ORTHOGONAL BEAMFORMING FOR SDMA DOWNLINK WITH LIMITED FEEDBACK

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

***Abstract*— On a multi-antenna downlink channel, separation of multiple users by transmit beamforming enables simultaneous transmission from the base station to the users, resulting in high sum throughput. This paper proposes and analyzes a practical algorithm for joint scheduling and orthogonal beamforming, which is enabled by feedback of quantized channel state information (CSI). In this approach, each user quantizes CSI using a codebook comprised of multiple orthonormal vector sets and sends back quantized CSI. Using feedback CSI, the base station jointly selects a set of orthogonal beamforming vectors and schedules a subset of feedback users for downlink transmission such that the throughput is maximized. For moderate to large numbers of users, the proposed algorithm achieves higher sum capacities than the conventional ones.**

***Index Terms*—** Array Signal Processing, Space Division Multiplexing, Multiuser Channels, Broadcast Channels, Scheduling

I. Introduction

In multi-antenna broadcast (downlink) channels, simultaneous transmission to multiple users, known as *space division multiple access* (SDMA), is capable of achieving very higher throughput [1]. The difficulty of implementing the optimal SDMA strategy known as *dirty paper coding* [2] motivates the development of more practical SDMA algorithms based on transmit beamforming, which are designed using different criteria and methods, including zero forcing [3]–[5], a signal-to-interference-plus-noise-ratio (SINR) constraint [6], minimum mean squared error (MMSE) [7], and channel decomposition [8]. These SDMA beamforming algorithms can be combined with scheduling to further increase the sum capacity by exploiting *multiuser diversity*, referring to scheduling a subset of users with good channels for each transmission [9], [10].

This paper is focused on beamforming and scheduling for SDMA systems with limited feedback [11], referring to techniques for quantization and feedback of *channel state information* (CSI) from users to a base station. A simple algorithm, named *opportunistic SDMA* (OSDMA), is proposed in [10], where users are scheduled based on their choices of beamforming vectors from an arbitrary set of orthonormal vectors. Despite being asymptotically optimal, for a small number of users, such arbitrary beamforming vectors are highly sub-optimal due to excessive interference between scheduled users. An improved algorithm, called *OSDMA with beam selection* (OSDMA-S), is proposed in [12] for increasing the sum capacity by iterative broadcast of orthonormal vector sets and feedback of users' choices of beamforming vectors. Another approach is proposed in [5], which combines limited feedback using *random vector quantization* [13], scheduling based on a greedy search, and zero-forcing beamforming. The combined algorithm is referred to as ZF-OSDMA.

In this paper, we propose a joint scheduling and beamforming algorithm for OSDMA with limited feedback, called limited feedback OSDMA (LF-OSDMA). First, the channel state information (CSI) is quantized by each user using a quantization codebook comprised of multiple orthonormal vector sets and then sent to the base station. Second, using quantized CSI feedback from users, the base station jointly schedules a subset of feedback users and chooses corresponding orthogonal beamforming vectors from the quantization codebook. It is worth mentioning that LF-OSDMA as well as ZF-OSDMA and OSDMA-S focuses on feedback reduction for individual users. We have addressed the issue of limiting the sum feedback rate from all users in a separate paper [14].

Besides the LF-OSDMA algorithm, our other contribution is the analysis of its sum capacity and the effect of limited feedback on the sum capacity. First, increasing CSI feedback is shown using a sum capacity upper bound to have the virtual effect of expanding the user set. Second, the capacity gain with respect to OSDMA is shown to increase logarithmically with the size of the quantization codebook but decrease also logarithmically with the number of users U. Third, we prove that given a fixed codebook size, CSI quantization prevents the sum capacity to grow with the increasing SNR. As also proved, this negative effect of CSI quantization can be avoided by increasing the codebook size with the SNR at a polynomial rate.

II. System Model

In Fig. 1, a base station with N_t antennas transmits data simultaneously to N_t active users chosen from a total of U users, each with one receive antenna. The base station separates the multi-user data streams by beamforming, i.e.

[†]Kaibin Huang is the recipient of a Motorola Partnerships in Research Grant. This work is funded by the National Science Foundation under grants CCF-514194 and CNS-435307.

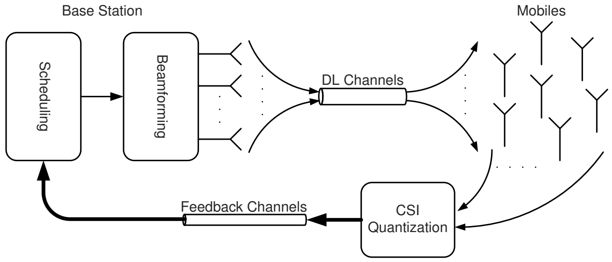

Fig. 1. Downlink system with limited feedback

assigning a beamforming vector to each of the N_t active users. The beamforming vectors $\{\mathbf{w}_n\}_{n=1}^{N_t}$ are selected from multiple sets of orthonormal vectors following the joint scheduling and beamforming algorithm in Section III-B. The received signal of the nth scheduled user is

$$y_n = \sqrt{P}\mathbf{h}_n^\dagger \sum_{k=1}^{N_t} \mathbf{w}_k x_k + \nu_n, \quad n = 1, \cdots, N_t. \quad (1)$$

N_t is the number of transmit antennas and also the number of scheduled users; $\mathbf{h}_n$ is the downlink channel vector ($N_t \times 1$); P is the transmit power; $\mathbf{w}_n$ is the beamforming vector ($N_t \times 1$) with $\|\mathbf{w}_n\|^2 = 1$; x_n is the transmitted symbol with $|x_n| = 1$; y_n is the received symbol; and ν_n is the AWGN sample with $\nu_n \in \mathcal{CN}(0, 1)$.

For the purpose of asymptotic analysis of LF-OSDMA, we make the following assumption:

AS 1: *The downlink channel $\mathbf{h}_u$ $\forall$ $u = 1, 2, \cdots, U$ is an i.i.d. vector with $CN(0, 1)$ coefficients.*

Given this assumption, which is commonly made in the literature of SDMA and multi-user diversity [10], [12], [15], the channel direction vector $\mathbf{h}_u/\|\mathbf{h}_u\|$ of each user follows a uniform distribution, which greatly simplifies the analysis of LF-OSDMA in Section IV.

III. Algorithms

The proposed LF-OSDMA algorithm is comprised of sub-algorithms for (i) CSI quantization at the mobiles and (ii) joint scheduling and beamforming at the base station, which are discussed in the following sections.

A. CSI Quantization

Without loss of generality, the discussion in this section is focused on the uth user and the same algorithm for CSI quantization is used by other users. We assume:

AS 2: *The uth user has perfect receive CSI $\mathbf{h}_u$.*

This assumption allows us to neglect channel estimation error at the uth mobile. For convenience, the CSI, $\mathbf{h}_u$, is decomposed into two components: the *gain* and the *shape*. Hence, $\mathbf{h}_u = g_u \mathbf{s}_u$ where $g_u = \|\mathbf{h}_u\|$ is the gain and $\mathbf{s}_u = \mathbf{h}_u/\|\mathbf{h}_u\|$ is the shape. The uth user quantizes and sends back to the base station two quantities: the *channel shape* and the *signal-to-noise ratio* (SINR).

The channel shape $\mathbf{s}_u$ is quantized using a quantizer whose codebook is comprised of multiple sets of orthonormal vectors in $\mathbb{C}^{N_t}$. Let $\mathcal{F}$ denote the codebook, $\mathcal{V}^{(m)}$ the mth orthonormal set in the codebook, and M the number of such sets. Thus, $\mathcal{F} = \cup_{m=1}^{M} \mathcal{V}^{(m)}$. For our design, the M orthonormal vector sets of $\mathcal{F}$ are generated randomly and independently. The quantized channel shape, represented by $\hat{\mathbf{s}}_u$, is the member of $\mathcal{F}$ that forms the smallest angle with the channel shape $\mathbf{s}_u$. Mathematically,

$$\hat{\mathbf{s}}_u = \mathcal{Q}(\mathbf{s}_u) = \arg\max_{\mathbf{f}\in\mathcal{F}} \left|\mathbf{f}^\dagger \mathbf{s}_u\right|^2, \quad (2)$$

where the function $\mathcal{Q}$ represents the channel shape quantization process. It follows that the *quantization error* can be defined as $\epsilon = \sin^2(\angle(\hat{\mathbf{s}}_u, \mathbf{s}_u))$. It is clear that $\epsilon = 0$ if $|\hat{\mathbf{s}}_u^\dagger \mathbf{s}_u| = 1$ and $\epsilon = 1$ if $\hat{\mathbf{s}}_u \perp \mathbf{s}_u$.

The quantized channel shape $\hat{\mathbf{s}}_u$ is sent back to the base station through a finite-rate feedback channel [11], [16]. Since the quantization codebook $\mathcal{F}$ can be known *a priori* to both the base station and mobiles, only the index of $\hat{\mathbf{s}}_u$ needs to be sent back. Therefore, the number of feedback bits per user is $\log_2 N$ since $|\mathcal{F}| = N$.

Besides the channel shape, the uth user also sends back to the base station the SINR, which serves as a channel quality indicator. For orthogonal beamforming, the SINR is [10]

$$\text{SINR}_u = \frac{P\rho_u(1-\epsilon_u)}{1 + P\rho_u\epsilon_u}. \quad (3)$$

Since the SINR is a scalar and requires much fewer feedback bits than the channel shape, we assume:

AS 3: *The SINR$_u$ is perfectly known to the base station through feedback.*

The same assumption is also made in [5], [10], [12].

B. Joint Scheduling and Beamforming

To maximize the sum capacity, N_t scheduled users must be selected through an exhaustive search, which is infeasible for a large user pool. Therefore, we propose a joint scheduling and beamforming algorithm that has low-complexity but outperforms conventional ones. In brief, the proposed algorithm schedules a subset of users with orthogonal quantized channel shapes and furthermore apples these channel shapes as the scheduled users' beamforming vectors.

The proposed joint scheduling and beamforming algorithm is elaborated as follows. Let $\mathbf{v}_n^{(m)}$ denote the nth vector of the mth orthonormal subset $\mathcal{V}^{(m)}$ of the codebook $\in \mathcal{F}$. First, each member of $\mathcal{F}$, for instance $\mathbf{v}_n^{(m)}$, is assigned to the user with the maximum SINR among the subset of feedback users whose quantized channel shapes are $\mathbf{v}_n^{(m)}$. Thus, the user assigned to $\mathbf{v}_n^{(m)}$ has the following SINR

$$\xi_n^{(m)} = \max_{u\in\mathcal{I}_n^{(m)}} \text{SINR}(\rho_n, \epsilon_n), \quad (4)$$

where the index set $\mathcal{I}_n^{(m)} = \left\{1 \leq u \leq U : \hat{\mathbf{s}}_u = \mathbf{v}_n^{(m)}\right\}$. If $|\mathcal{I}_n^{(m)}| = \emptyset$, $\xi_n^{(m)}$ is set as zero. Second, among the M orthonormal subsets in the codebook $\mathcal{F}$, the set that yields the maximum sum capacity is chosen. It follows that the maximum sum capacity conditional on the multi-user channel realization $\mathcal{H} = \{\rho_u, \mathbf{s}_u\}$ can be written as

$$\mathcal{C}(\mathcal{H}) = \max_{1 \leq m \leq M} \sum_{n=1}^{N_t} \log_2 \left(1 + \xi_n^{(m)}\right), \qquad (5)$$

where $\xi_n^{(m)}$ is in (4). Let $\tilde{m}$ denote the index chosen by the "max" operator in (5). Following the above algorithm, the users assigned to the vectors in $\mathcal{V}^{(\tilde{m})}$ are scheduled using beamforming vectors from $\mathcal{V}^{(\tilde{m})}$.

C. Algorithm Comparison

The proposed LF-OSDMA algorithm are compared with existing algorithms, namely ZF-OSDMA, OSDMA-S and OSDMA, as follows. First, LF-OSDMA requires no broadcast of beamforming vectors from the base station in contrast to OSDMA and OSDMA-S. In particular, the multiple rounds of broadcast required for OSDMA-S cause feedback delay that has negative impact on the sum capacity [17]. Second, using the joint scheduling and beamforming algorithm proposed in the last section, LF-OSDMA avoids the zero-forcing operation required by ZF-OSDMA for computing beamforming vectors for scheduled users.

IV. Capacity Analysis

In this section, we analyze (i) the gain in sum capacity for LF-OSDMA by increasing the amount of CSI feedback, and (ii) the effect of CSI quantization on the sum capacity as the SNR scales up.

A. Capacity Gain for Increasing CSI Feedback

This section considers a large number of users and focuses on the capacity gain for LF-OSDMA by increasing the number of CSI feedback bits of each user, denoted as B. Note that under the assumption AS 3, $B = \log_2(MN_t)$ where M is the number of orthonormal vector sets in the codebook $\mathcal{F}$. Therefore, it is equivalent to analyze how the value of M affects the capacity gain of LF-OSDMA.

The bounds of the sum capacity are derived for LF-OSDMA as follows. Let $\bar{\mathcal{C}}(U)$ denote the sum capacity of LF-OSDMA, where U is the number of users. Moreover, the sum capacity is defined as $\bar{\mathcal{C}}(U) = E[\mathcal{C}(\mathcal{H})]$ where $\mathcal{C}(\mathcal{H})$ is given in (5). We obtain an asymptotic upper bound for $\bar{\mathcal{C}}(U)$ as shown in Theorem 1.

Theorem 1: *For a fixed codebook size $N = MN_t$ and a large number of users ($U \to \infty$), the sum capacity of LF-OSDMA is bounded as*

$$N_t \log_2 \log_2 U \leq \lim_{U \to \infty} \bar{\mathcal{C}}(U) \leq N_t \log_2 \log_2 (MU). \qquad (6)$$

The proof is given in [18]. Consider OSDMA [10] as the special case of LF-OSDMA with the smallest codebook size $N = N_t$ or $M = 1$. As shown in [10], the lower bound in (6) gives the asymptotic sum capacity of OSDMA. Therefore, with respect to OSDMA, the capacity upper bound in (6) for LF-OSDMA implies that to some extent, increasing the value of M from one has the virtual effect of expanding the size of user pool from U to MU.

The capacity gain for LF-OSDMA, denoted as $\Delta\bar{\mathcal{C}}(U)$, is defined as the difference between the sum capacity of LF-OSDMA and that of OSDMA [10]. Mathematically,

$$\Delta\bar{\mathcal{C}}(U) = \bar{\mathcal{C}}(U) - N_t \log_2 \log_2 U, \qquad (7)$$

where the last term is the asymptotic sum capacity for OSDMA [10]. An upper-bound for $\Delta\mathcal{C}_U(I)$ is shown in the following corollary.

Corollary 1: *There exist an integer U_0 such that for $U \geq U_0$, the capacity gain is bounded as*

$$0 \leq \Delta\bar{\mathcal{C}}(U) \leq N_t \log_2 M / \log_2 U. \qquad (8)$$

The proof is given in [18]. Two remarks are in order. First, from (8), for a fixed M and equivalently a fixed number of feedback bits per user, the upper bound of the capacity gain $\Delta\bar{\mathcal{C}}(U)$ decreases logarithmically with the number of users U, indicating a diminishing capacity gain as U increases. Second, for a fixed U, the upper bound of $\Delta\bar{\mathcal{C}}(U)$ increases with M also logarithmically.

B. Effect of CSI Quantization

This section focuses on the effect of CSI quantization on the sum capacity of LF-OSDMA for the increasing SNR and a fixed number of users. The main results of this section are summarized as follows. First, despite the increase in SNR, the sum capacity of LF-OSDMA is upper bounded by a function that is approximately linear to the logarithm of the quantization codebook size. Second, by increasing the codebook size as a polynomial of the SNR, the sum capacity scales up with the SNR without being upper bounded. The analysis in this section is inspired by that in [15], where similar results as described above are derived for the broadcast channel with limited feedback and zero-forcing transmit beamforming. Note that our work differs from [15] by considering orthogonal beamforming and scheduling.

The main results of this section are summarized in the following theorem.

Theorem 2: *The sum capacity of LF-OSDMA is bounded as*

$$\bar{\mathcal{C}} > \left[N_t \log_2(1 + N_t P) - P \left(\frac{N}{N_t}\right)^{-\frac{1}{N_t - 1}}\right]^+, \qquad (9)$$

$$\bar{\mathcal{C}} < \frac{U 2^{N_t - 1}(\log_2 N - \log_2 N_t + \log_2 e)}{N_t(N_t - 1)} + 1, \qquad (10)$$

where the operator $[a]^+ = \max(a, 0)$.

The proof is given in [18]. The implications of the results in Theorem 1 are elaborated as follows. First, the capacity upper bound in (10) is a constant independent of the SNR P for the case considered in this section, where the codebook size N as well as the number of users U and the number of antennas at the base station N_t are fixed. In other words, the sum capacity of LF-OSDMA is limited due to CSI quantization despite the increasing SNR. A similar observation is made in [15] for SDMA with zero-forcing beamforming and limited feedback. Second, the capacity lower bound in (9) implies that increasing the codebook size N linearly with P^{N_t-1} allows the sum capacity of LF-SDMA to grow at least linearly with $N_t \log_2(1 + N_t P)$. To justify this claim, consider a constant $\alpha \in \mathbb{R}^+$ and set $P(N/N_t)^{-1/(N_t-1)} = \alpha$. It follows that N is proportional to P^{N_t-1} since

$$N = (P^{N_t-1})(\alpha^{N_t-1} N_t). \quad (11)$$

From (11) and (9) and by choosing $\alpha < N_t \log_2(1 + N_t P)$,

$$\bar{C} > N_t \log_2(1 + N_t P) - \alpha. \quad (12)$$

The last equation justifies the early claim that the sum capacity grows linearly with $N_t \log_2(1+N_t P)$ if the codebook size N is increases with the SNR P as in (11). Note that this capacity scaling also implies that the spatial multiplexing gain is achieved.

V. Performance Comparison

In Fig. 2, the sum capacities of LF-OSDMA are compared with those of ZF-OSDMA [5] and OSDMA-S [12] for a increasing number of users U. The number of transmit antenna is $N_t = 4$ and the SINR is 5 dB. For LF-OSDMA and ZF-OSDMA, the codebook sizes are $N = 64$ and hence the CSI feedback is 6 bits per user. For OSDMA-S, the CSI feedback summed over multiple iterations is also constrained to be no larger than 6 bits per user. For fair comparison, following the setup for OSDMA-S in [12], a CSI feedback overhead factor of $\lambda = 5\%$ is applied, which reduces the sum capacity by the factor λ for each round of CSI feedback. Hence, the sum capacity with feedback overhead is $\mathcal{C}_\lambda = (1 - K\lambda)\mathcal{C}$ where K is the number of rounds for CSI feedback and $\mathcal{C}$ is the sum capacity without considering feedback overhead. For OSDMA-S, K takes on the optimal values obtained in [12]. Note that $K = 1$ for LF-OSDMA and ZF-OSDMA.

As observed from Fig. 2, for $U \geq 40$, LF-OSDMA achieves a capacity gain up to 1.5 b/s/Hz with respect to ZF-OSDMA and OSDMA-S. For $U \leq 40$, LF-OSDMA is outperformed by ZF-OSDMA and OSDMA-S in terms of sum capacity.

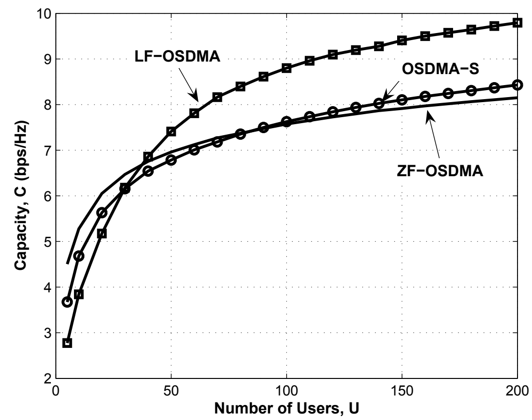

Fig. 2. Performance comparison between LF-OSDMA, ZF-OSDMA and OSDMA-S for the number of antennas $N_t = 4$, CSI feedback no larger than 6 bits/user and SNR = 5 dB.

References

[1] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. on Info. Theory*, vol. 49, no. 8, pp. 1912–21, 2003.

[2] M. Costa, "Writing on dirty paper," *IEEE Trans. on Info. Theory*, vol. 29, no. 3, pp. 439 – 441, 1983.

[3] K.-K. Wong, R. Murch, and K. Letaief, "A joint-channel diagonalization for multiuser MIMO antenna systems," *IEEE Trans. on Wireless Communications*, vol. 2, no. 4, pp. 773–786, 2003.

[4] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 461 – 471, 2004.

[5] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *to appear, IEEE Journal Sel. Areas in Communications*, 2007.

[6] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. on Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004.

[7] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. on Signal Processing*, vol. 52, pp. 214–226, Jan. 2004.

[8] L.-U. Choi and R. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. on Wireless Communications*, vol. 3, no. 1, pp. 20–24, 2004.

[9] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc., IEEE Intl. Conf. on Communications*, vol. 1, (Seattle, WA), pp. 331–335, 1995.

[10] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. on Info. Theory*, vol. 51, pp. 506–522, Feb. 2005.

[11] 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, pp. 54–59, Oct. 2004.

[12] W. Choi, A. Forenza, J. G. Andrews, and R. W. Heath Jr., "Opportunistic space division multiple access with beam selection," *to appear, IEEE Trans. on Communications*.

[13] C.-K. A. Yeung and D. J. Love, "Performance analysis of random vector quantization limited feedback beamforming," in *Proc., IEEE Asilomar*, pp. 408–412, Oct. 2005.

[14] K.-B. Huang, R. W. Heath Jr., and J. G. Andrews, "Space division multiple access with a sum feedback rate constraint," *accepted to IEEE Int. Conf. Acoust., Speech and Sig. Proc. 2007; available at: http://www.ece.utexas.edu/∼rheath/papers/2006/ICASSP1/*, Sept. 2006.

[15] N. Jindal, "MIMO broadcast channels with finite rate feedback," *submitted to IEEE Trans. Information Theory*, 2006.

[16] 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, pp. 2735–47, Oct. 2003.

[17] K.-B. Huang, B. Mondal, R. W. Heath, Jr., and J. G. Andrews, "Multi-antenna limited feedback systems for temporally-correlated channels: Feedback rate and effect of feedback delay," *submitted to IEEE Trans. on Info. Theory; available at ArXiv: http://arxiv.org/abs/cs.IT/0606022*, 2006.

[18] K.-B. Huang, R. W. Heath Jr., and J. G. Andrews, "Joint beamforming and scheduling for SDMA systems with limited feedback," *submitted to IEEE Trans. on Comm.; available at ArXiv: http://arxiv.org/abs/cs.IT/0606121*, June 2006.