Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems

David J. Love, *Student Member, IEEE*, Robert W. Heath, Jr., *Member, IEEE*, and Thomas Strohmer

***Abstract*—Transmit beamforming and receive combining are simple methods for exploiting the significant diversity that is available in multiple-input multiple-output (MIMO) wireless systems. Unfortunately, optimal performance requires either complete channel knowledge or knowledge of the optimal beamforming vector; both are hard to realize. In this correspondence, a quantized maximum signal-to-noise ratio (SNR) beamforming technique is proposed where the receiver only sends the label of the best beamforming vector in a predetermined codebook to the transmitter. By using the distribution of the optimal beamforming vector in independent and identically distributed Rayleigh fading matrix channels, the codebook design problem is solved and related to the problem of Grassmannian line packing. The proposed design criterion is flexible enough to allow for side constraints on the codebook vectors. Bounds on the codebook size are derived to guarantee full diversity order. Results on the density of Grassmannian line packings are derived and used to develop bounds on the codebook size given a capacity or SNR loss. Monte Carlo simulations are presented that compare the probability of error for different quantization strategies.**

***Index Terms*—Diversity methods, Grassmannian line packing, limited feedback, multiple-input multiple-output (MIMO) systems, Rayleigh channels.**

I. INTRODUCTION

Multiple-input multiple-output (MIMO) wireless systems make use of the spatial dimension of the channel to provide considerable capacity [1], [2], increased resilience to fading [3]–[5], or combinations of the two [6]–[8]. While the spectral efficiency improvement offered by MIMO communication is substantial, the reductions in fading obtained by trading capacity for spatial diversity should not be overlooked [9], [10]. In narrow-band Rayleigh-fading matrix channels, MIMO systems can provide a diversity in proportion to the product of the number of transmit and receive antennas. Diversity in a MIMO system can be obtained through the use of space–time codes (see e.g., [3]–[5]) or via intelligent use of channel state information at the transmitter (see, e.g., [11]–[17]). Transmit beamforming with receive combining is one of the simplest approaches to achieving full diversity and has been of interest recently [12]–[20]. Beamforming and combining in MIMO systems are a generalization of the vector channel beamforming/combining methods found in single-input–multiple-output (SIMO) combiners and multiple-input–single-output (MISO) beamformers which provide significantly more diversity. Compared with traditional space–time codes [3]–[5], beamforming and combining systems provide the same diversity order as well as significantly more array gain [21] at the expense of requiring channel state information at the transmitter in the form of the transmit beamforming vector (see, for example, [13]–[20]). Unfortunately, in systems where the forward and reverse channels are not reciprocal, this requires coarsely quantizing the channel or beamforming vector to accommodate the limited bandwidth of the feedback channel.

In this correspondence, we consider the problem of quantized beamforming for independent and identically distributed (i.i.d.) MIMO Rayleigh flat-fading channels when the transmitter has access to a low-bandwidth feedback channel from the receiver and the receiver employs maximum ratio combining (MRC). To support the limitations of the feedback channel, we assume the use of a codebook of possible beamforming vectors known to both the transmitter and receiver. The codebook is restricted to have fixed cardinality N and is designed off-line. The receiver is assumed to convey the best beamforming vector from the codebook over an error-free, zero-delay feedback channel. A primary contribution of this correspondence is to provide a constructive method for designing a quantized beamforming codebook. We show, using the distribution of the optimal unquantized beamforming vector, that the codebook design problem is equivalent to the problem of packing one-dimensional subspaces known as Grassmannian line packing.[1] These codebooks are a function of the number of transmit antennas and the size of the codebook but are independent of the number of receive antennas. We show that a sufficient condition for providing full diversity order is that the codebook cardinality is greater than or equal to the number of transmit antennas. We consider codebooks with additional constraints imposed on the beamforming vectors such as constant modulus entries or generalized subset selection.

The connection between Grassmannian line packing and quantized beamforming allows us to leverage results from the subspace packing literature to find constructive methods for deriving codebooks and also provides insight into codebook quality. In order to understand how the amount of feedback relates to system performance, we derive a new closed-form expression for the density of line packings based on a result from [22]. The density expression verifies the asymptotic subspace packing density presented in [23] and allows us to derive the Hamming bound and the Gilbert–Varshamov bound on codebook size. We use these results to obtain approximate bounds for choosing the codebook size based on a specific allowable capacity or average signal-to-noise ratio (SNR) loss.

Unquantized beamforming for MIMO systems was first proposed in [13]–[15]. Prior work on quantized beamforming, proposed in [24], addressed the problem of quantizing the maximum ratio transmission (MRT) [13]–[15] solution, which we call quantized maximum ratio transmission (QMRT). The beamforming codebooks proposed there were obtained using the Lloyd algorithm and a specific codebook design methodology was not developed. Additionally, the results were specialized only to MISO systems though, as we show, they are applicable to the MIMO case as well. The problem of quantizing the equal gain transmission (EGT) solution was proposed in [25]. The solution proposed there uniformly quantized the phases of the channel and does

Manuscript received October 30, 2002; revised June 21, 2003. This work is based in part upon work supported by the Texas Advanced Research (Technology) Program under Grant 003658-0614-2001, the National Science Foundation under Grant DMS–0208568, the National Instruments Foundation, and the Samsung Advanced Institute of Technology. The work of D. J. Love was supported by a Microelectronics and Computer Development Fellowship and a Cockrell Doctoral Fellowship through The University of Texas at Austin. The material in this correspondence was presented in part at the 40th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2002, the 36th IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2002, and the IEEE International Conference on Communications, Anchorage, AK, May 2003.

D. J. Love and R. W. Heath, Jr. are with the Wireless Networking and Communications Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: djlove@ece.utexas.edu; rheath@ece.utexas.edu).

T. Strohmer is with the Department of Mathematics, University of California, Davis, Davis, CA 95616 (e-mail: strohmer@math.ucdavis.edu).

Communicated by B. Hassibi, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.817466

[1]Recall that Grassmannian line packing is the problem of spacing N lines that pass through the origin in order to maximize the sine of the minimum angular separation between any two lines.

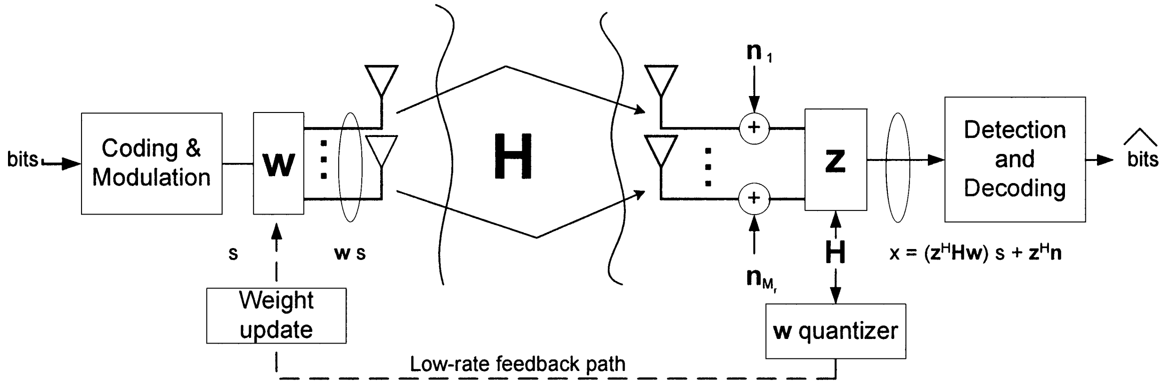

Fig. 1. Block diagram of a MIMO system.

not make the connection to line packing. Different codebooks were designed in [17], [20], extending the work in [25], but are still suboptimal since they are required to use codebooks containing a set of orthogonal vectors to satisfy the supposition for the proof of diversity order. Variations of QMRT and quantized EGT (QEGT) are part of the wide-band code division multiple access (WCDMA) closed-loop diversity mode [26]. The proposed solutions are specialized for two transmit antennas and essentially quantize the channel from one of the antennas. Transmit antenna selection for MIMO systems is a special case of quantized beamforming and has been proposed in [27], [28] for the MISO case and [11], [12] for the MIMO case.

The relationship between quantized beamforming and Grassmannian line packing was observed in [22], [29]–[31] in parallel and independently of our work in [32]–[35]. Their analysis, however, is explicitly for the MISO scenario and does not encompass MIMO beamforming and combining systems. Additionally, [22], [29]–[31] do not specifically address the design of hardware-constrained beamformers. Imposing additional constraints on the beamforming vector codebook, such as equal gain coefficients or selection columns, makes limited feedback precoding more practical than with arbitrary codebooks (e.g., see the closed-loop mode in the WCDMA standard [26]). In addition, we propose new results in Grassmannian line packing that are of use in judging the optimality of the designed quantized beamformers. Our analysis considers the amount of feedback required given acceptable losses in capacity or SNR.

This correspondence is organized as follows. Section II reviews beamforming and combining in MIMO systems and states the quantized beamforming problem. Grassmannian line packing is reviewed in Section III, and some results on the minimum distance and density are derived. Section IV examines the distribution of the optimal beamforming vector, proposes a distortion criterion, and then relates the problem of quantizing this vector to the problem of Grassmannian line packing. Different performance criteria are studied in Section V to provide some insight on selecting the codebook size. Section VI presents Monte Carlo simulation results that illustrate performance as a function of the amount of feedback available. The correspondence concludes in Section VII with some suggestions for future work.

II. System Overview

A MIMO system with transmit beamforming and receive combining, using M_t transmit antennas and M_r receive antennas, is illustrated in Fig. 1. Suppose that the bandwidth is much smaller than the coherence time of the channel thus the discrete-time equivalent channel can be modeled as an $M_r \times M_t$ matrix $\boldsymbol{H}$. Then the discrete-time input/output relationship at baseband, given a real or complex transmitted symbol s, for this system is given by[2]

$$x = \boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w} s + \boldsymbol{z}^H \boldsymbol{n}. \tag{1}$$

The vectors $\boldsymbol{w}$ and $\boldsymbol{z}$ are called the beamforming and combining vectors, respectively. The noise vector $\boldsymbol{n}$ has i.i.d. entries distributed according to $\mathcal{CN}(0, N_0)$. We model the channel $\boldsymbol{H}$ as having i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$. The channel is assumed to be known perfectly at the receiver. The symbol energy is given by $E_s[|s|^2] = \mathcal{E}_t$.

In a beamforming and combining system, the key question is how to design $\boldsymbol{w}$ and $\boldsymbol{z}$ to maximize performance. It has been shown [15], [16], [28] that $\boldsymbol{w}$ and $\boldsymbol{z}$ should be chosen to maximize the SNR in order to minimize the average probability of error and maximize the capacity. For the proposed system, the SNR γ_r, after combining at the receiver, is

$$\begin{aligned}\gamma_r &= \frac{\mathcal{E}_t |\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|^2}{\|\boldsymbol{z}\|_2^2 N_0} \\ &= \frac{\left(\mathcal{E}_t \|\boldsymbol{w}\|_2^2\right) \left| \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_2}^H \boldsymbol{H} \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2} \right|^2}{N_0}.\end{aligned} \tag{2}$$

Notice that in (2), $\|\boldsymbol{z}\|_2$ factors out, therefore, we fix $\|\boldsymbol{z}\|_2 = 1$ without loss of generality. As well, the transmitter transmits with total energy $\mathcal{E}_t \|\boldsymbol{w}\|_2^2$, therefore, we assume that $\|\boldsymbol{w}\|_2 = 1$ and that $\mathcal{E}_t$ is held constant for power constraint reasons. Using these assumptions

$$\gamma_r = \frac{\mathcal{E}_r}{N_0} = \frac{\mathcal{E}_t |\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|^2}{N_0} = \frac{\mathcal{E}_t \boldsymbol{\Gamma}_r}{N_0} \tag{3}$$

where $\boldsymbol{\Gamma}_r = |\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|^2$ is the effective channel gain.

In a MIMO system, unlike in a MISO system, both a transmit beamforming vector and a receive combining vector need to be chosen. A receiver where $\boldsymbol{z}$ maximizes $|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|$ given $\boldsymbol{w}$ is called an MRC receiver. The form of this vector follows from the vector norm inequality

$$|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|^2 \le \|\boldsymbol{z}\|_2^2 \|\boldsymbol{H} \boldsymbol{w}\|_2^2. \tag{4}$$

We already defined $\|\boldsymbol{z}\|_2^2 = 1$, thus, the MRC vector must set

$$|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|^2 = \|\boldsymbol{H} \boldsymbol{w}\|_2^2. \tag{5}$$

[2]We use w_i to refer to the ith entry of the vector $\boldsymbol{w}$, $H_{[k,l]}$ to refer to the (k, l) entry of a matrix $\boldsymbol{H}$, ^{T} to denote matrix transposition, ^{H} to denote matrix conjugate transposition, $|\cdot|$ to denote absolute value, $\|\cdot\|_2$ to denote the matrix two-norm, $\|\cdot\|_1$ to denote the matrix one-norm, $j = \sqrt{-1}$, $\mathbb{C}^m$ to denote the m-dimensional complex vector space, Ω_m to refer to the set of unit vectors in $\mathbb{C}^m$, $\mathcal{U}_m^N$ is the set of $m \times N$ complex matrices with unit vector columns, and $E_y[\cdot]$ to denote expectation with respect to a random variable y.

This is easily seen to be the unit vector $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{w}/\|\boldsymbol{H}\boldsymbol{w}\|_2$. We assume that the receiver always uses MRC.

The beamforming vectors $\boldsymbol{w}$ and $\boldsymbol{z}$ can be designed to maximize the SNR under different side constraints depending on implementation issues. Since we assume optimal combining at the receiver, we are primarily concerned with selecting $\boldsymbol{w}$. The four interesting cases are maximum ratio transmission, equal gain transmission, selection diversity transmission (SDT), and generalized subset selection (GSS). A transmitter where $\boldsymbol{w}$ maximizes $|\boldsymbol{z}^H\boldsymbol{H}\boldsymbol{w}|$ given $\boldsymbol{z}$ is called maximum ratio transmission (MRT). A transmitter where w_k satisfies $|w_k| = \frac{1}{\sqrt{M_t}}$ for $1 \le k \le M_t$ is called equal gain transmission (EGT). Note that this definition allows $\boldsymbol{w}$ to be expressed as $\boldsymbol{w} = \frac{1}{\sqrt{M_t}} e^{j\boldsymbol{\theta}}$ where $\boldsymbol{\theta} = [\theta_1 \theta_2 \cdots \theta_{M_t}]^T$ and $\theta_k \in [0, 2\pi)$. SDT requires that $\boldsymbol{w}$ be one of the columns of $\boldsymbol{I}_{M_t}$, the $M_t \times M_t$ identity matrix. A transmitter where $\tilde{\boldsymbol{w}}$ is the sum of columns of $\boldsymbol{I}_{M_t}$ and $\boldsymbol{w} = \tilde{\boldsymbol{w}}/\|\tilde{\boldsymbol{w}}\|_2$ is called generalized subset selection (GSS). This corresponds to vectors of the form

$$\boldsymbol{w} = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} (\boldsymbol{I}_{M_t})_{n_k}$$

where $(\boldsymbol{I}_{M_t})_{n_k}$ is the n_kth column of $\boldsymbol{I}_{M_t}$ and $n_k \neq n_l$ for $knTeql$. GSS is clearly a generalization of SDT when more than one radio chain is available. This method corresponds to transmitting on subsets of antennas depending on channel conditions.

Given no design constraints on the form of the unit vectors $\boldsymbol{w}$ or $\boldsymbol{z}$ and a fixed N_0, the optimal solutions in an average probability of symbol error sense are the beamforming and combining vectors, respectively, that maximize $\mathcal{E}_r$. For a combining scheme that solves for the beamforming vector $\boldsymbol{w}$ using the feasible set[3] $\mathcal{W}$ ($\mathcal{W} \subseteq \Omega_{M_t}$) with an MRC receiver, $\boldsymbol{w}$ is given by

$$\boldsymbol{w} = \arg\max_{\boldsymbol{x} \in \mathcal{W}} \|\boldsymbol{H}\boldsymbol{x}\|_2 \tag{6}$$

where $\arg\max$ returns a global maximizer. Note that this optimization returns only *one* out of possibly *many* global maximizers meaning that the global maximizer over most $\mathcal{W}$ is not unique. Notice that if $\mathcal{W} = \Omega_{M_t}$, the case for an MRT system, then $\boldsymbol{w}$ is the dominant right singular vector of $\boldsymbol{H}$, the right singular vector of $\boldsymbol{H}$ corresponding to the largest singular value of $\boldsymbol{H}$ [14], [15].

In this correspondence, we consider a communication link where channel state information is not available to the transmitter, but there exists a low-rate, error-free, zero-delay feedback link for the purpose of conveying $\boldsymbol{w}$ to the transmitter. Since $\boldsymbol{w}$ can be any unit vector in possibly a continuum of feasible vectors (Ω_{M_t} for MRT), it is essential to introduce some method of quantization due to the limited reverse-link feedback channel. A reasonable solution is to let the receiver and transmitter both use a codebook of N beamforming vectors [24], [25]. The receiver then quantizes the beamforming vector by selecting the best (according to (6)) beamforming vector from the codebook and conveys the index of this vector back to the transmitter. The main benefit of using a finite codebook is that the number of feedback bits can be kept to a manageable number given by $\lceil \log_2 N \rceil$. Unfortunately, it is not obvious which N vectors should be included in the codebook.

To compare the performance of different quantized and unquantized beamformers, we use the average probability of symbol error defined as $\overline{P_e} = E_{\boldsymbol{H}}[P_e]$, where the expected value of the probability of symbol error P_e is taken over the channel $\boldsymbol{H}$. Two measures that are relevant when comparing average probability of symbol error are array gain and diversity order. A system is said to have *array gain* A and *diversity order* D if for SNR $\gg 0$ the average probability of symbol error is inversely proportional to $A(\mathcal{E}_t/N_0)^D$ [6].

[3]A feasible set is the set that a cost function is maximized over.

III. Grassmannian Line Packing

Grassmannian line packing is the problem of optimally packing one-dimensional subspaces [36]. It is similar to the problem of spherical code design with one important difference: spherical codes are *points* on the unit sphere while Grassmannian line packings are *lines* passing through the origin in a vector space. Grassmannian line packing forms the basis for our quantized beamforming codebook design. In this section, we present a summary of key results on Grassmannian line packing and some new results. The terminology follows from the work of researchers in Grassmannian subspace packing (see, for example, Sloane's webpage [37]).

Consider the space of unit-norm transmit beamforming vectors Ω_m. Let use define an equivalence relation between two unit vectors $\boldsymbol{w}_1 \in \Omega_m$ and $\boldsymbol{w}_2 \in \Omega_m$ by $\boldsymbol{w}_1 \equiv \boldsymbol{w}_2$ if for some $\theta \in [0, 2\pi)$, $\boldsymbol{w}_1 = e^{j\theta}\boldsymbol{w}_2$. This equivalence relation says that two vectors are equivalent if they are on the same line in $\mathbb{C}^m$. The quotient space with respect to this equivalence relation is the set of all one-dimensional subspaces in $\mathbb{C}^m$ [38]. The complex *Grassmann manifold* $\mathcal{G}(m, 1)$ is the set of all one-dimensional subspaces of the space $\mathbb{C}^m$. We define a distance function on $\mathcal{G}(m, 1)$ by letting the distance between the two lines generated from unit vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ be the sine of the angle $\theta_{1,2}$ between the two lines. This distance is expressed as [23]

$$d(\boldsymbol{w}_1, \boldsymbol{w}_2) = \sin(\theta_{1,2}) = \sqrt{1 - |\boldsymbol{w}_1^H \boldsymbol{w}_2|^2}.$$

The *Grassmannian line packing problem* is the problem of finding the set, or packing, of N lines in $\mathbb{C}^m$ that has maximum minimum distance between any pair of lines. Because of the relation to Ω_m, the problem simplifies down to arranging N unit vectors so that the magnitude correlation between any two vectors is as small as possible. We represent a packing of N lines in $\mathcal{G}(m, 1)$ by an $m \times N$ matrix $\boldsymbol{W} = [\boldsymbol{w}_1 \boldsymbol{w}_2 \cdots \boldsymbol{w}_N]$, where $\boldsymbol{w}_i$ is the vector in Ω_m whose column space is the ith line in the packing. The packing problem is only of interest in nontrivial cases where $N > m$.

The minimum distance of a packing is the sine of the smallest angle between any pair of lines. This is written as

$$\delta(\boldsymbol{W}) = \min_{1 \le k < l \le N} \sqrt{1 - |\boldsymbol{w}_k^H \boldsymbol{w}_l|^2} = \sin(\theta_{\min}) \tag{7}$$

where $\theta_{\min}$ is the smallest angle between any pair of lines in the packing. The problem of finding algorithms to design packings for arbitrary m and N has been studied by many researchers in applied mathematics and information theory (see [36], [39], [40], etc.). The Rankin bound (see, for example, [23], [36], [39]) gives an upper bound on the minimum distance for line packings as a function of m and $N \ge m$ and is given by [23], [36]

$$\delta(\boldsymbol{W}) \le \sqrt{\frac{(m-1)N}{m(N-1)}}. \tag{8}$$

Another useful property of a packing besides the minimum distance is the density. To define the density of a line packing, consider a metric ball in $\mathcal{G}(m, 1)$. Let $\mathcal{P}_{\boldsymbol{v}}$ denote the line generated by a vector $\boldsymbol{v} \in \Omega_m$ (i.e., the column-space of the vector $\boldsymbol{v}$). The ball of radius γ in $\mathcal{G}(m, 1)$ around the line generated by $\boldsymbol{w}_i$ is defined as

$$\mathcal{B}_{\boldsymbol{w}_i}(\gamma) = \{\mathcal{P}_{\boldsymbol{v}} \in \mathcal{G}(m, 1) \colon d(\boldsymbol{v}, \boldsymbol{w}_i) < \gamma\}. \tag{9}$$

Note

$$\mathcal{B}_{\boldsymbol{w}_k}(\gamma) \cap \mathcal{B}_{\boldsymbol{w}_l}(\gamma) = \phi \tag{10}$$

for $k \neq l$ when $\gamma \le \delta(\boldsymbol{W})/2$ where ϕ is the empty set. Metric balls in $\mathcal{G}(m, 1)$ can be geometrically visualized as spherical caps on Ω_m.

Thus, the ball $\mathcal{B}_{\boldsymbol{w}_i}(\gamma)$ is the set of lines generated by all vectors on the unit sphere that are within a chordal distance of γ from any point in $\Omega_m \cap \mathcal{P}_{\boldsymbol{w}_i}$.

The normalized Haar measure on Ω_m introduces a normalized invariant measure μ on $\mathcal{G}(m, 1)$. This measure allows the computation of volumes of sets within $\mathcal{G}(m, 1)$ [38], and thus can be used to determine the percentage of $\mathcal{G}(m, 1)$ covered by the metric balls of a line packing, called the density of a line packing. The density of a line packing is defined as

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

where $\mathcal{B}(\delta(\boldsymbol{W})/2)$ is an arbitrarily centered metric ball of radius $\delta(\boldsymbol{W})/2$..

Closed-form expressions for the density of Grassmannian subspace packings are often difficult to obtain [23]. In the case of line packings, though, we have found a way to calculate the density exactly. The result is proved in Theorem 1.

Theorem 1: For any line packing in $\mathcal{G}(m, 1)$

$$\Delta(\boldsymbol{W}) = N\left(\delta(\boldsymbol{W})/2\right)^{2(m-1)}. \tag{11}$$

Proof: Let

$$\mathcal{C}_{\boldsymbol{w}_i}(\gamma) = \{\boldsymbol{v} \in \Omega_m : d(\boldsymbol{v}, \boldsymbol{w}_i) < \gamma\}.$$

Using our previous observation

$$\mu\left(\mathcal{B}_{\boldsymbol{w}_i}(\gamma)\right) = \frac{A(\mathcal{C}_{\boldsymbol{w}_i}(\gamma))}{A(\Omega_m)} \tag{12}$$

where $A(\cdot)$ is a function that computes area. It was shown in [22][4] that

$$\frac{A(\mathcal{C}_{\boldsymbol{w}_i}(\gamma))}{A(\Omega_m)} = \gamma^{2(m-1)}. \tag{13}$$

The result then follows. □

Theorem 1 provides insight into the rate at which the density grows as a function of the minimum distance. This result specifically verifies the asymptotic results in [23] for the one-dimensional subspace case.

The bound in Theorem 1 yields a new upper bound on the minimum distance of Grassmannian line packings. The Hamming bound on the maximum minimum distance achievable by a Grassmannian line packing of a fixed size N is the maximum radius of the metric balls before any two metric balls overlap.

Theorem 2: For any N line packing in $\mathcal{G}(m, 1)$

$$\delta(\boldsymbol{W}) \leq 2\left(\frac{1}{N}\right)^{1/(2(m-1))}. \tag{14}$$

Proof: This follows by using the Hamming bound on codesize [23],

$$N\mu(\mathcal{B}(\delta(\boldsymbol{W})/2)) \leq 1.$$ □

Bounds on the existence of line packings of arbitrary radius also follow from Theorem 1 using the Gilbert–Varshamov bound on codebook size. The Gilbert–Varshamov bound is obtained by finding the maximum number of metric balls of a desired minimum distance that can be packed without covering $\mathcal{G}(m, 1)$.

[4]Note that [22] evaluated the area ratio to derive the MISO outage probability of quantized beamformers.

Theorem 3: Let $N(m, \delta)$ be the maximum cardinality of a line packing in $\mathcal{G}(m, 1)$ with minimum distance δ. Then

$$\delta^{-2(m-1)} \leq N(m, \delta) \leq (\delta/2)^{-2(m-1)}. \tag{15}$$

Proof: The Gilbert–Varshamov bound applied to line packing says that a packing of size $N = M + 1$ exists when $M\mu(\mathcal{B}(\delta)) < 1$ [23]. Using the fact that $\mu(\mathcal{B}(\delta)) = \delta^{2(m-1)}$, the Hamming bound, and solving for N gives (15). □

Finding the global maximizer of the minimum distance for arbitrary m and N is not easy either analytically or numerically [36]. For this reason, it is often most practical to resort to random computer searches; for example, see the extensive tabulations on [37] that have been computed for the real case. In some cases, closed-form solutions are available, e.g., when $N = 2m = p^\alpha + 1$, where p is prime and α is a positive integer, conference matrices allow explicit constructions of packings [39].

IV. Codebook Analysis and Design

In [14], [15] it is shown that an optimal beamforming vector for MRT systems is the dominant right singular vector of $\boldsymbol{H}$ with $\boldsymbol{H}$ defined as in Section II. Therefore, $\boldsymbol{w}_{\mathrm{MRT}}$ that satisfies (6) ($\mathcal{W} = \Omega_{M_t}$) is an optimal MRT solution. A restatement of this is that the optimal vector solves

$$\boldsymbol{w}_{\mathrm{MRT}} = \arg\max_{\boldsymbol{x}\in\Omega_{M_t}} |\boldsymbol{x}^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{x}|. \tag{16}$$

Recall that $\arg\max$ in this case (as mentioned in Section II) returns only one out of possibly many global maximizers. Therefore, it is important to note that if $\boldsymbol{w}_{\mathrm{MRT}}$ satisfies (16), then $e^{j\phi}\boldsymbol{w}_{\mathrm{MRT}}$ also satisfies (16) since

$$|\boldsymbol{w}_{\mathrm{MRT}}^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{w}_{\mathrm{MRT}}| = |e^{-j\phi}\boldsymbol{w}_{\mathrm{MRT}}^H\boldsymbol{H}^H\boldsymbol{H}e^{j\phi}\boldsymbol{w}_{\mathrm{MRT}}|.$$

Thus, the optimal beamforming vector obtained from (16) is not unique.

This property can be restated in terms of points on a complex line. Because of the properties of the absolute value function, if $\boldsymbol{w} \equiv \tilde{\boldsymbol{w}}$ (using the equivalence relation defined in Section III) then $\boldsymbol{w}$ and $\tilde{\boldsymbol{w}}$ are both global maximizers and thus provide the same performance. The authors in [24] recognized this point and used this result in designing the vector quantization algorithm for codebook design.

Let $\boldsymbol{H}$ be defined as in Section II with all entries independent. The distribution of $\boldsymbol{X} = \boldsymbol{H}^H\boldsymbol{H}$ is the complex Wishart distribution [41]. An important property of complex Wishart distributed random matrices that we need is summarized in Lemma 1.

Lemma 1 (James [41], Edelman [42]): If $\boldsymbol{X}$ is complex Wishart distributed, then $\boldsymbol{X}$ is equivalent in distribution to $\boldsymbol{U\Sigma U}^H$ where $\boldsymbol{U}$ is Haar distributed on the group of $M_t \times M_t$ unitary matrices and $\boldsymbol{\Sigma}$ has distribution commonly found in [42].

Thus a matrix of i.i.d. complex normal distributed entries is invariant in distribution to multiplication by unitary matrices. From this, it is easily proven that the complex Wishart distribution is invariant to transformation of the form $\boldsymbol{V}^H(\cdot)\boldsymbol{V}$ where $\boldsymbol{V} \in \mathcal{U}_{M_t}$ where $\mathcal{U}_{M_t}$ is the group of $M_t \times M_t$ unitary matrices. This is a trivial property in the case of the single transmit antenna distribution because of the commutativity of complex numbers, but this property has highly nontrivial implications for $M_t > 1$. A very important property of Haar distributed matrices that will be exploited later is given in the following lemma.

Lemma 2 (Marzetta and Hochwald [43]) : Let $\boldsymbol{U}$ be a Haar distributed $M_t \times M_t$ unitary random matrix. If $\boldsymbol{v} \in \Omega_{M_t}$ then $\boldsymbol{Uv}$ is uniformly distributed on Ω_{M_t}.

One solution to (16) has a distribution equivalent to $\boldsymbol{U}^H \boldsymbol{w}_{\mathrm{MRT}} = [1\ 0\ 0\ \cdots\ 0]^T$ or rather $\boldsymbol{w}_{\mathrm{MRT}} = \boldsymbol{U}[1\ 0\ 0\ \cdots\ 0]^T$ with $\boldsymbol{U}$ given in Lemma 2. Since $\boldsymbol{U}$ is Haar distributed on $\mathcal{U}_{M_t}$ and $[1\ 0\ 0\ \cdots\ 0]^T$ is a unit vector, Lemma 2 states that $\boldsymbol{w}_{\mathrm{MRT}} = \boldsymbol{U}[1\ 0\ 0\ \cdots\ 0]^T$ is distributed uniformly on Ω_{M_t}. It similarly follows that the columns of $\boldsymbol{U}$ and any unit norm linear combination of columns of $\boldsymbol{U}$ are uniformly distributed on Ω_{M_t}.

This result, taken along with Lemma 1, reveals a fundamental result about quantized beamforming systems that, until this point to the best of the authors' knowledge, has never been shown. The distribution of the optimal beamforming vector is independent of the number of receive antennas. Thus, *the problem of finding quantized beamformers for MISO systems is the same problem as that of finding quantized beamformers for MIMO systems.* Therefore, the MISO quantized beamforming analysis contained in [24], [25] is directly applicable to MIMO systems.

A corollary to Lemma 2 follows from observing that the optimal beamformer is actually defined by a line.

Corollary 1: The line generated by the optimal beamforming vectors for a MIMO Rayleigh-fading channel is an isotropically oriented line in $\mathbb{C}^{M_t}$ passing through the origin. Therefore, the problem of quantized transmit beamforming in a MIMO communication system reduces to quantizing an isotropically oriented line in $\mathbb{C}^{M_t}$.

To find an optimal codebook we need to define an encoding function and a distortion measure. The optimal transmit beamformer and receiver combiner maximize the receive SNR by maximizing the equivalent channel gain $\boldsymbol{\Gamma}_r$ in (3). Therefore, we use an encoding function at the receiver $\mathcal{Q}_{\boldsymbol{w}}: \mathbb{C}^{M_r \times \mathrm{M}_t} \rightarrow \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_N\}$ that selects the element of the codebook that maximizes the equivalent channel gain. Thus,

$$\mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H}) = \arg \max_{1 \le i \le N} \|\boldsymbol{H}\boldsymbol{w}_i\|_2^2. \tag{17}$$

Notice that this encoding function is not solely a function of the maximum singular value direction in the matrix channel case. The explanation is that situations arise where it is better to use the quantized vector that is close to some unit norm linear combination of the M_t singular vectors. For example, certain channels where all of the singular values are equal would fall into this case.

To measure the average distortion introduced by quantization, we use the distortion function

$$G(\boldsymbol{W}) = E_{\boldsymbol{H}}\left[\lambda_1 - \|\boldsymbol{H}\,\mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\|_2^2\right]$$

where λ_1 is the maximum eigenvalue of $\boldsymbol{H}^H\boldsymbol{H}$ and the effective channel gain for an optimal MRT beamformer. An upper bound is

$$\begin{aligned} G(\boldsymbol{W}) &= E_{\boldsymbol{H}}\left[\lambda_1 - \sum_{i=1}^{M_t} \lambda_i \left|\boldsymbol{u}_i^H \mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\right|^2\right] \\ &\le E_{\boldsymbol{H}}\left[\lambda_1 - \lambda_1 \left|\boldsymbol{u}_1^H \mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\right|^2\right] \\ &= E_{\boldsymbol{H}}[\lambda_1]\, E_{\boldsymbol{H}}\left[1 - \left|\boldsymbol{u}_1^H \mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\right|^2\right] \end{aligned} \tag{18}$$

where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{M_t} \ge 0$ and $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{M_t}$ are the eigenvalues and corresponding eigenvectors of $\boldsymbol{H}^H\boldsymbol{H}$. The inequality in (18) follows from the independence of the eigenvalues and eigenvectors of complex Wishart matrices [38], [41].

The intuition behind the bound in (18) is that the first factor is an indication of channel quality on average while the second factor is an indication of the beamforming codebook quality. Using the interpretation of $\boldsymbol{W}$ as a line packing and that $\boldsymbol{u}_1$ is uniformly distributed on Ω_{M_t}, it follows that

$$\Pr\left(1 - \left|\boldsymbol{u}_1^H \mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\right|^2 < \frac{\delta^2(\boldsymbol{W})}{4}\right) = \Delta(\boldsymbol{W}). \tag{19}$$

Thus, by (19) and Theorem 1

$$\begin{aligned} G(\boldsymbol{W}) &\le E_{\boldsymbol{H}}[\lambda_1]\left(\frac{\delta^2(\boldsymbol{W})}{4}\,\Delta(\boldsymbol{W}) + (1 - \Delta(\boldsymbol{W}))\right) && (20) \\ &= E_{\boldsymbol{H}}[\lambda_1]\left(1 + N\left(\frac{\delta(\boldsymbol{W})}{2}\right)^{2(M_t-1)}\left(\frac{\delta^2(\boldsymbol{W})}{4} - 1\right)\right). && (21) \end{aligned}$$

The bound in (20) was obtained by observing that there are two cases of the channel corresponding to if the line generated by $\boldsymbol{u}_1$ is or is not a member of a metric ball of one of the codebook lines. The line generated by $\boldsymbol{u}_1$ is in a metric ball with probability $\Delta(\boldsymbol{W})$. When the line is inside of a metric ball we know that $1 - |\boldsymbol{u}_1^H \mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})|^2 < \frac{\delta^2(W)}{4}$, but when the line is not in a metric ball we can only state the trivial bound that $1 - |\boldsymbol{u}_1^H \mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})|^2 \le 1$. These two cases and Theorem 1 then give (21). In conclusion, minimizing (21) corresponds to maximizing the minimum distance between any pair of lines spanned by the codebook vectors. Thus, we propose the following criterion for designing quantized beamforming codebooks.

Grassmannian Beamforming Criterion: Design the set of codebook vectors $\{\boldsymbol{w}_i\}_{i=1}^N$ such that the corresponding codebook matrix $\boldsymbol{W}$ maximizes

$$\delta(\boldsymbol{W}) = \min_{1 \le k < l \le N} \sqrt{1 - |\boldsymbol{w}_k^H \boldsymbol{w}_l|^2}.$$

This criterion captures the essential point about quantized beamforming codebook design for Rayleigh-fading MIMO wireless systems: *Grassmannian line packings* are the key to codebook construction. Thus, beamforming codebooks can be designed without regard to the number of receive antennas by thinking of the codebook as an optimal packing of lines instead of a set of points on the complex unit sphere.

One benefit of making the connection between codebook construction and Grassmannian line packing is that it provides an approach for finding good codebooks, namely, leveraging work that has already been done on finding optimal line packings. In the real case, this problem has been thoroughly studied and the best known packings are cataloged at [37]. For the complex case, the single-antenna noncoherent codes in [44] often have large minimum distances (see the discussion in [22]). Other times it is possible to find codebooks using analytical [39] or numerical [36], [40] methods. Some example codebooks are given in Appendix A in Tables I–V. Notice that when $N \le M_t$ maximally spaced packings are trivial: simply take N columns of any $M_t \times M_t$ unitary matrix. It follows that selection diversity represents a special form of quantized beamformer designed using the *Grassmannian beamforming criterion.*

Another advantage of the connection to Grassmannian line packing is that the bounds in Theorems 2 and 3 and the Rankin bound can be used to judge the quality of any given codebook. For example, for a given M_t and $N \ge M_t$, the Rankin bound gives a firm upper bound on $\delta(\boldsymbol{W})$. Unfortunately, in most cases, the Rankin bound is not attainable and in effect quite loose [39]. The Hamming bound given in Theorem 2 can be useful for large N but is looser than the Rankin bound for small N. These bounds will be of further use in Section V for determining rules of thumb on the selection of N to meet specific performance requirements.

A QMRT codebook designed according to the *Grassmannian beamforming criterion* uses a codebook matrix $\boldsymbol{W}$ given by

$$\boldsymbol{W} = \arg \max_{\boldsymbol{X} \in \mathcal{U}_{M_t}^N} \delta(\boldsymbol{X}). \tag{22}$$

Practical considerations such as hardware complexity often motivate imposing additional constraints on the elements of the codebook. The *Grassmannian beamforming criterion* is still applicable to the design of these constrained codebooks. Solving for the optimum beamforming vector, however, requires restricting the line packing matrix $\boldsymbol{W}$ to be an element of a class of constrained beamforming vectors $\mathcal{V}_{M_t}^N$ where $\mathcal{V}_{M_t}^N \subset \mathcal{U}_{M_t}^N$.

One popular constraint, as discussed in Section II, is to impose the requirement that every coefficient of $\boldsymbol{W}$ have the same constant modulus. In these QEGT systems [17], [20]

$$\mathcal{V}_{M_t}^N = \left\{ \boldsymbol{V} \in \mathcal{U}_{M_t}^N : \forall\, k,\, l,\, |V_{[k,l]}| = \frac{1}{\sqrt{M_t}} \right\}.$$

A Grassmannian beamforming QEGT codebook is then designed by solving

$$\boldsymbol{W} = \arg \max_{\boldsymbol{X} \in \mathcal{V}_{M_t}^N} \delta(\boldsymbol{X}). \tag{23}$$

Numerical optimization techniques (such as those in [40]) are often ineffective in designing QEGT codebooks. For this reason, random search based designs often yield codebooks with the best minimum distance. Suboptimal methods for designing QEGT codebooks were proposed in [17], [20] but often perform worse than QEGT codebooks designed using (23) (see *QEGT Experiment* in Section VI for an example). Other QEGT codebooks are available from the codebooks designed from the noncoherent codes in [44]. As stated earlier, the codebooks in [44] are often optimal or near optimal even for the unconstrained QMRT case. Therefore, *there is often no difference in performance between QMRT and QEGT when using codebooks designed with the Grassmannian beamforming criterion.*

Another constraint of interest is to use only antenna combinations that transmit on subsets of antennas. This corresponds to using beamforming vectors that pick a number $1 \le M \le M_t$ and then select the best M antennas to transmit on. Thus, we choose one of the nonempty members of the power set of $\{1, \ldots, M_t\}$ and transmit on this antenna subset. Generalized subset selection, as we call this transmission method, is a discrete system that can be represented via an M_t-bit codebook. If M_t is large, we might wish to use fewer than M_t bits for our generalized subset selection system. In this case, we would pick the codebook matrix $\boldsymbol{W}$ that satisfies

$$\boldsymbol{W} = \arg \max_{\boldsymbol{X} \in \mathcal{I}_{M_t}^N} \delta(\boldsymbol{X}) \tag{24}$$

where $\mathcal{I}_{M_t}^N$ is given by the set of matrices in $\mathcal{U}_{M_t}^N$ where each column can be represented as the normalized sum of unique column vectors of $\boldsymbol{I}_{M_t}$. Since $\mathcal{I}_{M_t}^N$ has finite cardinality, the global maximum to (24) can be obtained by performing a brute-force search over all matrices in $\mathcal{I}_{M_t}^N$. GSS codebooks provide better performance than selection diversity because additional vectors are included to allow a better quantization of the optimal beamforming vector.

V. Bounds on Codebook Size

Codebook size naturally has an impact on the performance of a quantized beamforming system. To obtain a good approximation of the optimal beamforming vector, it is desirable to choose N large. On the other hand, minimizing the required feedback motivates choosing N small. In this section, we derive the minimum value of N required to achieve full diversity order with codebooks designed using the *Grassmannian beamforming criterion.* We also find approximate lower bounds on N given an acceptable loss in capacity or SNR due to quantization. These bounds function similarly to the Gilbert-Varshamov bound in Theorem 3 using an approximation to the Rankin bound as the minimum distance.

A. Diversity Order

Closed-form results on the average probability of symbol error for quantized beamforming and combining systems are difficult if not impossible to determine. Therefore, we use the diversity order performance metric defined in Section II which is indicative of the high-SNR performance of various linear modulation schemes. The following theorem, proved in Appendix B, determines a bound on N that guarantees a diversity order of $M_r M_t$ for codebooks designed according to the *Grassmannian beamforming criterion* in Section IV assuming MRC at the receiver. The trick in proving the theorem is to recognize that the codebook matrix resulting from the *Grassmannian beamforming criterion* is guaranteed to be full rank. This full rank assumption can be trivially satisfied if a rank degenerate codebook with a K-dimensional null space is designed for $N \ge M_t$ by replacing the K vectors in the codebook that can be written as linear combinations of the other $(N-K)$ vectors with the K orthogonal vectors that span the null space (see Lemma 3).

Theorem 4: If $N \ge M_t$, then the *Grassmannian beamforming criterion* yields QMRT, QEGT, and GSS codebooks that have full diversity order.

Equality in Theorem 4 is achieved when $N = M_t$. In this case, the codebook matrix is simply a unitary matrix (i.e., $\boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{I}_{M_t}$) and, thus, the codebook is any set of orthonormal vectors. Unfortunately, it can be readily shown (using the unitary invariance of the Gaussian distribution) that this is equivalent to selection diversity. While such codebooks provably obtain full diversity order, choosing $N > M_t$ will more closely approximate the optimal MRT solution and result in a larger array gain.

B. Capacity

The capacity loss associated with using quantized beamforming is an important indicator of the quality of the quantization method. To determine this loss, we compare the capacity assuming perfect beamforming with the capacity assuming the use of quantized beamforming. Using this difference we derive a criterion for choosing N based on an acceptable capacity loss.

Consider the system equation in (1) with the scalar effective channel produced by beamforming $\boldsymbol{z}^H\boldsymbol{H}\boldsymbol{w}$ (recall that $\boldsymbol{w}$ and $\boldsymbol{z}$ are unit norm). With MRT, the ergodic capacity of this scalar fading channel is given by

$$C_{\text{unquant}} = E_{\boldsymbol{H}} \left[\log_2 \left(1 + \frac{\lambda_1 \mathcal{E}_t}{N_0} \right) \right] \tag{25}$$

where λ_1 is the maximum eigenvalue of $\boldsymbol{H}^H\boldsymbol{H}$ while with quantization it is given by

$$C_{\text{quant}} = E_{\boldsymbol{H}} \left[\log_2 \left(1 + \frac{\|\boldsymbol{H}\mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\|_2^2 \mathcal{E}_t}{N_0} \right) \right]. \tag{26}$$

Notice that we are computing the ergodic capacity of the equivalent fading channel and we are not attempting to fully optimize over the input distribution given partial channel information as in [24].

To compute a rule of thumb for choosing N based on a desired capacity loss, we approximate the quantized and unquantized capacity as

$$C_{\text{quant}} \approx E_{\boldsymbol{H}} \left[\log_2 \left(\frac{\|\boldsymbol{H}\mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\|_2^2 \mathcal{E}_t}{N_0} \right) \right] \tag{27}$$

and

$$C_{\text{unquant}} \approx E_{\boldsymbol{H}} \left[\log_2 \left(\frac{\lambda_1 \mathcal{E}_t}{N_0} \right) \right]. \tag{28}$$

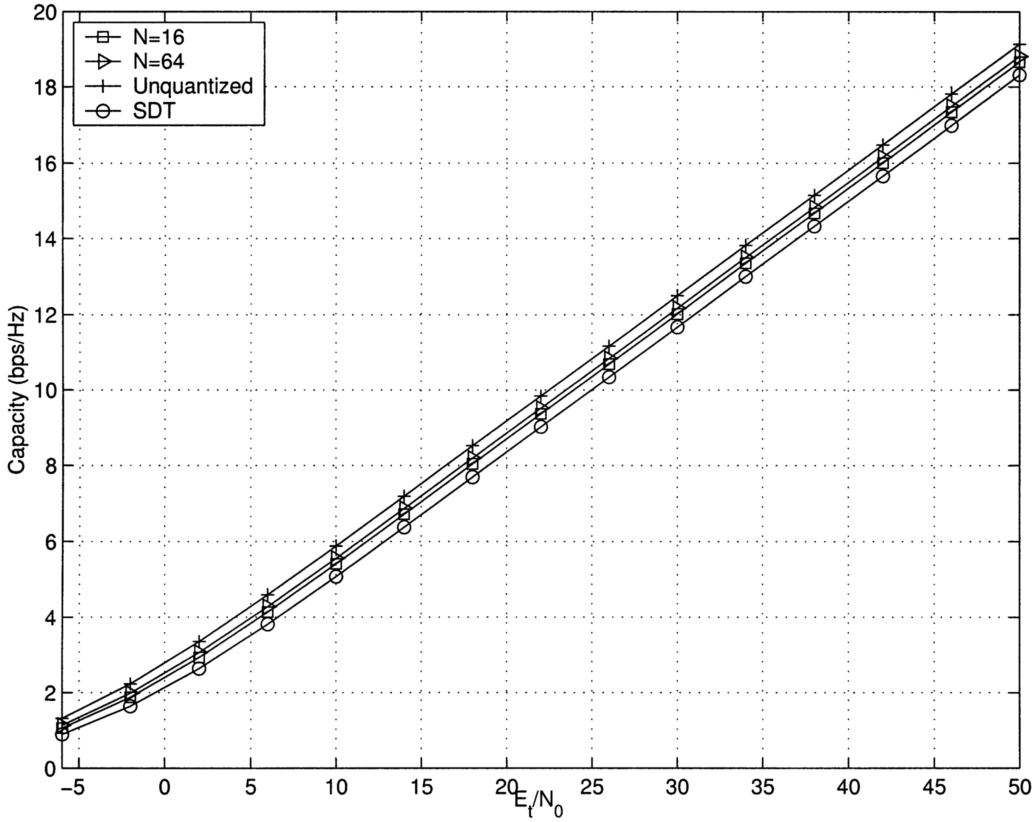

Fig. 2. Capacity comparison of unquantized beamforming with three Grassmannian beamforming schemes for a 4×2 system.

The capacity loss due to quantization for high SNR (using the techniques that bounded distortion) is given by

$$\begin{aligned} C_{\mathrm{loss}} &= E_{\boldsymbol{H}}\left[\log_2\left(1+\frac{\lambda_1 \mathcal{E}_t}{N_0}\right)\right] \\ &\quad - E_{\boldsymbol{H}}\left[\log_2\left(1+\frac{\|\boldsymbol{H}\,\mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})\|_2^2 \mathcal{E}_t}{N_0}\right)\right] \\ &\le E_{\boldsymbol{H}}\left[\log_2\left(1+\frac{\lambda_1 \mathcal{E}_t}{N_0}\right)\right] \\ &\quad - E_{\boldsymbol{H}}\left[\log_2\left(1+\frac{\lambda_1 \mathcal{E}_t}{N_0}\,|\boldsymbol{u}_1\,\mathcal{Q}_{\boldsymbol{w}}(\boldsymbol{H})|^2\right)\right] \qquad (29) \\ &\approx C_{\mathrm{unquant}}\left(1-N\left(\frac{\delta(\boldsymbol{W})}{2}\right)^{2(M_t-1)}\right) \\ &\quad - N\left(\frac{\delta(\boldsymbol{W})}{2}\right)^{2(M_t-1)}\log_2\left(1-\left(\frac{\delta(\boldsymbol{W})}{2}\right)^2\right) \qquad (30) \\ &\approx C_{\mathrm{unquant}}\left(1-N\left(\frac{\delta(\boldsymbol{W})}{2}\right)^{2(M_t-1)}\right). \qquad (31) \end{aligned}$$

The result in (29) follows from zeroing the other channel singular values, and (30) results from using the minimum-distance boundaries of the metric balls and the high-SNR approximation to C_{unquant}. Therefore, an approximate bound on the normalized capacity loss $\overline{C_{\mathrm{loss}}} = C_{\mathrm{loss}}/C_{\mathrm{unquant}}$ is given by $(1-N(\frac{\delta(\boldsymbol{W})}{2})^{2(M_t-1)})$.

Note that for the cases of large N, the Rankin bound on $\delta(\boldsymbol{W})$ in (8) in this case is approximately $\sqrt{\frac{M_t-1}{M_t}}$. Substituting for $\delta(\boldsymbol{W})$, we obtain a selection criterion (rule of thumb) on N based on capacity loss.

Capacity Loss Criterion: Given an acceptable normalized capacity loss $\overline{C_{\mathrm{loss}}}$, choose N such that

$$N \gtrsim \left(1-\overline{C_{\mathrm{loss}}}\right)\left(\frac{4M_t}{M_t-1}\right)^{M_t-1}. \qquad (32)$$

Equivalently, the corresponding number of bits of feedback ($b = \log_2 N$) should be chosen to be

$$b \gtrsim \log_2\left(1-\overline{C_{\mathrm{loss}}}\right) + 2(M_t-1) + (M_t-1)\log_2\left(\frac{M_t}{M_t-1}\right).$$

The last term corresponds to at most $M_t - 1$ thus, at most $3(M_t-1)$ bits of feedback or less are needed depending on the tolerable loss.

This bound is once again approximate, but it gives insight into the feedback amount required. Fig. 2 provides further intuition by showing a plot of the capacities for a 4×2 systems using unquantized beamforming and three different types of Grassmannian beamforming systems: QMRT with $N = 64$, QMRT with $N = 16$, and selection diversity. QMRT with $N = 64$ provides approximately a 1.5-dB gain compared to selection diversity and a 0.5-dB gain compared to QMRT with $N = 16$. This plot clearly shows the capacity benefit of increasing N.

C. SNR

Consider the expression for the normalized SNR loss $\overline{G}(\boldsymbol{W})$ obtained from (21)

$$\begin{aligned} \overline{G}(\boldsymbol{W}) &= \frac{G(\boldsymbol{W})}{E_{\boldsymbol{H}}[\lambda_1]} \\ &\le \left(1+N\left(\frac{\delta(\boldsymbol{W})}{2}\right)^{2(M_t-1)}\left(\frac{\delta^2(\boldsymbol{W})}{4}-1\right)\right). \qquad (33) \end{aligned}$$

Just as an approximate bound for N is derived in Section V-B given an acceptable capacity loss due to quantization, a criterion for choosing N, based on an acceptable normalized SNR loss $\overline{G}$ follows from (33). Substituting in the approximate Rankin bound of

$$\delta(\boldsymbol{W}) \lessapprox \sqrt{\frac{M_t-1}{M_t}}$$

we obtain the following approximate criterion.

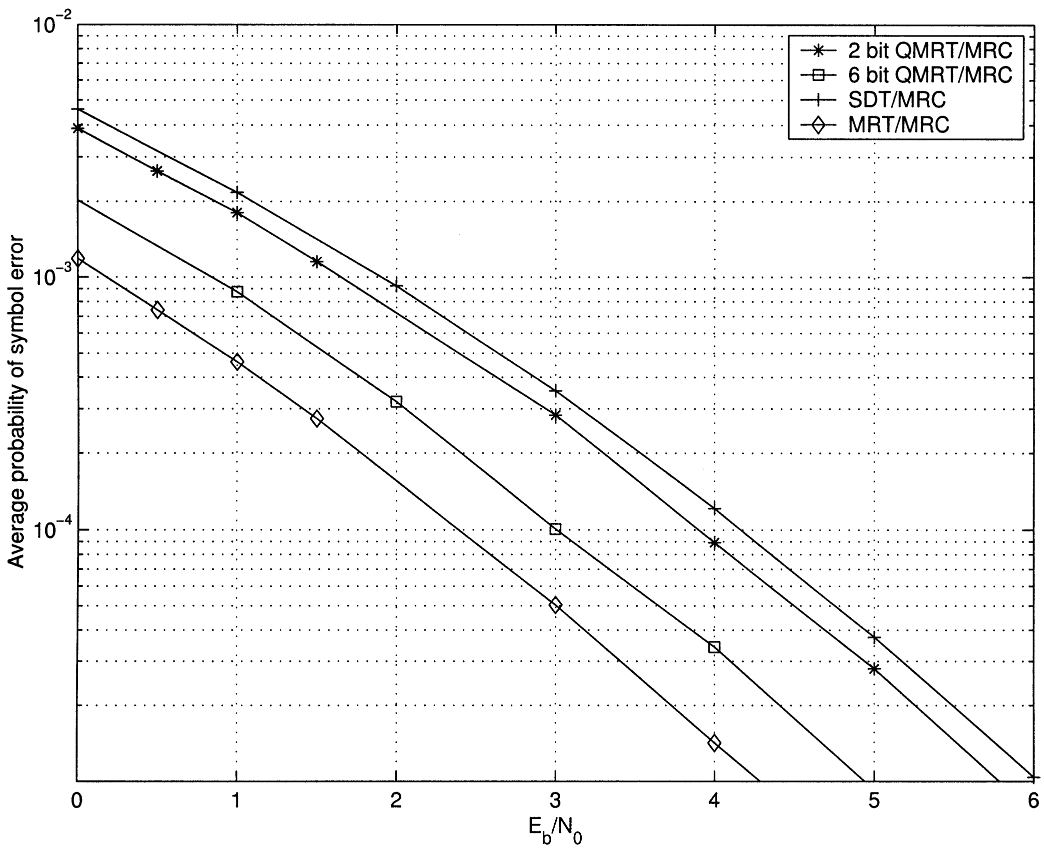

Fig. 3. Probability of symbol error for three transmit and three receive antenna systems using QMRT/MRC, SDT/MRC, and MRT/MRC.

SNR Criterion: Given an acceptable normalized SNR loss $\overline{G}$, choose N such that

$$N \gtrapprox \frac{1-\overline{G}}{\left(\frac{M_t-1}{4M_t}\right)^{M_t-1}\left(1-\frac{M_t-1}{4M_t}\right)}. \tag{34}$$

Once again, this bound is only approximate because it uses the Rankin bound approximation but yields intuition into the choice of N.

Following the analysis in Section V-B, the corresponding approximate number of bits of feedback $(b = \log_2 N)$ should be chosen to be

$$b \gtrapprox \log_2\left(1-\overline{G}\right) + 2(M_t-1) + (M_t-1)\log_2\left(\frac{M_t}{M_t-1}\right) - \log_2\left(1-\frac{M_t-1}{4M_t}\right). \tag{35}$$

In (35), at most $3(M_t-1) - \log_2(1-\frac{M_t-1}{4M_t})$ bits of feedback or less are needed depending on the tolerable loss.

As an aside, we should point out that $E_{\boldsymbol{H}}[\lambda_1]$ can be expressed in a closed-form integral expression using techniques from [15], [18], [45]. This is of particular interest if bounds on N were desired that were a function of an SNR loss that was not normalized. In [18], [45], the probability density function of the largest singular value of a central, complex Wishart distribution is derived, while the cumulative distribution function is derived in [15]. These results can also be used to derive integral expressions for the outage probability as a generalization of results in [22].

VI. Simulations

We simulate three different quantized beamforming schemes: quantized maximum ratio transmission, quantized equal gain transmission, and generalized subset selection. All simulations used binary phase-shift keying (BPSK) modulation and i.i.d. Rayleigh fading (where $H_{[k,l]}$ is distributed according to $\mathcal{CN}(0,1)$). The average probability of symbol error is estimated using at least 1.5 million iterations per SNR point. Codebooks for the QEGT and QMRT systems were designed based on the proposed *Grassmannian beamforming criterion.* The codebooks were found using the optimal constructions available in [39], [44]. GSS codebooks are globally optimal since searching over all possible codebooks is feasible. All of the simulations assume an MRC receiver.

QMRT Experiment 1: In the first experiment, an $M_r = M_t = 3$ system with QMRT is simulated with two different quantizations and the results shown in Fig. 3. The vectors in the 2–bit codebook are shown in Table IV in Appendix A. The 6-bit codebook has a maximum absolute correlation of 0.9399. The simulated error-rate curve of an optimal unquantized beamformer and the actual error-rate curve for a selection diversity system are shown for comparison. Notice that QMRT provides a 0.2-dB gain over selection diversity for the same amount of feedback. Using 6 bits instead of 2 bits of feedback provides around a 0.9-dB gain. The system using 6 bits performs within 0.6 dB of the optimal unquantized MRT system.

QMRT Experiment 2: In [24], vector quantization techniques were used to design QMRT codebooks. In this experiment, we compare a system using Grassmannian beamforming (i.e., quantized beamforming using a codebook designed with the *Grassmannian beamforming criterion*) with a system using a codebook designed by the Lloyd algorithm. The Grassmannian beamforming codebook is shown in Table III in Appendix A. Codebooks containing eight vectors were designed for an $M_r = M_t = 2$ system. The results are shown in Fig. 4. This simulation provides additional evidence of the validity of the proposed design criterion. Thus, we are able to design

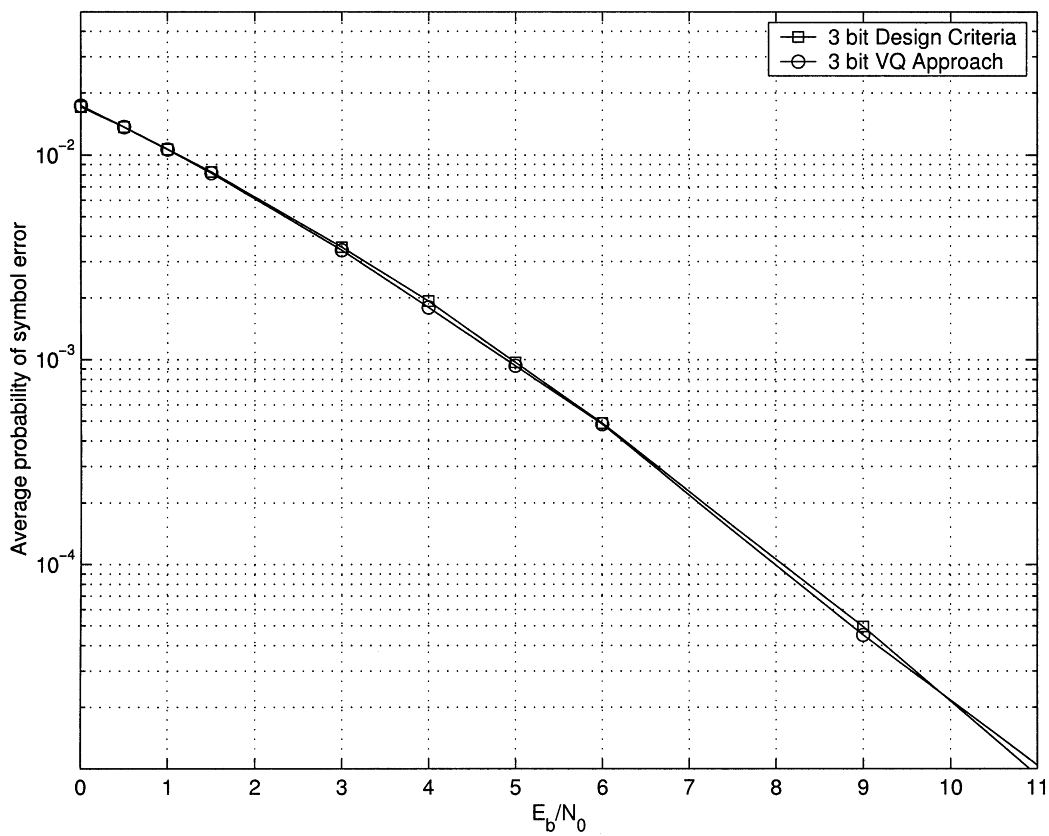

Fig. 4. Probability of symbol error for two transmit and two receive antenna systems using QMRT codebooks designed with the proposed criterion and with vector quantization.

codebooks that perform just as well as the codebooks performed using computationally complex vector quantization algorithms.

QEGT Experiment: In this experiment, two different methods of QEGT codebook design are compared on a three transmit and three receive antenna system. The results are shown in Fig. 5. The new method refers to codebooks constructed using the *Grassmannian beamforming criterion.* The 3-bit new codebook has maximum absolute correlation of 0.5774 and the 5-bit new codebook has maximum absolute correlation of 0.8836. The old method refers to the codebook design method outlined previously in [17], [20].

A 3-bit new design method QEGT codebook performs approximately the same as a 5-bit old design method QEGT codebook. Thus, we can use two fewer bits of feedback and actually maintain the average symbol error rate performance by using Grassmannian beamforming. Performance improves by 0.5 dB when changing from 3-bit new QEGT to 5-bit new QEGT. Thus, we can either gain 0.5 dB and use the same amount of feedback or keep the same performance and save 2 bits of feedback by using the *Grassmannian beamforming criterion*.

Comparison Experiment: The final experiment, shown in Fig. 6, compares GSS and QMRT codebooks for a four transmit and two receive antenna system. The 4-bit codebook has a maximum absolute correlation of 0.5817, while the 6-bit codebook has a maximum absolute correlation of 0.7973. A 4-bit QMRT system outperforms a 4-bit GSS system by approximately 0.5 dB. This illustrates that even a substantial restriction on the nature of the codebook does not severely impact performance when designed using the *Grassmannian beamforming criterion*. A 6-bit QMRT system has an array gain of approximately around 0.5 dB compared to a 4-bit QMRT system. This illustrates the benefits of increasing the amount of quantization even when a significant amount of quantization is already used.

VII. Conclusion and Future Work

In this correspondence, we derived a codebook design criterion for quantizing the transmit beamforming vectors in a MIMO wireless communication system. By bounding the SNR degradation for a given codebook size, we showed that the problem of designing beamformer codebooks is equivalent to Grassmannian line packing, which is the problem of maximally spacing lines in the Grassmann manifold. To approximate the feedback requirements, we used the Rankin bound along with several newly derived results for line packings such as a closed-form density expression, the Hamming upper bound on the minimum distance and codebook size, and the Gilbert–Varshamov lower bound on the codebook size.

A point that we did not address in detail pertains to implementation. Grassmannian beamforming will likely be implemented in a lookup table format. When the channel is slowly varying, it may be possible to reduce the necessary number of bits sent back by using some successive refinement techniques based on channel correlation. One solution is to have a series of codebooks for different values of N that support successive refinement along the lines of [46].

Another important point in a practical implementation is the effect of feedback error and delay in the feedback link. We did not address this issue in our work because we modeled the feedback link as error and delay free. An extensive simulation and/or analytical study of beamformer quantization such as that in [47] is needed. These effects will play an important role in performance in deployed MIMO systems using quantized beamforming.

Finally, one limitation of the work proposed here is that we considered only the transmission of a single data stream. It is well known, however, that MIMO channels can increase capacity by supporting the transmission of multiple data streams simultaneously ([1], [2], etc.). In

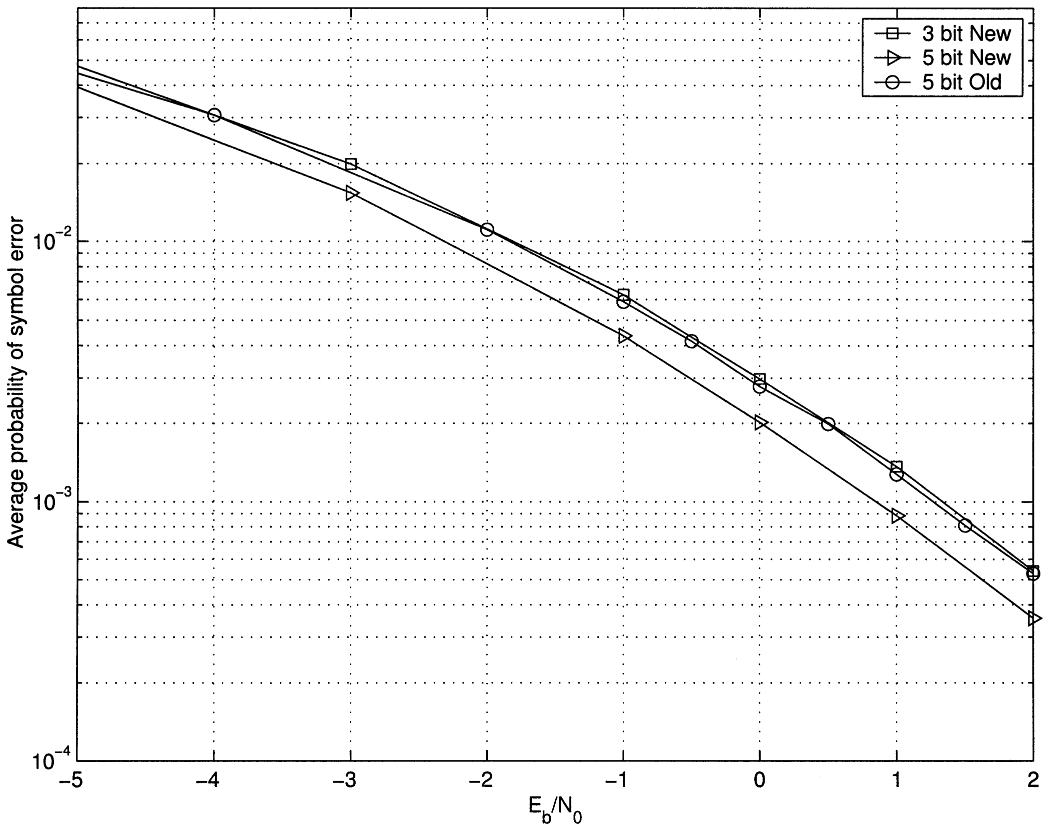

Fig. 5. Comparison of probability of symbol error for three transmit and three receive antenna systems using QEGT/MRC with the old and new codebook designs.

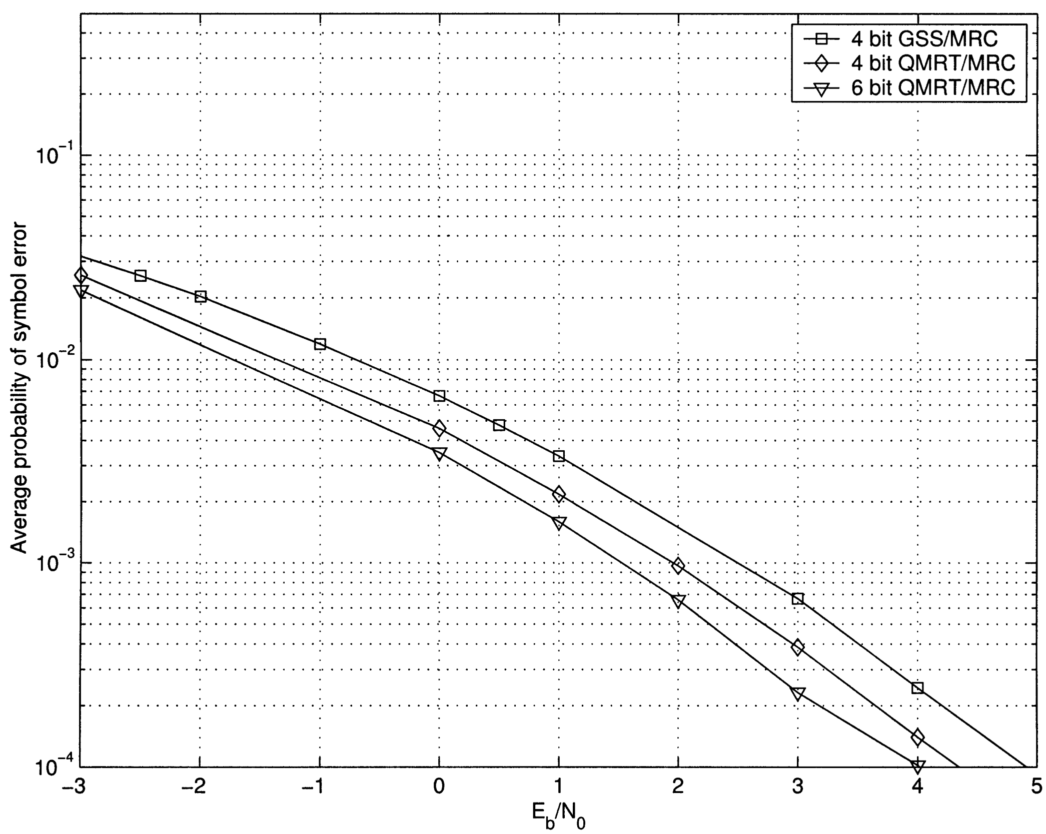

Fig. 6. Comparison of probability of symbol error for four transmit and two receive antenna systems using QMRT/MRC and GSS/MRC.

the general case with full channel knowledge at the transmitter, it is possible to transmit on multiple right singular vectors with the number of vectors and the power on each vector being determined by the desired optimization criterion. For example, the capacity-achieving solution is determined by waterpouring on the channel's nonzero singular values. A natural extension of our approach would be to derive codebooks for quantizing each of the singular vectors. While our Grassmannian codebooks could be used, they do not retain the orthogonality between the quantized singular vectors. Constructing codebooks for simultaneously quantizing multiple singular vectors is an interesting topic for future work.

APPENDIX

A. Tables of Line Packings

Examples of the packings found for various M_t are given in Tables I–V. The codebooks were found using random searches or through some of the constructions presented [39] depending on the choice of parameters.

B. Proof of Theorem 4

Before proving Theorem 4, we need to prove the following lemma that establishes that $\boldsymbol{W}$ is full rank for codes designed according to the *Grassmannian beamforming criterion.*

Lemma 3: If $N \geq M_t$ then $\boldsymbol{W}$ is full rank when designed using the *Grassmannian beamforming criterion* for QMRT, QEGT, or GSS.

Proof: Suppose that $N \geq M_t$ and all optimal maximum minimum-distance packings are not full rank. Let $\boldsymbol{W}$ be an optimal codebook matrix with a K-dimensional null space. Let $\boldsymbol{w}_{i_1}, \boldsymbol{w}_{i_2}, \ldots, \boldsymbol{w}_{i_{(N-K)}}$ be columns that form a basis for the column space of $\boldsymbol{W}$. Because the columns of $\boldsymbol{W}$ do not span $\mathbb{C}^{M_t}$, there exists an orthonormal basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_K$ for the null space. Then we can construct a new, full rank codebook matrix $\boldsymbol{X}$ with $\delta(\boldsymbol{W}) \leq \delta(\boldsymbol{X})$ by setting $\boldsymbol{x}_l = \boldsymbol{v}_l$ for $l \leq K$ and $\boldsymbol{x}_l = \boldsymbol{w}_{i_{(l-K)}}$ for $K < l \leq N$. This is a contradiction. We can therefore trivially construct a full rank codebook. □

Now we prove Theorem 4. This proof holds for any quantized beamforming technique (not just Grassmannian beamforming) that uses a codebook with at least M_t vectors and has a full rank codebook matrix.

Proof: First consider the receive SNR, $\mathcal{E}_t \boldsymbol{\Gamma}_r / N_0$. Since $\mathcal{E}_t$ and N_0 are assumed fixed, we only need to consider $\boldsymbol{\Gamma}_r = |\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|^2$. It has been shown for a fixed realization of $\boldsymbol{H}$ that the vectors $\boldsymbol{w}$ and $\boldsymbol{z}$ that maximize $|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{w}|$ are the left and right singular vectors of $\boldsymbol{H}$ corresponding to the largest singular value of $\boldsymbol{H}$ [14], [15]. This solution has been shown to achieve a diversity order of $M_r M_t$ [13], [15]. Quantized beamforming can perform only as well as the unquantized case, therefore, the achievable diversity order with quantized beamforming is upper-bounded by $M_r M_t$. To prove equality, we will now show that $M_r M_t$ is also the lower bound on the achievable diversity order.

For an N vector beamforming codebook system with MRC at the receiver, the effective channel gain is given by

$$\boldsymbol{\Gamma}_r = \max_{1 \leq i \leq N} \|\boldsymbol{H} \boldsymbol{w}_i\|_2^2. \tag{36}$$

Because the columns of the codebook matrix $\boldsymbol{W}$ span $\mathbb{C}^{M_t}$, $\boldsymbol{W}$ can be factored via a singular value decomposition (SVD) into the form $\boldsymbol{W} = \boldsymbol{U}_1[\boldsymbol{D}\ \boldsymbol{0}]\boldsymbol{U}_2$ where $\boldsymbol{U}_1$ is an $M_t \times M_t$ unitary matrix, $\boldsymbol{0}$ is an $M_t \times (N - M_t)$ matrix of zeros, $\boldsymbol{U}_2$ is an $N \times N$ unitary matrix, and $\boldsymbol{D}$ is a real diagonal matrix with $D_{[1,1]} \geq D_{[2,2]} \geq \cdots \geq D_{[M_t, M_t]} > 0$.

TABLE I
TRIVIAL CODEBOOK GENERATED FOR $M_t = 2$ AND $N = 2$ (1 BIT)

1 0	0 1

Absolute Max Correlation = 0
Absolute Theoretical Max Correlation = 0

TABLE II
CODEBOOK GENERATED FOR $M_t = 2$ and $N = 4$ (2 BITS)

-0.1612 - 0.7348j -0.5135 - 0.4128j	-0.0787 - 0.3192j -0.2506 + 0.9106j	-0.2399 + 0.5985j -0.7641 - 0.0212j	-0.9541 0.2996

Absolute Max Correlation = 0.57735
Absolute Theoretical Max Correlation = $\sqrt{1/3}$

TABLE III
CODEBOOK GENERATED FOR $M_t = 2$ and $N = 8$ (3 BITS)

0.8393 - 0.2939j -0.1677 + 0.4256j	-0.3427 + 0.9161j 0.0498 + 0.2019j	-0.2065 + 0.3371j 0.9166 + 0.0600j
0.3478 - 0.3351j 0.2584 + 0.8366j	0.1049 + 0.6820j 0.6537 +0.3106j	0.0347 - 0.2716j 0.0935 - 0.9572j
-0.7457 + 0.1181j -0.4553 - 0.4719j	-0.7983 + 0.3232j 0.5000 + 0.0906j	

Absolute Max Correlation = 0.84152
Absolute Theoretical Max Correlation = $\sqrt{3/7}$

TABLE IV
CODEBOOK GENERATED FOR $M_t = 3$ and $N = 4$ (2 BITS)

$1/\sqrt{3}$ $1/\sqrt{3}$ $1/\sqrt{3}$	$j/\sqrt{3}$ $-1/\sqrt{3}$ $-j/\sqrt{3}$	$-1/\sqrt{3}$ $1/\sqrt{3}$ $-1/\sqrt{3}$	$-j/\sqrt{3}$ $-1/\sqrt{3}$ $j/\sqrt{3}$

Absolute Max Correlation = 1/3
Absolute Theoretical Max Correlation = 1/3

TABLE V
CODEBOOK GENERATED FOR $M_t = 3$ and $N = 8$ (3 BITS)

$\frac{1}{\sqrt{2}}$ $\frac{1}{\sqrt{2}}$ 0	$\frac{1}{\sqrt{2}}$ 0 $\frac{1}{\sqrt{2}}$	0 $\frac{1}{\sqrt{2}}$ $\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}e^{2\pi j/3}$ $\frac{1}{\sqrt{2}}e^{4\pi j/3}$ 0
$\frac{1}{\sqrt{2}}e^{2\pi j/3}$ 0 $\frac{1}{\sqrt{2}}e^{4\pi j/3}$	$\frac{1}{\sqrt{2}}e^{4\pi j/3}$ 0 $\frac{1}{\sqrt{2}}e^{2\pi j/3}$	$\frac{1}{\sqrt{2}}e^{4\pi j/3}$ $\frac{1}{\sqrt{2}}e^{2\pi j/3}$ 0	0 $\frac{1}{\sqrt{2}}e^{4\pi j/3}$ $\frac{1}{\sqrt{2}}e^{2\pi j/3}$

Absolute Max Correlation = 0.5
Absolute Theoretical Max Correlation = $\sqrt{5/21}$

By the invariance of complex normal random matrices [41], $\boldsymbol{H}\boldsymbol{U}_1^H$ is equivalent in distribution to $\boldsymbol{H}$. Therefore,

$$\boldsymbol{\Gamma}_r \stackrel{d}{=} \tilde{\boldsymbol{\Gamma}}_r = \max_{1 \leq i \leq N} \|\boldsymbol{H}\boldsymbol{U}_1^H \boldsymbol{w}_i\|_2^2. \tag{37}$$

Now using matrix norm inequalities taken from [48], stated for the real case but easily seen to extend to the complex case, we find that

$$\tilde{\boldsymbol{\Gamma}}_r = \max_{1 \leq i \leq N} \|\boldsymbol{H}\boldsymbol{U}_1^H \boldsymbol{w}_i\|_2^2 \tag{38}$$

$$\geq \frac{1}{M_r} \|\boldsymbol{H}\boldsymbol{U}_1^H \boldsymbol{W}\|_1^2. \tag{39}$$

Using the SVD, (39) can be rewritten as

$$\tilde{\boldsymbol{\Gamma}}_r \geq \frac{1}{M_r} \|\boldsymbol{H}\boldsymbol{U}_1^H \boldsymbol{U}_1 [\boldsymbol{D}\boldsymbol{0}] \boldsymbol{U}_2\|_1^2$$

$$= \frac{1}{M_r} \left\| \left[\boldsymbol{H}\boldsymbol{D} \ \ \tilde{\boldsymbol{0}}\right] \boldsymbol{U}_2 \right\|_1^2 \tag{40}$$

where $\tilde{\boldsymbol{0}}$ is an $M_r \times (N - M_t)$ matrix of zeros.

By the matrix submultiplicative property [48]

$$\left\| \left[\boldsymbol{H}\boldsymbol{D} \ \ \tilde{\boldsymbol{0}}\right] \boldsymbol{U}_2 \right\|_1 \|\boldsymbol{U}_2^H\|_1 \geq \left\| \left[\boldsymbol{H}\boldsymbol{D} \ \ \tilde{\boldsymbol{0}}\right] \boldsymbol{U}_2 \boldsymbol{U}_2^H \right\|_1$$
$$= \|\boldsymbol{H}\boldsymbol{D}\|_1.$$

Then using an inequality property of the matrix one- and two-norm

$$\begin{aligned}\|\boldsymbol{HD}\|_1 &\leq \left\| \left[\boldsymbol{HD}\ \ \tilde{\mathbf{0}}\right] \boldsymbol{U}_2 \right\|_1 \sqrt{N} \|\boldsymbol{U}_2^H\|_2 \\ &= \sqrt{N} \left\| \left[\boldsymbol{HD}\ \ \tilde{\mathbf{0}}\right] \boldsymbol{U}_2 \right\|_1 \end{aligned} \tag{41}$$

or rather

$$\left\| \left[\boldsymbol{HD}\ \ \tilde{\mathbf{0}}\right] \boldsymbol{U}_2 \right\|_1^2 \geq \frac{1}{N} \|\boldsymbol{HD}\|_1^2. \tag{42}$$

Applying this bound we find that

$$\begin{aligned}\tilde{\Gamma}_r &\geq \frac{1}{NM_r} \|\boldsymbol{HD}\|_1^2 \\ &\geq \frac{D^2_{[M_t, M_t]}}{NM_r} \|\boldsymbol{H}\|_1^2 \\ &\geq \frac{D^2_{[M_t, M_t]}}{NM_r} \max_{i,j} |H_{[i,j]}|^2. \end{aligned} \tag{43}$$

The lower bound on $\tilde{\Gamma}_r$ is the effective channel gain of a system which selects the largest gain channel from among $M_r M_t$ i.i.d. complex Gaussian random variables with $D_{[M_t, M_t]} > 0$. Diversity systems of this form have been shown to achieve an $M_r M_t$ diversity order [28], [49]. The scale factor of $\frac{D^2_{[M_t, M_t]}}{NM_r}$ can simply be interpreted as a loss of array gain but not affecting the asymptotic diversity slope.

Combining the lower and upper bounds on diversity order, we have shown that at high SNR, quantized beamforming obtains a diversity order of $M_r M_t$. The guarantee of diversity order is an important performance indicator for the quantized beamforming system. Note that this proof also verifies the diversity results in [12], [13], [17]. □

REFERENCES

[1] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[3] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[6] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[7] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.

[8] R. W. Heath Jr. and A. J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Processing*, vol. 50, pp. 2429–2441, Oct. 2002.

[9] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. New York: Cambridge Univ. Press, 2003.

[10] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. New York: Cambridge Univ.Press, 2003.

[11] N. R. Sollenberger, "Diversity and automatic link transfer for a TDMA wireless access link," in *Proc. IEEE GLOBECOM*, vol. 1, Nov.–Dec. 1993, pp. 532–536.

[12] S. Thoen, L. Van der Perre, B. Gyselinckx, and M. Engels, "Performance analysis of combined transmit-SC/receive-MRC," *IEEE Trans. Commun.*, vol. 49, pp. 5–8, Jan. 2001.

[13] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, pp. 1458–1461, Oct. 1999.

[14] C.-H. Tse, K.-W. Yip, and T.-S. Ng, "Performance tradeoffs between maximum ratio transmission and switched-transmit diversity," in *Proc. 11th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communication*, vol. 2, Sept. 2000, pp. 1485–1489.

[15] P. A. Dighe, R. K. Mallik, and S. S. Jamuar, "Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, pp. 694–703, Apr. 2003.

[16] J. B. Anderson, "Antenna arrays in mobile communications: Gain, diversity, and channel capacity," *IEEE Antennas Propagat. Mag.*, vol. 42, pp. 12–16, April 2000.

[17] D. J. Love and R. W. Heath Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, pp. 1102–1110, July 2003.

[18] M. Kang and M.-S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 418–426, Apr. 2003.

[19] ——, "Performance analysis of MIMO MRC systems over Rician fading channels," in *Proc. IEEE Vehicular Technology Conf.*, vol. 2, Sept. 2002, pp. 869–873.

[20] D. J. Love and R. W. Heath Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," in *Proc. IEEE GLOBECOM*, vol. 2, Nov. 2002, pp. 1124–1128.

[21] G. Bauch and J. Hagenauer, "Smart versus dumb antennas-capacities and FEC performance," *IEEE Commun. Lett.*, vol. 6, pp. 55–57, Feb. 2002.

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

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

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

[25] R. W. Heath Jr. and A. Paulraj, "A simple scheme for transmit diversity using partial channel feedback," in *Proc. 32nd Annu. Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, Nov. 1998, pp. 1073–1078.

[26] H. Holma and A. Toskala, Eds., *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications, Revised Edition*. New York: Wiley, 2001.

[27] A. Wittneben, "Analysis and comparison of optimal predictive transmitter selection and combining diversity for DECT," in *Proc. IEEE GLOBECOM*, vol. 2, Nov. 1995, pp. 1527–1531.

[28] M. K. Simon and M.-S. Alouini, *Digital Communications Over Fading Channels*. New York: Wiley, 2000.

[29] K. K. Mukkavilli, A. Sabharwal, and B. Aazhang, "Finite rate feedback design for multiple transmit antennas," in *Proc. Allerton Conf. Communications, Control, and Computers*, Oct. 2002.

[30] K. K. Mukkavilli, A. Sabharwal, B. Aazhang, and E. Erkip, "Performance limits on beamforming with finite rate feedback for multiple antenna systems," in *Proc. 36th Annu. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, Nov. 2002, pp. 536–540.

[31] ——, "Beamformer design with feedback rate constraints: Criteria and constructions," in *Proc. Int. Symp. Information Theory*, Yokohama, Japan, July 2003.

[32] D. J. Love, "Transmit diversity quantization methods for multiple-input multiple-output wireless systems," M.S. thesis, Univ. Texas at Austin, May 2002.

[33] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Quantized antenna weighting codebook design for multiple-input multiple-output wireless systems," in *Proc. 40th Allerton Conf. Communications, Control, and Computing*, Moticello, IL, Oct. 2002.

[34] ——, "Quantized maximum ratio transmission for multiple-input multiple-output wireless systems," in *Proc. 36th Annu. Asilomar Conf. Signals, Systems and Computers*, vol. 1, Nov. 2002, pp. 531–535.

[35] ——, "Grassmannian beamforming for multiple-input multiple-output wireless systems," in *Proc. IEEE Int. Conf. Communications*, vol. 4, May 2003, pp. 2618–2622.

[36] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[37] N. J. A. Sloane. Packings in Grassmannian spaces. [Online] Available: http://www.research.att.com/~njas/grass/index.html

[38] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[39] T. Strohmer and R. W. Heath Jr., "Grassmannian frames with applications to coding and communications," *Appl. Comput. Harm. Anal.*, vol. 14, no. 3, pp. 257–275, to be published.

[40] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2618–2626, Sept. 2001.

[41] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, June 1964.

[42] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, MIT, Cambridge, MA, 1989.

[43] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

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

[45] C. G. Khatri, "Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations," *Ann. Math. Statist.*, vol. 35, pp. 1807–1810, Dec. 1964.

[46] B. C. Banister and J. R. Zeidler, "Tracking performance of a stochastic gradient algorithm for transmit antenna weight adaptation with feedback," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 5, May 2001, pp. 2965–2968.

[47] E. N. Onggosanusi, A. Gatherer, A. G. Dabak, and S. Hosur, "Performance analysis of closed-loop transmit diversity in the presence of feedback delay," *IEEE Trans. Commun.*, vol. 49, pp. 1618–1630, Sept. 2001.

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

[49] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw-Hill, 2001.

Rate–Diversity Tradeoff of Space–Time Codes With Fixed Alphabet and Optimal Constructions for PSK Modulation

Hsiao-feng Lu, *Student Member, IEEE,* and P. Vijay Kumar, *Fellow, IEEE*

***Abstract*—In this correspondence, we show that for any $(Q \times M)$ space–time code $\mathcal{S}$ having a fixed, finite signal constellation, there is a tradeoff between the transmission rate R and the transmit diversity gain ν achieved by the code. The tradeoff is characterized by $R \le Q - \nu + 1$, where Q is the number of transmit antennas. When either binary phase-shift keying (BPSK) or quaternary phase-shift keying (QPSK) is used as the signal constellation, a systematic construction is presented to achieve the maximum possible rate for every possible value of transmit diversity gain.**

***Index Terms*—Rate–diversity tradeoff, space–time codes.**

I. Introduction

Consider a space–time coded system with Q transmit and P receive antennas. Under the quasi-static Rayleigh fading assumption, the channel is fixed for a duration of M symbol transmissions. In general, we will assume $Q \le M$. Let $\mathcal{A}$ denote the signal alphabet (constellation) and $\mathcal{S} \subset \mathcal{A}^{QM}$ be a space–time code. Each codeword in the space–time code is thus a $(Q \times M)$ matrix.

Manuscript received July 18, 2002; revised June 17, 2003. This work was supported by the National Science Foundation under Grant CCR-0082987.

The authors are with the Department of Electrical Engineering–Systems, University of Southern California, Los Angeles, CA 90089-2565 USA (e-mail:hsiaofel;@usc.edu; vijayk@usc.edu).

Communicated by B. Hassibi, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.817469

Given the signal constellation $\mathcal{A}$, we follow [7] and define the rate R of a $(Q \times M)$ space–time code $\mathcal{S}$ by

$$R := \frac{1}{M} \log_{|\mathcal{A}|} |\mathcal{S}|. \tag{1}$$

Under this definition, a rate-one code corresponds to a space–time code of size $|\mathcal{A}|^M$, i.e., to a code which transmits on the average, one symbol from the signal constellation $\mathcal{A}$ per time slot. We say that a space–time code $\mathcal{S}$ achieves diversity gain $P\nu$ if the power-series expansion of the maximum pairwise-error probability (PEP) can be expressed as [4], [7]

$$\text{PEP} = c\rho^{-P\nu} + o(\rho^{-P\nu}) \tag{2}$$

where c is some constant independent of the signal-to-noise ratio (SNR) ρ. The quantity ν is termed the *transmit diversity gain* [1], [7]. It is shown in [7], that from the point of view of PEP, a space–time code $\mathcal{S}$ achieves transmit diversity gain ν if and only if for every $S_1 \ne S_2 \in \mathcal{S}$, the difference matrix $\triangle S = S_1 - S_2$ has rank at least ν over the field of complex numbers. In [4], Lu *et al.*showed that the transmit diversity gain equals ν even when one replaces the PEP criterion by either the codeword-error probability or else the symbol-error probability.

In this correspondence, we first show that for a fixed, finite-signal constellation $\mathcal{A}$ there is a tradeoff between the rate R and the transmit diversity gain ν of a space–time code $\mathcal{S}$. More specifically, given transmit diversity gain ν, the rate R is upper-bounded by

$$R \le Q - \nu + 1.$$

If either binary phase-shift keying (BPSK) or quaternary phase-shift keying (QPSK) is used as the signal constellation, i.e., $\mathcal{A} = \{\pm 1\}$ or $\mathcal{A} = \{\pm 1, \pm\sqrt{-1}\}$, we give a systematic code construction having rate R that achieves the upper bound $R = Q - \nu + 1$ for every $1 \le \nu \le Q$ and for any Q and M with $Q \le M < \infty$.

It should be noted that there is a distinction between the problem considered here and the one treated in Zheng and Tse [9]. In [9], the authors consider space–time codes which transmit at rates close to channel capacity, which calls for a signal alphabet that grows linearly with the logarithm of the SNR, whereas, we deal here with the more common situation of a fixed and finite-signal alphabet. There is also a difference in the definition of rate, the authors of [9] define rate in bits per channel use, i.e., they define the rate R' via

$$R' := \frac{1}{M} \log_2 |\mathcal{S}|.$$

Consequently, the authors of [9] arrive at a different tradeoff between rate and maximum-achievable diversity gain.

II. Rate–Diversity Tradeoff

We first show that when the signal constellation set $\mathcal{A}$ is finite, there is a tradeoff between the rate and transmit diversity gain of the space–time code.

Theorem 1: Given the desired transmit diversity gain ν and signal constellation $\mathcal{A}$ with $|\mathcal{A}| = a < \infty$, the size of the space–time code $\mathcal{S}$ is upper-bounded by

$$|\mathcal{S}| \le a^{M(Q-\nu+1)}. \tag{3}$$

Hence, the rate R has the upper bound

$$R \le Q - \nu + 1. \tag{4}$$

Proof: The difference matrix $\triangle S = S_1 - S_2$ between two distinct matrices S_1, S_2 drawn from $\mathcal{S}$ cannot have rank at least ν if the first $Q - \nu + 1$ rows of the two matrices S_1, S_2 are identical. It follows that the set $\mathcal{S}$ cannot have size larger than $|\mathcal{A}|^{M(Q-\nu+1)}$. The result follows. ■