T. L. Marzetta, T. L. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 9, pp. 1962–1973, Sep. 2000.
[13] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. Int. Symp. Information Theory*, Washington DC, Jun. 2001, p. 321.
[14] G. Jöngren and M. Skoglund, "Utilizing quantized feedback information in orthogonal space–time block coding," in *Proc. IEEE Global Telecommunications Conf. (IEEE GLOBECOM)*, vol. 2, San Francisco, CA, Nov. 27–Dec. 1 2000, pp. 995–999.
[15] V. K. N. 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, no. 1, pp. 62–70, Jan. 2004.
[16] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.
[17] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.
[18] J. L. Massey and T. Mittlelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in *Sequences II: Methods in Communication, Security and Computer Sciences*. New York: Springer, 1993, pp. 63–78.
[19] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. A. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
[20] 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. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
[21] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1261–1266, Jul. 1994.
[22] D. V. Sarwate, "Meeting the Welch bound with equality," in *Sequences and Their Applications*. New York: Springer, 1999, pp. 79–102.
[23] T. Strohmer and R. W. Heath Jr., "Grassmannian frames with applications to coding and communication," *Applied Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257–275, 2003.
[24] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.
[25] E. Visotsky and U. Madhow, "Space–time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
[26] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
[27] P. Xia, S. Zhou, and G. B. Giannakis, "Achieving the Welch bound with difference sets," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1900–1907, May 2005.
[28] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[29] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space–time block coding based on channel mean feedback," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.
[30] S. Zhou, Z. Wang, and G. B. Giannakis, "Quantifying the power loss when transmit-beamforming relies on finite rate feedback," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1948–1957, Jul. 2005.

Pengfei Xia (S'03–M'05) received the B.S. and M.S. degrees in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 1997 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Twin Cities, in 2005.

His research interest lies in the area of signal processing and communications, including MIMO wireless communications, phased antenna array and beamforming, OFDM and multicarrier transmissions, adaptive modulation and coding.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Southern California (USC), 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been a Professor with the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 250 journal papers, 400 conference papers, and two edited books. Current research focuses on diversity techniques for fading channels, complex-field and space–time coding, cross-layer designs, and sensor networks.

Dr. Giannakis is the (co)recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (1992, 1998, 2000, 2001, 2003, and 2004). He also received the Technical Achievement Awards from the SP Society in 2000 and from EURASIP in 2005. He served as Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the SP Conference Board, as Member of the SP Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the SP for Communications Technical Committee, and as a Member of the IEEE Fellows Election Committee. He has also served as a Member of the IEEE SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.