

When is Limited Feedback for Transmit Beamforming Useful?

Vasanthan Raghavan

Department of Electrical and Computer Engineering

raghavan@cae.wisc.edu

December 13, 2004

Abstract

Multi-antenna systems have been shown to offer tremendous capacity increase in ideal rich scattering conditions. Transmit beamforming and receive combining are low complexity techniques that help in achieving the full diversity afforded by the multi-antenna channel. Even if complete channel knowledge at the receiver maybe a realistic assumption in these systems, the same may not be true at the transmitter side. Thus, quantized beamforming with limited feedback on a reverse link has been a topic that has attracted great attention recently. But almost all of the work to date has focussed on modeling the channel with independent, identically distributed (i.i.d.) Rayleigh fading between antenna pairs. In this report, we first review the earlier works which focus on Grassmannian line packing. We then consider the correlated channel case, and show that Grassmannian line packing is an artificial artifact of the i.i.d. assumption. We show that there are dominant peaks in the eigen-domain when correlation is imposed and the code-book construction should be matched to the correlation in the channel, which renders the problem in this case easier than for the i.i.d. channels.

1 Introduction

Multi-antenna systems have come to the forefront of research with the seminal works of Telatar [1], and Foschini and Gans [2] who have shown that enormous capacity increase is possible under rich scattering conditions. Exploiting the spatial dimensions to achieve enhanced resilience to fading has been the other extreme objective. The tradeoff between achieving the full rate and the full diversity in a multi-antenna system has been analyzed in [3].

Diversity in a multi-antenna system is achievable either through space-time coding or using the channel state information (CSI) at the transmitter intelligently. Low complexity techniques like transmit beamforming and receive combining have attracted significant attention of late. This is because of the possibility to feed back at a lower rate, partial CSI from the receiver to the transmitter, both in deployed 3-G wireless systems and in ongoing standardization efforts for 4-G systems [4].

Recent works towards this goal has addressed the problem of code-book design for the transmit beamforming vector under the independent and identically distributed (i.i.d.) channel assumption [5], [6]. A natural code-book design criterion in the i.i.d. case was shown in [6] to be maximization of minimum distance between beamforming vectors. Love

et al. [5] showed that this criterion coincides with the objective of maximizing the average SNR at the receiver. This objective led to a solution of packing one-dimensional subspaces, also well known in algebra as the Grassmannian line packing problem.

Most realistic channels are far from having i.i.d. Rayleigh fading between antenna pairs. Therefore existing code-book designs have to be modified to account for channel correlation to achieve higher performance gains in realistic channel conditions. But before pursuing the objective of efficient code-book designs for the correlated channel case, one needs to address the fundamental question: When is limited feedback for transmit beamforming beneficial? This report concerns this important question.

We show that the i.i.d. channel assumption, far from making the code-book design problem easier, masks it due to the isotropic nature of the channel, i.e. the singular vectors of the channel are equally probable to point in any direction in the eigen-domain. We first consider a Kronecker product correlation model, and later an even more realistic virtual representation of channel matrices to show that correlation in the channel entries renders the code-book design problem easy. Unlike the i.i.d. channel which has singular vectors isotropically distributed, a correlated channel has peaks in the eigen-domain corresponding to the eigenvectors of the transmit covariance matrix. This observation leads to an important, albeit previously unnoticed, conclusion: Limited feedback for transmit beamforming is beneficial in the i.i.d. channel case, and in the correlated channel case provided the number of transmit and receive antennas are few.

Subsequent work would focus on sub-optimal code-book designs in the correlated channel case. The system model under consideration is described in the next section, while in Section 3 we introduce the reader to the existing literature on i.i.d. beamforming design. New results on design criterion and benefits of beamforming are discussed in Section 4, while conclusions are drawn in Section 5. Some of the proofs have been relegated to the appendix.

In this report we use the following notations: x_i to denote the i -th entry of the vector \mathbf{x} , $\mathbf{A}(i, j)$ for the i, j -th entry of the matrix \mathbf{A} , \mathcal{C}^m to denote the m -dimensional complex space, Ω_m to denote the set of unit vectors in \mathcal{C}^m , $\mathbf{E}_{\mathbf{H}}[\cdot]$ for the expectation operator with respect to the random variable \mathbf{H} , $\|\cdot\|_2$ to refer to the 2-norm of a vector or a matrix, $\mathbf{H}^{\mathbf{H}}$, \mathbf{T} for Hermitian transpose, and regular transpose respectively. We also remind the readers that we mean $\left| \frac{f(N)}{g(N)} \right| \leq K < \infty$ by the notation $f(N) = \mathcal{O}(g(N))$.

2 System Model

2.1 System Setup

We consider a single user communication system employing transmit beamforming and receive combining, and assume that signaling is done using N_T transmit and N_R receive antennas. The input-output relationship of this system is given by

$$y = \mathbf{z}^H \mathbf{H} \mathbf{w} x + \mathbf{z}^H \mathbf{n} \quad (1)$$

where \mathbf{H} is the $N_R \times N_T$ channel matrix connecting the transmitter and the receiver, \mathbf{z} is the receive combining vector, \mathbf{w} is the transmit beamforming vector, x is the transmitted symbol from a chosen constellation (QPSK, 16-QAM etc.), and \mathbf{n} is the independent noise added at the receiver.

We assume that the receiver knows the channel perfectly, while the transmitter knows the channel statistics. This assumption is quite common in existing multi-antenna system

designs. The receiver and the transmitter have the knowledge of a finite code-book, \mathbf{W} , of transmit beamforming vectors (dependent on the channel statistics) and the receiver feeds back for every independent channel realization, the optimal beamforming vector from this code-book via a low-rate feedback channel (viz. the vector index in the code-book).

The beamforming and combining vectors are chosen to maximize the received SNR, γ , which is,

$$\gamma = \frac{E_s |\mathbf{z}^H \mathbf{H} \mathbf{w}|^2}{\|\mathbf{z}\|_2^2 N_0} \quad (2)$$

where $E_s = \mathbf{E}|x|^2$ is the symbol energy, and N_0 is the noise variance. If the transmitter has CSI, then it is easy to see that the optimal beamforming vector is the dominant right singular vector of \mathbf{H} corresponding to the largest singular value¹. Given a transmit beamforming vector, it is also easy to see that the optimal receive combining vector is $\frac{\mathbf{H} \mathbf{w}}{\|\mathbf{H} \mathbf{w}\|_2}$.

The distortion, $G(\mathbf{W})$, between the system with complete CSI and the system employing limited feedback is given by [5]

$$G(\mathbf{W}) = \mathbf{E}_{\mathbf{H}} [\lambda_1 - \|\mathbf{H} \mathbf{Q}_{\mathbf{w}}(\mathbf{H})\|_2^2] \quad (3)$$

where λ_1 is the largest eigenvalue of $\mathbf{H}^H \mathbf{H}$, and

$$\mathbf{Q}_{\mathbf{w}}(\mathbf{H}) = \arg \max_{\mathbf{w} \in \mathbf{W}} \|\mathbf{H} \mathbf{w}\|_2^2. \quad (4)$$

Here λ_1 is the normalized gain achieved by a system that has complete CSI and $\|\mathbf{H} \mathbf{Q}_{\mathbf{w}}(\mathbf{H})\|_2^2$ is the gain corresponding to limited feedback. The design objective thus is to select a code-book based on the channel statistics so as to minimize the distortion measure. We will follow the terminologies of transmit beamforming as in [5].

2.2 Channel Models

In this sub-section we consider various statistical and parametric modeling schemes for the channel \mathbf{H} . The statistics of \mathbf{H} depend on antenna geometry, physical scattering environment, frequency of operation etc. Ideal channel modeling assumes that the entries of \mathbf{H} are i.i.d. Gaussian random variables [1], [2]. Analysis of an i.i.d. channel is tractable, but the i.i.d. channel assumption, however, maybe unrealistic for wireless applications where large antenna spacings or a rich scattering environment are not possible.

Parametric *physical* models, on the other hand, explicitly model signal copies from different directions [7]. Despite being highly accurate in capturing the scattering features, most of these models suffer from the problem of mathematical intractability. They do not offer insights into designing effective signaling strategies or space-time code construction in the general case. The need for channel modeling schemes that capture most of the statistical information of the underlying physical environment, and yet stay within the realms of mathematical tractability is therefore paramount.

Towards this goal, we first study the often-used Kronecker product (or separable) correlation model. This channel model is characterized by a stacked covariance matrix that has a Kronecker product structure. More precisely, the channel matrix \mathbf{H} is of the form

$$\mathbf{H} = \Sigma_R^{1/2} \mathbf{H}_{\text{id}} \Sigma_T^{1/2} \quad (5)$$

¹Application of Cauchy-Schwartz inequality.

where Σ_R and Σ_T correspond to receive and transmit covariance matrices respectively, and \mathbf{H}_{iid} is an i.i.d. random matrix as defined earlier. If $\mathbf{h} = \text{vec}(\mathbf{H})$ is the vector obtained by stacking the columns of \mathbf{H} , then it follows that

$$\mathbf{E} [\mathbf{h}\mathbf{h}^H] = \Sigma_T \otimes \Sigma_R. \quad (6)$$

The Kronecker product correlation model, despite its mathematical tractability and simplification of performance analysis of MIMO systems, suffers from deficiencies acquired by the separability property of channel correlation that limits the number of degrees of freedom in the model [8], [9]. Measurement campaigns have shown that the Kronecker product correlation model is accurate in capturing the underlying channel statistics under certain conditions [10], but in general, the model underestimates the measured channel capacity [8] and other system parameters.

Various statistical models have been proposed to increase the number of degrees of freedom in the channel model. When the transmitter and receiver antennas form a uniform linear array (ULA), the virtual channel representation [11] has been shown to capture the channel characteristics quite accurately. The generalization of the virtual representation to the non-ULA case, the canonical statistical model [12], [13], [14], has also been studied extensively of late, both mathematically and via measurement campaigns. We now introduce the virtual representation and the canonical statistical model.

The virtual representation exploits the finite dimensionality of the signal space, afforded by finite number of antenna elements and finite array aperture, to develop a linear channel representation that uses spatial beams in fixed, virtual directions. The virtual channel representation can be expressed as

$$\mathbf{H} = \sum_{m=1}^{N_R} \sum_{n=1}^{N_T} H_V(m, n) \mathbf{a}_R(\theta_{R,m}) \mathbf{a}_T^H(\theta_{T,n}) = \tilde{\mathbf{A}}_R \mathbf{H}_V \tilde{\mathbf{A}}_T^H \quad (7)$$

where $\left\{ \theta_{R,m} = \frac{m}{N_R} \right\}$ and $\left\{ \theta_{T,n} = \frac{n}{N_T} \right\}$ are fixed virtual angles that result in unitary $\tilde{\mathbf{A}}_R$ and $\tilde{\mathbf{A}}_T$. Here, $\tilde{\mathbf{A}}_R = [\mathbf{a}_R(\theta_{R,1}) \ \cdots \ \mathbf{a}_R(\theta_{R,N_R})]$, $\tilde{\mathbf{A}}_T = [\mathbf{a}_T(\theta_{T,1}) \ \cdots \ \mathbf{a}_T(\theta_{T,N_T})]$, and \mathbf{H}_V is the virtual channel matrix.

The virtual representation is linear since the virtual angles are fixed by the spatial resolution. The virtual channel coefficients represent the coupling between the transmit and the receive virtual angles. The most important advantage of the virtual representation is the fact that the virtual entries are approximately uncorrelated and this uncorrelatedness of channel coefficients gets better as number of antennas increase. The number of virtual channel entries with non-zero variance corresponds to the number of independent degrees of freedom afforded by the channel and provides a lower bound on the number of resolvable paths captured by the antenna arrays. The readers are referred to [11] for a detailed treatment of the virtual channel representation in the narrowband case and [15], [16] for the wideband case.

The generalization of the virtual representation to a non-ULA configuration leads to the canonical statistical channel model [12]. This model assumes that the auto and cross-correlation matrices on both transmitter and receiver sides have the same eigen-basis, and exploits this redundancy to decompose the channel as

$$\mathbf{H} = \mathbf{U}_R^{1/2} \mathbf{H}_{\text{ind}} \mathbf{U}_T^{1/2} \quad (8)$$

where \mathbf{U}_R and \mathbf{U}_T correspond to receive and transmit covariance matrices respectively, and \mathbf{H}_{ind} is a random matrix with independent entries that are zero mean, proper complex Gaussian and variance σ_{ij}^2 , which are not necessarily equal.

It can be shown that when ULAs are used at the transmitter and receiver ends, the Kronecker product model reduces to a special case of the virtual representation [17]. Besides, in the non-ULA setting, the Kronecker product correlation model is a special case of the canonical statistical model. The canonical statistical model not only incorporates the virtual representation and the Kronecker product model as special cases, but has also been shown to be quite accurate in predicting performance metrics of measured channels. Accuracy of the canonical statistical model in predicting measured channel capacity and other performance metrics is reported in [13], [18].

3 Grassmannian Line Packing Solution to I.I.D. Beamforming

In this section, we essentially review some of the known results in the case of i.i.d. beamforming which amounts to Grassmannian line packing [5].

3.1 Mathematical Preliminaries

Grassmannian line packing is the problem of optimally packing one-dimensional subspaces. It is similar to the problem of spherical code design with one big difference: spherical codes are points on the unit sphere while Grassmannian line packings are lines passing through the origin in a vector space. Here we present some key results that would help us develop the problem of limited feedback code-book design.

Consider the space of unit-norm transmit beamforming vectors Ω_m . An equivalence relation is defined on this space where $\mathbf{w}_1 \equiv \mathbf{w}_2$ if and only if there exists $\theta \in [0, 2\pi)$ s.t. $\mathbf{w}_1 = e^{i\theta} \mathbf{w}_2$. In other terms, two vectors are equivalent if they lie on the same line in \mathcal{C}^m . The quotient space under this equivalence relation is defined to be the complex Grassmann manifold $\mathcal{G}(m, 1)$, the set of all one-dimensional subspaces in \mathcal{C}^m . A standard distance function is defined on $\mathcal{G}(m, 1)$ where the distance between the two lines generated from unit vectors \mathbf{w}_1 and \mathbf{w}_2 is the sine of the angle between the two lines, i.e.

$$d(\mathbf{w}_1, \mathbf{w}_2) = \sqrt{1 - |\mathbf{w}_1^H \mathbf{w}_2|^2}. \quad (9)$$

The Grassmannian line packing problem is the problem of finding the set, or packing, of N lines in \mathcal{C}^m that has maximum minimum distance between any pair of lines. This optimal packing is represented by an $m \times N$ matrix $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_N]$, where \mathbf{w}_i corresponds to the i -th line in the packing. We consider non-trivial cases where $N > m$, for if $N \leq m$, then any set of N orthogonal directions could be chosen.

The minimum distance of the packing, $\delta(\mathbf{W})$ is defined as

$$\delta(\mathbf{W}) = \min_{k \neq l} d(\mathbf{w}_k, \mathbf{w}_l). \quad (10)$$

There is a vast amount of literature on algorithms to design packings for a given m and N . The reader is referred to [19], [20], [21] (and the references therein) for more information.

An upper bound to the best possible performance is obtained via the Rankin bound, which states that

$$\delta(\mathbf{W}) \leq \sqrt{\frac{(m-1)N}{m(N-1)}}. \quad (11)$$

There are refined bounds for $\delta(\mathbf{W})$ in the line of the well-known Gilbert Varshamov bound of coding theory.

Another useful property of a packing is its density. Let $\mathcal{P}_{\mathbf{v}}$ denote the column-space spanned by the vector \mathbf{v} (also known as the line generated by the vector \mathbf{v}). The ball of radius γ in $\mathcal{G}(m, 1)$ around the line generated by the vector \mathbf{w}_i is defined as

$$\mathcal{B}_{\mathbf{w}_i}(\gamma) = \{\mathcal{P}_{\mathbf{v}} \in \mathcal{G}(m, 1) \text{ s.t. } d(\mathbf{v}, \mathbf{w}_i) < \gamma\}. \quad (12)$$

We note that

$$\mathcal{B}_{\mathbf{w}_k}(\gamma) \cap \mathcal{B}_{\mathbf{w}_l}(\gamma) = \phi \quad (13)$$

for $k \neq l$ when $\gamma \leq \delta(\mathbf{W})/2$ with ϕ representing the empty set. Metric balls in $\mathcal{G}(m, 1)$ can be viewed as spherical caps on Ω_m . Thus, the ball $\mathcal{B}_{\mathbf{w}_i}(\gamma)$ is the set of lines generated by all vectors on the unit sphere that are within a chordal distance of γ from any point in $\Omega_m \cap \mathcal{P}_{\mathbf{w}_i}$.

The normalized Haar measure, the standard measure on unitary matrix space, on Ω_m introduces a normalized invariant measure μ on $\mathcal{G}(m, 1)$ allowing the computation of the density of a packing. The density of a packing is defined as

$$\begin{aligned} \Delta(\mathbf{W}) &= \mu \left(\bigcup_{i=1}^N \mathcal{B}_{\mathbf{w}_i}(\delta(\mathbf{W})/2) \right) \\ &= \sum_{i=1}^N \mu(\mathcal{B}_{\mathbf{w}_i}(\delta(\mathbf{W})/2)) \\ &= N\mu(\mathcal{B}(\delta(\mathbf{W})/2)) \end{aligned} \quad (14)$$

where $\mathcal{B}(\delta(\mathbf{W})/2)$ is an arbitrarily centered ball of radius $\delta(\mathbf{W})/2$. A result in [5] obtains the density of line packings as

$$\Delta(\mathbf{W}) = N(\delta(\mathbf{W})/2)^{2(m-1)}. \quad (15)$$

This result states that higher the minimum distance of the packing, the higher the density. Thus finding packings with large minimum distance is distinctively advantageous not only in terms of separation, but also in terms of the efficiency of the packing.

3.2 Code-Book Design

In this section, we apply the ideas of Grassmannian line packing to construct good code-books in the case of an i.i.d. channel \mathbf{H} . As stated in Section 2, the code-book design is based on minimizing the distortion in (3). In the case of an i.i.d. channel, this goal is simplified because of the fact that an i.i.d. random matrix has eigenvalues and eigenvectors that are mutually independent. More so, the eigenvectors are equally probable to point in any direction in Ω_m [22].

Using the eigen-decomposition of $\mathbf{H}^H \mathbf{H}$, we can write $G(\mathbf{W})$ as

$$\begin{aligned} G(\mathbf{W}) &= \mathbf{E}_{\mathbf{H}} \left[\lambda_1 - \sum_{i=1}^{N_T} \lambda_i |\mathbf{u}_i^H \mathcal{Q}_{\mathbf{w}}(\mathbf{H})|^2 \right] \\ &\leq \mathbf{E}_{\mathbf{H}} \left[\lambda_1 - \lambda_1 |\mathbf{u}_1^H \mathcal{Q}_{\mathbf{w}}(\mathbf{H})|^2 \right] \\ &= \mathbf{E}_{\mathbf{H}} [\lambda_1] \mathbf{E}_{\mathbf{H}} \left[1 - |\mathbf{u}_1^H \mathcal{Q}_{\mathbf{w}}(\mathbf{H})|^2 \right] \end{aligned} \quad (16)$$

where λ_i and \mathbf{u}_i are the eigenvalues and the eigenvectors of $\mathbf{H}^H \mathbf{H}$, and the last line follows from the independence of the eigenvalues and eigenvectors in the i.i.d. case. The $\mathbf{E}_{\mathbf{H}} [\lambda_1]$ corresponds to the channel contribution to distortion, and $\mathbf{E}_{\mathbf{H}} \left[1 - |\mathbf{u}_1^H \mathcal{Q}_{\mathbf{w}}(\mathbf{H})|^2 \right]$, the contribution due to code-book design.

Using the fact that \mathbf{u}_1 is uniformly distributed and the equivalent definition of the density of a packing which is $\Delta(\mathbf{W}) = \mathbb{P} \left(1 - |\mathbf{u}_1^H \mathcal{Q}_{\mathbf{w}}(\mathbf{H})|^2 < \frac{\delta(\mathbf{W})^2}{4} \right)$, we have an upper bound for $G(\mathbf{W})$ as

$$\begin{aligned} G(\mathbf{W}) &\leq \mathbf{E}_{\mathbf{H}} [\lambda_1] \left(\frac{\delta(\mathbf{W})^2}{4} \Delta(\mathbf{W}) + (1 - \Delta(\mathbf{W})) \right) \\ &\leq \mathbf{E}_{\mathbf{H}} [\lambda_1] \left(1 + N \left(\frac{\delta(\mathbf{W})}{2} \right)^{2N_T-2} \left(\frac{\delta(\mathbf{W})^2}{4} - 1 \right) \right). \end{aligned} \quad (17)$$

Thus to minimize the upper bound as above, one needs to maximize the minimum distance $\delta(\mathbf{W})$, since $\frac{\delta(\mathbf{W})}{2} < 1$. Thus the code-book design problem has been effectively reformulated to an algebraic problem of maximizing the minimum distance of a packing in \mathcal{C}^m . The best packings for a given m and N are tabulated in [23] and can be used to obtain the optimal code-book in the i.i.d. case.

4 Beamforming in the Correlated Channel

4.1 Mathematical Preliminaries

In this sub-section, we introduce the mathematical results that would be needed in our code-book design for the correlated channel case. The first lemma is the complement to Fatou's Lemma for positive random variables [24].

Lemma 1 (Durrett). Let $\{X_i\}$ be a sequence of positive random variables. Then

$$\limsup_{n \rightarrow \infty} \mathbf{E} X_n \leq \mathbf{E} \left[\limsup_{n \rightarrow \infty} X_n \right]$$

The law of large numbers and the central limit theorem play an ubiquitous role in the analysis of communication systems. The next lemma of this section pertains to the rate of convergence of the law of large numbers for a sum of independent (not necessarily identically distributed) random variables [25].

Lemma 2 (Tuyêñ, 1991). Let $\{X_i\}$ be a sequence of independent zero-mean random variables and let $\{a_{nk}\}$ be a double array of real numbers. Let $\|a_n\|_{\beta} = \left(\sum_k |a_{nk}|^{\beta} \right)^{1/\beta}$.

Suppose there exist real numbers $\alpha > 0$, $\beta \geq 1$ such that $0 < \|a_n\|_{\min(\beta, 2)} < \infty$ for all n , $\|a_n\|_\beta \rightarrow 0$, and there exists $0 < T_{\alpha\beta} \leq \infty$ such that $f_k(t) = \mathbf{E} [e^{tX_k}] \leq e^{\alpha|t|^\beta}$ for all k , and all real t such that $|t| \leq T_{\alpha\beta}$. If $\beta > 1$, then

$$\mathbb{P} \left(\left| \sum_{k=1}^n a_{nk} X_k \right| \geq \epsilon \right) \leq 2 \exp \left(-\frac{\epsilon \min(T_{\alpha\beta}, T)}{\|a_n\|_\beta} + \alpha [\min(T_{\alpha\beta}, T)]^\beta \right) \quad (18)$$

where $T = \left(\frac{\epsilon}{\alpha\beta\|a_n\|_\beta} \right)^{1/(\beta-1)}$.

Note that the theorem is stated for real random variables X_k in the paper, and its extension to the complex case is straightforward.

The last lemma of this section concerns the eigenvectors of perturbed matrices. The classical Gershgorin theorem [26] states that the eigenvalue of a matrix perturbed by an error matrix differs from the eigenvalue of the unperturbed matrix by the (spectral)-norm of the perturbation. However the connection between the eigenvectors of perturbed and unperturbed matrices is more subtle. The famous Davis-Kahan $\sin(\theta)$ theorem [26] states that the sine of the angle between eigenspaces of matrices that are close to one another is related to the norm of the difference between the matrices, provided the eigenvalues of the two matrices are well separated. In our setting, the following easier lemma [27] suffices.

Lemma 3 (Mathias, 1994). Let $\Lambda = \text{diag}(\lambda_i)$ be a positive definite diagonal matrix with distinct eigenvalues. Let X be a Hermitian matrix and consider $S(\epsilon) = \Lambda + \epsilon X$. Then for sufficiently small ϵ , we have $\lambda_j(\epsilon) = \lambda_j(S(\epsilon))$ distinct and one can choose $\hat{u}(\epsilon)$ to be an eigenvector of $S(\epsilon)$ such that

$$\begin{aligned} \hat{u}(\epsilon)_j &= 1 + \mathcal{O}(\epsilon^2) \\ |\hat{u}(\epsilon)_i| &\leq \epsilon \frac{|X_{ij}|}{|\lambda_i - \lambda_j|} + \mathcal{O}(\epsilon^2), \quad i \neq j. \end{aligned}$$

We note that in the above lemma, only the largest eigenvalue of Λ needs to be distinct to ensure that the eigenvector corresponding to the largest eigenvalue of $\Lambda + \epsilon X$ has a perturbation of ϵ over the eigenvector of Λ . This is an important point that we would exploit in the sequel.

4.2 Code-Book Design

4.2.1 Distortion Measure

Here, we propose a mathematically tractable upper bound for the distortion measure in (3) for a general correlated channel \mathbf{H} . We remind the readers that this goal is simplified in the case of an i.i.d. channel, since the eigenvalues and the eigenvectors of $\mathbf{H}^H \mathbf{H}$ are independent. However, this independence is no longer true in the case of a correlated \mathbf{H} . However, we have the following upper bound for the distortion measure.

Proposition 1. The distortion measure, $G(\mathbf{W})$ is upper bounded by U where

$$U = \sqrt{1 - \mathbf{E}_{\mathbf{H}}(A)} \left[\mathbf{E}_{\mathbf{H}}(\lambda_1) + \sqrt{2\text{Var}_{\mathbf{H}}(\lambda_1)} \right] \quad (19)$$

where λ_1 is the largest eigenvalue of $\mathbf{H}^H \mathbf{H}$, \mathbf{u}_1 is the corresponding eigenvector, and A denotes $\max_{\mathbf{w} \in \mathbf{W}} |\mathbf{u}_1^H \mathbf{w}|^2$.

Proof. See Appendix A. □

The first part of U , which is $\sqrt{1 - \mathbf{E}_{\mathbf{H}}(A)}$ represents the effect of code-book design on the distortion, and $\left[\mathbf{E}_{\mathbf{H}}(\lambda_1) + \sqrt{2\mathbf{Var}_{\mathbf{H}}(\lambda_1)} \right]$ corresponds to the channel contribution to distortion. Our focus here will be on minimizing the average fractional loss in SNR, $1 - \mathbf{E}_{\mathbf{H}}(A)$, so as to reduce distortion. We also note that the upper bound in Proposition 1 complements the upper bound on distortion in the i.i.d. channel case, (16). Another useful upper bound to (19) is obtained by employing Lemma 1 and this results in $\sqrt{\min_{\mathbf{w} \in \mathbf{W}} \mathbf{E}_{\mathbf{H}} [1 - |\mathbf{u}_1^H \mathbf{w}|^2]} \left[\mathbf{E}_{\mathbf{H}}(\lambda_1) + \sqrt{2\mathbf{Var}_{\mathbf{H}}(\lambda_1)} \right]$.

4.2.2 Conditions Under Which Limited Feedback is Beneficial

In this section, we explore the conditions under which limited feedback for transmit beamforming produces substantial benefits. To be precise, we consider the case when there is no feedback from the receiver to the transmitter and show that the average fractional loss in SNR diminishes with increasing receiver dimensions. It is easy to see that when there is no feedback between the receiver and the transmitter the optimal beamforming vector is the eigenvector corresponding to the largest eigenvalue of $\mathbf{H}^H \mathbf{H}$. We would then like to quantify the average loss in SNR given that our codebook has only one code vector, the eigenvector of $\mathbf{H}^H \mathbf{H}$ corresponding to the largest eigenvalue. Under this assumption, the part of the upper bound in Proposition 1 due to the code-book design can be written as $\mathbf{E}_{\mathbf{H}} [1 - |\mathbf{u}_1^H \mathbf{w}|^2]$. Our main conclusion stems from the following theorem.

Theorem 1. Let the channel \mathbf{H} be modeled via the Kronecker product correlation form. Also assume that the largest eigenvalue of Σ_t has an algebraic multiplicity 1. Then the average fractional loss of SNR in the case of no feedback is $\mathcal{O} \left(\frac{\log N_T N_R}{N_R^{1/2}} \right)$.

Proof. See Appendix B. □

The theorem shows that the average fractional loss in SNR diminishes to zero as the receiver dimensions increase. Intuitively, the independence of channel entries leads to channel hardening [28], which in turn results in the concentration of the dominant right singular vector of \mathbf{H} in the direction corresponding to the largest eigenvector of the transmit covariance matrix. The separability assumption of the largest eigenvalue of $\mathbf{H}^H \mathbf{H}$ makes intuitive sense since the presence of a dominant eigen-direction implies that when the transmitter has only statistical information about the channel, beamforming along that dominant direction would be optimal. It appears that this multiplicity assumption about the largest eigenvalue is superfluous, for if the eigenspace corresponding to the largest eigenvalue had a higher dimensionality than 1, beamforming along any vector in this space would be optimal.

Also the convergence rate proven in Theorem 1 is critically dependent on the convergence rate of Lemma 2. Better bounding techniques in the theorem and faster convergence rates for the lemma would help in increasing the rate of decay of the average fractional loss of SNR to 0 with antenna dimensions. We have the following generalization of Theorem 1 to the case when \mathbf{H} is modeled via the virtual representation or the canonical statistical model.

Theorem 2. Assume that the channel \mathbf{H} is expressed via a virtual representation or the canonical statistical model. Assume that the largest value in the ensemble given by $\{(D_T)_{ii} \sum_k (D_r)_{kk} \sigma_{ki}^2\}$ is well separated from the rest, where D_T and D_R are the diagonal

matrices in the eigenvalue decomposition of the transmit and receive covariance matrix. Then the average fractional loss of SNR in the case of no feedback is $\mathcal{O}\left(\frac{\log N_T N_R}{N_R^{1/2}}\right)$.

Proof. The proof is exactly similar to the proof of Theorem 1. The assumption of well-separability is used to show that with a very high probability the largest eigenvalue of Λ (as in Theorem 1) is distinct. \square

The theorems 1 and 2 show that in a correlated channel, gains obtained by limited feedback diminish as the receiver dimensions increase. This is however not the case in the case of an i.i.d. channel, where there are no dominant singular vectors due to the isotropicity of the channel matrix. In the case of correlated channels, there is a dominant peak in the eigen-domain and this peak precisely corresponds to the dominant right singular vector of \mathbf{H} . The limited feedback (code-book design) problem belongs to a category of problems where the solution is easier when the channel is correlated than if it is i.i.d. unlike capacity analysis, space-time code design etc. Thus substantial gains via limited feedback could be hoped for only in the small antenna dimensions.

The code-book design in the correlated channel case is more complicated than the i.i.d. design since not much is known about the probability distribution of the eigenvectors of a random non-isotropic matrix. This knowledge is crucial in constructing optimal code-books with constraints on feedback rate. However as shown in Theorem 1, the largest singular vector of \mathbf{H} tends to converge in some sense to the largest eigenvector of the transmit covariance matrix. Exploiting this knowledge, one could construct sub-optimal designs for the correlated channel. Subsequent studies will focus on this aspect of correlated beamforming.

5 Conclusions

In this report, we have studied the problem of limited feedback beamformer design for the multi-antenna channel. Limited feedback is an important problem which has secured immense attention in the wireless community of late. We first reviewed some of the known results for i.i.d. beamformer design which corresponds to a packing problem in the m -dimensional complex space. Minimum distance of a packing has to be increased for the code-book to be optimal. The vast amount of literature in the applied mathematics and algebra literature helps us in leveraging our understanding of transmit beamforming.

In this work, we considered the beamformer design problem in the correlated channel situation, and contrasting the i.i.d. case, we have showed that the loss in performance when no feedback is done vanishes as the number of receiver antennas increase. This result of ours shows that limited feedback offers substantial benefits in two scenarios: the i.i.d. channel case that has been already considered in the literature, and in the correlated case provided the antenna dimensions are few. As antenna dimensions increase, the singular vectors of the channel tend to get peaked around the largest eigenvector of the transmit covariance matrix in some sense and this renders the design problem trivial.

Appendix

5.1 Proof of Proposition 1

Proof. It is not difficult to see that $G(\mathbf{W})$ can be upper bounded by $\mathbf{E}_{\mathbf{H}} [\lambda_1(1 - |\mathbf{u}_1^H \mathbf{Q}_{\mathbf{W}}(\mathbf{H})|^2)]$ (see (18) in [5]). An easily tractable upper bound to the former is $\mathbf{E}_{\mathbf{H}} [\lambda_1(1 - \max_{\mathbf{w} \in \mathbf{W}} |\mathbf{u}_1^H \mathbf{w}|^2)]$. Let A denote the random variable $\max_{\mathbf{w} \in \mathbf{W}} |\mathbf{u}_1^H \mathbf{w}|^2$. We now employ the definition of the correlation coefficient between two random variables to obtain

$$\mathbf{E}_{\mathbf{H}} [\lambda_1(1 - A)] = \mathbf{E}_{\mathbf{H}}(\lambda_1)\mathbf{E}_{\mathbf{H}} [1 - A] + \rho\sqrt{\mathbf{Var}_{\mathbf{H}}(\lambda_1)}\sqrt{\mathbf{Var}_{\mathbf{H}}(1 - A)} \quad (20)$$

$$\leq \mathbf{E}_{\mathbf{H}}(\lambda_1)\mathbf{E}_{\mathbf{H}} [1 - A] + \sqrt{\mathbf{E}_{\mathbf{H}}(A^2) - [\mathbf{E}_{\mathbf{H}}(A)]^2}\sqrt{\mathbf{Var}_{\mathbf{H}}(\lambda_1)} \quad (21)$$

$$\leq \mathbf{E}_{\mathbf{H}}(\lambda_1) [1 - \mathbf{E}_{\mathbf{H}}(A)] + \sqrt{\mathbf{Var}_{\mathbf{H}}(\lambda_1)}\sqrt{1 - [\mathbf{E}_{\mathbf{H}}(A)]^2} \quad (22)$$

$$\begin{aligned} &= \sqrt{1 - \mathbf{E}_{\mathbf{H}}(A)} \left[\sqrt{1 - \mathbf{E}_{\mathbf{H}}(A)}\mathbf{E}_{\mathbf{H}}(\lambda_1) + \sqrt{1 + \mathbf{E}_{\mathbf{H}}(A)}\sqrt{\mathbf{Var}_{\mathbf{H}}(\lambda_1)} \right] \\ &\leq \sqrt{1 - \mathbf{E}_{\mathbf{H}}(A)} \left[\mathbf{E}_{\mathbf{H}}(\lambda_1) + \sqrt{2\mathbf{Var}_{\mathbf{H}}(\lambda_1)} \right] \end{aligned} \quad (23)$$

where in (21) we have used the fact that ρ is real and $|\rho| \leq 1$, in (22) we have used a trivial upper bound of 1 for $\mathbf{E}_{\mathbf{H}}(A^2)$, and in (23) we have used the fact that $\mathbf{E}_{\mathbf{H}}(A) \leq 1$. Thus we are done. \square

5.2 Proof of Theorem 1

Proof. Let $\mathbf{U}_T \mathbf{D}_T \mathbf{U}_T^H$ and $\mathbf{U}_R \mathbf{D}_R \mathbf{U}_R^H$ denote the eigen-decompositions of Σ_T and Σ_R respectively. We first note that exploiting the isotropicity property of an i.i.d. random matrix, $\mathbf{H}^H \mathbf{H}$ can be written as $\mathbf{H}^H \mathbf{H} = N_R \left(\frac{1}{N_R} \mathbf{H}^H \mathbf{H} \right) = N_R \mathbf{U}_T \mathbf{D}_T^{1/2} \left(\frac{1}{N_R} \mathbf{H}_{\text{iid}}^H \mathbf{D}_R \mathbf{H}_{\text{iid}} \right) \mathbf{D}_T^{1/2} \mathbf{U}_T^H = \mathbf{Y} + \Delta \mathbf{Y}$, where $\mathbf{Y} = N_R \mathbf{U}_T \mathbf{D}_T^{1/2} \mathbf{D} \mathbf{D}_T^{1/2} \mathbf{U}_T^H$, $\mathbf{D} = \text{diag} \left[\left(\frac{1}{N_R} \mathbf{H}_{\text{iid}}^H \mathbf{D}_R \mathbf{H}_{\text{iid}} \right)_{ii} \right]$, $\Delta \mathbf{Y} = N_R \mathbf{U}_T \mathbf{D}_T^{1/2} \mathbf{O} \mathbf{D}_T^{1/2} \mathbf{U}_T^H$, $\mathbf{O} = \frac{1}{N_R} \mathbf{H}_{\text{iid}}^H \mathbf{D}_R \mathbf{H}_{\text{iid}} - \mathbf{D}$. That is, we decompose $\mathbf{H}^H \mathbf{H}$ into a part constituted by the dominant diagonal terms, \mathbf{D} , and the off-diagonal components in \mathbf{O} , which vanish in the limit of large N_R due to the law of large numbers, placing us within the realm of Lemma 3.

Using the hypothesis that the largest eigenvalue of Σ_T is well separated from the other eigenvalues, and Lemma 2 (with $\beta = 2$, $f_k(t) = \frac{1}{(1-t(D_R)_{kk})^2}$ [29], $T_{\alpha\beta} = \frac{1}{2 \max_k (D_R)_{kk}}$, $\|a_n\|_{\beta} = \frac{1}{\sqrt{N_R}}$, and α sufficiently large), we have that the largest eigenvalue of $\mathbf{D}_T \mathbf{D}$ is distinct with probability greater than $p_0 = 1 - 2 \exp \left(-\frac{\epsilon^2 \sqrt{N_R}}{8 \max_k (D_R)_{kk}} \right)$. Further analysis is done modulo this high probability set. We use Λ to denote $\mathbf{D}_T \mathbf{D}$, and $X = \frac{1}{\max_{ij} |\mathbf{O}_{ij}|} \mathbf{D}_T^{1/2} \mathbf{O} \mathbf{D}_T^{1/2}$. Thus $\mathbf{H}^H \mathbf{H} = N_R \mathbf{U}_T [\Lambda + \max_{ij} |\mathbf{O}_{ij}| X] \mathbf{U}_T^H$. Applying Lemma 3, we have the following conclusion:

$$\begin{aligned} \hat{u}(j) &= 1 + e_j, \quad |e_j| = \mathcal{O}(\epsilon^2) \\ \hat{u}(i) &= e_i, \quad |e_i| \leq \epsilon K + \mathcal{O}(\epsilon^2), i \neq j \end{aligned} \quad (24)$$

with probability $p = \text{P}(\max_{ij} |\mathbf{O}_{ij}| \leq \epsilon)$, where \hat{u} is the eigenvector of $\Lambda + \max_{ij} |\mathbf{O}_{ij}| X$, and K is a positive constant.

Note that the fractional loss in SNR corresponds to the squared value of the sine of the angle between the eigenvectors of the perturbed and unperturbed matrices. Elementary

algebra using the orthogonality of columns of \mathbf{U}_T shows that the fractional loss in SNR random variable is $\mathcal{O}(\epsilon^2)$ with probability p . Thus the average fractional loss in SNR is upper bounded by $U_1 = 1 + (K_1\epsilon^2 - 1)(p + p_0 - 1)$ where K_1 is a positive constant. Note that $q = 1 - p \leq \sum_{ij} \mathbb{P}(|\mathbf{O}_{ij}| > \epsilon) \leq \frac{N_T(N_R-1)}{2} \mathbb{P}(|\mathbf{O}_{ij}| > \epsilon)$. We observe that a product of zero-mean Gaussian random variables (of variances σ_1^2 and σ_2^2 , respectively) has a Bessel-function of the zeroth order like probability density function and its Moment Generating Function is given by $\left(\frac{1}{1-\sigma_1^2\sigma_2^2 t^2}\right)^{1/2}$ [29]. Using Lemma 2 to estimate the right hand side of the above expression, with $\beta = 2$, $T_{\alpha\beta} = \frac{1}{2\max_k (D_R)_{kk}}$, $\|a_n\|_\beta = \frac{1}{\sqrt{N_R}}$, $\alpha = 2\sqrt{N_R}\max_k (D_R)_{kk}$, we obtain $U_1 \leq K_1\epsilon^2 p_0 + (1 - K_1\epsilon^2)q + (1 - p_0) \leq K_1\epsilon^2 + (1 - K_1\epsilon^2)(N_T(N_R - 1) + 2) \exp\left(-\frac{\epsilon^2\sqrt{N_R}}{8\max_k (D_R)_{kk}}\right)$. Choosing ϵ to minimize the above upper bound, we find that $U_1 = \mathcal{O}\left(\frac{\log(N_R N_T)}{N_R^{1/2}}\right)$. \square

References

- [1] Í. E. Telatar, "Capacity of Multi-antenna Gaussian channels," *AT& T-Bell Labs Internal Tech. Memo.*, June 1995.
- [2] G. J. Foschini, M. J. Gans "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Per. Comm.*, vol. 6, no. 3, pp. 311-335, 1998.
- [3] L. Zheng, D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073-1096, May 2003
- [4] M. Z. Ikram, E. N. Onggosanusi, V. Raghavan, A. G. Dabak, S. Hosur, "An enhanced closed-loop MIMO design for OFDM/OFDMA-PHY," IEEE 802.16 Broadband Wireless Access WG, IEEE doc. C802.16e-04/267, Seoul, Korea, Aug. 30 - Sept. 2, 2004.
- [5] D. J. Love, R. W. Heath Jr., T. Strohmer, "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735-2747, Oct. 2003
- [6] K. K. Mukkavilli, A. Sabharwal, E. Erkip, B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562-2579, Oct. 2003
- [7] A. J. Paulraj, C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, pp. 49-83, Nov. 1997.
- [8] H. Ozelik, M. Herdin, W. Weichselberger, J. Wallace, E. Bonek, "Deficiencies of 'Kronecker' MIMO Radio Channel Model," *Electronics Letters*, vol. 39, no. 16, pp. 1209-1210, Aug. 2003
- [9] J. Wallace, H. Ozelik, M. Herdin, E. Bonek, M. Jensen, "Power and Complex Envelope Correlation for Modeling Measured Indoor MIMO Channels: A Beamforming Evaluation," *IEEE 58th Fall Vehicular Technology Conference, 2003*, vol. 1, pp. 363-367, Oct. 2003

- [10] D. P. McNamara, M. A. Beach, P. N. Fletcher, "Spatial Correlation in Indoor MIMO Channels," *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, 2002*, vol. 1, pp. 290-294, 2002
- [11] A. M. Sayeed, "Deconstructing Multi-antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563-2579, Oct. 2002.
- [12] J. H. Kotecha, A. M. Sayeed, "Canonical Statistical Modeling and Capacity Analysis of Correlated MIMO Fading Channels," *Submitted to the IEEE Trans. Wireless Comm.*, Mar. 2004
- [13] W. Weichselberger, H. Ozelik, M. Herdin, E. Bonek, "A Novel Stochastic MIMO Channel Model and its Physical Interpretation," *Sixth International Symposium on Wireless Personal Multimedia Communications (WPMC'03)*, Oct. 2003, Yokosuka, Japan
- [14] A. M. Tulino, A. Lozano, S. Verdu, "Impact of Correlation on the capacity of multi-antenna channels," *Submitted to IEEE Trans. Inform. Theory*. Sept. 2003
- [15] K. Liu, V. Raghavan, A. M. Sayeed, "Capacity scaling and spectral efficiency in wideband correlated MIMO channels," *IEEE Trans. Inform. Theory, Special Issue on MIMO Systems*, vol. 49, no. 10, pp. 2504-2526, Oct. 2003.
- [16] A. M. Sayeed, V. V. Veeravalli, "A virtual representation for time and frequency selective MIMO channels: A bridge between physical and statistical models," *submitted to the IEEE Trans. Comm.*, available at <http://dune.ece.wisc.edu>.
- [17] J. H. Kotecha, A. M. Sayeed, "Optimal transmit signal design for estimation of correlated MIMO fading channels," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 546-557, Feb. 2004.
- [18] Y. Zhou, A. M. Sayeed, "Exploring the realistic MIMO channel through virtual channel representation," *to be submitted to IEEE. Trans. Antennas and Propagation*, May 2004.
- [19] J. H. Conway, R. H. Hardin, N. J. A. Sloane, "Packing lines, planes, etc.: Packing in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139-159. 1996
- [20] T. Strohmer, R. W. Heath Jr., "Grassmannian frames with applications to coding and communications," *Appl. Comput. Harm. Anal.*, vol. 14, no. 3, pp. 257-275
- [21] D. Agarwal, T. J. Richardson, R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inform. Theory*, vol. 47, no. 9, pp. 2618-2626, Sept. 2001
- [22] A. Edelman, "Eigenvalues and condition numbers of random matrices," *Ph. D. dissertation*, MIT, Cambridge, MA, 1989
- [23] N. J. A. Sloane, Packings in Grassmannian spaces [Online] <http://www.research.att.com/njas/grass/index.html>
- [24] R. A. Durrett, "Probability: Theory and Examples," Duxbury Press, 2nd edition, 1995.

- [25] D. Q. Tuyên, “On the convergence rate of weighted sums of independent random variables,” Available on the web
- [26] R. Bhatia, “Matrix analysis,” Springer, 1996
- [27] R. Mathias, “Spectral Perturbation bounds for positive definite matrices,” *SIAM Journ. Matrix Anal. Appl.*, vol. 18, no. 4, pp. 959-980, Oct. 1997
- [28] B. Hochwald, T. Marzetta, V. Tarokh, “Multi-antenna channel hardening and its implications for rate feedback and scheduling,” *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893-1909, Sept. 2004
- [29] M. K. Simon, “Probability distributions involving Gaussian random variables: A handbook for engineers and scientists,” Kluwer, 2002