Multi-user MIMO Linear Precoding Schemes in OFDM Systems

Peng Liu[#1], Mu Qing Wu[#2] Chun Xiu Xu[#3], Feng Zheng[#4]
[#]School of Information and Telecommunication Engineering Beijing University of Post and Telecommunication, Beijing 100876, China
e-mail: 1liupeng.pub@gmail.com
e-mail: [2] wumuqing@bupt.edu.cn

Abstract—**In this paper, several multi-user multiple-input multiple-output (MU-MIMO) linear precoding schemes with both perfect and imperfect channel state information (CSI) for downlink transmission are proposed. Compared to nonlinear precoding schemes which need perfect CSI to achieve the channel capacity, we focus on systems which are restricted to use only linear precoding schemes due to complexity constraints and limited feedback. The limited feedback and some improvement technologies have also been introduced. The proposed MU-MIMO schemes extend the point-to-point single-user precoding to point-to-multipoint multi-user precoding.**

Keywords-MU-MIMO; multi-user interference; linear precoding; limited feedback

I. Introduction

Multiple-input multiple-output (MIMO) antenna systems can greatly improve the spectral efficiency in wireless communication systems. Information theory reveals that under certain conditions, there is a linear relationship between the channel capacity and the number of antennas of MIMO systems. Such systems have received a lot of attention in the context of emerging cellular systems, such as the 3GPP long term evolution (LTE). MU-MIMO refers to a system where a transmitter equipped with multiple antennas is communicating with several users possibly equipped with multiple antennas simultaneously, on the same physical resources.

Information theory reveals that if there is full CSI at the transmitter (CSIT) and at the receiver (CSIR), the optimum transmit scheme for the MU-MIMO broadcast channel involves a theoretical pre-interference cancellation technique known as dirty paper coding (DPC) [1]. However, DPC is hard to implement, thus simpler and sub-optimal transmission schemes have been proposed in this paper.

In MU-MIMO systems, CSIT allows for multi-user spatial multiplexing and thus increases the system throughput. It can be achieved by exploiting channel reciprocity in a time division duplex (TDD) system. But in a frequency division duplex (FDD) system, different carrier frequencies are used for uplink and downlink. Perfect CSIT is almost impossible to achieve. However, it is possible to obtain partial CSIT by means of a limited feedback channel. There are two different ways to feed back the necessary information to build the precoding matrix. The one is based on an extension of the limited feedback single-user MIMO (SU-MIMO) scheme to a MU system, as proposed in [2]. Users determine their preferred precoding vectors based on a finite codebook of possible precoders and feed back the codebook index with the quality of the chosen precoding vector. The base station (BS) just need to find the high quality users which have chosen different vectors of the same precoding matrix, and schedule them for transmission. The other is based on channel vector quantization (CVQ) using a finite channel codebook as proposed in [3]. Each user quantizes its channel based on the codebook and feeds back the corresponding index with an approximative signal to interference and noise ratio (SINR) value. Finally, the BS uses the available SINR values to schedule the users by maximizing the sum rate or any other criteria, and uses the quantized channel information to derivate the precoding matrix based on its precoding scheme.

In this paper, we describe the MU-MIMO system model firstly. Then We focus on several widespread concerned linear precoding schemes such as per-user unitary rate control (PU2RC), zero forcing (ZF), regularized channel inversion (also called MMSE), block diagonalization (BD) and signal to leakage ratio (SLR). Some useful improvements are introduced in details too. Finally, the advantages and disadvantages of the above mentioned precoding schemes are analyzed.

II. System Model

In this section, we describe a system model of the MU-MIMO downlink channel. The BS employs M transmit antennas and communicates with K users simultaneously. User k, ($k = 1,\cdots, K$), has N_k receive antennas. The channel model from the BS to the k-th user is represented by a $N_k \times M$ channel matrix $\mathbf{H}_k$.

Let $\mathbf{s}_k \in {}^{N_k \times 1}$ denote the k-th user transmit symbol vector. The user k employs a linear transmit precoding matrix $\mathbf{W}_k \in {}^{M \times N_k}$, which transforms the data vector $\mathbf{s}_k$ to the $M \times 1$ transmitted vector $\mathbf{W}_k \times \mathbf{s}_k$. The received signal vector of the k-th user is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{W}_k \mathbf{s}_k + \mathbf{H}_k \sum_{i \neq k} \mathbf{W}_i \mathbf{s}_i + \mathbf{n}_k \quad (1)$$

Where $\mathbf{n}_k = \left[n_{k,1} \cdots n_{k,N_k} \right]^T$ denotes the noise vector for the k-th user. Here $(\cdot)^T$ denotes the transpose of a matrix (or vector). The components $n_{k,i}$ of the noise vector $\mathbf{n}_k$ are i.i.d. with zero mean and variance σ^2 for $k = 1,\cdots,K$ and $i = 1,\cdots,N_k$. Note that both the desired signal $\mathbf{H}_k \mathbf{W}_k \mathbf{s}_k$ and

978-1-4244-5539-3/10/$26.00 ©2010 IEEE

the interference $\mathbf{H}_k \sum_{i \neq k} \mathbf{W}_i \mathbf{s}_i$ are received by the user k. Defining the network channel as:

$$\mathbf{H} = \left[\mathbf{H}_1^T \cdots \mathbf{H}_K^T \right]^T \tag{2}$$

The corresponding signals at all the users can be arranged as

$$\mathbf{y} = \mathbf{HWs} + \mathbf{n} \tag{3}$$

Where $\mathbf{y} = \left[\mathbf{y}_1^T \mathbf{y}_2^T \cdots \mathbf{y}_K^T \right]^T$, $\mathbf{W} = \left[\mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_K \right]$, $\mathbf{s} = \left[\mathbf{s}_1^T \mathbf{s}_2^T \cdots \mathbf{s}_K^T \right]^T$ and $\mathbf{n} = \left[\mathbf{n}_1^T \mathbf{n}_2^T \cdots \mathbf{n}_K^T \right]^T$. The purpose of the linear precoder is to design the precoding matrix $\mathbf{W}$ based on the channel knowledge, so that the performance of the MU-MIMO system can be improved.

III. MU-MIMO DOWNLINK LINEAR PRECODING

The detailed steps of several linear precoding schemes of MU-MIMO downlink channel will be introduced in this section. In order to describe the precoding methods directly, we will firstly assume that the transmitter can get the CSI synchronously and perfectly. Then some useful improvements will be introduced.

A. Per-User Unitary Rate Control

Per-User Unitary Rate Control (PU^2RC) precoding scheme is proposed by Samsung Corporation [2]. Its transmitter structure is shown in Fig. 1. This scheme is designed to meet the restriction of limited feedback.

PU^2RC is a typical precoding scheme of the first method of limited feedback. In this scheme, users need to feed back some indexes of their preferred precoding vectors based on a finite codebook of possible precoders with the channel quality indicator (CQI). This scheme can support both the space division multiplexing (SDM) and space division multiple access (SDMA). The detailed steps of PU^2RC scheme are as follows:

- Receiver channel estimation: According to the pilot signal, each user estimates the CSI.
- Receiver feedback: Each user feedback a preferred precoding matrix's index with the CQI.
- Group users: The BS groups users which declare the same index of the preferred precoding matrix.
- Select a group: The BS selects a group with the highest group priority.
- Select codeword: The BS selects codeword of the highest priority users in the selected group.
- AMC and precoding: The BS applies adaptive modulation and coding (AMC) and precoding schemes corresponding to the selected group.

PU^2RC scheme is low computational complexity. The BS can directly use the feedback information to select users and precode. The only requirement for the precoder is that it must be unitary.

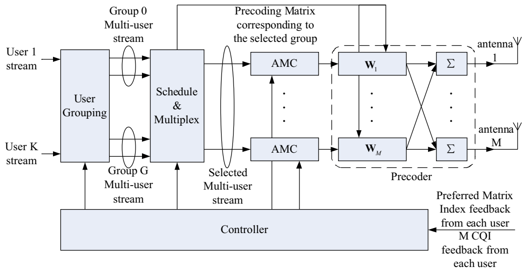

Figure 1. The transmitter structure of PU2RC.

The unitary constraint imposes benefits in some cases when it comes to CQI estimation but reduces the probability to find a MU-MIMO scheduling pair of UEs. A drawback of the unitary precoding is that the size of the codebook must be kept very small (1-2 matrices) to increase the probability to find at least two users that report preference for the same precoding matrix and simultaneously different columns to the matrix. But a too small codebook implies worse performance since the spatial mismatch is on average larger. Therefore, the performance of the scheme will increase with the number of active users but decrease with the size of the codebook.

B. Zero Forcing

Zero Forcing (ZF) precoding scheme which is a nonunitary scheme is proposed by Freescale Semiconductor Inc [4]. Under ideal conditions, each user needs single-antenna and experiences no multi-user interference. The system structure is shown in Fig. 2. The detailed steps of ZF scheme are as follows:

- Receiver channel estimation: According to the pilot signal, each user estimates its channel matrix $\mathbf{H}_k$.
- Receiver feedback: Each user feeds back its channel matrix $\mathbf{H}_k$ through the feedback channel.
- Precoding matrix derivation: To achieve zero multi-user interference, i.e., $\mathbf{H}_k \mathbf{W}_j = 0$ for $j \neq k$. The BS derivate the precoding matrix by the Moore-Penrose pseudoinverse of $\mathbf{H}$

  $$\mathbf{W} = \mathbf{H}^\dagger = \mathbf{H}^H \left(\mathbf{HH}^H \right)^{-1} \tag{4}$$

  Where $\mathbf{W}_k$ is obtained by normalizing the k-th column of $\mathbf{W}$.
- AMC and precoding: The BS applies AMC and precoding schemes.

Under ideal conditions, the received signal by the k-th user would be $\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k$. And the performance of ZF is very close to the performance of the optimum DPC. As K increases, the performance gap narrows because it is more likely to find a set of M users whose channels are increasingly mutually orthogonal

When the CSIT is imperfect, ZF cannot eliminate the multi-user interference, which causes the degradation of the SNR. And ZF scheme does not take into account the impact of noise, therefore, the system performance may be poor in the low SNR condition.

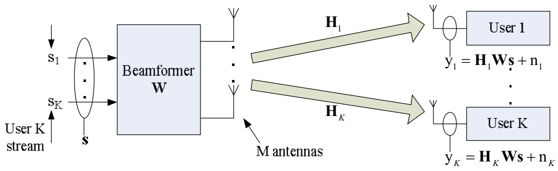

Figure 2. The system structure of ZF.

The feedback and power allocation improvement are as follows:

In practice, a feedback channel with limited bandwidth is available, it is impossible to get the perfect CSIT. The CVQ can be used to feedback a quantized version of the channel to provide partial CSIT.

In this scheme, each UE selects a quantization vector $\mathbf{c}_{I_k}$ from a codebook $C=\{\mathbf{c}_1,\cdots,\mathbf{c}_C\}$ of size $C=2^B$, so that the angle between the channel $\mathbf{H}_k$ and the codeword $\mathbf{c}_{I_k}$ is minimized. This is equivalent to writing

$$I_k=\underset{i=1,\cdots,C}{\arg\max}\left|\mathbf{c}_i^H\mathbf{H}_k\right| \tag{5}$$

Then the user feeds back the index I_k with the CQI and the value of $\|\mathbf{H}_k\|$, which requires B+CQI+$\|\mathbf{H}_k\|$ bits per user. The transmitter, which also knows the codebook, can reconstruct the channel by simply lookup table: $\hat{\mathbf{H}}_k=\mathbf{c}_{I_k}\|\mathbf{H}_k\|$. The codebook *C* is designed off-line and there are several well-known possibilities, such as Fourier codebook, Grassmannian codebook, random codebook and correlated random codebook.

We denote with $\hat{\mathbf{H}}=\left[\hat{\mathbf{H}}_1^T,\cdots,\hat{\mathbf{H}}_K^T\right]^T$ the concatenated quantized channel vectors of the selected users. Then, the ZF precoding matrix is given by

$$\mathbf{W}=\hat{\mathbf{H}}^H\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^H\right)^{-1}\mathrm{diag}(\mathbf{p})^{1/2} \tag{6}$$

Where $\mathbf{p}=(p_1,\cdots,p_K)^T$ is the vector of power normalization coefficients that impose the power constraint on the transmitted signal. Such as water-filling power allocation and equal power allocation

The above improvement of limited feedback using CVQ can be used in all of the following schemes, so we will omit it later. We will also omit the detailed steps of the following schemes, because they are similar with the above.

C. MMSE

To some extent, MMSE precoding scheme can be considered to be the improvement of ZF precoding scheme. For rank-deficient channels, the performance of ZF precoding can be improved by a regularization of the pseudo-inverse [5], which can be expressed as:

$$\mathbf{W}=\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H+\beta\mathbf{I}\right)^{-1} \tag{7}$$

Where β is a regularization factor. And $\mathbf{W}_k$ is obtained by normalizing the *k*-th column of $\mathbf{W}$.

A none-zero β value results in a measured amount of multi-user interference. In practice, the regularization factor is commonly chosen as $\beta=M\sigma^2/P$ so that it approximately maximizes the SINR at each user, and leads to linear capacity growth with *M*. The performance of MMSE is certainly better at low SNR and converges to that of ZF precoding at high SNR. However, MMSE does not provide orthogonal channels so that power allocation techniques cannot be performed in a straightforward manner.

D. Block Diagonalization

Block diagonalization (BD) is a generalization of channel inversion techniques where multiple antennas are at each user [6]. Under the BD scheme, the system can be equivalently regarded as a single-user MIMO environment. The system structure is shown in Fig. 3. We define $\tilde{\mathbf{H}}_k$ as:

$$\tilde{\mathbf{H}}_k=\left[\mathbf{H}_1^T\cdots\mathbf{H}_{k-1}^T\mathbf{H}_{k+1}^T\cdots\mathbf{H}_K^T\right]^T \tag{8}$$

Let the singular value decomposition (SVD) of $\tilde{\mathbf{H}}_k$ be:

$$\tilde{\mathbf{H}}_k=\tilde{\mathbf{U}}_k\tilde{\mathbf{D}}_k\left[\tilde{\mathbf{V}}_k^{(1)}\tilde{\mathbf{V}}_k^{(0)}\right]^H \tag{9}$$

Where $\tilde{\mathbf{U}}_k$ and $\tilde{\mathbf{D}}_k$ are the left singular vector matrix and the matrix of singular values of $\tilde{\mathbf{H}}_k$, respectively, and $\tilde{\mathbf{V}}_k^{(1)}$ and $\tilde{\mathbf{V}}_k^{(0)}$ denote the right singular matrices each corresponding to non-zero singular values and zero singular values. Thus, the last $\left(M-\mathrm{rank}\left(\tilde{\mathbf{H}}_k\right)\right)$ right singular vectors $\tilde{\mathbf{V}}_k^{(0)}$ forms an orthogonal basis for the null space of $\tilde{\mathbf{H}}_k$. Any precoder $\mathbf{W}_k$ which is a linear combination of the columns of $\tilde{\mathbf{V}}_k^{(0)}$ will lie in the null space of $\tilde{\mathbf{H}}_k$.

When using the BD scheme under ideal conditions, the system can be equivalently regarded as a single-user MIMO environment so that each user would experience no multi-user interference. But the computational complexity of BD scheme is slightly higher. Because the users need to know the equivalent channel to achieve detection, the BS needs to insert specific pilot frequently and it will reduce the effective information transmission rate in practice.

Feedback and power allocation improvement: We can also use the CVQ scheme to improve the BD scheme feedback. And we can enhance the throughput by performing water-filling as follows:

Let L_k represent $\mathrm{rank}\left(\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}\right)$. Define the SVD

$$\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}=\mathbf{U}_k\begin{bmatrix}\mathbf{D}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\left[\mathbf{V}_k^{(1)}\quad\mathbf{V}_k^{(0)}\right]^H \tag{10}$$

Where $\mathbf{D}_k$ is $L_k\times L_k$, and $\mathbf{V}_k^{(1)}$ represents the first L_k singular vectors. Denoting $\mathbf{D}$ is a diagonal matrix whose elements are $\mathbf{D}_k$ for $k=1,\cdots,K$. And $\mathbf{P}$ is a diagonal matrix whose elements λ_i scale the power transmitted into each of the columns of the precoding matrix. Then we can use water filling on the diagonal elements of $\mathbf{D}$ to determine the optimal power loading matrix $\mathbf{P}$ under power constraint *P*. The precoding matrix is

$$\mathbf{W}=\left[\tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(1)}\quad\tilde{\mathbf{V}}_2^{(0)}\mathbf{V}_2^{(1)}\quad\cdots\quad\tilde{\mathbf{V}}_K^{(0)}\mathbf{V}_K^{(1)}\right]\mathbf{P}^{1/2} \tag{11}$$

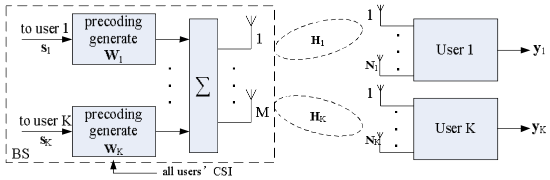

Figure 3. The system structure of BD

E. Signal to Leakage Ratio

Compared with the above schemes, signal to leakage ratio (SLR) scheme is an alternative approach, based on maximizing the signal to leakage ratio for designing transmit beamforming vectors in a multi-user system without eliminating the multi-user interference [7]. The system structure is shown in Fig. 4.

The SLR expression can be written as:

$$\text{SLR}=\frac{\mathbf{W}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_k}{\mathbf{W}_k^H\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\mathbf{W}_k} \quad (12)$$

Then the precoding matrices are:

$$\mathbf{W}_k^o \alpha \max \text{gen.eigenvector}\left(\mathbf{H}_k^H\mathbf{H}_k\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\right) \quad (13)$$

If $\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k$ is invertible, then the generalized eigenvalue problem reduces to

$$\mathbf{W}_k^o \alpha \max \text{eigenvector}\left(\left(\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\right)^{-1}\mathbf{H}_k^H\mathbf{H}_k\right) \quad (14)$$

This scheme can maximize the SLR at each user and it does not impose a restriction on the system configuration in terms of the number of antennas. But it is different from other schemes which can serve multiple-streams for each user simultaneously, the SLR scheme can only serve single-stream for each user at the same time without employing some improvements such as orthogonal space-time coding.

Feedback and performance Improvement: We can also use the CVQ scheme to improve the SLR scheme feedback. We can further take into account the influence of noise to improve the performance [8]. By doing so, the signal to leakage and noise ratio (SLNR) scheme outperforms ZF even when the dimension requirement for ZF scheme is satisfied. The precoding matrices are:

$$\mathbf{W}_k^o \alpha \max \text{eigenvector}\left(\left(N_k\sigma_k^2\mathbf{I}+\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\right)^{-1}\mathbf{H}_k^H\mathbf{H}_k\right) \quad (15)$$

Where σ_k^2 is the k-th user's noise variance per receive antenna.

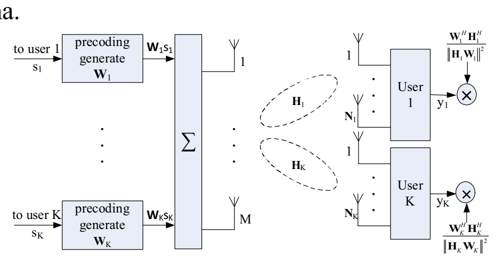

Figure 4. The system structure of SLR.

IV. CONCLUSIONS

In this paper, five different MU-MIMO linear precoding schemes are analyzed. We have introduced some useful improvement too, such as CVQ and power allocation. Generally, the PU^{2}RC scheme is low computational complexity and support both SDM and SDMA, but the unitary constraint requires us to balance the users matching and system performance. The ZF and MMSE schemes have been widely concerned for their high performance and low feedback rate. The BD and SLR schemes are high performance but high computational complexity too. And the SLR scheme can only serve single-stream for each user simultaneously, although it does not impose a restriction on the system configuration in terms of the number of antennas.

There are two important criteria need to be considered when using and designing MU-MIMO systems. Spatial separation of users has a very strong impact on the performance of linear precoding schemes. In particular, the performance of the ZF precoder drops significantly when the users are close together. Therefore it is necessary to design proper scheduling algorithms that select users with different spatial signatures. The performance of limited feedback MU-MIMO schemes crucially depends on the codebook, and the codebook design for MU-MIMO systems which can represent the channel more appropriately remains a hot topic.

ACKNOWLEDGMENT

This work was supported in part by ZTE Research and Development Foundation (No.2010110031000698). And the authors will also thank the anonymous reviewers.

REFERENCES

[1] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel", IEEE Trans. Inf. Theory, vol.49, no.7, pp.1691–1706, Jul. 2003.

[2] Samsung, "Downlink MIMO for EUTRA", Feb. 2006, 3GPP TSG-RAN WG1 #44 R1-060335.

[3] Philips, "Comparison between MU-MIMO codebook-based channel reporting techniques for LTE downlink", Oct. 2006, 3GPP TSG-RAN WG1 #46 R1-062483.

[4] Freescale Semiconductor, "Details of Zero-forcing MU-MIMO for DL EUTRA", Mar. 2007, 3GPP TSG-RAN WG1 #48bis R1-071510.

[5] C.B. Peel, B.M. Hochwald, and A.L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication-part I: channel inversion and regularization", IEEE Trans. Commun., vol.53, no.1, pp.195–202, Jan. 2005.

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

[7] A. Tarighat, M. Sadek, and A.H. Sayed, "A multi user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios", in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.3, Philadelphia, PA, pp.1129–1132, Mar. 2005.

[8] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," IEEE Trans. Wireless Commun., vol.6, no.5, pp.1711-1721, May. 2007.