Balance of Multiuser Diversity and Multiplexing Gain in Near-Orthogonal MIMO Systems with Limited Feedback

Ruben de Francisco*, Dirk T.M. Slock* and Ying-Chang Liang†
*Eurecom Institute
Sophia-Antipolis, France
Email: {defranci, slock}@eurecom.fr
†Institute for Infocomm Research
Singapore
Email: ycliang@i2r.a-star.edu.sg

Abstract—Near-orthogonal transmission over MIMO broadcast channels with limited feedback is addressed, identifying the intrinsic tradeoffs that govern the system performance in terms of sum rate. A low-complexity scheme is proposed for joint scheduling and beamforming, which employs feedback in the form of quantized channel directions and lower bounds on each user’s received SINR, given a maximum orthogonality factor $\epsilon$ between transmit beamforming vectors ($\epsilon = 0$ if orthogonal beamforming). We assume a simple power allocation, which consists of distributing the available power equally over the active beams. In a system with $K$ active users and a given average SNR, we show that the sum rate is a function of the number of antennas, the number of active beams (considered less than or equal to the number of antennas), the orthogonality factor $\epsilon$ and the quantization codebook size. A practical rate function is derived for the described system which approximates the average sum rate accurately as validated through computer simulations. Based on the proposed rate function, optimality of SDMA vs. TDMA systems with asymptotically large number of users is shown. In addition, we provide a performance comparison between the proposed algorithm with limited feedback and other approaches with different level of channel state information at the transmitter.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) systems can significantly increase the spectral efficiency by exploiting the spatial degrees of freedom created by multiple antennas. In point-to-point MIMO systems, the capacity increases linearly with the minimum of the number of transmit/receive antennas, irrespective of the availability of channel state information (CSI) [1],[2]. In the MIMO broadcast channel, it has recently been proven [3] that the sum capacity is achieved by dirty paper coding (DPC) [4]. However, the applicability of DPC is limited due to its computational complexity and the need for full channel state information at the transmitter (CSI). Several downlink techniques based on Space Division Multiple Access (SDMA) have been proposed [5], achieving the same asymptotic sum rate as that of DPC.

The capacity gain of multiuser MIMO systems is highly dependent on the available CSIT. While having full CSI at the receiver can be assumed, this assumption is not reasonable at the transmitter side. Several limited feedback approaches have been considered in point-to-point systems [6], [7], [8], where each user sends to the transmitter the index of a quantized version of its channel vector from a codebook. An extension for MIMO broadcast channels is made in [9], in which each mobile feeds back a finite number of bits regarding its channel realization at the beginning of each block based on a codebook. An SDMA extension of opportunistic beamforming [10] using partial CSIT in the form of individual signal-to-interference-plus-noise ratio (SINR) is proposed in [11], achieving optimum capacity scaling for large number of users.

In [12],[13], a simple scheme for joint scheduling and beamforming based on limited feedback is proposed. The receivers compute and feed back a scalar metric that can be interpreted as an upper bound on the SINR. Note that a scheme with similar metric is also reported in [14]. In this paper, we propose instead a system in which each mobile user feeds back a lower bound on the SINR given certain orthogonality conditions that need to be satisfied by the beamforming vectors. The proposed system can ensure a lower bound on each user’s achievable rate thus avoiding outage events, without need of link adaptation (a second step of feedback).

Near-orthogonality transmission was discussed in [15] for MIMO broadcast channels, as an efficient technique that can guarantee certain per-user SIR. In [16], a simple power allocation technique is proposed for Random Beamforming (RBF), which consists of activating/deactivating certain beams for transmission, performing equal power allocation among the active beams. In our work, we combine this type of simple power allocation policy with near-orthogonal transmission. The ability of this system to adapt the number of transmit beams and orthogonality properties gives rise to the need for
joint optimization in order to maximize the system throughput. This provides a simple - yet practical - way to balance the available multiuser diversity and multiplexing gain. The system increases the multiplexing gain by activating more beams, while a relaxation in the orthogonality conditions for user scheduling provides an increase in multiuser diversity. We introduce a rate function approximation that enables an easier analysis of such systems and provide simulations to illustrate its validity. Numerical results are further given to observe the behavior of the system sum rate with different levels of CSIT.

II. System Model

We consider a multiple antenna broadcast channel consisting of $M$ antennas at the transmitter and $K \geq M$ single-antenna receivers. The received signal $y_k$ of the $k$-th user is mathematically described as

$$y_k = h_k^T x + n_k, \quad k = 1, \ldots, K$$

(1)

where $x \in \mathbb{C}^{M \times 1}$ is the transmitted signal, $h_k \in \mathbb{C}^{M \times 1}$ is the channel vector, and $n_k$ is additive white Gaussian noise at receiver $k$. We assume that each of the receivers has perfect and instantaneous knowledge of its own channel $h_k$, and that $n_k$ is independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian with zero mean and unit variance. The transmitted signal is subject to an average transmit power constraint $P$, i.e., $\mathbb{E}([|x|^2]} = P$. For the throughput analysis, we consider an homogeneous network where all users have the same signal-to-noise ratio (SNR). Due to the noise variance normalization to one, $P$ takes on the meaning of average SNR.

We consider an i.i.d. block Rayleigh flat fading channel, whose parameters are considered invariant during each coded block, but are allowed to vary independently from block to block. We focus on the ergodic sum rate, which means that the capacity is averaged over the fading distribution, and thus the block size does not affect our results. Let $H \in \mathbb{C}^{K \times M}$ refer to the concatenation of all channels, $H = [h_1 \ h_2 \ldots \ h_K]^T$, where the $k$-th row is the channel of the $k$-th receiver. Define $Q$ as the set of all possible subsets of cardinality $M$ of disjoint indices among the complete set of user indices $K = \{1, \ldots, K\}$. Let $S \in Q$ be one such group of $M$ users selected for transmission at a given time slot. Then $H(S)$, $W(S)$, $s(S)$, $y(S)$ are the concatenated channel vectors, beamforming vectors, uncorrelated data symbols and received signals respectively for the set of scheduled users $S$. When concatenating the beamforming matrix $W(S)$ prior to transmission, the signal model can be described as follows

$$y(S) = H(S)W(S)s(S) + n$$

(2)

At the $k$-th mobile, the received signal is given by

$$y_k = \sum_{i \in S} h_k^T w_i s_i + n_k, \quad k = 1, \ldots, K$$

(3)

For the selected set of users $S$ scheduled for transmission, the beamforming matrix is given by

$$W(S) = V(S)\Lambda(S)^{1/2}$$

(4)

where the columns of $V(S)$ are the normalized beamforming vectors and $\Lambda(S)$ is a diagonal power allocation matrix.

Notation: We use bold upper and lower case letters for matrices and column vectors, respectively. $(\cdot)^T$, and $(\cdot)^H$ stand for transpose and Hermitian transpose, respectively. $\mathbb{E}(\cdot)$ denotes the expectation operator. The notation $||x||$ refers to the Euclidean norm of the vector $x$, and $\angle(x, y)$ refers to the angle between vectors $x$ and $y$.

III. Near-Orthogonal Beamforming with Limited Feedback

Near-orthogonal transmission was first studied in [15] as an efficient technique that can guarantee certain per-user SIR. We define a set of $\epsilon$-orthogonal vectors with orthogonality factor $\epsilon$ as follows

$$S_\epsilon = \{S \mid |v_i^H v_j| \leq \epsilon \ \forall i, j \in S\}$$

(5)

which in fact measures the degree of non orthogonality. Given an $\epsilon$ factor fixed by the Base Station (BS), the columns of the normalized beamforming matrix $V(S)$ are constrained to be $\epsilon$-orthogonal.

In our work, we perform joint beamforming and scheduling in a system where limited feedback is present at the transmitter side. In order to clearly identify the tradeoff between multiuser diversity and multiplexing gain in near-orthogonal transmission, we simplify the codebook design, beamforming strategy and scheduling algorithm. A greedy scheduling algorithm is introduced based on scalar feedback of each user’s lower bound on the SINR and quantized channel indexes, for the purpose of ensuring $\epsilon$-orthogonality and a minimum per-user SINR. The analysis and results provided in our work can indeed be applied for any $\epsilon \in [0, 1]$. However, we speak of near-orthogonal transmission since this value will tend to be small in order to obtain good rates and reduce the complexity of the scheduling algorithm.

The number of active beams for transmission $M_o$ and orthogonality factor $\epsilon$ are system parameters that can be adapted in order to maximize the system sum rate. As we show analytically and through simulations, given $K$ active users with a certain average SNR, a pair $\{\epsilon, M_o\}$ exists that maximizes the average sum rate. In a realistic system, these parameters can be set at every time slot for throughput maximization.

In what follows, we show in detail the different elements of the proposed approach: beamforming, power allocation, codebook design, feedback strategy and scheduling algorithm.
A. Beamforming and Power Allocation

As simple transmission technique we consider transmit matched filtering (TxMF), which consists of using as normalized beamforming vectors the quantized channel directions of users scheduled for transmission. Each user feeds back to the base station an index corresponding to its quantized channel direction from a codebook known to both the user and the base station. The design of quantization codebooks is discussed in the next subsection.

We propose a simple ON/OFF power allocation technique for transmission, which consists of transmitting equal power \( P/M \) over \( M_o \) active beams, with \( 1 \leq M_o \leq M \). Hence the value of the elements on the diagonal of the power allocation matrix \( \Lambda(S) \) is equal to \( P/M \) for active beams and zero otherwise. Even though this strategy is not optimal, it allows the system to rapidly adapt the number of beams to varying scenarios in order to improve the throughput.

B. Codebook Design

Consider a \( B \)-bit quantization codebook \( \mathcal{V}_k \) containing \( L = 2^B \) unit norm vectors in \( \mathbb{C}^M \), which is assumed to be known to both the receiver and the transmitter. The normalized channel vector of user \( k \) to be quantized is \( \overline{h}_k = h_k/ \| h_k \| \), which corresponds to the channel direction.

The optimal vector quantizer is difficult to find and the solution to this problem is not yet known. As codebook design goes beyond the scope of the paper, we adopt the geometrical framework presented in [8]. The unit norm sphere \( \mathcal{U} \) on which a random vector \( \overline{h}_k \) lies is partitioned into \( N \) decision regions \( \{ \tilde{C}_{ki} ; i = 1, \ldots, L, \forall k \} \), where \( \tilde{C}_{ki} = \{ \overline{h}_k \in \mathcal{U} : \| h_k^H v_{ki} \|^2 \geq \| h_k^H v_j \|^2, \forall j \neq i, 1 \leq j \leq L, \forall k \} \). If the channel \( \overline{h}_k \in \tilde{C}_{ki} \), the receiver \( k \) feeds back the index \( i \). Since \( \overline{h}_k \) is uniformly distributed over \( \mathcal{U} \), we have that \( \Pr\{ \overline{h}_k \in \tilde{C}_{ki} \} \approx 1/L, \forall i, k \). Thus, we can consider the following approximated quantization cell [8], [17]

\[
\tilde{C}_{ki} = \{ \overline{h}_k \in \mathcal{U} : 1 - \| h_k^H v_{ki} \|^2 \leq \delta \}, \forall i, k
\]

for \( \delta = 2^{-B/(M-1)} \). The quantization error is defined as

\[
\sin^2 \theta_k = \sin^2 (\angle(\overline{h}_k, v_k)) = 1 - \| h_k^H v_k \|^2
\]

[8], [17], where \( v_k \) is the quantized channel direction of user \( k \).

Using this framework, the cumulative distribution function (CDF) of the quantization error is given by [8], [17]

\[
F_{\sin^2 \theta_k}(x) = \begin{cases} 
\delta^{1-M} x^{M-1} & 0 \leq x \leq \delta \\
1 & x > \delta
\end{cases}
\]

C. Feedback Strategy

At each time slot, each receiver determines its ‘best’ vector from the codebook based on its current channel realization \( \overline{h}_k \), i.e., the codeword that optimizes a certain objective function. Similarly to [7], [8], we assume that each receiver quantizes its channel to the vector that maximizes the inner product

\[
v_k = \arg \max_{v \in \mathcal{V}_k} \| h_k^H v \|^2 = \arg \max_{v \in \mathcal{V}_k} \cos^2 (\angle(\overline{h}_k, v))
\]

(6)

Each user sends the corresponding quantization index back to the transmitter through an error-free and zero-delay feedback channel using \( B \) bits. Note that this model is equivalent to the finite rate feedback model proposed by [7], [9].

In order to perform the task of user scheduling, the users feed back a scalar value that corresponds to a lower bound on the expected SINR. This lower bound is a function of the channel realization and average transmit power (normalized noise power is assumed), as well as system parameters fixed by the base station:

- Maximum orthogonality factor \( \epsilon \)
- Number of active beams \( M_o \)

For user \( k \) and index set \( S \), the multiuser interference can be expressed as

\[
I_k(S) = \sum_{i \in S, i \neq k} \frac{P}{M_o} \| h_k^H v_i \|^2, \quad \text{where } T_k(S) \text{ denotes the interference over the normalized channel } \overline{h}_k.
\]

Define \( T_{UB_k} \) as the upper bound on \( T_k \) and \( \theta_k = \angle(\overline{h}_k, v_k) \). Based on the work developed in [18] for arbitrary orthogonality between beamforming vectors, we propose the following lower bound on the SINR of the \( k \)-th user:

\[
\text{SINR}_k \geq \frac{\| h_k \|^2 \cos^2 \theta_k}{\| h_k \|^2 T_{UB_k} + \frac{M_o}{M_o}}
\]

(7)

where

\[
T_{UB_k} = \alpha_k \cos^2 \theta_k + \beta_k \sin^2 \theta_k + 2 \gamma_k \sin \theta_k \cos \theta_k
\]

(8)

and

\[
\alpha_k = (M_o - 1) \epsilon^2 \\
\beta_k = \left\{ \begin{array}{ll} 0 & \text{if } M_o = 1 \\
1 + (M_o - 2) \epsilon & \text{otherwise} \end{array} \right.
\]

\[
\gamma_k = (M_o - 1) \epsilon
\]

D. Proposed Scheduling Algorithm

As our optimization criterion is to maximize the system throughput, it is desirable to schedule a set of \( M \) users with large channel gains and mutually orthogonal beamforming vectors. We propose a simple greedy approach, in which the first selected user corresponds to the one with the largest scalar metric \( \text{SINR}_k \). At the \( i \)-th selection step, \( i = 2, \ldots, M_o \), the algorithm selects the user with largest \( \text{SINR}_k \) among the remaining users that are \( \epsilon \)-orthogonal to the quantized channels of users selected in the \( i - 1 \) previous steps. An outline of the proposed scheduling algorithm is shown in Table I.

Note that this algorithm is equivalent to the one proposed in [5], [19], [20], with the difference that here we use as metric a lower bound on the SINR.
In the proposed system we assume that if the algorithm fails to find a set of $M_o$ $\epsilon$-orthogonal users, the BS sets $M_o \to M_o - 1$ and requests retransmission of $\text{SINR}_k$ feedback at the current time slot. The operation is repeated within a time slot until a set of $\epsilon$-orthogonal users is found. At the next time slot, the value of $M_o$ is restored to its initial value. This assumption makes easier the task of comparing the proposed rate function with a simulation scenario, ensuring that at each time slot there always exists an $\epsilon$-orthogonal set to be scheduled. It corresponds to a system with the ability to adapt the number of beams within each time slot, hence feedback is assumed to be exchanged fast between base and mobile station. In a system with $M = 2$ antennas, this would correspond to a switch between SDMA and TDMA transmission.

IV. RATE FUNCTION APPROXIMATION

In this section we derive a function to approximate the average sum rate that the system with near-orthogonal transmission can provide, given knowledge of each user’s SINR lower bound. Denoting the lower bound on SINR of all users with near-orthogonality condition needs to be satisfied. Hence, the $i$-th user is selected over $K$ i.i.d. random variables yielding a CDF for the maximum SINR given by $F_{s_i} = (F_s(s))^{K_i}$. Its mean value can be approximated as

$$\mathbb{E}(s_i) \approx \int_0^{1/\alpha} 1 - (F_s(s))^{K_i} ds$$

which approximates well the mean of $s_i$ for small values of $\epsilon$. An approximation of $K_i$ can be calculated through the probability that a random vector in $\mathbb{C}^{M \times 1}$ is $\epsilon$-orthogonal to a set with $i-1$ vectors in $\mathbb{C}^{M \times 1}$, which is equal to $L_{\epsilon^2}(i, M-i)$ [5]. $L_{\epsilon^2}(a, b)$ being the regularized incomplete beta function. By using the law of large numbers [20], we can find the following approximation:

$$K_i \approx K L_{\epsilon^2}(i - 1, M - i + 1)$$

The average sum rate in a system with $M_o$ active beams can be bounded as follows by using Jensen’s inequality

$$SR = \sum_{i=1}^{M_o} \mathbb{E} \left[ \log_2 (1 + s_i) \right] \leq \sum_{i=1}^{M_o} \log_2 \left[ 1 + \mathbb{E}(s_i) \right]$$

Using equation (13) and solving the integral in equation (11) for the CDF of $s$ described in (10), we obtain the following theorem after some approximations

**Theorem 1.** Given $\epsilon$-orthogonal transmission in a system with $M_o$ active beams, the average sum rate is approximated as follows

$$R_{M_o} \approx \sum_{i=1}^{M_o} \log_2 \left[ 1 + \sum_{n=1}^{K_i} \frac{(-1)^{n-1}}{n^{(M-1)}} \left( K_i \right) \frac{1}{n} \left( 1 + \frac{C_n}{\alpha} e^{\alpha} E_i \left( \frac{C_n}{\alpha} \right) \right) \right]$$

**TABLE I**

<table>
<thead>
<tr>
<th>Outline of Scheduling Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MS</strong></td>
</tr>
<tr>
<td>Compute &amp; Feedback <strong>SINR</strong> $k \geq \frac{|h_k|^2 \cos^2 \theta_k}{|h_k|^2 T_k + \frac{1}{\epsilon}}$</td>
</tr>
<tr>
<td>SNR $\approx i \approx 1$</td>
</tr>
<tr>
<td>BS</td>
</tr>
<tr>
<td>Initialize Set $S = \emptyset$</td>
</tr>
<tr>
<td>Loop: For $i : 1 \ldots M_o$ repeat</td>
</tr>
<tr>
<td>Set $SINR_i = 0$</td>
</tr>
<tr>
<td>Loop: For $k : 1 \ldots K, k \notin S$ repeat</td>
</tr>
<tr>
<td>If $SINR_k &gt; SINR_i$ $\max$ and $</td>
</tr>
<tr>
<td>$SINR_k \to SINR_i$ $\max$ and $k_i = k$</td>
</tr>
<tr>
<td>Select $k_i \to S$</td>
</tr>
</tbody>
</table>

Fig. 1. Comparison of analytical and simulated average sum rate for $K = 25$ users, $SNR = 10$ dB.
where \( C = \frac{M}{2} + (M - 1)\beta \), \( \alpha \) and \( \beta \) are as described in equation (9) and the exponential integral function is used, defined as \( E_1(x) = \int_x^\infty \frac{e^{-t}}{t} \, dt \).

In practice, since \( K \) has to be an integer, we round it to the nearest integer greater than or equal to \( K \). Note that, as a particular case of the equation above, a simpler expression can be derived for \( M_o = 1 \), given by

\[
R_1 \approx \log_2 \left[ 1 + \sum_{n=1}^{K} \frac{(-1)^{n-1}}{\delta_n(M-1)} \left( \frac{K}{n} \right)^{P} \right] \tag{15}
\]

Assuming retransmission of SINR values as discussed in the previous section, the BS sets \( M_o \rightarrow M_o - 1 \) if the algorithm fails to find a set of \( M_o \) \( \epsilon \)-orthogonal users. In a system with \( M = 2 \) antennas, the probability of not finding 2 \( \epsilon \)-orthogonal users is given by \( p = \left[ 1 - e^2 \right]^{K-1} \). Hence, the approximated rate in the described scenario is given by

\[
R \approx pR_1 + (1 - p)R_2 \tag{16}
\]

where \( R_1 \) and \( R_2 \) (\( R_{M_o} \) with \( M_o = 2 \)) are as described in equations (14) and (15) respectively. Figure 1 shows a comparison of analytical and simulated average sum rates in such system, with \( M = 2 \) antennas, \( K = 25 \) users and \( SNR = 10 \) dB. Each user has a simple codebook designed as described in previous section with \( B = 1 \) bit, different from user to user. Note that the jitter in the analytical curve is due to the rounding effect of \( K \).

V. TDMA vs. SDMA

In this section we provide asymptotical results showing that SDMA can provide higher rates than TDMA (\( M_o = 1 \)) in near-orthogonal MIMO systems as the number of users increases. We also illustrate by simulations the impact of average SNR values on the optimal choice of the system parameters \( M_o \) and \( \epsilon \). First, note that the number of available users at the \( i \)-th step can be bounded as \( K_i \geq e^{2(M-1)} \) as shown in [5]. The worst case scenario for SDMA is given by \( \epsilon = 1 \), since in that case the SINR bound becomes very pessimistic. For finite SNR, even when the proposed transmission scheme sets \( \epsilon = 1 \), we can easily obtain from equations (14) and (15) the following result

**Theorem 2.** Given an arbitrary \( \epsilon \), SDMA outperforms TDMA asymptotically with the number of users \( K \) by a factor

\[
\lim_{K \rightarrow \infty} \frac{R_{M_o}}{R_1} = M_o \tag{17}
\]

Hence SDMA exploits the multiplexing gain. In Figure 2 we show the evolution of the optimal value of \( \epsilon \) for varying SNR in a cell with large number of users, \( K = 1000 \), and \( M = 2 \) antennas. The simulated system performs switching from SDMA to TDMA when \( \epsilon \)-orthogonal sets are not found, as discussed in section III. A shift to the right in the position of the maximum implies that the number of \( \epsilon \)-orthogonal users found at the second step (\( K_2 \)) also increases, hence using 2 beams for transmission and thus exploiting the benefits of SDMA rather than TDMA. Therefore, Figure 2 shows that as the SNR decreases, a system based on near-orthogonal transmission tends to select SDMA over TDMA.

VI. NUMERICAL RESULTS

Figure 3 shows a performance comparison in terms of average sum rate versus orthogonality factor \( \epsilon \) for various transmission systems. The simulated system has \( M = 2 \) antennas and a simple codebook of \( B = 1 \) bits. The number of active users is \( K = 10 \) and the average \( SNR = 20 \) dB. Transmit matched filtering approaches with different channel state information at the transmitter (CSIT) are compared. The upper curve corresponds to optimal transmit matched filtering, with perfect CSIT and exhaustive search. Hence, its average rate is not a function of the orthogonality factor. The lower curve corresponds to the proposed system, whereas the third curve corresponds to the sum rate of a system with second step of full CSIT feedback. This means that, given a set of users selected for transmission, the BS requests full channel information from those users to perform transmit matched filtering. We can see that the bound becomes looser as \( \epsilon \) increases, since the bound on the SINR becomes more pessimistic. In the simulated system with \( K = 10 \) users, the maximum average sum rate occurs when the system sets orthogonality \( \epsilon = 0 \). This means that the system forces that at each time slot only one beam will be active, since there is zero probability of finding two quantized random channels perfectly orthogonal, assuming different quantization
codebooks for each user. Thus, in the simulated scenario with reduced number of users, TDMA (one active beam per time slot) is the optimal transmission technique while in systems with large number of users SDMA is optimal as shown in previous section.

VII. CONCLUSIONS

A scheme for near-orthogonal transmission in MIMO broadcast channels with limited feedback has been introduced. Based on simple scalar feedback and transmit matched filtering along quantized channels, joint beamforming and scheduling is performed. For the described system, a rate function approximation has been given, which has been shown to fit well the simulated average sum rate. This function is a powerful design tool for systems based on similar limited feedback scenarios, where a greedy algorithm reduces the amount of multiuser diversity at each selection step. In a system with $K$ active users and a given average SNR, fixing the number of antennas and codebook size, we have shown that the sum rate is a function of the number of active beams (considered less or equal than the number of antennas) and the orthogonality factor $\epsilon$. Hence, the benefits obtained from multiuser diversity and multiplexing gain can be conveniently balanced in a real scenario by appropriately choosing these parameters. In a realistic system, they can be set at every time slot for throughput maximization.

REFERENCES


