# Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels

Vasanthan Raghavan, Robert W. Heath, Jr., and Akbar M. Sayeed

***Abstract*— The full diversity gain provided by a multi-antenna channel can be achieved by transmit beamforming and receive combining. This requires the knowledge of channel state information (CSI) at the transmitter which is difficult to obtain in practice. Quantized beamforming where fixed codebooks known at both the transmitter and the receiver are used to quantize the CSI has been proposed to solve this problem. Most recent works focus attention on limited feedback codebook design for the uncorrelated Rayleigh fading channel. Such designs are sub-optimal when used in correlated channels. In this paper, we propose systematic codebook design for correlated channels when channel statistical information is known at the transmitter. This design is motivated by studying the performance of pure statistical beamforming in correlated channels and is implemented by maps that can rotate and scale spherical caps on the Grassmannian manifold. Based on this study, we show that *even* statistical beamforming is near-optimal if the transmitter covariance matrix is ill-conditioned and receiver covariance matrix is well-conditioned. This leads to a partitioning of the transmit and receive covariance spaces based on their conditioning with variable feedback requirements to achieve an operational performance level in the different partitions. When channel statistics are difficult to obtain at the transmitter, we propose a universal codebook design (also implemented by the rotation-scaling maps) that is robust to channel statistics. Numerical studies show that even few bits of feedback, when applied with our designs, lead to near perfect CSI performance in a variety of correlated channel conditions.**

***Index Terms*— Diversity methods, fading channels, Grassmannian line packing, limited feedback, MIMO systems, quantization**

## I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) wireless systems provide a mechanism to increase the reliability of signal reception (diversity), or rate of information transfer (multiplexing), or a combination of these aspects. The focus of this paper is on achieving the maximal possible diversity and array gain in a MIMO channel. Simple techniques like transmit beamforming and receive combining that need *only* a scalar codec (that is, they have a linear complexity in the antenna dimensions) are well-known to achieve this objective [1]. Optimal beamforming, however, requires *perfect* transmitter knowledge of the channel's dominant right singular vector. This can be obtained using the reciprocity of the channel in certain scenarios. In the most general case, one needs to rely on the use of a feedback channel for conveying limited channel state information (CSI) to the transmitter.

One approach for multi-antenna channels [2], [3], [4] is based on quantization of the instantaneous CSI at the receiver followed by its conveyance to the transmitter using a low-rate feedback channel. This is the *limited feedback* approach. There are other approaches [5], [6], [7] which are based on feedback of *partial* (or *statistical*) channel information, *e.g.,* the channel mean or the channel covariance matrices. These methods, in general, do not perform as well as the ones using instantaneous feedback since they do not track the rapid fluctuations of the channel. There are more general feedback schemes [8], [9], [10] that dynamically adapt to the distribution of the channel, but these methods are generally too complicated to implement in practice.

In [2], [3], [4], the beamforming vector is quantized at the receiver using a fixed codebook available at both the transmitter and the receiver. In the current literature, codebooks are predominantly designed and optimized for uncorrelated Rayleigh fading channels with various codeword selection criteria based on the instantaneous channel knowledge. These quantization strategies lack the ability to exploit knowledge of the channel correlation, and hence result in a degradation of the received signal-to-noise ratio (SNR) when the channel is not uncorrelated. There are designs which tailor the codebooks to spatially correlated Rayleigh fading channels (using knowledge of the channel correlation matrix) [11], [12], but these works are constrained to a single receiver antenna and transmitter-side correlation. To the best of our knowledge, no work exists on designing channel statistics-dependent codebooks in the most general MIMO setting.

In this work, we propose such a systematic limited feedback codebook construction for quantized beamforming in correlated channels. The workhorse of this design is the motivation that ill-conditioning of the transmit and well-conditioning of the receive covariance matrices aid eigenvector hardening which results in vanishing loss of pure statistical beamforming with respect to perfect CSI beamforming in the limit of receive

Manuscript received June 1, 2006; revised Jan. 2, 2007. This work was supported by the NSF Grants No. CCF-514194, CNS-435307, CNS-322957 through the University of Texas, Austin and Grant No. CCF-0431088 through the University of Wisconsin-Madison, and the ONR Grant No. N00014-05-1-0169 through the University of Texas, Austin. This paper was presented in part at the International Symposium on Information Theory 2005, Adelaide, Australia and 2006, Seattle, WA, USA and at the 2005 IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA with R. Samanta.

Vasanthan Raghavan was with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA while the work was done. He is currently with the Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA (e-mail: vasanthan_raghavan@ieee.org).

Robert W. Heath, Jr. is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA (e-mail: rheath@ece.utexas.edu).

Akbar M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA (e-mail: akbar@engr.wisc.edu).

Digital Object Identifier 10.1109/JSAC.2007.070804.

antenna dimensions [13]. Thus in the extreme case when there is a singularly dominant transmit eigen-mode and a receiver with a large number of antennas that see uncorrelated fading, statistical beamforming is near-optimal. On the other extreme when the transmitter sees uncorrelated statistics, but the receiver has a dominant eigen-mode, the performance loss between statistical beamforming and perfect CSI beamforming is expected to be large.

Based on the above motivation, we arrive at a design criterion that maximizes the average projection of the dominant right singular vector of the channel onto the limited feedback codebook. This criterion is met by a design that quantizes the local neighborhood around the dominant transmit eigen-modes and incorporates a quantization of the Grassmannian manifold for the rarer global perturbations of the singular vector from the statistical directions. An important contribution of this work is the construction of rotation and scaling maps that ease the implementation of the proposed codebook design. Besides, these maps can be characterized by few parameters and are hence particularly attractive in terms of codebook storage requirements. Numerical studies show that our scheme achieves near perfect CSI performance for a variety of correlated channel conditions, even with few bits of feedback. In the special case of a MISO setting, numerical studies indicate that our codebook design provides similar performance to that of [11], [12]. We develop our results using the separable (Kronecker) channel model and then provide generalizations to the more realistic canonical model for correlated MIMO channels [14], [15].

While in the case of time-division duplex systems the channel statistics can be learned via reverse link training, they *usually* have to be fed back in the case of frequency-division duplex systems. Thus it is important not to ignore the cost of statistical feedback in the design of practical communication systems. The earlier design assumes that the channel statistics remain static for a sufficiently long duration so that the receiver can feed back the statistical information at essentially zero cost. On the other hand, the channel statistics could vary at a fast rate, *e.g.,* highly mobile scenarios, or that accurate statistical feedback is not possible. In this situation, we propose a universal codebook design that is robust to channel statistics, very correlated through near i.i.d. scenarios. The construction of this universal codebook again hinges on the rotation and scaling maps, the building blocks behind our previous design.

This paper is organized as follows. The system setup is provided in Section II. The limited feedback preliminaries and the channel modeling framework for design when channel statistics are known at the transmitter are provided in Section III. The intuition for the systematic designs is obtained by studying the average received SNR loss in statistical beamforming when compared to perfect CSI beamforming in Section IV. In Section V, we propose a systematic codebook design when channel statistics are known at the transmitter and later in Section VI, we elucidate the universal design.

*Notation*: We use $X(i,j)$ and $X(i)$ to denote the $i,j$-th and $i$-th diagonal entries of a matrix $X$, $(.)^H$ to denote conjugate transposition, $\|\cdot\|$ for the Frobenius norm, $E[\cdot]$ to denote expectation, $\chi(\cdot)$ to denote the indicator function of a set, $\mathcal{CN}(\mu,\sigma^2)$ to denote the complex normal distribution with mean $\mu$ and variance $\sigma^2$, and $\mathbb{C}^t$ for the $t$-dimensional complex vector space. We also use the standard big-Oh ($\mathcal{O}$) and little-oh ($o$) notations.

We use $\lambda_\bullet^t$, $\lambda_\bullet^r$ and $\lambda_\bullet$ to denote the eigenvalues of the transmit and receive covariance matrices ($\Sigma_t = E[\mathbf{H}^H\mathbf{H}]$ and $\Sigma_r = E[\mathbf{H}\mathbf{H}^H]$) and $\mathbf{H}^H\mathbf{H}$ respectively under the assumption that they are arranged in the non-increasing order, that is, $\lambda_1^t \geq \lambda_2^t$ *etc*. Unit vectors $\mathbf{u}_i$ (dependent on channel statistics) and $\mathbf{v}_i$ (dependent on instantaneous realizations), $i = 1, \cdots, N_t$ are used to denote eigenvectors corresponding to $\lambda_i^t$ and $\lambda_i$, respectively. To stress the statistical nature of $\mathbf{u}_1$, we will also use $\mathbf{u}_{\rm stat}$ often. If $\frac{\lambda_1^t}{\lambda_{N_t}^t}$ is (or is not) significantly larger than 1, we loosely say that $\Sigma_t$ is ill-(or well-)conditioned.

## II. Limited Feedback Setup

We consider a single user communication system with $N_t$ transmit and $N_r$ receive antennas employing transmit beamforming and receive combining to maximize the diversity gain. Assuming a narrowband channel, the discrete-time baseband signal model at time instant $k$ is

$$y[k] = \mathbf{z}^H[k]\ \mathbf{H}[k]\ \mathbf{w}[k]\ x[k] + \mathbf{z}^H[k]\ \mathbf{n}[k], \qquad (1)$$

where $x[k]$ is the transmitted symbol, $\mathbf{H}[k]$ is the $N_r \times N_t$ channel matrix connecting the transmitter and the receiver and $y[k]$ is the combined observation at the receiver, all at time $k$. We assume a block-stationary model in which the channel realization remains fixed within a coherence time duration $T_{\rm coh}$ and based on the channel statistics, it then fades independently from block to block. Further, we assume that the channel statistics remain static over a duration $T_{\rm stat}$ where $T_{\rm stat} = L\,T_{\rm coh}$. The beamforming and combining vectors are given by $\mathbf{w}[k]$ and $\mathbf{z}[k]$, respectively. The additive white noise at the receiver is denoted by $\mathbf{n}[k]$ with independent and identically distributed (i.i.d.) entries from $\mathcal{CN}(0,N_0)$. Further, the symbol energy is given by $\mathcal{E}_t = E|x[k]|^2$. Henceforth, when no confusion can arise, we will skip the discrete-time indices to simplify the notation.

We assume that $\mathbf{H}$ is perfectly known at the receiver, and that the system employs maximal ratio transmission and maximal ratio combining to maximize the received SNR. The optimal beamforming and combining vectors are given by [1]

$$\mathbf{w}_{\rm opt} = \arg \max_{\mathbf{w}\,\in\,\mathbb{C}^{N_t},\,\|\mathbf{w}\|=1} \|\mathbf{H}\,\mathbf{w}\|, \mathbf{z}_{\rm opt} = \frac{\mathbf{H}\,\mathbf{w}_{\rm opt}}{\|\mathbf{H}\,\mathbf{w}_{\rm opt}\|}, \quad (2)$$

where $\|\mathbf{w}\| = 1$ reflects the power constraint at the transmitter. It follows from (2) that an optimal beamforming vector is a right singular vector corresponding to the largest singular value of $\mathbf{H}$. Note that this optimal vector is not unique as it is invariant to transformations of the form $\mathbf{w}_{\rm opt} \mapsto e^{j\phi}\mathbf{w}_{\rm opt}$, where $\phi \in \mathbb{R}$. The non-uniqueness can be incorporated by considering the quotient space[1] of all unit vectors under the above map.

To realize the maximum value of received SNR, it is evident that the transmitter must have knowledge of the dominant right singular vector of $\mathbf{H}$, which is extremely difficult. This

[1]The space of unit vectors is also known as the Stiefel manifold, and the quotient space, the Grassmann manifold. See Section V-B for details.

motivates the general approach of quantized beamforming or limited feedback [3]. We assume that a low-bandwidth, error-free, zero-delay feedback channel exists between the transmitter and the receiver and is used to convey quantized channel information to the transmitter. Codebook-based schemes proposed earlier in [3], [4] use a fixed codebook $\mathcal{W}$ available at the transmitter and the receiver to enable feedback of the optimal quantized beamforming vector. Specifically, the index of the codeword (from the codebook) that maximizes a performance metric like received SNR or outage probability is fed back to the transmitter. Current works study the impact of codebooks designed for channels with uncorrelated fading, when applied to correlated channels [16].

## III. Codebook Design Based on Channel Statistics: Preliminaries

The focus in this paper is on channels that are spatially correlated. The statistics of the channel $\mathbf{H}$ depend on the frequency of operation, physical propagation environment which controls the angular spreading function and the path distribution, antenna geometry (arrangement and spacing) *etc.* It has been well-documented that the assumption of zero mean Rayleigh fading is an accurate model for $\mathbf{H}$ in a non line-of-sight setting, but the commonly used i.i.d. model where the channel entries are i.i.d. is not accurate in describing realistic propagation environments. The complete channel statistics are thus described by the second-order moments.

### A. Statistical Channel Modeling

We now briefly review the different channel modeling frameworks to model the second-order statistics of the channel. Ideal channel modeling assumes that the entries of $\mathbf{H}$ are i.i.d. Gaussian random variables [17], [18]. Performance 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.

A more realistic channel model is the often-used separable correlation model where the correlation of channel entries is in the form of a Kronecker product of the transmit and receive covariance matrices. The channel is given by

$$\begin{aligned}\mathbf{H} &= \frac{1}{\sqrt{E\left[\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]}}\ \Sigma_r^{1/2}\ \mathbf{H}_{\mathrm{iid}}\ \Sigma_t^{1/2} \\ &\overset{(a)}{=} \frac{1}{\sqrt{E\left[\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]}}\ \mathrm{U}_r\ \Lambda_r^{1/2}\ \mathbf{H}_{\mathrm{iid}}\ \Lambda_t^{1/2}\ \mathrm{U}_t^H \end{aligned} \quad (3)$$

where $\Sigma_r = E\left[\mathbf{H}\mathbf{H}^H\right] = \mathrm{U}_r\Lambda_r\mathrm{U}_r^H$ and $\Sigma_t = E\left[\mathbf{H}^H\mathbf{H}\right] = \mathrm{U}_t\Lambda_t\mathrm{U}_t^H$ correspond to receive and transmit covariance matrices, respectively (alongwith their respective eigen-decompositions), $\mathbf{H}_{\mathrm{iid}}$ is an i.i.d. random matrix, and (a) follows from the isotropicity of an i.i.d. channel under a unitary transformation. The above normalization results in $\mathrm{Tr}\left(\Sigma_t\right) = \mathrm{Tr}\left(\Sigma_r\right) = E\left[\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]$.

The separable model, despite its mathematical tractability in terms of simplification of performance analysis of MIMO systems, suffers from deficiencies acquired by the separability property of channel correlation [19], [20]. Measurement campaigns have shown that the separable model is accurate in capturing the underlying channel statistics under certain conditions [21], but in general, the model results in misleading estimates of system performance [19], [14].

Various statistical models have been proposed to overcome the structural deficiency associated with the separable model. A canonical decomposition of the channel along the transmit and receive covariance bases [14], [15] has been shown to be a better fit than the separable model in predicting system metrics like capacity, probability of error *etc.,* [19], [15], [22]. The canonical 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 $\mathbf{H}$ as

$$\mathbf{H} = \mathrm{U}_r\,\mathbf{H}_{\mathrm{ind}}\,\mathrm{U}_t^H \quad (4)$$

where $\mathbf{H}_{\mathrm{ind}}$ has independent, but not necessarily identically distributed entries, and $\mathrm{U}_r$ and $\mathrm{U}_t$ are unitary matrices. The receive and transmit covariance matrices are given by $\Sigma_r = E\left[\mathbf{H}\mathbf{H}^H\right] = \mathrm{U}_r\,E\left[\mathbf{H}_{\mathrm{ind}}\mathbf{H}_{\mathrm{ind}}^H\right]\mathrm{U}_r^H = \mathrm{U}_r\Lambda_r\mathrm{U}_r^H$ and $\Sigma_t = E\left[\mathbf{H}^H\mathbf{H}\right] = \mathrm{U}_t\,\Lambda_t\,\mathrm{U}_t^H$, where $\Lambda_r = E\left[\mathbf{H}_{\mathrm{ind}}\mathbf{H}_{\mathrm{ind}}^H\right]$ and $\Lambda_t = E\left[\mathbf{H}_{\mathrm{ind}}^H\mathbf{H}_{\mathrm{ind}}\right]$ are diagonal. Under the assumption of separable channel statistics, the canonical model reduces to the Kronecker model. We now work towards systematic codebook designs for quantized beamforming with separable and canonical models, with the statistical information known at the transmitter. We note that for uniform linear arrays, the canonical model reduces to the virtual representation [23] with $\mathrm{U}_r$ and $\mathrm{U}_t$ replaced by DFT matrices.

### B. Distortion Metric for Reliability Loss

One of the main attractions of a transmit beamforming/receive combining scheme is the achievability of the full diversity (reliability) gain afforded by a MIMO channel. We now restate a well-known result [12], [9] which quantifies the performance gap between quantized beamforming and perfect CSI beamforming.

*Lemma 1 (Distortion Metric):* Let $D(\mathcal{W})$ be defined by $D(\mathcal{W}) \triangleq E_{\mathbf{H}}\left[\lambda_1 - \max_{\mathbf{w}\in\mathcal{W}} ||\mathbf{H}\mathbf{w}||_2^2\right]$ where $\lambda_1$ denotes the largest eigenvalue of $\mathbf{H}^H\mathbf{H}$ and the normalized[2] received SNR in a system with perfect CSI and with quantized feedback are given by $\lambda_1$ and $\max_{\mathbf{w}\in\mathcal{W}} ||\mathbf{H}\mathbf{w}||_2^2$, respectively. Thus the distortion metric measures the average normalized received SNR loss. Moreover, if there exists a family of codebooks (*e.g.,* a sequence of codebooks with number of bits of feedback $B \to \infty$) such that $D(\mathcal{W}) \to 0$, then the probability of error with quantized beamforming converges to that with perfect CSI beamforming. In addition to capturing the error probability, if $D(\mathcal{W}) \to 0$, then the average mutual information achieved by the quantized beamforming scheme converges to that of the perfect CSI beamforming case. ■

The utility of $D(\mathcal{W})$ would be considerably enhanced if a mathematically tractable and tight upper-bound to it can be obtained. In the case of an i.i.d. channel, this goal is simplified since the singular values and singular vectors of $\mathbf{H}$ are independent. This independence is no longer true in a correlated setting. Despite this difficulty, we propose a useful upper bound to $D(\mathcal{W})$ in the following proposition. See Appendix A for proof.

[2]The received SNR can be written as $\lambda_1 \frac{\mathcal{E}_t}{N_0}$.

*Proposition 1:* The distortion measure $D(\mathcal{W}) = E_{\mathbf{H}}\left[\lambda_1 - \max_{\mathbf{w}\in\mathcal{W}} ||\mathbf{H}\mathbf{w}||_2^2\right]$ can be upper bounded by

$$D_u(\mathcal{W}) \triangleq \underbrace{\sqrt{1 - E_{\mathbf{H}}(P)}}_{D_u^{code}(\mathcal{W})} \underbrace{\left[E_{\mathbf{H}}(\lambda_1) + \sqrt{2\,\mathrm{Var}_{\mathbf{H}}(\lambda_1)}\right]}_{D_u^{chan}(\mathcal{W})} \quad (5)$$

where $\mathbf{v}_1$ is an eigenvector corresponding to $\lambda_1$ and $P$ is a measure of the projection of $\mathbf{v}_1$ onto the codebook and is given by $P = \max_{\mathbf{w}\in\mathcal{W}} |\mathbf{v}_1^H \mathbf{w}|^2$. ■

First, note that $D_u(\mathcal{W})$ complements the upper bound on distortion in the i.i.d. channel case, $E_{\mathbf{H}}(\lambda_1)(1 - E_{\mathbf{H}}(P))$, see *e.g.,* equation (18) in [3]. The upper bound $D_u(\mathcal{W})$ is made of two components: 1) $D_u^{code}(\mathcal{W}) = \sqrt{1 - E_{\mathbf{H}}(P)}$ representing the effect of codebook design and 2) $D_u^{chan}(\mathcal{W}) = E_{\mathbf{H}}(\lambda_1) + \sqrt{2\,\mathrm{Var}_{\mathbf{H}}(\lambda_1)}$ representing the channel "amplification." *The overall design objective of a quantized beamforming scheme is thus the construction of a (channel statistics dependent) codebook so as to maximize the average projection* $E_{\mathbf{H}}(P)$.

Note that the lack of an i.i.d. property on the channel entries implies that we have a looser upper bound to $D(\mathcal{W})$ than in the i.i.d. case. However, analogous to the i.i.d. case, $D_u^{code}(\mathcal{W})$ captures the fundamental trend in limited feedback codebook design. It must be emphasized that our conclusions are obtained by studying $D_u(\mathcal{W})$ because of a lack of closed-form expressions for $D(\mathcal{W})$. It is remarkable that as long as $\lambda_1^t$, the largest eigenvalue of $\Sigma_t$, is distinct (that is, $\lambda_1^t > \lambda_2^t$), just statistical beamforming results in $E_{\mathbf{H}}(P) \to 1$ as $N_r \to \infty$. This is the content of the following discussion.

## IV. Performance Loss with Statistical Beamforming

In this section, we study the case of statistical beamforming [24], [25], [26] where the transmit symbol is beamformed along the statistically dominant[3] eigen-direction, that is, the codebook $\mathcal{W}$ is fixed with only one codeword, $\mathbf{u}_{\mathrm{stat}}$. This codebook requires "almost" zero feedback as it has to be updated only on time scales of statistical variations. We show that in the limit of receive antenna dimensions $N_r$, *just* knowledge of $\mathbf{u}_{\mathrm{stat}}$ is sufficient to achieve near perfect CSI performance, in other words, the error probability enhancement with statistical beamforming is negligible in the receive antenna asymptotics.

### A. *Received* SNR *Loss with Statistical Beamforming*

We assume that the channel $\mathbf{H}$ is characterized by a transmit covariance matrix $\Sigma_t \neq \mathrm{I}$ and that $\lambda_1^t > \lambda_2^t$. Measurement campaigns indicate that most realistic channels satisfy the above assumptions with the extreme case of $\lambda_2^t = 0$ corresponding to a highly non-isotropic scattering environment [27]. Using eigenvector perturbation theory, we now study the loss in reliability with statistical beamforming in the limit of receive dimensions. This analysis is eased by the observation that for large $N_r$, the dominant right singular vector of $\mathbf{H}$, $\mathbf{v}_1$, is close to $\mathbf{u}_{\mathrm{stat}}$ with a high probability. We state the main results of this section, first with the separable model and then with the canonical model. See Appendices B and C for proofs.

*Theorem 1:* Let $\mathbf{H}$ be given by the Kronecker model and let the eigenvalues of $\Sigma_t$ and $\Sigma_r$ be ordered as $\lambda_1^t \geq \cdots \geq \lambda_{N_t}^t$ and $\lambda_1^r \geq \cdots \geq \lambda_{N_r}^r$. Assume that $\lambda_1^t > \lambda_2^t\left(1 + \frac{2}{\mu_{r,1} N_r^a}\right)$ for some $a > 0$ and define $\mu_{r,1}$, $\mu_{r,2}$, $\mathrm{Gap}_{t,12}$ and $\mathrm{Gap}_t$ as follows:

$$\mu_{r,2} \triangleq \frac{\sum_{i=1}^{N_r} (\lambda_i^r)^2}{N_r}, \quad \mu_{r,1} \triangleq \frac{\sum_{i=1}^{N_r} \lambda_i^r}{N_r}, \quad \mathrm{Gap}_{t,12} \triangleq 1 - \frac{\lambda_2^t}{\lambda_1^t},$$

$$\mathrm{Gap}_t \triangleq \frac{1}{N_t - 1} \sum_{i=2}^{N_t} \frac{x_i}{(1 - x_i)^2}, \quad x_i = \frac{\lambda_i^t}{\lambda_1^t}. \quad (6)$$

Then,

$$\begin{aligned}\left(D_u^{code}(\mathcal{W})\right)^2 &\leq \left(K_1 \cdot \mathrm{Gap}_t \cdot \frac{\mu_{r,2}}{(\mu_{r,1})^2}\right) \cdot \frac{N_t \log(N_r)}{N_r} \\ &\leq K_1 \cdot \frac{\mu_{r,2}}{(\mathrm{Gap}_{t,12}\, \mu_{r,1})^2} \cdot \frac{N_t \log(N_r)}{N_r} \quad (7)\end{aligned}$$

where $K_1 > 0$ is a constant independent of $\Sigma_t$, $\Sigma_r$, $N_t$ and $N_r$. ■

Note that $\mu_{r,1}$ and $\mu_{r,2}$ are the first and the (non-centralized) second moments of $\{\lambda_i^r\}$ and $\mathrm{Gap}_t$ is a measure of the gap (or separation) between $\{\lambda_i^t\}$. Since $\mathrm{Gap}_t$ is dependent on all the $\lambda_i^t$, we propose a simplified constant $1/\left(\mathrm{Gap}_{t,12}\right)^2$ which is a measure of the well-conditioning of $\Sigma_t$ with a hold on $\lambda_1^t$ and $\lambda_2^t$ alone.

*Theorem 2:* Let $\mathbf{H}$ be given by the canonical model and let $\sigma_{ij}^2$ represent the variances of $\mathbf{H}_{\mathrm{ind}}(i,j)$. Then $\Lambda_t(i) = \sum_k \sigma_{ki}^2$ and $\Lambda_r(j) = \sum_k \sigma_{jk}^2$. Assume that $\Lambda_t(1) \geq \cdots \geq \Lambda_t(N_t)$ and $\Lambda_r(1) \geq \cdots \geq \Lambda_r(N_r)$ and define $\mathrm{Gap}_t^c$ and $\mu_{r,2}^c$ as follows:

$$\mathrm{Gap}_t^c \triangleq \frac{1}{N_t - 1} \sum_{i=2}^{N_t} \frac{N_r^2}{(\Lambda_t(1) - \Lambda_t(i))^2}, \; \mu_{r,2}^c \triangleq \max_{i>1} \frac{\sum_k \sigma_{ki}^2 \sigma_{k1}^2}{N_r}.$$

If $\frac{\Lambda_t(1)}{N_r} > \frac{\Lambda_t(2)}{N_r} + \frac{2}{N_r^a}$ for some $a > 0$, then there exists a $K_2$ such that

$$\left(D_u^{code}(\mathcal{W})\right)^2 \leq (K_2 \cdot \mathrm{Gap}_t^c \cdot \mu_{r,2}^c) \cdot \frac{N_t \log(N_r)}{N_r}. \quad ■$$

The constants $\mathrm{Gap}_t^c$ and $\mu_{r,2}^c$ are analogous to $\mathrm{Gap}_t$ and $\mu_{r,2}$ proposed in Theorem 1 for Kronecker channels.

### B. *Implications*

**Sufficiency of Statistics in Asymptotics of $N_r$:** Since $D_u^{code}(\mathcal{W}) = \sqrt{1 - E_{\mathbf{H}} |\mathbf{v}_1^H \mathbf{u}_{\mathrm{stat}}|^2}$ and Theorems 1 and 2 show that $\lim_{N_r\to\infty} D_u^{code}(\mathcal{W}) = 0$, we can conclude the following: As long as there is a distinct dominant transmit eigen-direction, *just* increasing $N_r$ "hardens" [24] the dominant right singular vector of $\mathbf{H}$, $\mathbf{v}_1$, to that of the statistical direction. Thus as $N_r$ increases, the gain obtained via a knowledge of perfect CSI cannot be significantly better than that obtained via a knowledge of *only* the channel statistics and in this asymptotic case, there is no need to construct a limited feedback codebook. Besides, numerical studies suggest faster rates of convergence than those proved in Theorems 1 and 2.

[3]By the statistically dominant eigen-direction, we mean along $\mathbf{u}_{\mathrm{stat}}$, an eigenvector corresponding to $\lambda_1^t$.

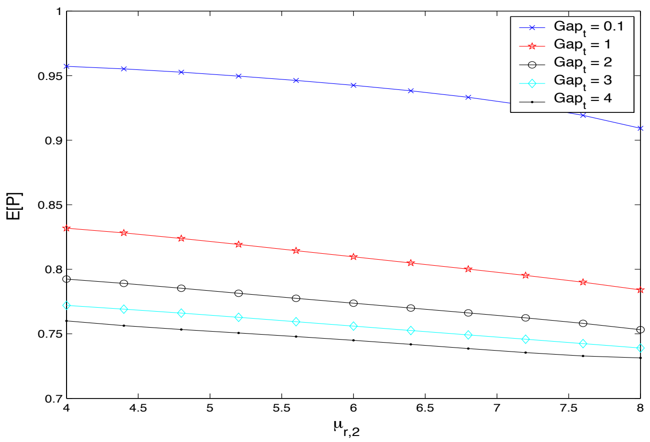


Fig. 1. Impact of $\mu_{r,\,2}$ on $E[P] = E[|\mathbf{u}_{\text{stat}}^H \mathbf{v}_1|^2]$ for a family of $2 \times 2$ channels with a separable model and different values of $\text{Gap}_t$.

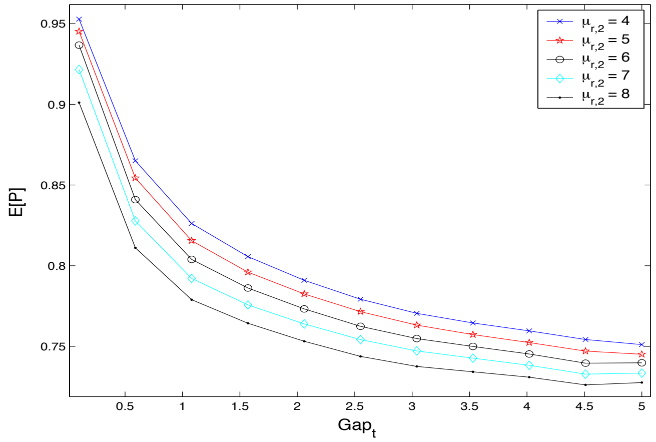


Fig. 2. Impact of $\text{Gap}_t$ on $E[P] = E[|\mathbf{u}_{\text{stat}}^H \mathbf{v}_1|^2]$ for a family of $2 \times 2$ channels with a separable model and different values of $\mu_{r,\,2}$.

**Non-Asymptotic Case:** We first consider the separable channel case. The constants $\mu_{r,\,2}$ and $\text{Gap}_t$ in Theorem 1 also address the performance loss when $N_r$ is not asymptotically large. It is easy to see that, since $\sum_k \lambda_k^r = E\left[\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]$, $\mu_{r,\,2}$ is minimized by the extreme in well-conditioning of $\Sigma_r$: $\Sigma_r = \frac{E\left[\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]}{N_r} \mathbf{I}_{N_r}$. Similarly, a large value of $\text{Gap}_t$ is equivalent to many of the $\lambda_i^t$ close to $\lambda_1^t$, that is, well-conditioning of $\Sigma_t$.

For the canonical model, a large value of $\text{Gap}_t^c$ reflects the closeness of $\Lambda_t(1)$ with respect to $\Lambda_t(i)$ for many values of $i$, or well-conditioning of $\Sigma_t$. On the other hand, it is very difficult to obtain insights on the conditioning of $\Sigma_r$ with $\mu_{r,\,2}^c$. Therefore, we define $\mu_{r,\,2}^{c,\,ub} \triangleq \left(\frac{\sum_k \sigma_{k1}^4}{N_r}\right)^{\frac{1}{2}} \max_{i>1} \left(\frac{\sum_k \sigma_{ki}^4}{N_r}\right)^{\frac{1}{2}}$. A standard application of the Cauchy-Schwarz inequality implies that $\mu_{r,\,2}^c \leq \mu_{r,\,2}^{c,\,ub}$. Using the slightly looser bound $\mu_{r,\,2}^{c,\,ub}$, it is easy to infer that $\max_i \frac{\sum_k \sigma_{ki}^4}{N_r}$ has to be minimized for near perfect CSI performance. By considering $\sum_{ki} \sigma_{ki}^4$ instead of the maximum, and using the fact that $\sum_{ki} \sigma_{ki}^2 = E\left[\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]$, as above, we see that well-conditioning of $\Sigma_r$ helps this goal. Thus we conclude that

- Ill-conditioning of $\Sigma_r$ and well-conditioning of $\Sigma_t$ slow down the rate of dominant eigenvector hardening (or concentration), and
- If $\Sigma_r$ is well-conditioned and $\Sigma_t$ is ill-conditioned, even statistical information is sufficient to achieve near perfect CSI performance. Ill-conditioning of $\Sigma_t$ enhances the likelihood that a dominant eigen-direction lies near the statistical peak while the well-conditioning of $\Sigma_r$ ensures that the perturbations of the dominant eigen-direction are minimized. Of these two factors, the more important contribution is that of $\Sigma_t$ as it has a direct bearing on the hardening process.

These conclusions are illustrated in Figures 1 and 2 where $E[|\mathbf{v}_1^H \mathbf{u}_{\text{stat}}|^2]$ is plotted (for a family of $2 \times 2$ channels with separable model) as a function of $\mu_{r,\,2}$ (for different values of $\text{Gap}_t$) and $\text{Gap}_t$ (for different values of $\mu_{r,\,2}$), respectively. We normalize the channel power so that $\mu_{r,\,1} = \frac{E\left[\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]}{N_r} = N_t$. Thus, $4 = N_t^2 \leq \mu_{r,\,2} \leq N_t^2 N_r = 8$, but $\text{Gap}_t$ can take any value in the open set $(0, \infty)$. Comparing Figures 1 and 2, we note that the most importance factor in eigenvector hardening is the degree of ill-conditioning of $\Sigma_t$.

Intuitively, the distinctness of $\lambda_1^t$ and the independence of entries in $\mathbf{H}_{\text{iid}}$ (and $\mathbf{H}_{\text{ind}}$) lead to hardening of $\mathbf{v}_1$ along $\mathbf{u}_{\text{stat}}$. The distinctness assumption of $\lambda_1^t$ is crucial to this intuition, for if this were not the case, then the singular vector is equally likely to point in any direction (isotropically distributed).

**Covariance Space Partitioning:** Our work, so far, highlights the importance of partitioning the transmit and receive eigen-spaces (based on their condition numbers) so as to gain an insight into the impact of feedback on achieving near perfect CSI performance. Based on this principle, we partition the eigen-spaces into four regions: 1) Well-conditioned $\Sigma_r$ and Ill-conditioned $\Sigma_t$, 2) Well-conditioned $\Sigma_r$ and $\Sigma_t$, 3) Ill-conditioned $\Sigma_r$ and $\Sigma_t$, and 4) Ill-conditioned $\Sigma_r$ and Well-conditioned $\Sigma_t$. In the first region, statistical information should lead to near-optimal performance and in terms of feedback requirement $F_{\bullet}$, to achieve a particular operational level of performance, we expect the following ordering relation: $F_1 < \{F_2, F_3\} < F_4$, that is, the feedback requirement should be the least for channels in Region 1 and maximum in Region 4 while that for channels in Regions 2 and 3 lie in between. Figures 1 and 2 lend credence to this observation.

## V. Codebook Design for Correlated Channels

Initial work [3], [4] on quantized beamforming codebook design in i.i.d. channels led to a design metric of maximizing the minimum distance between codewords and this resulted in a natural Grassmannian line packing (GLP) solution, thanks to the isotropicity of the maximal right singular vector. The GLP problem, where an optimal packing of $N > N_t$ vectors in the Grassmann manifold is desired, has been well-studied in the mathematical literature. Closed-form packings are known only for certain values of $N$ and $N_t$ [28]; in other cases numerical solutions are required. Vector quantization (VQ) [2], [9] is one such numerical codebook design technique, applicable for all $N$ or $N_t$, where the codewords are chosen via a Lloyd-type algorithm. Another commonly used design technique is that of the theory of equiangular frames [29].

In [11], [12], the i.i.d. channel solution was leveraged by a transmitter statistics induced transformation of the GLP (or VQ) codebook and was shown to offer improved performance over the GLP codebook. The disadvantages of this scheme are that it is constrained to a multiple transmitter, *single* receiver system and requires correlation at the transmitter. Although the scheme provides substantial gains, the construction is ad-hoc in the sense that there is no theoretical underpinning behind the construction. In this section, we assume that the channel statistics remain static for a *long* duration so that the transmitter and receiver can estimate this information at essentially zero cost. We address the problem of a systematic codebook construction for a general MIMO system in the non-asymptotic setting (see Section IV-B) where there may not be a singularly dominant eigen-direction and when the receive covariance matrix is ill-conditioned. We first motivate a codebook design heuristic and follow up with a systematic rotation and scaling-based local Grassmannian packing.

### A. Design Heuristic

In the non-asymptotic regime (of $N_r$), as can be seen from Figures 1 and 2, $E\left[\left|\mathbf{v}_1^H \mathbf{u}_{\rm stat}\right|^2\right]$ is usually sufficiently large even if it may not be close to 1. What this means is that with a high probability, the dominant right singular vector of $\mathbf{H}$ points in the direction corresponding to that given by the channel statistics. The discrepancy of this average from 1 is the occurrence of two events: 1) There are channel realizations where $\mathbf{v}_1$ points locally around $\mathbf{u}_{\rm stat}$ but not in the precise direction, and 2) There are some channel realizations, albeit relatively rare, where the angle between $\mathbf{v}_1$ and $\mathbf{u}_{\rm stat}$ is large. The probability that these two events occur is small and it converges to 0 in the asymptotics of $N_r$ as seen from Theorem 1. It is also important to note that reliable data reception at high SNR is governed by these events. Thus any codebook design optimized to minimize the probability of error should account for these channel realizations. We now propose a systematic design that has three components: $D, L$ and $G$ accounting for the *dominant* eigen-direction(s), *local* perturbations and *global* perturbations respectively.

### B. Systematic Construction

*1) Notations and Definitions:* We now establish the notations that would be needed in the systematic construction by briefly recalling some well-known facts about the Grassmann manifold. The unit sphere in $\mathbb{C}^{N_t}$, also known as the unidimensional complex Stiefel manifold $\mathrm{St}(N_t, 1)$, is defined as $\Omega_{N_t} = \left\{\mathbf{x} \in \mathbb{C}^{N_t} : \|\mathbf{x}\| = 1\right\}$. The invariance of any vector $\mathbf{x}$ to transformations of the form $\mathbf{x} \mapsto e^{j\phi}\mathbf{x}$, which is invisible to the optimizations of the type in (2), is incorporated by considering vectors to be points on $G(N_t, 1)$, the unidimensional complex Grassmann manifold that consists of the set of one-dimensional subspaces of $\Omega_{N_t}$.

A survey of literature on Grassmann manifold [30] shows that numerous distance metrics can be defined on $G(N_t, 1)$, however, we will content ourselves with the chordal distance $d(\cdot,\cdot)$, defined as $d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{1 - |\mathbf{x}_1^H \mathbf{x}_2|^2}$. Using this distance metric, for any $\gamma < 1$, we can define a *spherical cap* on $G(N_t, 1)$ with center $\mathbf{o}$ and radius $\gamma$ as the open set $\mathbb{O}(\mathbf{o}, \gamma) = \{\mathbf{x} \in G(N_t, 1) : d(\mathbf{x}, \mathbf{o}) < \gamma\}$. A spherical cap on $G(N_t, 1)$ induces a spherical cap on $\Omega_{N_t}$ via the equivalence partitioning described above.

*2) Rotation and Scaling Maps for Spherical Caps:* Since the dominant right singular vector of a correlated channel is localized in $G(N_t, 1)$, in contrast to an i.i.d. channel where it is isotropically distributed in $G(N_t, 1)$, an important component of a limited feedback codebook tailored to correlated channels is the quantization of a localized region (or a proper subset) of $G(N_t, 1)$. We will formalize the notion of a "local packing" very shortly. Theoretically, we can use VQ to design these local packings. However, this method suffers from two major disadvantages: storage and complexity. If VQ is used to design a codebook with many such local components, the local components have to be stored individually thereby increasing the storage requirements. Also, these components have to be designed individually and this increases the implementation complexity. From a practical perspective, the VQ-based design problem is more difficult to handle since channel statistics could change over the time-scales of codebook design.

To overcome these disadvantages, we now propose two operations (maps) that are restricted to a given spherical cap (or a finite-element subset of it). In particular, let us restrict attention to a spherical cap $\mathbb{O}(\mathbf{o}, \gamma)$ on $G(N_t, 1)$ for some $\gamma < 1$. The maps serve to 1) rotate the cap to $\mathbb{O}(\mathbf{o}_{\rm rot}, \gamma)$ and 2) shrink it to $\mathbb{O}(\mathbf{o}, \alpha\gamma)$ for some $\alpha < 1$. The two operations can then be combined to result in $\mathbb{O}(\mathbf{o}_{\rm rot}, \alpha\gamma)$. Note that the shrinking operation requires some care due to the constraints of the space. For example, an operation of the form $\mathbf{x} \mapsto \alpha\mathbf{x}$ where $\alpha \in \mathbb{R}$ yields a vector that is not unit-norm. Once these maps have been defined, they can be applied to a single finite-element subset of a spherical cap repeatedly to generate a limited feedback codebook. Thus, the implementation complexity is reduced to that of designing *only* one VQ-based local packing. Besides, the storage requirements are also reduced substantially as the maps can be characterized by few parameters.

*Proposition 2 (Rotation):* The map $r : \mathbb{O}(\mathbf{o}, \gamma) \mapsto \mathbb{O}(\mathbf{o}_{\rm rot}, \gamma)$ defined as

$$r(\mathbf{x}) = \mathbf{U}_{\rm rot}\, \mathbf{x} \triangleq \mathbf{x}_{\rm rot}, \tag{8}$$

where $\mathbf{U}_{\rm rot}$ is a unitary matrix that satisfies $\mathbf{o}_{\rm rot} = \mathbf{U}_{\rm rot}\, \mathbf{o}$ effects the necessary rotation.

*Proof:* For all $\mathbf{x} \in \mathbb{O}(\mathbf{o}, \gamma)$, check that $\mathbf{x}^H\mathbf{o} = \mathbf{x}_{\rm rot}^H \mathbf{o}_{\rm rot}$ and thus $d(\mathbf{o}_{\rm rot}, \mathbf{x}_{\rm rot}) = d(\mathbf{o}, \mathbf{x})$. Hence $\mathbf{x}_{\rm rot} \in \mathbb{O}(\mathbf{o}_{\rm rot}, \gamma)$. ■

The rotation map is illustrated in Figure 3. Since the dimensionality of an $N_t \times N_t$ complex matrix over the field of reals is $2N_t^2$ and there are $N_t^2 + 2N_t$ real equations associated with the definition of $\mathbf{U}_{\rm rot}$, there are infinitely many solutions for $\mathbf{U}_{\rm rot}$ when $N_t > 2$. Any criterion used at the receiver to select $\mathbf{U}_{\rm rot}$ out of its possibilities is assumed to be known at the transmitter. This insures that the rotation map of a vector $\mathbf{x}$ at both the transmitter and the receiver results in the same $\mathbf{x}_{\rm rot}$. Also, once $\mathbf{U}_{\rm rot}$ is fixed, the inverse map $r^{-1}$ is generated by $\mathbf{U}_{\rm rot}^H$.

*Proposition 3 (Scaling):* Let $\alpha$ be such that $\alpha \in (0, 1)$. Following Proposition 2, define a rotation map $r_{\rm vertex}$ generated by a unitary matrix $\mathbf{U}_{\rm vertex}$ that effects the rotation of

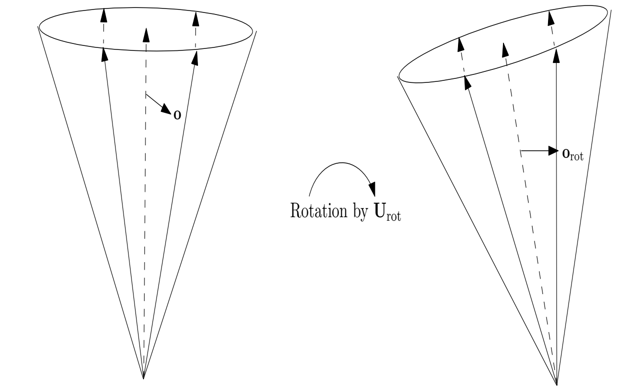


Fig. 3. Illustration of rotation of a five element subset of a spherical cap with center at $\mathbf{o}$ by $\mathbf{U}_{\rm rot}$ to generate a corresponding subset of a spherical cap with center at $\mathbf{o}_{\rm rot}$.

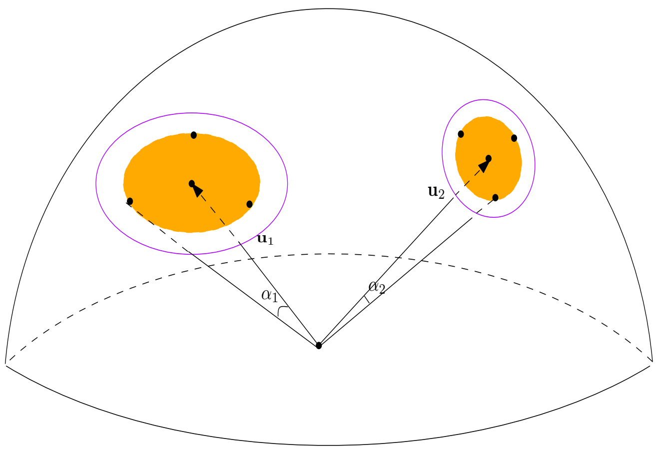


Fig. 4. A 3-bit codebook design with $M_D = 2$, $M_{L_1} = M_{L_2} = 3$ and $M_G = 0$. Recall that the feedback rate is $\log_2(M_D + M_{L_1} + M_{L_2} + M_G)$ bits. The centers of the local packings $\mathbf{u}_1$ and $\mathbf{u}_2$ and their corresponding scaling factors $\alpha_1$ and $\alpha_2$ are also shown.

$\mathbf{o}$ to a vertex of the unit cube. For simplicity[4], consider the vertex to be $\mathbf{o}_{\rm vertex} = [1, 0, \cdots, 0]^T$ and $\mathbf{o}_{\rm vertex} = \mathbf{U}_{\rm vertex}\, \mathbf{o}$. Then, define a vertex scaling map $s_{\rm vertex} : \mathbb{O}(\mathbf{o}_{\rm vertex}, \gamma) \mapsto \mathbb{O}(\mathbf{o}_{\rm vertex}, \alpha\gamma)$ by

$$s_{\rm vertex}\left([r_1 e^{j\theta_1}, r_2 e^{j\theta_2}, \cdots, r_{N_t} e^{j\theta_{N_t}}]^T\right) = \left[\sqrt{1-\alpha^2(1-r_1^2)}e^{j\theta_1}, \alpha r_2 e^{j\theta_2}, \cdots, \alpha r_{N_t} e^{j\theta_{N_t}}\right]^T,$$

where we have denoted the vector $\mathbf{x}$ in its polar form in the above equation. The map $s$ defined as a composition $s = r_{\rm vertex}^{-1} \circ s_{\rm vertex} \circ r_{\rm vertex}$ effects the necessary scaling.

*Proof:* First, we check that the map $s_{\rm vertex}$ is well-defined resulting in unit norm vectors. Then, we note that

[4] A simple choice of $\mathbf{U}_{\rm vertex}$ that effects this transformation is $\mathbf{U}_{\rm vertex}^T = \left[\begin{array}{cc} \mathbf{o}^\star & \mathbf{O}^\perp \end{array}\right]$ where $\mathbf{O}^\perp$ refers to a matrix representative from the $N_t \times (N_t - 1)$ dimensional null-space of $\mathbf{o}$, that is, the columns of $\mathbf{O}$ are unit norm and orthogonal to $\mathbf{o}$. It is crucial to note that for any such matrix $\mathbf{O}^\perp$ and (hence) any such choice of $\mathbf{U}_{\rm vertex}$, the first column *should* necessarily be $\mathbf{o}^\star$. This is useful in Step (e) of the ensuing proof.

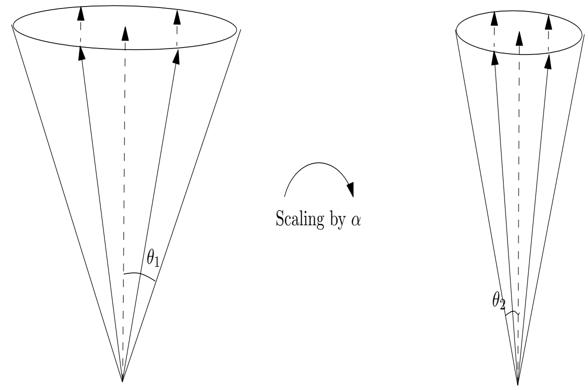


Fig. 5. Illustration of the scaling map (by an $\alpha < 1$). A five element subset of the spherical cap shown on the left side is scaled to generate the set on the right side. In this operation, the following relationship holds true: $\sin(\theta_2) = \alpha \sin(\theta_1)$.

$s_{\rm vertex}(\mathbf{o}_{\rm vertex}) = \mathbf{o}_{\rm vertex}$ and thus $s(\mathbf{o}) = \mathbf{o}$. We see that

$$\begin{aligned}
\left(d(s(\mathbf{x}), s(\mathbf{o}))\right)^2 &= 1 - |s(\mathbf{o})^H s(\mathbf{x})|^2 \;=\; 1 - |\mathbf{o}^H s(\mathbf{x})|^2 \\
&\overset{(a)}{=} 1 - |\mathbf{o}^H\, \mathbf{U}_{\rm vertex}^H\, s_{\rm vertex}(\mathbf{U}_{\rm vertex}\, \mathbf{x})|^2 \\
&\overset{(b)}{=} 1 - |\mathbf{o}_{\rm vertex}^H\, s_{\rm vertex}(\mathbf{U}_{\rm vertex}\, \mathbf{x})|^2 \\
&\overset{(c)}{=} 1 - |\left(s_{\rm vertex}(\mathbf{U}_{\rm vertex}\, \mathbf{x})\right)_1|^2 \\
&\overset{(d)}{=} \alpha^2(1 - |\left(\mathbf{U}_{\rm vertex}\, \mathbf{x}\right)_1|^2) \\
&\overset{(e)}{=} \alpha^2(1 - |\mathbf{o}^H \mathbf{x}|^2) \;=\; \alpha^2\left(d(\mathbf{x}, \mathbf{o})\right)^2
\end{aligned}$$

where (a), (b) and (c) follow from the definitions of $s(\cdot)$ and $\mathbf{o}_{\rm vertex}$, the 1 under the subscript refers to the first element of the array and (d) follows from the definition of $s_{\rm vertex}(\cdot)$ and (e) follows from Footnote 4 at the bottom of the previous column. Thus we are done. ■

The scaling map is illustrated in Figure 5. Note that $s_{\rm vertex}$ and $\mathbf{o}_{\rm vertex}$ could have been defined with respect to any of the $N_t$ coordinates. As with the rotation map, we assume that the choice of $s_{\rm vertex}$ is known to both the transmitter and the receiver, thus avoiding any ambiguity in codebook design. Also note that the scaling map can only reduce the radius of a spherical cap. Thus, by combining the rotation and the scaling maps, any spherical cap of radius $\gamma$ can be centered about any vector on $\Omega_{N_t}$ and its radius reduced to any value, smaller than $\gamma$.

*3) Construction:* We now rigorously define the notion of a local packing. A local Grassmannian packing (LGP) with parameters $N_t, N, \mathbf{o}, \gamma$ is a set of $N$ vectors, $\mathbf{x}_i$, $i = 1, \cdots, N$ (all different from $\mathbf{o}$), constrained to a spherical cap $\mathbb{O}(\mathbf{o}, \gamma)$ in $G(N_t, 1)$ such that $d(\mathbf{x}_i, \mathbf{o}) = \gamma$ for all $i$ and $\min_{i \neq j} d(\mathbf{x}_i, \mathbf{x}_j)$ is maximized. In other words, an LGP is a uniform quantization of a particular localized region (boundary of a spherical cap) in $G(N_t, 1)$. For simplicity, we will use "center of the LGP" to denote the vector $\mathbf{o}$.

Given any triplet $(M_D, M_L, M_G)$ of number of vectors, the design for correlated channels elucidated here leads to a codebook of $M_D + M_L + M_G$ codewords. The indices $D, L$ and $G$ stand for dominant, local and global respectively and will be explained shortly. We will not concern ourselves with the optimal choice of the triplet, but a systematic codebook design for a particular choice. The optimal choice depends on the trade-off between a larger codebook size and the gain in the probability of error performance and has to be studied more carefully. An understanding of the performance of the limited feedback system as a function of $B$ would be crucial towards this goal. Even in the simple i.i.d. case, such an understanding is as yet incomplete. It is natural to expect that the correlated channel case would be more difficult than the i.i.d. case. Due to the lack of closed-form expressions, we will restrict ourselves to some design guidelines in this work.

Step 0 - *Root Codeset*: An LGP of $N$ vectors (for some $N$ sufficiently large) and restricted to a cap $\mathbb{O}(\mathbf{o}, \gamma)$ for some $\gamma < 1$, also sufficiently large, forms the root codeset from which we derive the codebook design for correlated channels. Such a packing can be easily designed via VQ [2], [9]. The reason for choosing $N$ and $\gamma$ large is that if only a few vectors are needed in the design process, it is easier to scale down the size (by discarding the unused vectors) and the scaling map, by definition, can only reduce the radius of a sphere cap. For the sake of concreteness, we set $\gamma = 0.95$ and $N = 16$ even though the precise numbers can be adapted as necessary.

Step 1 - *Dominant Statistical Directions*: The first step in the codebook design is to identify the $M_D$ dominant eigenvectors of $\Sigma_t$, that is, eigenvectors $\mathbf{u}_i$ corresponding to the $M_D$ largest eigenvalues of $\Sigma_t$. As the family of channels range through Regions 1-4 (see Section IV-B), that is, as $\Sigma_t$ becomes more well-conditioned, $M_D$ increases. A heuristic we use in this work is to pick $M_D$ as $M_D = \arg\max_{i=1,\cdots,N_t} \frac{\lambda_i^t}{\lambda_1^t} > 0.2$. The threshold 0.2 is arbitrary and could be replaced by any appropriate choice.

Step 2 - *Local Perturbations*: The proposed codebook has $M_D$ local components and $\mathbf{u}_i, i = 1, \cdots, M_D$ form the centers of these local packings. These local packings are in turn generated from the root codeset (designed in Step 0) via the rotation and scaling maps of the previous section. More precisely, in this step, we pick $M_{L_i},\ i = 1, \cdots,\ M_D$ vectors around each dominant statistical direction, $\mathbf{u}_i$, to account for those channel realizations that steer the dominant singular vector in a local neighborhood of $\mathbf{u}_i$. The $\{M_{L_i}\}$ satisfy $M_L \triangleq \sum_{i=1}^{M_D} M_{L_i}$, and are non-increasing in $i$ (since the less dominant an eigenvector, the smaller the level of local perturbations around it that is relevant). The local packing for each $\mathbf{u}_i$ is generated by performing the rotation of the root codeset to $\mathbf{o}_{\rm rot} = \mathbf{u}_i$ and scaling it by $\alpha_i$ for some appropriate $\alpha_i$. Just as with $M_{L_i}$, $\alpha_i$ are non-increasing in $i$. Since we do not know the optimal values of $\alpha_i$, we will provide numerical illustrations later. Note that the rotated and scaled set has $N$ codevectors, where $N$ is the size of the root codeset, of which $M_{L_i}$ codevectors are picked[5] (and the rest discarded) via a protocol known at both the transmitter and the receiver. As $\Sigma_r$ and $\Sigma_t$ become more ill and well-conditioned respectively, $M_{L_i}$ increases.

Step 3 - *Global Perturbations*: Global perturbations are accommodated by a GLP (or VQ) solution of $M_G$ vectors in $N_t$ dimensions [28], [2], [29]. Similar to the case of local perturbations, $M_G$ increases as we move towards Region 4 of the eigen-space partitioning.

The design of a codebook which has only local codewords and no global codewords is illustrated in Figure 4.

*4) Description of the Feedback Scheme:* Recall that in the block fading model the channel statistics remain static for $T_{\rm stat} = LT_{\rm coh}$. The underlying assumption behind the design proposed here is that $L \gg 1$; if not, this is handled in the next section. The receiver uses a fraction $T_{\rm train}$ of $T_{\rm stat}$ coherence blocks to estimate specific statistical information about the channel using training data and conveys this information to the transmitter. The cost of this statistical information transfer is "essentially zero" if $L$ is large.

Since the transmitter and the receiver have knowledge of channel statistics at the end of the training phase, following the above design methodology, a well-defined (and identical) codebook can be constructed at both ends. At the starting of every new coherence block in the post-training phase, the receiver feeds back the index of the codeword from the codebook that maximizes the received SNR. The complexity of this search is directly proportional to the size of the codebook. The transmitter, upon knowledge of the optimal codeword, beamforms the transmit symbol along this direction and the receiver decodes by combining with the corresponding combining vector.

*5) Numerical Simulations:* Figure 6 shows the performance of our codebook design in a $2 \times 2$ channel generated with the separable model with $\lambda_1^r = 3.75$, $\lambda_2^r = 0.25$, $\lambda_1^t = 2.08$, $\lambda_2^t = 1.92$. Thus $\mu_{r,2} \approx 7.05$ and ${\rm Gap}_t \approx 148$. Note that these values of $\mu_{r,2}$ and ${\rm Gap}_t$ correspond to Region 4 where $\Sigma_r$ and $\Sigma_t$ are respectively ill-and well-conditioned (see Section IV-B) and which is expected to have the highest (worst-case) feedback requirement for any operational level of near perfect CSI performance. As expected from Theorem 1, due to the closeness of the two transmit eigenvalues, we see that pure statistical beamforming is more than 5 dB away from perfect CSI beamforming. However with an optimal

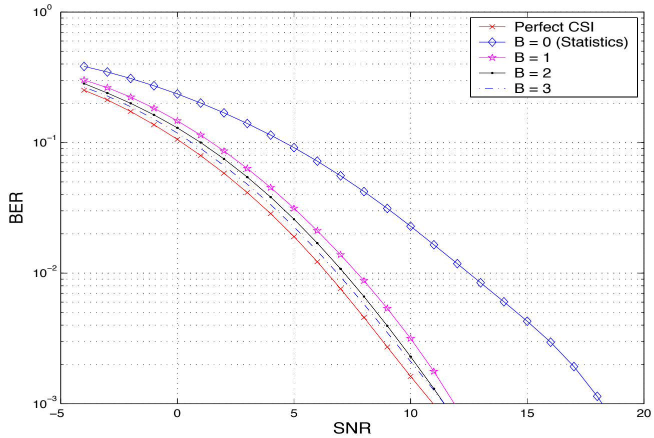


Fig. 6. Feedback performance in a $2 \times 2$ channel generated with the separable model with $\mu_{r,2} \approx 7.05$ and ${\rm Gap}_t \approx 148$. The figure shows that even with as low as 1-bit feedback, substantial gains can be achieved with our codebook designs in extremely correlated scenarios.

[5]Due to the non-increasing property of $M_{L_i}$ and the fact that a single root codeset is used to generate the local component of the codebook, not all of the $M_{L_i}$ can equal $N$.

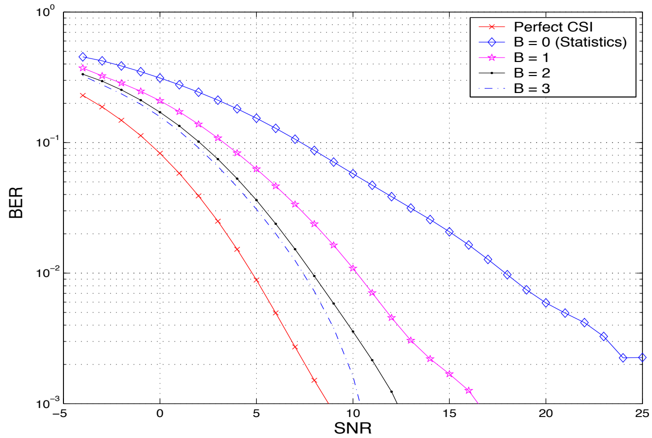


Fig. 7. Feedback performance in a $1 \times 4$ channel generated with the separable model with transmit eigenvalues $1.38, 1.37, 0.99$ and $0.26$ resulting in $\mu_{r,\,2} = 16$ and $\text{Gap}_t \approx 10^4$.

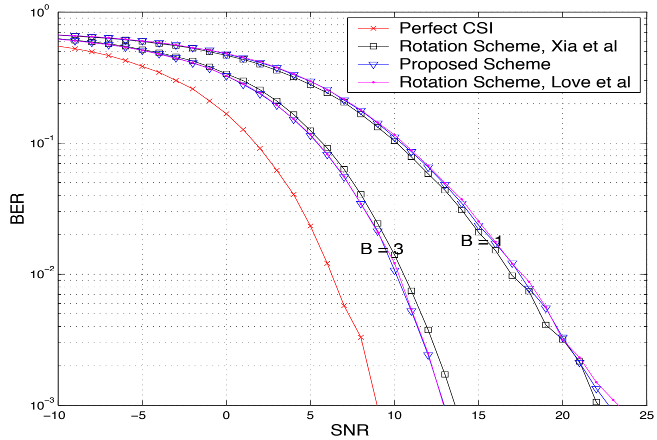


Fig. 8. Comparison of codebook designs of [12], [11] and the design proposed here in the MISO case with $N_t = 8$.

1-bit codebook design which has the two statistical eigenmodes as codewords, we see that we are within 1.5 dB of perfect feedback. The optimal 2-bit codebook which is given by $M_D = 2, M_{L_1} = 2, M_{L_2} = M_G = 0$ leads to within a 0.5 dB loss. (Recall that the feedback requirement is $\log_2(M_D + \sum_{i=1}^{M_D} M_{L_i} + M_G)$ bits.) Increasing the number of bits beyond this point leads to diminishing returns as can be seen from the performance of the optimal 3-bit codebook which has $M_D = 2, M_{L_1} = M_{L_2} = 1$ and $M_G = 4$.

We now provide a short description on how the codebooks are designed. As per prior discussion, given a root codeset of $N$ codewords, it is always possible to obtain another root codeset of $K$ codewords for all $K \leq N$, simply by discarding some of the codewords. Similarly, the scaling map can only reduce the radius of the cap and not *vice versa*. Thus we start with a root codeset (LGP) of 16 codewords restricted to a cap radius of 0.95. This root codeset is designed via VQ [2], [9]. By using the rotation map, we rotate the root codeset to the statistical eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ and the scaling map is used with the parameters $\alpha_1 = 0.7$ and $\alpha_2 = 0.5$. For the two bit feedback scheme with $M_D = 2$ and $M_{L_1} = 2$, we retain only two of the sixteen codewords in the rotated codeset corresponding to $\mathbf{u}_1$ as center.

In Figure 7, we plot the feedback performance of a $1 \times 4$ (MISO) channel with transmit eigenvalues equal to $1.38, 1.37, 0.99, 0.26$. The fact that $\Sigma_t$ is well-conditioned is characterized by $\text{Gap}_t$ which is approximately $10^4$. The plot shows that 3-bit feedback with our codebook design leads to within a dB of perfect CSI performance. The design parameters for this simulation are: $N = 16$, $\gamma = 0.95$, $M_D = 3$, $M_{L_1} = M_{L_2} = M_{L_3} = 1$, $M_G = 2$, $\alpha_1 = \alpha_2 = 0.7$, $\alpha_3 = 0.5$.

We then focus on an $1 \times 8$ case where we compare the performance of our codebook design with those proposed in [12] and [11]. In these works, a Grassmannian codebook designed for i.i.d. channels is skewed by the transmit covariance matrix $\Sigma_t$. In particular, in [11], the codewords have the form $\frac{\Sigma_t^{1/2}\mathbf{x}}{\|\Sigma_t{}^{1/2}\mathbf{x}\|}$ while in [12], it is of the form $\frac{\Sigma_t \mathbf{x}}{\|\Sigma_t \mathbf{x}\|}$ with $\mathbf{x}$ an i.i.d. channel codeword. Apart from the common theme that all three designs assign priorities to dominant spatial direction(s), the method in which this priority is assigned seems to be drastically different. While [12] and [11] first assign weightages to the coefficients of a Grassmannian codebook based on the transmit eigenvalues, and then rotate the codebook to the statistical directions, we choose an appropriate number of codewords around the statistical direction(s) based on the transmit eigenvalues (which determines $\{M_{L_i}\}$).

One of our future works would be on understanding the connections between the codebook designs in a more comprehensive way. In Figure 8, we plot the performance of the three feedback schemes on an $1 \times 8$ channel with transmit eigenvalues $[1.59\ 1.59\ 1.52\ 1.11\ 0.67\ 0.67\ 0.60\ 0.24]$. For both $B = 1$ as well as $B = 3$, we see that all three codebook designs perform almost similarly. Extensive numerical simulations seem to show this trend even though it is unclear as to what the connections between the different codebook designs are. However, note that while the schemes in [12] and [11] specifically assume a MISO setting, our scheme extends to the more general MIMO case. Moreover, the heuristic behind our scheme is easily explainable and is motivated by studying statistical beamforming. On the other hand, it is not clear why a transformation by $\Sigma_t$ as in [12] or $\Sigma_t^{1/2}$ as in [11] should lead to optimal performance. Finally, note that neither [12] or [11] study the impact of correlation on received SNR loss, akin to Theorems 1 and 2.

## VI. Universal Codebook Design

Now, we consider the case where channel statistical information is difficult to obtain at the transmitter, either due to a rapid change in statistics (that is, $L$ is not large) or due to a feedback constraint which prevents accurate statistical information transfer to the transmitter. In this setting, we propose a universal design that **is independent of channel statistics** and is aimed at robustness for all channel correlations. For every frame where the channel statistics remain static, we first estimate whether we should use a near-i.i.d. solution or restrict ourselves to a localized codebook. In the latter case, we also estimate the best radius and direction of a spherical cap (used for quantization) by using some training observations.

### *A. Construction*

The universal codebook $\mathcal{C}$ proposed here consists of two components: a GLP component $\mathcal{C}_0$ of $N$ codewords denoted by $\mathcal{C}_0 = [\mathbf{c}_1 \cdots \mathbf{c}_N]$, and a family of $K = N_{\rm cen} N_{\rm rad}$ local packings $\mathcal{C}_k$, $k = 1, \cdots, K$ with $N-1$ local codewords. The first component is suitable for channels that are not very correlated and the second component partitions the abstract space of channel correlations and the $K$ local packings serve as quantizations of this partition. We now explain the construction of the second component in more detail.

The first step in the design of the local component is the design of a root codeset of $N-1$ codevectors that corresponds to a LGP. Similar to the design in "Step 0" of the previous section, this can be done via VQ. Then, a set $\mathbf{O} = \{\mathbf{o}_1 \cdots \mathbf{o}_{N_{\rm cen}}\}$ of $N_{\rm cen}$ isotropically distributed unit vectors in $\Omega_{N_t}$ is designed using either GLP, VQ or equiangular frames. These $N_{\rm cen}$ vectors serve as the centers of the local Grassmannian packings to be constructed now. We further constrain the radii of these local packings to take values from a finite set of $N_{\rm rad}$ values given by $\Gamma = \{\gamma_1 \cdots \gamma_{N_{\rm rad}}\}$.

With the techniques expounded in the previous section, the $K$ LGPs that form the local component of the codebook are obtained by rotation and scaling of the root codeset (to $\mathbf{o}_i$ and by $\gamma_j$, respectively) and are denoted by $\mathcal{C}_{(\mathbf{o}_i, \gamma_j)}, i = 1 \cdots N_{\rm cen}, j = 1 \cdots N_{\rm rad}$ for simplicity. The most important advantage of this construction is that using the elementary operations of rotation and scaling, the second component can be constructed from a single root codeset without resorting to a $K$-fold VQ construction of local packings. The codebook design is illustrated in Figure 9.

### *B. Codeword Selection*

Here we present the criteria for selection of the *best* center for quantization of the optimal beamforming vector. Define $\mu_{GLP}$ and $\mu_{\rm loc}$ as

$$\begin{aligned} \mu_{GLP} &\triangleq \frac{1}{T_{\rm train}} \sum_{k=1}^{T_{\rm train}} \max_{\mathbf{c}_j \in \mathcal{C}_0} \|\mathbf{H}[k]\, \mathbf{c}_j\|^2, \\ \mu_{\rm loc} &\triangleq \max_{\mathbf{o}_i \in \mathbf{O}} \frac{1}{T_{\rm train}} \sum_{k=1}^{T_{\rm train}} \|\mathbf{H}[k]\, \mathbf{o}_i\|^2. \end{aligned} \quad (9)$$

The quantity $\mu_{GLP}$ is a measure of the projection of the dominant right singular vector of $\mathbf{H}[k]$ (in the training phase) onto a GLP codebook, that is, $\mu_{GLP}$ reflects if these singular vectors are isotropically distributed, which would be the case if the statistics of $\mathbf{H}[k]$ are near i.i.d. On the other hand, $\mu_{\rm loc}$ measures the maximal projection of these singular vectors onto a fixed direction (given by $\mathbf{o}_i$) and thus $\mu_{\rm loc}$ measures if the singular vectors are localized in some spherical cap in $G(N_t, 1)$. Thus $\mu_{GLP}$ and $\mu_{\rm loc}$ are measures of correlation in the channel statistics.

If the channel statistics are near-i.i.d. or not-so correlated, then $\mu_{GLP} > \mu_{\rm loc}$ and the optimal codeword is the one from $\mathcal{C}_0$ that maximizes the received SNR. Otherwise, the center of the LGP defining the optimal codeword is chosen (as below) as the one that maximizes the average received SNR over all possible centers during the training phase. Thus, we have

$$\mathbf{o}_{\rm opt} = \arg \max_{\mathbf{o}_i \in \mathbf{O}} \frac{1}{T_{\rm train}} \sum_{k=1}^{T_{\rm train}} \|\mathbf{H}[k]\, \mathbf{o}_i\|^2. \quad (10)$$

Similarly, the optimal radius $\gamma_{\rm opt}$ is the argument that maximizes the average received SNR during the training phase of the locally-optimal codewords from all local packings centered at $\mathbf{o}_{\rm opt}$. Once $\mathbf{o}_{\rm opt}$ and $\gamma_{\rm opt}$ have been chosen, we use instantaneous channel information (known at the receiver) in the post-training phase to select the locally-optimal codeword from the specific LGP as

$$\mathbf{w}_{\rm opt}[k] = \arg \max_{\mathbf{w}_j \in \mathcal{C}_{(\mathbf{o}_{\rm opt}, \gamma_{\rm opt})}} \|\mathbf{H}[k]\, \mathbf{w}_j\|^2, \; k > T_{\rm train}. \quad (11)$$

Since there are $K+1$ packings (one GLP and $K$ local ones), the feedback requirement for packing identification is $\log_2(K+1)$ bits per $T_{\rm stat}$ duration, and since there are $N$ codewords in each packing, the overhead corresponding to transmission of the codeword index is $\log_2(N)$ bits for every coherence block.

## VII. CONCLUSION

Open loop schemes in which the transmitter does not *explicitly* use the channel state information lead to substantial losses in performance when compared with closed-loop schemes. Instantaneous channel state information at the transmitter is extremely costly to obtain in practice. However, if the *channel statistics* are static for a sufficiently long duration, this information can be fed back from the receiver to the transmitter with essentially zero cost. In such situations, a limited feedback codebook could be designed by exploiting the statistical information and be used to achieve better performance than the open loop case. This work proposes a systematic channel statistics-dependent codebook construction for quantized beamforming in correlated channels. Numerical studies show that our scheme achieves near perfect CSI performance for a variety of correlated channel conditions, even with few bits of feedback. On the other hand, the channel statistics could vary at a fast rate, *e.g.,* highly mobile scenarios, or that accurate statistical feedback is not possible. In this situation, we propose a universal codebook design that is robust for all channel statistics, very correlated through near-i.i.d. channels.

These constructions borrow heavily from studying the impact of transmit/receive covariance matrix conditioning on performance loss with statistical beamforming in correlated channels. Partitioning the covariance matrices based on their condition numbers seems to be a fruitful technique as has been shown by another recent work on adaptive MIMO transmission [31]. The construction of the non-trivial rotation and scaling maps enable the easy implementation of the codebook design.

This work is a first attempt at a systematic codebook design for MIMO channels by exploiting spatial correlation. More work needs to be done to understand the impact of spatial correlation on error probability for a limited feedback scheme. A source coding perspective similar to that in [32] could be useful in this regard. This study would help in the

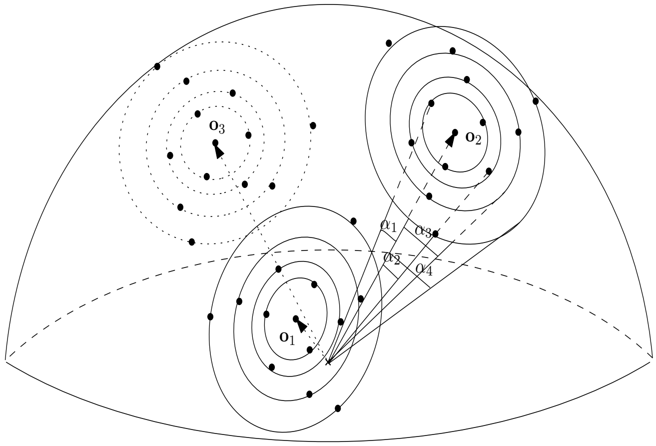


Fig. 9. Illustration of the local component of a universal codebook with $N_{\rm cen} = 3, N_{\rm rad} = 4$ and $N = 4$. The centers of the LGP are denoted by $\mathbf{o}_1, \mathbf{o}_2$ and $\mathbf{o}_3$ while the angles of these local packings are denoted by $\alpha_1, \cdots, \alpha_4$. These local packings are obtained by scaling down a root codeset by appropriate radius values. The feedback requirement is $\log_2(13)$ bits per $T_{\rm stat}$ duration and 2-bits per coherence block in the post-training phase.

design of more efficient feedback schemes. Further work is also necessary to study the impact of the codebook designs proposed here on the rate of information transfer. Besides, limited feedback studies usually focus on either the spatial multiplexing aspect or the diversity gain aspect. Thus, there is a need to systematically study limited feedback in the intermediate regime as in [33], [34]. More interesting is the insight that channel hardening can provide on multi-user/distributed communications with limited feedback. For *e.g.,* (capacity) order-optimal schemes on the downlink channel like zero-forcing beamforming can be studied using ideas similar to those explored in this paper [35]. Distributed beamforming in sensor networks is another area where these ideas can be further leveraged.

## Acknowledgment

The authors would like to thank Roopsha Samanta for the codebook design from [36] and Bishwarup Mondal for the rotation and scaling construction.

## Appendix

### A. Performance Loss with Quantized Feedback

Let the eigen-decomposition of $\mathbf{H}^H\mathbf{H}$ be given by $\sum_{i=1}^{N_t} \lambda_i \mathbf{v}_i \mathbf{v}_i^H$ where $\lambda_1 \geq \cdots \geq \lambda_{N_t} \geq 0$. Then, $D(\mathcal{W})$ can be written as

$$\begin{aligned} D(\mathcal{W}) &= E_{\mathbf{H}}\left[\lambda_1 - \max_{\mathbf{w}\in\mathcal{W}} \sum_i \lambda_i \left|\mathbf{v}_i^H \mathbf{w}\right|^2\right] \\ &\overset{(a)}{\leq} E_{\mathbf{H}}\Big[\lambda_1 (1 - \underbrace{\max_{\mathbf{w}\in\mathcal{W}} \left|\mathbf{v}_1^H \mathbf{w}\right|^2}_{P})\Big] \end{aligned} \tag{12}$$

where (a) follows from ignoring $\lambda_2, \cdots, \lambda_{N_t}$. We now employ the definition of the correlation coefficient between two random variables to obtain $E_{\mathbf{H}}\left[\lambda_1(1-P)\right]$

$$\begin{aligned} &= E_{\mathbf{H}}(\lambda_1) E_{\mathbf{H}}\left[1-P\right] + \rho\sqrt{\mathrm{Var}_{\mathbf{H}}(\lambda_1)}\sqrt{\mathrm{Var}_{\mathbf{H}}(1-P)} \\ &\overset{(b)}{\leq} E_{\mathbf{H}}(\lambda_1) E_{\mathbf{H}}\left[1-P\right] + \sqrt{E_{\mathbf{H}}(P^2) - \left[E_{\mathbf{H}}(P)\right]^2}\sqrt{\mathrm{Var}_{\mathbf{H}}(\lambda_1)} \\ &\overset{(c)}{\leq} E_{\mathbf{H}}(\lambda_1)\left[1 - E_{\mathbf{H}}(P)\right] + \sqrt{\mathrm{Var}_{\mathbf{H}}(\lambda_1)}\sqrt{1 - \left[E_{\mathbf{H}}(P)\right]^2} \\ &= \left(\sqrt{1 - E_{\mathbf{H}}(P)}\right)^2 E_{\mathbf{H}}(\lambda_1) \\ &\quad + \sqrt{1 - E_{\mathbf{H}}(P)}\left[\sqrt{1 + E_{\mathbf{H}}(P)}\sqrt{\mathrm{Var}_{\mathbf{H}}(\lambda_1)}\right] \\ &\overset{(d)}{\leq} \sqrt{1 - E_{\mathbf{H}}(P)}\left[E_{\mathbf{H}}(\lambda_1) + \sqrt{2\,\mathrm{Var}_{\mathbf{H}}(\lambda_1)}\right] \end{aligned}$$

where in (b) we have used the fact that $\rho$ is real and $|\rho| \leq 1$, in (c) we have used a trivial upper bound of 1 for $E_{\mathbf{H}}(P^2)$, and in (d) we have used $E_{\mathbf{H}}(P) \leq 1$. Thus we are done. ■

### B. Proof of Theorem 1

We first introduce some mathematical results that would be needed in the proof of Theorem 1.

The well-known Bernstein's inequality [37, Theorem 2.8, p. 57], provides an exponentially decaying bound on the concentration of sums of independent random variables.

*Lemma 2 (Bernstein):* Consider a sequence of independent random variables $\{X_i\}$ such that $E[X_i] = 0$ and $\sigma_i^2 = E[X_i^2] < \infty$ for all $i$. Let $S_n = \sum_{i=1}^n X_i$, $B_n = \sum_{i=1}^n \sigma_i^2$ and assume that there exists a positive constant $H$ such that

$$\left|E[X_i^m]\right| \leq \frac{1}{2}\, m!\, \sigma_i^2\, H^{m-2} \tag{13}$$

for all $i$ and all integers $m \geq 2$. If $x \leq \frac{B_n}{H}$, then

$$\Pr\left(|S_n| \geq x\right) \leq 2\exp\left(-\frac{x^2}{4B_n}\right). \quad ■$$

Note that the moment condition in (13) is a condition on the tail probability of $\{X_i\}$. Also, the lemma is usually applied with $X_i = \frac{Y_i}{n}$ for some "nice" random variables $\{Y_i\}$ and the lemma can be easily extended to the complex case. The next lemma of this section concerns the eigenvectors of matrix perturbations [38].

*Lemma 3 (Mathias, 1994):* Let $\Lambda = \mathrm{diag}(\lambda_j)$ be an $n \times n$ positive definite diagonal matrix with distinct eigenvalues. Let $X$ be a Hermitian matrix and consider $S(\epsilon) = \Lambda + \epsilon X$. Then for sufficiently small $\epsilon$ and $j = 1 \ \cdots \ n$, 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

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

We note that in the above lemma, only the largest eigenvalue of $\Lambda$ needs to be distinct to ensure that the eigenvector corresponding to the largest eigenvalue of $\Lambda + \epsilon X$ has a perturbation of $\epsilon$ over the corresponding eigenvector of $\Lambda$. This is an important point that we would exploit in the sequel. We are now prepared to prove Theorem 1.

*Proof:* Let $\mu_{r,1}$, $\mu_{r,2}$ and $\text{Gap}_t$ be defined as in the main statement. From (3), we have $\frac{E[\text{Tr}(\mathbf{H}\mathbf{H}^H)]}{N_r} \; \mathbf{H}^H\mathbf{H}$

$$\begin{aligned} &= \text{U}_t \, \Lambda_t^{1/2} \, \left( \frac{\mathbf{H}_{\text{iid}}^H \Lambda_r \mathbf{H}_{\text{iid}}}{N_r} \right) \Lambda_t^{1/2} \, \text{U}_t^H \\ &= \text{U}_t \, \Lambda_t \, \mathbf{\Lambda} \, \text{U}_t^H + \text{U}_t \, \Lambda_t^{1/2} \, \mathbf{O} \, \Lambda_t^{1/2} \, \text{U}_t^H \end{aligned} \quad (14)$$

where $\mathbf{\Lambda}$ is a diagonal matrix with $\mathbf{\Lambda}(i) = \frac{1}{N_r}\sum_k |\mathbf{H}_{\text{iid}}(k,i)|^2 \lambda_k^r$ and $\mathbf{O} = \frac{1}{N_r}\mathbf{H}_{\text{iid}}^H \Lambda_r \mathbf{H}_{\text{iid}} - \mathbf{\Lambda}$. That is, we decompose $\mathbf{H}^H\mathbf{H}$ into a part constituted by the dominant diagonal terms, $\mathbf{\Lambda}$, and the off-diagonal components in $\mathbf{O}$ which vanish in the limit of large $N_r$ due to the law of large numbers. Our program is as follows: First, we show that the eigenvalues of $\text{U}_t \Lambda_t \mathbf{\Lambda} \text{U}_t^H$ are distinct with a high probability. Then, performing analysis modulo this high-probability set, we show that the contributions to the dominant eigenvector from the off-diagonal entries are small with a very high probability. In the last step, we compute the rate of convergence of $D_u^{code}(\mathcal{W})$.

For any given pair of values of $i$ (denoted by $i_1$ and $i_2$), with $X_k = \frac{1}{N_r}\left( |\mathbf{H}_{\text{iid}}(k,i)|^2 \lambda_k^r - \lambda_k^r \right)$, we now check that the assumptions in Lemma 2 hold. It is straightforward to see that $E[X_k] = 0$. Simple consequences of Gaussian moment factoring theorem [39] are the following: $E[X_k^2] = \left(\frac{\lambda_k^r}{N_r}\right)^2$ and $E[X_k^m] \le m! \left(\frac{\lambda_k^r}{N_r}\right)^m$. Thus $B_n = \frac{\mu_{r,2}}{N_r}$ and $H = \frac{\max_k \lambda_k^r}{N_r}$. Note that the only way a pair $\Lambda_t(i_1)\mathbf{\Lambda}(i_1)$ and $\Lambda_t(i_2)\mathbf{\Lambda}(i_2)$ can be non-distinct is if $\mathbf{\Lambda}(i_1) < \mu_{r,1} - \epsilon$ and if $\mathbf{\Lambda}(i_2) > \mu_{r,2} + \epsilon$ for some $\epsilon > 0$. From Lemma 2, this happens with a probability bounded by $2\exp\left(-\frac{\epsilon^2 N_r}{4\mu_{r,2}}\right)$ as long as $\epsilon < \frac{\mu_{r,1}}{\max_k \lambda_k^r}$. Thus, if $\frac{\lambda_1^t}{\lambda_2^t} > 1 + \frac{2\epsilon N_r}{\sum_i \lambda_i^r - \epsilon N_r}$, then $\Lambda_t(i_1)\mathbf{\Lambda}(i_1)$ and $\Lambda_t(i_2)\mathbf{\Lambda}(i_2)$ are distinct with probability larger than $1 - 2\exp\left(-\frac{\epsilon^2 N_r}{4\mu_{r,2}}\right)$. Doing the above for all the $i$'s, the eigenvalues of $\Lambda_t\mathbf{\Lambda}$ are distinct with a probability

$$p_1 \ge 1 - 2N_t \exp\left( -\frac{\epsilon^2 N_r}{4\mu_{r,2}} \right). \quad (15)$$

Modulo this high-probability set, we compute $p_2 \triangleq \Pr(1 - |\mathbf{v}_1^H \mathbf{u}_{\text{stat}}|^2 < \delta)$ for some $\delta$ arbitrary. Towards this, we apply Lemma 3 on the matrix $\mathbf{S} = \Lambda_t\mathbf{\Lambda} + \Lambda_t^{1/2}\mathbf{O}\Lambda_t^{1/2}$. The following represents the components of the (unnormalized) eigenvector matrix of $\mathbf{S}$: $i$-th diagonal entry is $1 + K_i$ where $K_i$ corresponds to the $\mathcal{O}(\epsilon^2)$ term in the lemma, and for $i \neq j$, the $(i,j)-$th entries are $\frac{\Lambda_t(i)^{1/2}\Lambda_t(j)^{1/2}}{\Lambda_t(i)\mathbf{\Lambda}(i) - \Lambda_t(j)\mathbf{\Lambda}(j)} |\mathbf{O}(i,j)|$. Note that in the high probability set that is considered, $\mathbf{\Lambda}(i)$ can be well-approximated by $\mu_{r,1}$. Normalizing the eigenvectors of $\text{U}_t\mathbf{S}\text{U}_t^H$ to be unit normed, and computing $|\mathbf{v}_1^H\mathbf{u}_{\text{stat}}|^2$, we have $p_2 \ge \Pr\left( \sum_{i=2}^{N_t} \frac{x_i}{(1-x_i)^2} |\mathbf{O}(i,1)|^2 < \delta\,(\mu_{r,1})^2 \right)$ where $x_i = \frac{\lambda_i^t}{\lambda_1^t}$ and we have assumed that $\mathbf{H}$ has been permuted such that $\lambda_1^t$ is the leading diagonal element of $\Lambda_t$, that is, the indexing in $\Lambda_t(1) = \lambda_1^t$ and $\mathbf{O}(i,1)$ reflects this permutation.

After some standard simplifications, we have

$$\begin{aligned} p_2 \;\; &\ge \;\; 1 - (N_t - 1)\Pr\left( \frac{1}{N_r} \left| \sum_{k=1}^{N_r} \mathbf{H}_{\text{iid}}(k,1)^\star \, \mathbf{H}_{\text{iid}}(k,j)\, \lambda_k^r \right| \right. \\ &\qquad \left. > \sqrt{\frac{\delta\,(\mu_{r,1})^2}{(N_t-1)\,\text{Gap}_t}} \right), \text{ for any } j \neq 1. \end{aligned}$$

Another application of Lemma 2 results in $p_2 \ge 1 - 2(N_t - 1)\exp\left( -\frac{\delta\,(\mu_{r,1})^2 N_r}{4(N_t-1)\,\text{Gap}_t\,\mu_{r,2}} \right)$.

Now, we compute $D_u^{code}(\mathcal{W})$ as follows:

$$\begin{aligned} \left(D_u^{code}(\mathcal{W})\right)^2 \;\; &= \;\; E_{\mathbf{H}}\left[ (1 - |\mathbf{v}_1^H\mathbf{u}_{\text{stat}}|^2)\; \chi(\mathcal{A}) \right] \\ &\quad + E_{\mathbf{H}}\left[ (1 - |\mathbf{v}_1^H\mathbf{u}_{\text{stat}}|^2)\; \chi(\mathcal{A}^c) \right] \end{aligned}$$

where $\mathcal{A}$ is the set over which the eigenvalues are distinct and $1 - |\mathbf{v}_1^H\mathbf{u}_{\text{stat}}|^2 < \delta$. We thus have $\Delta = \left(D_u^{code}(\mathcal{W})\right)^2$

$$\begin{aligned} \Delta \;\; &\le \;\; \delta p_1 p_2 + (1 - p_1 p_2) \\ &= \;\; (\delta - 1) p_1 p_2 + 1 \\ &\overset{(a)}{\le} \;\; (\delta - 1)(p_1 + p_2 - 1) + 1 \\ &\approx \;\; \delta + 2(1-\delta) N_t \exp\left( -\frac{\delta\,(\mu_{r,1})^2\, N_r}{4\;\text{Gap}_t \cdot \mu_{r,2} \cdot N_t} \right) \end{aligned} \quad (16)$$

where in (a) we have used the elementary fact that $p_1 p_2 \ge p_1 + p_2 - 1$. By letting $\epsilon \to 0$ as $\epsilon = \frac{1}{N_r^a}$ for some $a > 0$, the $p_1$ term is sub-dominant with respect to the $p_2$ term (that is, $p_1$ can be made to go to 0 much faster than $p_2$). This is reflected in (16) and since $\delta$ is arbitrary, by balancing the two $\delta$-terms, we have $\left(D_u^{code}(\mathcal{W})\right)^2 \le K_1 \, \frac{\mu_{r,2}}{(\mu_{r,1})^2} \cdot \text{Gap}_t \cdot \frac{N_t \log(N_r)}{N_r}$. Thus we are done. ■

### *C. Proof of Theorem 2*

The proof follows on analogous lines to Theorem 1 by considering $\frac{\mathbf{H}^H\mathbf{H}}{N_r}$ and noting that Lemma 2 holds for independent random variables and replacing $\mathbf{O}(i,j)$ in Lemma 3 by the appropriate random variable. ■

## References

[1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, NY, 2003.
[2] 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 Jour. Select. Areas in Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
[3] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.
[4] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On Beamforming with Finite Rate Feedback in Multiple Antenna Systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562 – 2579, Oct. 2003.
[5] E. Visotsky and U. Madhow, "Space-Time Transmit Precoding with Imperfect Feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
[6] S. Zhou and G. B. Giannakis, "Optimal Transmitter Eigen-Beamforming And Space-Time Block Coding Based On Channel Mean Feedback," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.
[7] S. Zhou and G. B. Giannakis, "Optimal Transmitter Eigen-Beamforming And Space-Time Block Coding Based On Channel Correlations," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1673–1690, July 2003.
[8] B. C. Banister and J. R. Zeidler, "A Simple Gradient Sign Algorithm for Transmit Antenna Weight Adaptation with Feedback," *IEEE Trans. Sig. Proc.*, vol. 51, no. 5, pp. 1156–1171, May 2003.

[9] J. C. Roh and B. D. Rao, "Transmit Beamforming in Multiple-Antenna Systems with Finite Rate Feedback: A VQ-Based Approach," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

[10] B. Mondal and R. W. Heath, Jr., "Channel Adaptive Quantization for Limited Feedback MIMO Beamforming Systems," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4731–4740, Dec. 2006.

[11] P. Xia and G. B. Giannakis, "Design and Analysis of Transmit Beamforming based on Limited-Rate Feedback," *IEEE Trans. Sig. Proc.*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[12] D. J. Love and R. W. Heath, Jr., "Limited Feedback Diversity Techniques for Correlated Channels," *IEEE Trans. Veh. Tech.*, vol. 55, no. 2, pp. 718–722, Mar. 2006.

[13] V. Raghavan, A. M. Sayeed, and N. Boston, "Near-Optimal Codebook Constructions for Limited Feedback Beamforming in Correlated MIMO Channels with Few Antennas," *Proc. ISIT*, 2006.

[14] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Canonical Statistical Models for Correlated MIMO Fading Channels and Capacity Analysis," *To be submitted to IEEE Trans. Inform. Theory*, 2007.

[15] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Link Ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[16] J. Zheng and B. D. Rao, "Analysis of Vector Quantizers using Transformed Codebooks with Application to Feedback-Based Multiple Antenna Systems," *Proc. EUSIPCO*, 2006.

[17] Í. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 2172–2178, 2000.

[18] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[19] H. Ozcelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of 'Kronecker' MIMO Radio Channel Model," *Electronics Letters*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.

[20] J. Wallace, H. Ozcelik, M. Herdin, E. Bonek, and M. Jensen, "Power and Complex Envelope Correlation for Modeling Measured Indoor MIMO Channels: A Beamforming Evaluation," *IEEE Fall VTC, 2003*, vol. 1, pp. 363–367, Oct. 2003.

[21] D. P. McNamara, M. A. Beach, and P. N. Fletcher, "Spatial Correlation in Indoor MIMO Channels," *IEEE Intern. Symp. Pers. Ind. Mob. Radio Comm.*, 2002.

[22] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental Study of MIMO Channel Statistics and Capacity via the Virtual Channel Representation," *To be submitted*, 2007.

[23] A. M. Sayeed, "Deconstructing Multi-Antenna Fading Channels," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[24] V. Raghavan, A. M Sayeed, and N. Boston, "When is Limited Feedback for Transmit Beamforming Beneficial?," *Proc. ISIT*, 2005.

[25] M. R. McKay, I. B. Collings, A. Forenza, and R. W. Heath, Jr., "A Throughput-Based Adaptive MIMO-BICM Approach for Spatially-Correlated Channels," *Proc. of the IEEE ICC, Istanbul*, 2006.

[26] A. Forenza, M. R. McKay, A. Pandharipande, R. W. Heath, Jr., and I. B. Collings, "Capacity Enhancement via Multi-Mode Adaptation in Spatially Correlated MIMO Channels," *IEEE Intern. Symp. Pers. Indoor, Mobile Radio Commun.*, 2005.

[27] P. Almers, F. Tufvesson, and A. F. Molisch, "Measurement of Keyhole Effect in a Wireless Multiple-Input Multiple-Output Channel," *IEEE Fall VTC 2003*, Oct. 2003.

[28] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing Lines, Planes etc., Packings in Grassmannian Spaces," *Experimental Math.*, vol. 5, pp. 139–159, 1996.

[29] B. Mondal, R. Samanta, and R. W. Heath, Jr., "Frame-Theoretic Quantization for Limited Feedback MIMO Beamforming Systems," *Proc. Wirelesscom*, May 2005.

[30] A. Barg and D. Yu. Nogin, "Bounds on Packings of Spheres in the Grassmann Manifolds," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2450–2454, Sep. 2002.

[31] A. Forenza, A. Pandharipande, H. Kim, and R. W. Heath, Jr., "Adaptive MIMO Transmission Scheme: Exploiting the Spatial Selectivity of Wireless Channels," *IEEE Fall VTC*, pp. 2188–2192, 2005.

[32] J. Zheng and B. D. Rao, "Capacity Analysis of Correlated Multiple-Antenna Systems with Finite Rate Feedback," *Proc. IEEE ICC*, 2006.

[33] V. Raghavan and A. M. Sayeed, "Impact of Spatial Correlation on Statistical Precoding in MIMO Channels with Linear Receivers," *Proc. Annual Allerton Conference on Communications, Control and Computing*, 2006.

[34] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Limited Feedback Precoder Design for Spatially Correlated MIMO Channels," *Proc. Conf. on Inform. Sci and Sys.*, 2007.

[35] V. Raghavan and V. V. Veeravalli, "On Quantized Multi-User Beamforming in Spatially Correlated Broadcast Channels," *Proc. ISIT*, 2007.

[36] R. Samanta and R. W. Heath, Jr., "Codebook Adaptation for Quantized MIMO Beamforming Systems, *Proc. IEEE Asilomar Conf. Signals, Systems, and Computers*, pp. 376–380, 2005.

[37] V. V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Clarendon Press, Oxford, 1995.

[38] R. Mathias, "Spectral Perturbation Bounds for Positive Definite Matrices," *SIAM Journ. Matrix Anal. Appl.*, vol. 18, no. 4, pp. 959–980, Oct. 1997.

[39] I. S. Reed, "On a Moment Theorem for Complex Gaussian Processes," *IRE Trans. Inform. Theory*, vol. 8, pp. 194–195, Apr. 1962.


**Vasanthan Raghavan** (S'01 - M'06) received the B.Tech degree in electrical engineering from the Indian Institute of Technology at Madras, Chennai, India in 2001, the M.S. and the Ph.D. degrees in electrical and computer engineering in 2004 and 2006 respectively, and the M.A. degree in mathematics in 2005, all from the University of Wisconsin-Madison, Madison, WI. He is currently a postdoctoral research associate with the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, Urbana, IL. His research interests span multi-antenna communication techniques, information theory, multihop networking, stochastic control, and random matrix theory.

**Robert W. Heath, Jr.** (S'96 - M'01 - SM'06) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, in 1996 and 1997 respectively, and the Ph.D. from Stanford University, Stanford, CA, in 2002, all in electrical engineering.

From 1998 to 2001, he was a Senior Member of the Technical Staff and then, a Senior Consultant at Iospan Wireless Inc, San Jose, CA where he worked on the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. In 2003 he founded MIMO Wireless Inc, consulting company dedicated to the advancement of MIMO technology. Since January 2002, he has been with the Department of Electrical and Computer Engineering at The University of Texas at Austin where he is currently an Assistant Professor and member of the Wireless Networking and Communications Group. His research interests cover a broad range of MIMO communication including limited feedback techniques, multihop networking, multiuser MIMO, antenna design, and scheduling algorithms as well as 60GHz communication techniques.

Dr. Heath serves as an Editor for the IEEE Transactions on Communication, an Associate Editor for the IEEE Transactions on Vehicular Technology, and is a member of the Signal Processing for Communications Technical Committee of the IEEE Signal Processing Society.

**Akbar M. Sayeed** (S'96 - M'97 - SM'02) received the B.S. degree from the University of Wisconsin-Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 1993 and 1996, all in Electrical Engineering. He was a postdoctoral fellow at Rice University from 1996 to 1997 and he has been with the University of Wisconsin-Madison since 1997 where he is currently Associate Professor of Electrical and Computer Engineering. His current research interests include wireless communications, statistical signal processing, multi-dimensional communication theory, information theory, and applications in wireless communication networks and sensor networks. Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the NSF CAREER Award (1999), the ONR Young Investigator Award (2001), and the UW Grainger Junior Faculty Fellowship (2003). He is a Senior Member of the IEEE, served as an Associate Editor for the IEEE Signal Processing Letters from 1999 to 2002, and is currently serving on the Signal Processing for Communications technical committee of the IEEE Signal Processing Society.