ORTHOGONAL BEAMFORMING FOR SDMA DOWNLINK WITH LIMITED FEEDBACK


The University of Texas at Austin
{huangkb, 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 orthogonal beamforming and scheduling, which is enabled by feedback of quantized channel state information. In this approach, the base station generates multiple sets of orthogonal beamforming vectors and associates each vector with a specific user. Subsequently, the base station selects one set of beamforming vectors and the associated users for downlink transmission such that the throughput is maximized. Unlike an exhaustive approach, the proposed algorithm has computational complexity that increases only linearly with the number of users since it does not require an exhaustive search over all possible subsets of users.

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

1. 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 [9–11].

This paper is focused on joint beamforming and scheduling for SDMA systems with the objective of maximizing sum capacity. To avoid an exhaustive search, a practical algorithm, named opportunistic SDMA (OSDMA), is proposed in [10], where an arbitrary set of orthogonal beamforming vectors is chosen from the optimal asymptotic scaling rate for the sum capacity with the number of users. Nevertheless, for a small number of users, such arbitrary beamforming vectors are highly sub-optimal due to excessive interference between scheduled users. As proposed in [12], the beamforming vectors for OSDMA can be improved using beam and user selection, in which a set of orthogonal beamforming vectors is chosen from multiple sets (beam selection) and assigned to a subset of users (user selection) for downlink transmission. The drawback of this algorithm, named OSDMA with beam selection (OSDMA-S), is the required numerous iterations of broadcast and feedback due to beam selection distributed at mobiles. The feedback cost incurred by the iterations reduces the capacity gain of OSDMA-S over OSDMA.

To overcome the drawback of OSDMA-S due to distributed beam selection, we propose an OSDMA algorithm with centralized beam and user selection. First, the required channel state information (CSI) is sent to the base station through quantized CSI feedback, also known as limited feedback [13], which gives the name of the proposed algorithm as limited feedback OSDMA (LF-OSDMA). Second, using quantized CSI feedback from users, the base station iteratively searches for an optimal set of scheduled users and their optimal orthogonal beamforming vectors to maximize sum capacity. It is worth mentioning that LF-OSDMA as well as 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 required amount of quantized feedback for ensuring optimal asymptotic sum capacity. First, the beam and user selection is shown using a sum capacity upper bound to have the virtual effect of expanding the user set. Second, the capacity gain contributed by beam and user selection is shown to increase logarithmically with the number of iterations for beam and user selection but decrease also logarithmically with the number of users $U$. Third, we prove that increasing the number of CSI feedback bits in proportion to $U$ guarantees optimal growth in sum capacity with $U$, avoiding the potential saturation of sum capacity due to CSI quantization [15]. Numerical results show that LF-OSDMA achieves significant gains in sum capacity with respect to OSDMA and OSDMA-S.

2. SYSTEM MODEL
In Fig. 1, a base station with $N_r$ 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. assigning a beamforming vector to each of the $N_t$ active users. The beamforming vectors $\{w_n\}_{n=1}^{N_t}$ are selected from multiple sets of unitary orthogonal vectors following

*Kaibin Huang is the recipient of a Motorola Partnerships in Research Grant. This work is funded the National Science Foundation under grants CCF-514194 and CNS-435307.
the beam and user selection algorithm described in Section 3.2. The received signal of the \( n \)th scheduled user is expressed as

\[
y_n = \sqrt{P} h_n^\dagger \sum_{k=1}^{N_t} \mathbf{w}_k x_k + \nu_n, \quad n = 1, \ldots, N_t, \tag{1}
\]

where we use the following notation:

- \( N_t \) number of transmit antennas and also number of scheduled users;
- \( h_n \) \((N_t \times 1)\) vector downlink channel;
- \( x_n \) transmitted symbol with \(|x_n| = 1\);
- \( y_n \) received symbol;
- \( \mathbf{w}_n \) \((N_t \times 1)\) beamforming vector with \( ||\mathbf{w}_n||^2 = 1\);
- \( P \) transmit SNR; and
- \( \nu_n \) AWGN sample with \( \nu_n \sim CN(0, 1) \).

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

**AS 1** The downlink channel \( h_n \forall u = 1, 2, \ldots, U \) 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 \( h_n/||h_n|| \) of each user follows a uniform distribution, which greatly simplifies the analysis of LF-OSDMA in Section 4.

## 3. ALGORITHMS

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

### 3.1. CSI Quantization

Without loss of generality, the discussion in this section is focused on the \( u \)th user and the same algorithm for CSI quantization is used by each user. We assume:

**AS 2** The \( u \)th user has perfect receive CSI \( h_u \).

This assumption allows us to neglect channel estimation error at the \( u \)th mobile. For convenience, the CSI \( h_u \), is decomposed into two components: the gain and the shape, which are quantized separately. Hence, \( h_u = g_u s_u \) where \( g_u = ||h_u|| \) is the gain and \( s_u = h_u/||h_u|| \) is the shape. Since the channel gain is a scalar and hence easy to quantize, we make the following assumption

\[
\text{AS 3} \quad \text{The channel gain } g_u \text{ is perfectly known to the base station through feedback.}
\]

Similar assumptions are also made in [10, 12]. AS 3 allows us to focus discussion on quantization of the channel shape.

The random vector quantization (RVQ) method [15] is applied for quantizing the channel shape \( s_u \). First, this method requires random generation of \( N \) unitary vectors, which follow the uniform distribution. These vectors as a group are called a codebook, denoted as \( \mathcal{F} \). Second, RVQ selects from the codebook \( \mathcal{F} \) a member that forms the smallest angle with the channel shape \( s_u \) as in [16]. The selected member gives the quantized channel shape, denoted as \( \hat{s}_u \). Mathematically,

\[
\hat{s}_u = Q(s_u) = \arg \max_{f \in \mathcal{F}} \left| f^\dagger s_u \right|, \tag{2}
\]

where the function \( Q \) represents the CSI quantization process. We define the quantization error as the angle between \( \hat{s}_u \) and \( s_u \), denoted as \( \angle(s_u, \hat{s}_u) \). It is clear that the quantization error is zero if \( \hat{s}_u = s_u \). Again, due to the ease of sending back a scalar quantity, we make the following assumption similar to AS 3:

**AS 4** The quantization error \( \angle(s_u, \hat{s}_u) \) is perfectly known to the base station through feedback.

The quantized channel shape \( \hat{s}_u \) is sent back to the base station through a finite-rate feedback channel [13, 16]. Since the quantization codebook \( \mathcal{F} \) can be known a priori to both the base station and mobiles, only the index of \( \hat{s}_u \) needs to be sent back. Therefore, the number of feedback bits per user is \( \log_2 N \) since \( |\mathcal{F}| = N \).

### 3.2. Centralized Beam and User Selection

Having collected quantized CSI from all \( U \) users, the base station schedules \( N_t \) users for transmission and computes their beamforming vectors. 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 more practical scheduling algorithm whose complexity is scalable with the base station’s computational power.

The proposed algorithm consists of \( I \) independent iterations for beam and user selection. The procedure of each iteration is described as follows. \( N_t \) unitary beamforming vectors \( \{\mathbf{w}_u^{(i)}\} \) are randomly generated. From these vectors and the \( U \) users, a combination of a user and a beamforming vector is chosen such that the following SINR lower bound\(^1\) is maximized [17]

\[
\text{SINR}(\rho^*, \angle(s^*, \mathbf{w}^*)) \geq \text{SINR}(\rho^*, \angle(s^*, \hat{s}^*) + \angle(\hat{s}^*, \mathbf{w}^*)) \tag{3}
\]

where \( \rho^* \), \( s^* \) and \( \hat{s}^* \) are the channel power, the original and the quantized channel shapes of the chosen user respectively and \( \mathbf{w}^* \) is the chosen beamforming vector. The function SINR(\( \cdot, \cdot \)) is given as [17]

\[
\text{SINR}(\alpha, \beta) = (1 + P\alpha)/(1 + P\alpha \sin^2(\beta)) - 1. \tag{4}
\]

The above beam and user selection process is repeated on the remaining beamforming vectors and users till each of \( N_t \) vectors is associated with a specific user.

Let \( i_1, i_2, \ldots, i_{N_t} \) denote the indices of the chosen users for the \( i \)th iteration. The expected sum capacity\(^2\) for the \( i \)th iteration can be

\[1\text{The actual SINR is unknown to the base station due to CSI quantization.}

\[2\text{The expected sum capacity computed at the base station differs from the actual value due to CSI quantization error.}\]
written as
\[ R^{(i)} = \sum_{k=1}^{N_I} \log_2 \left[ 1 + \text{SINR} \left( \rho_{i,k}, \angle \left( \mathbf{s}_{i,k}, \mathbf{w}_k \right) \right) \right]. \] (5)

After \( I \) iterations, the expected sum capacity is \( R^* = \max_{1 \leq i \leq I} R^{(i)} \).

### 3.3. Algorithm Comparison

The key differences between the proposed LF-OSDMA algorithm and existing algorithms, namely OSDMA-S and OSDMA, are described 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 [18]. Second, while OSDMA and OSDMA-S require feedback of each user’s choice of a beamforming vector, LF-OSDMA demands feedback of quantized CSI from each user. Therefore, the feedback for LF-OSDMA can be compressed by exploiting channel temporal correlation [18], which is however inapplicable for OSDMA and OSDMA-S. Last, compared with OSDMA and OSDMA-S, LF-OSDMA requires more computation at the base station, reflected in the multiple (\( I \)) iterations. This may be a desirable tradeoff given the higher computational power at the base station.

### 4. CAPACITY ANALYSIS

In this section, we analyze the gain in sum capacity due to beam and user selection for LF-OSDMA, and also the requirement on CSI quantization for ensuring that the sum capacity increases with the number of users following the optimal rate.

### 4.1. Capacity Gain for Beam and User Selection

The analysis in this section is focused on the capacity gain for beam and user selection in LF-OSDMA with respect to OSDMA [10], which can be considered as LF-OSDMA with only one iteration for beam and user selection. For simplicity, CSI quantization error is neglected and its analysis is postponed to Section 4.2. In other words, we make the following assumption just for Section 4.1.

**AS 5** In the capacity gain analysis of beam and user selection for LF-OSDMA, CSI feedback is assumed perfect.

Hence, \( \hat{s}_u = s_u \) for \( u = 1, \ldots, U \). Roughly speaking, the effect of CSI quantization is to offset the capacity gain of beam and user selection. The analysis in this section is carried out by first obtaining an upper-bound for the sum capacity of LF-OSDMA and subsequently an upper-bound for the capacity gain of LF-OSDMA due to beam and user selection.

To simplify notation, let \( \text{SINR}_{u,n}^{(i)} \) denote the SINR for the \( u \)th user who is assigned the \( n \)th beamforming vector generated in the \( i \)th iteration, hence \( \text{SINR}_{u,n}^{(i)} = \text{SINR}(\rho_u, \angle(s_u, \mathbf{w}_n^{(i)})) \) where the function \( \text{SINR}(\cdot) \) is given in (4). With AS 5 and beam and user selection as described in Section 3.2, the sum capacity for LF-OSDMA is given as

\[ C_U(I) = E \left[ \max_{1 \leq i \leq I} \sum_{n=1}^{N_I} \log_2 \left( 1 + \max_{1 \leq u \leq U} \text{SINR}_{u,n}^{(i)} \right) \right]. \] (6)

Using these lemmas, we obtain an asymptotic upper bound for the sum capacity of LF-OSDMA as shown in Theorem 1.

**Theorem 1** Let the number of iterations for beam and user selection \( I \) be fixed and the CSI feedback be perfect. Then the sum capacity for LF-OSDMA is bounded as

\[ N_t \log_2 U \leq \lim_{U \to \infty} C_U(I) \leq N_t \log_2 \log_2(U). \] (7)

The proof is given in [17]. The upper bound in (7) is equal to the sum capacity of OSDMA with \( U \) IT users. By treating OSDMA as the special case of LF-OSDMA with one iteration for beam and user selection, this upper bound implies that to some extent, the beam and user selection in LF-OSDMA has the virtual effect of expanding the size of user pool from \( U \) to \( U \).

Next, we investigate the capacity gain of beam and user selection with respect to no beam selection (hence OSDMA), which is defined as

\[ \Delta C_U(I) = C_U(I) - N_t \log_2 \log_2 U, \] (8)

where \( C_U(I) \) is given in (6) and the last term is the asymptotic sum capacity for OSDMA [10]. An upper-bound for \( \Delta C_U(I) \) is shown in the following corollary.

**Corollary 1** There exist an integer \( U_0 \) such that for \( U \geq U_0 \), the capacity gain due to beam and user selection is bounded as

\[ 0 \leq \Delta C_U(I) \leq N_t \log_2 I/I \log_2 U. \] (9)

The proof is given in [17]. Two remarks are in order:

1. From (9), for a fixed number of iterations for beam and user selection, the upper bound of the capacity gain \( \Delta C_U(I) \) decreases logarithmically with the number of users \( U \).
2. For a fixed number of users, the upper bound of \( \Delta C_U(I) \) increases with the number of iterations also logarithmically.

### 4.2. Requirement on CSI Quantization

An interesting question to ask is: *how much CSI feedback is sufficient for LF-OSDMA?* We answer this question by proving that increasing CSI feedback bits with the number of users ensures the optimal sum capacity scaling law [10]. This result is stated as the following theorem.

**Theorem 2** Consider LF-OSDMA with finite iterations for beam and user selection. If the quantization codebook size \( N \) increases with the number of users \( U \) as

\[ N = \alpha (\log_2 U)^{N_t - 1} + \beta, \] (10)

with \( 0 < \alpha < \infty \) and \( |\beta| < \infty \), the sum capacity \( C_U(I) \) in (6) converges as:

\[ \lim_{U \to \infty} \frac{C_U(I)}{N_t \log_2 \log_2 U} = 1. \] (11)

The proof is given in [17]. Even though the values of \( \alpha \) and \( \beta \) in (10) have no effect on the asymptotic value of the sum capacity in (11) as shown by the above theorem, they should be small, e.g. \( \alpha = 1 \) and \( \beta = 0 \), to give reasonable values for the quantization codebook size \( N \). For a large number of users (\( U \to \infty \)) and perfect CSI feedback, since the capacity gain of LF-OSDMA over OSDMA converges to zero (cf. Corollary 1), the sum capacity of LF-OSDMA increases with the number of users at the same optimal rate as OSDMA, namely \( N_t \log_2 \log_2 U \) [10]. Theorem 2 shows that as long as the size of the quantization codebook increases following (10), asymptotically CSI quantization does not affect the optimal growth rate of the sum capacity \( C_U \) with \( U \).
Fig. 2. Performance comparison between LF-OSDMA, OSDMA and OSDMA-S for four transmit antennas. For LF-OSDMA, the number of iterations is \( I = U \); the quantization codebook size is \( N = 5^\frac{1}{2U} \).

5. PERFORMANCE COMPARISON

The sum capacities of LF-OSDMA are compared with those of OSDMA and OSDMA-S in Fig. 2. The number of transmit antenna is \( N_t = 4 \) and the SINR is 10 dB. For LF-OSDMA, the codebook size and the number of iterations for beam and user selection scale with the number of users as \( N = 5^\frac{1}{2U} \) and \( I = U \), respectively. 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 \( \lambda \) for each round of CSI feedback. Hence, the sum capacity with feedback overhead is \( C = (1 - K \lambda) C \) where \( K > 1 \) is the number of rounds for CSI feedback and \( C \) is the sum capacity without considering feedback overhead. For OSDMA-S, the number of feedback rounds, hence the beam and user selection iterations, follows the optimal values obtained in [12]. As observed from Fig. 2, LF-OSDMA achieves a capacity gain ranging from 0.8 to 1.6 b/s/Hz with respect to OSDMA-S and 1.1 to 1.7 b/s/Hz with respect to OSDMA. The required numbers of feedback bits per user for LF-OSDMA, OSDMA-S and OSDMA can be computed to be 6–10 bits, 4–6 bits and 2 bits, respectively.

6. CONCLUSION

This paper proposes a new algorithm for SDMA downlink with centralized beamforming and scheduling, called LF-OSDMA. We derived an upper bound for its capacity gain due to beam selection with respect to no beam selection. Moreover, we showed that increasing the quantization codebook size ensures that the sum capacity grows optimally with the number of users. Numerical results showed that LF-OSDMA can achieve significant gains in sum capacity with respect to OSDMA-S and OSDMA at the cost of modest computational complexity (base station) and additional feedback.

7. REFERENCES


