Efficient Feedback for Precoder Design in Single- and Multi-User MIMO Systems

Natasha Dharamdial and Raviraj Adve
Dept. of Elec. and Comp. Eng., Univ. of Toronto,
10 King's College Road
Toronto, Ontario M5S 3G4, Canada

***Abstract*— Recently several researchers have proposed downlink precoding to enhance the performance of multiple-input, multiple-output (MIMO) systems. Precoding has been used, for example, to transmit multiple data streams to the same user, data streams to multiple users and/or to improve the diversity order on a single data transmission. These schemes, proposed for single- and multi-user scenarios, combine data streams to achieve a chosen optimality criterion. The precoding is based on some channel knowledge, requiring *feedback* of channel parameters from a mobile receiver to the transmitter. Clearly, minimizing this overhead (measured here in terms of an overall allocation for feedback bits) would be an important goal of any implementation. This paper develops a practical, flexible and efficient strategy for feedback of the channel eigen-modes, generally the most useful required information. An analysis-synthesis approach to channel quantization greatly reduces the number of parameters that must be conveyed. This scheme optimally distributes the allowed bit allocation over all feedback parameters. Distinguishing features from previous feedback schemes include the ability to handle both single- and multi-user situations, flexibility and simplicity in distributing the bit allocation and flexibility in meeting data rate/diversity demands made by individual users.**

I. Introduction

It is now well established that using a multiple-input multiple output (MIMO) system at either end of a wireless transmission allows for spatial diversity over fading channels and/or multiplexing multiple data streams to an individual user. While there exist schemes that do not require information about the transmission channel, in the single user case, system performance can be significantly enhanced given channel information [1]. More recently, there has been increasing focus on multiplexing to multiple users, requiring channel knowledge at the transmitter [2], [3]. In [3], the authors introduce a linear precoder that achieves the expected trade off between diversity and intra- and/or inter-user multiplexing. All these schemes depend on the availability of channel information, especially the channel eigenvalues and eigen-modes at the transmitter.

In wireless systems the uplink and downlink channels are generally non-reciprocal. By channel estimation techniques, it can be assumed that the receiver has full and accurate knowledge of the channel and can therefore employ optimal combining. The transmitter, however, does not have knowledge of its downlink channel and must be provided with information to allow proper precoding. Previous works on effective channel feedback have generally focused on two approaches - the design of a precoder codebook [4], [5] or feedback of the channel eigen-modes [6], [7].

The codebook approach requires predesign of a set of precoders and the feedback only of the index of best choice from this set. This is a very effective scheme requiring very limited feedback. However, it is not very flexible. The predesign, based on Grassmanian beamforming [5], are known only up to 8 transmit antennas. More restrictively the codebook must be designed for every possible combination of numbers of transmit and receive antennas and data streams being transmitted. Finally, these schemes have been restricted to the single user case with exponentially increasing complexity if extended to the multiuser case.

The more flexible scheme, though requiring greater feedback than Grassmanian beamforming, recognizes that most precoding schemes depend only on the channel right singular vectors and singular values. This feedback scheme uses the Givens decomposition of the right singular vectors to represent the necessary information by a reduced set of parameters. Roh [6] explores suboptimal scalar quantization with adaptive delta modulation for the single user case in slowly time-varying channels. In [7], a trellis diagram is used to allocate the available bits across the Givens parameters. However, the scheme appears rather complex and the authors indicate use of an incorrect parameter probability distribution.

In this paper, the principles of transform coding are applied to the feedback problem at hand. Transform coding is a special case of vector quantization within the larger family of analysis-synthesis coding schemes. The statistical distribution of each Givens parameter of interest is derived and used to allocate the available bits. Transform coding allows for bit allocation with very

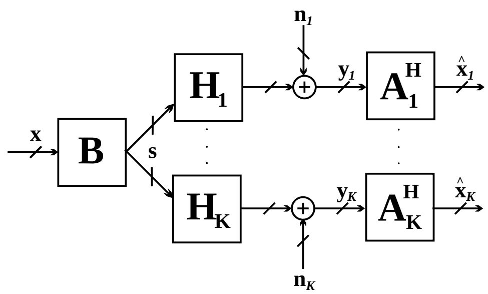

Fig. 1. The MIMO system under consideration.

low complexity, but with minimum distortion [8]. An interesting conclusion based on the analysis developed here is that, if a reasonably large number of bits are available, the simplest scheme allocating bits uniformly is close to optimal. The scheme presented here enables pre-coding in general, with data in both single- and multi-user scenarios. The scheme is tested in the single-user case using the design of [1] and in the multi-user case with the design of [3].

This paper is organized as follows. Section II briefly describes the MIMO system model and precoder designs used. Section III discusses the Givens decomposition and the quantization strategy used. Section IV presents results of simulations illustrating the performance of the proposed scheme. The paper ends with some conclusions in Section V.

II. MIMO System Model

The channel feedback scheme is developed in context of the multiuser scenario in Fig. 1, though may be used in other applications as well. This system, used to develop linear multiuser precoding in [3], reverts to the precoder of [1] if restricted to a single user. Consider a MIMO antenna system with N antennas at the transmitter servicing K users, each with $M_k, k = 1, \ldots, K$ receive antennas for a total of $M = \sum_{k=1}^{K} M_k$ antennas. The transmitter transmits L_k data streams over a flat-fading channel to the k-th user for a total of $L = \sum_{l=1}^{K} L_k$ streams. The entries of the $M_k \times N$ channel matrix $\mathbf{H}_k$ to each user and the additive noise are modelled as zero mean, unit variance independent and identically distributed (i.i.d.) complex Gaussian variables($CN(0,1)$). Each user uses the Wiener decoding matrix $\mathbf{A}_k$ to minimize the sum MSE) over its own data streams. The transmitter precoding matrix $\mathbf{B}$ is a $N \times L$ matrix designed to minimize the sum MSE over all data streams. In [1], [3], this matrix is derived assuming complete knowledge of all channels, $\mathbf{H}_k$, $k = 1, \ldots, K$ via feedback. In both cases, the precoder transmits data streams on eigen-modes of the channel matrix with power allocated by water-filling over the channel singular values. The goal of this paper is to enable efficient feedback of the needed information with limited complexity. We assume that each receiver can estimate its own channel without error.

Denote as $\mathbf{V}$ the $N \times N$ matrix of right singular vectors of the channel matrix $\mathbf{H}$. In the single user case, the optimal precoding and decoding matrices, $\mathbf{B}$ and $\mathbf{A}$ are given by [1]:

$$\mathbf{B} = \mathbf{V}_L \mathbf{\Phi}_b \quad (1)$$
$$\mathbf{A} = \mathbf{\Phi}_a \mathbf{V}^H \mathbf{H}^H \quad (2)$$

where $\mathbf{V}_L$ is the $N \times L$ matrix formed using the first L columns of $\mathbf{V}$ and $\mathbf{\Phi}_a$ is a diagonal matrix. The $L \times L$ diagonal power allocation matrix $\mathbf{\Phi}_b$ is obtained by waterfilling across the L largest channel singular values. This need for eigenvalue information is ignored in [7]. Based on Eqns. (1) and (2) the length-L soft-decision received signal is

$$\begin{aligned} \mathbf{y} &= \mathbf{AHBx} + \mathbf{n}, \\ &= \mathbf{\Phi}_a \mathbf{V}^H \mathbf{H}^H \mathbf{H} \mathbf{V}_L \mathbf{\Phi}_b \mathbf{x} + \mathbf{\Phi}_a \mathbf{V}^H \mathbf{H}^H \mathbf{n}, \\ &= \mathbf{\Phi}_a \mathbf{D} \mathbf{\Phi}_b \mathbf{x} + \mathbf{\Phi}_a \mathbf{V}^H \mathbf{H}^H \mathbf{n}, \end{aligned} \quad (3)$$

where $\mathbf{x}$ is the length-L vector of the L data streams transmitted simultaneously and $\mathbf{n}$ represents the additive noise. $\mathbf{D}$ is a diagonal matrix of channel singular values.

In the multiuser case, a similar equation applies, though needing an iterative solution. The precoder matrix $\mathbf{B}_k$, for user-k, forming L_k columns of the overall matrix $\mathbf{B}$ is given by [3]

$$\mathbf{B}_k = \mathbf{V}_{k,L_k} \mathbf{\Phi}_k \quad (4)$$

where $\mathbf{V}_{k,L_k}$ is the matrix formed using the L_k eigenvectors corresponding to the L_k largest eigenvalues of $\mathbf{R}_{H_k}$ given by

$$\mathbf{R}_{H_k} = \mathbf{H}_k^H \mathbf{R}_{(n+I)_k}^{-1} \mathbf{H} \quad (5)$$

$$\text{where} \quad \mathbf{R}_{(n+I)_k} = \mathbf{R}_{n_k} + \sum_{j=1, j\neq k}^{K} \mathbf{H}_k \mathbf{B}_j \mathbf{B}_j^H \mathbf{H}_k^H . \quad (6)$$

Note that in both the single- and multi-user cases each user must feed back only the L_k largest singular values and their corresponding singular vectors. The next section focuses on feedback of a limited number of columns of a unitary matrix and the associated eigenvalues.

III. Feedback Strategy

For ease of exposition, the feedback strategy is presented here in terms of the single user case. This strategy is the central theme of this paper. In a multiuser scenario, this strategy would be implemented by each user to feedback the same information.

A. Givens Rotations

As seen in Section II, it is only necessary to provide the transmitter with information about the strongest L eigen-modes. An efficient quantization of $\mathbf{V}_L$ can be accomplished by recognizing that any unitary matrix can be decomposed into a product of Givens rotations as [9]

$$\mathbf{V}_L = \left(\prod_{l=1}^{(2NL-L^2-L)/2} \mathbf{G}_l^H \right) \mathbf{R} \tag{7}$$

where the matrix $\mathbf{R}$ is an $N \times L$ matrix with phase rotations along the main diagonal and $\mathbf{G}_l$ are $N \times N$ unitary matrices, each of which zeroes an element below the main diagonal of $\mathbf{V}_L$. Each $\mathbf{G}_l$ can be characterized by two parameters, c and θ as follows

$$\mathbf{G}_l = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & c & |s|e^{j\theta} & \cdots & 0 \\ 0 & \cdots & -|s|e^{-j\theta} & c & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{8}$$

where c is a real number, $c \in (0,1)$ and $|s| = \sqrt{1-c^2}$. The position of the rotational component of $\mathbf{G}_l$ shifts along the main diagonal in order to zero elements in $\mathbf{V}_L$ below the main diagonal starting with the bottom left element and moving up each column sequentially. Using the Givens decomposition allows for the matrix $\mathbf{V}_L$ to be characterized by $\left(2NL - L^2\right)$ real numbers as opposed to NL complex (or $2NL$ real) numbers.

The role of the feedback scheme, described below, is to optimally assign the total available B bits for feedback to the $\left(2NL - L^2\right)$ parameters with minimum distortion. This bit budget is then used to quantize the parameters for feedback. The design of the precoder, and thus the quality of the received signal is directly affected by quantization of $\mathbf{V}_L$. If $\Delta\mathbf{V}_L$ represents error due to quantization, the received signal in Eqn. (3) becomes

$$\mathbf{y} = \Phi_a \mathbf{D} \Phi_b \mathbf{x} + \Phi_a \mathbf{D} \mathbf{V}^H \Delta\mathbf{V}_L \Phi_b \mathbf{x} + \Phi_a \mathbf{V}^H \mathbf{H}^H \mathbf{n} \tag{9}$$

where the second term represents the resulting error. To minimize the effect of this extraneous term on the decoding of the received symbols, it is desirable to minimize its variance. The sum variance of the error term can be expressed as

$$\begin{aligned} E\Big\{ \mathrm{Tr} \left[\left(\Phi_a \mathbf{D} \mathbf{V}^H \Delta\mathbf{V}_L \Phi_b \mathbf{x} \right) \left(\Phi_a \mathbf{D} \mathbf{V}^H \Delta\mathbf{V}_L \Phi_b \mathbf{x} \right)^H \right] \Big\} \\ \propto \quad E\left\{ \mathrm{Tr}\left[\Delta\mathbf{V}_L \Delta\mathbf{V}_L^H \right] \right\}, \end{aligned} \tag{10}$$

where $E\{\}$ represents the expectation and Tr the trace operator. In the next section, we will discuss a quantization strategy that minimizes this error.

B. Bit Allocation and Quantization

Due to the unitary nature of this matrix, we know that $\sum_{n=1}^{N} |\mathbf{V}_L(n,m)|^2 = 1$, i.e., the N random variables forming the m-th column of $\mathbf{V}_L$ lie on a N-dimensional sphere with unit radius. The marginal probability density function (PDF) of the real or imaginary part of each matrix element is uniform between (0,1). Using the fact that (a) the Givens decomposition zeros elements sequentially and (b) at each stage the power in the element zeroed out is added to that in the next element, the parameter θ is uniformly distributed between $(0, 2\pi)$ as is the angle associated with the phase rotation in $\mathbf{R}$. The parameter c_n, corresponding to introducing zero in row number n, has PDF

$$p(c_n) = 2nc_n(1 - c_n^2)^{n-1}, \tag{11}$$

where $n = 1$ corresponds to the bottom row, counting up and $c_n \in [0,1]$. Furthermore, all parameters are statistically independent. The details of the proof are omitted for lack of space. Finally, the PDF of the channel eigenvalues are known [10]. Since the parameters are not all identically distributed, it is desirable to allocate the feedback bits intelligently to minimize distortion.

The statistical independence of all parameters being quantized significantly simplifies the bit allocation problem, eliminating any need for a complicated joint vector quantization scheme. An optimal bit allocation scheme followed by scalar quantization significantly reduces feedback complexity [8]. This simplified vector quantization scheme is significantly less complex than the scheme proposed [7], but more effective than the scalar quantization in [6]. The bit allocation scheme minimizes the overall squared error distortion in Eqn. (10).

The overall distortion, D_{V_L}, can be expressed as a weighted sum of the distortion of each Givens matrix and the eigenvalues

$$D_{V_L} = \sum_{l=1}^{2NL-L^2-L} g_l d(\mathbf{G}_l) + \sum_{l=1}^{L} g_l d(\mathbf{R}_{ll}), \tag{12}$$

where $d(\mathbf{G}_l)$ represents the distortion to to the l-th Givens matrix and $d(\mathbf{R}_{ll})$ due to the l-th phase rotation. The weighting g_l recognizes the fact that not all Givens parameters equally impact the reconstructed unitary matrix. The first order approximation to the distortion of $\mathbf{G}_l$ is given by

$$d(\mathbf{G}) = d_c + (1-c^2) d_\theta = |c - \hat{c}|^2 + (1-c^2)|\theta - \hat{\theta}|^2, \tag{13}$$

where $\hat{c}$ and $\hat{\theta}$ are the quantized versions of c and θ respectively. Given an adequate bit budget, the optimal bit allocation to the i-th parameter, with PDF $f(x)$, is given by [8]

$$b_i = \bar{b} + \frac{1}{2}\log_2 \frac{\sigma_i^2}{\rho^2} + \frac{1}{2}\log_2 \frac{h_i}{H} + \frac{1}{2}\log_2 \frac{g_i}{G}, \quad (14)$$

$$h_i = \frac{1}{12}\left[\int_{-\infty}^{\infty} f_i^{1/3}(x)dx\right]^3, \quad (15)$$

where σ_i^2 is the variance of the parameter, ρ, H and G are the geometric means of the σ_i, h_i and g_i parameters, respectively and $\bar{b} = B/\left(2NL - L^2\right)$ the average number of bits per parameter. Given the bit allocation, the Lloyd algorithm sets the quantization levels.

To summarize the feedback scheme, the receiver estimates the channel matrix and decomposes it into its singular values right singular vectors. It then decomposes the singular vectors into the Givens components. The available bits are allocated using Eqn. (14) and the parameters quantized. This information is then fed back to the transmitter.

An important consideration in the scheme above is that with the $(1 - c^2)$ term in the distortion equation of Eqn. (13), the bit allocation and quantization levels are dependent on the instantaneous channel realization. However, as shown in Section IV, the impact of this term is negligible and the bit allocation can be made based on the (known a priori) statistical distribution.

IV. PERFORMANCE

In this section we evaluate the results of applying the bit allocation and quantization scheme to serve as feedback for precoder design in a MIMO system. All simulations use BPSK for data modulation. Due to a low statistical variance, it was found that eigenvalues could each be quantized with an allocation of 1 bit and provide adequate information for waterfilling. All eigenvalues were quantized in this manner.

A. *Effect of* $(1 - c^2)$ *term*

In Eqn. (13), the relative distortion between the c and θ parameters are weighted by a factor $(1 - c^2)$ introducing significant complexity. Ideally, the bit allocation scheme should depend only on the antenna configuration and number of bit streams being transmitted.

Figure 2 illustrates the effect of excluding the $(1-c^2)$ parameter from the bit allocation scheme. The figure plots the bit error rate (BER) versus signal-to-noise ratio (SNR) for several values of B, the total bit allocation. Clearly, the impact of the $(1 - c^2)$ term is negligible. Furthermore, with $(2NL - L^2) = 21$ parameters to be quantized, an average of approximately 4 bits per parameter yields close-to-ideal performance.

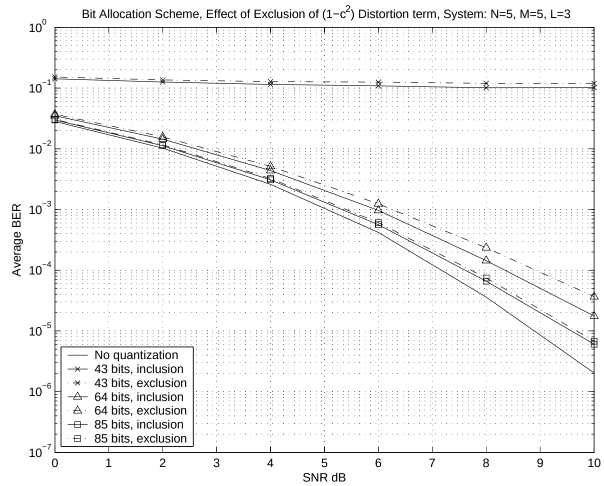

Fig. 2. Impact of excluding the $(1 - c^2)$ term

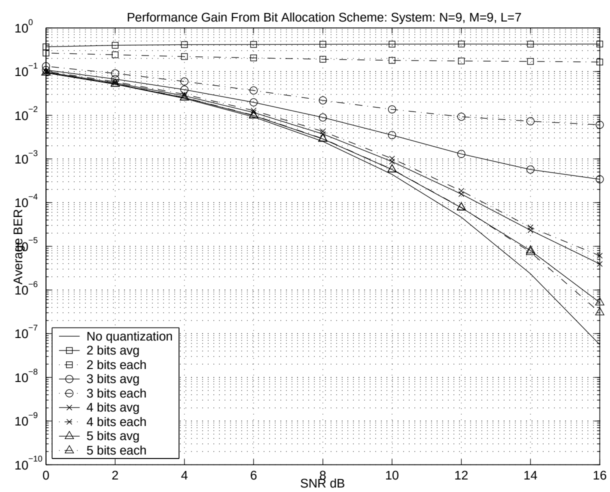

Fig. 3. Performance of Bit Allocation Scheme

B. *Bit Allocation*

In this simulation the performance of the bit allocation scheme is compared to procedure of allocating an equal number of bits to each of the $2NL - L^2$ Givens parameters, illustrating the importance of optimal bit allocation with low bit budgets. Figure 3 shows the results of this comparison. It was found if an average of 4 or more bits are allocated to each parameter, equally allocating bits performs as effectively as the optimal scheme of Eqn. (14).

C. *Number of Feedback Bits Required*

This simulation examines the achievable bit error rates when varying the number of feedback bits at fixed SNR values. Figure 4 shows the results for a system of $N = 5$ antennas at each end of the wireless link, with $L = 3$ bitstreams. From the results shown, to achieve

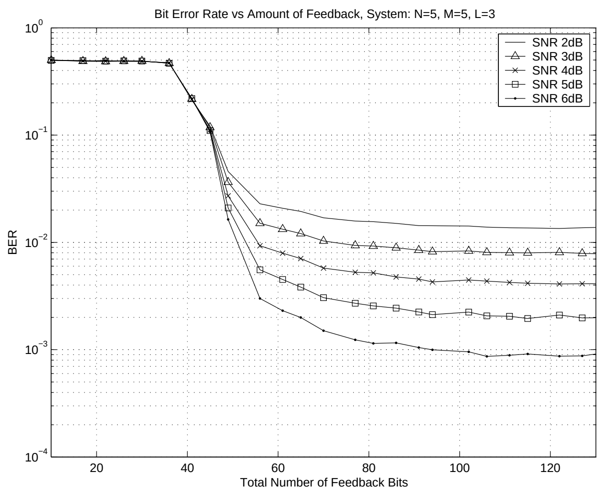

Fig. 4. Determining How Much Feedback is Necessary

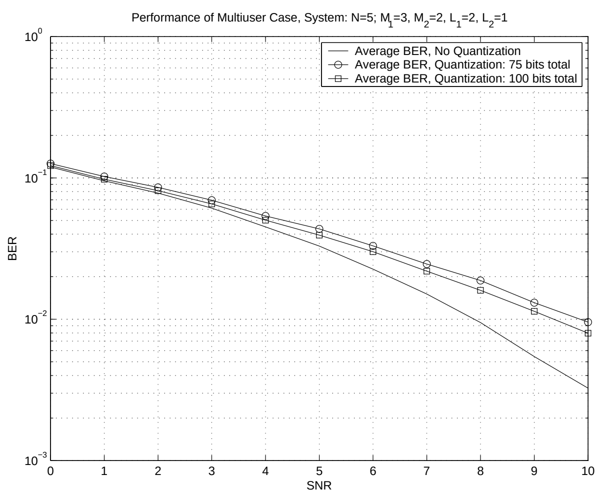

Fig. 5. Quantized Feedback Scheme in Multiuser Case

near optimal bit error rates, approximately 70 bits must be used for quantization of 21 Givens parameters. This means that an average of only 3.33 bits are need for each parameter.

D. Application to Multiuser Case

An important factor motivating this research was to enable *multiuser* precoding. This example simulates a system with $N = 5$ antennas serving two users simultaneously. User 1 uses $M_1 = 3$ antennas and receives $L_1 = 2$ data streams and user 2 uses $M_2 = 2$ and receives $L_2 = 1$ data stream. The transmitter uses the linear precoder of [3]. Figure 5 plots the overall BER with various levels of feedback, comparing the results to having true channel information available at the transmitter. As seen, the feedback mechanism works quite well and is able to accommodate multiple users, varying numbers of receive antennas and receive data streams without redesign. This is a crucial advance from the, otherwise more efficient, Grassmanian beamforming approach.

V. Conclusions

This paper addresses the need for an efficient, but flexible, feedback mechanism to enable precoding in downlink multiuser communications. The use of Givens decomposition is an effective means of reducing the number of parameters that require quantization. Deriving the true PDF of the parameters to be returned allows for a simple bit allocation scheme.

In terms of practical implementation, this quantization scheme can easily be applied to a MIMO system transmitting a variable number of bit streams. One can also adjust the amount of feedback given information about acceptable losses in bit error rate and channel conditions without additional complexity. An interesting extension of the scheme proposed here would be to incorporate the quantization error into the design of the precoder.

References

[1] H. Sampath, P. Stoica, and A. Paulraj, “Generalized linear precoder and decoder design for mimo channels using the weighted mmse criterion,” *IEEE Transactions on Communications*, vol. 49, pp. 2198–2206, Dec. 2001.

[2] R. F. H. Fischer and C. A. Windpassinger, “Improved MIMO precoding for decentralized receivers resembling concepts from lattice reduction,” in *Proc. of IEEE Globecom*, vol. 4, (San Francisco, CA), pp. 1852–1856, Dec. 2003.

[3] A. J. Tenenbaum and R. S. Adve, “Joint multiuser transmit-receive optimization using linear processing,” *IEEE International Conference on Communications*, vol. 1, pp. 588 – 592, June 2004.

[4] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, “On beamforming with finite rate feedback in multiple-antenna systems,” *IEEE Transactions on Information Theory*, vol. 49, pp. 2562 – 2579, Oct. 2003.

[5] D. J. Love and R. W. Heath Jr., “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Transactions on Information Theory*, vol. 49, Oct. 2003.

[6] J. C. Roh and B. D. Rao, “An efficient feedback method for mimo systems with slowly time-varying channels,” *Wireless Communications and Networking Conference*, vol. 3, pp. 760–764, Mar. 2004.

[7] M. A. Sadrabadi, A. K. Khandani, and F. Lahouti, “A new method of channel feedback quantization for high data rate mimo systems,” *GlobeCom*, vol. 999, p. 99999999, Dec. 2004.

[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. The Kluwer Academic Publishers, 1992.

[9] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 2 ed., 1989.

[10] E. Telatar, “Capacity of multi-antenna gaussian channels,” *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.