LINEAR TRANSCEIVER DESIGN IN A MULTIUSER MIMO SYSTEM WITH QUANTIZED CHANNEL STATE INFORMATION

Muhammad Nazmul Islam and Raviraj Adve

Department of Electrical and Computer Engineering, University of Toronto
10 King's College Road, Toronto, Ontario, M5S 3G4, Canada

ABSTRACT

We design an end-to-end linear transceiver in the downlink of a multi-user multiple-input multiple-output (MIMO) system with multiple data streams per user and quantized channel state information at the transmitter. We minimize the sum mean squared error (SMSE) under a sum power constraint with quantization based on the mean squared inner product. We make three contributions: (i) we remove dimensionality constraints on the MIMO configuration and the resulting feedback overhead scales linearly with the number of data streams; (ii) we use the combination of eigenmode combining and minimum mean square error receiver that makes user's feedback mutually independent; (iii) we analyze SMSE at high signal-to-noise ratio and large number of transmit antennas and ~~show the flooring effect of limited feedback systems in terms of SMSE~~.

Index Terms: MU precoding, SMSE bounds, MSIP

1. INTRODUCTION

The benefits of multi-user (MU) multiple-input-multiple-output(MIMO) systems are now well accepted [1–3]. These systems need to precode before transmitting the signals to combat receiver noise and MU interference; this in turn requires channel state information (CSI) at the transmitter. In both frequency division duplex and broadband time division duplex systems [4] the receiver ~~can~~ estimate, quantize and ~~feedback~~ the CSI to the transmitter. This paper deals with ~~the~~ feedback in a MU MIMO system with each user possibly receiving multiple data streams.

Most of the relevant works in the literature deal with zero-forcing (ZF) beamforming ~~[3]~~ and therefore assume that the number of transmit antennas are greater than that of receive antennas (the dimensionality constraint) [3]. The authors in [2] use coordinated beamforming which ~~avoid~~ this constraint. However, in [2], the feedback overhead is proportional to $(M^2 - 1)$ where M is the number of transmit antennas. Motivated by the need to minimize the overhead, in our work, each user projects its MIMO channel to its dominant eigenvectors and then quantizes the effective unit norm multiple-input-single-output (MISO) channel for the

data streams. Therefore, feedback overhead is spent only on the direction of those MISO channels and scales linearly with the number of transmit antennas and data streams.

Grassmanian line packing [5], vector quantization (VQ) based on mean square error (MSE) [6] and random vector quantization (RVQ) [3, 7] have been the popular approaches for codebook generation in limited feedback. As an alternative, the authors in [8] propose maximizing MSIP as the codebook design criterion. We use this approach here.

This paper makes the following contributions: we eliminate the dimensionality constraint and, using an eigenvector based combining, tie the feedback overhead to the number of data streams, which is necessarily smaller than the number of receiver antennas. This allows for the feedback scheme to be tied to an individual user. Finally we ~~show the flooring effect of the SMSE~~ for high SNR and large M; previous work in this area has focused on the ceiling effect in terms of capacity [3, 9] and signal-to-interference-plus-noise ratio [6].

2. ~~SYSTEM MODEL~~

We consider a single base station equipped with M transmit antennas and K independent users. User k has N_k antennas and receives L_k data streams. ~~We have~~ $L = \sum_k L_k$ and $N = \sum_k N_k$. The ith data stream is processed by a unit norm linear precoding vector $\mathbf{u}_i$ with the global precoder $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_L]$. Let $\mathbf{p} = [p_1, p_2, .., p_L]^T$ be the allocated power to the data streams and define the downlink power matrix $\mathbf{P} = diag(\mathbf{p})$. $||\mathbf{p}||_1 \leq P_{max}$. The overall data vector is $\mathbf{x} = [x_1, ..., x_L]^T = [\mathbf{x}_1^T, \mathbf{x}_2^T, ..., \mathbf{x}_K^T]^T$. The $N_k \times M$ block fading channel, $\mathbf{H}_k^H$, between the BS and the user is assumed to be flat. The global channel matrix is $\mathbf{H}^H$, with $\mathbf{H} = [\mathbf{H}_1, ..., \mathbf{H}_k]$. User k receives

$$\mathbf{y}_k^{DL} = \mathbf{H}_k^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{x} + \mathbf{n}_k, \tag{1}$$

where $\mathbf{n}_k$ represents the additive white Gaussian noise at the receiver with $E[\mathbf{n}\mathbf{n}^H] = \sigma^2 \mathbf{I}_{N_k}$. Also, $E[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_L$. To estimate its own transmitted symbols, from $\mathbf{y}_k^{DL}$, user k forms $\hat{\mathbf{x}}_k = \mathbf{V}_k^H \mathbf{y}_k^{DL}$. Let $\mathbf{V}$ be the $N \times L$ block diagonal global decoder matrix, $\mathbf{V} = diag(\mathbf{V}_1, ..., \mathbf{V}_K)$. Overall

$$\hat{\mathbf{x}} = \mathbf{V}^H \mathbf{H}^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{x} + \mathbf{V}^H \mathbf{n} = \mathbf{F}^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{x} + \mathbf{V}^H \mathbf{n} \tag{2}$$

where, to facilitate our analysis, we define the $M \times L$ matrix $\mathbf{F} = \mathbf{HV}$ with $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_L]$. The vectors $\mathbf{f}_1, \ldots, \mathbf{f}_L$ are the effective $M \times 1$ channels for the individual data streams.

In designing the precoder $\mathbf{U}$, it is computationally efficient to use a ~~virtual dual~~ uplink [1]. ~~Assume that in~~ this uplink the transmit powers are $\mathbf{q} = [q_1, .., q_L]^T$ for the L data streams. The global virtual uplink power allocation matrix $\mathbf{Q}$ is defined as, $\mathbf{Q} = diag(\mathbf{q})$. So,

$$\hat{x}_i^{UL} = \mathbf{u}_i^H \left(\sum_{j=1}^{L} \mathbf{f}_j \sqrt{q_j} x_j + \mathbf{n} \right) \qquad (3)$$

To ensure resolvability, we ~~assume~~ $L \le M$ and $L_k \le N_k$.

3. EIGENMODE BASED COMBINING AND MSIP

We assume that the receivers have perfect channel knowledge. For the purposes *of quantization only*, each user uses, as $\mathbf{V}_k$, the L_k singular vectors corresponding to the maximum singular values of $\mathbf{H}_k$. ~~This is unlike the MMSE solution in [1]~~, but allows for the *channel feedback to be independent of the other users' actions*. The channels are quantized *after* this eigenmode based combining (EBC).

The quantization codebook is generated as a VQ problem using the MSIP optimality criterion [8]. Different quantization codebooks are generated for different data streams so that channels for multiple data streams are not quantized to the same codevector. Each codebook consists of 2^B unit norm vectors $\hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_{2^B}$ and each user feeds back B bits for each of its data streams. The receivers individually normalize and then quantize the effective channels for the data streams as,

$$\widehat{\mathbf{f}}_i = \arg \max_{\hat{\mathbf{w}} \in \{\hat{\mathbf{w}}_1, .., \hat{\mathbf{w}}_{2^B}\}} |\overline{\mathbf{f}}_i^H \hat{\mathbf{w}}| \qquad (4)$$

$\overline{\mathbf{f}}_i = \frac{\mathbf{f}_i}{||\mathbf{f}_i||}$ where $|| \cdot ||$ denotes the euclidean norm. The quantization error $\widetilde{\mathbf{f}}_i$ can be defined as,

$$\widetilde{\mathbf{f}}_i = \overline{\mathbf{f}}_i - \left(\widehat{\mathbf{f}}_i^H \overline{\mathbf{f}}_i \right) \widehat{\mathbf{f}}_i$$

The channel model at the receiver takes the following form,

$$\begin{aligned} \mathbf{f}_i &= ||\mathbf{f}_i|| \overline{\mathbf{f}}_i \\ &= ||\mathbf{f}_i|| \left(\left(\widehat{\mathbf{f}}_i^H \overline{\mathbf{f}}_i \right) \widehat{\mathbf{f}}_i + \widetilde{\mathbf{f}}_i \right) \qquad (5) \end{aligned}$$

Since we only send back the index of the unit norm $\widehat{\mathbf{f}}_i$ to the BS, we consider the following channel model at the BS,

$$\mathbf{f}_i = \widehat{\mathbf{f}}_i + \widetilde{\mathbf{f}}_i \quad \text{or} \quad \mathbf{F} = \widehat{\mathbf{F}} + \widetilde{\mathbf{F}} \qquad (6)$$

Here $\mathbf{F}$ consists of L unit norm effective channel vectors. $\widehat{\mathbf{F}}$ and $\widetilde{\mathbf{F}}$ denote the quantized feedback and error in feedback respectively. Although there is a phase shift of $\left(\widehat{\mathbf{f}}_i^H \overline{\mathbf{f}}_i \right)$ between the original $\mathbf{f}_i$ at the receiver and the assumed $\mathbf{f}_i$ at the BS, since we propose MMSE decoder while receiving data, the system performance remains invariant to this phase shift.

We approximate that $\widetilde{\mathbf{F}}$ has $M \times L$ independent identically distributed (i.i.d.) elements with zero mean and σ_E^2/M variance. In [8], σ_E^2 is measured in terms of the angle spread between the original and the quantized vector, $\sigma_E^2 = E\left[\sin^2 \left(\angle \left(\mathbf{f}_k, \hat{\mathbf{f}}_k \right) \right) \right]$. We also assume that $\widetilde{\mathbf{F}}$ is independent of $\mathbf{x}$, $\mathbf{n}$ and $\widehat{\mathbf{F}}$. The use of MSIP ensures ~~less~~ angular spread than ~~than~~ the use of MSE. This is mainly because σ_E^2 for MSIP is *upper bounded* by $2^{\frac{-B}{M-1}}$ [8] whereas in VQ MSE, the average euclidean distance is *lower bounded* by $2^{\frac{-B}{M}}$ [6]. ~~MSIP therefore improves~~ on the error exponent.

4. SMSE PRECODER DESIGN

Let e_i^{DL} be the MSE of the data stream i in the downlink where $e_i^{DL} = E\left[(\hat{x}_i - x_i)(\hat{x}_i - x_i)^H \right]$. Then the SMSE minimization problem can be formulated as,

$$\min_{\mathbf{p}, \mathbf{U}} \sum_{i=1}^{L} e_i^{DL} \qquad \text{subject to} : ||p||_1 \le P_{max} \qquad (7)$$

It is computationally efficient to first solve the problem in the virtual uplink and then transfer the solution using duality. Using (3) for $\hat{x}_i$ the MSE in the virtual uplink is

$$e_i^{UL} = \mathbf{u}_i^H \left(\mathbf{FQF}^H + \sigma^2 \mathbf{I} \right) \mathbf{u}_i + 1 - \left(\mathbf{u}_i^H \mathbf{f}_i + \mathbf{f}_i^H \mathbf{u}_i \right) \sqrt{q_i}$$

Using the fact that $\mathbf{F} = \widehat{\mathbf{F}} + \widetilde{\mathbf{F}}$,

$$\begin{aligned} e_i^{UL} &= \mathbf{u}_i^H \widehat{\mathbf{F}} Q \widehat{\mathbf{F}}^H \mathbf{u}_i + \sigma^2 \mathbf{u}_i^H \mathbf{u}_i + 1 - \mathbf{u}_i^H \widehat{\mathbf{f}}_i \sqrt{q_i} \\ &\quad - \sqrt{q_i} \widehat{\mathbf{f}}_i^H \mathbf{u}_i + E\left[\mathbf{u}_i^H \widetilde{\mathbf{F}} Q \widetilde{\mathbf{F}}^H \mathbf{u}_i | \widehat{\mathbf{F}} \right] \qquad (8) \end{aligned}$$

The last term in (8) can be written as,

$$E_{\widetilde{\mathbf{F}}} \left[E_{\widehat{\mathbf{F}}} \left[\mathbf{u}_i^H \widetilde{\mathbf{F}} Q \widetilde{\mathbf{F}}^H \mathbf{u}_i | \widehat{\mathbf{F}}, \widetilde{\mathbf{F}} \right] \right] = \frac{\sum_{i=1}^{L} q_i}{M} \sigma_E^2 \mathbf{u}_i^H \mathbf{u}_i \quad (9)$$

$$\begin{aligned} e_i^{UL} &= \mathbf{u}_i^H \widehat{\mathbf{F}} Q \widehat{\mathbf{F}}^H \mathbf{u}_i + \sigma^2 \mathbf{u}_i^H \mathbf{u}_i + 1 - \mathbf{u}_i^H \widehat{\mathbf{f}}_i \sqrt{q_i} \\ &\quad - \sqrt{q_i} \widehat{\mathbf{f}}_i^H \mathbf{u}_i + \frac{\sigma_E^2}{M} (q_1 + .. + q_L) \mathbf{u}_i^H \mathbf{u}_i \qquad (10) \end{aligned}$$

The uplink MSE receiver is given by $\mathbf{u}_i^{MSE} = \mathbf{J}^{-1} \widehat{\mathbf{f}}_i \sqrt{q_i}$

$$\mathbf{J} = \widehat{\mathbf{F}} Q \widehat{\mathbf{F}}^H + \sigma^2 \mathbf{I}_M + \frac{\sigma_E^2}{M} (q_1 + .. + q_L) \mathbf{I}_M \quad (11)$$

$$e_i^{UL,MSE} = 1 - \sqrt{q_i} \widehat{\mathbf{f}}_i^H \mathbf{J}^{-1} \widehat{\mathbf{f}}_i \sqrt{q_i} \qquad (12)$$

$$\begin{aligned} SMSE^{UL} &= \sum_{i=1}^{L} e_i^{UL,MSE} = \sum_{i=1}^{L} 1 - \sum_{i=1}^{L} \sqrt{q_i} \widehat{\mathbf{f}}_i^H \mathbf{J}^{-1} \widehat{\mathbf{f}}_i \sqrt{q_i} \\ &= L - M + \left(\sigma^2 + \frac{\sigma_E^2 \sum_{i=1}^{L} q_i}{M} \right) tr\left[\mathbf{J}^{-1} \right] \quad (13) \end{aligned}$$

where tr $[\cdot]$ denotes the trace operator. For fixed $\hat{\mathbf{F}}$, the SMSE is a function of $\mathbf{Q}$. Since, (13) remains a nonincreasing function of SNR with equal expected quantization error [10], power allocation becomes a convex optimization problem. A recent result shows that at the optimal solution ~~and with the same power constraint, the same MSE can be achieved with~~ $\mathbf{p} = \mathbf{q}$ [11]. We use this in our work.

Receiver Design: As mentioned earlier, the eigenbased combining is used for quantization purposes only. The base station determines $\mathbf{p}$ and $\mathbf{U}$ based on $\hat{\mathbf{F}}$. However, by sending dedicated pilot symbol for each data stream, the receiver can implement the MMSE decoder through training and thus improve performance compared to the eigenbased decoder. Therefore, ~~decoder~~ for the ith stream is formed as,

$$\mathbf{v}_i = \left(\mathbf{H}_i^H \mathbf{U}\mathbf{P}\mathbf{U}^H \mathbf{H}_i + \sigma^2 I\right)^{-1} \mathbf{H}_i^H \mathbf{u}_i \sqrt{p_i}. \quad (14)$$

Here, $||\mathbf{v}_i|| = 1$. $\mathbf{H}_i$ is the channel of the user ~~receiving~~ i^{th} stream.

Overall Algorithm: The steps in the overall algorithm for MU MIMO SMSE minimization are:

1. BS sends common pilots to the users so that each user can estimate its own channel.

2. Each user converts its estimated MIMO channel to effective MISO channels using EBC.

3. Each user generates a codebook of 2^B unit norm vectors using MSIP based VQ and then quantizes their effective channels and informs the BS.

4. Virtual uplink power allocation:
$$\mathbf{Q}^{opt} = \min_{\mathbf{Q}} \left(\sigma^2 + \frac{\sigma_E^2 \sum_{i=1}^L q_i}{M}\right) tr(\mathbf{J}^{-1}) \quad \text{s.t.} \quad tr[\mathbf{Q}] = P_{max}; q_k \ge 0 \text{ is convex in } \mathbf{Q}$$

5. Uplink beamforming: $\mathbf{u}_i = \mathbf{J}^{-1}\hat{\mathbf{f}}_i \sqrt{q_i}, ||\mathbf{u}_i|| = 1$

6. Downlink power allocation $\mathbf{p} = \mathbf{q}$

7. BS sends dedicated pilot symbols for each of the data streams. Thereafter, each user finds $\mathbf{v}_i$ using (14) and through training.

The algorithm above results in a precoder, $\mathbf{U}$, decoder, $\mathbf{V}$ and power allocation, $\mathbf{p}$. Note that the solution is sub-optimal ~~because~~ $\mathbf{U}$ and $\mathbf{p}$ are designed based on the EBC receiver, not an SMSE criterion.

5. ASYMPTOTIC SMSE ANALYSIS

In this section we analyze the estimated SMSE at the base station, at high SNR and for large M. To simplify the analysis we assume equal power allocation. Using (12) and (11) and the fact that $q_i = P_{max}/L$,

$$e_i^{UL} = 1 - \hat{\mathbf{f}}_i^H \left(\hat{\mathbf{F}}\hat{\mathbf{F}}^H + \frac{\sigma^2}{q_i}\mathbf{I}_M + \frac{\sigma_E^2}{M}L\mathbf{I}_M\right)^{-1} \hat{\mathbf{f}}_i, \quad (15)$$

since $\mathbf{Q} = q_i \mathbf{I}_L$. At high SNR, the second term in the bracket is negligible. Let

$$\hat{\mathbf{F}}\hat{\mathbf{F}}^H + \frac{\sigma^2}{q_i}\mathbf{I}_M = \mathbf{U}\mathbf{D}\mathbf{U}^H$$

We have

$$\begin{aligned} e_i^{UL,MSE} &= 1 - \hat{\mathbf{f}}_i^H \left(\mathbf{U}\mathbf{D}\mathbf{U}^H + \frac{\sigma_E^2}{M}L\mathbf{I}_M\right)^{-1} \hat{\mathbf{f}}_i \\ &= 1 - \hat{\mathbf{f}}_i^H \mathbf{U}(\mathbf{D} + \mathbf{C})^{-1}\mathbf{U}^H \hat{\mathbf{f}}_i \quad (16) \end{aligned}$$

In (16) we assume that every diagonal element in $\mathbf{C}$ is $\frac{\sigma_E^2}{M}L$. All the $\hat{\mathbf{f}}$ vectors fed back to the transmitter are unit norm and statistically independent with each other. Therefore, using the channel model at the base station (6), if we keep increasing M while keeping L fixed, then by the law of large numbers, $\lim_{M\to\infty} \hat{\mathbf{f}}_i^H \hat{\mathbf{F}}\hat{\mathbf{F}}^H \hat{\mathbf{f}}_i = 1$ [6].

This, in turn, leads to, $\lim_{M\to\infty} \hat{\mathbf{f}}_i^H \mathbf{U}\mathbf{D}\mathbf{U}^H \hat{\mathbf{f}}_i = 1$. Without loss of generality, we can assume that $\hat{\mathbf{f}}_i$ constitutes one of the eigenvectors of $\mathbf{U}$ and the corresponding eigenvalue in $\mathbf{D}$ is 1. Therefore,

$$\begin{aligned} e_i^{UL,MMSE} &= 1 - \frac{1}{1 + \frac{L\sigma_E^2}{M}} \\ &= \frac{L\sigma_E^2}{M + L\sigma_E^2} \quad (17) \end{aligned}$$

Since, this finds the uplink MSE for one data stream,

$$\lim_{SNR,M\to\infty} SMSE = \frac{L^2\sigma_E^2}{M + L\sigma_E^2} \quad (18)$$

The estimated SMSE at the BS (18) leads to the following result. *For a fixed quantization error, the SMSE of a multiuser system is lower bounded by a fixed value which does not depend on SNR.* We call this the *flooring effect* of multiuser broadcast systems. This is similar to the ceiling effect, in terms of capacity and SINR, seen previously in limited feedback literature [3, 6].

6. SIMULATION RESULTS

In Fig. 1 the proposed algorithm's performance is compared with that of coordinated beamforming [2]. Since [2] implements joint transceiver design, it performs better than the proposed algorithm with full CSIT. However, coordinated beamforming needs at least $\left(M^2 - 1\right)$ bits for the feedback of $\hat{\mathbf{H}}\hat{\mathbf{H}}^H$. 15 bits per user means 1 bit per unique scalar entry of that matrix which is very low. On the other hand, the proposed algorithm projects the data into the most dominant eigenvector and quantizes that with an error of only $2^{\frac{-B}{M-1}} = 0.03125$. Thus, the proposed algorithm performs very close to its full CSIT curve and outperforms [2] with limited feedback.

In Fig. 2, the performance of the proposed algorithm is compared with that of ZF and random vector quantization (RVQ) based QBC algorithm which projects the data streams in the least quantization error directions [9]. ~~Since, [9]~~ does not ~~require second phase training,~~ we also show the proposed
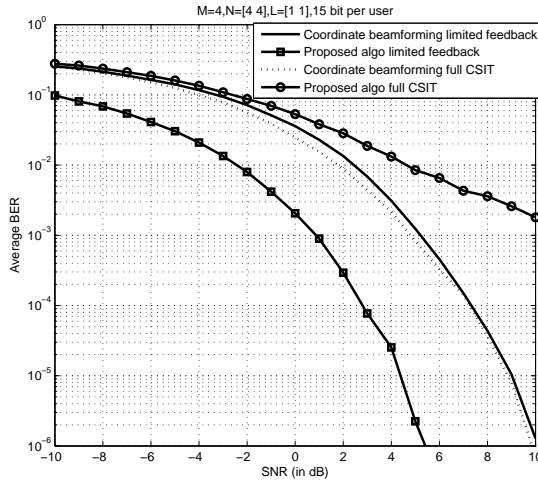
Fig. 1. Comparison with the co-ordinated beamforming M = 4, N = [4 4], L = [1 1], 15 bits per user, QPSK

Fig. 2. Comparison with the ZF-QBC, M = 4, N = [3 3], L = [2 2], QPSK, 12 bit per user

algorithm's performance with an estimated MMSE receiver, implemented with 8 training symbols. By using SMSE precoder and MMSE receiver, coming at the cost of training symbols, the proposed algorithm performs better than [9].

7. CONCLUSION

In this work, we proposed linear transceiver design in a MU MIMO system to minimize SMSE under a sum power constraint with limited feedback. We used an eigenmode based combining scheme to reduce the MIMO problem to a MISO situation and use MSIP for effective channel quantization. However, this is done only for the feedback and the receiver improves performance by using a MMSE decoder while receiving data. We analyzed SMSE asymptotically under some simplifying conditions. Our future works will incorporate channel gain feedback along with the proposed limited shape feedback and analyze the overall scheme.

8. REFERENCES

[1] A. M. Khachan, A. J. Tenenbaum, and R. S. Adve, "Linear processing for the downlink in multiuser MIMO systems with multiple data streams," in *Proc. IEEE ICC'2006*, June 2006, vol. 9, pp. 4113–4118.

[2] C. Chae, D. Mazzarese, N. Jindal, and R. W. Heath Jr, "Coordinated beamforming with limited feedback in the MIMO broadcast channel," *IEEE Journal of Selected Areas in Communications*, vol. 26, pp. 1505–1515, Oct. 2008.
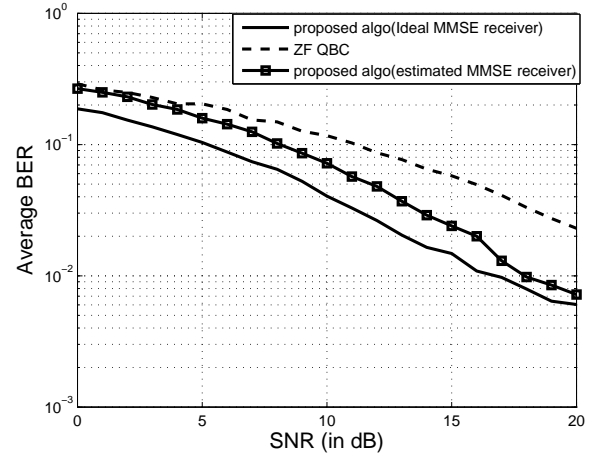
[3] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Communications*, vol. 52, pp. 5045–5060, Nov. 2006.

[4] J. C. Haartsen, "Impact of non-reciprocal channel conditions in broadband TDD systems," in *Proc. IEEE PIMRC'2008*, Sept. 2008, pp. 1–5.

[5] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmanian beam-forming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2735–2747, Oct. 2003.

[6] A. D. Dabbagh and D. J. Love, "Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch," *IEEE Transactions on Communications*, vol. 7, pp. 1859–1868, Nov. 2008.

[7] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 1473–1482, Oct. 2008.

[8] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna system with finite rate feedback: A VQ-based approach," *IEEE Transactions on Information Theory*, vol. 52, pp. 1101–1112, Mar. 2006.

[9] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 3834–3844, Oct. 2008.

[10] M. Ding, *Multiple-input multiple-output system design with imperferct channel knowledge*, Ph.D. thesis, Queen's University, 2008.

[11] A. Tenenbaum and R. S. Adve, "Minimizing sum-MSE implies identical downlink and dual uplink power allocations," IT Archive, Dec. 2009.