Quantized Multimode Precoding in Spatially Correlated Multi-Antenna Channels

Vasanthan Raghavan, Venugopal V. Veeravalli*, Akbar M. Sayeed

***Abstract*— Multimode precoding, where the number of independent data-streams is adapted optimally, can be used to maximize the achievable throughput in multi-antenna communication systems. Motivated by standardization efforts embraced by the industry, the focus of this work is on systematic precoder design with realistic assumptions on the spatial correlation, channel state information (CSI) at the transmitter and the receiver, and implementation complexity. For the spatial correlation of the channel matrix, we assume a general channel model, based on physical principles, that has been verified by many recent measurement campaigns. We also assume a coherent, linear minimum mean-squared error (MMSE) receiver and knowledge of the spatial statistics at the transmitter along with the presence of an ideal, low-rate feedback link from the receiver to the transmitter. The reverse link is used for codebook-index feedback and the goal of this work is to construct precoder codebooks, adaptable in response to the statistical information, such that the achievable throughput is significantly enhanced over that of a fixed, non-adaptive, i.i.d. codebook design. We illustrate how a codebook of semiunitary precoder matrices localized around some fixed center on the Grassmann manifold can be skewed in response to the spatial correlation via low-complexity maps that can rotate and scale submanifolds on the Grassmann manifold. The skewed codebook in combination with a low-complexity statistical power allocation scheme is then shown to bridge the gap in performance between a perfect CSI benchmark and an i.i.d. codebook design.**

***Index Terms*— Limited feedback communication, quantized feedback, adaptive coding, low-complexity signaling, MIMO systems, channel state information at transmitter, multimode signaling**

I. Introduction

Research over the last decade has firmly established the utility of multiple antennas at the transmitter and the receiver in providing a mechanism to increase the reliability of signal reception [1], or the rate of information transfer [2], or a combination of the two. The focus of this work is on maximizing the achievable rate under certain communication models that are motivated by practical wireless systems [3]. In particular, we assume a *limited (or quantized) feedback* model [4] with perfect channel state information (CSI) at the receiver, perfect statistical knowledge of the channel at the transmitter, and a low-rate feedback link from the receiver to the transmitter.

In this setting, the fundamental problem is to determine the optimal signaling/feedback scheme that maximizes the average mutual information given a statistical description of the channel, the signal-to-noise ratio (SNR), the number of antennas, and the quality of feedback. A first step to solve this problem is to identify the rank of the optimal precoder as a function of the statistics, SNR, and the feedback quality [5]. In practice, the implementation of such a solution is often constrained by the need for low-complexity techniques that limit the number of radio-frequency (RF) link chains (and consequently, the rank of the precoder). Thus the design of the optimal scheme under low-complexity constraints is, in principle, essentially the same as that of the optimal design of a fixed rank limited feedback precoder.

Motivated by this line of reasoning, the main theme of this work is the construction of a systematic, yet low-complexity, limited feedback precoding scheme (of a fixed rank) that results in significantly improved performance over an open-loop[1] scheme. Towards this goal, we consider a simple block fading/narrowband setup where spatial correlation is modeled by a mathematically tractable channel decomposition [6], [7], and includes as special cases the well-studied *i.i.d. model* [2], the *separable correlation model* [8], and the *virtual representation framework* [9], [10]. Furthermore, we assume a simple, linear minimum mean-squared error (MMSE) receiver architecture in this work.

While precoding has been studied extensively under the i.i.d. model [11]–[19], considerable theoretical gaps exist in the limited feedback setting. The extreme case of limited feedback beamforming has been studied in the i.i.d. setting where the isotropicity[2] of the dominant right singular vector of the channel can be leveraged to uniformly quantize the space of unit-normed beamforming vectors, a problem well-studied in mathematics literature as the Grassmannian line packing (GLP) problem [20], [21]. Alternate constructions based on Vector Quantization (VQ)/Random Vector Quantization (RVQ) are also possible [22], [23]. Spatial correlation, however, skews the isotropicity of the right singular vector, and hence poses a fundamentally more challenging problem.

V. Raghavan and V. V. Veeravalli are with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. A. M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA. Email: vasanthan_raghavan@ieee.org, vvv@uiuc.edu, akbar@engr.wisc.edu. *Corresponding author.

This work was partly supported by the NSF under grant #CCF-0049089 through the University of Illinois, and grant #CCF-0431088 through the University of Wisconsin. This paper was presented in part at the 41st Annual Conference on Information Sciences and Systems, Baltimore, MD, 2007.

[1]There is no channel statistical information at the transmitter in an open-loop scheme. That is, the channel is assumed to be i.i.d. and an i.i.d. codebook design is used.

[2]Here, isotropic means that the dominant right singular vector is equally likely to point along any direction in the space of all possible right singular vector(s), which is referred to as the Grassmann manifold. Precise definitions follow later.

While VQ codebooks can be constructed for the correlated channel case, the construction suffers from high computational complexity and the codebook has to be reconstructed from scratch every time the statistics change, thus rendering VQ-type solutions impractical. Recently, beamforming codebooks that can easily be adapted to statistical variations (with low-complexity transformations) have been proposed [24]–[26]. The other extreme, limited feedback spatial multiplexing, has also been studied recently [23], [27].

In the intermediate setting[3] of rank-M precoding, under the i.i.d. assumption, the isotropicity property of the dominant right singular vector of the channel extends to the subspace spanned by the M-dominant right singular vectors thereby allowing a Grassmannian subspace packing solution [28]. In the correlated case, the fundamental challenge on how to quantize the space of M-dominant right singular vectors non-uniformly remains the same as in the beamforming case. However, unlike the beamforming case, it is not even clear how a codebook designed for i.i.d. channels can be skewed in response to the channel correlation. In fact, using an i.i.d. codebook design in a correlated channel can lead to a dramatic degradation in performance (see Figs. 3 and 4).

In contrast to VQ codebooks [23], [29], our systematically constructed semiunitary[4] precoder codebooks are tailored to the spatial correlation, and are easily adaptable in response to changes in statistics. The heuristic behind our construction comes from our previous study of the asymptotic (in antenna dimensions) performance of statistical precoders [30]. We showed in [30] that the performance of a statistical precoder is closest to the optimal precoder when the number of dominant transmit eigenvalues is equal to the rank of the precoder, these dominant eigenvalues are well-conditioned, and the receive covariance matrix is also well-conditioned. A channel satisfying the above conditioning properties is said to be *matched* to the precoding scheme. Measurement campaigns (*e.g.,* see [31, Figs. 9-11]) show that in many realistic situations, the number of dominant transmit eigen-modes is much larger than the precoder rank (which is limited by complexity constraints) indicating that mismatched channels, where the above channel conditions are not met, are quite common in practice. Thus, while limited (or even perfect) feedback can only lead to marginal performance improvement in *matched channels*, in the case of *mismatched channels* where the relative gap in performance between the statistical and the optimal precoders is usually large (see Figs. 3 and 4), the potential benefits of limited feedback are more significant.

Our study [30] suggests that spatial correlation orients the directivity of the M-dominant right singular vectors of the channel towards the statistically dominant subspaces, and hence, a non-uniform quantization of the local neighborhood around the statistically dominant subspaces is necessary. The realizability of such non-uniform quantization with low-complexity, as well as its adaptability, are eased by constructing mathematical maps that can be used to rotate a root codeset (or a submanifold) centered at some arbitrary location on the Grassmann manifold $\mathcal{G}(N_t,\, M)$ towards an arbitrary center and scale it arbitrarily.

Our design includes a statistical component of dominant M-dimensional subspaces of the transmit covariance matrix, a component corresponding to local quantization around the codewords in the statistical component, and an RVQ component that can be constructed with low-complexity. In this context, our construction mirrors and generalizes our recent work in the beamforming case [26]. By combining a semiunitary codebook (of a small enough cardinality) with a low-complexity power allocation scheme that is related to statistical waterfilling, we show via numerical studies that significant performance gains can be achieved, and that the gap to the perfect CSI scheme can be bridged considerably.

Organization: The system setup is introduced in Section II. In Section III, we introduce the notion of mismatched channels where limited feedback precoding results in significant performance improvement. In Section IV, limited feedback codebooks that enhance performance are proposed, and in Section V, mathematical maps are constructed to realize these designs with low-complexity. Numerical studies are provided in Section VI, with a discussion of our results and conclusions in Section VII.

Notation: The M-dimensional identity matrix is denoted by $\mathbf{I}_M$. We use $\mathbf{X}(i,j)$ and $\mathbf{X}(i)$ to denote the i,j-th and i-th diagonal entries of a matrix $\mathbf{X}$, respectively. In more complicated settings (*e.g.,* when the matrix $\mathbf{X}$ is represented as a product or sum of many matrices), we use $\mathbf{X}_{i,j}$ to denote the i,j-th entry. The complex conjugate, conjugate transpose, regular transpose and inverse operations are denoted by $(\cdot)^\star$, $(\cdot)^H$, $(\cdot)^T$ and $(\cdot)^{-1}$ while $E[\cdot]$, $\mathrm{Tr}(\cdot)$ and $\det(\cdot)$ stand for the expectation, the trace and the determinant operators, respectively. The t-dimensional complex vector space is denoted by $\mathbb{C}^t$. We use the ordering $\lambda_1(\mathbf{X}) \geq \cdots \geq \lambda_n(\mathbf{X})$ for the eigenvalues of an $n \times n$-dimensional Hermitian matrix $\mathbf{X}$. The notations $\lambda_{\max}(\mathbf{X})$ and $\lambda_{\min}(\mathbf{X})$ also stand for $\lambda_1(\mathbf{X})$ and $\lambda_n(\mathbf{X})$, respectively.

II. System Setup

We consider a communication model with N_t transmit and N_r receive antennas where M ($1 \leq M \leq \min(N_t, N_r)$) independent data-streams are used in signaling. That is, the M-dimensional input vector $\mathbf{s}$ is precoded into an N_t-dimensional vector via the $N_t \times M$ precoding matrix $\mathbf{F}$ and transmitted over the channel. The discrete-time baseband signal model used is

$$\mathbf{y} = \mathbf{H}\,\mathbf{F}\,\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ is the N_r-dimensional received vector, $\mathbf{H}$ is the $N_r \times N_t$ channel matrix, and $\mathbf{n}$ is the N_r-dimensional zero mean, unit variance additive white Gaussian noise.

A. Channel Model

We assume a block fading, narrowband model for the correlation of the channel in time and frequency. The main emphasis in this work is on the channel correlation in the spatial domain. The spatial statistics of $\mathbf{H}$ depend on the

[3]Here, $1 < M < \min(N_t, N_r)$ with N_t and N_r denoting the transmit and the receive antenna dimensions.

[4]An $N_t \times M$ matrix $\mathbf{X}$ with $M \leq N_t$ is said to be semiunitary if it satisfies $\mathbf{X}^H\mathbf{X} = \mathbf{I}_M$.

operating frequency, the physical propagation environment that controls the angular spreading function, and other factors such as the path distribution and antenna geometry (arrangement and spacing). It is well-known that Rayleigh fading (zero mean complex Gaussian) is an accurate model for $\mathbf{H}$ in a non line-of-sight setting, and hence the complete spatial statistics are described by the second-order moments.

The most general, mathematically tractable spatial correlation model is a *canonical decomposition*[5] of the channel along the transmit and receive covariance bases [6], [7]. In this model, we assume that the auto- and the cross-covariance matrices of all the rows and the columns of $\mathbf{H}$ have the same eigen-bases (denoted by unitary matrices $\mathbf{U}_t$ and $\mathbf{U}_r$, respectively), and thus we can decompose $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U}_r \, \mathbf{H}_{\mathsf{ind}} \, \mathbf{U}_t^H \tag{2}$$

where $\mathbf{H}_{\mathsf{ind}}$ has independent, but not necessarily identically distributed entries. The transmit and the receive covariance matrices are given by

$$\begin{aligned} \mathbf{\Sigma}_t = E\left[\mathbf{H}^H\mathbf{H}\right] &= \mathbf{U}_t \, E\left[\mathbf{H}_{\mathsf{ind}}^H\mathbf{H}_{\mathsf{ind}}\right] \, \mathbf{U}_t^H = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H \\ \mathbf{\Sigma}_r = E\left[\mathbf{H}\mathbf{H}^H\right] &= \mathbf{U}_r \, E\left[\mathbf{H}_{\mathsf{ind}}\mathbf{H}_{\mathsf{ind}}^H\right] \, \mathbf{U}_r^H = \mathbf{U}_r \, \mathbf{\Lambda}_r \, \mathbf{U}_r^H \end{aligned} \tag{3}$$

where $\mathbf{\Lambda}_t = E\left[\mathbf{H}_{\mathsf{ind}}^H\mathbf{H}_{\mathsf{ind}}\right]$ and $\mathbf{\Lambda}_r = E\left[\mathbf{H}_{\mathsf{ind}}\mathbf{H}_{\mathsf{ind}}^H\right]$ are diagonal. Under certain special cases, the model in (2) reduces to some well-known spatial correlation models [6]:

- The case of *ideal channel modeling* assumes that the entries of $\mathbf{H}_{\mathsf{ind}}$ are i.i.d. standard complex Gaussian random variables [2]. The i.i.d. model corresponds to an extreme where the channel is characterized by a single independent parameter, the common variance.
- When $\mathbf{H}_{\mathsf{ind}}$ is assumed to have the form $\frac{1}{\sqrt{\rho_c}} \cdot \mathbf{\Lambda}_r^{1/2} \, \mathbf{H}_{\mathsf{iid}} \, \mathbf{\Lambda}_t^{1/2}$ with $\mathbf{H}_{\mathsf{iid}}$ an i.i.d. channel matrix and the channel power $\rho_c = \mathrm{Tr}(\mathbf{\Lambda}_t) = \mathrm{Tr}(\mathbf{\Lambda}_r)$, the canonical model reduces to the often-studied normalized *separable correlation framework* where the correlation of channel entries is in the form of a Kronecker product of the transmit and the receive covariance matrices [8]. The separable model is described by no more than $N_t + N_r$ independent parameters corresponding to the eigenvalues $\{\mathbf{\Lambda}_t(i)\}$ and $\{\mathbf{\Lambda}_r(i)\}$.
- When uniform linear arrays (ULAs) of antennas are used at the transmitter and the receiver, $\mathbf{U}_t$ and $\mathbf{U}_r$ are well-approximated by discrete Fourier transform (DFT) matrices and the canonical model reduces to the virtual representation framework [9], [10], [32]. In contrast to the general model in (2), the virtual representation offers many attractive properties: a) The matrices $\mathbf{U}_t$ and $\mathbf{U}_r$ are *fixed* and independent of the underlying scattering environment and the spatial eigenfunctions are beams in the virtual directions. Thus, the virtual representation is physically more intuitive than the general model in (2). b) It is only necessary that the entries of $\mathbf{H}_{\mathsf{ind}}$ be independent, but not necessarily Gaussian, a criterion important as antenna dimensions increase. For example, this is relevant as signaling moves towards the 60 GHz frequency regime. c) The case of specular (or line-of-sight) scattering can be easily incorporated within the virtual representation framework [32]. In contrast to the separable model, the virtual representation can support up to N_tN_r independent parameters corresponding to the variances of $\{\mathbf{H}_{\mathsf{ind}}(i,j)\}$.

While performance analysis is tractable in the i.i.d. case, it is unrealistic for applications where large antenna spacings or a rich scattering environment are not possible. Even though the separable model may be an accurate fit under certain channel conditions [33], deficiencies acquired by the separability property result in misleading estimates of system performance [6], [34], [35]. The readers are referred to [7], [31], [34], [36] for more details on how the canonical/virtual models fit measured data better.

B. Channel State Information

If the fading is sufficiently slow, perfect CSI at the receiver is a reasonable assumption for practical communication architectures that use a "training followed by signaling" model. Even in scenarios where this may not be true (*e.g.,* a highly mobile setting), the performance with imperfect CSI at the receiver can be approximated reasonably accurately by the perfect CSI case along with an SNR-offset corresponding to channel estimation. Thus in this work, we will assume a perfect CSI (coherent) receiver architecture. However, obtaining perfect CSI at the transmitter is usually difficult due to the high cost associated with channel feedback/reverse-link training[6].

On the other hand, the statistics of the fading process change over much longer time-scales and can be learned reliably at both the ends. Hence, we assume that the transmitter has perfect knowledge of the channel statistics. In addition, recent technological advances have enabled the possibility of a few bits of quantized channel information to be fed back from the receiver to the transmitter at regular intervals. The most common form of quantized channel information is via a limited feedback codebook $\mathcal{C}$ of 2^B codewords known at both the ends. In this setup, the receiver estimates the channel at the start of a coherence block and computes the index of the optimal codeword from $\mathcal{C}$ for that realization of the channel according to some optimality criterion. It then feeds back the index of the optimal codeword with B bits over the limited feedback link which is assumed to have negligible delay and essentially no errors (since B is usually small [3]). The transmitter exploits this information to convey useful data over the remaining symbols of the coherence block.

C. Transceiver Architecture

The transmitted vector $\mathbf{Fs}$ (see (1)) has a power constraint ρ. Assuming that the input symbols $\mathbf{s}(k)$ have equal energy $\frac{\rho}{M}$, $\mathbf{F}$ satisfies $\mathrm{Tr}(\mathbf{F}^H\mathbf{F}) \leq M$. Non-linear maximum likelihood (ML) decoding of the transmitted data symbols

[5]This model is referred to as the "eigenbeam/beamspace model" in [7].

[6]In case of Time-Division Duplexed (TDD) systems, the reciprocity of the forward and the reverse links can be exploited to train the channel on the reverse link. In case of Frequency-Division Duplexed (FDD) systems, the channel information acquired at the receiver has to be fed back.

using knowledge of $\mathbf{H}$ at the receiver is optimal. However, ML decoding suffers from exponential complexity, in both antenna dimensions and coherence length. Thus, a simple linear MMSE receiver is preferred in practice. With this receiver, the symbol corresponding to the k-th data-stream is recovered by projecting the received signal $\mathbf{y}$ on to the $N_r \times 1$ vector

$$\mathbf{g}_k \quad = \quad \sqrt{\frac{\rho}{M}} \left(\frac{\rho}{M} \mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H + \mathbf{I}_{N_r} \right)^{-1} \mathbf{H}\mathbf{f}_k \qquad (4)$$

where $\mathbf{f}_k$ is the k-th column of $\mathbf{F}$. That is, the recovered symbol is $\widehat{\mathbf{s}}(k) = \sqrt{\frac{\rho}{M}}\, \mathbf{g}_k^H \mathbf{y}$. The signal-to-interference-noise ratio (SINR) at the output of the linear filter, $\mathbf{g}_k$, is

$$\mathsf{SINR}_k = \frac{1}{\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F}\right)^{-1}_{k,k}} - 1. \qquad (5)$$

The outputs $\widehat{\mathbf{s}}(k)$ are passed to the decoder and we assume separate encoders/decoders for each data-stream, as well as independent interleavers and de-interleavers, which reduces the correlation among the interference terms at the outputs of the receiver filters. The performance measure is the mutual information between $\mathbf{s}$ and $\widehat{\mathbf{s}}$. Assuming that the interference plus noise at the output of the linear filter has a Gaussian distribution, which is true with Gaussian inputs and is a good approximation in the non-Gaussian setting when $\{M, N_t, N_r\}$ are large, the mutual information is given by

$$I(\mathbf{s};\widehat{\mathbf{s}}) \quad = \quad \sum_{k=1}^{M} \log_2 \left(1 + \mathsf{SINR}_k\right) \qquad (6)$$

$$\quad = \quad -\sum_{k=1}^{M} \log_2 \left(\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F} \right)^{-1}_{k,k} \right). \qquad (7)$$

When perfect CSI is available at the transmitter and no constraints are imposed on the structure of the precoder, the optimal precoder $\mathbf{F}_{\mathsf{perf}}$ is channel diagonalizing and is of the form $\mathbf{F}_{\mathsf{perf}} = \widetilde{\mathbf{V}}_{\mathbf{H}} \mathbf{\Lambda}_{\mathsf{wf}}^{1/2}$ where $\mathbf{V}_{\mathbf{H}}\mathbf{\Lambda}_{\mathbf{H}}\mathbf{V}_{\mathbf{H}}^H$ is an eigen-decomposition of $\mathbf{H}^H\mathbf{H}$ with the eigenvalues arranged in non-increasing order, $\widetilde{\mathbf{V}}_{\mathbf{H}}$ is the $N_t \times M$ principal submatrix of $\mathbf{V}_{\mathbf{H}}$, and $\mathbf{\Lambda}_{\mathsf{wf}}$ is an $M \times M$ matrix with non-negative entries only along the leading diagonal and these entries are obtained by waterfilling. In this setting, the mutual information is given by

$$I_{\mathsf{perf}}(\mathbf{s};\widehat{\mathbf{s}}) = \sum_{k=1}^{M} \log_2 \left(1 + \frac{\rho}{M} \mathbf{\Lambda}_{\mathbf{H}}(k) \mathbf{\Lambda}_{\mathsf{wf}}(k)\right). \qquad (8)$$

The optimality of $\mathbf{F}_{\mathsf{perf}}$ with other choices of objective functions is also known; see [11]–[19].

D. Limited Feedback Framework

The focus of this work is on understanding the implications of partial CSI at the transmitter on the performance of the precoding scheme. In particular, there exists a precoder codebook

$$\mathcal{C} = \left\{ \mathbf{F}_i, i = 1, \cdots, 2^B \ : \ \mathrm{Tr}(\mathbf{F}_i^H \mathbf{F}_i) \leq M \right\}. \qquad (9)$$

The most general structure for $\mathbf{F}_i$ is $\mathbf{F}_i = \mathbf{V}_i \mathbf{\Lambda}_i^{1/2}$ where $\mathbf{V}_i$ is an $N_t \times M$ semiunitary matrix and $\mathbf{\Lambda}_i$ is an $M \times M$ non-negative definite, diagonal power allocation matrix. While the structure of the optimal limited feedback codebook of B bits could involve allocating some fraction of B to the power allocation component of $\mathbf{F}_i$, numerical studies [5], [30] indicate that the degradation in performance is minimal when $\mathbf{\Lambda}_i$ is chosen to be fixed (say, $\mathbf{\Lambda}_{\mathsf{stat}}$ with $\mathrm{Tr}(\mathbf{\Lambda}_{\mathsf{stat}}) \leq M$), but designed appropriately, as a function of SNR if necessary, so that it can be adapted easily to statistical variations without recourse to Monte Carlo methods[7].

Motivated by this heuristic, in this work, all the B bits in limited feedback are allocated to quantize the eigenspace of the channel. That is, the codebook is $\mathcal{C} = \{\mathbf{V}_i : \mathbf{V}_i^H\mathbf{V}_i = \mathbf{I}_M\}$ and the index of the codeword that is fed back is

$$j^\star \quad = \quad \arg\max_j \left\{ -\sum_{k=1}^{M} \log_2 \left(\left(\mathbf{I}_M + \frac{\rho}{M} \mathbf{\Lambda}_{\mathsf{stat}}^{1/2} \mathbf{V}_j^H \mathbf{H}^H \mathbf{H} \mathbf{V}_j \mathbf{\Lambda}_{\mathsf{stat}}^{1/2} \right)^{-1}_{k,k} \right) \right\}. \qquad (10)$$

Although computing $j^\star$ is straightforward, the design of an optimal codebook to maximize $I(\mathbf{s};\widehat{\mathbf{s}})$ is achieved via VQ-based codebook constructions in the literature [23], [29]. The high-complexity of such VQ designs leads us to adopt a suboptimal strategy in Sec. IV where the goal is to maximize the average projection of the best quantizer from $\mathcal{C}$ onto $\widetilde{\mathbf{V}}_{\mathbf{H}}$. Towards the precise mathematical formulation of this problem, we need a metric to define distance between two semiunitary matrices.

E. Distance Metrics and Spherical Caps on the Grassmann Manifold

We now recall some well-known facts about the Grassmann manifold. The unit sphere in $\mathbb{C}^{N_t}$, also known as the uni-dimensional[8] complex Stiefel manifold $\mathrm{St}(N_t, 1)$, is defined as $\mathrm{St}(N_t, 1) = \left\{\mathbf{x} \in \mathbb{C}^{N_t} : \|\mathbf{x}\| = 1\right\}$. The invariance of any vector $\mathbf{x}$ to transformations of the form $\mathbf{x} \mapsto e^{j\phi}\mathbf{x}$ in the above definition is incorporated by considering vectors modulo the above map. The partitioning of $\mathrm{St}(N_t, 1)$ by this equivalence map results in the uni-dimensional Grassmann manifold $\mathcal{G}(N_t, 1)$. In short, a point on the Grassmann manifold represents a linear subspace of an Euclidean space. Similarly, the class of $N_t \times M$ semiunitary matrices forms the M-dimensional complex Stiefel manifold $\mathrm{St}(N_t, M)$ and points on the M-dimensional complex Grassmann manifold $\mathcal{G}(N_t, M)$ are identified modulo the M-dimensional unitary space.

A literature survey of packings on $\mathcal{G}(N_t, 1)$ [37]–[39] shows that the dot-product metric defined as,

$$d\left(\mathbf{x}_1, \mathbf{x}_2\right) = \sqrt{1 - |\mathbf{x}_1^H \mathbf{x}_2|^2}, \qquad (11)$$

is the most natural metric from an engineering perspective. Using this distance metric, for any $\gamma < 1$, we can define a

[7] The low-complexity design of $\mathbf{\Lambda}_{\mathsf{stat}}$ will be described in Sec. IV.

[8] Uni-dimensional because its definition is based on the norm of an $N_t \times 1$ vector.

spherical cap with center $\mathbf{o}$ and radius γ (as a *submanifold* on $\mathcal{G}(N_t, 1)$) as the open set

$$\mathbb{O}(\mathbf{o}, \gamma) = \{\mathbf{x} \in \mathcal{G}(N_t, 1) : d(\mathbf{x}, \mathbf{o}) < \gamma\}. \quad (12)$$

In the more general $M > 1$ case, there is no unique distance metric extension [38]. While various well-defined distance metrics can be pursued depending on the space in which the Grassmann manifold is embedded [38], [39], we will focus on the *projection* 2*-norm distance metric* in this work. Here, the distance between two $N_t \times M$ semiunitary matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ is defined as

$$d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_2) = \|\mathbf{V}_1\mathbf{V}_1^H - \mathbf{V}_2\mathbf{V}_2^H\|_2 \quad (13)$$
$$= \max_i |\lambda_i\left(\mathbf{V}_1\mathbf{V}_1^H - \mathbf{V}_2\mathbf{V}_2^H\right)|. \quad (14)$$

If the principal angles[9] between the subspaces spanned by columns of $\mathbf{V}_1$ and $\mathbf{V}_2$ are denoted by $\{\theta_m,\, m = 1, \cdots, M\}$, it is known that [38] the singular values of $\mathbf{V}_2^H\mathbf{V}_1\mathbf{V}_1^H\mathbf{V}_2$ are $\{\cos^2(\theta_m),\, m = 1, \cdots, M\}$. Furthermore, the projection 2-norm distance can be written as

$$d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_2) = \max_{m=1,\cdots,M} |\sin(\theta_m)|. \quad (15)$$

A particular choice of the distance metric is not extremely critical in precoder optimization since codebooks designed with different choices of distance metrics result in near-identical performance for SNR regimes of practical interest [28], [30]. We now state some properties of the projection 2-norm metric.

Lemma 1: In the $M = 1$ case, the projection 2-norm metric reduces to the standard dot-product metric.

Proof: Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two unit-normed $N_t \times 1$ vectors. The proof follows trivially from the principal angle interpretation of $d_{\mathrm{proj},\,2}(\cdot)$ in (15). In the $M = 1$ case, $\cos(\theta_1) = |\mathbf{v}_1^H\mathbf{v}_2|$ and from (15), it follows that

$$d_{\mathrm{proj},\,2}(\mathbf{v}_1, \mathbf{v}_2) = |\sin(\theta_1)| = \sqrt{1 - |\mathbf{v}_1^H\mathbf{v}_2|^2}. \quad (16)$$

In Appendix A, we present an alternate proof of the above claim rooted in matrix algebra. ∎

Proposition 1: The following are true.
1) $0 \leq d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_2) \leq 1$,
2) More precisely,

$$d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_2) = \sqrt{1 - \lambda_{\min}(\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H\mathbf{V}_1)}, \quad \text{and} \quad (17)$$

3) Equality in the lower bound of 1) occurs if and only if $\mathbf{V}_1 = \mathbf{V}_2$ on $\mathcal{G}(N_t, M)$ while equality is possible in the upper bound if and only if $\lambda_{\min}(\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H\mathbf{V}_1) = 0$.

Proof: The proof is obvious following the geometric interpretation of $d_{\mathrm{proj},\,2}(\cdot)$. The first claim follows immediately since the principal angles are in $[0, \pi/2]$ by definition. The second claim follows from the connection between singular values and the principal angles in (15). See Appendix A for an alternate proof of the above claims. For the third claim, note that if $d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_2) = 0$, then $\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H\mathbf{V}_1 = \mathbf{I}_M$. Thus, $\mathbf{V}_1^H\mathbf{V}_2$ is $M \times M$, unitary and hence, $\mathbf{V}_1 = \mathbf{V}_2$ on $\mathcal{G}(N_t, M)$. The other direction of the statement follows trivially. Both the directions of the upper bound follow from the expression in 2). ∎

Lemma 2: Let $\mathbf{U}$ be a fixed $N_t \times N_t$ unitary matrix. Then, $d_{\mathrm{proj},\,2}(\mathbf{U}\mathbf{V}_1, \mathbf{U}\mathbf{V}_2) = d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_2)$.

Proof: The proof is obvious from the geometric interpretation of (15) since rotation by $\mathbf{U}$ does not change the relative orientation of the subspaces and hence, the distance properties. Alternately, using the fact that the eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ are the same, the proof is also obvious from (13). ∎

Once a choice of distance metric has been settled, the notion of a spherical cap in (12) can be generalized as a submanifold on $\mathcal{G}(N_t, M)$ with center $\mathbf{O}$ and radius γ as the open set

$$\mathbb{O}(\mathbf{O}, \gamma) = \{\mathbf{X} \in \mathcal{G}(N_t, M) : d_{\mathrm{proj},\,2}(\mathbf{X}, \mathbf{O}) < \gamma\}. \quad (18)$$

The codebook design problem can now be simply stated as:

$$\text{Construct } \mathcal{C} = \{\mathbf{V}_i, i = 1, \cdots, 2^B\}$$
$$\text{s.t. } E_{\mathbf{H}}\left[\min_{i=1,\cdots,2^B} d_{\mathrm{proj},\,2}(\mathbf{V}_i, \widetilde{\mathbf{V}}_{\mathbf{H}})\right] \text{ is minimized.}$$

Towards a systematic solution to this problem, we now describe some related prior work [30].

III. Matched Versus Mismatched Channels

The case of unconstrained precoding with genie-aided perfect CSI and the optimality of $\mathbf{F}_{\mathsf{perf}} = \widetilde{\mathbf{V}}_{\mathbf{H}}\mathbf{\Lambda}_{\mathsf{wf}}^{1/2}$ was summarized in Sec. II-C. Knowledge of $\widetilde{\mathbf{V}}_{\mathbf{H}}$ and $\mathbf{\Lambda}_{\mathsf{wf}}$ at the transmitter necessitate the tracking of the evolution of $\mathbf{H}$ which is difficult in practice. To avoid this problem and to reduce the complexity of precoding, *structured precoding*[10] was introduced in [30]. The readers are referred to [30] for the following results.

- When the precoder is assumed to be structured, the optimal choice of $\mathbf{V}$ under perfect CSI is still $\widetilde{\mathbf{V}}_{\mathbf{H}}$. This optimality is assured for many different classes of objective functions apart from the case of maximizing mutual information. When only statistical information is available at the transmitter, the optimal choice of $\mathbf{V}$ is $\mathbf{V}_{\mathsf{stat}}$ where

$$\mathbf{V}_{\mathsf{stat}} = [\mathbf{u}_1 \cdots \mathbf{u}_M], \quad \mathbf{U}_t = [\mathbf{u}_1 \cdots \mathbf{u}_{N_t}], \quad (19)$$

  where $\mathbf{\Sigma}_t = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H$ is as in (3). We call these two schemes *optimal structured* and *statistical structured precoding schemes*, respectively.
- We studied the performance loss between these two schemes as a function of the channel statistics. The notion of *matched* and *mismatched channels*, introduced in [30], correspond to the cases where the relative performance of the statistical structured precoder is closest to and farthest from the perfect CSI structured precoder, respectively. Even knowledge of only channel statistics at the transmitter results in near-optimality for the matched channel

[9]The principal angles are quantities in $[0, \pi/2]$ describing the relative orientation of one subspace with the other, more precisely to their bi-orthogonal basis expansions (see [40]), which are independent of the given representation of the subspaces. We thank one of the reviewers for this observation.

[10]In structured precoding, the precoder has the form $\mathbf{F} = \mathbf{V}\mathbf{\Lambda}_{\mathsf{stat}}^{1/2}$ where $\mathbf{V}$ is an $N_t \times M$ semiunitary matrix that can be optimized, and $\mathbf{\Lambda}_{\mathsf{stat}}$ is an $M \times M$ fixed, rank-M power allocation matrix chosen *a priori*.

case! Thus, it is important to note that any feedback (limited or otherwise) is helpful *only* in mismatched channels. This conclusion is a generalization of our earlier beamforming result [26], [41].

- More specifically, a matched channel is one where the channel is effectively matched to the precoding scheme (in particular, the precoding rank M) with the following two conditioning properties being true: 1) The M-dominant eigenvalues of $\mathbf{\Sigma}_t$ are *well-conditioned*[11], whereas the remaining $(N_t - M)$ eigenvalues are *ill-conditioned* away from the dominant ones, and 2) $\mathbf{\Sigma}_r$ is also *well-conditioned*. A mismatched channel is one where both $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_r$ are ill-conditioned, with the additional condition that $\mathsf{rank}(\mathbf{H}) \geq M$ with probability 1. We also proposed metrics to capture the degree of channel-to-precoder scheme matching continuously and showed that these metrics can be used to compare two channels in terms of their average mutual information performance [30, Fig. 2].

Henceforth, the focus is on mismatched channels primarily because the potential to bridge the performance gap between the statistical and perfect CSI schemes is maximum (see numerical studies in [30] for performance of statistical schemes over matched channels). Our goal is to construct a systematic, statistics-dependent codebook (of a fixed size 2^B) that ensures this bridging.

IV. Quantized Feedback Designs to Bridge the Performance Gap

In contrast to the i.i.d. case where the isotropicity of $\widetilde{\mathbf{V}}_{\mathbf{H}}$ leads to a design [28] based on Grassmannian subspace packings [39], spatial correlation skews this isotropicity and poses fundamental challenges. The optimality of $\mathbf{V}_{\mathsf{stat}}$ in the statistics-only case and its near-optimality in the matched channel case suggests that when we have the freedom to pick more than one codeword ($B > 0$), the codewords should correspond to a local quantization[12] of $\mathbf{V}_{\mathsf{stat}}$. This heuristic is also motivated by numerical studies [26, Figs. 1 and 2], [30] that show that for most reasonable channels, the probability density function of $d_{\mathrm{proj},2}(\widetilde{\mathbf{V}}_{\mathbf{H}}, \mathbf{V}_{\mathsf{stat}})$ is concentrated around 0, suggesting that a local quantization could lead to improved performance.

Building on our prior work in the beamforming case [26], [41] where we designed codebooks on a local quantization principle, we now develop a multi-mode generalization. The main difference here is in packing subspaces instead of lines and in the choice of an appropriate distance metric. The proposed design has three components: 1) A statistical component, 2) Local perturbation components, and 3) An RVQ component. The cardinalities of these components are denoted by $N_{\mathsf{stat}}, N_{\mathsf{loc}}$ and N_{rvq} with the feedback rate (per channel use) defined by $B = \log_2(N_{\mathsf{stat}} + N_{\mathsf{loc}} + N_{\mathsf{rvq}})$.

[11] If $\mathbf{\Lambda}_t(1) \geq \cdots \geq \mathbf{\Lambda}_t(M)$ denote the first M eigenvalues of $\mathbf{\Sigma}_t$ and $\frac{\mathbf{\Lambda}_t(1)}{\mathbf{\Lambda}_t(M)}$ is (or is not) significantly larger than 1, we loosely say that these eigenvalues are ill-(or well-)conditioned.

[12] The notion of local quantization will be made mathematically precise shortly.

Statistical Component: We first need to identify the dominant M-dimensional subspaces of $\mathbf{\Sigma}_t$. Note that this identification cannot be based on distance metrics because if $\mathbf{V}_1$ and $\mathbf{V}_2$ denote any two distinct M-dimensional subspaces of $\mathbf{\Sigma}_t$, then from Prop. 1 we have

$$d_{\mathrm{proj},2}(\mathbf{V}_1, \mathbf{V}_2) = 1. \tag{20}$$

Since there exists no granularity in subspace dominance based on distance metrics, we define the *generalized eigenvalue of a M-dimensional subspace* as follows. The generalized eigenvalue of the subspace $[\mathbf{u}_{i_1} \cdots \mathbf{u}_{i_M}]$ of $\mathbf{\Sigma}_t$ is

$$\mu_{[\mathbf{u}_{i_1} \cdots \mathbf{u}_{i_M}]} \triangleq \prod_{j=1}^{M} \mathbf{\Lambda}_t(i_j). \tag{21}$$

In the above definition, $\{i_1, \cdots, i_M\}$ is a distinct M-tuple of $\{1, \cdots, N_t\}$ which implies that there are $\binom{N_t}{M}$ distinct M-dimensional subspaces of $\mathbf{\Sigma}_t$. Note that the subspace spanned by $\mathbf{V}_{\mathsf{stat}} = [\mathbf{u}_1 \cdots \mathbf{u}_M]$ results in the largest generalized eigenvalue, and we will denote $\mu_{[\mathbf{u}_1 \cdots \mathbf{u}_M]}$ by μ_1 (for short). Similarly, when necessary, we will use μ_i to denote the i-th dominant generalized eigenvalue of $\mathbf{\Sigma}_t$.

The precise probability distribution of $d_{\mathrm{proj},2}(\mathbf{V}_{\mathsf{stat}}, \widetilde{\mathbf{V}}_{\mathbf{H}})$ is dependent on the separation (gap) between the generalized eigenvalues of $\mathbf{\Sigma}_t$. For example, if μ_1 and μ_2 are close to each other, there is a non-negligible probability for the event that the best quantizer is the subspace corresponding to μ_2 and hence, the distance between $\mathbf{V}_{\mathsf{stat}}$ and $\widetilde{\mathbf{V}}_{\mathbf{H}}$ could be arbitrarily close (follows from Prop. 1 and the triangle inequality for the $d_{\mathrm{proj},2}(\cdot)$ metric) to 1. On the other hand, if μ_1 is much larger than the other generalized eigenvalues, it is intuitive to expect that $d_{\mathrm{proj},2}(\mathbf{V}_{\mathsf{stat}}, \widetilde{\mathbf{V}}_{\mathbf{H}})$ is concentrated around zero and hence, statistical precoding is near-optimal. Thus the gap between μ_1 and the rest of the generalized eigenvalues heuristically determines the statistical component.

In our work, a threshold β is chosen *a priori* and the statistical component $\mathcal{S}$ consists of all M-dimensional subspaces such that their generalized eigenvalues exceed β. That is,

$$\mathcal{S} = \Bigg\{ [\mathbf{u}_{i_1} \cdots \mathbf{u}_{i_M}] : 1 \leq i_1 < i_2 < \cdots < i_M \leq N_t \text{ and } \frac{\mu_{[\mathbf{u}_{i_1} \cdots \mathbf{u}_{i_M}]}}{\mu_1} = \prod_{j=1}^{M} \frac{\mathbf{\Lambda}_t(i_j)}{\mathbf{\Lambda}_t(j)} > \beta \Bigg\}. \tag{22}$$

The cardinality of $\mathcal{S}$ is denoted by N_{stat}. Note that if β is large, N_{stat} is small and $\mathcal{S}$ may not accurately quantize the space of statistical eigenvectors of $\mathbf{H}$. On the other hand, if β is small, the feedback overhead for the statistical component may be too large. This heuristical trade-off governs the choice of β. In our numerical studies, we found setting $\beta \approx 0.2$ to result in a good trade-off. Future work will study this optimization problem more carefully.

Local Components: Let $\mathbf{V}_1$ denote the i-th member of the statistical component. The local component, $\mathcal{L}_i$, around $\mathbf{V}_1$ is defined as

$$\mathcal{L}_i = \left\{ \mathbf{V}_j, j = 2, \cdots, N_{\mathsf{loc}}^i + 1 : \mathbf{V}_j^H \mathbf{V}_j = \mathbf{I}_M \right\} \tag{23}$$

such that these N_{loc}^i codewords are localized and well-packed around $\mathbf{V}_1$. The notions of *localized and well-packed* are made mathematically precise in Sec. V. The heuristic here is to choose N_{loc}^i in proportion to μ_i. The following observation motivates this heuristic: the larger the separation of μ_1 (corresponding to $\mathbf{V}_{\mathsf{stat}}$) from μ_2 or the more matched $\mathbf{\Sigma}_t$ is, the lesser the relevance of the sub-dominant subspaces in terms of precoding and hence, the smaller the values of $\{N_{\mathsf{loc}}^i\}, i > 1$ need to be. These $N_{\mathsf{loc}} = \sum_{i=1}^{N_{\mathsf{stat}}} N_{\mathsf{loc}}^i$ codewords form the *local component* of our codebook design.

While $\mathcal{L}_i$ can be constructed brute-force via VQ or a Monte Carlo method as theoretically conceived in [23], [29], we provide low-complexity alternatives in Sec. V.

RVQ Component: If B is sufficiently large, there is a need to refine the quantization of $\widetilde{\mathbf{V}}_{\mathbf{H}}$. This is because while the statistical and the local components lead to significant limited feedback gains when B is small, their marginal utility diminishes as B increases. Empirical observations suggest that the addition of a few codewords obtained via RVQ can help in boosting performance significantly. In this context, we set

$$N_{\mathsf{rvq}} \triangleq 2^B - N_{\mathsf{stat}} - N_{\mathsf{loc}}. \tag{24}$$

Random channel matrices $\{\mathbf{H}_i, i = 1, \cdots, N_{\mathsf{rvq}}\}$ are generated according to the relationship in (2) and the RVQ component, $\mathcal{RVQ}$, is given as

$$\mathcal{RVQ} = \Big\{\widetilde{\mathbf{V}}_{\mathbf{H}_i} : \mathbf{H}_i = \mathbf{U}_{\mathbf{H}_i}\mathbf{\Lambda}_{\mathbf{H}_i}\mathbf{V}_{\mathbf{H}_i}^H \text{ and } \widetilde{\mathbf{V}}_{\mathbf{H}_i} \text{ is the principal } N_t \times M \text{ submatrix of } \mathbf{V}_{\mathbf{H}_i}\Big\}. \tag{25}$$

Note that the RVQ component can be generated with low-complexity once the statistics are known perfectly.

In practice, the choice of B is determined by the application. For example, in the design of Third Generation wireless systems, B is on the order of 1 to 4 per sub-carrier chunk (which usually consists of 20-30 OFDM tones) [3]. The choice of β determines what value N_{stat} should take, whereas N_{loc} is determined by the relative strength of the eigen-modes and the above guidelines. The choice of B then dictates N_{rvq} as per the relationship in (24).

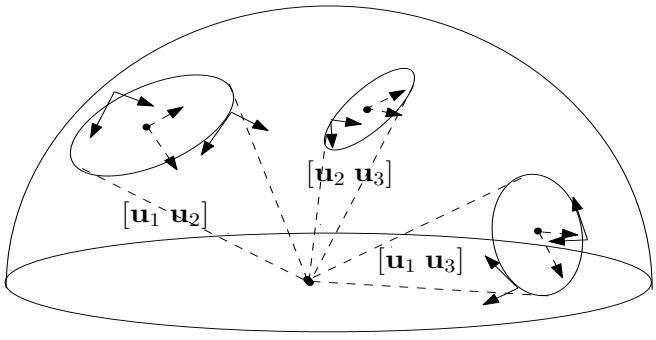


Fig. 1. Proposed Codebook Design for $N_t = 3, M = 2$, and $B = 3$ with only the statistical and local components.

In Fig. 1, we illustrate the design of a codebook with statistical and local components (no RVQ component) for $N_t = 3, M = 2$, $N_{\mathsf{stat}} = 3$, $N_{\mathsf{loc}}^1 = N_{\mathsf{loc}}^2 = 2$ and $N_{\mathsf{loc}}^3 = 1$. If $\mathbf{U}_t = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$, then the three statistical transmit eigenspaces with $M = 2$ are those spanned by $[\mathbf{u}_1\ \mathbf{u}_2]$, $[\mathbf{u}_1\ \mathbf{u}_3]$ and $[\mathbf{u}_2\ \mathbf{u}_3]$. The "directions" corresponding to these subspaces are symbolically represented in the figure with dashed lines. The first local component consists of two codewords around $[\mathbf{u}_1\ \mathbf{u}_2]$ and so on. Since there are eight codewords in our design, this codebook can be parameterized with $B = 3$ bits.

A. Power Allocation

It is preferred that the power allocation matrix $\mathbf{\Lambda}_{\mathsf{stat}}$ be only dependent on the channel statistics and be easily adaptable to statistical variation. The optimal choice of $\mathbf{\Lambda}_{\mathsf{stat}}$ needs to be constructed via a Monte Carlo algorithm [29] which is difficult to implement as well as adapt to statistical variations with low-complexity. As an alternative, we consider three low-complexity power allocations: 1) uniform power allocation across the excited modes, 2) waterfilling based on $\mathbf{\Lambda}_t(i),\ i = 1, \cdots, M$, and 3) power allocation proportional to the transmit eigenvalues. The last two schemes have near-identical performances and are near-optimal in the low-SNR regime while uniform power allocation is more useful in the high-SNR regime [5], [30].

B. Codeword Selection

The receiver acquires the channel information at the start of a coherence block and it computes the index of the optimal codeword from the codebook that maximizes the instantaneous mutual information. The receiver then communicates to the transmitter the index of the optimal codeword with B bits. The transmitter uses the optimal codeword along with an appropriate power allocation to communicate over the remaining period in the coherence block.

V. Rotating and Scaling Spherical Caps on $\mathcal{G}(N_t, M)$

We now construct mathematical maps to ensure that the codebook design proposed in Sec. IV can be realized with low-complexity. For this, we need the notion of a *root codeset*[13]. A root codeset, $\mathcal{R}(\mathbf{V}_1, N, \theta, \gamma)$, is a set of N semiunitary matrices ($\mathbf{V}_i,\ i = 1, \cdots, N$) satisfying the following properties that signify a 'good' local quantization:

1) ***Localization:*** The root codeset is localized (centered) around $\mathbf{V}_1$, which is labeled as the *center of the root codeset*. There exists a $\theta \in (0, \pi/2)$ such that $d_{\mathrm{proj},2}(\mathbf{V}_1, \mathbf{V}_i) \leq \sin(\theta)$ for all $i \neq 1$. The smaller the value of θ, the more localized a packing. This is illustrated in Fig. 2 where a set of $N = 5$ precoders form the localized root codeset in the $N_t = 3, M = 2$ setting.
2) ***Well-Packing:*** The codewords in $\mathcal{R}(\mathbf{V}_1, N, \theta, \gamma)$ are well-packed (well-separated). That is, the minimum distance of the packing $d_{\min}(\mathcal{R})$ satisfies

$$d_{\min}(\mathcal{R}) \triangleq \min_{i \neq j} d_{\mathrm{proj},2}(\mathbf{V}_i, \mathbf{V}_j) \geq \gamma \quad \text{for some } \gamma \in (0, \gamma_{\max}(N_t, M, N, \theta)). \tag{26}$$

[13]We use the term root codeset to indicate that the construction of $\mathcal{C}$ is rooted in the design of a 'good' $\mathcal{R}$.

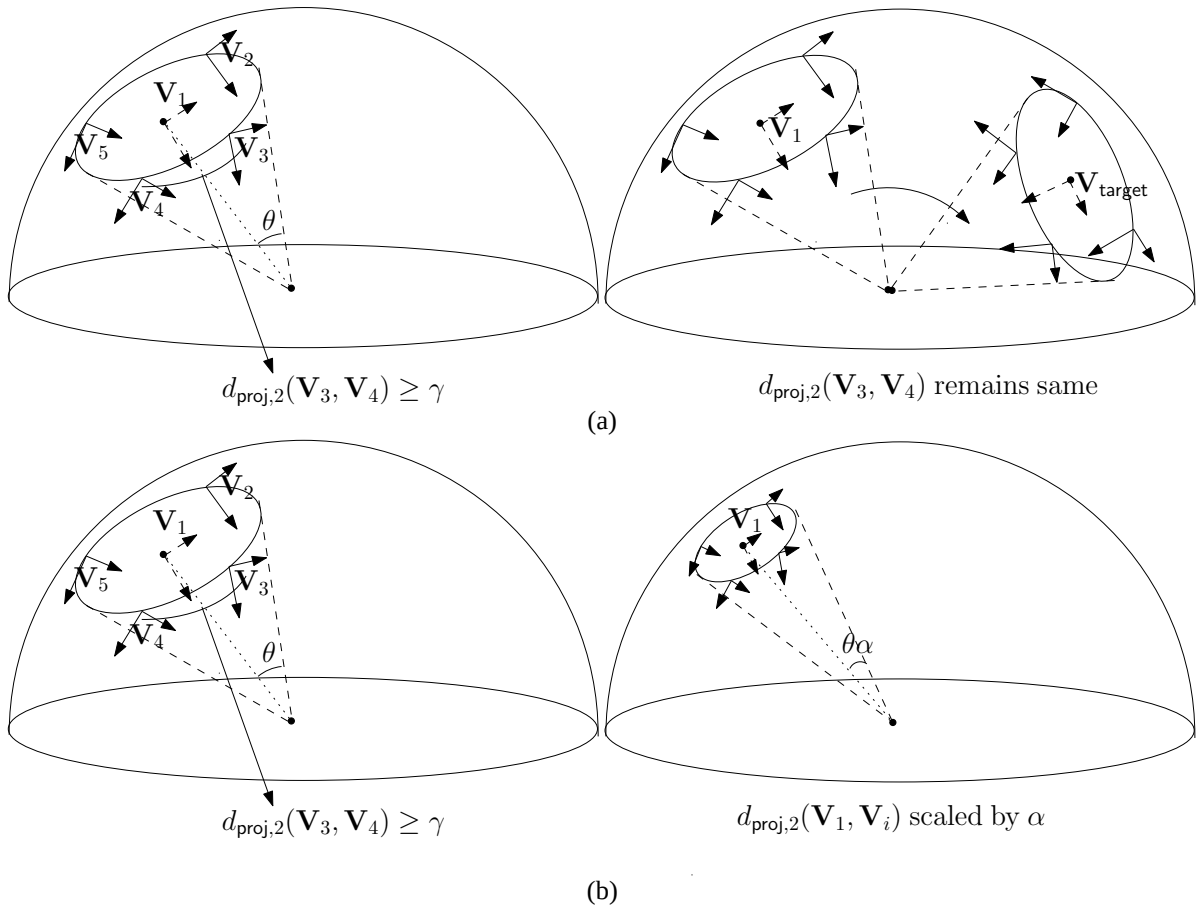


Fig. 2. (a) Rotation of a root codeset of semiunitary precoders $\{\mathbf{V}_i, i = 1, \cdots, 5\}$ with $N_t = 3$ and $M = 2$. The root codeset satisfies the localization and well-packing properties described in Sec. V. The distance between any two precoders remains unchanged after rotation. (b) Scaling of the root codeset by α. The position of $\mathbf{V}_1$ remains unchanged after scaling.

> The larger the value of γ, the more well-packed $\mathcal{R}$ is. Hence γ can also be viewed as an abstract measure of the packing density. Here, $\gamma_{\max}(N_t, M, N, \theta)$ is the maximum possible packing density[14] achievable in the Grassmann manifold $\mathcal{G}(N_t, M)$ with N codewords localized in a cap of radius $\sin(\theta)$.

Note that for any fixed choice of N_t, M and N, it is intuitive to expect that $\gamma_{\max}(N_t, M, N, \theta)$ decreases as θ decreases. In other words, the above two properties are in some sense conflicting with a root codeset that is more localized necessarily forced to have a small packing density and *vice versa*.

Despite this apparent difficulty, it is important to note that a packing with the above properties can *always* be constructed, either via algebraic methods or via a vector quantization [22], [23] approach (that is, a brute-force search via Monte Carlo methods). Furthermore, $\mathcal{R}$ needs to be constructed (offline) just once, and once this has been done, $\mathcal{C}$ can be designed for any statistics starting from $\mathcal{R}$. For this, we now show how mathematical operations can be constructed to perform the following two tasks:
1) Given $\mathcal{R}(\mathbf{V}_1, N, \theta, \gamma)$, how can we center $\mathcal{R}$ around $\mathbf{V}_{\text{target}}$ to obtain $\mathcal{R}(\mathbf{V}_{\text{target}}, N, \theta, \gamma)$ without having to resort to a VQ-type codebook construction *again*? That is, we seek a map to rotate the center of $\mathcal{R}$ to $\mathbf{V}_{\text{target}}$ without changing the packing density, and
2) Given $\mathcal{R}(\mathbf{V}_1, N, \theta, \gamma)$ and some fixed $\alpha \in (0, 1)$, how can we scale $\mathcal{R}$ to obtain $\mathcal{R}(\mathbf{V}_1, N, \theta\alpha, \gamma')$ for some $\gamma' \leq \gamma$? That is, we seek a map to reduce the radius of $\mathcal{R}$ without changing its center.

While we develop such maps for spherical caps/submanifolds, we will state the results as applicable to finite element subsets of $\mathcal{G}(N_t, M)$. But prior to that, we recall results from recent work [42], [43] where rotation and scaling maps to solve 1) and 2) (as above) have been proposed in the beamforming case ($M = 1$). The rotation map is straightforward and is effected by an appropriately chosen unitary matrix. In contrast to the rotation operation, the scaling map 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. It is to be noted that both rotation and scaling maps are non-unique. We summarize the maps of [42], [43] in the following lemma[15] for $M = 1$.

Lemma 3 (See [42], [43]): Let $\mathcal{R}(\mathbf{v}_1, N, \theta, \gamma) = \{\mathbf{v}_i, i = 1, \cdots, N\}$ be a root codeset in $\mathcal{G}(N_t, 1)$. Rotation of $\mathcal{R}$ to $\mathbf{v}_{\text{target}}$ is trivially achieved by the map $r(\mathbf{v}_i) = \mathbf{U}_{\text{target}}\mathbf{v}_i$

[14]While the exact characterization of $\gamma_{\max}(N_t, M, N, \theta)$ remains an open problem for general values of N_t, M, N and θ, some bounds have been established; see [20], [25], [38], [39] and references therein.

[15]The readers are referred to [26] for details of the proof.

with $\mathbf{U}_{\mathsf{target}}$ satisfying[16] $\mathbf{v}_{\mathsf{target}} = \mathbf{U}_{\mathsf{target}}\mathbf{v}_1$. Scaling by α is achieved by the composition

$$s_{\mathsf{bf}} = r_{\mathsf{vertex}}^{-1} \circ s_{\mathsf{vertex}} \circ r_{\mathsf{vertex}} \tag{27}$$

$$s_{\mathsf{bf}}(\mathbf{v}_i) = \mathbf{v}_1\sqrt{1-\alpha^2(1-|\mathbf{v}_1^H\mathbf{v}_i|^2)}e^{j\angle \mathbf{v}_1^H\mathbf{v}_i} + \alpha\mathbf{v}_1^{\perp}\mathbf{v}_1^{\perp,\,H}\mathbf{v}_i \tag{28}$$

■

In (27), $r_{\mathsf{vertex}}(\cdot)$ is the map that induces the rotation of $\mathcal{R}$ to $\mathbf{v}_{\mathsf{vertex}} = [1,\, 0, \cdots, 0]^T$ (a vertex of the unit cube), $s_{\mathsf{vertex}} : \mathbb{O}(\mathbf{v}_{\mathsf{vertex}}, \gamma) \mapsto \mathbb{O}(\mathbf{v}_{\mathsf{vertex}}, \alpha\gamma)$ is a vertex scaling map with

$$s_{\mathsf{vertex}}\left([r_1 e^{j\theta_1},\, r_2 e^{j\theta_2}, \cdots, r_{N_t}e^{j\theta_{N_t}}]^T\right)$$
$$= \left[\sqrt{1-\alpha^2(1-r_1^2)}e^{j\theta_1},\, \alpha\, r_2 e^{j\theta_2}, \cdots, \alpha\, r_{N_t}e^{j\theta_{N_t}}\right]^T, \tag{29}$$

and the argument in the above equation is in its polar form. We now generalize the above result to the precoding scenario, $M > 1$.

Theorem 1: Let $\mathcal{R}(\mathbf{V}_1, N, \theta, \gamma) = \{\mathbf{V}_i, i = 1, \cdots, N\}$ be a root codeset in $\mathcal{G}(N_t, M)$. Rotation of $\mathcal{R}$ to $\mathbf{V}_{\mathsf{target}}$ is achieved by

$$\begin{aligned}\mathcal{R}(\mathbf{V}_{\mathsf{target}}, N, \theta, \gamma) &= \{\mathbf{G}_i, i = 1, \cdots, N\} \\ \text{where } \mathbf{G}_i &= \mathbf{U}_{\mathbf{V}_{\mathsf{target}}}\,\mathbf{U}_{\mathbf{V}_1}^H\,\mathbf{V}_i\end{aligned} \tag{30}$$

with unitary matrices $\mathbf{U}_{\mathbf{V}_1}$ and $\mathbf{U}_{\mathbf{V}_{\mathsf{target}}}$ defined as $\mathbf{U}_{\mathbf{V}_1} = \left[\mathbf{V}_1 \;\; \mathbf{V}_1^{\mathsf{null}}\right]$ and $\mathbf{U}_{\mathbf{V}_{\mathsf{target}}} = \left[\mathbf{V}_{\mathsf{target}} \;\; \mathbf{V}_{\mathsf{target}}^{\mathsf{null}}\right]$. Here, $\mathbf{V}_1^{\mathsf{null}}$ and $\mathbf{V}_{\mathsf{target}}^{\mathsf{null}}$ are $N_t \times (N_t - M)$-dimensional representatives of the null-spaces of $\mathbf{V}_1$ and $\mathbf{V}_{\mathsf{target}}$, respectively.

Proof: The proof follows trivially from Lemma 2 since rotation by a common unitary matrix $\mathbf{U}_{\mathbf{V}_{\mathsf{target}}}\,\mathbf{U}_{\mathbf{V}_1}^H$ does not alter the distance properties of the root codeset. ■

Note that there exists more than one basis for the null-space and therefore the usage of the term "representative" in the statement of the theorem. The lack of a unique representative for the null-space is responsible for the non-uniqueness of the rotation map that can effect a desired rotation.

Before we get into the most general form of the scaling map in Appendix B, we illustrate a special case of it so as to provide insights into its construction. Let $\mathbf{V}_1 = [\mathbf{v}_1 \cdots \mathbf{v}_M]$ where $\mathbf{v}_i$ is an $N_t \times 1$ vector and is the i-th column of $\mathbf{V}_1$. Define the map $s(\cdot)$ by

$$s(\mathbf{V}_i) = \left[\begin{array}{ccccc}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{M-1} & \kappa\mathbf{v}_M + \delta\mathbf{v}_{M+1}\end{array}\right] \tag{31}$$

$$\kappa = \sqrt{1-\alpha^2\left(1-\lambda_{\min}\left(\mathbf{V}_1^H\mathbf{V}_i\mathbf{V}_i^H\mathbf{V}_1\right)\right)}, \quad \delta = \sqrt{1-\kappa^2}$$

and $\mathbf{v}_{M+1}$ is orthogonal to $\mathbf{V}_1$ (that is, $\mathbf{v}_{M+1}^H\mathbf{V}_1 = \mathbf{0}_{1\times M}$). We illustrate three properties satisfied by $s(\cdot)$ which ensures that it can scale submanifolds. A straightforward consequence of the orthonormality of $\mathbf{v}_i, i = 1, \cdots, M+1$ in $\mathbb{C}^{N_t}$ and $\kappa^2 + \delta^2 = 1$ is that $s(\mathbf{V}_i)^H s(\mathbf{V}_i) = \mathbf{I}_M$. For $s(\mathbf{V}_1)$, note that $\kappa = 1$ and $\delta = 0$ which results in $s(\mathbf{V}_1) = \mathbf{V}_1$.

[16]One possible choice of $\mathbf{U}_{\mathsf{target}}$ is $\mathbf{U}_{\mathsf{target}} = \left[\mathbf{v}_{\mathsf{target}}\;\mathbf{v}_{\mathsf{target}}^{\perp}\right]\left[\begin{array}{cc}\mathbf{v}_1 & \mathbf{v}_1^{\perp}\end{array}\right]^H$ where $\mathbf{v}_{\mathsf{target}}^{\perp}$ and $\mathbf{v}_1^{\perp}$ refer to matrix representatives from the $N_t \times (N_t - 1)$ dimensional null-space of $\mathbf{v}_{\mathsf{target}}$ and $\mathbf{v}_1$, respectively. That is, $\mathbf{v}_1^{\perp,\,H}\mathbf{v}_1^{\perp} = \mathbf{I}_{N_t-1}$ and $\mathbf{v}_1^{\perp,\,H}\mathbf{v}_1 = \mathbf{0}_{N_t-1\times 1}$.

Proposition 2: We also have $d_{\mathrm{proj},\,2}(s(\mathbf{V}_1), s(\mathbf{V}_i)) = \alpha d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_i)$ for any $i \neq 1$. Thus, $s(\cdot)$ induces the scaling of $\mathcal{R}$ by α.

Proof : Note that $d_{\mathrm{proj},\,2}(s(\mathbf{V}_1), s(\mathbf{V}_i))$

$$\stackrel{(a)}{=} d_{\mathrm{proj},\,2}(\mathbf{V}_1, s(\mathbf{V}_i)) = \max_i \left|\lambda_i\left(\mathbf{V}_1\mathbf{V}_1^H - s(\mathbf{V}_i)s(\mathbf{V}_i)^H\right)\right|$$
$$\stackrel{(b)}{=} \max_i \left|\lambda_i\left(\mathbf{v}_M\mathbf{v}_M^H - (\kappa\mathbf{v}_M + \delta\mathbf{v}_{M+1})(\kappa\mathbf{v}_M + \delta\mathbf{v}_{M+1})^H\right)\right|$$

where in (a) we have used $s(\mathbf{V}_1) = \mathbf{V}_1$ and (b) follows from (31). Using the eigenvalue computation trick of Lemma 1 in Appendix A, observe that the square of $\max_i |\lambda_i|$ in the above equation satisfies

$$\begin{aligned}\max_i |\lambda_i|^2 &= 1 - \left|\mathbf{v}_M^H(\kappa\mathbf{v}_M + \delta\mathbf{v}_{M+1})\right|^2 = 1 - \kappa^2 &(32)\\ &= \alpha^2(1 - \lambda_{\min}(\mathbf{V}_1^H\mathbf{V}_i\mathbf{V}_i^H\mathbf{V}_1)) &(33)\\ &= \left(\alpha d_{\mathrm{proj},\,2}(\mathbf{V}_1, \mathbf{V}_i)\right)^2. &(34)\end{aligned}$$

■

The choice of $\mathbf{v}_{M+1}$ is not unique and it is not clear whether the map in (31) is unique modulo the choice of $\mathbf{v}_{M+1}$. Furthermore, note that when $(N_t - M) \geq M$, $s(\mathbf{V}_i)$ can be written as

$$s(\mathbf{V}_i) = \mathbf{V}_1\mathbf{A}_i + \mathbf{V}_1^{\mathsf{null}}\mathbf{B}_i \tag{35}$$

where $\mathbf{A}_i = \mathrm{diag}([1, \cdots, 1, \kappa])$ and $\mathbf{B}_i$ has only one non-zero entry which is at the (M, M)-th location and its value is δ. In Appendix B, we resolve the uniqueness issue and construct the most general form of $s(\cdot)$. We also show that the most general form of $s(\mathbf{V}_i)$ is of the form in (35) for a suitable choice of $\mathbf{A}_i$ and $\mathbf{B}_i$.

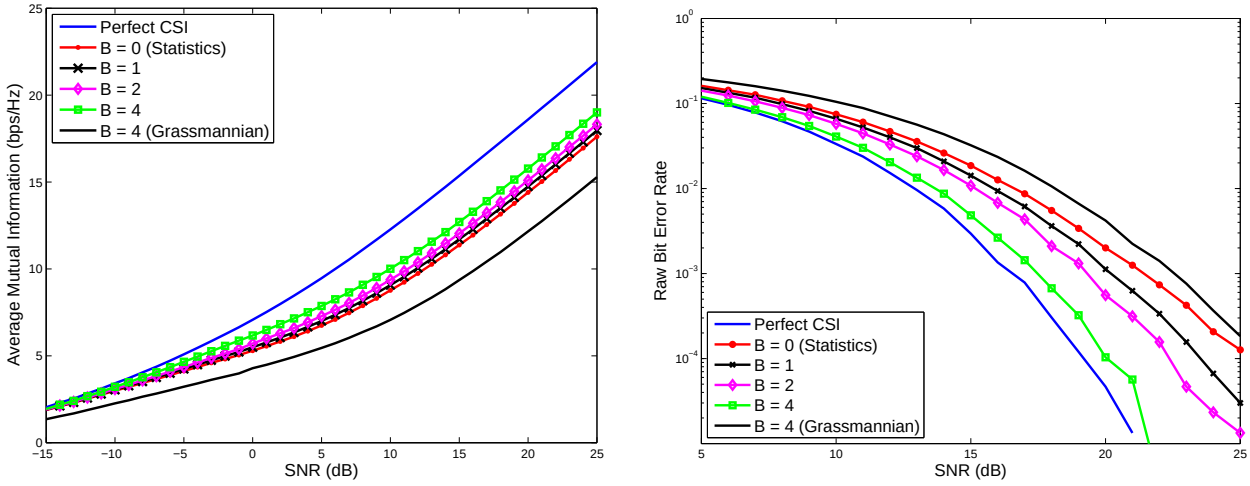
Corollary 1: It can also be easily checked that in the special case of $M = 1$, the scaling map proposed in (31) (and extended in Theorem 2 of Appendix B) is a generalization of the map proposed in Lemma 3 (see (28)). ■

A. Low-Complexity Generation of Local Components

We now illustrate how the theory of rotation and scaling maps can be used to construct precoding codebooks with low-complexity.

Root Codeset Generation: A root codeset that satisfies the localization and well-packing conditions as described above is constructed via VQ and stored offline. The number of codewords in the root codeset is larger than N_{loc}^1 so as to ensure that any local component has a cardinality smaller than that of the root codeset. Furthermore, since the scaling map can only ensure that the output packing is more localized than the input packing, we need to pick θ sufficiently large, but smaller than $\pi/2$. The quantity $\gamma_{\max}(N_t, M, N, \theta)$ corresponding to the choices of N_t, M, N and θ is determined via Monte Carlo techniques and some γ is chosen in the interval $(0, \gamma_{\max}(N_t, M, N, \theta))$.

Local Components: For each member of the statistical component, we rotate the root codeset (via the rotation map of Theorem 1) to the $N_t \times M$ matrix corresponding to the subspace of $\mathbf{\Sigma}_t$ in the statistical component. Then, each rotated



(a) (b)

Fig. 3. (a) Average mutual information with Gaussian inputs in a 4×4 *mismatched* channel following a separable model. Two data-streams are used in signaling and a limited feedback codebook designed along the principle elucidated in Sec. IV is used. (b) Error probability performance with the same codebook under QPSK inputs.

codeset is scaled by a shrinking factor $\alpha_i \triangleq \frac{\mu_i}{\mu_1}$. That is, we scale each rotated codeset in proportion to the generalized eigenvalue of that subspace. From each rotated codeset of N codewords, we retain $N_{\mathsf{loc}}^i, i = 1, \cdots, N_{\mathsf{stat}}$ codewords. The heuristic behind the choice of N_{loc}^i has been explained in the previous section. The same heuristic can be used to justify the choice of α_i as well.

B. Exploiting the General Structure of the Scaling and Rotation Maps

We now delve into why a general form of the maps in Appendix B is useful. In many practical systems, it is desirable for the precoder codebook to have more structure so as to facilitate implementation [3]. For example, two commonly desired properties are:

1) *Bounded Gain Power Amplifier Architecture* where we require

$$\max_{\mathbf{V}_i \in \mathcal{C}} \frac{\max_{mn} |\mathbf{V}_i(m,n)|}{\min_{mn} |\mathbf{V}_i(m,n)|} \leq \eta. \tag{36}$$

The above condition is useful in ensuring that the power amplifiers used in the RF link chain are not driven to their operational limits. The most general form of the rotation and scaling maps allows one to search for a codebook that satisfies the above property in addition to the localization and well-packing properties,

2) *Recursive/Nested Codebook Structure* where a codebook of rank-N_{small} can be generated from a codebook of rank-N_{large} (with $N_{\mathsf{large}} > N_{\mathsf{small}}$) by retaining only a subset of N_{small} columns from every precoder in the rank-N_{large} codebook. This property is desired so as to minimize the algorithmic complexity of generating a family of codebooks of different ranks *on the fly*. The low-complexity property of the proposed maps and the *offline* generation of the root codesets of different ranks ensure that this issue is redundant with our codebook design.

Thus, we strongly generalize the maps of [42], [43] and as a by-product observe that even in the $M = 1$ case, a rich family of maps can effect the scaling operation other than (28). Additional structure in the codebook can also be accommodated to ease implementation complexity.

VI. Numerical Results

We now illustrate via numerical studies the performance gains possible with our codebook construction and the consequent bridging of the gap between statistical and optimal precoding. In the first study, we consider a 4×4 channel under the separable model with $\mathbf{\Lambda}_t = \mathrm{diag}\left([14.98 \;\; 0.50 \;\; 0.26 \;\; 0.26]\right)$ and $\mathbf{\Lambda}_r = \mathrm{diag}\left([15.5 \;\; 0.25 \;\; 0.15 \;\; 0.10]\right)$. This choice ensures that the transmit/receive covariance matrices are both ill-conditioned and with $M = 2$, note that the channel is *not* matched to the precoder. We first generate a root codeset of $N = 4$ codewords with $\sin(\theta) \approx 0.76$ and $\gamma \approx 0.75$ via VQ. Let $\{\mathbf{u}_i, \; i = 1, \cdots, 4\}$ denote the column vectors of $\mathbf{U}_t$. The codebook used for $B = 1$ satisfies $N_{\mathsf{stat}} = 1$ with the codeword corresponding to $[\mathbf{u}_1 \; \mathbf{u}_2]$ and $N_{\mathsf{rvq}} = 1$ while with $B = 2$, the codebook has an additional RVQ codeword and a local codeword around $[\mathbf{u}_1 \; \mathbf{u}_2]$. Similarly, with $B = 4$, $N_{\mathsf{stat}} = 3, N_{\mathsf{loc}}^1 = N_{\mathsf{loc}}^2 = 3, N_{\mathsf{loc}}^3 = 2$ and $N_{\mathsf{rvq}} = 5$. The statistical codewords correspond to $[\mathbf{u}_1 \; \mathbf{u}_i]$, $i = 2, \cdots, 4$. Since we are mainly interested in illustrating the performance gains in the high-SNR regime, uniform power allocation is used for $\mathbf{\Lambda}_{\mathsf{stat}}$.

Fig. 3(a) shows the average mutual information with a Gaussian input for statistical and limited feedback precoding. In addition to the mutual information, raw bit error rate (BER) is useful as well. Fig 3(b) shows the improvement in error probability for the same channel with QPSK inputs. In the error probability case, the index of the codeword that minimizes the distance to the instantaneous $\widetilde{\mathbf{V}}_{\mathbf{H}}$ is fed back. Note that while the performance gap between the optimal and the statistical schemes is significantly bridged in the error

probability case, further improvement in mutual information is possible. Nevertheless, both the figures show that substantial gains are possible with a few bits of feedback. For example, with $B = 4$ bits of feedback, a 3 dB gain is possible at a rate of 10 bps/Hz while a 6 dB gain is possible at a BER of 10^{-3}. Also, note that an i.i.d. codebook design incurs a dramatic loss in performance in correlated channels.

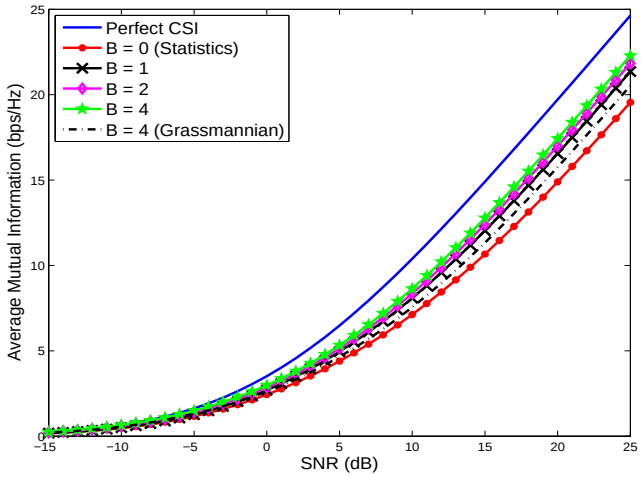


Fig. 4. Average mutual information with Gaussian inputs in a 4×4 *mismatched* channel with non-separable correlation and $M = 3$.

In the second study, we consider a 4×4 channel with non-separable correlation following the virtual representation framework. The variance matrix $\sigma(i,j) \triangleq E\left[|\mathbf{H}_{\mathsf{ind}}(i,j)|^2\right]$ used in the study is

$$\sigma = \begin{bmatrix} 1.24 & 1.42 & 7.49 & 0.23 \\ 0.41 & 0.14 & 0.42 & 0.03 \\ 0.72 & 1.39 & 0.07 & 0.02 \\ 0.28 & 0.13 & 0.50 & 1.51 \end{bmatrix}. \tag{37}$$

Note that the channel has a single dominant transmit (as well as receive) eigen-mode and is hence mismatched when $M = 3$ data-streams are used in signaling. The parameters of the root codeset are $N = 4$, $\sin(\theta) \approx 0.87$ and $\gamma \approx 0.84$. As before, let $\{\mathbf{u}_i,\ i = 1, \cdots, 4\}$ denote the column vectors of the DFT matrix $\mathbf{U}_t$. The codebook for $B = 1$ has the two statistical codewords $[\mathbf{u}_3\ \mathbf{u}_2\ \mathbf{u}_1]$ and $[\mathbf{u}_3\ \mathbf{u}_2\ \mathbf{u}_4]$. For $B = 2$, we use two additional RVQ codewords and for $B = 4$, we use $N_{\mathsf{stat}} = 3, N^1_{\mathsf{loc}} = N^2_{\mathsf{loc}} = 3, N^3_{\mathsf{loc}} = 2$ and $N_{\mathsf{rvq}} = 5$. The third statistical codeword when $B = 4$ is $[\mathbf{u}_3\ \mathbf{u}_1\ \mathbf{u}_4]$. Fig. 4 illustrates the bridging of the gap in mutual information between the optimal and the statistical schemes. It is important to note that both the channels studied here are so constructed to result in a substantial performance gap between perfect CSI and statistical signaling. Thus our studies illustrate that substantial gains can be achieved even with few bits of feedback.

VII. Concluding Remarks

In this work, we have studied linear precoding under a realistic system model. In particular, the focus is on the impact of spatial correlation when perfect CSI is available at the receiver, statistical information is available at both the ends, and quantized channel information is fed back from the receiver to the transmitter. While initial works on precoding assume perfect CSI at both the ends and hence do not impose any particular structure on the precoder matrices, under the model studied here, we see that structure can help in minimizing the reverse-link feedback as well as reduce the implementation complexity.

We further developed the notions of matched and mismatched channels, introduced in [30], in this work. The study of statistical precoding in [30] motivates the proposed limited feedback design where we quantize the space of semiunitary matrices with a non-uniform bias towards the statistically dominant eigen-modes. The design as well as its adaptability are rendered practical by the construction of mathematical maps (operations) that can be used to rotate and scale submanifolds on the Grassmann manifold. More importantly, numerical studies show that the proposed designs yield significant improvement in performance when the channel is mismatched to the communication scheme.

This work is a first attempt at systematic precoder codebook design in single-user multi-antenna channels that exploits spatial correlation and channel structure explicitly. Possible extensions are the study of more complex receiver architectures and performance analysis in the finite antenna, arbitrary SNR setting, along the lines of [30]. More work also needs to be done to understand the impact of spatial correlation on the performance of the proposed limited feedback scheme which could in turn drive the development of more efficient codebook constructions. Other open issues that need further study include practical aspects such as codebook designs for wideband channels, codebook designs based on Fourier/Hadamard matrices that are useful in achieving the bounded gain power amplifier architecture and hence, have found much interest in the standardization community, incorporating the cost of statistics acquisition in performance analysis [44], and more general scattering environment-independent channel decompositions [45] that mimic the physical model closely. The case of multi-user systems with feedback, where the impact of different users' channel structure is more critical in system performance, is another area for study.

We close the paper by drawing attention to the philosophy that has guided this work. While deducing the structure of the optimal signaling scheme under general assumptions on spatial correlation and channel information seems extremely difficult, an alternative approach that partitions this problem into smaller sub-problems could be quite fruitful. The general idea of matching the rank of the precoding scheme to the number of dominant transmit eigenvalues with the resolution necessary to decide whether an eigenvalue is "dominant" or not being a function of the SNR reminds one of the classical source-channel matching paradigm [46]. Initial evidence seen in this paper also suggests that this partitioning provides a natural framework to understand the performance of limited feedback schemes.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for suggesting the use of the geometrical (principal angle) interpretation of the projection 2-norm distance metric. This approach considerably simplified the exposition of this paper.

APPENDIX

A. Matrix-Algebraic Proofs of Distance Properties

Proof of Lemma 1: The projection 2-norm distance between $\mathbf{v}_1$ and $\mathbf{v}_2$ is defined as $d_{\mathrm{proj},2}(\mathbf{v}_1,\mathbf{v}_2) = \max_i |\lambda_i\left(\mathbf{v}_1\mathbf{v}_1^H - \mathbf{v}_2\mathbf{v}_2^H\right)|$. We can write the matrix within the above operation as $[\mathbf{v}_1\ \mathbf{v}_2]\,[\mathbf{v}_1\ \ -\mathbf{v}_2]^H$. Since the non-trivial eigenvalues of a matrix product $\mathbf{AB}$ are the same as those of $\mathbf{BA}$, we need the largest eigenvalue of

$$\mathbf{X} = \begin{bmatrix} \mathbf{v}_1^H \\ -\mathbf{v}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{v}_1^H\mathbf{v}_2 \\ -\mathbf{v}_2^H\mathbf{v}_1 & -1 \end{bmatrix}. \quad (38)$$

Expanding the characteristic equation of $\mathbf{X}$, $\det(\mathbf{X}-\lambda\mathbf{I}_2)=0$, we have $\max_i|\lambda_i|^2 = 1-|\mathbf{v}_1^H\mathbf{v}_2|^2$. Using the positive root for $\lambda_{\max}$, the lemma follows immediately. ∎

Proof of Prop. 1: The matrix-algebraic proof is as follows.
1) Note that

$$\begin{aligned} \lambda_i(\mathbf{V}_1\mathbf{V}_1^H - \mathbf{V}_2\mathbf{V}_2^H) &\le \lambda_i(\mathbf{V}_1\mathbf{V}_1^H) \le 1 \ \text{ and} \quad (39)\\ -\lambda_i(\mathbf{V}_1\mathbf{V}_1^H - \mathbf{V}_2\mathbf{V}_2^H) &= \lambda_{N_t-i+1}(\mathbf{V}_2\mathbf{V}_2^H - \mathbf{V}_1\mathbf{V}_1^H) \\ &\le \lambda_{N_t-i+1}(\mathbf{V}_2\mathbf{V}_2^H) \le 1. \quad (40)\end{aligned}$$

The claim in 1) follows from the above two inequalities.
2) We the need the following result [47] that helps in computing the determinant of partitioned matrices.

Lemma 4: If $\mathbf{X},\mathbf{Y},\mathbf{Z}$ and $\mathbf{W}$ are $n\times n$ matrices and $\mathbf{W}$ is invertible, we have

$$\det\begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \det(\mathbf{X}-\mathbf{Y}\mathbf{W}^{-1}\mathbf{Z})\cdot\det(\mathbf{W}). \quad (41)$$

∎

Using this fact and the trick (in the above version of the proof of Lemma 1) of rewriting the eigenvalues of $\mathbf{AB}$ in terms of eigenvalues of $\mathbf{BA}$, 2) follows trivially. ∎

B. Generalized Scaling Map

Let $\mathcal{R}(\mathbf{V}_1,N,\theta,\gamma)$ be a root codeset on $\mathcal{G}(N_t,M)$. Let $\mathbf{U_A}$ and $\mathbf{W}$ be arbitrary $M\times M$ unitary matrices and let $\mathbf{U_B}$ be an arbitrary $(N_t-M)\times(N_t-M)$ unitary matrix. Given $\alpha\in(0,1)$ and $M\le(N_t-M)$, for any $\mathbf{V}_i\in\mathcal{R}$, generate an $M\times M$ diagonal, positive-definite matrix $\mathbf{\Lambda}_i$ with

$$\begin{aligned} \mathbf{\Lambda}_{\min} &\triangleq \min_j \mathbf{\Lambda}_i(j) = 1-\alpha^2\left(1-\lambda_{\min}(\mathbf{V}_1^H\mathbf{V}_i\mathbf{V}_i^H\mathbf{V}_1)\right)\\ \mathbf{\Lambda}_{\max} &\triangleq \max_j \mathbf{\Lambda}_i(j) \le 1.\end{aligned}$$

Then, define $\mathbf{A}_i$ as $\mathbf{A}_i = \mathbf{U_A}\,\mathbf{\Lambda}_i^{1/2}\,\mathbf{W}^H$. Define the $M\times M$ principal component of the $(N_t-M)\times M$ diagonal matrix $\mathbf{\Lambda_B}$ as $(\mathbf{I}_M-\mathbf{\Lambda}_i)^{1/2}$ and $\mathbf{B}_i$ as $\mathbf{B}_i = \mathbf{U_B}\,\mathbf{\Lambda_B}^{1/2}\,\mathbf{W}^H$.

If $M>(N_t-M)$, for any $\mathbf{V}_i\in\mathcal{R}$, generate an $(N_t-M)\times(N_t-M)$ diagonal, positive-semidefinite matrix $\mathbf{\Gamma}_i$ with: $\mathbf{\Gamma}_{\max} \triangleq \max_j\mathbf{\Gamma}_i(j) = \alpha^2\left(1-\lambda_{\min}(\mathbf{V}_1^H\mathbf{V}_i\mathbf{V}_i^H\mathbf{V}_1)\right)$ and $\mathbf{\Gamma}_{\min}\triangleq\min_j\mathbf{\Gamma}_i(j)\ge 0$. Then, define $\mathbf{B}_i$ as $\mathbf{U_B}\,\mathbf{\Lambda_B}^{1/2}\,\mathbf{W}^H$ with the principal $(N_t-M)\times(N_t-M)$ component of $\mathbf{\Lambda_B}$ being $\mathbf{\Gamma}_i$. Define $\mathbf{A}_i$ as $\mathbf{A}_i = \mathbf{U_A}\,\mathbf{\Lambda_A}\,\mathbf{W}^H$ with the principal $(N_t-M)\times(N_t-M)$ component of $\mathbf{\Lambda_A}$ being $\mathbf{I}_{N_t-M}-\mathbf{\Gamma}_i$ and the principal southeast component being $\mathbf{I}_{2M-N_t}$.

Theorem 2: The map $s(\cdot)$ that leads to scaling of $\mathcal{R}(\mathbf{V}_1,N,\theta,\gamma)$ by α is given by

$$s(\mathbf{V}_i) = \mathbf{V}_1\,\mathbf{A}_i + \mathbf{V}_1^{\mathsf{null}}\,\mathbf{B}_i \quad (42)$$

where $\mathbf{V}_1^{\mathsf{null}}$ is a representative of the null-space corresponding to $\mathbf{V}_1$.

Proof: As in the beamforming case, we can decompose $s(\cdot)$ as $s(\cdot) = r_{\mathbf{U}_{\mathbf{V}_1}}\odot s_{\mathsf{vertex}}\odot r_{\mathbf{U}_{\mathbf{V}_1}^H}(\cdot)$ with

$$\mathbf{U}_{\mathbf{V}_1}^H s(\mathbf{V}_i) \;=\; s_{\mathsf{vertex}}(\mathbf{U}_{\mathbf{V}_1}^H\mathbf{V}_i) \quad (43)$$

where $r_{\mathbf{U}_{\mathbf{V}_1}^H}(\cdot)$ rotates $\mathcal{R}$ to the canonical precoder $[\mathbf{I}_M\ \ \mathbf{O}_{M\times(N_t-M)}]^T$ while $s_{\mathsf{vertex}}(\cdot)$ scales (shrinks) the canonical codeset by a factor α and $r_{\mathbf{U}_{\mathbf{V}_1}}$ rotates it back to the direction corresponding to $\mathbf{V}_1$.

Let $\mathbf{A}_i$ be an $M\times M$ full rank matrix and $\mathbf{B}_i$ be an $(N_t-M)\times M$ matrix such that

$$s_{\mathsf{vertex}}\left(\mathbf{U}_{\mathbf{V}_1}^H\mathbf{V}_i\right) \;=\; \begin{bmatrix} \mathbf{A}_i \\ \mathbf{B}_i \end{bmatrix} \quad (44)$$

and $s(\mathbf{V}_i)$ is as in (42). We now show that $\mathbf{A}_i$ and $\mathbf{B}_i$ have to be as in the statement of the theorem so that $s(\cdot)$ results in scaling by α. First, note that the semiunitarity of $s(\mathbf{V}_i)$ and the facts that $\mathbf{V}_1^H\mathbf{V}_1=\mathbf{I}_M$, $\mathbf{V}_1^H\mathbf{V}_1^{\mathsf{null}}=\mathbf{O}_{M\times(N_t-M)}$ and $\mathbf{V}_1^{\mathsf{null},H}\mathbf{V}_1^{\mathsf{null}}=\mathbf{I}_{N_t-M}$ imply that $\mathbf{A}_i^H\mathbf{A}_i+\mathbf{B}_i^H\mathbf{B}_i=\mathbf{I}_M$. Thus, we have

$$\begin{aligned} 1 &= \lambda_{\max}(\mathbf{A}_i^H\mathbf{A}_i+\mathbf{B}_i^H\mathbf{B}_i)\\ &\ge \lambda_{\max}(\mathbf{A}_i^H\mathbf{A}_i)+\lambda_{\min}(\mathbf{B}_i^H\mathbf{B}_i)\\ &\ge \max\left(\lambda_{\max}(\mathbf{A}_i^H\mathbf{A}_i),\lambda_{\max}(\mathbf{B}_i^H\mathbf{B}_i)\right). \quad (45)\end{aligned}$$

The map $s(\cdot)$ should satisfy
1) $s(\mathbf{V}_1)=\mathbf{V}_1$, and
2) $d_{\mathrm{proj},2}(s(\mathbf{V}_1),s(\mathbf{V}_i)) = \alpha d_{\mathrm{proj},2}(\mathbf{V}_1,\mathbf{V}_i)$ for all i.

First, let us consider the distance scaling property. We need

$$\begin{aligned} &\max_{m=1,\cdots,M}|\sin(\theta_m)|\\ \overset{(a)}{=}\ &\alpha d_{\mathrm{proj},2}(\mathbf{V}_1,\mathbf{V}_i) \overset{(b)}{=} d_{\mathrm{proj},2}(s(\mathbf{V}_1),s(\mathbf{V}_i))\\ \overset{(c)}{=}\ &d_{\mathrm{proj},2}\left(\mathbf{U}_{\mathbf{V}_1}^H s(\mathbf{V}_1),\mathbf{U}_{\mathbf{V}_1}^H s(\mathbf{V}_i)\right)\\ \overset{(d)}{=}\ &d_{\mathrm{proj},2}\left(s_{\mathsf{vertex}}(\mathbf{U}_{\mathbf{V}_1}^H\mathbf{V}_1),s_{\mathsf{vertex}}(\mathbf{U}_{\mathbf{V}_1}^H\mathbf{V}_i)\right)\\ \overset{(e)}{=}\ &d_{\mathrm{proj},2}\left(\begin{bmatrix}\mathbf{I}_M\\ \mathbf{0}_{N_t-M\times M}\end{bmatrix},\begin{bmatrix}\mathbf{A}_i\\ \mathbf{B}_i\end{bmatrix}\right)\\ \overset{(f)}{=}\ &\max_{m=1,\cdots,M}|\sin(\eta_m)| \quad (46)\end{aligned}$$

where $\{\theta_m\}$ and $\{\eta_m\}$ are the principal angles between $\mathbf{V}_1$ and $\mathbf{V}_i$, and $\begin{bmatrix}\mathbf{I}_M\\ \mathbf{0}_{N_t-M\times M}\end{bmatrix}$ and $\begin{bmatrix}\mathbf{A}_i\\ \mathbf{B}_i\end{bmatrix}$, respectively.

In the above series of equations, (a) and (f) follow from (15), (b) from the distance scaling property of $s(\cdot)$, (c) from Lemma 2, (d) from (43), and (e) from (44). Now using the property that $\{\cos(\theta_m)\}$ and $\{\cos(\eta_m)\}$ are the singular values

of $\mathbf{V}_1^H\mathbf{V}_i$ and $\mathbf{A}_i$, respectively, and the relation between sines and cosines, we have

$$\mathbf{\Lambda}_{\min} \triangleq \min_j \mathbf{\Lambda}_i(j) = 1 - \alpha^2\left(1 - \lambda_{\min}(\mathbf{V}_1^H\mathbf{V}_i\mathbf{V}_i^H\mathbf{V}_1)\right). \quad (47)$$

The constraints elucidated above are the only constraints to be imposed on the singular values of $\mathbf{A}_i$.

We now describe the complete decomposition as in the statement of the theorem. Assume a singular value decomposition for $\mathbf{A}_i$ and $\mathbf{B}_i$ of the form: $\mathbf{A}_i = \mathbf{U}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}^{1/2}\mathbf{W}_\mathbf{A}^H$ and $\mathbf{B}_i = \mathbf{U}_\mathbf{B}\mathbf{\Lambda}_\mathbf{B}^{1/2}\mathbf{W}_\mathbf{B}^H$, respectively where $\mathbf{U}_\mathbf{A}, \mathbf{W}_\mathbf{A}$ and $\mathbf{W}_\mathbf{B}$ are $M \times M$ unitary matrices, and $\mathbf{U}_\mathbf{B}$ is an $(N_t - M) \times (N_t - M)$ unitary matrix. The full-rankness of $\mathbf{A}_i$ means that the $M \times M$ diagonal matrix $\mathbf{\Lambda}_\mathbf{A}$ is positive definite while the $(N_t - M) \times M$ matrix $\mathbf{\Lambda}_\mathbf{B}$ has non-negative entries only along the leading diagonal. Since $\mathbf{A}_i^H\mathbf{A}_i + \mathbf{B}_i^H\mathbf{B}_i = \mathbf{I}_M$, we have $\mathbf{I}_M - \mathbf{\Lambda}_\mathbf{A} = \mathbf{W}_\mathbf{A}^H\mathbf{W}_\mathbf{B}(\mathbf{\Lambda}_\mathbf{B}^T\mathbf{\Lambda}_\mathbf{B})^{1/2}\mathbf{W}_\mathbf{B}^H\mathbf{W}_\mathbf{A}$. Comparing the two sides, we see that $\mathbf{W}_\mathbf{A} = \mathbf{W}_\mathbf{B}$ (we set both to be $\mathbf{W}$) and $\mathbf{I}_M - \mathbf{\Lambda}_\mathbf{A} = (\mathbf{\Lambda}_\mathbf{B}^T\mathbf{\Lambda}_\mathbf{B})^{1/2}$. Note that since there are no constraints on/relationship between $\mathbf{U}_\mathbf{A}$ and $\mathbf{U}_\mathbf{B}$, the leading diagonal entries of $\mathbf{\Lambda}_\mathbf{A}$ and $\mathbf{\Lambda}_\mathbf{B}$ can be in any order. This is because either unitary matrix can be appropriately adjusted by a permutation matrix.

If $M \leq (N_t - M)$, without loss in generality assume that the diagonal entries of $\mathbf{\Lambda}_\mathbf{A}$ are in non-increasing order while those of $\mathbf{\Lambda}_\mathbf{B}$ may be not. Given a choice of $\mathbf{\Lambda}_\mathbf{A}$, the condition $\mathbf{I}_M - \mathbf{\Lambda}_\mathbf{A} = (\mathbf{\Lambda}_\mathbf{B}^T\mathbf{\Lambda}_\mathbf{B})^{1/2}$ can be met by choosing the principal $M \times M$ component of $\mathbf{\Lambda}_\mathbf{B}$ to be $(\mathbf{I}_M - \mathbf{\Lambda}_\mathbf{A})^{1/2}$. If $M > (N_t - M)$, assume that the diagonal entries of $\mathbf{\Lambda}_\mathbf{B}$ are in non-increasing order while those of $\mathbf{\Lambda}_\mathbf{A}$ may be not. Then, the condition $\mathbf{I}_M - \mathbf{\Lambda}_\mathbf{A} = (\mathbf{\Lambda}_\mathbf{B}^T\mathbf{\Lambda}_\mathbf{B})^{1/2}$ can be met if $2M - N_t$ entries of $\mathbf{\Lambda}_\mathbf{A}$ are 1. The additional constraint on the smallest diagonal entry (see discussion above) ensures distance scaling.

To close the theorem, it is necessary to verify that $s(\mathbf{V}_1) = \mathbf{V}_1$. This can be done by checking that $\mathbf{\Lambda}_i$ can be computed in closed-form. For this, note that $\mathbf{\Lambda}_{\min} = 1$ and since $\mathbf{\Lambda}_{\max} \leq 1$, we have $\mathbf{\Lambda}_i = \mathbf{I}_M$. From here, it can be checked that $\mathbf{B}_i = \mathbf{O}_{(N_t-M)\times M}$ and from (42), we thus have $s(\mathbf{V}_1) = \mathbf{V}_1\,\mathbf{U}_\mathbf{A}\,\mathbf{W}^H$. On the Grassmann manifold $\mathcal{G}(N_t, M)$, multiplication by an $M \times M$ unitary matrix results in the same "point." Thus $s(\mathbf{V}_1) = \mathbf{V}_1$ and the proof is complete. ∎

Note that the choice of the scaling map is non-unique due to freedom in the choice of $\mathbf{U}_\mathbf{A}$, $\mathbf{U}_\mathbf{B}$ and $\mathbf{W}$ as well as the eigenvalues of $\mathbf{\Lambda}_i$ and $\mathbf{\Gamma}_i$. The case of $\mathbf{V}_i = \mathbf{V}_1$ is special where $\mathbf{\Lambda}_i$ turns out to be $\mathbf{I}_M$. With almost any other choice of $\mathbf{V}_i$, these matrices are non-identity, in general. Besides these choices, non-uniqueness of the representative of $\mathbf{V}_1^{\mathsf{null}}$ also leads to non-uniqueness of the map.

REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

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

[3] Third Generation Partnership Project, "TSG-RAN: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Overall Description, Stage 2, Release 8," *3GPP TS 36.300 v8.4.0*, Mar. 2008.

[4] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig, "What is the Value of Limited Feedback for MIMO Channels?," *IEEE Commun. Magaz.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[5] V. Raghavan, V. V. Veeravalli, and R. W. Heath, Jr., "Reduced Rank Signaling in Spatially Correlated MIMO Channels," *Proc. IEEE Intern. Symp. Inform. Theory*, July 2007.

[6] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Why Does the Kronecker Model Result in Misleading Capacity Estimates?," *Submitted, IEEE Trans. Inform. Theory*, Nov. 2007, Available: [Online]. `http://www.arxiv.org/abs/0808.0036`.

[7] 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.

[8] C.-N. Chuah, J. M. Kahn, and D. N. C. Tse, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

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

[10] A. M. Sayeed and V. V. Veeravalli, "Essential Degrees of Freedom in Space-Time Fading Channels," *Proc. 13th IEEE Intern. Symp. Personal Indoor and Mobile Radio Commun.*, vol. 4, pp. 1512–1516, Sept. 2002.

[11] K. H. Lee and D. P. Petersen, "Optimal Linear Coding for Vector Channels," *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1283–1290, Dec. 1976.

[12] J. Salz, "Digital Transmission over Cross-Coupled Linear Channels," *AT&T Tech. J.*, vol. 64, no. 6, pp. 1147–1159, July-Aug. 1985.

[13] J. Yang and S. Roy, "On Joint Transmitter and Receiver Optimization for Multiple-Input-Multiple-Output (MIMO) Transmission Systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[14] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant Filterbank Precoders and Equalizers Part I: Unification and Optimal Designs," *IEEE Trans. Sig. Proc.*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[15] H. Sampath, P. Stoica, and A. Paulraj, "Generalized Linear Precoder and Decoder Design for MIMO Channels Using the Weighted MMSE Criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[16] H. Sampath and A. Paulraj, "Linear Precoding for Space-Time Coded Systems with Known Fading Correlations," *IEEE Commun. Letters*, vol. 6, no. 6, pp. 239–241, June 2002.

[17] J. Yang and S. Roy, "Joint Transmitter-Receiver Optimization for Multi-input Multi-output Systems with Decision Feedback," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1334–1347, Sept. 1994.

[18] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal Designs for Space-Time Linear Precoders and Decoders," *IEEE Trans. Sig. Proc.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[19] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization," *IEEE Trans. Sig. Proc.*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[20] 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.

[21] 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.

[22] W. Santipach, Y. Sun, and M. L. Honig, "Benefits of Limited Feedback for Wireless Channels," *Proc. Annual Allerton Conf. Commun. Control and Computing*, Oct. 2003.

[23] J. C. Roh and B. D. Rao, "Multiple Antenna Channels with Partial Channel State Information at the Transmitter," *IEEE Trans. Wireless Commun*, vol. 3, no. 2, pp. 677–688, Mar. 2004.

[24] 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.

[25] 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.

[26] V. Raghavan, R. W. Heath, Jr., and A. M. Sayeed, "Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels," *IEEE Journ. Selected Areas in Commun.*, vol. 25, no. 7, pp. 1298–1310, Sept. 2007.

[27] D. J. Love and R. W. Heath, Jr., "Limited Feedback Unitary Precoding for Spatial Multiplexing," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[28] D. J. Love and R. W. Heath, Jr., "Multimode Precoding for MIMO Wireless Systems," *IEEE Trans. Sig. Proc.*, vol. 53, no. 10, pp. 3674–3687, Oct. 2005.

[29] T. Kim, M. Bengtsson, E. G. Larsson, and M. Skoglund, "Combining Long-Term and Low-Rate Short-Term Channel State Information over Correlated MIMO Channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2409–2414, July 2008.
[30] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Low-Complexity Structured Precoding for Spatially Correlated MIMO Channels," *Submitted, IEEE Trans. Inform. Theory*, May 2008, Available: [Online]. `http://www.arxiv.org/abs/0805.4425`.
[31] E. Bonek, "Experimental Validation of Analytical MIMO Channel Models," *Elektrotechnik und Informationstechnik*, vol. 122, no. 6, pp. 196–205, 2005.
[32] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO Rayleigh Fading Channels: Capacity, Optimal Signaling and Asymptotics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2058–2072, June 2005.
[33] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second Order Statistics of NLOS indoor MIMO Channels based on 5.2 GHz Measurements," *Proc. IEEE Global Telecommun. Conf.*, vol. 1, pp. 25–29, Nov. 2001.
[34] 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.
[35] 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," *Proc. 58th IEEE Fall Vehicular Technology Conference, 2003*, vol. 1, pp. 363–367, Oct. 2003.
[36] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental Study of MIMO Channel Statistics and Capacity via the Virtual Channel Representation," Tech. Rep., University of Wisconsin-Madison, May 2006, Available: [Online]. `http://dune.ece.wic.edu`.
[37] 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, Sept. 2002.
[38] A. Edelman, T. Arias, and S. Smith, "The Geometry of Algorithms with Orthogonality Constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Apr. 1999.
[39] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing Lines, Planes etc., Packings in Grassmannian Spaces," *Experimental Math.*, vol. 5, no. 2, pp. 139–159, 1996.
[40] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd edition, 1996.
[41] V. Raghavan, A. M. Sayeed, and N. Boston, "Near-Optimal Codebook Constructions for Limited Feedback Beamforming in Correlated MIMO Channels with Few Antennas," *Proc. IEEE Intern. Symp. Inform. Theory*, pp. 2622–2626, July 2006.
[42] 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, Nov. 2005.
[43] I. Kammoun, A. M. Cipriano, and J.-C. Belfiore, "Non-coherent Codes over the Grassmannian," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3657–3667, Oct. 2007.
[44] V. Raghavan, A. S. Y. Poon, and V. V. Veeravalli, "Non-Robustness of Statistics Based Beamformer Design in Correlated MIMO Channels," *Proc. IEEE Intern. Conf. Acoustics, Speech and Sig. Proc.*, 2008.
[45] V. Raghavan, A. S. Y. Poon, and V. V. Veeravalli, "MIMO Systems with Arbitrary Antenna Array Architectures: Channel Modeling, Capacity and Low-Complexity Signaling," *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, Nov. 2007.
[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, 1991.
[47] J. R. Silvester, "Determinants of Block Matrices," *The Mathematical Gazette*, vol. 84, no. 501, pp. 460–467, Nov. 2000.

Vasanthan Raghavan (S'01–M'06) received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology at Madras, 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 at Urbana-Champaign, Urbana, IL. His research interests span multi-antenna communication techniques, information theory, multihop networking, robust control, and random matrix theory.

Venugopal V. Veeravalli (S'86–M'92–SM'98–F'06) received the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the B.Tech degree in 1985 from the Indian Institute of Technology, Bombay, (Silver Medal Honors), all in Electrical Engineering. He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor in the department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science Laboratory. He served as a program director for communications research at the U.S. National Science Foundation in Arlington, VA from 2003 to 2005. He has previously held academic positions at Harvard University, Rice University, and Cornell University.

His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, and information theory. He is a Fellow of the IEEE and was on the Board of Governors of the IEEE Information Theory Society from 2004 to 2007. He was an Associate Editor for Detection and Estimation for the IEEE Transactions on Information Theory from 2000 to 2003, and an associate editor for the IEEE Transactions on Wireless Communications from 1999 to 2000. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award, the National Science Foundation CAREER Award, and the Presidential Early Career Award for Scientists and Engineers (PECASE).

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.