Multi-Antenna Limited Feedback for Temporally-Correlated Channels: Feedback Compression

Kaibin Huang, Bishwarup Mondal, Robert W. Heath, Jr and Jeffrey G. Andrews
Wireless Networking and Communications Group
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX 78712-0240
Email: {khuang, mondal, rheath, jandrews}@ece.utexas.edu

***Abstract*— A novel method is proposed for reducing the feedback rate of a transmit beamforming system with feedback of quantized channel state information. Specifically, the channel is modeled as a Markov chain and the feedback bits are compressed by truncating the Markov chain transition probabilities. Using the proposed method, the feedback rate can be compressed by more than 100% without degrading the system performance.**

I. INTRODUCTION

For a multi-antenna downlink (DL) system, the benefits of transmit (TX) channel state information (CSI), such as improvements on the data throughput and the link reliability, can be attained by using intelligent CSI quantization techniques, called *limited feedback*. In the literature of limited feedback, *block fading* is commonly assumed for the DL channel [1], [2]. Thereby, different channel realizations are assumed independent and potential temporal correlation is ignored [1]. Under this assumption, different methods for quantizing CSI have been developed such as line packing [3] and vector quantization [4], and different types of limited feedback systems have been investigated including quantized beamforming [5] and spatial multiplexing [6], [7].

In practical systems, the block fading assumption for the DL channel is inaccurate as channel temporal correlation often exists. Hence, the CSI feedback in a limited feedback system can be compressed by exploiting channel temporal correlation. Intuitively, data compression techniques [8], [9] can be employed for compressing CSI. Unfortunately, by processing CSI in blocks, these techniques incur feedback delay that severely degrades the system performance as shown in [10]. Alternatively, efficient CSI feedback algorithms are proposed in [11], [12]. In [11], a multiple-input-multiple-output (MIMO) channel is parameterized and the feedback of each parameter is compressed to be one bit. Nevertheless, the multiplicity of the channel parameters compromises the feedback efficiency. In [12], the CSI feedback is compressed to be one bit but requires the transmitter to broadcast channel subspace matrices. The drawbacks of existing algorithms motivate us to develop an CSI feedback method that exploits channel temporal correlation more efficiently. Last, it is worth mentioning that the channel spatial correlation is used in [13] for CSI feedback compression.

In this paper, a novel method is proposed for compressing the feedback of a TX beamforming system with limited feedback. First, the DL channel is modeled as a finite-state Markov chain. Second, the feedback bits are compressed by ignoring transitions between Markov states that occur with small probabilities, hence truncating the transition probabilities. The feedback rate and the DL capacity for the proposed method are derived. Numerical results show that the effectiveness of the feedback compression method increases with the quantization resolution. For example, for two transmit antennas, the reduction of the feedback rate with respect to (w.r.t.) the case of no compression is about 100% for 64 quantization points, or 200% for 256 points.

II. SYSTEM DESCRIPTION

The beamforming system with limited feedback and a TC channel is illustrated in Fig. 1. This system is discrete-time where continuous-time signals are sampled at the symbol rate $1/T_s$ with T_s being the symbol duration. Consequently, each signal is represented by a sequence of samples with the subscript n denoting the sample index. Furthermore, the signals in this system are defined and denoted as follows:

$\mathbf{h}_n$ ($L\times 1$ vector) DL channel with L TX antennas;
J_n channel state, $J_n \in \mathbb{J} = \{1, \cdots, N\}$;

$\mathcal{H}$ codebook containing N possible beamforming vectors;

$\mathbf{f}_i$ ith member of the codebook $\mathcal{H}$ where $1 \leq i \leq N$;

K_n TX channel state, $K_n \in \mathbb{J}$;

$\mathbf{w}_n$ ($L \times 1$ vector) TX beamforming vector with $\|\mathbf{w}_n\|^2 = 1$;

The following assumption about the DL channel[1] is made:

AS 1: *The DL channel* $\mathbf{h}_n$ *is an i.i.d. vector whose coefficients are* $\mathcal{CN}(0, 1)$.

Let the DL channel be decomposed as $\mathbf{h}_n = g_n \mathbf{u}_n$ where $g_n = \|\mathbf{h}_n\|$ and $\mathbf{u}_n = \mathbf{h}_n / \|\mathbf{h}_n\|$ are named *the channel gain* and *the channel shape*, respectively. The purpose of limited feedback is to send the channel shape $\mathbf{u}_n$ to the base-station for choosing the TX beamforming vector $\mathbf{w}_n$. In Fig. 1, the RX CSI represents the channel shape as the result of the following assumption (AS):

AS 2: *The mobile has perfect knowledge of the DL channel.*

By making this assumption, we ignore the channel estimation error at the mobile.

The process of CSI feedback is described as follows. First, the *CSI quantizer* maps the RX CSI $\mathbf{u}_n$ onto a member of the codebook $\mathcal{H}$ [2] and outputs its index J_n, called the *channel state*. Second, the time instants for sending back J_n are determined by an *aperiodic feedback scheme*, which initiates a feedback whenever the channel state changes ($J_n \neq J_{n-1}$). Third, the feedback channel conveys the channel state to the base-station and is specified by the following assumption:

AS 3: *The feedback channel is free of error and delay.*

The error-free assumption in AS 3 is justifiable as the feedback channel is usually protected using error-correction coding and hence has a very low error probability. Given this feedback channel, the TX channel state is equal to the channel state by D samples: $K_n = J_n$. Last, the *beamforming vector generator* simply performs the codebook lookup function using the TX channel state K_n, where the codebook is also $\mathcal{H}$.

III. FEEDBACK COMPRESSION METHOD

The proposed feedback compression method is comprised of two steps: modeling the DL channel using a Markov chain and truncating the Markov chain transition probabilities. These two steps are discussed in Section III-A and Section III, respectively.

A. Channel State Markov Chain

The channel state J_n is modeled using a finite-state and discrete-time Markov chain. The channel-state Markov chain is identical to that constructed in [14] and its construction and properties are summarized as follows.

The states of the channel-state Markov chain are one-to-one mapped to the members of the codebook $\mathcal{H}$. The one-to-one mapping also exists between the codebook members and the quantized values of the channel shape. The codebook $\mathcal{H}$ is designed by partitioning the channel shape space (unit hyper sphere) into N regions, called *Voronoi cells* and denoted as $\{\mathbb{V}_i\}$ [15]. Let $\mathcal{Q}$ denote the quantization function of the channel shape $\mathbf{u}_n$. We have the following one-to-one relationships:

$$\mathbf{u}_n \in \mathbb{V}_i \iff \mathcal{Q}(\mathbf{u}_n) = \mathbf{f}_i \iff J_n = i, \qquad (1)$$

where $1 \leq i \leq N$ and $\mathbf{f}_i$ is the ith codebook member.

The stationary and transition probabilities of the channel-state Markov chain are defined as

$$\begin{aligned} P_i &= \Pr\{J_n = i\}, \quad \text{and} && (2)\\ P_{i,j} &= \Pr\{J_n = i \mid J_{n-1} = j\}, && (3)\end{aligned}$$

where $1 \leq i, j \leq N$. As shown in [14], the channel-state Markov chain has the following symmetric properties:

$$\begin{aligned} P_1 &= \cdots = P_N = \frac{1}{N}, && (4)\\ P_{1,1} &= \cdots = P_{N,N}. && (5)\end{aligned}$$

Given difficulties in obtaining their closed-form expressions [14], the transition probabilities are computed using the Monte Carlo method.

B. Truncation of Transition Probabilities

Consider the channel-state Markov chain. Given the current state, the next state belongs to only a subset of the Markov state space with a high probability and the rest of the state space with a negligible probability due to channel temporal correlation. This motivates the feedback compression by truncating the transition probabilities as described shortly.

Define the ϵ-neighborhood of the Markov state i as

$$\mathcal{N}_\epsilon(i) = \{1 \leq j \leq N \mid P_{i,j} \geq \epsilon\}, \quad i = 1, \cdots, N, \qquad (6)$$

where $\epsilon > 0$ is a small positive number. Therefore, the required number of feedback bits for identifying the current channel state, say $J_n = i$, is defined as

$$B_i = \lceil \log_2 |\mathcal{N}_i(\epsilon)| \rceil \text{ bits}, \quad i = 1, \cdots, N, \qquad (7)$$

---

[1] Note that our analysis can be extended to the case of spatially correlated channels, which mainly changes the properties of the Markov chain modeling the RX CSI.

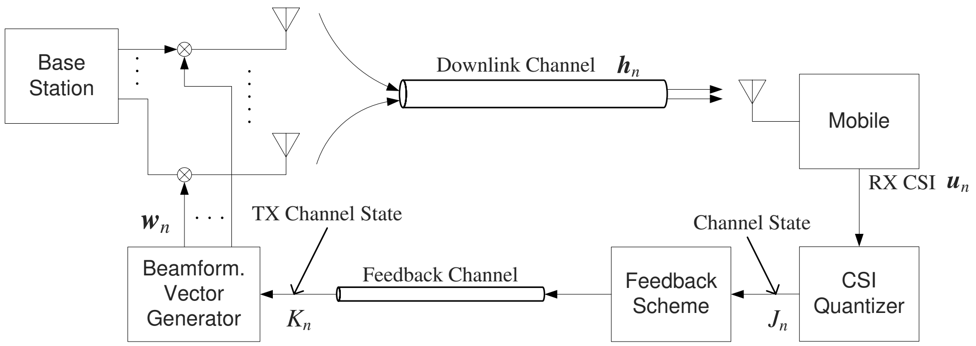


Fig. 1. Discrete-time transmit beamforming system with limited feedback and a TC channel (time is represented by the sample index n)

where $|\cdot|$ denotes the cardinality of a set and $\lceil a \rceil$ is the smallest integer larger than a. It follows from (6) that $|\mathcal{N}_i(\epsilon)| < N$ and hence $B_i < \log_2 N$. Note that $\log_2 N$ is the number of feedback bits for the case of no feedback compression. Compared with the case of no feedback compression, the above feedback compression method incurs an additional cost – storage of tables defining ϵ-neighborhoods at both the mobile and the base-station.

IV. ANALYSIS

The feedback rate and the DL ergodic capacity for the proposed feedback compression method are derived in Section IV-A and Section IV-B, respectively.

A. Feedback Rate

Consider the limited feedback system in Fig. 1 and a duration of M symbols. The average feedback rate is defined as

$$\mathcal{R}(M) = \lim_{M\rightarrow\infty} \frac{1}{MT_s} \sum_{n=0}^{M-1} B_{J_n} \mathcal{I}\{J_n \neq J_{n+1}\}, \quad (8)$$

where $\mathcal{I}\{\cdot\}$ is the indicator function, J_n is the channel state and B_i is defined in (7). The average feedback rate is derived and given in the following proposition, whose proof is given in [10].

Proposition 1: *With feedback compression, the average feedback rate is given as*

$$\textit{(feedback compression)} \quad \mathcal{R} = \frac{(1-P_{1,1})}{NT_s} \sum_{i=1}^{N} B_i, \quad (9)$$

where B_i is in (7).

Note that the symmetric properties of the channel-state Markov chain in (4) and (5) are used in deriving the average feedback rate in (9).

By replacing B_i in (9) with $\log_2 N$, we can obtain the average feedback rate for the case of no feedback compression as

$$\text{(feedback compression)} \quad \mathcal{R} = \frac{\log_2 N}{T_s}(1-P_{1,1}). \quad (10)$$

This result has been also obtained in [14].

B. Downlink Ergodic Capacity

In this section, the downlink ergodic capacity for the feedback compression method in Section III will be defined and derived.

The instantaneous capacity for the TX beamforming system in Fig. 1 is given as

$$\mathcal{C}(\mathbf{h}_n, \mathbf{w}_n) = \log_2\left(1 + \rho \left|\mathbf{w}_n^H \mathbf{h}_n\right|^2\right), \quad (11)$$

where ρ, $\mathbf{h}_n$ and $\mathbf{w}_n$ are the SNR, the DL channel and the TX beamforming vector, respectively. Define the DL ergodic capacity as

$$\bar{\mathcal{C}}_\epsilon = \lim_{N\rightarrow\infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{C}(\mathbf{h}_n, \mathbf{w}_n), \quad (12)$$

where the instantaneous capacity $\mathcal{C}$ is given in (11). We derive the DL ergodic capacity as shown in the following proposition.

Proposition 2: *Using the feedback compression method, the DL ergodic capacity defined in (12) is obtained as*

$$\bar{\mathcal{C}}_\epsilon = \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{j\in\mathcal{N}_\epsilon(i)} \mathcal{D}(j,j) P_{j,i} + \sum_{l\in\mathcal{N}_\epsilon^C(i)} \mathcal{D}(l,i) P_{l,i} \right], \quad (13)$$

where

$$\mathcal{D}(j,i) = \int_{\mathbb{V}_j} \int_g \mathcal{C}(g\mathbf{u}, \mathbf{f}_i) f_g(g) f_{\mathbf{u}}(\mathbf{u}) dg d\mathbf{u}, \quad (14)$$

with $\mathcal{C}$ given in (11), $\mathbf{f_i}$ being the i member of the codebook $\mathcal{H}$, and f_g and $f_{\mathbf{u}}$ the PDF's of the channel gain g and the channel shape $\mathbf{u}$, respectively.

Proof: See Appendix. ☐

By replacing the ϵ-neighborhood $\mathcal{N}_\epsilon(i)$ in (13) with the channel state space $\mathcal{N} = \{1, \cdots, N\}$, we obtain the DL ergodic capacity for the case of no feedback compression in the following corollary.

Corollary 1: Without feedback compression, the DL ergodic capacity is given as

$$\bar{\mathcal{C}} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{D}(j,j), \qquad (15)$$

where $\mathcal{D}(j,j)$ is in (14).

V. NUMERICAL RESULTS

For numerical results, the following assumption is made about the DL channel in addition to AS 1:

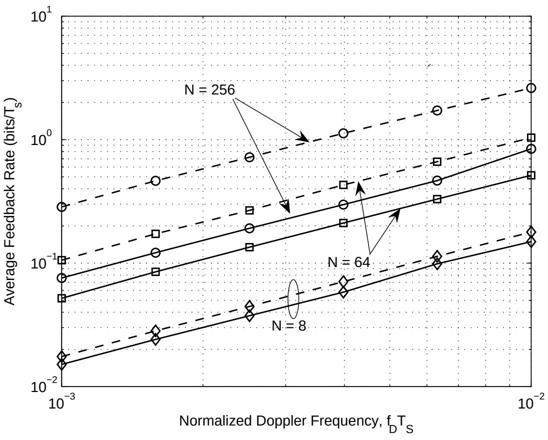
AS 4: *The scattering environment is uniform [16].* The direct result of the above assumption is that the temporal correlation of the DL channel coefficients are specified by Clarke's function [17] and hence characterized by the Doppler frequency. Moreover, the value of ϵ used in the feedback compression method is chosen to be 10^{-6}.

The CSI feedback rates with and without feedback compression are compared in Fig. 2, which shows the *average feedback rate vs. Doppler frequency* curves the cases of two ($L = 2$) and four ($L = 4$) TX antennas. The transition probabilities of the channel-state Markov chain are computed using the Monte Carlo method. Various sizes of the codebook $\mathcal{H}$ are considered. Several observations can be made. First, the compression ratio of the feedback rate w.r.t. the case of no compression is significant especially for large codebook sizes (N). For example, the ratio is as high as 200% for the case of two transmit antennas and the codebook size $N = 256$. Nevertheless, the compression ratio decreases with the codebook size N. Second, the feedback compression is more effective for lower Doppler frequencies, as clearly reflected in Fig. 2(b). Third, regardless of whether feedback is compressed, the feedback rates increase with the Doppler frequency.

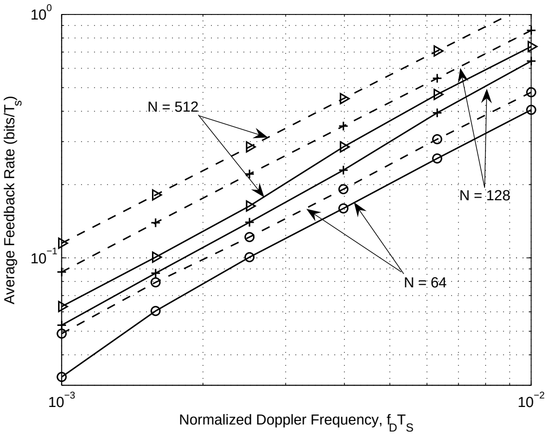
In Fig. 3, the DL ergodic capacities with feedback compression (plotted using Markers) are compared with those without feedback compression (plotted as solid curves) for various combinations of the codebook size and the number of TX antennas. As observed from Fig. 3, feedback compression does not decrease the DL ergodic capacity since the lines (without compression) and markers (with compression) overlap. Last, note that the DL ergodic capacities are independent of the Doppler frequency given no feedback delay and hence the curves in Fig. 3 are horizontal lines.

VI. CONCLUSION

We have proposed a method for compressing the CSI feedback in a TX beamforming system with limited feedback. The corresponding feedback rate and DL ergodic



(a) $L = 2$



(b) $L = 4$

Fig. 2. Comparison of average feedback rates with (solid curves) and without (dashed curves) feedback compression

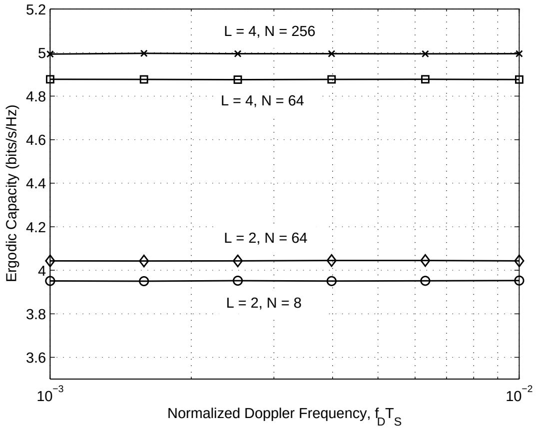


Fig. 3. Comparison of DL ergodic capacities with (markers) and without (solid curves) feedback compression.

capacity have been derived. Numerical results show that the proposed feedback compression method reduces the feedback rate significantly w.r.t. the case of no feedback compression. This gain on feedback compression does not compromise the DL ergodic capacity.

ACKNOWLEDGMENT

B. Mondal is the recipient of a Motorola Partnerships in Research Grant. This work is funded by Freescale Inc. and the National Science Foundation under grants CCF-514194 and CNS-435307.

APPENDIX

PROOF OF PROPOSITION 2

The TX beamforming vector $\mathbf{w}_n$ is chosen based on the feedback information on the quantized channel shape. Given the feedback compression method in Section III and by using (1), the beamforming vector can be expressed as

$$\mathbf{w}_n = \begin{cases} \mathcal{Q}(\mathbf{u}_n) = \mathbf{f}_{J_n}, & J_n \in \mathcal{N}_\epsilon(J_{n-1}), \\ \mathcal{Q}(\mathbf{u}_{n-1}) = \mathbf{f}_{J_{n-1}}, & \text{otherwise}, \end{cases} \tag{16}$$

where $\mathcal{N}_\epsilon(J_{n-1})$ is the ϵ-neighborhood of J_{n-1}.

By substituting $\mathbf{h}_n = g_n \mathbf{u}_n$ (cf Section II) into (12),

$$\bar{\mathcal{C}}_\epsilon = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{C}(g\mathbf{u}_n, \mathbf{w}_n). \tag{17}$$

Next, substitute (16) into (17) and we obtain that,

$$\begin{aligned} \bar{\mathcal{C}}_\epsilon = & \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{C}(g\mathbf{u}_n, \mathcal{Q}(\mathbf{u}_n)) \mathcal{I}\{J_n \in \mathcal{N}_\epsilon(J_{n-1})\} \\ & + \frac{1}{N} \sum_{n=1}^{N} \mathcal{C}(g\mathbf{u}_n, \mathcal{Q}(\mathbf{u}_n)) \mathcal{I}\{J_n \in \mathcal{N}_\epsilon^C(J_{n-1})\}. \end{aligned} \tag{18}$$

By applying Birkhoff's ergodic theorem [18] on (18),

$$\begin{aligned} \bar{\mathcal{C}}_\epsilon = & \mathbb{E}\left[\mathcal{C}(g\mathbf{u}_n, \mathcal{Q}(\mathbf{u}_n)) \mathcal{I}\{J_n \in \mathcal{N}_\epsilon(J_{n-1})\}\right] \\ & + \mathbb{E}\left[\mathcal{C}(g\mathbf{u}_n, \mathcal{Q}(\mathbf{u}_n)) \mathcal{I}\{J_n \in \mathcal{N}_\epsilon^C(J_{n-1})\}\right]. \end{aligned} \tag{19}$$

It follows from (19) that

$$\begin{aligned} \bar{\mathcal{C}}_\epsilon = & \sum_{i=1}^{N} \sum_{j=1}^{N} P_{j,i} P_i \int_{\mathbb{V}_j} \int_0^\infty \mathcal{C}(g\mathbf{u}, \mathbf{f}_j) \left(\mathcal{I}\{j \in \mathcal{N}_\epsilon(i)\}\right. \\ & \left. + \mathcal{I}\{j \in \mathcal{N}_\epsilon^C(i)\}\right) f_g(g) f_{\mathbf{u}}(\mathbf{u}) dg d\mathbf{u}, \end{aligned}$$

where f_g and $f_{\mathbf{u}}$ are the PDF's of the channel gain g and shape $\mathbf{u}$, respectively. Therefore,

$$\begin{aligned} \bar{\mathcal{C}}_\epsilon = & \sum_{i=1}^{N} P_i \left[\sum_{j \in \mathcal{N}_\epsilon(i)} P_{j,i} \int_{\mathbb{V}_j} \int_0^\infty \mathcal{C}(g\mathbf{u}, \mathbf{f}_j) f_g(g) f_{\mathbf{u}}(\mathbf{u}) dg d\mathbf{u} \right. \\ & \left. + \sum_{l \in \mathcal{N}_\epsilon^C(i)} P_{l,i} \int_{\mathbb{V}_l} \int_0^\infty \mathcal{C}(g\mathbf{u}, \mathbf{f}_l) f_g(g) f_{\mathbf{u}}(\mathbf{u}) dg d\mathbf{u} \right]. \end{aligned} \tag{20}$$

Substitution of (14) into (20) leads to (13).

REFERENCES

[1] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. on Info. Theory*, vol. 45, no. 5, pp. 1468–89, July 1999.

[2] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Comm. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[3] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. on Info. Theory*, vol. 49, no. 10, pp. 2735–47, Oct. 2003.

[4] J. C. Roh and B. D. Rao, "Performance analysis of multiple antenna systems with VQ based feedback," in *Proc. of the Asil. Conf. on Sig. Sys. and Comp.*, Nov. 2004, pp. 1978–82.

[5] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. on Info. Theory*, vol. 49, no. 10, pp. 2562–79, Oct. 2003.

[6] D. J. Love and R. W. Heath Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. on Info. Theory*, vol. 51, no. 8, pp. 1967–76, Aug. 2005.

[7] J. C. Roh and B. D. Rao, "Channel feedback quantization methods for MISO and MIMO systems," in *Proc., IEEE PIMRC*, Barcelona, Spain, Sept. 2004, pp. 805–9.

[8] S. A. Savari, "Renewal theory and source coding," *Proceedings of the IEEE*, vol. 88, no. 11, pp. 1692–1702, Nov. 2000.

[9] I. Tabus and J. Rissanen, "Asymptotics of greedy algorithms for variable-to-fixed length coding of Markov sources," *IEEE Trans. on Info. Theory*, vol. 48, no. 7, pp. 2022–35, July 2002.

[10] K.-B. Huang, B. Mondal, R. W. Heath, Jr., and J. G. Andrews, "Mutli-antenna limited feedback systems for temporally-correlated channels: Feedback rate and effect of feedback delay," 2006, in preparation.

[11] J. C. Roh and B. D. Rao, "An efficient feedback method for MIMO systems with slowly time-varying channels," in *Proc., IEEE Wireless Communications and Networking Conf.*, Atlanta, GA, Mar. 2004, pp. 11–14.

[12] B. C. Banister and J. R. Zeidler, "A simple gradient sign algorithm for transmit antenna weight adaptation with feedback," *IEEE Trans. on Sig. Proc.*, vol. 51, no. 5, pp. 1156 – 1171, May 2003.

[13] B. Mondal and R. W. Heath, Jr., "Channel adaptive quantization for limited feedback MIMO beamforming systems," submitted to *IEEE Trans. on Sig. Proc.*, 2005.

[14] K.-B. Huang, B. Mondal, R. W. Heath, Jr., and J. G. Andrews, "Markov models for multi-antenna limited feedback systems," to appear in *Proc., IEEE Int. Conf. Acoust., Speech and Sig. Proc., 2006*.

[15] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, 1992.

[16] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, pp. 957–1000, 1974.

[17] Jr. W. C. Jakes, *Microwave mobile comunications*, Willey, New York, 1974.

[18] P. Billingsley, *Ergodic Theory and Information*, R. E. Krieger Pub. Co, 1978.