Unitary Precoders for CQI Reliability in Closed Loop MU-MIMO OFDM Systems

Hemanth Acharya¹, Sivakishore Reddy Yerrapareddy², Kiran Kumar Kuchi², Bhaskar Ramamurthi³
1. Qualcomm India Private Limited, Bangalore, India 560 066
2. Centre of Excellence in Wireless Technology, Chennai, India 600 113
3. Department of Electrical Engineering, IIT Madras, Chennai, India 600 036
hemantha@qualcomm.com, {kishore, kkuchi}@cewit.org.in, bhaskar@tenet.res.in

Abstract—In Multi-user MIMO (MU-MIMO), the Base Station (BS) schedules multiple users simultaneously on the same resources. Closed loop MU-MIMO uses explicit feedback from each of these users for scheduling and allocation of resources. In this scheme, an estimate of the SINR experienced is fed back by each user and is known as the CQI (Channel Quality Information). This CQI is used by the BS for link adaptation to the scheduled users. When Base Stations change their precoders in the next scheduling epoch, interference levels and therefore CQI estimates made previously may no longer be accurate, leading to reduction in system throughput. In this paper, we discuss the benefits of using unitary precoders to reduce this mismatch in CQI and thus improve system performance.

I. INTRODUCTION

Next-generation wireless networks such as Wimax (IEEE 802.16m) [1] and LTE [2] offer data rates in excess of 100 Mbps. MIMO (Multiple In Multiple Out) technology, which involves the use of multiple antennas at both transmitter and receiver, plays a key role in improving spectral efficiency to achieve such high data rates.

Single-user MIMO schemes such as spatial multiplexing [3], exploit additional degrees of freedom gained through multiple antennas, to increase data speeds. However, spatial correlation between the antennas reduce the gains obtained through spatial multiplexing, which requires full-rank channel conditions. Multi-user MIMO (MU-MIMO) is an extension of single-user MIMO spatial multiplexing, where the Base Station (BS) communicates with multiple users using the same resources. In MU-MIMO, the additional spatial degrees of freedom are shared between multiple users, and system capacity increases on account of users being scheduled to reuse the same time and frequency resources. Hence, unlike single-user schemes that rely on good channel conditions for separation of data-streams, an MU-MIMO system exploits the physical separation between users and scheduling to obtain the required degrees of freedom.

Substantial gains can be achieved in a MIMO system if channel state information (CSI) is available at the transmitter. Based on feedback, MU-MIMO can operate in two modes - Closed Loop MU-MIMO and Open Loop MU-MIMO. Closed Loop mode involves explicit feedback from all users while Open loop mode does not. In this paper, we focus on Closed Loop MU-MIMO in the IEEE 802.16m standard. In this mode, each user feeds back information comprising of a) PMI (Preferred Matrix Index) and b) CQI (Channel Quality Information), which are used in the subsequent scheduling epoch. PMI is used by the BS for precoder design. CQI is used for link adaptation where the BS varies the MCS allocated to a user to suit the channel conditions.

Based on the uplink rate allowed for CQI feedback, we can have different schemes such as full-CQI feedback, best-M-band feedback, or thresholded feedback [4]. In this paper, we assume full-CQI feedback with an error-free feedback link, where frequency selective PMI and CQI information are returned for every subband (band of contiguous subcarriers defined by IEEE 802.16m). CQI is an estimate of the post-processing SINR a user is likely to experience in the next scheduling epoch. In this paper, we define the scheduling epoch to be one frame duration, with new users being scheduled every frame. Using channel estimates obtained from pilots, a user can estimate co-channel interference (CCI) levels in the current frame. However, when scheduling is done in the next frame, neighboring BSs change their precoders based on PMI fed back by the newly scheduled users. This leads to a fluctuation in the CCI levels across frames, and accurate CQI modeling even for users with low channel Doppler becomes difficult. This interference fluctuation causes a mismatch between the SINR experienced in one frame and the CQI fed back in the previous frame. This mismatch can be seen as an error in the reported CQI, and the effects of such errors on throughput performance have been reported in [5].

The inability of a user to predict in which directions the neighboring BSs would beam-form in the next frame can be seen as the reason behind this mismatch, which motivates us to look for a solution to this problem through precoder design. In particular, we present the advantages of using unitary precoders to reduce the fluctuation in interference levels and thereby improve system performance. However unitary precoders are feasible only when the number of transmit antennas is equal to the number of receive antennas. For example, in systems with 4-antennas at BS and 2-antennas at the user equipment (UE), when 2 users are scheduled with a single stream per user, it is impossible to form a unitary precoding matrix, and performance degradation due to CQI...
mismatch could negate the improvement occurring from a larger number of transmit antennas. System level simulations have been carried out and results have been provided to show the improvement due to CQI stability from using unitary precoders.

This paper is organised as follows: Section II describes the downlink MU-MIMO system and downlink frame structure, Section III describes the algorithms proposed for closed loop MU-MIMO operation, Section IV gives results of simulation and Section V concludes the paper.

II. SYSTEM MODEL

Consider an MU-MIMO system with \( N_T \) BS antennas and \( N_R \) receive antennas per UE. The BS schedules a set of \( K \) users from a pool of \( M \) users per sector. Each user in the system experiences CCI from \( I \) strong interferers. The MU-MIMO system equation for the \( j^{th} \) user is given by:

\[
y_{jk} = H_{jk} W_k x_k + \sum_{i=1}^{I} \sqrt{\gamma_{ij}} G_{ijk} W_i k z_{ik} + n_{jk} \tag{1}
\]

where \( j = 1, ..., K \).

The IEEE 802.16m standard employs OFDM on the downlink, where the broadband signal to a user is transmitted using several narrow band subcarriers, with each subcarrier experiencing flat fading. Hence equation (1) is at a subcarrier level, with all vectors being those that are transmitted or received using subcarrier \( k \). For example, \( y_{jk} \in \mathbb{C}^{N_T \times 1} \), is the vector received on the \( k^{th} \) subcarrier by the \( j^{th} \) user.

As specified in the IEEE 802.16m downlink MIMO mode 4 [6], we maintain a single stream per user. Hence, \( x_k \in \mathbb{C}^{K \times 1} \), contains the \( K \) symbols transmitted to the \( K \) scheduled users. We assume that symbols transmitted to different users are uncorrelated, and the total transmit power \( P \) is equally distributed amongst all scheduled users, i.e. \( \mathbb{E} [x_k x_k^H] = \left( \frac{P}{K} \right) I_K \).

The summation term in (1) denotes the CCI from \( I \) neighboring BSs, where \( z_{ik} \in \mathbb{C}^{K \times 1} \), represents the interference from the \( i^{th} \) interferer, with \( \mathbb{E} [z_{ik} z_{ik}^H] = \left( \frac{P}{K} \right) I_K \). We assume the interfering vectors \( z_{ik} \) are uncorrelated with \( x_k \), i.e. \( \mathbb{E} [x_k z_{ik}^H] = 0 \), \( \forall i=1,...,I \).

Vector \( n_{jk} \in \mathbb{C}^{K \times 1} \) represents i.i.d circular symmetric additive complex Gaussian noise at the \( j^{th} \) receiver. Elements of \( n_{jk} \) are zero mean with variance \( \sigma^2 \), with \( \mathbb{E} [n_{jk} n_{jk}^H] = \sigma^2 I_K \).

\( H_{jk} \in \mathbb{C}^{N_R \times N_T} \) represents the channel matrix for user \( j \) at subcarrier \( k \). The entries of \( H_{jk} \) are i.i.d zero-mean complex Gaussian with unit variance, and antennas are assumed to be spaced far enough to render them uncorrelated. Each element of the \( N_R \times N_T \) matrix is generated using Jakes model [11], which introduces the required temporal correlation across frames. The channel between user \( j \) and interferer \( i \) on subcarrier \( k \) is represented by \( G_{ijk} \). The entries of \( G_{ijk} \) are also i.i.d zero-mean complex Gaussian with unit variance.

Stream-to-antenna mapping is done through precoder \( W_k \in \mathbb{C}^{N_T \times K} \). The precoder used by the \( i^{th} \) interferer is represented by \( W_{ik} \in \mathbb{C}^{N_T \times K} \).

A. Downlink Frame structure

We consider an OFDM system that uses a 10 MHz band with 1024 subcarriers(\( N_c \)). Continuous transmission of an OFDM signal can be seen as a block of \( N_c \) symbols being transmitted, one block after another. When we stack up these \( N_c \) symbols that are transmitted during successive time intervals in a column-wise fashion, we can populate a time-frequency grid, called the downlink OFDM frame. The \( (i,j) \) entry of this frame represents the modulation symbol carried by the \( i^{th} \) subcarrier in the \( j^{th} \) OFDM symbol duration. This frame can be divided in time and frequency as shown in figure 1.

The smallest time-frequency resource that can be allocated to a scheduled user is known as a Physical Resource Unit (PRU). A PRU consists of a set of 18 contiguous subcarriers allocated for a duration of 6 OFDM symbols. The OFDM frame is divided in frequency into a number of subbands, where a subband is defined as a group of 4 adjacent PRUs. In this work, we do not use any minibands, and all 48 PRUs are used to form 12 subbands. The time duration of the downlink frame is 5 ms and it is divided into 5 subframes, where the duration of each subframe is 1 ms (6 OFDM symbols).

B. Feedback structure

Feedback information is returned every frame. Feedback at a subcarrier level is not feasible since it would require a high-capacity feedback link. We use subband-level feedback, where each user returns PMI and CQI information for every subband. Thus PMI selection is made such that it represents the best choice of codebook vector for all subcarriers in the subband. The CQI estimate that is returned is also averaged over all subcarriers in a subband.

III. CLOSED LOOP MU-MIMO

In order to simplify the analysis, we describe a 2x2 system, with two BS antennas and two receive antennas per UE. Two users are scheduled by the BS for transmission, and single streams are maintained for both users. Perfect CSI is assumed at each UE. Closed-loop operation requires PMI.
evaluation, CQI modelling and multi-user Proportional Fair (PF) scheduling. Each of these are described individually in the following sections.

A. MMSE post-processing

Each terminal is assumed to use an MMSE filter for post-processing. Let \( b_j \) represent the 2x1 MMSE filter used by user \( j \) to filter the 2x1 received vector \( y_j \). The expression for the MMSE filter \( b_j \), is given by

\[
b_j = R^{-1}p
\]  

(2)

where,

- \( R \) : Autocorrelation matrix (2x2) of received vector \( y_j \).
- \( p \) : Cross-correlation vector (2x1) between the desired symbol and received vector \( y_j \).

B. PMI evaluation

With 2 transmit antennas and a single stream per user, the IEEE 802.16m standard’s codebook that is used for quantisation is C(2,1,3) [6]. This codebook has eight 2x1 unit norm vectors thus requiring 3 bits for PMI feedback. Different strategies could be adopted for choosing the best codebook vector to be fed back [7][9][8]. Since each user employs MMSE post-processing, the codebook vector which minimizes the MSE is chosen, and its codebook index, known as preferred matrix index, is fed back to the BS. The PMI returned for subband \( s \) is given by:

\[
PMI_s = \arg \min_l \left( \sum_k |MSE_{l,k}|^2 \right)
\]

(3)

where the summation is over all subcarriers in the \( s^{th} \) subband. Minimising the sum-MSE ensures that we find the best codebook vector for all subcarriers in a subband.

C. Unitary Precoders & CQI modeling

The CQI is an indication of the SINR a user is likely to experience, if scheduled in the next frame with the precoder of his choice (PMI returned). Assume user \( j \) finds \( c_a \in \mathbb{C}^{2x1} \) to be the codebook vector which minimizes the MSE for subband \( s \) per equation (3). In the next frame, the BS would form the precoder

\[
W_k = \begin{bmatrix} c_a & c_b \end{bmatrix}
\]

(4)

where user \( j \) assumes \( c_b \in \mathbb{C}^{2x1} \) to be the precoding vector returned by the interfering user. Expanding equation (1) in terms of \( c_{a,b} \),

\[
y_j = H_{jk}c_ax_1 + H_{jk}c_bx_2 + \sum_{i=1}^{L} \sqrt{L_{ij}} G_{ijk} W_{ik} z_{ik} + n_{jk}
\]

(5)

User \( j \) uses an MMSE filter \( b_j \),

\[
b_j^H y_j = b_j^H H_{jk} c_a x_1 + b_j^H H_{jk} c_b x_2 + \sum_{i=1}^{L} \sqrt{L_{ij}} b_j^H G_{ijk} W_{ik}^H z_{ik} + b_j^H n_{jk}
\]

(6)

Hence, CQI for subband \( s \), which is fed back along with user \( j \)'s choice of PMI for subband \( s (c_b) \), is given by:

\[
CQI_s = \frac{1}{N_s} \sum_{k=1}^{N_s} \frac{L_{ij}}{2} \left| b_j^H H_{jk} c_b \right|^2 + b_j^H \left( \frac{L_{ij}}{2} I_{cov} + \sigma^2 I \right) b_j
\]

(7)

where,

- \( N_s \) : Total number of subcarriers in subband \( s \).
- \( I_{cov} \) : \( \sum_{i=1}^{I} L_{ij} G_{ijk} W_{ik}^H W_{ik} G_{ijk}^H \), the Interference Covariance matrix.

There is an inherent latency of one frame in the feedback-scheduling process. PMI and CQI information evaluated by users in Frame \( n \), is used by the BS-scheduler in Frame \( n+1 \). Evaluation of equation (7) in Frame \( n \) becomes difficult, since

- User \( j \) will be unaware of the interfering co-channel user who will be scheduled in Frame \( n+1 \), and will not be able to predict \( c_b \).
- On the same resource \( I_{cov} \) now reduces to \( \sum_{i=1}^{I} L_{ij} G_{ijk} W_{ik}^H W_{ik} G_{ijk}^H \) (the Interference covariance matrix) is a function of \( W_{ik} \). Since the interfering BSs would also change their precoders in the subsequent frame, CCI levels will fluctuate, and any interference measurements made in Frame \( n \) become unreliable in Frame \( n+1 \).

However, when unitary precoders are used,

- \( c_b \) would be a codebook vector that is orthogonal to \( c_a \). Once the PMI for a particular subband is decided, user \( j \) knows that the interfering user in the next frame would be one who has requested for an orthogonal PMI. Exploiting this unitary structure of the precoder to be formed, \( c_b \) can be evaluated by user \( j \) for a given \( c_a \).
- \( I_{cov} \) now reduces to \( \sum_{i=1}^{I} L_{ij} G_{ijk} G_{ijk}^H \) since the interfering BSs are also using unitary precoders. The use of unitary precoders removes the dependency of \( I_{cov} \) on the precoders of the interfering BSs, thereby making estimates of CCI levels much more reliable across frames.

Thus, by maintaining a look-up table for orthogonal vectors to each codebook vector, equation (7) can be used to evaluate the CQI that must be returned for any PMI.

D. 2x2 versus 4x2

Spatial de-correlation is achieved by spacing antennas sufficiently far apart. Due to size constraints at the UE, we assume only 2 antennas at the UE, but extend the transmit (BS) antennas to 4. In order to be able to compare throughputs with the 2x2 case, we work within the same MU-MIMO-framework of scheduling 2 users with single stream.
per user. Increasing the number of transmit antennas provides a gain by virtue of better beamforming. However, in the context of unitary precoding, moving to a 4x2 system becomes problematic because:

- In the 4x2 system, the precoding matrix dimension is 4x2. Hence, it is not possible to form a unitary precoding matrix. This results in CCI fluctuation, poor CQI estimates and consequently, a loss in user throughputs.
- The codebook used in the 4x2 case is C(4,1,6) [6], which consists of 64 unit-norm 4x1 vectors. In the C(2,1,3) codebook, each vector is orthogonal to only one other codebook vector (2-dimensional space) and the look-up table mentioned earlier stores this mapping. However, in the C(4,1,6) case, some vectors are orthogonal to multiple vectors within the codebook. Assume codebook vector \( c_a \) is orthogonal to codebook vectors \( c_a + c_d \). If user \( j \) uses one possible orthogonal pair while evaluating equation (7) (eg. \( c_a, c_d \)), but the precoder is designed with another pair (eg. \( c_a, c_d \)), it again results in CQI mismatch.

These problems are found to limit the gains when moving away from an N×N system, where unitary precoders cannot be formed.

E. Two-user Proportional Fair (PF) scheduler

The single-user per data-region PF-scheduler [10] has been extended to handle the two-user case. We define a data-region to be a group of 4 adjacent PRU’s, and is taken as the smallest unit of allocation to any user by the scheduler. The two-user PF scheduler must schedule two users for every data-region in the frame such that the resultant precoder is unitary, and at the same time, PF-metric maximisation also takes place.

The algorithm for the two-user PF scheduler is given below:

1) Evaluate PF-metrics for all users
   Assume data-region \( i \) belongs to subband \( s \). PF-metric for user \( j \) : \( PF_j = \frac{R_j}{\bar{R}_j} \). \( R_j \) is evaluated as a function of CQI returned for subband \( s \).

2) User-pairing
   Go through PMI’s returned by all users for subband \( s \). Use the look-up table, pair users who have returned PMI’s of codebook vectors that are orthogonal to each other.

3) If orthogonal-user-pairs are found,
   - Add the PF-metrics of users of every pair obtained in step 2 to obtain a sum-PF-metric for each user pair.
   - Schedule the user pair with highest sum-PF-metric for data-region \( i \).
   - Let \( c_{a,b} \) be the codebook vectors requested by the user pair with highest sum-PF-metric. Then, precoder for data-region \( i \):
     \[
     W_i = \begin{bmatrix} c_a^H & c_b^H \end{bmatrix}^{-1}
     \]
     Columns of \( W_i \) are normalised to maintain the power constraint.

4) Average throughput update
   
   Average throughput of user \( j \), \( \bar{R}_j \) is updated according to:
   \[
   \bar{R}_j = \begin{cases} 
   (1-\alpha)\bar{R}_j + \alpha R_j & \text{if scheduled} \\
   (1-\alpha)\bar{R}_j & \text{if not scheduled}
   \end{cases}
   \]
   where \( \alpha \) is the PF-factor.

A few points to note about this algorithm are:

- Average throughput updating procedure in step 4 results in fairness and throughput maximization taking place across all subbands and subframes.
- Orthogonal user-pairing of step 2 ensures that the resulting precoder is unitary.
- The else condition in step 3 of the algorithm deals with the case when orthogonal-user-pairing is not possible. Simulations have been run with 10 users in each sector, and such a case is found to occur very rarely. However, in such a situation, we group users to minimize the condition number, so that the resultant ZF-precoder is almost unitary. This maintains accuracy of the CQI estimates and also reduces the power penalty incurred by zero-forcing.

IV. RESULTS

System level simulations were performed with a 19 cell-per-cluster model. Each cell is divided into 3 sectors with 10 active
subscribers per sector. The 10 MHz downlink OFDM frame transmitted per sector contains 12 subbands and 5 subframes as shown in Figure 1. Each carrier is modulated using QPSK, 16QAM or 64QAM depending on the CQI returned. Losses due to log-normal shadowing and pathloss are introduced for each user. Modified PED-B power delay profile is used and the required temporal correlation is achieved using Jakes model with a Doppler of 7 Hz. No antenna correlation is used and the channel to each user is spatially white.

Figure 2 shows the reduction in CQI mismatch upon using unitary precoders. A histogram of the difference between CQI returned by a user in frame \(n\) and post-processing SINR experienced by the user in frame \(n\) has been plotted at low Doppler. If the CQI fed back was much higher than the post-processing SINR the user subsequently experiences, the MCS allocated would be beyond what the channel can support, leading to frame errors. If the mismatch is large, even HARQ will not be able to ensure successful transmission for a limited number of re-transmissions. On the other hand, if the CQI fed back was much lower than the post-processing SINR the user subsequently experiences, the MCS allocated would undermine the capacity of the channel, thus resulting in a loss in throughput. Hence, for good throughput-performance, we would require a narrow histogram which is centered about 0 dB. As shown in Figure 2, the 2x2 (unitary) histogram is narrower than the 4x2 (non-unitary) histogram, due to higher CCI fluctuations and CQI mismatch in the 4x2 case.

Figure 3 compares the CDF’s of user throughputs obtained in the 2x2 and 4x2 cases. The CCI estimated by each user depends on midamble design. If the midamble supports accurate CCI estimates, \(I_{cov}\) can be evaluated as per equation [7]. If the midamble does not support an accurate estimate of \(I_{cov}\), a scaled Identity matrix can be used as the CCI estimate, i.e. \(I_{cov} = \sigma^2_{cov} I\), where \(\sigma^2_{cov}\) is the measured interference power. Throughput were evaluated for both scenarios and the CDF’s of user-throughputs obtained are plotted in Figure 3.

We adapt a target BLER of 10% without HARQ in our simulations. HARQ, if introduced, will reduce this to 1% or less with only 10% loss in capacity. In order to maintain a BLER less than 10%, a backoff of about 3 dB was applied on CQI returned in the 4x2 case. However, no backoff was required in the 2x2 case due to more accurate CQI estimates. This is reflected in throughput-performance, where throughputs for the ideal CCI 2x2 case are 27% (at 50% CDF) higher than throughputs for the ideal CCI 4x2 case. When a scaled Identity matrix is used, throughputs in both cases have reduced due to the inaccuracy introduced in CCI estimates. However, the 2x2 throughputs are still higher than the 4x2 throughputs.

V. CONCLUSION

In this paper, we studied the advantages of using unitary precoders to improve reliability of CQI estimates in a closed-loop MU-MIMO system. We compared the 2x2 case with unitary precoders, and 4x2 case where it was not possible to use unitary precoders. Results of simulation show that:

- There is higher CCI fluctuation in the 4x2 case when compared to the 2x2 case.
- Fluctuation in CCI levels lead to higher CQI mismatch which leads to drop in user-throughputs.

In conclusion, moving from a 2x2 to 4x2 system would give us gains at a link-level due to better beam-forming. However, when we look at it from a system-level, moving away from the 2x2 unitary-precoding system introduces fluctuations in interference levels, thereby limiting system performance.

ACKNOWLEDGMENT

The authors would like to thank Padmanabhan M.S. and Dhivagar B for their help on simulations that were carried out.

REFERENCES


