Quantization Bounds on Grassmann Manifolds and Applications to MIMO Systems*

Wei Dai, Youjian Liu and Brian Rider
University of Colorado at Boulder
425 UCB, Boulder, CO, 80309, USA

Abstract

This paper considers the quantization problem on the Grassmann manifold with dimension n and p. The unique contribution is the derivation of a closed-form formula for the volume of a metric ball in the Grassmann manifold when the radius is sufficiently small. This volume formula holds for Grassmann manifolds with arbitrary dimension n and p, while previous results are only valid for either $p = 1$ or a fixed p with asymptotically large n. Based on the volume formula, the Gilbert-Varshamov and Hamming bounds for sphere packings are obtained. Assuming a uniformly distributed source and a distortion metric based on the squared chordal distance, tight lower and upper bounds are established for the distortion rate tradeoff. Simulation results match the derived results. As an application of the derived quantization bounds, the information rate of a Multiple-Input Multiple-Output (MIMO) system with finite-rate channel-state feedback is accurately quantified for arbitrary finite number of antennas, while previous results are only valid for either Multiple-Input Single-Output (MISO) systems or those with asymptotically large number of transmit antennas but fixed number of receive antennas.

Index Terms

the Grassmann manifold, distortion rate tradeoff, MIMO communications

I. Introduction

The *Grassmann manifold* $\mathcal{G}_{n,p}(\mathbb{L})$ is the set of all p-dimensional planes (through the origin) of the n-dimensional Euclidean space $\mathbb{L}^n$, where $\mathbb{L}$ is either $\mathbb{R}$ or $\mathbb{C}$. It forms a compact Riemann manifold of real dimension $\beta p(n-p)$, where $\beta = 1/2$ when $\mathbb{L} = \mathbb{R}/\mathbb{C}$ respectively. The Grassmann manifold provides a useful analysis tool for multi-antenna communications (also known as multiple-input multiple-output (MIMO) communication systems). For non-coherent MIMO systems, sphere packings on the $\mathcal{G}_{n,p}(\mathbb{L})$ can be viewed as a generalization of spherical codes [1]–[3]. For MIMO systems with partial channel state information at the transmitter (CSIT), which is obtained by finite-rate channel-state feedback, the quantization of beamforming matrices is related to the quantization on the Grassmann manifold [4]–[6].

The basic quantization problems addressed in this paper are the sphere packing bounds and distortion rate tradeoff. Roughly speaking, a quantization is a representation of a source in the $\mathcal{G}_{n,p}(\mathbb{L})$. In particular, it maps an element in the $\mathcal{G}_{n,p}(\mathbb{L})$ into a subset of the $\mathcal{G}_{n,p}(\mathbb{L})$, known as the code $\mathcal{C}$. Define the minimum distance of a code $\delta \triangleq \delta(\mathcal{C})$ as the minimum distance between any two codewords in a code $\mathcal{C}$. The sphere packing bound relates the size of a code and a given minimum distance δ. The rate distortion tradeoff is another important property of quantizations. A distortion metric is a mapping from the set of element pairs in the $\mathcal{G}_{n,p}(\mathbb{L})$ into the set of non-negative real numbers. Given a source distribution and a distortion metric, the rate distortion tradeoff is described by the minimum expected distortion achievable for a given code size or the minimum code size required to achieve a particular expected distortion.

There are several papers addressing the quantization problem in the Grassmann manifold. In [7], an isometric embedding of the $\mathcal{G}_{n,p}(\mathbb{R})$ into a sphere in Euclidean space $\mathbb{R}^{\frac{1}{2}(m-1)(m+2)}$ is given. Then using the Rankin bound in Euclidean space, the Rankin bound on the $\mathcal{G}_{n,p}(\mathbb{R})$ is obtained. However, this bound is not tight when the code size is large. Instead of resorting to some isometric embedding, sphere packing

*This work is partially supported by the Junior Faculty Development Award, University of Colorado at Boulder. It is Submitted to IEEE Transactions on Information Theory, Aug. 2005.

bounds can also be derived from analysis in the Grassmann manifold directly. Let $B(\delta)$ denote a metric ball of radius δ in the $\mathcal{G}_{n,p}(\mathbb{L})$. The sphere packing bounds can be derived from the volume of a $B(\delta)$ [3]. The exact volume formula for a $B(\delta)$ in the $\mathcal{G}_{n,p}(\mathbb{C})$ where $p=1$ is derived in [4]. An asymptotic volume formula for a $B(\delta)$ in the $\mathcal{G}_{n,p}(\mathbb{L})$, where $p\geq 1$ is fixed and n approaches infinity, is derived in [3]. Based on those volume formulas, the corresponding sphere packing bounds are developed in [3], [5]. Besides the sphere packing bounds, the rate distortion tradeoff is also treated in [8], where approximations to the distortion rate function are derived by the sphere packing bounds on the $\mathcal{G}_{n,p}(\mathbb{L})$. However, the derived approximations are based on the volume formulas in [3], [4] which are only valid for some special choices of n and p: either $p=1$ or fixed $p\geq 1$ with asymptotic large n.

This paper derives quantization bounds for the Grassmann manifold with arbitrary n and p when the code size is large. An explicit volume formula for a metric ball in the $\mathcal{G}_{n,p}(\mathbb{L})$ is derived when the radius is sufficiently small. It holds for Grassmann manifolds with arbitrary dimensions while previous results are only valid for either $p=1$ or a fixed p with asymptotically large n. Based on the derived volume formula, the Gilbert-Varshamov and Hamming bounds for sphere packings are obtained. For the rate distortion tradeoff, this paper starts from a new method employing optimization argument and extreme order statistics. Tight lower and upper bounds are established. Simulation results match the derived bounds. As an application of the derived quantization bounds, the information rate of a MIMO system with finite-rate channel-state feedback and power on/off strategy is accurately quantified for the first time. Since the corresponding Grassmann manifold for most practical MIMO systems has $p>1$ and small n, the quantization bounds derived in this paper are necessary.

The paper is organized as the follows. Section II provides some preliminaries on the Grassmann manifold. Section III derives the explicit volume formula for a metric ball in the Grassmann manifold. The corresponding sphere packing bounds and rate distortion tradeoff is accurately approximated in Section IV. An application of the quantization bounds to MIMO systems with finite-rate channel-state feedback is detailed in Section V. Section VI contains the conclusions.

II. PRELIMINARIES

This section presents a brief introduction to the Grassmann manifold. A metric and a measure on the Grassmann manifold are defined, and the problems relevant to quantization on the Grassmann manifold are formulated.

For the sake of applications [4]–[6], the projection Frobenius metric (*chordal distance*) is employed throughout the paper although the corresponding analysis is also applicable to the geodesic metric [3]. For any two planes $P,Q\in\mathcal{G}_{n,p}(\mathbb{L})$, we define the principle angles and the chordal distance between P and Q as follows. Let $\mathbf{u}_1\in P$ and $\mathbf{v}_1\in Q$ be the unit vectors such that $\left|\mathbf{u}_1^\dagger\mathbf{v}_1\right|$ is maximal. Inductively, let $\mathbf{u}_i\in P$ and $\mathbf{v}_i\in Q$ be the unit vectors such that $\mathbf{u}_i^\dagger\mathbf{u}_j=0$ and $\mathbf{v}_i^\dagger\mathbf{v}_j=0$ for all $1\leq j<i$ and $\left|\mathbf{u}_i^\dagger\mathbf{v}_i\right|$ is maximal. The principle angles are defined as $\theta_i=\arccos\left|\mathbf{u}_i^\dagger\mathbf{v}_i\right|$ for $i=1,\cdots,n$ [7], [9]. The chordal distance between P and Q is given by

$$d_c(P,Q)\triangleq\sqrt{\sum_{i=1}^{p}\sin^2\theta_i}.$$

The invariant measure on the $\mathcal{G}_{n,p}(\mathbb{L})$ is defined as follows. Let $O(n)/U(n)$ be the groups of $n\times n$ orthogonal/unitary matrices respectively. Let $\mathbf{A}\in O(n)/U(n)$ and $\mathbf{B}\in O(p)/U(n)$ when $\mathbb{L}=\mathbb{R}/\mathbb{C}$ respectively. An invariant measure μ on the $\mathcal{G}_{n,p}(\mathbb{L})$ satisfies, for any measurable set $\mathcal{M}\subset\mathcal{G}_{n,p}(\mathbb{L})$ and arbitrarily chosen $\mathbf{A}$ and $\mathbf{B}$,

$$\mu(\mathbf{A}\mathcal{M})=\mu(\mathcal{M})=\mu(\mathcal{M}\mathbf{B}).$$

The invariant measure defines the uniform distribution on the $\mathcal{G}_{n,p}(\mathbb{L})$ [9].

With a metric and a measure defined on the $\mathcal{G}_{n,p}(\mathbb{L})$, there are several bounds well known for sphere packings. Let a code $\mathcal{C}$ be a discrete subset of the $\mathcal{G}_{n,p}(\mathbb{L})$, δ be the minimum distance between any two codewords of a code $\mathcal{C}$ and $B(\delta)$ be the metric ball of radius δ in the $\mathcal{G}_{n,p}(\mathbb{L})$. If K is any positive integer such that $K\mu(B(\delta)) < 1$, then there exists a code $\mathcal{C}$ of size $K+1$ with minimum distance δ. This principle is called as the *Gilbert-Varshamov* lower bound [3], i.e.,

$$|\mathcal{C}| > \frac{1}{\mu(B(\delta))}. \tag{1}$$

On the other hand, $|\mathcal{C}|\,\mu(B(\delta/2)) \leq 1$ for any code $\mathcal{C}$. The *Hamming* upper bound captures this fact as [3]

$$|\mathcal{C}| \leq \frac{1}{\mu(B(\delta/2))}. \tag{2}$$

These two bounds relate the size of a code and a given minimum distance δ.

Distortion rate tradeoff gives another important property of quantization. A quantization is a mapping from the $\mathcal{G}_{n,p}(\mathbb{L})$ to a code $\mathcal{C}$, i.e.

$$q : \mathcal{G}_{n,p}(\mathbb{L}) \to \mathcal{C}.$$

A distortion metric is a mapping

$$d : \mathcal{G}_{n,p}(\mathbb{L}) \times \mathcal{C} \to [0, +\infty)$$

from the set of the element pairs in the $\mathcal{G}_{n,p}(\mathbb{L})$ and $\mathcal{C}$ into the set of non-negative real numbers. Assume that a source Q is randomly distributed in the $\mathcal{G}_{n,p}(\mathbb{L})$. The distortion associated with a quantization q is defined as

$$D \triangleq \mathrm{E}\left[d(Q, q(Q))\right].$$

The rate distortion tradeoff can be described by the infimum achievable distortion given a code size, which is called distortion rate function, or the infimum code size required to achieve a particular distortion, which is called rate distortion function. In this paper, the source Q is assumed to be uniformly distributed in the $\mathcal{G}_{n,p}(\mathbb{L})$. Define the distortion metric as the square of the chordal distance. For a given code $\mathcal{C} \subset \mathcal{G}_{n,p}(\mathbb{L})$, the optimal quantization to minimize the distortion is given by[1]

$$q(Q) = \arg\min_{P\in\mathcal{C}} d_c(P, Q).$$

The distortion associated with this quantization is

$$D(\mathcal{C}) = \mathrm{E}\left[\min_{P\in\mathcal{C}} d_c^2(P, Q)\right].$$

For a given code size K where K is a positive integer, the distortion rate function is then given by[2]

$$D^*(K) = \inf_{\mathcal{C}:|\mathcal{C}|=K} D(\mathcal{C}).$$

The rate distortion function is given by

$$K^*(D) = \inf_{D(\mathcal{C})\leq D} |\mathcal{C}|.$$

[1]The ties, i.e. the case that $\exists P_1, P_2 \in \mathcal{C}$ such that $d_c(P_1, Q) = \min_{P\in\mathcal{C}} d_c(P, Q) = d_c(P_2, Q)$, are broken arbitrarily because the probability of ties is zero.

[2]The standard definition of the distortion rate function is a function of the code rate defined by $\log_2 K$. The definition in this paper is equivalent to the standard one.

III. Metric Balls in the Grassmann Manifold

In this section, an explicit volume formula for a metric ball $B(\delta)$ in the $\mathcal{G}_{n,p}(\mathbb{L})$ is derived. The volume formula is essential for the quantization bounds in Section IV.

The volume calculation depends on the relationship between the measure and the metric defined on the Grassmann manifold. For the invariant measure μ and the chordal distance d_c, the volume of a metric ball $B(\delta)$ can be calculated by

$$\mu(B(\delta)) = \int \cdots \int_{\substack{\sqrt{\sum_{i=1}^{p} \sin^2\theta_i} \leq \delta \\ \frac{\pi}{2} \geq \theta_1 \geq \cdots \geq \theta_p \geq 0}} d\mu_\theta, \tag{3}$$

where $\theta_1, \cdots, \theta_p$ are the principle angles and the differential form $d\mu_\theta$ is given in [9]–[11] and Appendix A below.

The following theorem expresses the volume formula as an exponentiation of the radius δ.

Theorem 1: Let $B(\delta)$ be a ball of radius δ in the $\mathcal{G}_{n,p}(\mathbb{L})$. When $\delta \leq 1$,

$$\mu(B(\delta)) = \begin{cases} c_{n,p,\beta}\delta^{\beta p(n-p)}(1 + o(\delta)) & \text{if } \mathbb{L} = \mathbb{R} \\ c_{n,p,\beta}\delta^{\beta p(n-p)} & \text{if } \mathbb{L} = \mathbb{C} \end{cases}, \tag{4}$$

where $\beta = 1/2$ when $\mathbb{L} = \mathbb{R}/\mathbb{C}$ respectively and $c_{n,p,\beta}$ is a constant depending on n, p and β. When $\mathbb{L} = \mathbb{C}$, $c_{n,p,2}$ can be explicitly calculated

$$c_{n,p,2} = \begin{cases} \frac{1}{(np-p^2)!}\prod_{i=1}^{p}\frac{(n-i)!}{(p-i)!} & \text{if } 0 < p \leq \frac{n}{2} \\ \frac{1}{(np-p^2)!}\prod_{i=1}^{n-p}\frac{(n-i)!}{(n-p-i)!} & \text{if } \frac{n}{2} \leq p \leq n \end{cases}. \tag{5}$$

When $\mathbb{L} = \mathbb{R}$, $c_{n,p,1}$ is given by

$$c_{n,p,1} = \begin{cases} \frac{V_{n,p,1}}{2^p}\int \cdots \int_{\substack{\sum_{i=1}^{p} x_i \leq 1 \\ x_1 \geq \cdots \geq x_p \geq 0}} \left[\left|\prod_{i<j}^{p}(x_i - x_j)\right| \prod_{i=1}^{p}\left(x_i^{\frac{1}{2}(n-2p+1)-1}dx_i\right)\right] & \text{if } 0 < p \leq \frac{n}{2} \\ \frac{V_{n,n-p,1}}{2^{n-p}}\int \cdots \int_{\substack{\sum_{i=1}^{n-p} x_i \leq 1 \\ x_1 \geq \cdots \geq x_{n-p} \geq 0}} \left[\left|\prod_{i<j}^{p}(x_i - x_j)\right| \prod_{i=1}^{n-p}\left(x_i^{\frac{1}{2}(2p-n+1)-1}dx_i\right)\right] & \text{if } \frac{n}{2} \leq p \leq n \end{cases}, \tag{6}$$

where

$$V_{n,p,1} = \prod_{i=1}^{p}\frac{A^2(p-i+1)\,A(n-p-i+1)}{2A(n-i+1)}$$

and

$$A(p) = \frac{2\pi^{p/2}}{\Gamma\left(\frac{p}{2}\right)}.$$

Proof: See Appendix A. ■

Theorem 1 provides an explicit volume approximation for real Grassmann manifolds and an exact volume formula for complex Grassmann manifolds when $\delta \leq 1$. Although (4) is derived for $\delta \leq 1$, simulations show that this approximation remains good for relatively large δ (Fig. 1).

Theorem 1 is of course consistent with the previous results of [4] and [3], which pertain to special choices of n and p and are stated as follows.

Example 1: Consider the volume formula for a $B(\delta)$ in the $\mathcal{G}_{n,p}(\mathbb{C})$ where $p = 1$. Without normalization, the total volume of $\mathcal{G}_{n,1}(\mathbb{C})$ is $2\pi^n/(n-1)!$ and the volume of the $B(\delta)$ is $2\pi^n\delta^{2(n-1)}/(n-1)!$ [4]. Therefore,

$$\mu(B(\delta)) = \delta^{2(n-1)},$$

agreeing with Theorem 1 where $\beta = 2$ and $c_{n,1,2} = 1$.

Example 2: For the $\mathcal{G}_{n,p}(\mathbb{L})$ where p is fixed and $n \to +\infty$, an asymptotic volume formula for a $B(\delta)$ is derived in [3], which reads

$$\mu(B(\delta)) = \left(\frac{\delta}{\sqrt{p}}\right)^{\beta np + o(n)}. \tag{7}$$

On the other hand, Theorem 1 contains an asymptotic formula for $\mathbb{L} = \mathbb{C}$, $\delta \le 1$, fixed p and asymptotically large n in the form

$$\mu(B(\delta)) = \left(\frac{\delta}{\sqrt{p}}\right)^{2p(n-p)+\frac{1}{2}\left(p^2-1\right)\ln n/\ln\left(\delta/\sqrt{p}\right)+\ln c_p/\ln\left(\delta/\sqrt{p}\right)},$$

where c_p is a constant depending on p. This follows from (4) and Stirling approximation.

Importantly though, Theorem 1 is distinct from the previous results of [4] and [3] in that it holds for arbitrary n and p, $1 \le p \le n$. For the Grassmann manifold whose n is not asymptotically large or n and p are comparable, it is not appropriate to use (7) to estimate the volume of a $B(\delta)$. A trivial example is that the $n = p$ case. If $n = p$, the exact volume of $B(\delta)$ for $\forall \delta > 0$ is the constant 1. The volume formula in this paper (4) gives $c_{n,n,\beta}\delta^0 = 1$ because $c_{n,n,\beta} = 1$. However, the approximation $(\delta/\sqrt{n})^{\beta n^2}$ (from Barg's formula in (7)) will give a small number much less than 1 when δ is small.

Theorem 1 does not provide a simple formula to calculate $c_{n,p,\beta}$ with $\beta = 1 (\mathbb{L} = \mathbb{R})$ for general n and p. Although it contains

$$c_{n,p,1} = \begin{cases} \frac{1}{n-1}\frac{\pi^{1/2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^2\Gamma\left(\frac{n-1}{2}\right)} & \text{if } p = 1 \text{ or } p = n-1 \\ 1 & \text{if } p = n \end{cases},$$

the calculation of $c_{n,p,1}$ involves a complicate integral (6) for general n and p. As a supplement of Theorem 1, the following proposition considers fixed p and asymptotically large n. This proposition is an improvement of Barg's approximation in that it gives a more detailed exponential term. Moreover, the proof is also fundamental different.

Proposition 1: Let $B(\delta)$ be a ball of radius δ in the $\mathcal{G}_{n,p}(\mathbb{L})$. Let $p > 0$ be fixed and n approach infinity. Then

$$\mu(B(\delta)) \approx \frac{\Gamma\left(\frac{\beta}{2}np\right)}{\Gamma\left(\frac{\beta}{2}p(n-p)\right)\Gamma\left(\frac{\beta}{2}p^2\right)}\left(\frac{\delta^2}{p}\right)^{\frac{\beta}{2}p(n-p)}\left(1-\frac{\delta^2}{p}\right)^{\frac{\beta}{2}p^2-1},$$

where $\beta = 1/2$ when $\mathbb{L} = \mathbb{R}/\mathbb{C}$ respectively.

Proof: See Appendix B. ■

Fig. 1 compares the exact volume of a metric ball (3) and the volume approximated by (4). To calculate the exact volume of a metric ball, Monte Carlo simulations are employed to evaluate the complicate integrals in (3). Since

$$\mu(B(\delta)) = \Pr\{Q : d_c(P,Q) \le \delta\}$$

where $P \in \mathcal{G}_{n,p}(\mathbb{L})$ is chosen arbitrarily and Q is uniformly distributed in the $\mathcal{G}_{n,p}(\mathbb{L})$, simulation for the probability of the event $\{Q : d_c(P,Q) \le \delta\}$ gives $\mu(B(\delta))$. For the volume approximation $c_{n,p,\beta}\delta^{\beta p(n-p)}$, the constant $c_{n,p,\beta}$ is calculated either by (5) if $\mathbb{L} = \mathbb{C}$ or by Monte Carlo numerical integral of (6) if $\mathbb{L} = \mathbb{R}$. The simulation results for the real and complex Grassmann manifolds are presented in Fig. 1(a) and 1(b) respectively. Simulations show that the volume approximation (solid lines) is close to the exact volume (circles) when the radius of the metric ball is not large. We also compare our approximation with Barg's approximation $\left(\delta/\sqrt{p}\right)^{\beta np}$ for the case $n = 10$ and $p = 2$. Simulations show that the exact volume and Barg's approximation (dash-dot lines) may not be of the same order while the approximation in this paper is more accurate.

IV. Quantization Bounds

Based on the volume formula given in Theorem 1, the sphere packing bounds are derived and the rate distortion tradeoff is characterized by establishing tight lower and upper bounds on the distortion rate function. The results developed in this section hold for Grassmann manifolds with arbitrary n and p.

A. Sphere Packing Bounds

The Gilbert-Varshamov and Hamming bounds on the $\mathcal{G}_{n,p}(\mathbb{L})$ are given in the following corollary.

Corollary 1: When δ is sufficiently small, there exists a code in the $\mathcal{G}_{n,p}(\mathbb{L})$ with size K and the minimum distance δ such that

$$c_{n,p,\beta}^{-1}\delta^{-\beta p(n-p)} \lesssim K.$$

For any code with the minimum distance δ,

$$K \lesssim c_{n,p,\beta}^{-1}\left(\frac{\delta}{2}\right)^{-\beta p(n-p)}.$$

Here and throughout, the symbol $\lesssim$ indicates that the inequality holds up to $(1+o(1))$ error.

Proof: The corollary is proved by substituting the volume approximation (4) into (1) and (2). ■

B. Rate Distortion Tradeoff

Assume that the source is uniformly distributed in the Grassmann manifold and the distortion metric is defined as the square of the chordal distance. The distortion rate tradeoff is characterized in this section.

The following theorem gives a lower bound on the distortion rate function.

Theorem 2: Let $t = \beta p(n-p)$ be the number of the real dimensions of the Grassmann manifold $\mathcal{G}_{n,p}(\mathbb{L})$. When K is sufficient large, the distortion rate function is lower bounded by

$$\frac{t}{t+2}\left(c_{n,p,\beta}K\right)^{-\frac{2}{t}} \lesssim D^{*}(K).$$

Proof: See Appendix C. ■

The proof of this theorem is given by an optimization argument. The key is to construct an ideal quantizer, which may not exist, to minimize the distortion. Suppose that there exists K metric balls of the same radius x_0 covering the whole $\mathcal{G}_{n,p}(\mathbb{L})$ completely without any overlap. Then the quantizer which maps each of those balls into its center gives the minimum distortion among all quantizers. Of course such a covering may not exist, providing a lower bound on the distortion rate function.

An upper bound on the distortion rate function is given in the following theorem.

Theorem 3: Let $t = \beta p(n-p)$ be the number of the real dimensions of the Grassmann manifold $\mathcal{G}_{n,p}(\mathbb{L})$. Define a random code with size K as $\mathcal{C}_{\text{rand}} = \{P_1, \cdots, P_K\}$ where P_i's are independently drawn from the uniform distribution on the $\mathcal{G}_{n,p}(\mathbb{L})$. Then

$$\lim_{K\to+\infty} \mathrm{E}\left[K^{\frac{t}{2}} \cdot D\left(\mathcal{C}_{\text{rand}}\right)\right] = \frac{2\Gamma\left(\frac{2}{t}\right)}{t} c_{n,p,\beta}^{-\frac{2}{t}},$$

where the expectation is on the ensemble of random codes. Thus when K is sufficiently large, the distortion rate function can be upper bounded by

$$D^{*}(K) \le \mathrm{E}\left[D\left(\mathcal{C}_{\text{rand}}\right)\right] \approx \frac{2\Gamma\left(\frac{2}{t}\right)}{t}\left(c_{n,p,\beta}K\right)^{-\frac{2}{t}}.$$

Proof: See Appendix D. ■

The basic idea behind this theorem is that the distortion of any particular code is an upper bound of the distortion rate function and so is the average distortion of an ensemble of codes. Towards the proof, the ensemble of random codes $\mathcal{C}_{\text{rand}} = \{P_1, \cdots, P_K\}$ are employed. For any given plane $Q \in \mathcal{G}_{n,p}(\mathbb{L})$, define $X_i = d_c^2(P_i, Q)$ and $W_K \triangleq \min(X_1, \cdots, X_K) = \min\limits_{P_i \in \mathcal{C}_{\text{rand}}} d_c^2(P_i, Q)$. Since the codewords P_i's $1 \le i \le K$

are independently drawn from the uniform distribution on the $\mathcal{G}_{n,p}(\mathbb{L})$, X_i's $1 \le i \le K$ are independent and identically distributed (*i.i.d.*) random variables with the cumulative distribution function (CDF) given by Theorem 1. According to X_i's CDF, the CDF of W_K can be calculated by extreme order statistics. In appendix D, we prove that for any given $Q \in \mathcal{G}_{n,p}(\mathbb{L})$, $K^{\frac{t}{2}} \cdot \mathrm{E}_{W_K}[W_K]$ converges to a constant as K approaches infinity. Thus, $K^{\frac{t}{2}} \cdot \mathrm{E}_Q[\mathrm{E}_{W_K}[W_K]] = K^{\frac{t}{2}} \cdot \mathrm{E}_{\mathcal{C}_{\mathrm{rand}}}[D(\mathcal{C}_{\mathrm{rand}})]$ converges to the same constant. In this way, an upper bound of the distortion rate function is obtained for asymptotically large K.

The difference between the lower bound and the upper bound is studied as follows. Since both bounds have the same exponential term, we focus on the coefficients before the exponential terms. The difference between the two bounds is depending on the number of real dimensions $t = \beta p(n-p)$ of the underlying Grassmann manifold. There are three cases needed to consider.

Case 1: $t = 0$. This case happens if and only if $n = p$. For this case, the whole $\mathcal{G}_{n,n}(\mathbb{L})$ contains only one element and no quantization is needed essentially.

Case 2: $t = 1$. This case happens if and only if $\mathbb{L} = \mathbb{R}$, $n = 2$ and $p = 1$. In this case, it can be verified that the principle angle θ between a uniformly distributed $Q \in \mathcal{G}_{2,1}(\mathbb{R})$ and any fixed $P \in \mathcal{G}_{2,1}(\mathbb{R})$ is uniformly distributed in $\left[0, \frac{\pi}{2}\right]$. The optimal quantization can be explicitly constructed. Since there exists K metric balls with radius $\sin\frac{\pi}{4K}$ such that those balls not only pack but also cover the whole $\mathcal{G}_{2,1}(\mathbb{R})$, the quantizer mapping those balls into its center is optimal. The distortion rate function can be explicitly calculated as

$$D^*(K) = \frac{1}{2K} - \frac{1}{\pi}\sin\frac{\pi}{2K}.$$

Case 3: $t \ge 2$. For this general case, an elementary calculation shows that

$$\frac{1}{2} \le \frac{t}{t+2} \le \frac{2}{t}\Gamma(2/t) \le 1,$$

and

$$\lim_{t\to+\infty} \frac{t}{t+2} = 1 = \lim_{t\to+\infty} \frac{2}{t}\Gamma(2/t).$$

That is, the difference between the upper and lower bounds is uniformly bounded and converges to zero as $t \to +\infty$. This phenomenon is also observed from the simulation results in Fig. 2. Since the upper bound and the lower bound are close, combining these two bounds accurately quantifies the distortion rate function, as stated in the following corollary.

Corollary 2: Let $t = \beta p(n-p)$ be the number of the real dimensions of the Grassmann manifold $\mathcal{G}_{n,p}(\mathbb{L})$. When K is sufficiently large, the distortion rate function is bounded by

$$\frac{t}{t+2}\left(c_{n,p,\beta}K\right)^{-\frac{2}{t}} \lesssim D^*(K) \lesssim \frac{2\Gamma\left(\frac{2}{t}\right)}{t}\left(c_{n,p,\beta}K\right)^{-\frac{2}{t}}. \tag{8}$$

The following corollary bounds the rate distortion function.

Corollary 3: Let $t = \beta p(n-p)$ be the number of the real dimensions of the Grassmann manifold $\mathcal{G}_{n,p}(\mathbb{L})$. When the required distortion is sufficiently small, the rate distortion function is bounded by

$$\frac{1}{c_{n,p,\beta}}\left(\frac{t}{2\Gamma\left(\frac{2}{t}\right)}D\right)^{\frac{t}{2}} \lesssim K^*(D) \lesssim \frac{1}{c_{n,p,\beta}}\left(\frac{t+2}{t}D\right)^{\frac{t}{2}}.$$

As a comparison, we cite the distortion rate function approximation derived in [8]. For $\mathcal{G}_{n,p}(\mathbb{C})$ with $p = 1$, an approximation for the distortion rate function is given as

$$D^*(K) \approx \left(\frac{n-1}{n}\right) K^{-\frac{1}{n-1}} \tag{9}$$

by asymptotic arguments [8]. According to our results in Theorem 2, the approximation (9) is indeed a lower bound for the distortion rate function and valid for all possible n's. For $\mathcal{G}_{n,p}(\mathbb{C})$ with p fixed and $n \gg p$, a lower bound of an upper bound on the distortion rate function is given in [8] based on an estimation of the minimum distance of a code. It is less robust than the result in Theorem 2 in that it is neither a lower bound nor an upper bound and that it holds only for the case $n \gg p$ (see Fig. 2 for an empirical comparison).

Besides characterizing the rate distortion tradeoff, we are also interested in designing a code to minimize distortion for a given code size K. Generally speaking, it is computational complex to design a code to minimize distortion directly. In [5] and [12], a suboptimal design criterion, i.e. maximization the minimum distance between codeword pairs, is proposed to reduce the computational complexity. Refer this suboptimal criterion as *max-min* criterion. According to our volume formula (4), the same criterion can be verified. Let the minimum distance of a code $\mathcal{C}$ be δ. Note that the metric balls of radius $\frac{\delta}{2}$ and centered at $P_i \in \mathcal{C}$ are disjoint. Then the corresponding distortion is upper bounded by

$$D(\mathcal{C}) \leq \frac{\delta^2}{4} K\mu\left(B\left(\delta/2\right)\right) + p\left(1 - K\mu\left(B\left(\delta/2\right)\right)\right). \quad (10)$$

Apply the volume formula (4). An elementary calculation shows that the first derivative of the upper bound is negative when

$$\delta < \sqrt{\frac{4\beta p^2 (n-p)}{2 + \beta p (n-p)}}.$$

This property implies the upper bound (10) is a decreasing function of δ when δ is small enough. Thus, max-min criterion is an appropriate design criterion to obtain codes with small distortion. Since this criterion only requires to calculate the distance between codeword pairs, the computational complexity is less than that of designing a code to minimize the distortion directly.

Fig. 2 compares the simulated distortion rate function (the plus markers) with its lower bound (the dashed lines) and upper bound (the solid lines) in (8). To simulate the distortion rate function, we use the max-min criterion to design codes and use the minimum distortion of the designed codes as the distortion rate function. Simulation results show that the bounds in (8) hold for large K. When K is relatively small, the formula (8) can serve as good approximations to the distortion rate function as well. Simulations also verify the previous discussion on the difference between the two bounds. The difference between the bounds is small and it becomes smaller as the number of the real dimensions of the Grassmann manifold increases. In addition, we compare our bounds with the approximation (the "x" markers) derived in [8]. Simulations show that the approximation in [8] is neither an upper bound nor a lower bound. It works for the case that $n = 10$ and $p = 2$ but doesn't work when $n \leq 8$ and $p = 2$. As a comparison, the bounds (8) derived in this paper hold for arbitrary n and p.

V. An Application to MIMO Systems with Finite Rate Channel State Feedback

As an application of the derived quantization bounds on the Grassmann manifold, this section discusses the information theoretical benefit of finite-rate channel-state feedback for MIMO systems using power on/off strategy. We will show that the benefit of the channel state feedback can be accurately characterized by the distortion of a quantization on the Grassmann manifold.

The effect of finite-rate feedback on MIMO systems using power on/off strategy has been widely studied. MIMO systems with only one on-beam are discussed in [4] and [5], where the beamforming codebook design criterion and performance analysis are derived by geometric arguments in the Grassmann manifold $\mathcal{G}_{n,1}(\mathbb{C})$. MIMO systems with multiple on-beams are considered in [8], [13]–[16]. Criteria to select the beamforming matrix are developed in [13] and [14]. The signal-to-noise ratio (SNR) loss due to quantized beamforming is discussed in [8]. The corresponding analysis is based on Barg's formula (7) and only valid for MIMO systems with asymptotically large number of transmit antennas. The effect of beamforming quantization on information rate is investigated in [15] and [16]. The loss in information rate is quantified for high SNR region in [15]. The analysis is based on an approximation to the logdet function in the high SNR region and a metric in the Grassmann manifold other than the chordal distance. In [16], a formula to calculate the information rate for all SNR regimes is proposed by letting the numbers of transmit antennas, receive antennas and feedback rate approach infinity simultaneously. But this formula overestimates the performance in general.

The system model of a wireless communication system with L_T transmit antennas, L_R receive antennas and finite-rate channel state feedback is given in Fig. 3. The information bit stream is encoded into the Gaussian signal vector $\mathbf{X} \in \mathbb{C}^{s\times 1}$ and then multiplied by the beamforming matrix $\mathbf{P} \in \mathbb{C}^{L_T\times s}$ to generate the transmitted signal $\mathbf{T} = \mathbf{PX}$, where s is the dimension of the signal $\mathbf{X}$ satisfying $1 \leq s \leq L_T$ and the beamforming matrix $\mathbf{P}$ satisfies $\mathbf{P}^\dagger\mathbf{P} = \mathbf{I}_s$. In power on/off strategy, $\mathrm{E}\left[\mathbf{XX}^\dagger\right] = P_{\text{on}}\mathbf{I}_s$ where P_{on} is a positive constant to denote the on-power. Assume that the channel $\mathbf{H}$ is Rayleigh flat fading, i.e., the entries of $\mathbf{H}$ are independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian variables with zero mean and unit variance ($\mathcal{CN}(0,1)$) and $\mathbf{H}$ is i.i.d. for each channel use. Let $\mathbf{Y} \in \mathbb{C}^{L_R\times 1}$ be the received signal and $\mathbf{W} \in \mathbb{C}^{L_R\times 1}$ be the Gaussian noise, then

$$\mathbf{Y} = \mathbf{HPX} + \mathbf{W},$$

where $E\left[\mathbf{WW}^\dagger\right] = \mathbf{I}_{L_R}$. We also assume that there is a beamforming codebook $\mathcal{B} = \left\{\mathbf{P}_i \in \mathbb{C}^{L_T\times s} : \mathbf{P}_i^\dagger\mathbf{P}_i = \mathbf{I}_s\right\}$ declared to both the transmitter and the receiver before the transmission. At the beginning of each channel use, the channel state $\mathbf{H}$ is perfectly estimated at the receiver. A message, which is a function of the channel state, is sent back to the transmitter through a feedback channel. The feedback is error-free and rate limited. According to the channel state feedback, the transmitter chooses an appropriate beamforming matrix $\mathbf{P}_i \in \mathcal{B}$. Let the feedback rate be R_{fb}bits/channel use. Then the size of the beamforming codebook $|\mathcal{B}| \leq 2^{R_{\text{fb}}}$. The feedback function is a mapping from the set of channel state into the beamforming matrix index set, $\varphi: \{\mathbf{H}\} \to \{i : 1 \leq i \leq |\mathcal{B}|\}$. This section will quantify the corresponding information rate

$$\mathcal{I} = \max_{\mathcal{B}:|\mathcal{B}|\leq 2^{R_{\text{fb}}}} \max_{\varphi} \mathrm{E}\left[\log\left|\mathbf{I}_{L_R} + P_{\text{on}}\mathbf{HP}_{\varphi(\mathbf{H})}\mathbf{P}_{\varphi(\mathbf{H})}^\dagger\mathbf{H}\right|\right],$$

where $P_{\text{on}} = \rho/s$ and ρ is the average received SNR.

Before discussing the finite-rate feedback case, we consider the case that the transmitter has full knowledge of the channel state $\mathbf{H}$. In this setting, the optimal beamforming matrix is given by $\mathbf{P}_{\text{opt}} = \mathbf{V}_s$ where $\mathbf{V}_s \in \mathbb{C}^{L_T\times s}$ is the matrix composed by the right singular vectors of $\mathbf{H}$ corresponding to the largest s singular values [6]. The corresponding information rate is

$$\mathcal{I}_{\text{opt}} = \mathrm{E}_{\mathbf{H}}\left[\sum_{i=1}^{s}\ln\left(1 + P_{\text{on}}\lambda_i\right)\right], \tag{11}$$

where λ_i is the i^{th} largest eigenvalue of $\mathbf{HH}^\dagger$. In [6, Section III], we derive an asymptotic formula to approximate a quantity of the form $\mathrm{E}_{\mathbf{H}}\left[\sum_{i=1}^{s}\ln\left(1 + c\lambda_i\right)\right]$ where $c > 0$ is a constant. Apply the asymptotic formula in [6]. $\mathcal{I}_{\text{opt}}$ can be well approximated.

The effect of finite-rate feedback can be characterized by the quantization bounds in the Grassmann manifold. For finite-rate feedback, we define a suboptimal feedback function

$$i = \varphi(\mathbf{H}) \triangleq \arg\min_{1\leq i\leq|\mathcal{B}|} d_c^2\left(\mathcal{P}(\mathbf{P}_i), \mathcal{P}(\mathbf{V}_s)\right), \tag{12}$$

where $\mathcal{P}(\mathbf{P}_i)$ and $\mathcal{P}(\mathbf{V}_s)$ are the planes in the $\mathcal{G}_{L_T,s}(\mathbb{C})$ generated by $\mathbf{P}_i$ and $\mathbf{V}_s$ respectively. In [6], we show that this feedback function is asymptotically optimal as $R_{\text{fb}} \to +\infty$ and near optimal when $R_{\text{fb}} < +\infty$. With this feedback function and assuming that the feedback rate R_{fb} is large, it has been shown in [6] that

$$\mathcal{I} \approx \mathrm{E}_{\mathbf{H}}\left[\sum_{i=1}^{s}\ln\left(1 + \eta_{\text{sup}}P_{\text{on}}\lambda_i\right)\right], \tag{13}$$

where

$$\begin{aligned}\eta_{\text{sup}} &\triangleq 1-\frac{1}{s}\inf_{\mathcal{B}:|\mathcal{B}|\leq 2^{R_{\text{fb}}}} \mathrm{E}_{\mathbf{V}_s}\left[\min_{1\leq i\leq|\mathcal{B}|} d_c^2\left(\mathcal{P}\left(\mathbf{P}_i\right),\mathcal{P}\left(\mathbf{V}_s\right)\right)\right]\\ &= 1-\frac{1}{s}D^*\left(2^{R_{\text{fb}}}\right). \end{aligned} \tag{14}$$

Thus, the difference between perfect beamforming case (11) and finite-rate feedback case (13) is quantified by η_{sup}, which depends on the distortion rate function on the $\mathcal{G}_{L_T,s}\left(\mathbb{C}\right)$. Substitute quantization bounds (8) into (14) and apply the asymptotic formula in [6] for $\mathrm{E}_{\mathbf{H}}\left[\sum_{i=1}^{s}\ln\left(1+c\lambda_i\right)\right]$. Approximations to the information rate $\mathcal{I}$ are derived as functions of the feedback rate R_{fb}.

Simulations verify the above approximations. Let $m=\min\left(L_T,L_R\right)$. Fig. 4 compares the simulated information rate (circles) and approximations as functions of R_{fb}/m^2. The information rate approximated by the lower bound (solid lines) and the upper bound (dotted lines) in (8) are presented. The simulation results show that the performances approximated by the bounds (8) match the actual performance almost perfectly. As a comparison, the approximation proposed in [16], [17], which is based on asymptotic analysis and Gaussian approximation, overestimates the information rate. Furthermore, we compare the simulated information rate and the approximations for a large range of SNRs in Fig. 5. Without loss of generality, we only present the lower bound in (8) because it corresponds to the random codes and can be achieved by appropriate code design. Fig. 5(a) shows that the difference between the simulated and approximated information rate is almost unnoticeable. To make the performance difference clearer, Fig. 5(b) gives the relative performance as the ratio of the considered performance and the capacity of a 4×2 MIMO achieved by water filling power control. The difference in relative performance is also small for all SNR regimes.

VI. Conclusion

This paper considers the quantization problem on the Grassmann manifold. Based on the explicit volume formula for a metric ball in the $\mathcal{G}_{n,p}\left(\mathbb{L}\right)$, the sphere packing bounds are obtained and the distortion rate tradeoff is accurately characterized by establishing bounds on the distortion function. Simulations verify the developed results. As an application of the derived quantization bounds, the information rate of a MIMO system with finite-rate channel-state feedback and power on/off strategy is accurately quantified for the first time.

Appendix

A. Proof of Theorem 1

The proof is divided into two parts. In the first part, we show that when $\delta\leq 1$,

$$\mu\left(B\left(\delta\right)\right)=\begin{cases} c_{n,p,\beta}\delta^{\beta p(n-p)}\left(1+o\left(\delta\right)\right) & \text{if } \mathbb{L}=\mathbb{R}\\ c_{n,p,\beta}\delta^{\beta p(n-p)} & \text{if } \mathbb{L}=\mathbb{C}\end{cases},$$

and give the general formula to calculate $c_{n,p,\beta}$. In the second part, we simplify the formula to calculate $c_{n,p,2}$.

1) General Volume Formula: First, we derive the general formula for the volume of a metric ball in the Grassmann manifold.

Since it holds

$$\mu\left(B\left(\delta\right)\right)=\underset{\substack{\sqrt{\sum_{i=1}^{p}\sin^2\theta_i}\leq\delta\\ \frac{\pi}{2}\geq\theta_1\geq\cdots\geq\theta_p\geq 0}}{\int\cdots\int} d\mu_\theta,$$

we wish to calculate the above integral.

We introduce the following notations. Define $r_i \triangleq \cos\theta_i$ and order r_i's such that $r_i \le r_j$ ($\theta_i \ge \theta_j$) if $i<j$. Define $\mathbf{r} = [r_1, \cdots, r_p]$ and

$$\left|\Delta_p\left(\mathbf{r}^2\right)\right| = \prod_{i<j}^{p}\left(r_j^2 - r_i^2\right).$$

Let $\beta = 1$ for $\mathbb{L} = \mathbb{R}$ and $\beta = 2$ when $\mathbb{L} = \mathbb{C}$. Define

$$V_{n,p,\beta} = \begin{cases} \prod_{i=1}^{p} \frac{A^2(p-i+1)A(n-p-i+1)}{2A(n-i+1)} & \text{if } \beta = 1 \\ \prod_{i=1}^{p} \frac{2(n-i)!}{((p-i)!)^2(n-p-i)!} & \text{if } \beta = 2 \end{cases}$$

where

$$A\left(p\right) = \frac{2\pi^{p/2}}{\Gamma\left(\frac{p}{2}\right)}.$$

When $1 \le p \le \frac{n}{2}$, the distribution of the principle angles $\theta_1, \theta_2, \cdots, \theta_p$ can be expressed as the following differential form [11],

$$\begin{aligned} d\mu_\theta \; = \; & \frac{1}{2^p} V_{n,p,\beta} \left|\Delta_p\left(\mathbf{r}^2\right)\right|^\beta \\ & \prod_{i=1}^{p}\left(\left(r_i^2\right)^{\frac{\beta}{2}-1}\left(1-r_i^2\right)^{\frac{\beta}{2}(n-2p+1)-1} dr_i^2\right) \end{aligned}$$

Thus

$$\begin{aligned} & \mu\left(B\left(\delta\right)\right) \\ & = \frac{1}{2^p} \underset{\substack{\sum_{i=1}^{p}\left(1-r_i^2\right)\le\delta^2 \\ 0\le r_1^2\le\cdots\le r_p^2\le 1}}{\int\cdots\int} \left[V_{n,p,\beta}\prod_{i<j}^{p}\left(r_j^2-r_i^2\right)^\beta\right. \\ & \qquad \left.\prod_{i=1}^{p}\left(\left(r_i^2\right)^{\frac{\beta}{2}-1}\left(1-r_i^2\right)^{\frac{\beta}{2}(n-2p+1)-1} dr_i^2\right)\right] \\ & \overset{(a)}{=} \frac{1}{2^p} \underset{\substack{\sum_{i=1}^{p} x_i\le 1 \\ 1/\delta^2\ge x_1\ge\cdots\ge x_p\ge 0}}{\int\cdots\int} \left[V_{n,p,\beta}\prod_{i<j}^{p}\left(x_i-x_j\right)^\beta \delta^{2\beta\frac{p(p-1)}{2}}\right. \\ & \qquad \left.\prod_{i=1}^{p}\left(\left(1-\delta^2 x_i\right)^{\frac{\beta}{2}-1}\left(\delta^2 x_i\right)^{\frac{\beta}{2}(n-2p+1)-1}\left(-\delta^2 dx_i\right)\right)\right] \\ & \overset{(b)}{=} \begin{cases} c_{n,p,\beta}\delta^{\beta p(n-p)}\left(1+o\left(\delta\right)\right) & \text{if } \beta = 1 \\ c_{n,p,\beta}\delta^{\beta p(n-p)} & \text{if } \beta = 2 \end{cases}, \end{aligned}$$

where (a) follows the variable change $\delta^2 x_i = 1 - r_i^2$ and (b) follows from the definition

$$\begin{aligned} c_{n,p,\beta} \; \triangleq \; & \frac{1}{2^p} V_{n,p,\beta} \underset{\substack{\sum_{i=1}^{p} x_i\le 1 \\ 1\ge x_1\ge\cdots\ge x_p\ge 0}}{\int\cdots\int} \left[\left|\Delta_p\left(\mathbf{x}\right)\right|^\beta\right. \\ & \left.\prod_{i=1}^{p}\left(x_i^{\frac{\beta}{2}(n-2p+1)-1} dx_i\right)\right] \end{aligned} \tag{15}$$

and the fact that

$$\left(1-\delta^2 x_i\right)^{\frac{\beta}{2}-1} = \begin{cases} 1+o\left(\delta\right) & \text{if } \beta = 1 \\ 1 & \text{if } \beta = 2 \end{cases}$$

for small δ. Note that when $\delta \leq 1$, the integral domain $\sum_{i=1}^{p} x_i \leq 1$ and $1/\delta^2 \geq x_1 \geq \cdots \geq x_p \geq 0$ can be simplified as $\sum_{i=1}^{p} x_i \leq 1$ and $1 \geq x_1 \geq \cdots \geq x_p \geq 0$, which is independent of δ.

If instead, $\frac{n}{2} \leq p \leq n$, there are $2p-n$ principle angles always identical to zero and $n-p$ principle angles, say $\theta_1, \theta_2, \cdots, \theta_{n-p}$, different from zero. The distribution of the non-zero principle angles, expressed by the differential form, is [11]

$$
\begin{aligned}
d\mu_\theta &= \frac{1}{2^{n-p}} V_{n,p,\beta} \left| \Delta_{n-p}\left(\mathbf{r}^2\right) \right|^\beta \\
&\quad \prod_{i=1}^{n-p} \left(\left(r_i^2\right)^{\frac{\beta}{2}-1} \left(1-r_i^2\right)^{\frac{\beta}{2}(2p-n+1)-1} dr_i^2 \right).
\end{aligned}
$$

By calculation similar to the $0 < p \leq \frac{n}{2}$ case, the volume formula can be shown as

$$
\mu\left(B\left(\delta\right)\right) = \begin{cases} c_{n,p,\beta} \delta^{\beta p(n-p)} \left(1 + o\left(\delta\right)\right) & \text{if } \beta = 1 \\ c_{n,p,\beta} \delta^{\beta p(n-p)} & \text{if } \beta = 2 \end{cases},
$$

where

$$
\begin{aligned}
c_{n,p,\beta} &= \frac{1}{2^{n-p}} V_{n,n-p,\beta} \underset{\substack{\sum_{i=1}^{n-p} x_i \leq 1 \\ 1 \geq x_1 \geq \cdots \geq x_{n-p} \geq 0}}{\int \cdots \int} \left[\left| \Delta_{n-p}\left(\mathbf{x}\right) \right|^\beta \right. \\
&\quad \left. \prod_{i=1}^{n-p} \left(x_i^{\frac{\beta}{2}(2p-n+1)-1} dx_i \right) \right].
\end{aligned}
$$

2) Calculation of $c_{n,p,2}$: To calculate $c_{n,p,2}$, the key step is the following theorem.

Theorem 4: For unordered $x_1, x_2, \cdots, x_p$, define

$$
\Omega_p = \left\{ \left(x_1, \cdots, x_p\right) : \sum_{i=1}^{p} x_i \leq 1,\ x_i \geq 0\ 1 \leq i \leq p \right\}.
$$

Then, for any $t > 0$,

$$
\begin{aligned}
I_\Delta\left(t,p\right) &\triangleq \underset{\Omega_p}{\int \cdots \int} \left|\Delta\left(\mathbf{x}\right)\right|^2 \prod_{i=1}^{p} \left(x_i^t dx_i\right) \\
&= \frac{\prod_{i=1}^{p} \Gamma\left(i+1\right) \prod_{i=1}^{p} \Gamma\left(t+i\right)}{\Gamma\left(pt+p^2+1\right)}.
\end{aligned}
$$

Once this theorem is proved, substituting this theorem into the general formula for $c_{n,p,2}$ (15) gives the formula for $c_{n,p,2}$ by elementary calculation.

It is noteworthy that the although integral $I_\Delta\left(t,p\right)$ is similar to the Selberg's integral [18], they are different in that the integration of $I_\Delta\left(t,p\right)$ is a truncation of that of Selberg's integral. The calculation of $I_\Delta\left(t,p\right)$ involves new techniques.

To prove Theorem 4, we first express the integral as the determinant of a particular matrix and then evaluate the determinant of that matrix.

Lemma 1 expresses $I_\Delta\left(t,p\right)$ as the determinant of a matrix. Note that

$$
\begin{aligned}
&I_{\Delta}(t,p) \\
&= \int \cdots \int_{\Omega_p} \left(\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_p \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{p-1} & x_2^{p-1} & \cdots & x_p^{p-1} \end{bmatrix} \prod_{i=1}^{p} x_i^{t/2} \right) \\
&\quad \cdot \left(\det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{p-1} \\ 1 & x_2 & \cdots & x_2^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_p & \cdots & x_p^{p-1} \end{bmatrix} \prod_{i=1}^{p} x_i^{t/2} \right) \prod_{i=1}^{p} dx_i \\
&= \int \cdots \int_{\Omega_p} \det \begin{bmatrix} x_1^{\frac{t}{2}} & x_2^{\frac{t}{2}} & \cdots & x_p^{\frac{t}{2}} \\ x_1^{\frac{t}{2}+1} & x_2^{\frac{t}{2}+1} & \cdots & x_p^{\frac{t}{2}+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{\frac{t}{2}+p-1} & x_2^{\frac{t}{2}+p-1} & \cdots & x_p^{\frac{t}{2}+p-1} \end{bmatrix} \\
&\quad \cdot \det \begin{bmatrix} x_1^{\frac{t}{2}} & x_1^{\frac{t}{2}+1} & \cdots & x_1^{\frac{t}{2}+p-1} \\ x_2^{\frac{t}{2}} & x_2^{\frac{t}{2}+1} & \cdots & x_2^{\frac{t}{2}+p-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_p^{\frac{t}{2}} & x_p^{\frac{t}{2}+1} & \cdots & x_p^{\frac{t}{2}+p-1} \end{bmatrix} \prod_{i=1}^{p} dx_i \\
&= \int \cdots \int_{\Omega_p} \det \begin{bmatrix} \sum x_i^{t} & \sum x_i^{t+1} & \cdots & \sum x_i^{t+p-1} \\ \sum x_i^{t+1} & \sum x_i^{t+2} & \cdots & \sum x_i^{t+p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_i^{t+p-1} & \sum x_i^{t+p} & \cdots & \sum x_i^{t+2p-2} \end{bmatrix} \prod_{i=1}^{p} dx_i
\end{aligned}
$$

Define

$$
\begin{aligned}
G_{\mathbf{x}}(t,p) &\triangleq |\Delta(\mathbf{x})|^2 \prod_{i=1}^{p} x_i^{t} \\
&= \det \begin{bmatrix} \sum x_i^{t} & \sum x_i^{t+1} & \cdots & \sum x_i^{t+p-1} \\ \sum x_i^{t+1} & \sum x_i^{t+2} & \cdots & \sum x_i^{t+p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_i^{t+p-1} & \sum x_i^{t+p} & \cdots & \sum x_i^{t+2p-2} \end{bmatrix}
\end{aligned}
$$

and

$$
G(t,p) \triangleq \det \begin{bmatrix} \Gamma(t+1) & \Gamma(t+2) & \cdots & \Gamma(t+p) \\ \Gamma(t+2) & \Gamma(t+3) & \cdots & \Gamma(t+p+1) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(t+p) & \Gamma(t+1) & \cdots & \Gamma(t+2p-1) \end{bmatrix}.
$$

We have the following lemma.

Lemma 1:

$$
\begin{aligned}
I_{\Delta}(t,p) &= \int \cdots \int_{\Omega_p} G_{\mathbf{x}}(t,p) \prod_{i=1}^{p} dx_i \\
&= \frac{\Gamma(p+1)}{\Gamma(pt+p^2+1)} G(t,p).
\end{aligned}
$$

Proof: The polynomial $G_{\mathbf{x}}(t,p) = |\Delta(\mathbf{x})|^2 \prod_{i=1}^{p} x_i^t$ is a homogeneous polynomial of degree $pt + p(p-1)$. For every non-zero product term, i.e., $a_{t_1,\cdots,t_p} x_1^{t_1} \cdots x_p^{t_p}$ where $a_{t_1,\cdots,t_p} \neq 0$, we also define the order of that term as

$$\mathrm{O}\left(a_{t_1,\cdots,t_p} x_1^{t_1} \cdots x_p^{t_p}\right) = \min_{1 \le i \le p} t_i.$$

Then the minimum order of all the terms of the polynomial $G_{\mathbf{x}}(t,p)$ is greater than or equals to t. Let's call the terms with order greater than or equivalent to t as terms with appropriate order, and the terms with order less than t as terms with inappropriate order. When we expand

$$G_{\mathbf{x}}(t,p) = \det \begin{bmatrix} \sum x_i^t & \sum x_i^{t+1} & \cdots & \sum x_i^{t+p-1} \\ \sum x_i^{t+1} & \sum x_i^{t+2} & \cdots & \sum x_i^{t+p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_i^{t+p-1} & \sum x_i^{t+p} & \cdots & \sum x_i^{t+2p-2} \end{bmatrix},$$

the terms with inappropriate order will be cancelled out finally. Therefore, in the following, we ignore those terms even when they appear in some intermediate steps. This will simplify the analysis.

For any homogeneous polynomial of $x_1, \cdots, x_p$ that can be expressed in the form[3]

$$\bar{G}_{\mathbf{x}}(t,p) = \det \begin{bmatrix} \sum x_i^{t+a_{1,1}} & \sum x_i^{t+a_{1,2}} & \cdots & \sum x_i^{t+a_{1,p}} \\ \sum x_i^{t+a_{2,1}} & \sum x_i^{t+a_{2,2}} & \cdots & \sum x_i^{t+a_{2,p}} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_i^{t+a_{p,1}} & \sum x_i^{t+a_{p,2}} & \cdots & \sum x_i^{t+a_{p,p}} \end{bmatrix}$$

where $a_{k,l}$'s $1 \le k \le p, 1 \le l \le p$ are positive constants, we claim that

$$\int \cdots \int_{\Omega_p} \bar{G}_{\mathbf{x}}(t,p) \prod_{i=1}^{p} dx_i = \frac{\Gamma(p+1)}{\Gamma(pt+a+p+1)} \bar{G}(t,p) \tag{16}$$

where $a = \sum_{i=1}^{p} a_{i,i}$ and

$$\bar{G}(t,p) = \det \begin{bmatrix} \Gamma(t+a_{1,1}+1) & \Gamma(t+a_{1,2}+1) & \cdots & \Gamma(t+a_{1,p}+1) \\ \Gamma(t+a_{2,1}+1) & \Gamma(t+a_{2,2}+1) & \cdots & \Gamma(t+a_{2,p}+1) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(t+a_{p,1}+1) & \Gamma(t+a_{p,2}+1) & \cdots & \Gamma(t+a_{p,p}+1) \end{bmatrix}.$$

It is easy to verify this claim for $p = 1$ case. Suppose that this claim is true for all homogeneous polynomial $\bar{G}_{\mathbf{x}}(t,p-1)$. Then

$$\bar{G}_{\mathbf{x}}(t,p) = \sum_{l=1}^{p} \left(\sum x_i^{t+a_{1,l}}\right) (-1)^{1+l} M_{1,l},$$

where $M_{1,l}$ is the minor formed by eliminating row 1 and column l. It can be verified that $M_{1,l}$ is also a homogeneous polynomial of degree $pt + a - t - a_{1,l}$ (ignoring the terms with inappropriate order). Note that

$$\int \cdots \int_{\Omega_p} \prod_{i=1}^{p} x_i^{t_i} dx_i = \frac{\prod_{i=1}^{p} \Gamma(t_i+1)}{\Gamma\left(\sum_{i=1}^{p} t_i + p + 1\right)}$$

[3]In this expression, all the terms in the determinant expansion with inappropriate order are ignored.

where the final result is only dependent on the value of t_i's but independent of the order of those t_i's. According to determinant expansion by minors, we have

$$\begin{aligned}
&\underset{\Omega_p}{\int\cdots\int}\bar{G}_{\mathbf{x}}\left(t,p\right)\prod_{i=1}^{p}dx_i\\
&=\underset{\Omega_p}{\int\cdots\int}\left(\sum_{l=1}^{p}\left(\sum x_i^{t+a_{1,l}}\right)\left(-1\right)^{1+l}M_{1,l}\right)\prod_{i=1}^{p}dx_i\\
&\overset{(a)}{=}\sum_{l=1}^{p}\left(-1\right)^{1+l}\underset{\Omega_p}{\int\cdots\int}p\cdot x_1^{t+a_{1,l}}M_{1,l}\prod_{i=1}^{p}dx_i\\
&\overset{(b)}{=}\sum_{l=1}^{p}\left(-1\right)^{1+l}\underset{\Omega_p}{\int\cdots\int}p\cdot x_1^{t+a_{1,l}}M'_{1,l}\prod_{i=1}^{p}dx_i,
\end{aligned}$$

where $M'_{1,l}$ is the minor formed by eliminating row 1 and column l of the matrix

$$\begin{bmatrix}
\sum_{i=2}^{p}x_i^{t+a_{1,1}} & \sum_{i=2}^{p}x_i^{t+a_{1,2}} & \cdots & \sum_{i=2}^{p}x_i^{t+a_{1,p}}\\
\sum_{i=2}^{p}x_i^{t+a_{2,1}} & \sum_{i=2}^{p}x_i^{t+a_{2,2}} & \cdots & \sum_{i=2}^{p}x_i^{t+a_{2,p}}\\
\vdots & \vdots & \ddots & \vdots\\
\sum_{i=2}^{p}x_i^{t+a_{p,1}} & \sum_{i=2}^{p}x_i^{t+a_{p,2}} & \cdots & \sum_{i=2}^{p}x_i^{t+a_{p,p}}
\end{bmatrix},$$

(a) is because of the symmetry that

$$\underset{\Omega_p}{\int\cdots\int}x_1^{t+a_{1,l}}M_{1,l}\prod_{i=1}^{p}dx_i \quad = \quad \underset{\Omega_p}{\int\cdots\int}x_2^{t+a_{1,l}}M_{1,l}\prod_{i=1}^{p}dx_i$$

(ignoring the terms with inappropriate order), and
(b) is followed the observation that $x_1^{t+a_{1,l}}\cdot x_1^{t+a_{k,l'}}$ will result in a term with inappropriate order.

Therefore,

$$\begin{aligned}
&\underset{\Omega_p}{\int\cdots\int}\bar{G}_{\mathbf{x}}\left(t,p\right)\prod_{i=1}^{p}dx_i \\
&=\sum_{l=1}^{p}\left(-1\right)^{1+l}\underset{\Omega_p}{\int\cdots\int}p\cdot x_1^{t+a_{1,l}}M_{1,l}^{'}\prod_{i=1}^{p}dx_i \\
&=p\sum_{l=1}^{p}\left(-1\right)^{1+l}\int_0^1 x_1^{t+a_{1,l}}\left(\underset{\substack{\sum_{i=2}^{p}x_i\leq 1-x_1\\ 0\leq x_2,\cdots,0\leq x_p}}{\int\cdots\int}M_{1,l}^{'}\prod_{i=2}^{p}dx_i\right)dx_1 \\
&\overset{(a)}{=}p\sum_{l=1}^{p}\left(-1\right)^{1+l}\int_0^1 x_1^{t+a_{1,l}}\left(1-x_1\right)^{pt+a-t-a_{1,l}+p-1}dx_1 \\
&\qquad\cdot\left(\underset{\Omega_{p-1}}{\int\cdots\int}M_{1,l}^{'}\prod_{i=2}^{p}dx_i\right) \\
&\overset{(b)}{=}p\sum_{l=1}^{p}\left(-1\right)^{1+l}\frac{\Gamma\left(t+a_{1,l}\right)\Gamma\left(pt+a-t-a_{1,l}+p\right)}{\Gamma\left(pt+a+p+1\right)} \\
&\qquad\frac{\Gamma\left(p\right)}{\Gamma\left(pt+a-t-a_{1,l}+p\right)}\bar{G}_{1,l}\left(t,p-1\right) \\
&=\frac{\Gamma\left(p+1\right)}{\Gamma\left(pt+a+p+1\right)}\sum_{l=1}^{p}\left(-1\right)^{1+l}\Gamma\left(t+a_{1,l}\right)\bar{G}_{1,l}\left(t,p-1\right) \\
&\overset{(c)}{=}\frac{\Gamma\left(p+1\right)}{\Gamma\left(pt+a+p+1\right)}\bar{G}\left(t,p\right),
\end{aligned}$$

where

(a) follows the variable changes that $x_i^{'}=\frac{x_i}{1-x_1}$ and the fact that $M_{1,l}^{'}$ is a homogeneous polynomial with degree $pt+a-t-a_{1,l}$,

(b) is derived by calculating the integral with respect to x_1 and the assumption that

$$\left(\underset{\Omega_{p-1}}{\int\cdots\int}M_{1,l}^{'}\prod_{i=2}^{p}dx_i\right)=\frac{\Gamma\left(p\right)}{\Gamma\left(pt+a-t-a_{1,l}+p\right)}\bar{G}_{1,l}\left(t,p-1\right),$$

and

(c) follows the determinant expansion by minors of $\bar{G}\left(t,p\right)$.

Thus the claim (16) is proved. Lemma 1 immediately follows from this claim. ■

Therefore, the original integral $I_{\Delta}\left(t,p\right)$ is just a scaling determinant of a particular matrix, denoted as $G\left(t,p\right)$. The explicit formula to calculate $G\left(t,p\right)$ is given in the following lemma.

Lemma 2: $G\left(t,p\right)=\prod_{i=1}^{p}\Gamma\left(i\right)\prod_{i=1}^{p}\Gamma\left(t+i\right)$.

Proof: Since $G(t,1)=\Gamma(t+1)$ for any nonnegative number t, it is sufficient to prove $G(t,p)=\Gamma(t+1)\Gamma(p)G(t+1,p-1)$. Note that

$$\begin{aligned}
&G(t,p)\\
&=\det\begin{bmatrix}\Gamma(t+1) & \Gamma(t+2) & \cdots & \Gamma(t+p)\\ \vdots & \vdots & \ddots & \vdots\\ \Gamma(t+p-1) & \Gamma(t+p)! & \cdots & \Gamma(t+2p-2)\\ \Gamma(t+p) & \Gamma(t+p+1) & \cdots & \Gamma(t+2p-1)\end{bmatrix}\\
&=(p-1)!\Gamma(t+1)\det\begin{bmatrix}\Gamma(t+2) & \Gamma(t+3) & \cdots & \Gamma(t+p)\\ \vdots & \vdots & \ddots & \vdots\\ \Gamma(t+p-1) & \Gamma(t+p) & \cdots & \Gamma(t+2p-3)\\ \Gamma(t+p) & \Gamma(t+p+1) & \cdots & \Gamma(t+2p-2)\end{bmatrix}\\
&=\Gamma(p)\Gamma(t+1)G(t+1,p-1),
\end{aligned}$$

which follows from elementary row operations. This Lemma is therefore proved. ■

Substituting Lemma 2 into Lemma 1, we have

$$I_\Delta(t,p)=\frac{\prod_{i=1}^{p}\Gamma(i+1)\prod_{i=1}^{p}\Gamma(t+i)}{\Gamma(pt+p^2+1)}.$$

Theorem 4 is therefore proved. With elementary calculation,

$$c_{n,p,2}=\begin{cases}\frac{1}{(np-p^2)!}\prod_{i=1}^{p}\frac{(n-i)!}{(p-i)!} & \text{if } 0<p\le\frac{n}{2}\\ \frac{1}{(np-p^2)!}\prod_{i=1}^{n-p}\frac{(n-i)!}{(n-p-i)!} & \text{if } \frac{n}{2}\le p\le n\end{cases}.$$

B. Proof of Proposition 1

The volume of $B(\delta)$ is independent of the choice of the center. Let $\mathbf{I}_{n,p}$ be the $n\times p$ matrix formed by truncating the first p columns from the $n\times n$ identity matrix. Let $Q\in\mathcal{G}_{n,p}(\mathbb{L})$ be the plane generated by $\mathbf{I}_{n,p}$ and $P\in\mathcal{G}_{n,p}(\mathbb{L})$ be a uniformly distributed plane in the $\mathcal{G}_{n,p}(\mathbb{L})$. Then

$$\mu(B(\delta))=\Pr(d_c(P,Q)\le\delta).$$

The probability $\Pr(d_c(P,Q)\le\delta)$ can be calculated as follows. Let $\mathbf{H}\in\mathbb{L}^{n\times p}$ be a random matrix whose entries are i.i.d. Gaussian r.v. with unit variance per real dimension, i.e.,

$$H_{i,j}\sim\begin{cases}\mathcal{N}(0,1) & \text{if } \mathbb{L}=\mathbb{R}\\ \mathcal{CN}(0,1) & \text{if } \mathbb{L}=\mathbb{C}\end{cases},$$

where $H_{i,j}$ is the element in the i^{th} row and the j^{th} column of $\mathbf{H}$. Consider the singular value decomposition $\mathbf{H}\mathbf{H}^\dagger=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ where $\mathbf{U}\in\mathbb{L}^{n\times p}$, $\mathbf{U}^\dagger\mathbf{U}=\mathbf{I}_p$, $\mathbf{\Lambda}\in\mathbb{R}^{p\times p}$ is diagonal. It is well known that $\mathbf{\Lambda}$ and $\mathbf{U}$ are independent distributed, $\mathbf{U}$ is uniformly distributed in the Stiefel manifold $\mathcal{S}_{n,p}(\mathbb{L})$ [2] and $\mathcal{P}(\mathbf{U})$, the plane generated by $\mathbf{U}$, is uniformly distributed in the $\mathcal{G}_{n,p}(\mathbb{L})$. Thus,

$$\begin{aligned}
&\Pr(d_c(P,Q)\le\delta)\\
&=\Pr\left(d_c^2(\mathcal{P}(\mathbf{U}),Q)\le\delta^2\right)\\
&=\Pr\left(p-\mathrm{tr}\left(\mathbf{U}^\dagger\mathbf{I}_{n,p}\mathbf{I}_{n,p}^\dagger\mathbf{U}\right)\le\delta^2\right).
\end{aligned}$$

In order to calculate the distribution of $d_c^2(\mathcal{P}(\mathbf{U}),\mathcal{P}(\mathbf{I}_{n,p}))$, we use the following manipulations. It is easy to verify that

$$\mathrm{tr}\left(\frac{1}{\beta}\mathbf{H}^\dagger\mathbf{I}_{n,p}\mathbf{I}_{n,p}^\dagger\mathbf{H}\right)=\frac{1}{\beta}\sum_{i,j=1}^{p}|H_{i,j}|^2. \tag{17}$$

On the other hand,

$$\begin{aligned}
&\mathrm{tr}\left(\frac{1}{\beta}\mathbf{H}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{H}\right)\\
&=\mathrm{tr}\left(\frac{1}{\beta}\mathbf{\Lambda}\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)\\
&=\mathrm{tr}\left(\frac{1}{\beta}\left[\frac{1}{p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathbf{I}_{p}+\left(\mathbf{\Lambda}-\frac{1}{p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathbf{I}_{p}\right)\right]\right.\\
&\qquad\left.\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)\\
&=\frac{1}{\beta p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathrm{tr}\left(\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)\\
&\qquad+\mathrm{tr}\left(\frac{1}{\beta}\left(\mathbf{\Lambda}-\frac{1}{p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathbf{I}_{p}\right)\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)\\
&=\frac{1}{\beta p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathrm{tr}\left(\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)\\
&\qquad\left(1+\frac{\frac{1}{n}\mathrm{tr}\left(\left(\mathbf{\Lambda}-\frac{1}{p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathbf{I}_{p}\right)\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)}{\frac{1}{np}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathrm{tr}\left(\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)}\right).
\end{aligned}\tag{18}$$

Define $\mathbf{A}\triangleq\frac{1}{n}\left(\mathbf{\Lambda}-\frac{1}{p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathbf{I}_{p}\right)$ and $\mathbf{B}\triangleq\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}$. Since $\mathbf{B}$ is positive define, it can be proved that $\lambda_{\min}^{\mathbf{A}}\mathrm{tr}\left(\mathbf{B}\right)\leq\mathrm{tr}\left(\mathbf{AB}\right)\leq\lambda_{\max}^{\mathbf{A}}\mathrm{tr}\left(\mathbf{B}\right)$. Note that $\lambda_{\max}^{\mathbf{A}}$ and $\lambda_{\min}^{\mathbf{A}}$ converge to zero almost surely and $\frac{1}{np}\mathrm{tr}\left(\mathbf{\Lambda}\right)$ converges to β almost surely as $n\rightarrow+\infty$. Then (18) converges to $\frac{1}{\beta p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\mathrm{tr}\left(\mathbf{U}^{\dagger}\mathbf{I}_{n,p}\mathbf{I}_{n,p}^{\dagger}\mathbf{U}\right)=\frac{1}{\beta p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\left(p-d_c^2\right)$ in distribution as $n\rightarrow+\infty$. Substituting this conclusion into (17),

$$\frac{1}{\beta}\sum_{i,j=1}^{p}\left|H_{i,j}\right|^2\approx\frac{1}{\beta p}\mathrm{tr}\left(\mathbf{\Lambda}\right)\left(p-d_c^2\right),$$

where the symbol $\approx$ denotes that the random variables on the two sides of the equality have the same asymptotically distribution. Since $\mathrm{tr}\left(\mathbf{\Lambda}\right)=\sum_{i=1}^{n}\sum_{j=1}^{p}\left|H_{i,j}\right|^2$, we have

$$p-d_c^2\approx\frac{\frac{1}{\beta n}\sum_{i=1}^{p}\sum_{j=1}^{p}\left|H_{i,j}\right|^2}{\frac{1}{\beta np}\sum_{i=1}^{n}\sum_{j=1}^{p}\left|H_{i,j}\right|^2}.$$

After some manipulation,

$$\frac{d_c^2}{p}\approx1-\frac{1}{1+\frac{\sum_{i=p+1}^{n}\sum_{j=1}^{p}\left|H_{i,j}\right|^2/\beta p(n-p)}{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|H_{i,j}\right|^2/\beta p^2}\frac{\beta p(n-p)}{\beta p^2}}.$$

For notational convenience, let

$$a=\frac{\beta}{2}p\left(n-p\right),$$

$$b=\frac{\beta}{2}p^2,$$

and

$$X=\frac{\sum_{i=p+1}^{n}\sum_{j=1}^{p}\left|H_{i,j}\right|^2/2a}{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|H_{i,j}\right|^2/2b}.$$

Then

$$\frac{d_c^2}{p} \approx \frac{\frac{a}{b}X}{1+\frac{a}{b}X}.$$

We can calculate the distribution of d_c^2/p according to the distribution of X. Note that $\sum_{i=p+1}^{n}\sum_{j=1}^{p}\left|H_{i,j}\right|^2 \sim \chi^2_{\beta p(n-p)}$, $\sum_{i=1}^{p}\sum_{j=1}^{p}\left|H_{i,j}\right|^2 \sim \chi^2_{\beta p^2}$ and they are independent. The random variable X is F-distributed. Therefore,

$$\begin{aligned} f_{d_c^2/p}\left(y\right) &\approx f_X\left(x\right)\left|\frac{dx}{dy}\right| \\ &= f_X\left(\frac{b}{a}\frac{y}{1-y}\right)\frac{b}{a}\frac{1}{\left(1-y\right)^2} \\ &= \frac{\left(2b\right)^b\left(2a\right)^a\left(\frac{b}{a}\frac{y}{1-y}\right)^{a-1}}{B\left(a,b\right)\left(2b+2a\frac{b}{a}\frac{y}{1-y}\right)^{a+b}}\frac{b}{a}\frac{1}{\left(1-y\right)^2} \\ &= \frac{1}{B\left(a,b\right)}y^{a-1}\left(1-y\right)^{b-1}, \end{aligned}$$

where

$$B\left(a,b\right) = \frac{\Gamma\left(a\right)\Gamma\left(b\right)}{\Gamma\left(a+b\right)}$$

is the beta function.

With the density function for d_c^2/p, the volume of the metric ball $B\left(\delta\right)$ can be computed.

$$\begin{aligned} \mu\left(B\left(\delta\right)\right) &= F_{d_c^2/p}\left(\delta^2/p\right) \\ &\approx \frac{1}{B\left(a,b\right)}\int_0^{\delta^2/p} y^{a-1}\left(1-y\right)^{b-1}dy \\ &\approx \frac{\Gamma\left(a+b\right)}{\Gamma\left(a+1\right)\Gamma\left(b\right)}\left(\frac{\delta^2}{p}\right)^a\left(1-\frac{\delta^2}{p}\right)^{b-1}. \end{aligned}$$

C. Proof of Theorem 2

Assume a source Q is uniformly distributed in the $\mathcal{G}_{n,p}\left(\mathbb{L}\right)$. For any codebook $\mathcal{C}$, define the empirical cumulative distribution function as

$$F_{d_c^2,\mathcal{C}}\left(x\right) = \Pr\left\{Q:\ \left(\min_{P\in\mathcal{C}} d_c^2\left(P,Q\right)\right) \leq x\right\}.$$

Then the distortion associated with the codebook $\mathcal{C}$ is given by

$$D\left(\mathcal{C}\right) = \int_0^p x\cdot dF_{d_c^2,\mathcal{C}}\left(x\right). \tag{19}$$

To prove the lower bound, we want to find the optimal empirical cumulative distribution function (CDF) (corresponding to find the optimal code) to minimize the distortion. That is to find a $F_{d_c^2,\mathcal{C}}\left(x\right)$ to minimize

$$D\left(\mathcal{C}\right) = \int_0^{d_{c,\max}^2} x\cdot dF_{d_c^2,\mathcal{C}}\left(x\right),$$

with the constraint

$$F_{d_c^2,\mathcal{C}}\left(x\right) \leq \min\left(1, K\cdot\mu\left(B\left(\sqrt{x}\right)\right)\right),$$

where $\mu\left(B\left(\sqrt{x}\right)\right)$ is the volume of a metric ball of radius $\sqrt{x}$. The constraint follows from the fact that

$$\begin{aligned} F_{d_c^2,\mathcal{C}}\left(x\right) &= \Pr\left\{Q:\ \left(\min_{P\in\mathcal{C}} d_c^2\left(P,Q\right)\right)\leq x\right\} \\ &= \Pr\left(\cup_{i=1}^K\left\{Q:\ d_c^2\left(P_i,Q\right)\leq x\right\}\right) \\ &\leq \sum_{i=1}^K \Pr\left\{Q:\ d_c^2\left(P_i,Q\right)\leq x\right\} \\ &= K\cdot\mu\left(B\left(\sqrt{x}\right)\right), \end{aligned}$$

and the fact that $F_{d_c^2,\mathcal{C}}\left(x\right)\leq 1$.

It can be proved that the optimal empirical CDF to minimize the distortion is given by

$$F^*_{d_c^2,\mathcal{C}}\left(x\right)=\begin{cases} 0 & \text{if } x<0 \\ K\cdot\mu\left(\sqrt{x}\right) & \text{if } 0\leq x\leq x_0 \\ 1 & \text{if } x_0<x \end{cases},$$

where x_0 satisfies $K\cdot\mu\left(\sqrt{x_0}\right)=1$. To prove this claim, we calculate the difference between the distortions corresponding to $F^*_{d_c^2,\mathcal{C}}\left(x\right)$ and to any possible empirical CDF $F_{d_c^2,\mathcal{C}}\left(x\right)$. By an integration by parts,

$$\begin{aligned} D\left(\mathcal{C}\right) &= \int_0^p x\cdot dF_{d_c^2,\mathcal{C}}\left(x\right) \\ &= p-\int_0^p F_{d_c^2,\mathcal{C}}\left(x\right)dx. \end{aligned}$$

Then, with the same formula holding for $F^*_{d_c^2,\mathcal{C}}\left(x\right)$,

$$\begin{aligned} &\int_0^p x\cdot dF_{d_c^2,\mathcal{C}}\left(x\right)-\int_0^p x\cdot dF^*_{d_c^2,\mathcal{C}}\left(x\right) \\ &=\int_0^p F^*_{d_c^2,\mathcal{C}}\left(x\right)dx-\int_0^p F_{d_c^2,\mathcal{C}}\left(x\right)dx \\ &=\int_0^{x_0}\left(K\cdot\mu\left(B\left(\sqrt{x}\right)\right)-F_{d_c^2,\mathcal{C}}\left(x\right)\right)dx+\int_{x_0}^p\left(1-F_{d_c^2,\mathcal{C}}\left(x\right)\right)dx \\ &\geq 0, \end{aligned}$$

where the inequality in the last line follows from the constraint $F_{d_c^2,\mathcal{C}}\left(x\right)\leq\min\left(1,K\cdot\mu\left(B\left(\sqrt{x}\right)\right)\right).$

Substitute the optimal $F^*_{d_c^2,\mathcal{C}}\left(x\right)$ into (19), the corresponding distortion can be calculated. Suppose that the code size K is sufficiently large so that

$$\mu\left(B\left(\sqrt{x}\right)\right)\approx c_{n,p,\beta}x^{\frac{\beta p(n-p)}{2}}\qquad\text{for } x\leq x_0,$$

where

$$x_0=\left(c_{n,p,\beta}K\right)^{-\frac{2}{\beta p(n-p)}}.$$

Then

$$\begin{aligned} D^* &= \int_0^p x\cdot dF^*_{d_c^2,\mathcal{C}}\left(x\right) \\ &\approx \int_0^{x_0} x\cdot d\left(c_{n,p,\beta}x^{\frac{\beta p(n-p)}{2}}\right) \\ &\approx \frac{\beta p\left(n-p\right)}{\beta p\left(n-p\right)+2}\left(c_{n,p,\beta}K\right)^{-\frac{2}{\beta p(n-p)}}. \end{aligned}$$

D^* is a lower bound of the distortion rate function. D^* is the distortion corresponding to the optimal CDF $F^*_{d_c^2,\mathcal{C}}\left(x\right)$, which exists only when there exists K metric balls $B\left(\sqrt{x_0}\right)$ completely covers the whole $\mathcal{G}_{n,p}\left(\mathbb{L}\right)$ without any overlap. However, such a covering may or may not exist. Thus D^* is a lower bound of the actual distortion rate function.

D. Proof of Theorem 3

Let $\mathcal{C}_{\text{rand}} = \{P_1, \cdots, P_K\}$ be a random code whose codewords P_i's are independently drawn from the uniform distribution on the $\mathcal{G}_{n,p}(\mathbb{L})$. For any given element $Q \in \mathcal{G}_{n,p}(\mathbb{L})$, define $X_i = d_c^2(P_i, Q)$, $1 \le i \le K$. Then X_i's are independent and identically distributed (i.i.d.) random variables with CDF $F(x) = \mu\left(B\left(\sqrt{x}\right)\right)$. Define $W_K = \min(X_1, \cdots, X_K)$. Assume the source Q is uniformly distributed on the $\mathcal{G}_{n,p}(\mathbb{L})$ and independent with $\mathcal{C}_{\text{rand}}$. Then

$$\begin{aligned}
&\mathrm{E}_{\mathcal{C}_{\text{rand}}}\left[D\left(\mathcal{C}_{\text{rand}}\right)\right] \\
&= \mathrm{E}_{\mathcal{C}_{\text{rand}}}\left[\mathrm{E}_Q\left[\min_{P_i \in \mathcal{C}_{\text{rand}}} d_c^2(Q, P_i)\right]\right] \\
&= \mathrm{E}_Q\left[\mathrm{E}_{\mathcal{C}_{\text{rand}}}\left[\min_{P_i \in \mathcal{C}_{\text{rand}}} d_c^2(Q, P_i)\right]\right] \\
&= \mathrm{E}_Q\left[\mathrm{E}_{W_K}\left[W_K\right]\right].
\end{aligned}$$

To calculate $\mathrm{E}_{W_K}[W_K]$, the following lemma about the CDF of W_K is needed.

Lemma 3: Let X_i's $1 \le i \le K$ be i.i.d. random variables with CDF $F(x)$. Let $W_K = \min(X_1, \cdots, X_K)$. Then

$$\begin{aligned}
U(x) - 4KF^2(x)\left[1 - F(x)\right]^K &< \Pr\left(W_K \ge x\right) \\
&= \left(1 - F(x)\right)^K < U(x), \qquad (20)
\end{aligned}$$

where $U(x) = \exp(-KF(x))$, the first inequality holds for x such that $F(x) < \frac{1}{2\sqrt{K}}$ and the second inequality holds for all x.

Proof: See [19, page 10]. ∎

According to Lemma 3, some bounds on the CDF of W_K are derived. Define

$$t \triangleq \frac{\beta p (n-p)}{2}.$$

Define $y_{0,K}$ such that $y_{0,K}^t = \frac{1}{2c_{n,p,\beta}} K^{1/4}$. Then when $y \le y_{0,K}$,

$$U\left(\frac{y}{K^{1/t}}\right) > \Pr\left(W_K \ge \frac{y}{K^{1/t}}\right) > U\left(\frac{y}{K^{1/t}}\right) - K^{-1/2}, \qquad (21)$$

and when $y \ge y_{0,N}$,

$$\Pr\left(W_K \ge \frac{y}{K^{1/t}}\right) \le \exp\left(-\frac{1}{2}K^{1/4}\right). \qquad (22)$$

To prove (21), note that when $y \le y_{0,K}$,

$$F\left(\frac{y}{K^{1/t}}\right) \le F\left(\frac{y_{0,N}}{K^{1/t}}\right) = \frac{1}{2}K^{-3/4} < \frac{1}{2\sqrt{K}}.$$

According to Lemma 3 and the fact $\left[1 - F\left(y/K^{1/t}\right)\right]^K \le 1$, when $y \le y_{0,K}$,

$$\begin{aligned}
U\left(\frac{y}{K^{1/t}}\right) &> \Pr\left(W_K \ge \frac{y}{K^{1/t}}\right) \\
&> U\left(\frac{y}{K^{1/t}}\right) - 4K\frac{1}{4K^{3/2}} = U\left(\frac{y}{K^{1/t}}\right) - K^{-1/2}.
\end{aligned}$$

For $y \ge y_{0,N}$,

$$\begin{aligned}
&\Pr\left(W_K \ge \frac{y}{K^{1/t}}\right) \le \Pr\left(W_K \ge \frac{y_{0,K}}{K^{1/t}}\right) < \exp\left(-K \cdot c_{n,p,\beta}\frac{y_{0,K}^t}{K}\right) \\
&= \exp\left(-c_{n,p,\beta}\, y_{0,K}^t\right) = \exp\left(-\frac{1}{2}K^{1/4}\right).
\end{aligned}$$

Furthermore, according to the bounds in (21), the limit of the CDF of W_K as K approaches infinity can be derived. Noting that $U\left(y/K^{1/t}\right) = \exp\left(-c_{n,p,\beta}y^t\right)$ and $\lim\limits_{K\to+\infty} K^{-1/2} = 0$,

$$\lim_{K\to+\infty} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right) = \exp\left(-c_{n,p,\beta}y^t\right), \tag{23}$$

where the analysis holds for $y \leq y_{0,K} \to +\infty$ as $K \to +\infty$.

With the bounds on the CDF of W_K, we calculate $\mathrm{E}_{W_K}\left[W_K\right]$. In the following, we use $\mathrm{E}\left[W_K\right]$ instead of $\mathrm{E}_{W_K}\left[W_K\right]$ to simplify the notations. Employing the variable change $y = K^{1/t}x$,

$$\begin{aligned}
&\mathrm{E}\left[K^{1/t}W_K\right] \\
&= \int_0^{pK^{1/t}} yd\Pr\left(W_K < \frac{y}{K^{1/t}}\right) \\
&= -\int_0^{pK^{1/t}} yd\Pr\left(W_K \geq \frac{y}{K^{1/t}}\right) \\
&= -y\Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)\Big|_0^{pK^{1/t}} + \int_0^{pK^{1/t}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy.
\end{aligned}$$

Since $\Pr\left(W_K \geq \frac{y}{K^{1/t}}\right) = 0$ when $y = pK^{1/t}$, the first term vanishes. Thus,

$$\begin{aligned}
&\mathrm{E}\left[K^{1/t}W_K\right] \\
&= \int_0^{pK^{1/t}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy \\
&= \int_0^{y_{0,K}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy + \int_{y_{0,K}}^{pK^{1/t}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy,
\end{aligned}$$

where $y_{0,K}$ is taken such that $y_{0,K}^t = \frac{1}{2c_{n,p,\beta}}K^{1/4}$.

We perform the limit of $\mathrm{E}\left[K^{1/t}W_K\right]$ as $K \to +\infty$ by splitting the limit into two parts

$$\begin{aligned}
&\lim_{K\to+\infty} \mathrm{E}\left[K^{1/t}W_K\right] \\
&= \lim_{K\to+\infty}\left[\int_{y_{0,K}}^{pK^{1/t}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy\right] \\
&\quad + \lim_{K\to+\infty}\left[\int_0^{y_{0,K}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy\right].
\end{aligned} \tag{24}$$

The splitting is validated by the facts that

$$\begin{aligned}
0 &\leq \lim_{K\to+\infty}\int_{y_{0,K}}^{pK^{1/t}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right)dy \\
&\overset{(a)}{\leq} \lim_{K\to+\infty} pK^{\frac{1}{t}}\exp\left(-\frac{1}{2}K^{1/4}\right) \\
&= 0,
\end{aligned}$$

where (a) follows from (22), and that

$$
\begin{aligned}
&\lim_{K\to+\infty}\left[\int_0^{y_{0,K}} \Pr\left(W_K \geq \frac{y}{K^{1/t}}\right) dy\right] \\
&\overset{(b)}{=} \int_0^{+\infty} \lim_{K\to+\infty}\left[\Pr\left(W_K \geq \frac{y}{K^{1/t}}\right) 1\left(y < y_{0,K}\right)\right] dy \\
&\overset{(c)}{=} \int_0^{+\infty} \exp\left(-c_{n,p,\beta} y^t\right) dy \\
&= c_{n,p,\beta}^{-1/t} \cdot \frac{1}{t} \cdot \left[\frac{1}{1/t}\int_0^{+\infty} \exp\left(-c_{n,p,\beta} y^t\right) d\left(c_{n,p,\beta} y^t\right)^{1/t}\right] \\
&= c_{n,p,\beta}^{-1/t}\frac{\Gamma\left(t\right)}{t},
\end{aligned}
$$

where

$$
1\left(y < y_{0,K}\right) = \begin{cases} 1 & \text{if } y < y_{0,K} \\ 0 & \text{otherwise} \end{cases},
$$

(b) follows from the Lebesgue's Dominated Convergence Theorem and the facts that $\Pr\left(W_K \geq y/K^{1/t}\right) < \exp\left(-c_{n,p,\beta} y^t\right)$ and that $\exp\left(-c_{n,p,\beta} y^t\right)$ is integrable, and

(c) follows from (23).

Therefore,

$$
\begin{aligned}
&\lim_{K\to+\infty} \mathrm{E}_{\mathcal{C}_{\mathrm{rand}}}\left[D\left(\mathcal{C}_{\mathrm{rand}}\right)\right] \\
&= \lim_{K\to+\infty} \mathrm{E}_Q\left[\mathrm{E}_{W_K}\left[W_K\right]\right] \\
&\overset{(d)}{=} \lim_{K\to+\infty} \mathrm{E}_{W_K}\left[W_K\right] \\
&= c_{n,p,\beta}^{-1/t}\frac{\Gamma\left(t\right)}{t},
\end{aligned}
$$

where (d) holds because $\mathrm{E}_{W_K}\left[W_K\right]$ is independent of the choice of Q.

REFERENCES

[1] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Info. Theory*, vol. 47, no. 6, pp. 2618 – 2626, 2001.

[2] Z. Lizhong and D. N. C. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Info. Theory*, vol. 48, no. 2, pp. 359 – 383, 2002.

[3] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Info. Theory*, vol. 48, no. 9, pp. 2450–2454, 2002.

[4] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.

[5] D. J. Love, J. Heath, R. W., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[6] W. Dai, Y. Liu, V. K. N. Lau, and B. Rider, "On the information rate of MIMO systems with finite rate channel state feedback using beamforming and power on/off strategy," *IEEE Trans. Info. Theory*, submitted. [Online]. Available: http://ece-www.colorado.edu/~liue/publications/for_reviewer_onoff.pdf

[7] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc., packing in Grassmannian spaces," *Exper. Math.*, vol. 5, pp. 139–159, 1996.

[8] B. Mondal, R. W. H. Jr., and L. W. Hanlen, "Quantization on the Grassmann manifold: Applications to precoded MIMO wireless systems," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2005, pp. 1025–1028.

[9] A. T. James, "Normal multivariate analysis and the orthogonal group," *Ann. Math. Statist.*, vol. 25, no. 1, pp. 40 – 75, 1954.

[10] R. J. Muirhead, *Aspects of multivariate statistical theory*. New York: John Wiley and Sons, 1982.

[11] M. Adler and P. v. Moerbeke, "Integrals over Grassmannians and random permutations," *ArXiv Mathematics e-prints*, 2001.

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

[13] J. Heath, R. W. and D. J. Love, "Multi-mode antenna selection for spatial multiplexing with linear receivers," *IEEE Trans. Signal Processing*, accepted.

[14] D. Love and J. Heath, R.W., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Info. Theory*, to appear.

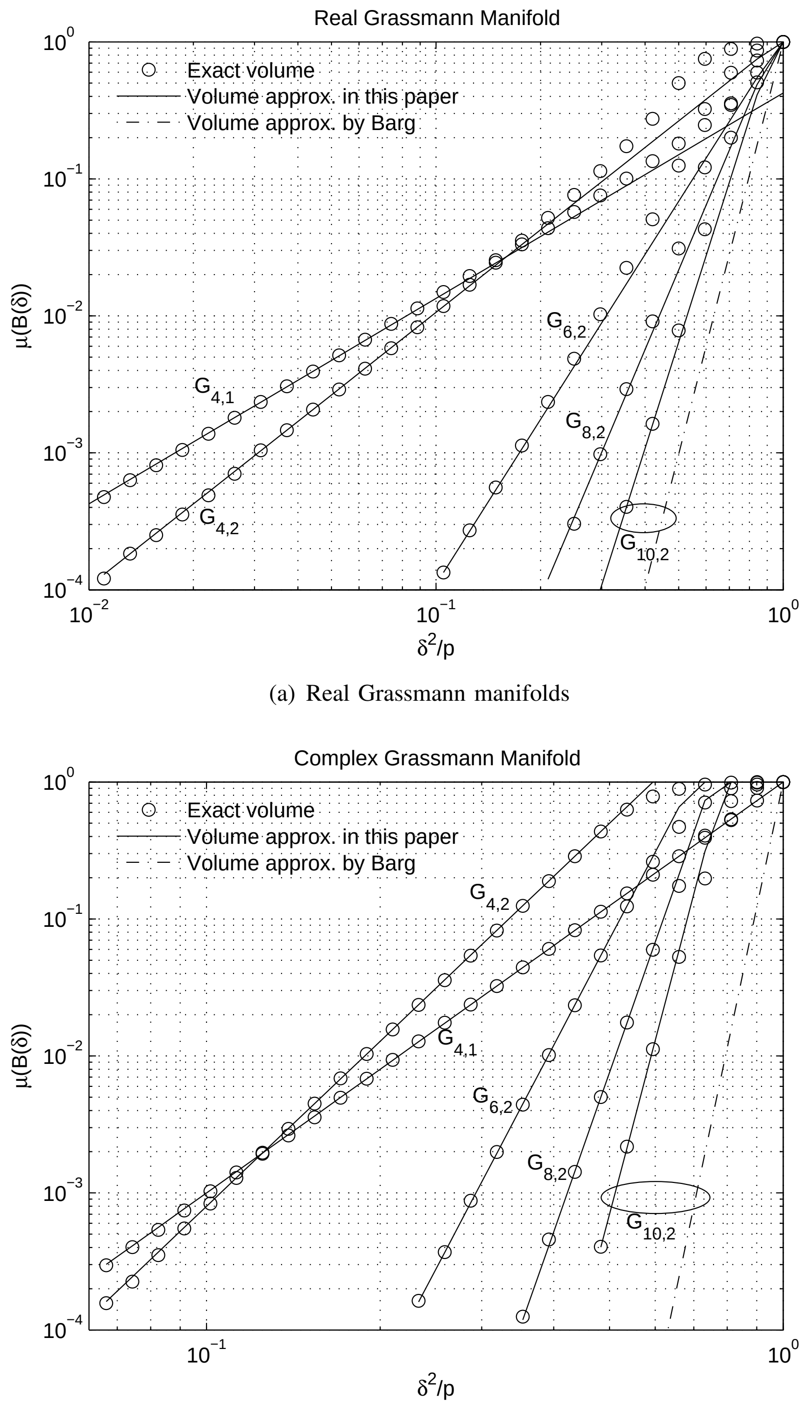
[15] J. C. Roh and B. D. Rao, "MIMO spatial multiplexing systems with limited feedback," in *Proc. IEEE International Conference on Communications (ICC)*, 2005.

[16] W. Santipach and M. L. Honig, "Asymptotic performance of MIMO wireless channels with limited feedback," in *Proc. IEEE Military Comm. Conf.*, vol. 1, 2003, pp. 141– 146.

[17] W. Santipach, Y. Sun, and M. L. Honig, "Benefits of limited feedback for wireless channels," in *Proc. Allerton Conf. on Commun., Control, and Computing*, 2003.

[18] M. L. Mehta, *Random Matrices*. Elsevier Academic Press, 2004.

[19] J. Galambos, *The asymptotic theory of extreme order statistics*, 2nd ed. Roberte E. Krieger Publishing Company, 1987.



(a) Real Grassmann manifolds

(b) Complex Grassmann manifolds

Fig. 1. The volume of a metric ball in the Grassmann manifold

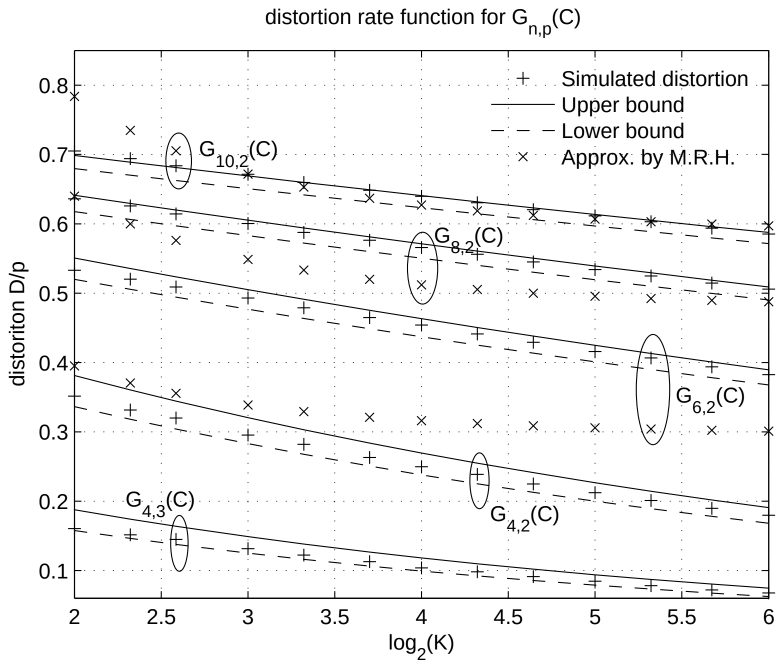


Fig. 2. Bounds on the distortion rate function

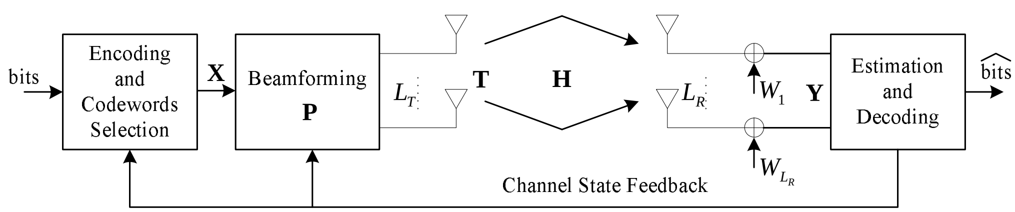


Fig. 3. System model for a MIMO system

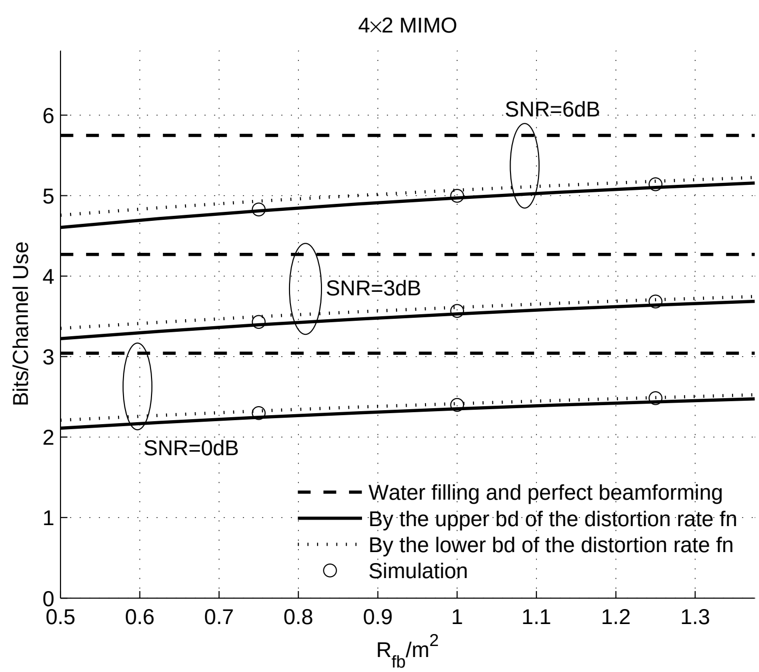
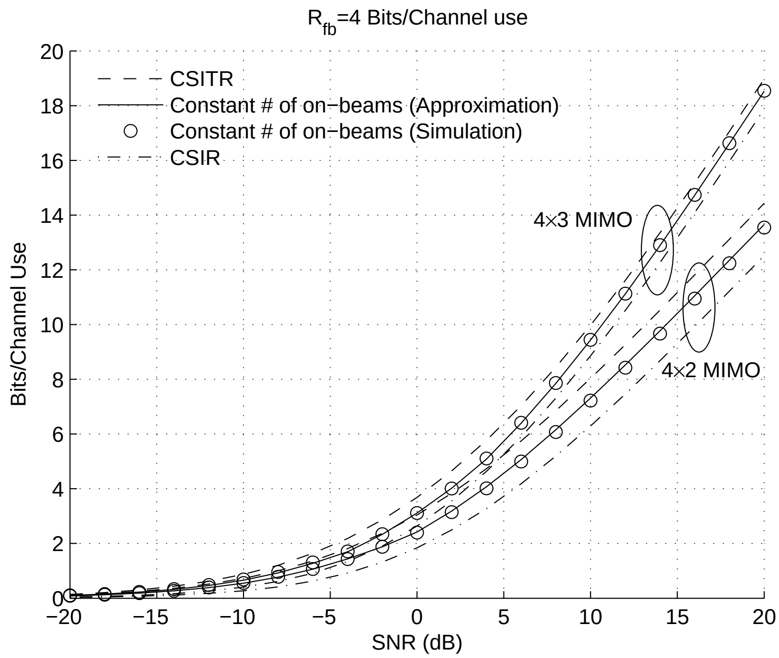
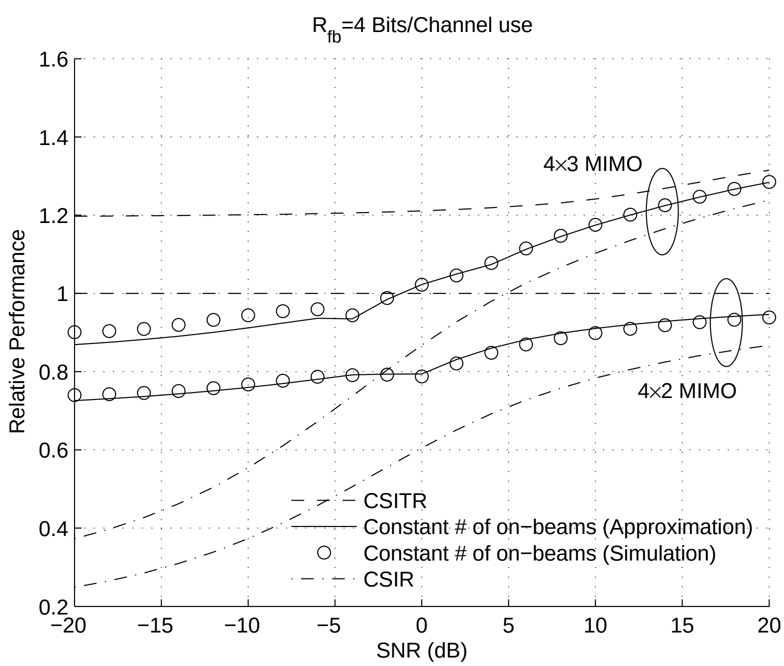


Fig. 4. Performance of a constant number of on-beams v.s. feedback rate



(a) Average mutual information



(b) Relative performance

Fig. 5. Performance of a constant number of on-beams v.s. SNR