Design and Analysis of MIMO Spatial Multiplexing Systems With Quantized Feedback

June Chul Roh, *Member, IEEE*, and Bhaskar D. Rao, *Fellow, IEEE*

***Abstract*—This paper investigates the problem of transmit beamforming in multiple-antenna spatial multiplexing (SM) systems employing a finite-rate feedback channel. Assuming a fixed number of spatial channels and equal power allocation, we propose a new criterion for designing the codebook of beamforming matrices that is based on minimizing an approximation to the capacity loss resulting from the limited rate in the feedback channel. Using the criterion, we develop an iterative design algorithm that converges to a locally optimum codebook. Under the independent identically distributed channel and high signal-to-noise ratio (SNR) assumption, the effect on channel capacity of the finite-bit representation of the beamforming matrix is analyzed. Central to this analysis is the complex multivariate beta distribution and tractable approximations to the Voronoi regions associated with the code points. Furthermore, to compensate for the degradation due to the equal power allocation assumption, we propose a multimode SM transmission strategy wherein the number of data streams is determined based on the average SNR. This approach is shown to allow for effective utilization of the feedback bits resulting in a practical and efficient multiple-input multiple-output system design.**

***Index Terms*—Channel capacity, channel information feedback, matrix quantization, multiple antennas, multiple-input multiple-output (MIMO) systems, spatial multiplexing, transmit beamforming.**

I. Introduction

RECENTLY, multiple-antenna communication systems have received much attention because of the potential improvements in data transmission rates and/or link reliability. The performance of a multiple-antenna channel depends on the nature of channel information available at the transmitter and at the receiver. When the transmitter has perfect knowledge of channel, a higher capacity link can be achieved in the single user case, and there are other benefits such as lower complexity receivers and better system throughput in a multiuser environment. However, the assumption that the transmitter has perfect knowledge of the multidimensional channel is unrealistic, as in many practical systems the channel information is provided to the transmitter through a finite-rate feedback channel. In a general multiple-input multiple-output (MIMO) flat-fading setting, the feedback information is a beamforming (or precoding) matrix, usually an orthonormal column matrix, and the power allocation along the beams. There have been several studies recently dealing with how to feed back the channel information. Some researchers have worked on feedback of channel information in vector forms, for example, for multiple-input single-output (MISO) channels [1]–[3] and for the principal eigenmode of MIMO channels [4]. Only recently has feedback of channel information in matrix forms for MIMO channels begun to be addressed [5]–[11].

In this paper, we investigate the problem of transmit beamforming in multiple-antenna spatial multiplexing systems with finite-rate feedback channel. It is well known that, with perfect channel state information at the transmitter (CSIT), the first n principal eigenmodes are the optimum beamforming vectors for a MIMO spatial multiplexing system for n data streams. However, in systems with finite-rate feedback, perfect CSIT is not possible. The receiver usually selects, based on the channel observation, the best beamforming matrix from a codebook (a set of a finite number of beamforming matrices), which is designed in advance and shared with the transmitter.

For designing the codebook for the beamforming matrix, in [12] the orthonormal matrix to be fed back was reparameterized leading to quantizing of a set of independent vectors with decreasing dimensionality. The MISO method was sequentially applied and shown to be reasonably effective. A drawback of the scheme is the loss that arises due to the sequential nature of the quantization when compared to the more optimal joint quantization approach, and the difficult bit assignment task that naturally arises with sequential quantization. To avoid these drawbacks, a good design criterion is joint quantization of the beamforming directions and power assignments by maximizing the average mutual information as studied in [9] and [13]. However, such a formulation leads to a difficult quantizer design problem. Using average mutual information as a measure directly does not lead to an iterative design algorithm with monotonic convergence property.[1] This necessitates approximations in the design algorithm that make it hard to guarantee the optimality of the resulting codebook (details will be discussed in Section III). This difficulty motivates consideration and development of appropriate design criterion that leads to effective quantizer design. In this paper, first a tractable measure of capacity loss due to finite feedback rate is derived in the context of a fixed number of spatial channels with equal power allocation, leading to a new quantizer design criterion. Using the proposed criterion, an iterative design algorithm is developed with guaranteed monotonic convergence. The design algorithm is similar in nature to the

Manuscript received December 10, 2004; revised August 28, 2005. This work was supported by the University of California under Discovery Grant core02-10109 sponsored by Ericsson. The work in this paper was presented in part at the IEEE International Conference on Communications (ICC) 2005, Seoul, Korea, May 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Helmut Boelcskei.

J. C. Roh was with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA. He is now with Texas Instruments, Inc., Sunnyvale, CA 94089 USA (e-mail: jroh@ti.com).

B. D. Rao is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail:brao@ece.ucsd.edu).

Digital Object Identifier 10.1109/TSP.2006.875392

[1]Monotonic convergence means that an improved design is guaranteed at every iteration.

Lloyd algorithm in vector quantization (VQ) study [14], but it is now for matrix quantization and with a new design objective. Using the design method, a locally optimum beamforming codebook can be designed for an arbitrary set of system parameters including the number of transmit antennas, data streams, and feedback bits. The efficacy of the approach is demonstrated by using it to design quantizers in a variety of MIMO contexts and evaluating their performance through computer simulations.

Next, the challenging task of analyzing the performance of MIMO systems employing such finite-rate feedback techniques is undertaken, and interesting analytical results are provided. Several researchers have considered the performance analysis of MISO systems with quantized beamforming [2], [3]. The methodology employed in our MISO-related work [3] is extended to MIMO systems in this paper. In the analysis, we exploit results from multivariate statistics. In particular, the complex matrix-variate beta distribution is found to be relevant and useful in the capacity loss analysis. Also another interesting aspect of the analysis is the approximation made of the Voronoi region associated with each code point for analytical tractability. Recently, Love and Heath [4] considered feedback of the principal eigenmode of MIMO channel and derived bounds on the codebook size required to achieve some performance target (diversity order, capacity, and SNR loss). Santipach and Honig [7] also presented a performance analysis by considering random vector quantization (thereby, the codebook design problem was circumvented). Their analysis is asymptotic in the number of feedback bits and the number of antennas with the ratio of the two being fixed. Our approach is quite different in flavor and provides a direct way to characterize the performance for a given antenna configuration and feedback bits.

To compensate for the degradation due to the equal power allocation assumption, we also propose a multimode spatial multiplexing (SM) transmission scheme. The key features of the transmission scheme are that the number of data streams n is determined based on the average SNR, and in each mode, simple equal power allocation over n spatial channels is employed. One of the motivations for the transmission strategy is that all the feedback bits are utilized for representing only effective beamforming vectors without being concerned about the power allocation over the spatial channels. For example, in low SNR region, since only the principal eigenvector is used most of the time, there is little reason to waste the limited number of bits in representing the other eigenvectors. This approach calls for generalization of the codebook design methodology, and the overall approach is shown to allow for effective utilization of the feedback bits. Another form of multimode SM transmission scheme was developed in [15]. The difference between the two schemes will be discussed later in Section V.

This paper is organized as follows. Section II describes the system model and assumptions. The design of the codebook for the beamforming matrix is presented in Section III. In Section IV, we analyze the effect of finite-bit representation of beamforming matrix on the channel capacity. In Section V, we develop a variant of the multimode spatial multiplexing transmission scheme.

We use the following notations. $A^\dagger$ and A^T indicate the conjugate transpose and the transpose of matrix A, respectively. I_n is the $n \times n$ identity matrix and $\mathrm{diag}(a_1, \ldots, a_n)$ is a square diagonal matrix with $a_1, \ldots, a_n$ along the diagonal. $\mathrm{tr}A$ indicates the trace of matrix A, and $\|A\|_F$ denotes the Frobenius norm of matrix A defined as $\|A\|_F = (\sum_{i,j} |a_{ij}|^2)^{1/2}$. An $m \times n$ $(m \geq n)$ matrix A with orthonormal columns, i.e., $A^\dagger A = I_n$, will be called *orthonormal column matrix*. $A > 0$ means that matrix A is positive definite, and $C \otimes D$ represents the Kronecker product of C and D. $\tilde{\mathcal{N}}_r(\mu, \Sigma)$ is the r-dimensional proper complex Gaussian random vector with mean μ and covariance Σ. Uniform distribution over a set S is denoted by $\mathcal{U}(S)$. The function $\log(\cdot)$ is the natural logarithm unless otherwise specified. Thus, mutual information and capacity are in "*nats per channel use*."

II. System Model

A multiple antenna channel with t transmit and r receive antennas that is denoted as (t, r) MIMO, assuming flat fading in each antenna pair, is modeled by the channel matrix $H \in \mathbb{C}^{r \times t}$. That is, the channel input $x \in \mathbb{C}^t$ and the channel output $y \in \mathbb{C}^r$ have the following relationship:

$$y = Hx + \eta \tag{1}$$

where η is the additive white Gaussian noise distributed by $\tilde{\mathcal{N}}_r(0, I)$. For this paper, we assume[2] $t \geq r$ and the rank of H is denoted by m. The singular value decomposition (SVD) of H is given by $H = U_H \Sigma_H V_H^\dagger$, where $U_H \in \mathbb{C}^{r \times m}$ and $V_H \in \mathbb{C}^{t \times m}$ are orthonormal column matrices and $\Sigma_H \in \mathbb{R}^{m \times m}$ contains the singular values $\sigma_1 \geq \ldots \geq \sigma_m > 0$ of H. The average transmit power is denoted by P_T, i.e., $E[x^\dagger x] = P_T$.

When perfect channel information is known at the receiver and the transmitter, a water-filling based SM strategy is known to be optimal. This means that information about V_H and the power allocation on the different channels/beams represented by the columns of V_H has to be fed back to the transmitter. With only a finite number of bits available for channel information feedback, we concentrate first on representing the first n $(1 \leq n \leq m)$ column vectors of V_H. For notational convenience, let us denote the ith column vector of V_H by v_i and define the first n column vectors of V_H as V, i.e., $V = [v_1, \ldots, v_n]$. We also define $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, an $n \times n$ submatrix of Σ_H.

To formalize the problem, we assume that the MIMO system has a feedback channel with a finite rate of B bits per channel update. A codebook $\mathcal{C}$ containing N, where $N = 2^B$, candidate beamforming matrices, i.e., $\mathcal{C} = \{\hat{V}_1, \ldots, \hat{V}_N\}$ with $\hat{V}_i$ being $t \times n$ orthonormal column matrices, is assumed known to both the receiver and the transmitter. The design of the codebook is a topic of this paper. The receiver selects the optimum beamforming matrix $\hat{V} = \mathcal{Q}(H)$ from the codebook $\mathcal{C}$ based on the current channel H and transmits the index of $\hat{V}$ in $\mathcal{C}$ to the transmitter through the feedback channel. In practical systems, the error in the feedback channel and the delay resulting from finite-rate channel update impact the overall system performance. However, this paper assumes feedback with no error and no delay, and focuses solely on the effect of finite-bit representation of the channel information.

The channel information $\hat{V} = \mathcal{Q}(H)$ is employed as the beamforming matrix at the transmitter. That is, an information-

[2]The case $t < r$ can also be handled in a similar manner.

bearing symbol vector $s = [s_1, \ldots, s_n]^T$ is transmitted as $x = \hat{V}s$, resulting in the received signal

$$y = H\hat{V}s + \eta. \quad (2)$$

Here we assume that $s \sim \tilde{\mathcal{N}}_n(0, \Phi)$ and $\Phi = P_T \cdot \mathrm{diag}(\varphi_1, \ldots, \varphi_n)$ with $\varphi_i \geq 0$ and $\sum_i \varphi_i = 1$. The vector $\varphi = [\varphi_1, \ldots, \varphi_n]$ is the *power allocation information*. With perfect knowledge of channel at the receiver, the optimum power allocation for the equivalent channel $H\hat{V}$ can be calculated at the receiver [16]. We can also consider quantizing and feeding back the power allocation information. The mutual information between s and y for a given channel H, when the transmitter uses the channel information $(\hat{V}, \hat{\varphi})$ for transmission, is given by

$$I(H, \hat{V}, \hat{\varphi}) = \log\det\left[I + \Sigma_H^2 V_H^\dagger \hat{V}\Phi(\hat{\varphi})\hat{V}^\dagger V_H\right] \quad (3)$$

where $\Phi(\hat{\varphi}) := P_T \cdot \mathrm{diag}(\hat{\varphi})$, the power allocation associated with $\hat{\varphi}$.

Instead of feeding back the power allocation information, which reduces the bit budget for the beamforming matrix, one can employ a simple equal power allocation strategy, i.e., $\varphi_i = 1/n\ \forall i$. For a given n, with equal power allocation, the mutual information is given by[3]

$$I(H, \hat{V}) = \log\det\left(I + \rho\Sigma_H^2 V_H^\dagger \hat{V}\hat{V}^\dagger V_H\right) \quad (4)$$

where $\rho := P_T/n$. This paper mainly focuses on equal power allocation because it is more amenable to optimum codebook design, and results in minor performance degradation when coupled with an efficient multimode SM transmission wherein n is chosen based on the average SNR. The multimode SM scheme is discussed more fully in Section V.

III. Codebook Design for Beamforming Matrix

In this section, we develop a general matrix quantization (MQ) based design method for constructing the codebook of the beamforming matrices. We consider the case of $n = m$ in this section, and $n < m$ is discussed in Section V. For designing a codebook, first a suitable criterion is needed, which is developed next.

A. Capacity Loss Due to Finite Rate Feedback

The mutual information (4) can be written as

$$I(H, \hat{V}) = \log\det\left(I + \rho\Sigma_H^2\right) + \log\det\left[I - \left(I + \rho\Sigma_H^2\right)^{-1}\rho\Sigma_H^2\left(I - V_H^\dagger\hat{V}\hat{V}^\dagger V_H\right)\right] \quad (5)$$

since $I + \rho\Sigma_H^2 V_H^\dagger \hat{V}\hat{V}^\dagger V_H = (I + \rho\Sigma_H^2) - \rho\Sigma_H^2(I - V_H^\dagger\hat{V}\hat{V}^\dagger V_H) = (I + \rho\Sigma_H^2)[I - (I + \rho\Sigma_H^2)^{-1}\rho\Sigma_H^2(I - V_H^\dagger\hat{V}\hat{V}^\dagger V_H)]$. Here we notice that when $n = m$, the first term is $I(H, V)$, the mutual information with the perfect beamforming matrix V at the transmitter, and the second term accounts for the loss due to finite-bit representation of V. Let us define the *capacity loss* as the difference between the ergodic capacities associated with V and $\hat{V}$, that is

$$\begin{aligned} C_L(H, \hat{V}) &= E\left[I(H, V)\right] - E\left[I(H, \hat{V})\right] \\ &= E\left[I_L(H, \hat{V})\right] \end{aligned} \quad (6)$$

where $I_L(H, \hat{V}) := I(H, V) - I(H, \hat{V})$, which is from (5)

$$I_L(H, \hat{V}) = -\log\det\left[I - \left(I + \rho\Sigma_H^2\right)^{-1} \cdot\rho\Sigma_H^2\left(I - V_H^\dagger\hat{V}\hat{V}^\dagger V_H\right)\right]. \quad (7)$$

Because of the difficulty associated with directly working with (7) for the codebook design and analysis, the following two approximations to $I_L(H, \hat{V})$ are developed.

i) High-Resolution Approximation[4]: When $V_H^\dagger\hat{V}\hat{V}^\dagger V_H$ is close to I (which is valid when the number of feedback bits B is reasonably large) *or* when $P_T \ll 1$ (low SNR), we use the approximation $-\log\det(I - A) \simeq -\log(1 - \mathrm{tr}A) \simeq \mathrm{tr}A$ when matrix A is small. More specifically, when the eigenvalues $\lambda_i(A) \ll 1\ \forall i$, we have

$$I_L(H, \hat{V}) \simeq \mathrm{tr}\left[\left(I + \rho\Sigma_H^2\right)^{-1}\rho\Sigma_H^2\left(I - V_H^\dagger\hat{V}\hat{V}^\dagger V_H\right)\right]. \quad (8)$$

ii) High-SNR Approximation: When $P_T \gg 1$ (high SNR), since $(I + \rho\Sigma_H^2)^{-1}\rho\Sigma_H^2 \simeq I$

$$I_L(H, \hat{V}) \simeq -\log\det\left(V_H^\dagger\hat{V}\hat{V}^\dagger V_H\right). \quad (9)$$

The first approximation (8) will be used for codebook design next and the second approximation (9) for performance analysis in Section IV.

B. Codebook Design Criterion and Design Algorithm

For designing the beamforming-matrix codebook, a natural design criterion is maximizing the expected mutual information $E[I(H, \hat{V})]$ or, equivalently, minimizing the capacity loss defined in (6). However, using it directly does not lead to an iterative design algorithm with monotonic convergence property. Lau *et al.* [9] consider *covariance-matrix* quantization, i.e., joint quantization of the beamforming directions and power assignments by maximizing the average mutual information. The difficulty with the approach is that generally there is no analytical expression for the optimum code matrix as a function of a given partition region[5] in the channel space. This necessitates approximation to the solution to the optimization problem as in [9]. Thus, the resulting iterative design algorithm does have a convergence problem, which in turn makes it hard to guarantee the optimality of the resulting codebook.

Instead of using the direct form of the capacity expression, we utilize the approximation to the capacity loss: from (8), when B is reasonably large *or* when $P_T \ll 1$

$$C_L \simeq E\left[\mathrm{tr}\tilde{\Sigma}_H^2 - \mathrm{tr}\left(\tilde{\Sigma}_H^2 V_H^\dagger\hat{V}\hat{V}^\dagger V_H\right)\right] \quad (10)$$

[3]When the equal power allocation is employed, we will drop the third argument in (3) and use a notation $I(H, W)$ for the mutual information for channel H with beamforming matrix W.

[4]The underlying assumption in this section is that the quantizer is of high resolution (B is reasonably large), therefore, the approximation (8) is *valid in all range of SNR*, even in high SNR region.

[5]This is called the centroid condition in quantizer design.

where $\tilde{\Sigma}_H^2 := (I+\rho\Sigma_H^2)^{-1}\rho\Sigma_H^2$. The second term inside the bracket in (10) can be written as $\|(V_H\tilde{\Sigma}_H)^\dagger \hat{V}\|_F^2$. Therefore, minimizing the expectation of (10) is equivalent to the following codebook design criterion.

New Design Criterion: Design a mapping $\mathcal{Q}$ (mathematically, $\mathcal{Q}: \mathbb{C}^{r\times t} \to \mathcal{C}$) such that

$$\max_{\mathcal{Q}(\cdot)} E\left\|(V_H\tilde{\Sigma}_H)^\dagger \mathcal{Q}(H)\right\|_F^2 \tag{11}$$

where $\hat{V} = \mathcal{Q}(H)$ is the quantized beamforming matrix ($\hat{V}^\dagger\hat{V} = I_n$).

This design criterion will be called the *generalized mean squared weighted inner product (MSwIP) criterion* since it can be viewed as a generalization for MIMO channels of the MSwIP criterion that was developed for beamforming codebook design (for MISO systems) in [3]. In MISO system (when $r=1$), the design criterion (11) reduces to

$$\max_{\mathcal{Q}(\cdot)} E\left|(\tilde{\sigma}v)^\dagger \mathcal{Q}(h)\right|^2 \tag{12}$$

where $H = h^\dagger$ ($h \in \mathbb{C}^{t\times 1}$) with $\sigma = \|h\|$, $v = h/\|h\|$, $\tilde{\sigma} = [\sigma^2 P_T/(1+\sigma^2 P_T)]^{1/2}$, and $\hat{v} = \mathcal{Q}(h)$ is the quantized beamforming vector ($\|\hat{v}\| = 1$). For details refer to [3].

One of the virtues of the new design criterion is that it *does* lead to an iterative design algorithm with guaranteed monotonic convergence. The design algorithm is essentially similar to the Lloyd algorithm in vector quantization (VQ) study, which is based on two necessary conditions for optimality: the nearest neighborhood condition (NNC) and the centroid condition (CC) [14], [17]. The same approach is used here for designing the codebook of beamforming matrices.

Design Algorithm:

1) NNC: For given code matrices $\{\hat{V}_i; i=1,\ldots,N\}$, the optimum partition cells satisfy

$$\mathcal{H}_i = \left\{H \in \mathbb{C}^{r\times t} : \left\|(V_H\tilde{\Sigma}_H)^\dagger \hat{V}_i\right\|_F^2 \geq \left\|(V_H\tilde{\Sigma}_H)^\dagger \hat{V}_j\right\|_F^2, \quad \forall j \neq i\right\} \tag{13}$$

for $i=1,\ldots,N$, where $\mathcal{H}_i$ is the partition cell of the channel matrix space $\mathbb{C}^{r\times t}$ for the ith code matrix $\hat{V}_i$.

2) CC: For a given partition $\{\mathcal{H}_i; i=1,\ldots,N\}$, the optimum code matrices satisfy

$$\hat{V}_i = \arg\max_{\hat{V}(t\times n):\hat{V}^\dagger\hat{V}=I_n} E\left[\left\|(V_H\tilde{\Sigma}_H)^\dagger \hat{V}\right\|_F^2 | H \in \mathcal{H}_i\right] \tag{14}$$

for $i=1,\ldots,N$. Fortunately, the above optimization problem has a closed-form solution as

$$\hat{V}_i = (n \text{ principal eigenvectors) of } E\left[V_H\tilde{\Sigma}_H^2 V_H^\dagger | H \in \mathcal{H}_i\right]. \tag{15}$$

Proof (Proof of the CC Solution): The following lemma is essential in this proof.

Lemma 1 [18, p. 191]: Let $A \in \mathbb{C}^{t\times t}$ be Hermitian and let n be a given integer with $1 \leq n \leq t$. Then

$$\max_{U(t\times n):U^\dagger U=I_n} \mathrm{tr}(U^\dagger AU) = \sum_{i=1}^{n}\lambda_i(A) \tag{16}$$

where $\lambda_i(A)$ is the ith largest eigenvalue of A and the maximum is achieved when the columns of U are chosen to be orthonormal eigenvectors corresponding to the n largest eigenvalues of A.

The Frobenius norm in (14) can be expressed as $\|(V_H\tilde{\Sigma}_H)^\dagger\hat{V}\|_F^2 = \mathrm{tr}(\hat{V}^\dagger V_H\tilde{\Sigma}_H^2 V_H^\dagger \hat{V})$. By noting that under the condition $H \in \mathcal{H}_i$, $\hat{V}$ is a constant matrix, (14) can be rewritten as

$$\hat{V}_i = \arg\max_{\hat{V}(t\times n):\hat{V}^\dagger\hat{V}=I_n} \mathrm{tr}\left\{\hat{V}^\dagger E\left[V_H\tilde{\Sigma}_H^2 V_H^\dagger | H \in \mathcal{H}_i\right]\hat{V}\right\}.$$

By directly applying Lemma 1, we arrive at the desired solution given in (15). □

The above two conditions are iterated until the design objective $E\|(V_H\tilde{\Sigma}_H)^\dagger\mathcal{Q}(H)\|_F^2$ converges. In practice, the codebook is designed off-line using a sufficiently large number of training samples (channel realizations). In that case, the statistical correlation matrix $E[V_H\tilde{\Sigma}_H^2 V_H^\dagger | H \in \mathcal{H}_i]$ in (15) is estimated with a sample average.

Beamforming Matrix Selection (Encoding): For a given codebook $\mathcal{C} = \{\hat{V}_1,\ldots,\hat{V}_N\}$, the receiver selects the optimum beamforming matrix from the codebook based on the observed channel H so that the mutual information is maximized, i.e., $\hat{V} = \arg\max_{\hat{V}_i\in\mathcal{C}} I(H,\hat{V}_i)$. This is also considered in [8]. It is useful later in performance analysis to note that this encoding scheme is equivalent to

$$\hat{V} = \mathcal{Q}(H) = \arg\min_{\hat{V}_i\in\mathcal{C}} I_L(H,\hat{V}_i) \tag{17}$$

where $I_L(H,\hat{V}_i)$ is given in (7). By the encoding scheme, the channel matrix space $\mathbb{C}^{r\times t}$, where random channel H lies, is partitioned into $\{\mathcal{R}_i; i=1,\ldots,N\}$, where

$$\mathcal{R}_i := \left\{H \in \mathbb{C}^{r\times t} : I_L(H,\hat{V}_i) \leq I_L(H,\hat{V}_j), \forall j \neq i\right\}. \tag{18}$$

The encoding scheme (17) can be restated simply as $\mathcal{Q}(H) = \hat{V}_i$ if $H \in \mathcal{R}_i$.

C. Two Related Design Methods

The codebook design methodology developed above is based on the generalized MSwIP criterion, a meaningful criterion, and is quite effective as will be found through experimental evaluations. However, a drawback of the generalized MSwIP design method is that the codebook is optimized for a particular SNR (or P_T). That is, for a given channel H, the orthonormal column matrix V multiplied by a diagonal matrix $\tilde{\Sigma}_H$ (each eigenvector v_i is weighted by $\tilde{\sigma}_i = [\rho\sigma_i^2/(1+\rho\sigma_i^2)]^{1/2}$, which is dependent on P_T) is used as a training sample matrix. As a result, we may need more than one codebook if the system has multiple operating SNR points. Therefore, it would be interesting to find other design methods that do not depend on P_T.

1) Low SNR Optimization: When $\rho \ll 1$ (in low SNR region) $\tilde{\sigma}_i \simeq \sqrt{\rho}\sigma_i$; hence $\tilde{\Sigma}_H \simeq \sqrt{\rho}\,\Sigma_H$. Then, the original criterion (11) becomes

$$\max_{\mathcal{Q}(\cdot)} E \left\| (V_H \Sigma_H)^\dagger \mathcal{Q}(H) \right\|_F^2 . \tag{19}$$

It is interesting to see that the above design criterion is equivalent to maximizing the mean squared channel norm (MSCN), that is

$$\max_{\mathcal{Q}(\cdot)} E \left\| H \mathcal{Q}(H) \right\|_F^2. \tag{20}$$

This can be seen by noticing that from (2), the norm of the composite channel $H\hat{V}$ can be written as $\|H\hat{V}\|_F^2 = \mathrm{tr}(H\hat{V}\hat{V}^\dagger H^\dagger) = \|\Sigma_H V_H^\dagger \hat{V}\|_F^2$, a consequence of plugging in the SVD of H and using the property $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. This design criterion (20) will be referred as the *MSCN criterion.*

The MSCN criterion is a reasonable choice by itself, because naturally we can benefit by maximizing the gain of the composite channel $(H\hat{V})$. The design algorithm can be obtained by replacing $V_H \tilde{\Sigma}_H$ in the NNC and the CC in Section III-B with $V_H \Sigma_H$ or H. In particular, the conditional correlation matrix in (15) is replaced with $E[V_H \Sigma_H^2 V_H^\dagger | H \in \mathcal{H}_i]$ or $E[H^\dagger H | H \in \mathcal{H}_i]$.

2) High SNR Optimization: Another simplification of the generalized MSwIP criterion can be found by considering the other extreme case. As the SNR increases $(\rho \to \infty)$, $\tilde{\Sigma}_H \to I$; hence, the original criterion (11) reduces to

$$\max_{\mathcal{Q}(\cdot)} E \left\| V_H^\dagger \mathcal{Q}(V_H) \right\|_F^2 \tag{21}$$

where $\hat{V} = \mathcal{Q}(V_H)$ is the quantized beamforming matrix $(\hat{V}^\dagger \hat{V} = I_n)$. One can see that this criterion is based only on V_H and independent of Σ_H and ρ. A design algorithm corresponding to this criterion is obtained by setting $\tilde{\Sigma}_H = I$ in the NNC and the CC in Section III-B. In particular, the conditional correlation matrix in (15) is replaced with $E[V_H V_H^\dagger | H \in \mathcal{H}_i]$.

Interestingly, the above design criterion (21) can be intuitively explained using a distance between two subspaces. The columns of V_H form an orthogonal basis for an m-dimensional subspace, and $\hat{V}_H = \mathcal{Q}(V_H)$ for another m-dimensional subspace. The principal angles $\theta_1, \ldots, \theta_m$ between the two subspaces (for definition, see [19, p. 603]) can be represented with the singular values of $V_H^\dagger \hat{V}_H$. If the SVD of $V_H^\dagger \hat{V}_H$ has the form $\tilde{U}\tilde{\Sigma}\tilde{V}^\dagger$, we have

$$\tilde{\Sigma} = \mathrm{diag}(\cos\theta_1, \ldots, \cos\theta_m).$$

There are several possibilities for the distance between subspaces. One of them is the *chordal distance*, which is defined as [20]

$$d_c(V_H, \hat{V}_H) = \left[\sum_{k=1}^{m} \sin^2 \theta_k \right]^{1/2}.$$

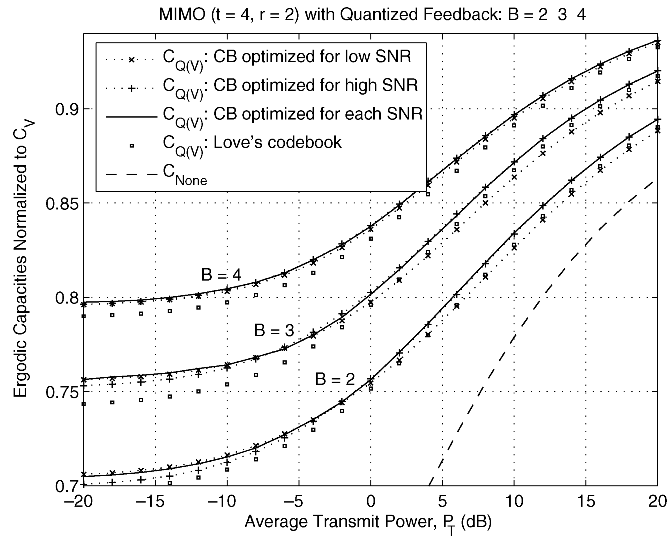

Fig. 1. Ergodic capacities of MIMO channels with quantized beamforming matrix for the different codebook design methods ($t = 4$, $r = 2$, $n = 2$, and $B = 2, 3, 4$).

We can easily express $\|V_H^\dagger \hat{V}_H\|_F^2$ in terms of the chordal distance

$$\left\| V_H^\dagger \hat{V}_H \right\|_F^2 = \sum_{k=1}^{m} \cos^2 \theta_k = m - d_c^2(V_H, \hat{V}_H). \tag{22}$$

Therefore, the design criterion (21) is equivalent to minimizing the *mean squared chordal distance* (MSCD) between the two subspaces specified by V_H and $\mathcal{Q}(V_H)$, i.e.,

$$\min_{\mathcal{Q}(\cdot)} E\left[d_c^2\left(V_H, \mathcal{Q}(V_H)\right) \right]. \tag{23}$$

Love and Heath consider the chordal distance in their structured codebook design method based on Grassmann subspace packing [21].

D. Design Examples

With the design algorithms developed above, we can obtain an optimum codebook for any set of system parameters, i.e., the number of antennas t and r, the number data streams n, and the number of feedback bits B $(N = 2^B)$.

The performances of codebooks designed with the three different design methods described in Section III-B and -C are compared in Fig. 1 in terms of the ergodic channel capacity, $C_{\mathcal{Q}(V)} = E[I(H, \hat{V})]$, for (4, 2) MIMO channel with independent identically distributed (i.i.d.) $\mathcal{N}(0,1)$ entries. For ease of comparison, all the capacities are normalized with respect to $C_V = E[I(H, V)]$, the ergodic capacity with the perfect beamforming matrix. As expected, it turns out that in high SNR region, the codebook optimized for high SNR region (Section III-C-2) performs better than that optimized for low SNR region (Section III-C-1); and in low SNR region, the reverse relation holds. Moreover, the generalized MSwIP method (Section III-B) always results in a performance better than or equal to any of the two in all the SNR range. The performance difference decreases as the number of bits B increases. As noted before, in the generalized MSwIP method, the codebook is optimized for each SNR point and the transmitter needs to know the operating SNR of the system *a priori*. This can be implemented in most practical systems by selecting a set of SNR points ahead and maintaining a particular SNR point over time using some form of power control mechanism. A simpler approach, with a minor performance loss, is to partition the SNR region in two and use the two codebooks optimized for high and low SNR regions. Additionally, we also compare with the codebook design method considered in [8], [15], and [21], wherein a codebook is designed using the algorithm developed for the noncoherent space-time constellation design in [22]. Details of the structured codebook design method can be found in [21]. We can see that the MSwIP method outperforms it in all the cases considered, especially in the low SNR range.

IV. Capacity Loss With Quantized Beamforming

In this section, we will attempt to analytically quantify the effect of quantization of beamforming matrix with a finite number of bits on the channel capacity for the i.i.d. MIMO channel. With a given number of feedback bits, we want to know how close one can approach the performance of the perfect beamforming-matrix feedback system. We consider a (t, r) MIMO system with quantized beamforming matrix $\hat{V}$ employed at the transmitter and with equal power allocation over the transmit symbols. The considered MIMO system is modeled as

$$y = H\hat{V}s + \eta$$

where H has i.i.d. $\tilde{\mathcal{N}}(0,1)$ entries, $s \sim \tilde{\mathcal{N}}_n(0, (P_T/n)I)$, $\eta \sim \tilde{\mathcal{N}}_r(0, I)$, and they are all independent.

We will analyze the capacity loss defined in (6), which is repeated here

$$\begin{aligned} C_L &= E_H\left[I_L\left(H, \mathcal{Q}(H)\right)\right] \\ &= E_H\left[-\log\det\left[I - \tilde{\Sigma}_H^2\left(I - V_H^\dagger \hat{V}\hat{V}^\dagger V_H\right)\right]\right]. \end{aligned}$$

Since $\hat{V} = \mathcal{Q}(H) = \hat{V}_i\ \forall H \in \mathcal{R}_i$ from (17), the capacity loss can be expressed as

$$\begin{aligned} C_L &= \sum_{i=1}^{N} P(H \in \mathcal{R}_i) E_{H\in\mathcal{R}_i}\left[I_L(H, \hat{V}_i)\right] \\ &= \sum_{i=1}^{N} P(H \in \mathcal{R}_i) \cdot \\ &\quad E_{H\in\mathcal{R}_i}\left[-\log\det\left[I - \tilde{\Sigma}_H^2\left(I - V_H^\dagger \hat{V}_i\hat{V}_i^\dagger V_H\right)\right]\right]. \end{aligned} \tag{24}$$

We need to calculate the expectation in (24). Due to the complex nature of the random variables involved and the complicated shape of the Voronoi region associated with a code point, it does not look analytically tractable. Instead of dealing with the general case, we therefore only consider the high-SNR approximation to the capacity loss using (9), which will be found to be more tractable. Also, we confine our attention to the case where the number of data stream equals to the rank of channel matrix, i.e., $n = m$. For the rest of this section, for notational simplicity, the subscript in V_H will be dropped. The high-SNR approximation to the capacity loss will be denoted by

$$\bar{C}_L = E_H\left[-\log\det(V^\dagger\hat{V}\hat{V}^\dagger V)\right]. \tag{25}$$

Since the high-SNR approximation (25) depends only on V and $\hat{V}$ (not on Σ_H and P_T), it makes the problem analytically more tractable.

In high SNR region, since $I_L(H, \hat{V}_i)$ is well approximated by $-\log\det(V^\dagger\hat{V}_i\hat{V}_i^\dagger V)$, the encoding scheme given in (17) can be rewritten as

$$\hat{V} = \mathcal{Q}(V) = \arg\max_{\hat{V}_i\in\mathcal{C}} \det\left(V^\dagger\hat{V}_i\hat{V}_i^\dagger V\right). \tag{26}$$

We denote the Stiefel manifold where the random matrix V lies by $\mathcal{V}_{n,t} = \{V \in \mathbb{C}^{t\times n} : V^\dagger V = I_n\}$. With the encoding scheme (26), the Stiefel manifold is partitioned into $\{\bar{\mathcal{R}}_i; i = 1, \ldots, N\}$, where[6]

$$\begin{aligned} \bar{\mathcal{R}}_i := \Big\{V \in \mathcal{V}_{n,t} : &\det\left(V^\dagger\hat{V}_i\hat{V}_i^\dagger V\right) \\ &\geq \det\left(V^\dagger\hat{V}_j\hat{V}_j^\dagger V\right), \quad \forall j \neq i\Big\}. \end{aligned} \tag{27}$$

Similar to (24), the high-SNR approximation (25) can be expressed as

$$\bar{C}_L = \sum_{i=1}^{N} P(V \in \bar{\mathcal{R}}_i) E_{V\in\bar{\mathcal{R}}_i}\left[-\log\det\left(V^\dagger\hat{V}_i\hat{V}_i^\dagger V\right)\right]. \tag{28}$$

In order to calculate (28), we need to know the conditional statistics of a random matrix $U_i := V^\dagger\hat{V}_i\hat{V}_i^\dagger V$ given that $V \in \bar{\mathcal{R}}_i$ ($\hat{V}_i$ is the code matrix for the partition cell $\bar{\mathcal{R}}_i$). For that purpose, we start with a simpler but related problem of the *unconditional* statistics of a random matrix $U_0 := V^\dagger V_0 V_0^\dagger V$, where V is uniformly distributed over $\mathcal{V}_{n,t}$ and V_0 is a *fixed* matrix in $\mathcal{V}_{n,t}$.[7]

A. Related Multivariate Statistics

Before continuing with the derivation of the statistics of U_0, we summarize related definitions and theorems from multivariate statistical analysis that are germane to the analysis. Since the multivariate statistical literature mainly considers real-valued matrices and it is hard to find results for the complex matrices, we summarize the related theorems for the complex random matrices.

Definition 1 (Complex Multivariate Gaussian Distribution): The complex random matrix $X(p \times n)$ is said to have normal distribution $\tilde{\mathcal{N}}_{p,n}(M, C\otimes D)$ if $E[X] = M$ and $C \otimes D$ is the covariance matrix of $\text{vec}(X^\dagger)$, that is, $\text{vec}(X^\dagger) \sim \tilde{\mathcal{N}}_{pn}(\text{vec}(M^\dagger), C\otimes D)$, where $C(p\times p) > 0$ and $D(n \times n) > 0$.

[6]The partitioning by (27) is subspace invariant in a sense that if $V \in \bar{\mathcal{R}}_i$, then $VQ \in \bar{\mathcal{R}}_i$ for some unitary matrix Q. Thus, we can also think of a similar partition of the Grassmann manifold.

[7]It can be shown that for the channel H with i.i.d. $\tilde{\mathcal{N}}(0,1)$ entries, the random matrix V is uniformly distributed over $\mathcal{V}_{n,t}$ [23].

Definition 2 (Complex Wishart Distribution): If $p \times p$ matrix $A = ZZ^\dagger$, where $Z \sim \tilde{\mathcal{N}}_{p,n}(0, \Sigma \otimes I_n)$, then A is said to have the Wishart distribution with n degrees of freedom and covariance matrix Σ. It will be denoted as $A \sim \tilde{\mathcal{W}}_p(n, \Sigma)$, and when $n \geq p$ the density function of A is

$$f(A) = \frac{1}{\tilde{\Gamma}_p(n)(\det \Sigma)^n} \mathrm{etr}(-\Sigma^{-1}A)(\det A)^{n-p} \tag{29}$$

where $\mathrm{etr}(\cdot) = \exp \mathrm{tr}(\cdot)$ and $\tilde{\Gamma}_p(n)$ is the complex multivariate gamma function

$$\tilde{\Gamma}_p(n) = \pi^{p(p-1)/2} \prod_{k=1}^{p} \Gamma(n-k+1).$$

See, e.g., [24, Th. 5.1] for the derivation of the density function.

Lemma 2: If $X \sim \tilde{\mathcal{N}}_{p,n}(0, \Sigma \otimes I_n)$ and P is an $n \times n$ Hermitian idempotent matrix of rank k, then $XPX^\dagger \sim \tilde{\mathcal{W}}_p(k, \Sigma)$.

Proof: See Theorem 3.2.5 of [25] or Problem 3.10 of [23]. The proof for the real matrix case in [25] also can be readily applied to the complex case. □

Definition 3 (Beta Distribution): A random variable x has the beta distribution with parameter (a, b), denoted as $x \sim \beta(a, b)$, if its density function is

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \ 0 < x < 1 \tag{30}$$

for $a > 0$, $b > 0$, and $f(x) = 0$ elsewhere.

Definition 4 (Complex Multivariate Beta Distribution): A $p \times p$ random Hermitian positive definite matrix U is said to have a multivariate beta distribution with parameter (a, b), denoted as $U \sim \tilde{\mathcal{B}}_p(a, b)$, if its density function is given by

$$f(U) = \frac{\tilde{\Gamma}_p(a+b)}{\tilde{\Gamma}_p(a)\tilde{\Gamma}_p(b)} (\det U)^{a-p} \det(I-U)^{b-p}, \ 0 < U < I_p \tag{31}$$

for $a \geq p$, $b \geq p$, and $f(U) = 0$ elsewhere. $0 < U < I_p$ means that U and $I_p - U$ are positive definite.

Theorem 1: Let A and B be independent, where $A \sim \tilde{\mathcal{W}}_p(n_1, \Sigma)$ and $B \sim \tilde{\mathcal{W}}_p(n_2, \Sigma)$, with $n_1 \geq p$, $n_2 \geq p$. Let $A + B = T^\dagger T$ where T is an upper triangular $p \times p$ matrix and U be the $p \times p$ Hermitian matrix defined from $A = T^\dagger U T$. Then $A + B$ and U are independent; $A + B \sim \tilde{\mathcal{W}}_p(n_1 + n_2, \Sigma)$ and $U \sim \tilde{\mathcal{B}}_p(n_1, n_2)$.

Proof: The proof is similar to that of [23, Th. 3.3.1] except that now we consider the complex matrix case. For a detailed proof, the reader is referred to [26]. □

Remark 1: The condition $n_1 \geq p$, $n_2 \geq p$ for the definition of the multivariate beta distribution can be relaxed in order to include the case when the Wishart matrix B in Theorem 1 is singular (i.e., $n_2 < p$). Mitra [27] assumed $n_1 + n_2 \geq p$ and then studied certain properties of the random matrix U (for the real case) using a density-free approach. This method can be extended to the complex matrix case.

Theorem 2: If $U \sim \tilde{\mathcal{B}}_p(n_1, n_2)$ with $n_1 \geq p$ and $U = T^\dagger T$, where T is upper triangular, then $t_{11}^2, \ldots, t_{pp}^2$ are all independent and $t_{ii}^2 \sim \beta(n_1 - i + 1, n_2)$ for $i = 1, \ldots, p$.

Proof: See the Appendix.

B. Statistics of $U_0 = V^\dagger V_0 V_0^\dagger V$ and $\gamma_0 = \det(U_0)$

Now, returning to the problem of statistics of $U_0 = V^\dagger V_0 V_0^\dagger V$, using the theorems above we arrive at the following result.

Theorem 3 (Distribution of $V^\dagger V_0 V_0^\dagger V$): Let random matrix $H \sim \tilde{\mathcal{N}}_{r,t}(0, \Sigma \otimes I_t)$ and $V(t \times m)$ be the right singular matrix of H as defined in Section II. For a fixed orthonormal column matrix $V_0(t \times m)$, the random matrix $U_0 = V^\dagger V_0 V_0^\dagger V$ has a multivariate beta distribution with parameters m and $t-m$, that is, $U_0 \sim \tilde{\mathcal{B}}_m(m, t-m)$.

Proof: We know that since $H \sim \tilde{\mathcal{N}}_{r,t}(0, \Sigma \otimes I_t)$ and $t \geq r = m$, $HH^\dagger \sim \tilde{\mathcal{W}}_m(t, \Sigma)$. By considering a $t \times t$ idempotent matrix $P = V_0 V_0^\dagger$, using Lemma 2, we can show that

$$HPH^\dagger \sim \tilde{\mathcal{W}}_m(m, \Sigma)$$
$$H(I-P)H^\dagger \sim \tilde{\mathcal{W}}_m(t-m, \Sigma).$$

Now we define $A = HPH^\dagger$ and $B = H(I-P)H^\dagger$ to use Theorem 1. Let the QR decomposition of $H^\dagger$ be given by $H^\dagger = QR$ (in economy-size representation), where $Q(t \times m)$ is an orthonormal column matrix and $R(m \times m)$ is an upper triangular matrix. Then, $A + B = R^\dagger R$ and $A = HPH^\dagger = R^\dagger Q^\dagger PQR$. From Theorem 1

$$U = Q^\dagger PQ = Q^\dagger V_0 V_0^\dagger Q \sim \tilde{\mathcal{B}}_m(m, t-m). \tag{32}$$

By noticing that Q and V have the same distribution, we arrive at the result that $U_0 = V^\dagger V_0 V_0^\dagger V$ has the distribution $\tilde{\mathcal{B}}_m(m, t-m)$. □

If $U_0 = T^\dagger T$, where T is upper triangular, $\det(U_0) = \prod_{i=1}^m t_{ii}^2$. But, it is known from Theorem 2 that $t_{ii}^2 \sim \beta(m - i + 1, t - m)$ and they are independent. This is summarized in the following corollary.

Corollary 1: $\det(U_0)$ is distributed as the product of independent beta variables, that is

$$\gamma_0 = \det(U_0) \sim \prod_{i=1}^{m} \beta_i$$

where $\beta_i \sim \beta(m - i + 1, t - m)$ and are independently distributed.

We can obtain the density function for a given m although it has a long and complicated form (e.g., [28]). For $m = 2$, we have a relatively concise expression.

Corollary 2: When $m = 2$, $\gamma_0 = \det(U_0)$ has the following distribution:

$$f_{\gamma_0}(x) = \frac{\Gamma(t)\Gamma(t-1)}{\Gamma(2t-4)} (1-x)^{2t-5} \cdot {}_2F_1(t-2, t-3; 2t-4; 1-x). \tag{33}$$

Proof: See [28].

C. Approximate Density of $\gamma = \det(V^\dagger \hat{V} \hat{V}^\dagger V)$

Now let us look at the conditional density of $\gamma_i := \det(V^\dagger \hat{V}_i \hat{V}_i^\dagger V)$ given $V \in \bar{\mathcal{R}}_i$. From the high-SNR encoding given in (26) or (27), generally each partition cell has a complicated shape defined by neighboring code matrices, and the partition cells have all different shapes. This geometrical complexity in the partition cells makes it difficult to obtain the exact analytical form for the conditional density of γ_i.

However, when N is large, since V is uniformly distributed over $\mathcal{V}_{n,t}$, then a reasonable approximation is that $P(V \in \bar{\mathcal{R}}_i) \simeq 1/N$ for all i, and that the shapes of the partition cells will be approximately identical. For analytical tractability, we consider the following approximation for the partition cell:

$$\bar{\mathcal{R}}_i \simeq \tilde{\mathcal{R}}_i := \left\{V \in \mathcal{V}_{n,t} : \det\left(V^\dagger \hat{V}_i \hat{V}_i^\dagger V\right) \geq 1 - \delta_B\right\} \quad (34)$$

for some $\delta_B > 0$, which will be determined as a function of B using $P(V \in \bar{\mathcal{R}}_i) = 1/N$ for all i, i.e.,

$$P(V \in \bar{\mathcal{R}}_i) = \int_{1-\delta_B}^{1} f_{\gamma_i}(x)dx = \frac{1}{2^B}. \quad (35)$$

With the partition cell approximation (34) and from the symmetrical property in the partition cells, we can use f_{γ_0} in the place of f_{γ_i}. Using the density function for γ_0 (e.g., the one derived in Corollary 2), δ_B can be numerically calculated for a given B. Although in general there are overlaps in the approximated partition cells, the analytical results from the approximation turn out to be quite accurate even when N is small. The quantization cell approximation of (34) is similar in some sense to those in quantization error analysis for high-rate VQ in source coding study. In the high-rate VQ study, a quantization cell is approximated with a hyperellipsoid having the same volume of the cell. In our case, each quantization cell is approximated to the simple geometrical region defined in (34) having the same *probability* as the cell. The partition cell approximation is also similar to what was introduced in [2] and [3] for MISO systems. In [2], Mukkavilli *et al.* considered a geometrical region (called *spherical cap*) on a constant-norm sphere to obtain a union bound for the area of no outage region.

With the partition cell approximation, since the approximated cells $\tilde{\mathcal{R}}_i$ have the identical geometrical shape and the probabilities $P(V \in \tilde{\mathcal{R}}_i)$ are the same, the random variables γ_i given $V \in \tilde{\mathcal{R}}_i$ have the same density of

$$\begin{aligned} f_{\gamma_i}(x|V \in \tilde{\mathcal{R}}_i) &= \frac{f_{\gamma_0}(x)1_{[1-\delta_B,1)}(x)}{P(V \in \tilde{\mathcal{R}}_i)} \\ &= 2^B f_{\gamma_0}(x)1_{[1-\delta_B,1)}(x), \text{ for all } i \end{aligned} \quad (36)$$

where $1_A(x)$ is the indicator function having one if $x \in A$ and zero otherwise. Therefore, it is enough to focus on a particular partition cell. Thus, we arrive at the following result.

Approximate Density: With the partition cell approximation described in (34), the density function for $\gamma := \det(V^\dagger \hat{V} \hat{V}^\dagger V)$ is approximated by a *truncated* distribution of $f_{\gamma_0}(x)$, that is

$$f_\gamma(x) \simeq \tilde{f}_\gamma(x) = 2^B f_{\gamma_0}(x)1_{[1-\delta_B,1)}(x). \quad (37)$$

D. Capacity Loss With Quantized Beamforming

Now we utilize the statistical results developed above to analyze the problem of interest, the capacity loss analysis (25). This can be written as

$$\bar{C}_L = \int_0^1 [-\log x] f_\gamma(x)dx. \quad (38)$$

For the above expectation, we will use the approximate density $\tilde{f}_\gamma$ given in (37) instead of the real density f_γ. The new approximate is denoted by $\tilde{C}_L$, that is

$$\tilde{C}_L = \int_{1-\delta_B}^1 [-\log x]\tilde{f}_\gamma(x)dx. \quad (39)$$

Lemma 3: The capacity loss approximation $\tilde{C}_L$ in (39) is a *lower bound* on $\bar{C}_L$ in (38). That is

$$\tilde{C}_L \leq \bar{C}_L. \quad (40)$$

Proof: This can be proved by directly following the proof of a similar theorem for the MISO system case [3] (or [29]). It can shown that $f_\gamma(x) \leq \tilde{f}_\gamma(x)$ for $1 - \delta_B \leq x < 1$. By using this and the fact that the function $-\log x$ is a monotonically decreasing function, the result (40) can be proved. □

Using (39), a closed-form expression can be derived for $m = 2$, and an efficient computational procedure is developed in the general case ($m \geq 3$). For $m = 2$ (i.e., when $t \geq 2$ and $r = 2$), since we have a relatively simple form of density function, we can derive a closed-form expression for the capacity loss. That is, when $m = 2$, using the density f_{γ_0} in Corollary 2 and the definition of the hypergeometric function

$$\begin{aligned} \tilde{C}_L = 2^B \frac{\Gamma(t)\Gamma(t-1)}{\Gamma(2t-4)} \sum_{k=0}^{\infty} \frac{(t-2)_k(t-3)_k}{(2t-4)_k k!} \\ \cdot \int_{1-\delta_B}^1 (1-x)^{k+2t-5}[-\log x]dx \end{aligned} \quad (41)$$

where $(a)_k := a(a+1)\cdots(a+k-1)$, called the Pochhammer symbol. The integral in (41) can be solved using [30, (2.729.1)]. Finally, for $m = 2$, we have

$$\begin{aligned} \tilde{C}_L = 2^B \frac{\Gamma(t)\Gamma(t-1)}{\Gamma(2t-4)} \sum_{k=0}^{\infty} \frac{(t-2)_k(t-3)_k}{(2t-4)_k k!(k+2t-4)} \\ \cdot \left[\left(1-\delta_B^{k+2t-4}\right)\log(1-\delta_B) + \sum_{l=1}^{k+2t-4} \frac{\delta_B^{k-l+2t-3}}{k-l+2t-3} \right]. \end{aligned} \quad (42)$$

For the general case of $m \geq 3$, one can obtain the approximate capacity loss (39) easily with an efficient Monte Carlo integration method which is described as follows: i) generate a large number of samples for the random variable γ; each is just a product of independent beta distributed random variables as shown in Corollary 1; ii) take a subset $S = \{\gamma > 1-\delta_B : P(\gamma > 1-\delta_B) = 1/2^B\}$; and iii) average over the subset S for $E_S[-\log \gamma]$, which is an estimate (39).

1) Analysis Through Gamma Distribution Fitting: We now develop an alternate approach for $m \geq 3$ based on distribution fitting. In the multivariate analysis literature, a random variable $\nu := -\log\det(U)$, where $U \sim \tilde{\mathcal{B}}_p(n_1, n_2)$, is approximately distributed as a χ^2 distribution with appropriate degrees of freedom that is dependent on the parameters of U (see, e.g., [31, Ch. 8]). This is called Bartlett's approximation. But, the approximation is accurate only when n_1 is moderately large, which unfortunately does not hold in our case. To circumvent this problem, we attempt to match the distribution using a gamma distribution. Since the gamma distribution is a larger

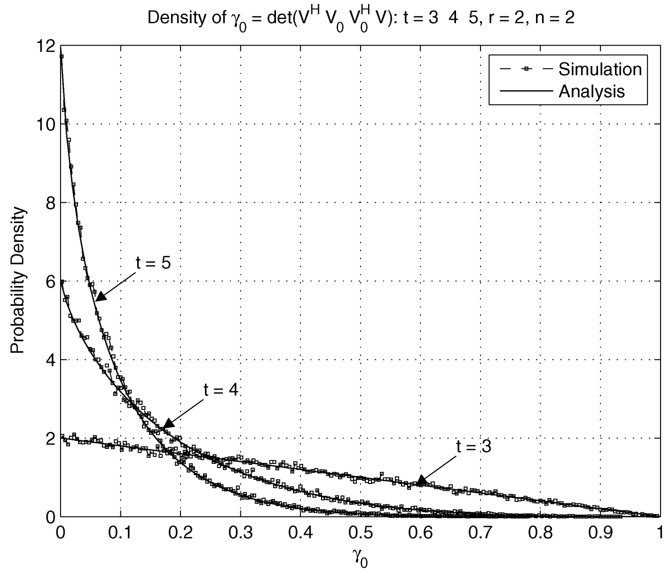

Fig. 2. Probability density of γ_0 ($t = 3, 4, 5$; $r = 2$; $n = 2$).

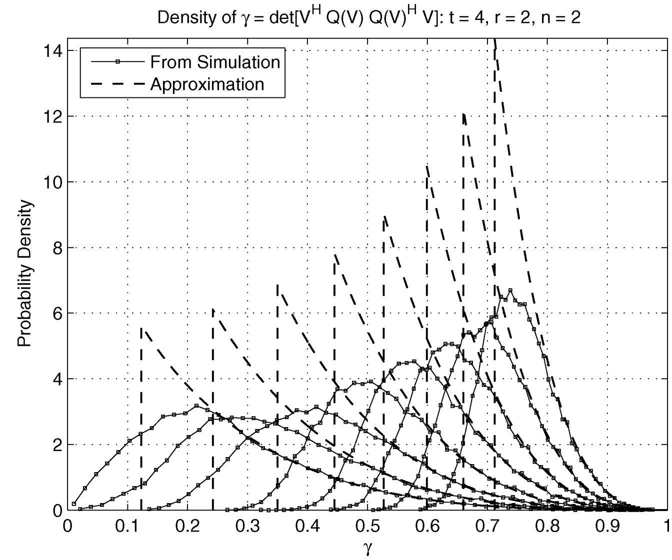

Fig. 3. Probability density of $\gamma = \det[V^\dagger \mathcal{Q}(V)\mathcal{Q}(V)^\dagger V]$ ($t = 4$, $r = 2$, $n = 2$; $B = 1, \ldots, 8$).

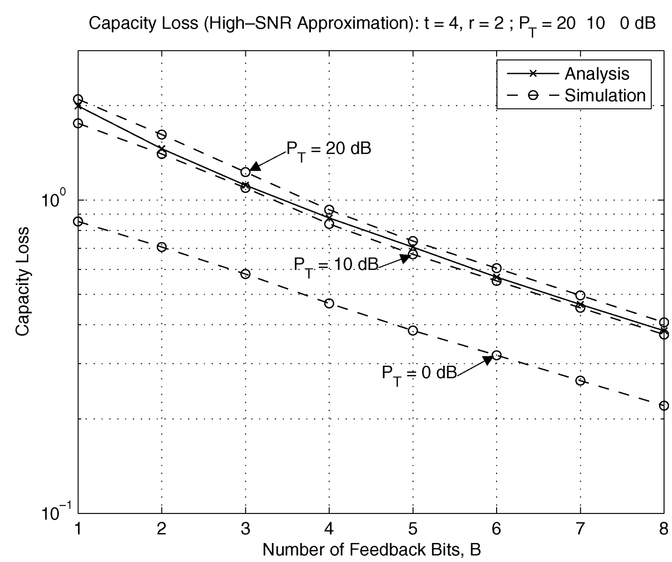

Fig. 4. Capacity loss (in bits per channel use) of MIMO channel due to quantized beamforming matrix ($t = 4$, $r = 2$; $P_T = 20, 10, 0$ dB; $B = 1, 2, \ldots, 8$).

family of distribution that includes the χ^2 distribution, we expect a better match.

The gamma distribution has the form

$$f_\Gamma(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$$

characterized by two parameters (α, β) that have to be estimated. The maximum likelihood estimate for (α, β), denoted by $(\hat{\alpha}, \hat{\beta})$, is employed in this paper because of its known attractive statistical properties as an estimator (details can be found in the statistics literature, and also it is implemented in Matlab Statistics Toolbox). To obtain a better fit in the lower end, which is the important region in our analysis, we also consider a samples truncation technique: For the distribution fitting, use the subset of samples $\{\nu \le \nu_U : P(\nu \le \nu_U) = 1 - \epsilon\}$ for small $\epsilon > 0$.

Once the approximate distribution is found, we can solve for ν_B that satisfies $F_\Gamma(\nu_B; \hat{\alpha}, \hat{\beta}) = 1/2^B$ in a numerical way, where $F_\Gamma(x; \hat{\alpha}, \hat{\beta})$ is the cumulative distribution corresponding to the gamma density $f_\Gamma(x; \hat{\alpha}, \hat{\beta})$. Then, we can approximately compute the capacity loss as

$$\begin{aligned} \tilde{C}_L &\simeq 2^B \int_0^{\nu_B} x f_\Gamma(x; \hat{\alpha}, \hat{\beta}) dx \\ &= \frac{2^B \hat{\beta}}{\Gamma(\hat{\alpha})} \gamma(\hat{\alpha} + 1, \nu_B/\hat{\beta}) \end{aligned} \tag{43}$$

where $\gamma(\alpha, x) := \int_0^x t^{\alpha-1} e^{-t} dt$ is the incomplete gamma function [30, (8.350)].

E. Numerical Results

In Fig. 2, the analytical result for the density of γ_0 derived in Section IV-B using a multivariate statistical approach is verified by comparing with simulation results. Fig. 3 shows the approximate density $\tilde{f}_\gamma$ for $t = 4$, $r = 2$, and various B, together with actual densities from simulations using the codebooks designed by the MSCN method. We can see that as B increases, the distribution of γ moves towards one. The approximate density functions follow this behavior of the simulation results. Also, if we look at a particular approximate density, it is very close to its counterpart (from simulation) in the area around $\gamma = 1$, but they do not agree around $\gamma = \delta_B$. This is a result of the partition cell approximation.

Fig. 4 shows the capacity loss in bits per channel use for ($t = 4$, $r = 2$) MIMO channels when the beamforming matrix is represented with $B = 1, \ldots, 8$. The analytical result is from $\tilde{C}_L$ given in (42), which was derived for $m = 2$ under the high-SNR approximation. The figure also includes simulation results using the beamforming-matrix codebooks. It shows that the analytical result is close to the simulation result at high SNR (e.g., $P_T = 20, 10$ dB). However, at low SNR (e.g., $P_T = 0$ dB) it deviates from the simulation result. This is to be expected because in our analysis, with the high-SNR assumption, the effect of the term $\tilde{\Sigma}_H^2 = (I + \rho\Sigma_H^2)^{-1} \rho\Sigma_H^2$ is ignored, which is on average quite

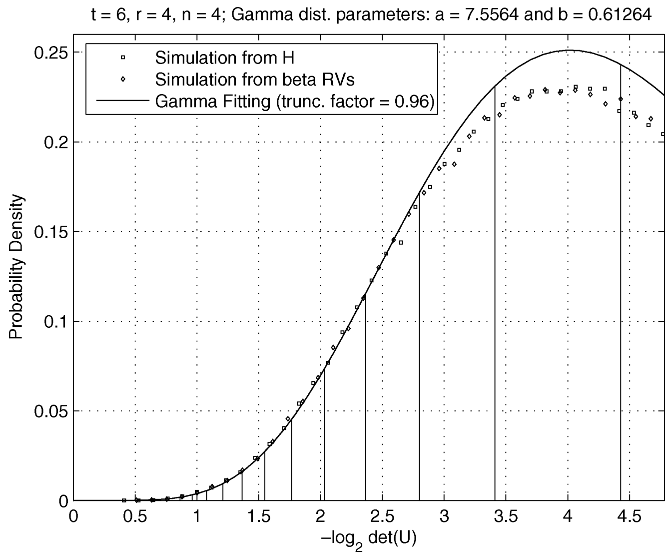

Fig. 5. Gamma distribution fitting of $\nu = -\log_2 \det(U)$ with samples truncation ($t = 6, r = 4, n = 4; 1 - \epsilon = 0.96$).

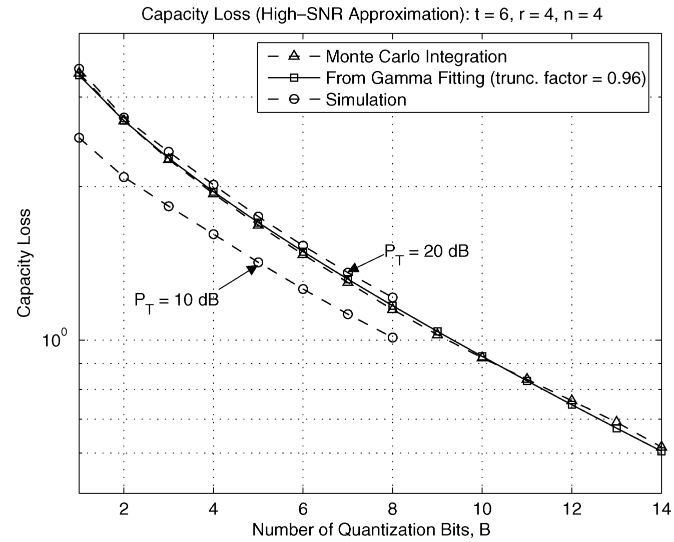

Fig. 6. Capacity loss (in bits per channel use) of MIMO channel due to quantized beamforming matrix ($t = 6, r = 4, n = 4$; $P_T = 20, 10$ dB; $B = 1, 2, \ldots, 8$).

different from I at low SNR (P_T). Therefore, the high-SNR assumption results in higher values for capacity loss at low SNR.

For $m \geq 3$, we considered fitting $\nu = -\log\det(U)$ to a gamma distribution with the samples truncation technique applied. Fig. 5 shows three densities of ν for $t = 6$ and $r = 4$: from random channel matrix H, from beta distributed independent random variables, and from the gamma distribution fitting. The truncation factor $1-\epsilon$ for the gamma fitting is set to 0.96 (after several trials), which was selected for a good fit in the low end. The two estimated parameters for the gamma distribution in this example are $\hat{a} = 7.5564$ and $\hat{b} = 0.6126$. The vertical lines in the figure indicate the ν_B that satisfies $P(\nu \leq \nu_B) = 1/2^B$ for different B. Fig. 6 shows the capacity loss for $t = 6, r = 4$, and $n = 4$. From the figure we can see that the result from the gamma fitting is very close to that from the Monte Carlo integration, and both predict the performance from simulation with good accuracy.

V. Multimode Spatial Multiplexing Transmission Strategy

Equal power allocation on each of the parallel channels is clearly inefficient at lower SNR. To overcome this limitation, in this section, we present a multimode MIMO spatial multiplexing transmission scheme that allows for efficient utilization of the feedback bits. The transmission strategy is described as follows:

1) The number of data streams n is determined based on the average SNR: n changes from one to m, the rank of the channel, as the average SNR increases (see the example in Section V-B).
2) In each mode, the simple equal power allocation over n spatial channels is employed; thereby the entire feedback bits are utilized in representing only useful beamforming vectors without concerning about the power allocation over the spatial channels.

The average SNR is assumed to change at a much slower rate than the beamforming vectors, and so has to be updated less frequently consuming much fewer bits. Therefore it can be assumed to be negligible overhead. The effectiveness of the transmission strategy can be understood with a water-filling argument. That is, in low SNR region, only the principal spatial channel is useful most of the time; and as the SNR increases, more spatial channels become involved. The proposed scheme can be viewed as a rough and indirect implementation of water-filling power allocation over multiple spatial channels. With the multimode transmission scheme, it will be shown later that with perfect knowledge of $V(t \times n)$ and equal power allocation, we can achieve most of the capacity with perfect CSIT over all the range of SNR. For finite-rate feedback, only the relevant beamforming vectors are encoded using the entire feedback bits. From a quantization point of view, the multimode scheme increases the quantization resolution by reducing the dimension of the beamforming matrix in low SNR region.

A similar idea of optimizing the number of data streams is also discussed in [7] for asymptotically large systems (the number of transmit and receive antennas goes infinite). In [15], another form of multimode SM scheme was studied in which the mode is adapted using current channel condition and the feedback bits are divided among multiple codebooks, one for each mode. Changing the SM mode at the same rate as the beamforming vectors is less effective for the scenario envisioned here, as it lowers the quantization resolution compared to the scheme mentioned above.

A. Codebook Design for Beamforming Matrix (When $n < m$)

In Section III, we have discussed codebook design for the full-rank beamforming matrix ($n = m$). Here we will discuss codebook design for the case $n < m$. Let us define submatrices

$$\Sigma_H = \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma_2 \end{bmatrix} \text{ and } V_H = [V \quad V_2]$$

where Σ $(n \times n)$ and V $(t \times n)$ and also define

$$V_H^\dagger \hat{V}\hat{V}^\dagger V_H = \begin{bmatrix} V^\dagger\hat{V}\hat{V}^\dagger V & V^\dagger\hat{V}\hat{V}^\dagger V_2 \\ V_2^\dagger\hat{V}\hat{V}^\dagger V & V_2^\dagger\hat{V}\hat{V}^\dagger V_2 \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}.$$

When the transmitter uses $\hat{V}(t \times n)$ as beamforming matrix and equal power allocation, the mutual information given in (4) can be written as

$$\begin{aligned} I(H,\hat{V}) &= \log\det\left(I_m + \rho \begin{bmatrix} \Sigma^2 & 0 \\ 0 & \Sigma_2^2 \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}\right) \\ &= \log\det(I_n + \rho\Sigma^2 U_{11}) \\ &\quad + \log\det\left[I_{m-n} + \rho\Sigma_2^2 U_{22} - \rho\Sigma_2^2 U_{21} \right. \\ &\quad \left. \cdot (I_n + \rho\Sigma^2 U_{11})^{-1} \rho\Sigma^2 U_{12}\right] \end{aligned} \tag{44}$$

where (44) is obtained using $\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B)$. If the number of feedback bits is reasonably large, the second term in (44) is relatively small since the submatrices U_{12}, U_{21}, and U_{22} are close to zeros matrices; therefore

$$I(H,\hat{V}) \gtrsim \log\det(I_n + \rho\Sigma^2 U_{11}) =: \check{I}(H,\hat{V}). \tag{45}$$

Then, the capacity loss defined in (6) is bounded as

$$C_L(H,\hat{V}) \lesssim E\left[\check{I}_L(H,\hat{V})\right] =: \check{C}_L(H,\hat{V}) \tag{46}$$

where $\check{I}_L(H,\hat{V}) := I(H,V) - \check{I}(H,\hat{V})$.

We employ the upper bound $\check{C}_L(H,\hat{V})$ as the design objective for the codebook design. In a manner similar to that in Section III, we have the following codebook design criterion:

$$\max_{\mathcal{Q}(\cdot)} E\left\|(V\tilde{\Sigma})^\dagger \mathcal{Q}(H)\right\|_F^2 \tag{47}$$

where $\hat{V} = \mathcal{Q}(H)$ is the $t \times n$ quantized beamforming matrix ($\hat{V}^\dagger\hat{V} = I_n$). The corresponding design algorithm is similar to that for $n = m$ in Section III-B except now submatrices V and Σ are used in places of V_H and Σ_H. Also, as in Section III-C, it leads to two related design methods, each optimized for low- and high-SNR region

$$\max_{\mathcal{Q}(\cdot)} E\left\|(V\Sigma)^\dagger \mathcal{Q}(H)\right\|_F^2 \tag{48}$$

$$\max_{\mathcal{Q}(\cdot)} E\left\|V^\dagger \mathcal{Q}(V)\right\|_F^2. \tag{49}$$

B. Examples

Fig. 7 shows the ergodic capacities of MIMO channel ($t = 6$, $r = 4$) in different transmission modes ($n = 1, \ldots, 4$) when the beamforming matrix is *perfectly* known to the transmitter and equal power allocation over spatial channels is employed. Even though each of the modes are by themselves not adequate for the entire SNR range, they are near optimal over a range of SNR. By switching modes based on the SNR, one can then make the best of each of the modes. Even though there are losses especially around the crossing points, the loss is small. The equal power allocation scheme is particularly attractive in the quantized scenario since no bits have to be wasted to convey power allocation information. The figure shows that with the proposed multimode SM scheme, we can achieve most of the capacity of full feedback without feeding back the power allocation information.

In Fig. 8, the performance of the multimode SM scheme with a finite number of feedback bits ($B = 8$) is shown for the same

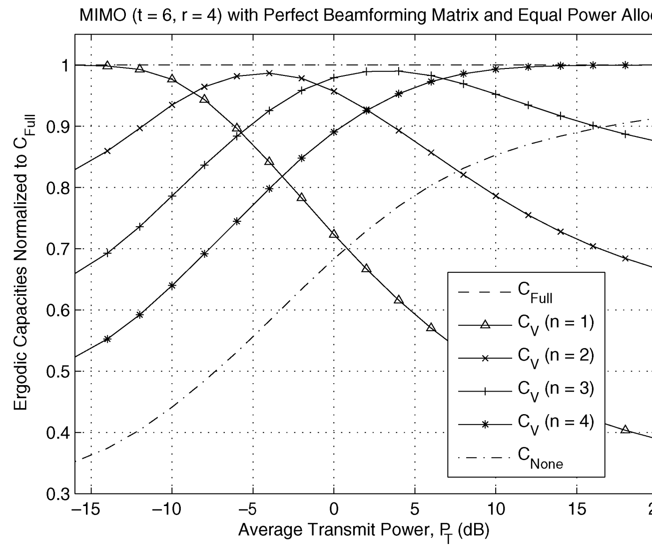

Fig. 7. Ergodic capacity of multimode MIMO spatial multiplexing systems with perfect beamforming matrix and equal power allocation at the transmitter ($t = 6, r = 4, n = 1, \ldots, 4$).

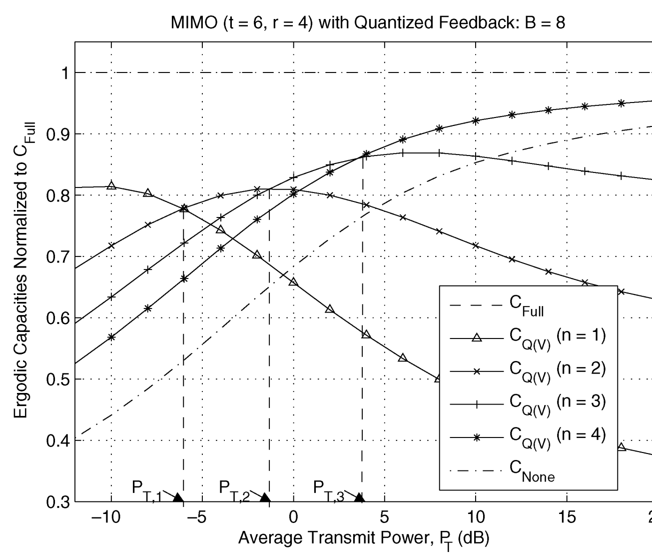

Fig. 8. Ergodic capacity of multimode MIMO spatial multiplexing systems with quantized beamforming matrix ($t = 6, r = 4, n = 1, \ldots, 4; B = 8$).

MIMO configuration as in Fig. 7. The number of data streams n is determined as follows: $n = 1$ when $P_T < P_{T,1}$, $n = 2$ when $P_{T,1} \le P_T < P_{T,2}$, and $n = 3$ when $P_T \ge P_{T,2}$, where $P_{T,1}$ and $P_{T,2}$ are the boundary points (see the figure). The gap from the performance with perfect beamforming matrix results from the finite-bit representation of the beamforming matrix. To determine the mode, the transmitter needs to know the operating SNR of the system, which is assumed to be changing at a much slower rate than the channel itself in most time-varying channel environments. Therefore, only small additional feedback is necessary (e.g., to notify the transmitter to increase or decrease the number of data streams n).

VI. Conclusion

We have investigated the codebook design problem associated with transmit beamforming in MIMO spatial multiplexing systems with finite-rate feedback. Assuming a fixed number of spatial channels and equal power allocation, we designed the beamforming codebook by striving to minimize the capacity loss resulting from the finite-rate feedback. The capacity loss, under the assumption of a reasonably large number of feedback bits or low SNR, was suitably approximated leading to an iterative codebook design algorithm with monotonic convergence property. The developed design algorithm is based on the Lloyd algorithm in vector quantization study but now has as its objective matrix quantization to minimize capacity loss. With the proposed method, we can design the optimum beamforming codebook for arbitrary number of transmit and receive antennas, feedback bits, and any spatial correlation structure in the channel. The effect on the MIMO channel capacity of finite-rate feedback was analyzed assuming high-SNR and equal power allocation over the spatial channels. Central to the analysis is the complex multivariate beta distribution and simplifications of the Voronoi regions resulting from the codebook. To compensate for the degradation due to the equal power allocation assumption, we also proposed a multimode spatial multiplexing transmission strategy that allows for efficient utilization of the feedback bits by quantizing only relevant beamforming vectors. The multimode transmission scheme can be viewed as a rough and indirect implementation of water-filling power allocation over the multiple spatial channels.

APPENDIX
PROOF OF THEOREM 2

Proof: First, we provide a proof based on the multivariate beta density $\tilde{\mathcal{B}}_p(n_1, n_2)$, which is valid under the condition $n_1 \geq p$ and $n_2 \geq p$. The proof follows that of [23, Th. 3.3.3], which is for the case of real-valued matrices. In the density function for U

$$f(U) = \frac{\tilde{\Gamma}_p(n_1+n_2)}{\tilde{\Gamma}_p(n_1)\tilde{\Gamma}_p(n_2)} (\det U)^{n_1-p} \det(I-U)^{n_2-p}$$

make change of variables $U = T^\dagger T$; then

$$\det U = \det(T^\dagger T) = \prod_{i=1}^{p} t_{ii}^2$$

and, from [24]

$$(dU) = 2^p \prod_{i=1}^{p} t_{ii}^{2p-2i+1} (dT)$$

so that the density of T is $f(T; p, n_1, n_2)$, where

$$f(T;p,n_1,n_2) = \frac{\tilde{\Gamma}_p(n_1+n_2)}{\tilde{\Gamma}_p(n_1)\tilde{\Gamma}_p(n_2)} 2^p \cdot \prod_{i=1}^{p} t_{ii}^{2n_1-2i+1} \det\left(I - T^\dagger T\right)^{n_2-p}. \quad (50)$$

Now partition T as

$$T = \begin{bmatrix} t_{11} & t^\dagger \\ 0 & T_{22} \end{bmatrix}$$

where t is $(p-1)\times 1$ and T_{22} is $(p-1)\times(p-1)$ and upper triangular; note that

$$\begin{aligned} \det(I - T^\dagger T) &= \det \begin{bmatrix} 1-t_{11}^2 & -t_{11}t^\dagger \\ -t_{11}t & I - tt^\dagger - T_{22}^\dagger T_{22} \end{bmatrix} && (51) \\ &= \left(1-t_{11}^2\right) \det\left(I - T_{22}^\dagger T_{22} - \frac{1}{1-t_{11}^2} tt^\dagger\right) && (52) \\ &= \left(1-t_{11}^2\right) \det\left(I - T_{22}^\dagger T_{22}\right) \\ &\quad \cdot \det\left[I - \frac{1}{1-t_{11}^2}\left(I - T_{22}^\dagger T_{22}\right)^{-1} tt^\dagger\right] && (53) \\ &= \left(1-t_{11}^2\right) \det\left(I - T_{22}^\dagger T_{22}\right) \\ &\quad \cdot \left[1 - \frac{1}{1-t_{11}^2} t^\dagger \left(I - T_{22}^\dagger T_{22}\right)^{-1} t\right] && (54) \end{aligned}$$

where in (52) an identity $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B)$ is used. Now make a change variables from t_{11}, T_{22}, t to t_{11}, T_{22}, v, where

$$v = \frac{1}{\left(1-t_{11}^2\right)^{1/2}} \left(I - T_{22}^\dagger T_{22}\right)^{-1/2} t.$$

Then

$$\begin{aligned}(dT) &= dt_{11} \wedge (dT_{22}) \wedge (dt) \\ &= \left(1-t_{11}^2\right)^{p-1} \det\Big(I - T_{22}^\dagger T_{22}\Big) dt_{11} \wedge (dT_{22}) \wedge (dv)\end{aligned}$$

by the Jacobian of transformation of t to v, and hence the density of t_{11}, T_{22} and v is

$$\frac{2^p\tilde{\Gamma}_p(n_1+n_2)}{\tilde{\Gamma}_p(n_1)\tilde{\Gamma}_p(n_2)} t_{11}^{2n_1-1}\left(1-t_{11}^2\right)^{n_2-1} \cdot \prod_{i=2}^{p} t_{ii}^{2n_1-2i+1} \det\left(I - T_{22}^\dagger T_{22}\right)^{n_2-p+1} (1 - v^\dagger v)^{n_2-p}.$$

This shows immediately that t_{11}, T_{22}, and v are all independent and t_{11} has the density

$$f(t_{11}) = \frac{2\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} t_{11}^{2n_1-1}\left(1-t_{11}^2\right)^{n_2-1}.$$

By a transformation of t_{11} to t_{11}^2, it is shown that $t_{11}^2 \sim \mathcal{B}(n_1, n_2)$. The density function of T_{22} is proportional to

$$\prod_{i=2}^{p} t_{ii}^{2n_1-2i+1} \det\left(I - T_{22}^\dagger T_{22}\right)^{n_2-p+1}$$

which has the same form as the density function of (50) for T, with p replaced by $p-1$ and n_1 replaced by $n_1 - 1$. Hence the density function of T_{22} is $f(T_{22}; p-1, n_1-1, n_2)$. Repeating the argument above on this density function then shows that $t_{22}^2 \sim \mathcal{B}(n_1 - 1, n_2)$ and is independent of $t_{33}, \ldots, t_{pp}$. The proof is completed in an obvious way by repetition of this argument.

For a more general case $n_1 \geq p$, [27, Lemma 3.9] is relevant (see also [32, 8b.2(xi)]) and can be easily extended to the complex matrix case, which is stated as follows. *Let U have the*

$\tilde{\mathcal{B}}_p(n_1, n_2)$ distribution with $n_1 \geq p$, and let $U^{[i]}$ be the matrix consisting of the first i rows and columns of U, with $\det(U^{[0]}) = 1$. Then $\vartheta_i = \det(U^{[i]})/\det(U^{[i-1]})$ is $\beta(n_1 - i + 1, n_2)$ and $\vartheta_1, \ldots, \vartheta_p$ are independent. By noticing that $U^{[i]} = (T^{[i]})^\dagger T^{[i]}$, where $T^{[i]}$ is defined as the matrix consisting of the first i rows and columns of T, we can easily see that $\vartheta_i = t_{ii}^2$. Therefore, we arrive at the desired result $t_{ii}^2 \sim \beta(n_1 - i + 1, n_2)$. □

References

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

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

[3] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

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

[5] E. N. Onggosanusi and A. G. Dabak, "A feedback-based adaptive multi-input multi-output signaling scheme," in *Proc. Asilomar Conf. 2002*, Pacific Grove, CA, Nov. 2002, pp. 1694–1698.

[6] R. S. Blum, "MIMO with limited feedback of channel state information," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 2003, pp. IV-89–IV-92.

[7] W. Santipach and M. L. Honig, "Asymptotic performance of MIMO wireless channels with limited feedback," in *Proc. IEEE MILCOM 2003*, Boston, MA, Oct. 2003, pp. 141–146.

[8] D. J. Love and R. Heath, Jr., "Limited feedback precoding for spatial multiplexing systems," in *Proc. IEEE GLOBECOM 2003*, San Francisco, CA, Dec. 2003, pp. 1857–1861.

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

[10] J. C. Roh and B. D. Rao, "An efficient feedback method for MIMO systems with slowly time-varying channels," in *Proc. IEEE Wireless Communications Networking Conf. (WCNC) 2004*, Atlanta, GA, Mar. 2004, pp. 805–809.

[11] G. Jöngren and M. Skoglund, "Quantized feedback information in orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2473–2482, Oct. 2004.

[12] J. C. Roh and B. D. Rao, "Channel feedback quantization methods for MISO and MIMO systems," in *Proc. IEEEEEE Int. Symp. Personal, Indoor, Mobile Radio Communications (PIMRC)*, Barcelona, Spain, Sep. 2004, pp. 760–764.

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

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

[15] D. J. Love and R. Heath, Jr., "Multi-mode precoding using linear receivers for limited feedback MIMO systems," in *Proc. IEEE Int. Conf. Communications (ICC)*, Paris, France, Jun. 2004, pp. 448–452.

[16] J. C. Roh and B. D. Rao, "Multiple antenna channels with partial channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 677–688, Mar. 2004.

[17] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84–95, Jan. 1980.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[20] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2450–2454, Sep. 2002.

[21] D. J. Love and R. Heath, Jr., "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 64–73, Jan. 2005.

[22] B. M. Hochwald, T. Marzetta, T. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1962–1973, Sep. 2000.

[23] R. J. Muirhead, *Aspects of Multivariate Statsitical Theory*. New York: Wiley, 1982.

[24] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152–177, 1963.

[25] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall/CRC Press, 2000.

[26] J. C. Roh, "Multiple-antenna communication systems with finite rate feedback," Ph.D. dissertation, Univ. of California at San Diego, San Diego, CA, Apr. 2005.

[27] S. K. Mitra, "A density-free approach to the matrix variate beta distribution," *Sannkhya*, vol. A32, pp. 81–88, 1970.

[28] A. M. Mathai and D. S. Tracy, "Random beta matrix and the distribution of its determinant," *Metron*, vol. 44, no. 1–4, pp. 101–110, 1986.

[29] J. C. Roh and B. D. Rao, "Performance analysis of multiple antenna systems with VQ-based feedback," in *Proc. Asilomar Conf. 2004*, Pacific Grove, CA, Nov. 2004, pp. 1978–1982.

[30] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed. New York: Academic, 1994.

[31] A. M. Kshirsagar, *Multivariate Analysis*. Berlin, Germany: Marcel Dekker, 1972.

[32] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.

June Chul Roh (S'03–M'06) received the B.S. and M.S. degrees in electronics engineering from the University of Seoul, Seoul, Korea, in 1993 and 1995, respectively, and the Ph.D. degree in electrical engineering from the University of California at San Diego, La Jolla, in 2005.

From 1995 to 2000, he was a Member of Technical Staff with the Korea Telecom R&D Group, Seoul, where he worked on third-generation cellular systems and was involved in developing a wide-band CDMA system. Since 2005, he has been a Systems Engineer with Texas Instruments, Inc., Sunnyvale, CA, where he is working on high-throughput wireless LAN systems. His research interests include communication theory, information theory, statistical signal processing, multiple-antenna systems, and CDMA and OFDM systems.

Bhaskar D. Rao (S'80–M'83–SM'91–F'00) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, in 1979 and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1981 and 1983, respectively.

Since 1983, he has been with the University of California at San Diego, La Jolla, where he is currently a Professor with the Electrical and Computer Engineering Department. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human–computer interactions.

Dr. Rao has been a member of the Statistical Signal and Array Processing Technical Committee of the IEEE Signal Processing Society. He is currently a member of the Signal Processing Theory and Methods Technical Committee.