Channel Feedback Quantization for High Data Rate MIMO Systems

Mehdi Ansari Sadrabadi, Amir K. Khandani, and Farshad Lahouti

Coding & Signal Transmission Laboratory
Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L 3G1

Technical Report UW-E&CE#2005-10

July 30, 2005

Channel Feedback Quantization for High Data Rate MIMO Systems

Mehdi Ansari Sadrabadi, Amir K. Khandani and Farshad Lahouti

Abstract

In this work, we study a multiple-input multiple-output (MIMO) wireless system where the channel state information is partially available at the transmitter through a feedback link. Based on singular value decomposition, the MIMO channel is split into independent sub-channels. Effective feedback of the required spatial channel information entails efficient quantization/encoding of a unitary matrix. We propose two schemes for quantizing unitary matrices via Givens matrices and examine the performance for a scenario where the rates allocated to the sub-channels are selected according to their corresponding gains. Numerical results show that the proposed schemes offer a significant performance improvement as compared to that of MIMO systems without feedback, with a negligible increase in the complexity.

Index Terms:MIMO wireless systems, singular value decomposition, Givens decomposition, matrix quantization

I. Introduction

In recent years, researchers have examined the transmission strategies for MIMO systems in which the transmitter and/or the receiver have full or partial knowledge of the channel state information (CSI). It is shown that the capacity is substantially improved through even partial CSI at the transmitter [1]. Subject to finite rate feedback, optimal MIMO signaling is studied in [2] [3]

This work is financially supported by Communications and Information Technology Ontario (CITO), Nortel Networks, and Natural Sciences and Engineering Research Council of Canada (NSERC). Mehdi Ansari and Amir K. Khandani are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (e-mail: {mehdi, khandani}@cst.uwaterloo.ca) Farshad Lahouti is with the Department of Electrical and Computer Engineering, Tehran University, Tehran, Iran. (e-mail:lahouti@ieee.org)

to maximize the average channel capacity, while precoder design for MIMO systems with linear receivers is investigated in [4].

Transmit beamforming can considerably improve the performance of MIMO systems [5]. Assuming partial CSI is available at the transmitter, the authors in [6] design a codebook of beamformer vectors to minimize the outage probability. Reference [7] addresses the problem of codebook design with partial CSI where the criterion is to maximize the received signal to noise ratio (SNR). A beamforming method is presented in [8] which relies on the method of [9] for the quantization of the channel spatial information (singular vectors of the channel matrix). In [10], the authors use the Givens parameters to represent the singular matrix of the channel in a slowly time-varying environment. The adaptive delta modulation is applied to quantize each parameter with a one-bit quantizer.

In this paper, assuming a block fading channel model, we consider the situation in which a MIMO channel is split into several independent sub-channels by means of singular value decomposition (SVD). In this scheme, the spatial information of the channel and the constellation index of each sub-channel is needed at the transmitter. The modulation format is selected to match the SNR on each sub-channel. We use Givens rotation to decompose the spatial information of the channel (a unitary matrix). We develop quantization methods by expressing the distortion function of the unitary matrix in terms of the Givens matrices using the first order approximation. The quantizer design and the optimum bit allocation among the quantizers are achieved based on the interference measure defined in Section III. The simulation results are presented in Section IV. Finally, Section V concludes the paper.

II. System Model

We consider an independent and identically distributed (*i.i.d.*) block fading channel model. For a MIMO system with M transmit and M receive antennas, the model leads to the following complex baseband representation of the received signal

$$\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{n}, \tag{1}$$

where **x** is the $M \times 1$ vector of the transmitted symbols, $\mathbf{H}$ is the $M \times M$ channel matrix, $\mathbf{W}$ is an $M \times M$ precoder matrix, **n** is the $M \times 1$ zero mean Gaussian noise vector with the autocorrelation $\sigma^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix, and **y** is the received signal. Matrix $\mathbf{H}$ consists of circularly symmetric complex Gaussian elements with zero mean and unit variance.

The SVD of $\mathbf{H}$ is defined as $\mathbf{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^*$, where $\mathbf{V}$ and $\mathbf{U}$ are the unitary matrices, $\mathbf{\Lambda}$ is a diagonal matrix [11], and $(.)^*$ denotes the hermitian of $(.)$. We assume that CSI is known to the receiver and a noiseless feedback link from the receiver to the transmitter is available. By the SVD of $\mathbf{H}$ at the receiver, $\mathbf{U}$ is computed, quantized and sent to the transmitter. The transmitter uses the quantized version of $\mathbf{U}$ as a precoder, i.e. $\mathbf{W} = \mathbf{U} + \mathbf{\Delta U}$, where $\mathbf{\Delta U}$ represents the quantization error.

$$\mathbf{y} = \mathbf{H}(\mathbf{U} + \mathbf{\Delta U})\mathbf{x} + \mathbf{n}. \tag{2}$$

The receiver multiplies the received vector $\mathbf{y}$ by $\mathbf{V}^*$,

$$\mathbf{r} = \mathbf{V}^*\mathbf{y} = \mathbf{\Lambda}\mathbf{x} + \mathbf{\Lambda}\mathbf{U}^*\mathbf{\Delta U}\mathbf{x} + \mathbf{n}. \tag{3}$$

We consider a case in which data is transmitted and received separately in each sub-channel with different rates and with equal energy. The power constraint of the transmitted signal is defined as $E(\mathbf{x}\mathbf{x}^*) = M\mathcal{E}\mathbf{I}$, where $\mathcal{E}$ is the energy per data stream and E represents the expectation. Under the assumption of continuous approximation [12], it can be shown that the use of equal energy maximizes the rate for a cubical shaping region (subject to a constraint on total energy) as follows. If C is a lattice code of reasonably large size, then the distribution of its points in N dimensional space is well approximated by a uniform continuous distribution over the shaping region $\mathtt{R}$ bounding the constellation (shaping region). This is called the continuous approximation [12]. With the assumption of the continuous approximation, the average power $P(C)$ of the constellation is approximately equal to the average power $P(\mathtt{R})$ of a continuous distribution that is uniform with $\mathtt{R}$ and zero elsewhere, and the number of constellation points is approximately proportional to the volume of the shaping region [12]. Cubical shaping region is a cube bounded between $-A_i$ and A_i along the ith dimension. For a cubical shaping region, we have

$$P(C) \simeq P(\mathtt{R}) = \frac{1}{3}\sum_{i=1}^{n} A_i^2 \leq P. \tag{4}$$

Therefore, with the assumption of the continuous approximation, maximizing the rate is equivalent with maximizing the volume of the shaping region,i.e.

$$V(\mathtt{R}) = \prod_{i=1}^{n} A_i \tag{5}$$

It is well known that maximizing the term in (5) subject to (4) is achievd by setting $A_1 = A_2 = ... = A_n$. Therefore, using equal energy is optimum in this sense.

At the receiver, a modulation scheme for each sub-channel is selected such that a target bit error rate (BER), P_b, is achieved. The indices of the corresponding modulation schemes are sent to the transmitter. The received SNR at the kth sub-channel is

$$SNR_k = \frac{\mathcal{E}\lambda_k^2}{\sigma^2 + \widehat{\sigma}_k^2}, \tag{6}$$

where $\widehat{\sigma}_k^2$ is the corresponding noise variance caused by the quantization error in the kth sub-channel. We consider a set of QAM modulation formats. The rate of the kth sub-channel, r_k, is computed such that $r_k = \max_{P(r,SNR_k)\leq P_b} r$, where $P(r, SNR)$ is the BER function of a QAM modulation scheme in terms of the rate r and SNR. An approximation formula for $P(r, SNR)$ is given in [13]. If none of the modulation formats meets the desired BER in a given sub-channel, no data stream is sent over that sub-channel.

III. Feedback Design: Channel Singular Matrix Quantization

Noting that the receiver detects the sub-channels separately, the quantizers are designed to minimize the interference between the sub-channels. The variance of the interference signal is

$$\begin{aligned} E(\|\mathbf{\Lambda}\mathbf{U}^*\mathbf{\Delta U}\mathbf{x}\|^2) &= \lambda^2 E\text{Tr}(\mathbf{U}^*\mathbf{\Delta U}\mathbf{x}\mathbf{x}^*\mathbf{\Delta U}^*\mathbf{U}) \\ &= \lambda^2 E\text{Tr}(\mathbf{\Delta U}\mathbf{\Delta U}^*\mathbf{x}\mathbf{x}^*) \\ &= \lambda^2\mathcal{E}E(\|\mathbf{\Delta U}\|^2), \end{aligned} \tag{7}$$

where $E(\mathbf{\Lambda}^2) = \lambda^2\mathbf{I}$ the first M columns of $\mathbf{U}$, $\mathbf{\Delta U}$ is the corresponding quantization error and Tr denotes the trace function. In deriving (7), we use the property that the singular values of a Gaussian matrix with *i.i.d.* entries are independent from the corresponding singular vectors [14]. In the following, we develop two methods to quantize a unitary matrix to minimize (7).

We consider Givens rotation which decomposes a unitary matrix to the minimum number of parameters ($n^2 - n$ parameters for an $n \times n$ matrix) [11]. An $n \times n$ unitary matrix $\mathbf{U}$ can be decomposed in terms of the products of Givens matrices [11], i.e.

$$\mathbf{U} = \prod_{k=1}^{n-1}\prod_{i=k+1}^{n}\mathbf{G}(k,i). \tag{8}$$

Each $\mathbf{G}(k,i)$ consists of two parameters, $c_{k,i}$ and $s_{k,i}$, where $c_{k,i}$ is in both the (k,k)th and the (i,i)th positions, $s_{k,i}$ is in the (k,i)th position and $-s_{k,i}^*$ is in the (i,k)th position. The other diagonal elements of the matrix $\mathbf{G}(k,i)$ are 1 and the remaining elements are zero. Since $\mathbf{G}(k,i)$ is a unitary matrix, then $|c_{k,i}|^2 + |s_{k,i}|^2 = 1$. In this work, we assume that the procedure of decomposing the unitary matrix is performed such that $c_{k,i}$ is real and non-negative (See Appendix A).

Theorem 1 *If* **U** *is a singular matrix derived from an $n \times n$ Gaussian matrix with i.i.d. entries, the set of Givens matrices $\mathbf{G}(k,i)$, $1 \leq k < i \leq n$, in (8) will be statistically independent of each other and the probability distribution function (PDF) of the elements of $\mathbf{G}(k,i)$ is*

$$p_{c_{k,i},\angle s_{k,i}}(c,\angle s) = p_{c_{k,i}}(c)p_{\angle s_{k,i}}(\angle s) = \frac{i-k}{\pi}c^{2(i-k)-1}, 0 \leq c \leq 1, \quad \angle s \in [-\pi,\pi]. \tag{9}$$

Proof: See Appendix A.

Based on the criterion presented for the quantizer design in (7), we define the distortion measure as follows

$$D(\mathbf{Q},\widehat{\mathbf{Q}}) = \frac{1}{2}E(\|\mathbf{Q}-\widehat{\mathbf{Q}}\|^2), \tag{10}$$

where $\widehat{\mathbf{Q}}$ is the quantized version of $\mathbf{Q}$ and $\mathcal{R}(.)$ is the real part of $(.)$. Using (8) and (10), we can easily derive the first order approximation of the distortion measure for unitary matrix **U** as follows

$$\begin{aligned}
D(\mathbf{U},\widehat{\mathbf{U}}) \quad &= E\mathrm{Tr}\left(\mathbf{I}-\mathcal{R}\left(\prod_{l=1}^{n-1}\prod_{j=l+1}^{n}\widehat{\mathbf{G}}(l,j)(\prod_{l'=1}^{n-1}\prod_{j'=l'+1}^{n}\mathbf{G}(l',j'))^*\right)\right)\\
&= E\mathrm{Tr}\left(\mathbf{I}-\mathcal{R}\left(\prod_{l=1}^{n-1}\prod_{j=l+1}^{n}(\widehat{\mathbf{G}}(l,j)-\mathbf{G}(l,j)+\mathbf{G}(l,j))(\prod_{l'=1}^{n-1}\prod_{j'=l'+1}^{n}\mathbf{G}(l',j'))^*\right)\right)\\
&\simeq E\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}\mathrm{Tr}\mathcal{R}\left(\prod_{l=1}^{k}\prod_{j=l+1}^{i-1}\mathbf{G}(l,j)\right.\\
&\left.\left(\mathbf{G}(k,i)-\widehat{\mathbf{G}}(k,i)\right)\prod_{j=i+1}^{n}\mathbf{G}(k,j)\prod_{l=k+1}^{n-1}\prod_{j=l+1}^{n}\mathbf{G}(l,j)(\prod_{l'=1}^{n-1}\prod_{j'=l'+1}^{n}\mathbf{G}(l',j'))^*\right)\\
&= E\mathrm{Tr}\left(\mathcal{R}\left(\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}\left(\mathbf{G}(k,i)-\widehat{\mathbf{G}}(k,i)\right)\mathbf{G}^*(k,i)\right)\right)\\
&= E\mathrm{Tr}\left(\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}\left(\mathbf{I}-\mathcal{R}(\widehat{\mathbf{G}}(k,i)\mathbf{G}^*(k,i))\right)\right)\\
&= \sum_{k=1}^{n-1}\sum_{i=k+1}^{n}D(\mathbf{G}(k,i),\widehat{\mathbf{G}}(k,i)).
\end{aligned} \tag{11}$$

We assess the accuracy of approximations in (11). For simplicity, we change the notation of Givens matrices as follows,

$$\mathbf{U} = \prod_{k=1}^{N} \mathbf{G}_k, \tag{12}$$

where $N = \frac{n^2-n}{2}$. We define $\delta_{k1} = c_k - \widehat{c}_k$, $\delta_{k2} = c_k - \widehat{c}_k$, $\delta_k = \sqrt{\delta_{k1}^2 + \delta_{k2}^2}$, and $\mathbf{\Delta}_k = (\mathbf{G}_k - \widehat{\mathbf{G}}_k)/\delta_k$. Note that $\mathbf{\Delta}_k$ is a unitary matrix. The distortion of the unitary matrix is re-written as follows

$$D(\mathbf{U}, \widehat{\mathbf{U}}) = E\mathrm{Tr}(\mathbf{I} - \mathcal{R}(\prod_{k=1}^{N} \widehat{\mathbf{G}}_k (\prod_{i=1}^{N} \mathbf{G}_i)^*)) \simeq \sum_{k=1}^{N} D(\mathbf{G}_k, \widehat{\mathbf{G}}_k). \tag{13}$$

We evaluate $Er_{D(\mathbf{U},\widehat{\mathbf{U}})}$, the approximation error in (13). $Er_{D(\mathbf{U},\widehat{\mathbf{U}})}$ consists of the terms with multipliers $(\mathbf{G}_k - \widehat{\mathbf{G}}_k)$ of order two or higher.

$$\begin{aligned} Er_{D(\mathbf{U},\widehat{\mathbf{U}})} = E\mathcal{R}\mathrm{Tr} \Bigg(& \sum_{m=1}^{N-1} \sum_{k=m+1}^{N} \prod_{i=1}^{m-1} \mathbf{G}_i (\widehat{\mathbf{G}}_m - \mathbf{G}_m) \prod_{j=m+1}^{k-1} \mathbf{G}_j (\widehat{\mathbf{G}}_k - \mathbf{G}_k) \prod_{l=k+1}^{N} \mathbf{G}_l (\prod_{r=1}^{N} \mathbf{G}_r)^* \\ & + ... + \prod_{i=1}^{N} (\widehat{\mathbf{G}}_k - \mathbf{G}_k)(\prod_{r=1}^{N} \mathbf{G}_r)^* \Bigg) \\ = E\mathcal{R}\mathrm{Tr} \Bigg(& \sum_{m=1}^{N-1} \sum_{k=m+1}^{N} \prod_{i=1}^{m-1} \mathbf{G}_i \delta_m \mathbf{\Delta}_m \prod_{j=m+1}^{k-1} \mathbf{G}_j \delta_k \mathbf{\Delta}_k \prod_{l=k+1}^{N} \mathbf{G}_l (\prod_{r=1}^{N} \mathbf{G}_r)^* \Bigg) \\ & + ... + E\mathcal{R}\mathrm{Tr} \left(\prod_{i=1}^{N} \delta_i \mathbf{\Delta}_i (\prod_{r=1}^{N} \mathbf{G}_r)^* \right). \end{aligned} \tag{14}$$

Using the fact that $\mathcal{R}(\mathrm{Tr}(\mathbf{W})) \leq n$, where $\mathbf{W}$ is an $n \times n$ unitary matrix and assuming $\delta_k \leq \delta$ for $1 \leq k \leq N$, the approximation error is bounded as follows

$$Er_{D(\mathbf{U},\widehat{\mathbf{U}})} \leq \sum_{k=2}^{N} \frac{N!}{(N-k)!k!} \delta^k.$$

For example, for $n = 4$, we have

$$Er_{D(\mathbf{U},\widehat{\mathbf{U}})} \leq 15\delta^2 + 20\delta^3 + 15\delta^4 + 6\delta^5 + \delta^6.$$

From our expriments, when the number of allocated bits to each Givens matrix is moderately high (4 bits), $Er_{D(\mathbf{U},\widehat{\mathbf{U}})}$ is negligible.

1) Method A: The parameters of $\mathbf{G}(k, i)$, $c_{k,i}$ and $\theta_{k,i} = \angle s_{k,i}$, are quantized as $\widehat{c}_{k,i}$ and $\widehat{\theta}_{k,i}$, independently, for $1 \leq k < i \leq n$. The matrix $\widehat{\mathbf{G}}(k, i)$ with the corresponding parameters $\widehat{c}_{k,i}$ and $\widehat{s}_{k,i}$ is constructed at the transmitter, using

$$\widehat{s}_{k,i} = \sqrt{1 - \widehat{c}_{k,i}^2} e^{j\widehat{\theta}_{k,i}}, \tag{15}$$

which forces $\widehat{\mathbf{G}}(k, i)$ to be a unitary matrix. Alternatively, one can quantize the underlying complex values using polar representation [15]. Using (15) and applying the first order approximation, we have (See Appendix B)

$$\|\mathbf{G}(k, i) - \widehat{\mathbf{G}}(k, i)\|^2 \simeq \frac{2}{1 - c_{k,i}^2}(c_{k,i} - \widehat{c}_{k,i})^2 + 2(1 - c_{k,i}^2)(\theta_{k,i} - \widehat{\theta}_{k,i})^2 \quad 1 \leq k < i \leq n \tag{16}$$

Substituting (16) in (11) and using (9), we have

$$D(\mathbf{U}, \widehat{\mathbf{U}}) \simeq \sum_{k=1}^{n-1} \sum_{i=k+1}^{n} E\left(\frac{(c_{k,i} - \widehat{c}_{k,i})^2}{1 - c_{k,i}^2}\right) + \frac{1}{2(i-k)+1} E(\theta_{k,i} - \widehat{\theta}_{k,i})^2. \tag{17}$$

Noting (17), we design Linde-Buzo-Gray (LBG) quantizers for $c_{k,i}$ and $\theta_{k,i}$ to minimize $E\left(\frac{(c_{k,i}-\widehat{c}_{k,i})^2}{1-c_{k,i}^2}\right)$ and, $E(\theta_{k,i} - \widehat{\theta}_{k,i})^2$, $1 \leq k < i \leq n$, respectively. The quantizer for $\theta_{k,i}$ follows the conventional approach to iterative design of a scalar LBG quantizer [16], while for parameter $c_{k,i}$ the iterative design procedure should use the following reconstruction value, $\widehat{c}_{k,i} = \frac{E(\frac{c_{k,i}}{1-c_{k,i}^2})}{E(\frac{1}{1-c_{k,i}^2})}$, which is easily derived by setting $\frac{\partial}{\partial \widehat{c}} E\left(\frac{(c-\widehat{c})^2}{1-c^2}\right) = 0$.

We utilize dynamic programming to find the optimum bit allocation among the quantizers. First, we design b-bit quantizers for $c_{k,i}$ and $\theta_{k,i}$, for $1 \leq k < i \leq n$ and $0 \leq b \leq B$. Then, we calculate $\mu^b(c_{k,i}) = E\left(\frac{(c_{k,i}-\widehat{c}_{k,i})^2}{1-c_{k,i}^2}\right)$, and $\mu^b(\theta_{k,i}) = \frac{1}{2(i-k)+1} E(\theta_{k,i} - \widehat{\theta}_{k,i})^2$ using the PDF of $c_{k,i}$ and $\theta_{k,i}$ given in 9, for $1 \leq k < i \leq n$ and $0 \leq b \leq B$. We use a trellis diagram with B states and $n^2 - n$ stages to allocate B bits to the quantizers corresponding to $c_{k,i}$ and $\theta_{k,i}$, $1 \leq k < i \leq n$. In the trellis diagram, each branch represents the difference between the number of bits corresponding to the two ending states on the branch. The metric of hte branch connecting the lth state at the $(j-1)$th trellis stage to the $(l+b)$th state at the jth trellis stage is $\mu^b(\vartheta_j)$, where ϑ_j is the quantization parameter corresponding to the jth stage. The search through the trellis determines the path with the minimum overall distortion and the corresponding number of bits for each parameter. The overall additive metric along a given trellis path is equal to the overall distortion given in (11). Examples of the bit allocation for the Givens parameters of a 3×3 unitary matrix is provided in Table I.

TABLE I

THE BIT ALLOCATION FOR DIFFERENT GIVENS PARAMETERS.

$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{2,3}$	$c_{1,2}$	$c_{2,3}$	$c_{1,3}$	Total bits
3	3	3	2	2	1	14
4	4	4	3	3	2	20

2) Method B: In this method, we quantize each Givens matrix as one unit. Let us define a new parameterization as follows: $c = \cos(\eta)$ and $s = e^{j\theta}\sin(\eta)$, where $0 \le \theta \le 2\pi$ and $0 \le \eta \le \pi$. We use the LBG algorithm to determine the regions and centroids of the two-dimensional quantizers corresponding to various (η, θ) for each Givens matrix. Using (10), the distortion function of a Givens matrix is

$$D = \sum_{m=1}^{T} \int_{R_m} \left(1 - \cos(\eta)\cos(\eta_m) + \sin(\eta)\sin(\eta_m)\cos(\theta - \theta_m)\right) p(\eta, \theta) d\eta d\theta, \qquad (18)$$

where R_m is the mth quantization region and T is the number of quantization partitions. The centroid (η_m, θ_m) is determined iteratively by minimizing the distortion function in the region R_m (See Appendix C)

$$\theta_m = \tan^{-1}(\frac{\varsigma_m}{\gamma_m}), \qquad (19)$$

$$\eta_m = \tan^{-1}\left(\frac{\sqrt{\varsigma_m^2 + \gamma_m^2}}{\int_{R_m} \cos^{l+1}(\eta)\sin(\eta) d\eta d\theta}\right), \qquad (20)$$

where $\gamma_m = \int_{R_m} \cos^l(\eta)\sin^2(\eta)\cos(\theta) d\eta d\theta$, $\varsigma_m = \int_{R_m} \cos^l(\eta)\sin^2(\eta)\sin(\theta) d\eta d\theta$, and $l = 2(i - k) - 1$, in the case of quantizing $\mathbf{G}(k, i)$ in (8). In this method, similar to the earlier case, a trellis diagram is used for the optimum bit allocation. The trellis diagram contains $\frac{n^2-n}{2}$ stages, each corresponding to a Givens component of an $n \times n$ unitary matrix, and B states where B is the total number of bits.

IV. Performance Evaluation

Fig. 1 shows the average bit rate versus SNR for different MIMO systems with $M = 3$ at the target BER$= 5 \times 10^{-3}$. Method B outperforms method A at the cost of a higher complexity for the codebook search. It is observed that the performance gain, compared to the gain of a 3×3 open loop MIMO system with the ML decoding, is noticeable. We also compare the performance of the proposed system with that of a V-BLAST system which is proposed as a solution to overcome the decoding complexity at the receiver. Note that as the receiver in the proposed method decodes the sub-channels separately, the decoding complexity is similar to that of the V-BLAST. Fig. 1 displays a significant improvement in comparison with the V-BLAST, showing the gain achieved through feedback.

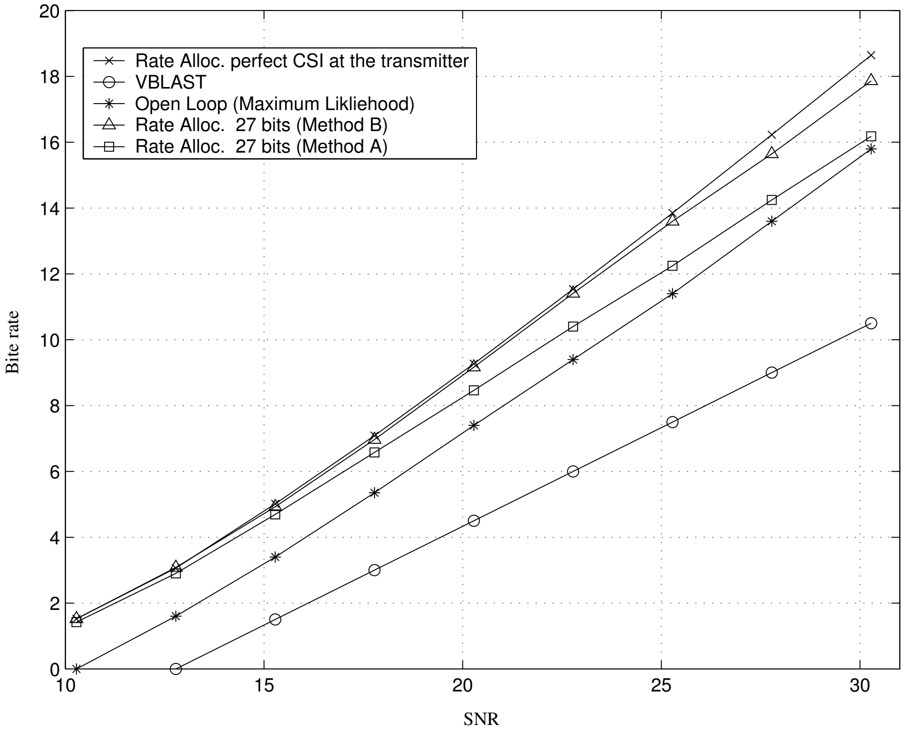

Fig. 1.   The average bit rate for different schemes where $M = 3$. The target BER= 5×10^{-3}.

In [9], the authors use Householder reflections to decompose an $n \times m, m \leq n$ unitary matrix into m unit-norm vectors with different dimensions, $q_1 \in S_n$, $q_2 \in S_{n-1}$, ..., $q_m \in S_{n-m+1}$, where $S_t = \{u \in \mathbb{C}^t : \parallel u \parallel = 1\}$. Then, vector quantization is applied to separately quantize q_1 to q_m. In [4], a method which has been proposed in [17] (to design unitary space-time constellations) is used to directly quantize the precoding unitary matrices.

We transmit two independent streams of 64-QAM symbols over the two sub-channels with the higher SNR and the third sub-channel is left empty. In Fig. 2, we plot the BER of this system using different quantization methods. In this scenario, the right singular matrix is fed back by 9 bits in each block. The bit allocation for different methods is shown in Table II, and the codebook

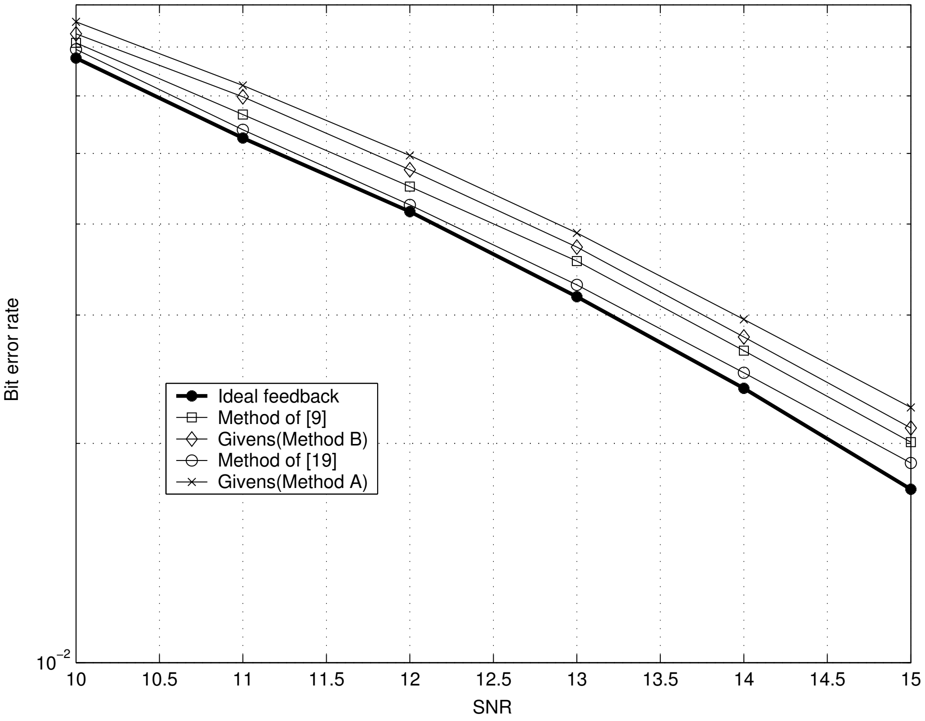

Fig. 2. The bit error rate for different schemes in a 3×3 MIMO system sending 2 64-QAM streams

search complexity of different quantization methods is compared in Table III. In order to search a codebook in the method proposed in [9], one needs to perform 32 vector multiplications of size 3, 32 norm calculations and 32 comparisons to select the corresponding q_1. Similarly, 16 vector multiplications of size 2, 16 norm calculations and 16 comparisons is needed to select the corresponding q_2. In the method used in [4], one needs to perform exhaustive search among 2^9, 3×3 matrices, requiring 2^9 matrix multiplications and 2^9 trace calculations. Note that the complexity of SVD, Givens rotations and Householder reflection is in the order of n^3 for an $n \times n$ matrix [11]. Although the method in [9] and the unitary space-time constellation design used in [4] outperform our quantization schemes, our proposed methods have a much lower complexity.

V. Conclusion

In this work, we have presented efficient methods for the channel information quantization in a MIMO system. We have developed efficient algorithms for the quantization of the underlying unitary matrices. Simulation results show a significant improvement as compared to a MIMO system without feedback, at the cost of a low-rate feedback link and a small increase in the com-

TABLE II

THE BIT ALLOCATION FOR METHOD A, METHOD B AND HOUSEHOLDER REFLECTION METHOD

$c_{1,2}$	$c_{1,3}$	$c_{2,3}$	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{2,3}$
1	1	1	2	2	2

$G(1,2)$	$G(1,3)$	$G(2,3)$
3	3	3

q_1	q_2
5	4

TABLE III

THE CODEBOOK SEARCH COMPLEXITY OF DIFFERENT METHODS ARE COMPARED.

	Givens(Method A)	Givens(Method B)	Householder [9]	Space-time Constel. [17]
Multiplications	18	72	768	9216
Additions	0	48	384	8704
Comparisons	18	24	48	512

putational complexity.

APPENDIX A

It is a simple matter to zero a specified entry in a vector by using a Givens matrix. Based on this fact, there is an iterative algorithm to find the Givens matrices of a unitary matrix. In each iteration step, a Givens matrix is multiplied by the resulting matrix of the last step to zero a specified entry of the product matrix. The process of decomposing $n \times n$ unitary matrix $\mathbf{U}$ into its first $n-1$ Givens component can be formulated as follows [11]:

$$\begin{aligned} &\mathbf{Q}^0 = \mathbf{U}, \\ &\mathbf{Q}^i = \mathbf{G}^*(1, i+1)\mathbf{Q}^{i-1}, \quad 1 \leq i \leq n-1 \end{aligned} \tag{21}$$

where the superscript represents the iteration number. The elements of $\mathbf{G}(1, i+1)$ are determined such that $q^i_{i+1,1} = 0, 1 \leq i \leq n-1$, where q^i_{kl} is the corresponding (k,l) element of $\mathbf{Q}^i$. Therefore, the first column of $\mathbf{Q}^{n-1}$ is zero except for the first element. We signify the parameters of $\mathbf{G}(1, i+1)$ by c_i and s_i, which can be easily computed regarding to the procedure in (21) as follows:

$$c_i = \frac{\sqrt{\sum_{l=1}^{i} |u_{l1}|^2}}{\sqrt{\sum_{l=1}^{i+1} |u_{l1}|^2}}, \tag{22}$$

and

$$s_i = \frac{-e^{-j\angle u_{11}} u_{(i+1)1}}{\sqrt{\sum_{l=1}^{i+1} |u_{l1}|^2}}, \tag{23}$$

where $u_{kl} = |u_{kl}|e^{j\angle u_{kl}}$ is the corresponding (k,l) element of $\mathbf{U}$. This process can be repeated for the following columns of the matrix $\mathbf{Q}^{n-1}$, until it becomes a diagonal matrix $\mathbf{D}_n$. Note that in this decomposition parameter c in each Givens matrix is real and non-negative. Suppose $\mathbf{U}$ is decomposed to its Givens components based on the above algorithm such that $\mathbf{U} = \prod_{k=1}^{n-1} \prod_{i=k+1}^{n} \mathbf{G}(k,i)\mathbf{D}_n$. Then, the SVD of the channel matrix can be written as follows

$$\begin{aligned} \mathbf{H} &= \mathbf{V}\boldsymbol{\Lambda} \left(\prod_{k=1}^{n-1} \prod_{i=k+1}^{n} \mathbf{G}(k,i)\mathbf{D}_n \right)^* \\ &= (\mathbf{V}\mathbf{D}_n^*)\, \boldsymbol{\Lambda} \left(\prod_{k=1}^{n-1} \prod_{i=k+1}^{n} \mathbf{G}(k,i) \right)^* . \end{aligned} \tag{24}$$

Note that diagonal matrices $\boldsymbol{\Lambda}$ and $\mathbf{D}_n$ are commutative and $\mathbf{V}\mathbf{D}_n^*$ is a unitary matrix. It can be inferred from (24) that the set of Givens matrices with the format we have introduced provide enough information to represent the required spatial information of the channel at the transmitter.

We are interested in the probability distribution of the singular matrices[1] of the complex Gaussian channel matrix in the space of $M(n)$, namely the group of $n \times n$ unitary matrices. It is known that such a random unitary matrix takes its values uniformly from $M(n)$ in the sense of the following property [18].

Theorem 2 *Let us assume that* $\boldsymbol{U}$ *is a singular matrix of a random Gaussian matrix. For all* $\boldsymbol{V} \in M(n)$*, the distribution of* $\boldsymbol{U}$ *and* $\boldsymbol{VU}$ *are the same.*

[1] The probability distribution of a matrix is the joint PDF of its elements.

Such a distribution is called the Haar distribution and the corresponding unitary matrices are called Haar unitary matrices [18]. We refer to this property as the right invariance property. Also, the left invariance property can be easily derived from Theorem 2.

By using the right invariance property, the joint probability density $p(u_{11}, u_{21}, ..., u_{n1})$ of the elements of the first column of **U** is [18]

$$p(u_{11}, u_{21}, ..., u_{n1}) = \frac{2\pi^n}{\Gamma(n)} \delta \left(1 - \sum_{l=1}^{n} |u_{l1}|^2\right), \tag{25}$$

where $\delta(.)$ is Kronecker delta function. By integrating (25) over the variables $u_{(k+1)1}, ..., u_{n1}$, we find the following expression for the joint probability density of elements $u_{11}, u_{21}, ..., u_{k1}$:

$$p(u_{11}, u_{21}, ..., u_{k1}) = K \left(1 - \sum_{l=1}^{k} |u_{l1}|^2\right)^{n-k-1}, \qquad k < n \tag{26}$$

where K is a constant. The joint probability density of the absolute values of the elements of the first column of **U** can be easily derived as follows:

$$\begin{aligned} p(|u_{11}|, |u_{21}|, ..., |u_{k1}|) &= \int p(u_{11}, u_{21}, ..., u_{k1}) |u_{11}||u_{21}|..|u_{k1}| d\theta_{11} d\theta_{21}..d\theta_{k1} \\ &= (2\pi)^k K \left(1 - \sum_{l=1}^{k} |u_{l1}|^2\right)^{n-k-1} \prod_{l=1}^{k} |u_{l1}|. \end{aligned} \tag{27}$$

To determine the probability density of c_i in (22), we define the parameters $v_0, v_1, .., v_i$ as follows:

$$v_0 = |u_{11}|, v_1 = |u_{21}|, ..., v_i = |u_{(i+1)1}|.$$

First, we compute the joint density function of c_i in (22) and $|u_{l1}|, 2 \leq l \leq i + 1$,

$$p(v_1, .., v_i, c_i) = \frac{p_{(|u_{21}|,..,|u_{(i+1)1}|)} \left(v_i, .., v_1, \sqrt{\frac{c_i^2 v_i^2}{1-c_i^2} - \sum_{l=1}^{i} v_l^2}\right)}{\text{Joc}_{|u_{21}|,..,|u_{(i+1)1}|}(v_1, .., v_i, c_i)}, \tag{28}$$

where Joc is as the Jacobian representation and can easily be calculated,

$$\begin{aligned} \text{Joc}_{|u_{21}|,...,|u_{(i+1)1}|}(v_1, v_2, ..., v_i, c_i) &= |\frac{dc_i}{dv_0}| \\ &= \frac{v_0(1 - c_i^2)2}{v_i^2 c_i}. \end{aligned} \tag{29}$$

If (27) and (29) are substituted into (28), then

$$p(|u_{21}|,...,|u_{(i+1)1}|,c_i) = (2\pi)^k K \left(1-\frac{|u_{(i+1)1}|^2}{1-c_i^2}\right)^{n-i-1} \frac{|u_{(i+1)1}|3}{(1-c_i^2)2} c_i \prod_{l=2}^{i+1} |u_{l1}|. \qquad (30)$$

The marginal distribution of c_i in (22) can be calculated from the joint distribution of $\{|u_{l1}|\}_{l=1}^{i+1}$ using (27),

$$p(c_i) = 2ic_i^{2i-1}. \quad 0 \leq c_i \leq 1 \qquad (31)$$

The joint distribution of $\{c_l\}_{l=1}^{i}$ is computed by using (22) and (27),

$$p(v_0, c_1, ..., c_i) = \frac{p(v_0, v_1, ..., v_i)}{\text{Joc}_{v_0,v_1,...,v_i}(v_0, c_1, ..., c_i)}, \quad i \leq n-1 \qquad (32)$$

where

$$\text{Joc}_{v_0,v_1,...,v_i}(v_0, c_1, ..., c_i) = \Pi_{l=1}^{i}(\frac{v_l c_l^{2(i-l)+3}}{v_0 2}). \qquad (33)$$

Substituting (27) and (33) in (32), we can write

$$p(|u_{11}|, c_1, ..., c_i) = (2\pi)^k K \frac{\left(1-\frac{|u_{11}|2}{\prod_{l=1}^{i} c_l 2}\right)^{n-i-1} |u_{11}|^{2i+1}}{\prod_{l=1}^{i} c_l^{2(i-l)+3}}, \qquad (34)$$

and then,

$$p(c_1, ..., c_i) = \int p(|u_{11}|, c_1, ..., c_i) d|u_{11}| = \prod_{l=1}^{i}(2lc_l^{2l-1}). \qquad (35)$$

The comparison of the joint distribution and the marginal distribution of $\{c_i\}_{i=1}^{n-1}$ in (35) and (31), respectively, implies that $\{c_i\}_{i=1}^{n-1}$ are statistically independent of each other.

In order to parameterize a Givens matrix in the format we stated, it is only necessary to have c_i and the angle of complex s_i (Note that $c_i^2 + |s_i|^2 = 1$). The probability distribution of the angles of the elements of a column of the Haar unitary matrix is uniformly distributed and independent [14]. Considering this argument and (23), we have

$$p(\angle s_i) = \frac{1}{2\pi}. \qquad -\pi \leq \angle s_i \leq \pi \qquad (36)$$

After $n-1$ step of the decomposition process, the first column of $\mathbf{Q}^{n-1}$, defined in (21), has one non-zero element, and $\mathbf{Q}^{n-1}$ is in the following format:

$$\mathbf{Q}^{n-1} = \begin{pmatrix} e^{j\angle u_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix}, \qquad (37)$$

where $\mathbf{V}$ is an $(n-1) \times (n-1)$ unitary matrix. In the following, we prove that $\mathbf{V}$ and the set $\{\mathbf{G}(1,i)\}_{i=2}^{n}$ are statistically independent. Since matrices $\{\mathbf{G}(1,i)\}_{i=2}^{n}$ are derived from the first column of $\mathbf{U}$, namely $\mathbf{u}_1$, it is sufficient to show that $\mathbf{V}$ and $\mathbf{u}_1$ are independent.

It is easy to show that the set of unitary matrices with a fixed first column in $M(n)$ and the group $M(n-1)$ are bijective. To show this, consider a given $n \times n$ unitary matrix $\mathbf{A}$ with a fixed first column $\mathbf{a}_1$, and an $n \times n$ unitary matrix $\mathbf{W} = \mathrm{diag}(1, \mathbf{Y})$, where $\mathbf{Y}$ is an arbitrary $(n-1) \times (n-1)$ unitary matrix. Using $\mathbf{B} = \mathbf{AW}$ and noting that $\mathbf{B}$ and $\mathbf{W}$ are invertible, we conclude that there exists a one-to-one correspondence between $\mathbf{B}$ and $\mathbf{W}$. On the other hand, according to Theorem 2, the probability distribution of $\mathbf{A}$ and $\mathbf{B}$ are the same.

Noting the above arguments, we conclude that the probability density of $\mathbf{U}$ conditioned on $\mathbf{u}_1$ is distributed uniformly in $M(n-1)$. This means the probability density of $\mathbf{V}$ in (37) conditioned on $\mathbf{u}_1$ is distributed uniformly in $M(n-1)$ and therefore is statistically independent of $\mathbf{u}_1$.

Similarly, the decomposition algorithm described in (21) is applied on $\mathbf{V}$ and all the statistical arguments about $\mathbf{U}$ can be extended to $\mathbf{V}$.

APPENDIX B

In this part, the first order approximation of $\|\mathbf{G} - \widehat{\mathbf{G}}\|^2$ is derived.

$$\begin{aligned}
\|\mathbf{G} - \widehat{\mathbf{G}}\|^2 &= 2(c-\widehat{c})^2 + 2|\sqrt{1-c^2}e^{j\theta} - \sqrt{1-\widehat{c}^2}e^{j\widehat{\theta}}|^2 \\
= 2(c-\widehat{c})^2 &+ 2|\sqrt{1-c^2}e^{j\theta} - \sqrt{1-c^2}e^{j\widehat{\theta}} + \sqrt{1-c^2}e^{j\widehat{\theta}} - \sqrt{1-\widehat{c}^2}e^{j\widehat{\theta}}|^2 \\
&= 2(c-\widehat{c})^2 + 2(1-c^2)|e^{j\theta} - e^{j\widehat{\theta}}|^2 + 2|e^{j\widehat{\theta}}|^2|\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2}|^2 \\
&\quad +4\mathcal{R}((e^{j(\theta-\widehat{\theta})} - 1)\sqrt{1-c^2}(\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2})) \\
&\simeq 2(c-\widehat{c})^2 + 2(1-c^2)(\theta-\widehat{\theta})^2 + \frac{2c^2}{1-c^2}(c-\widehat{c})^2 \\
&\quad +2(\theta-\widehat{\theta})^2\sqrt{1-c^2}(\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2}) \\
&\simeq 2(c-\widehat{c})^2 + 2(1-c^2)(\theta-\widehat{\theta})^2 + \frac{2c^2}{1-c^2}(c-\widehat{c})^2 \\
&= \frac{2}{1-c^2}(c-\widehat{c})^2 + 2(1-c^2)(\theta-\widehat{\theta})^2. \qquad (38)
\end{aligned}$$

The approximation error in (38) is

$$Er = \|\mathbf{G} - \widehat{\mathbf{G}}\|^2 - \frac{2}{1-c^2}(c-\widehat{c})^2 + 2(1-c^2)(\theta-\widehat{\theta})^2. \qquad (39)$$

Or equally,

$$Er = 2(1-c^2)(|e^{j\theta} - e^{j\widehat{\theta}}|^2 - (\theta - \widehat{\theta})^2) + 2(|\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2}|^2 - \frac{2c^2}{1-c^2}(c-\widehat{c})^2)$$
$$+4\mathcal{R}((e^{j(\theta-\widehat{\theta})} - 1)\sqrt{1-c^2}(\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2})) \quad (40)$$

The error can be bounded as follows

$$\begin{aligned}|Er| &\leq |2(1-c^2)(|e^{j\theta} - e^{j\widehat{\theta}}|^2 - (\theta - \widehat{\theta})^2)| + |2(|\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2}|^2 - \frac{2c^2}{1-c^2}(c-\widehat{c})^2)| \\ &\quad +4|\mathcal{R}((e^{j(\theta-\widehat{\theta})} - 1)|\sqrt{1-c^2}(\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2})) \\ &= |2(1-c^2)(4\sin^2(\frac{\theta-\widehat{\theta}}{2}) - (\theta - \widehat{\theta})^2)| + |2(|\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2}|^2 - \frac{2c^2}{1-c^2}(c-\widehat{c})^2)| \\ &\quad +8\sin^2(\frac{\theta-\widehat{\theta}}{2})\sqrt{1-c^2}(\sqrt{1-c^2} - \sqrt{1-\widehat{c}^2})).\end{aligned}$$

Using the Tylor series to expand $\sqrt{1-\widehat{c}^2}$ and $\sin^2(\frac{\theta-\widehat{\theta}}{2})$, we have

$$\begin{aligned}|Er| &\lesssim \frac{1-c^2}{6}(\theta-\widehat{\theta})^4 + \frac{2c|c-\widehat{c}|^3}{1-c^2} \\ &\quad +c\sqrt{1-c^2}(c-\widehat{c})(\theta-\widehat{\theta})^2.\end{aligned}$$

Appendix C

The joint probability of the parameters η and θ of the Givens matrix $\mathbf{G}(k,i)$ can be easily derived regarding to (9) as follows:

$$p(\eta,\theta) = \frac{i-k}{\pi}\sin(\eta)\cos(\eta)^{2(i-k)-1}. \quad (41)$$

We find the centroid (η_m, θ_m) in mth region for the LBG quantizer in Section III-.2, which is the minimizing point of the following function:

$$\begin{aligned}D_{R_m} &= \int_{R_m} D_m(G)p(\eta,\theta)d\eta d\theta \\ &= \int_{R_m}(1-(\cos(\eta)\cos(\eta_m)+\sin(\eta)\sin(\eta_m)\cos(\theta-\theta_m)))p(\eta,\theta)d\eta d\theta \quad (42)\end{aligned}$$

We find the centroid (η_m, θ_m) by forcing the partial derivatives of D_{R_m} to zero. The partial derivative of D_{R_m} respect to θ_m is,

$$\frac{\partial}{\partial\theta_m}D_{R_m} = \sin(\eta_m)\int_{R_m}(\sin(\theta_m)\cos(\theta) - \cos(\theta_m)\sin(\theta))\sin(\eta)p(\eta,\theta)d\eta d\theta. \quad (43)$$

By setting (43) to zero, we have,

$$\tan(\theta_m) = \frac{\int_{R_m} \sin(\eta)\sin(\theta)p(\eta,\theta)d\eta d\theta}{\int_{R_m} \sin(\eta)\cos(\theta)p(\eta,\theta)d\eta d\theta}. \tag{44}$$

The partial derivative of D_{R_m} respect to η_m is,

$$\frac{\partial}{\partial \eta_m} D_{R_m} = \int_{R_m} (\cos(\eta)\sin(\eta_m) - \sin(\eta)\cos(\eta_m)\cos(\theta-\theta_m))p(\eta,\theta)d\eta d\theta. \tag{45}$$

Similarly, we set (45) to zero to find η_m. Therefore, we have,

$$\tan(\eta_m) = \frac{\cos(\theta_m)\int_{R_m} \sin(\eta)\cos(\theta)p(\eta,\theta)d\eta d\theta + \sin(\theta_m)\int_{R_m} \sin(\eta)\sin(\theta)p(\eta,\theta)d\eta d\theta}{\int_{R_m} \cos(\eta)p(\eta,\theta)d\eta d\theta}. \tag{46}$$

By using the results in (41), (44) and (45), (20) can be easily derived.

References

[1] E. Vistosky and U. Madhow, “Space-Time Transmit Precoding with Imperfect Feedback ,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[2] R. S. Blum, “MIMO with limited feedback of channel state information,” *IEEE Int. Conf. Speech and Signal Processing*, vol. 4, pp. IV–89–92, April 2003.

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

[4] D. J. Love and R. W. Heath Jr. , “Limited feedback precoding for spatial multiplexing systems,” *Proc. IEEE Global Telecommun. Conf.*, pp. 1857–1861, Dec. 2003.

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

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

[7] David J. Love,Robert W. Heath, Jr., and Thomas Strohmer, “Grassmannian beamforming for Multiple-Input Multiple-Output systems,” *IEEE Trans. Inform. Theory*, vol. 49, Oct. 2003.

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

[9] J.C. Roh and B. D. Rao, “Channel feedback quantizationmethods for MISO and MIMO systems,” *IEEE Int. Symp. on PIMRC*, vol. 2, pp. 805 – 809, Sept. 2004.

[10] J. C. Roh and B. D. Rao, “An efficient feedback method for MIMO systems with slowly time-varying channels,” *IEEE Wireless Communication and Networking Conf.*, vol. 2, pp. 760 – 764, March 2004.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, third ed., 1996.

[12] G.D. Forney and L. Wei, “Multidimensional constellations- Part I: Introduction, figures of merit, and generalized cross constellations,” *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877 – 892, Aug. 1989.

[13] J. G. Proakis, *Digital Communication*. McGraw-Hill, 4th ed., 2000.

[14] V. L. Girko, *Theory of Random Determinants*. Kluwer Academic Publishers, 1990.

[15] Peter F. Swaszek and J. B. Thomas, “Multidimensional spherical coordinates quantization,” *IEEE Trans. Inform. Theory,*, vol. 29, pp. 570–576, July 1983.

[16] Allen Gersho and Robert M. Gray, *Vector Quantization and Vector Compression*. Kluwer Academic Publication, 1992.

[17] 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.

[18] F. Hiai and D. Petz, “The Semicircle Law, Free Random Variables and Entropy,” *American Mathematical Society*, vol. 77, 2000. Mathematical Surveys and Monographs.