Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 947869, 13 pages
doi:10.1155/2008/947869

Research Article

A Limited Feedback SDMA for Downlink of Multiuser MIMO Communication System

Yongming Huang,[1] Luxi Yang,[1] and Ju Liu[2]

[1] *School of Information Science and Engineering, Southeast University, Nanjing 210096, China*
[2] *School of Information Science and Engineering, Shandong University, Jinan 250100, China*

Correspondence should be addressed to Yongming Huang, huangym@seu.edu.cn

Received 14 February 2008; Accepted 13 August 2008

Recommended by Christoph F. Mecklenbräuker

This paper proposes a limited feedback SDMA scheme of combining opportunistic scheduling and codebook-based multiuser precoding. A new systematic construction for SDMA codebook, which comprises a set of precoders for multiple simultaneously active users, is first presented. Different from conventional Grassmannian codebook, the proposed codebook is designed in terms of array processing and has a cluster-based structure, with each cluster generated using a perturbation method. In order to tackle the intractable interuser interference issue inhered in limited feedback SDMA, this paper further proposes two novel opportunistic scheduling algorithms, which are able to fully exploit the cluster structure of the proposed codebook. The first proposed algorithm schedules the simultaneous users and their preferred precoders in a successive way, and is implemented in a Markov-like fashion. The second proposed algorithm is capable of rapidly finding a group of channel-matching users along with their preferred precoders. Simulation results demonstrate that in sparse networks, the proposed SDMA exhibits a better throughput performance than the conventional limited feedback SDMA does, while both having a comparable feedback overhead.

1. INTRODUCTION

Space division multiple access (SDMA) is capable of considerably improving the throughput of multiple antenna broadcast channels, in comparison with time division multiple access (TDMA). Thus, SDMA has been adopted by IEEE 802.20 and other standard bodies. Through multiuser precoding, SDMA enables base station to simultaneously communicate with multiple users using the same time-frequency resource. The optimal performance of SDMA can be achieved by combining dirty paper coding (DPC) [1] with appropriate user scheduling. However, DPC is infeasible since it is extremely computationally intensive, even some simplified versions of DPC, such as Tomlinson-Harashima precoding [2–4], are still very difficult for implementation due to high computational complexity. In contrast, several low-complexity multiuser precoding techniques have been developed, including zero-forcing precoding [5], block diagonalized precoding [6], MMSE precoding [7], the generalized eigenvalue-based solutions [8] and the iterative algorithms [9], and so forth, which can achieve a large portion of DPC capacity. To implement these schemes, perfect channel state information (CSI) of each user is acquired at the base station. In FDD systems, the downlink CSI should be fed back to the base station, which is infeasible in practice due to the limited bandwidth. In TDD systems, CSI can be obtained by exploiting the reciprocity between the downlink and uplink channels. However, for some reasons such as different RF circuits at the base station and user terminal, the obtained CSI may suffer from severe estimation error. Thus, it is important to improve the robustness of the existing multiuser precoding techniques to CSI estimation error [10], which, however, will dramatically increase the complexity. Recently, a type of SDMA based on quantized CSI feedback has been extensively studied [11–13]. Most of these works concentrate on the MISO-SDMA, their extensions for the MIMO-SDMA commonly require much more feedback bits to guarantee a reasonable performance. Alternatively, a type of opportunistic SDMA using a random unitary matrix as the multiuser precoder is proposed in [14], which can asymptotically obtain the optimal scaling law of sum capacity when the number of users is sufficiently large. However, in sparse networks where the number of users is small, the performance of this scheme severely degrades due to excessive mutual interferences between simultaneously active users. To conquer this problem, a modified opportunistic SDMA is proposed in [15], which requires the scheduled users to feed back full CSI using an extra feedback round. Although this scheme is able to deal with the interference problem in sparse networks, the feedback overhead is dramatically increased as well. Comparatively, the MIMO-SDMA scheme adopted by IEEE 802.20 standard [16] uses a set of predefined precoding matrices, that is, an SDMA codebook, to guarantee a reasonable performance in sparse networks. This scheme needs a small amount of feedback information and has a simple user scheduling algorithm. However, this codebook-based SDMA scheme is oversimplified since its codebook is essentially packed with only two Grassmannian subspaces. Therefore, the performance improvement from codebook employment is still limited.

In this paper, we first present a systematic design for a new SDMA codebook. Unlike the conventional precoder codebook used for point-to-point communication systems [16–19], the SDMA codebook design should take into account more performance metrics such as the facilitation of user scheduler and the degree of interuser interference suppression [20, 21]. Aiming at an SDMA codebook that is more general than the one adopted in IEEE 802.20, this paper proposes a systematic construction for SDMA codebook from the viewpoint of array processing. In terms of subspace packing [18], the proposed SDMA codebook can be packed with arbitrary number of subspaces. Instead of using the conventional criterion of maximizing the minimum distance, our employed subspace packing is designed with a uniform separation of beam direction, that is more suitable for multiuser precoding in terms of interuser interference suppression. In addition, each subspace included in the SDMA codebook experiences a number of unitary perturbations [22], with each perturbation generating a distinct precoding matrix. The set of the resultant precoding matrices from each subspace forms a cluster, and the collection of all the clusters forms the codebook. The perturbation diversity associated with each subspace is capable of compensating for the performance loss caused by a linear receiver [22]. Based on the proposed SDMA codebook, we further propose two opportunistic scheduling algorithms, both of which acquire only limited feedback information from users. The first algorithm is called Markov opportunistic scheduling (MOS), which utilizes the feedback information from both the current and previous scheduling intervals, and is implemented in a Markov-like fashion such that the interuser interference is fully suppressed. In this proposed algorithm, the scheduled users and the preferred precoders will be selected successively, and the feedback information in each scheduling can be divided into two parts, with one part used in the current scheduling and the other part used in the next scheduling. The second algorithm is called quick matching opportunistic scheduling (QMOS), which is able to rapidly find a group of best-matching users along with their preferred precoders. The channels of the matching users have a potential of good interference suppression. Thus, a simultaneous transmission to the matching users can fully exploit the spatial degree of freedom.

The rest of this paper is organized as follows. Section 2 briefly describes the proposed system model. Section 3 presents the construction for a new SDMA codebook. Section 4 proposes two novel opportunistic scheduling algorithms. Simulation results are given in Section 5 and conclusions are drawn in Section 6.

2. SYSTEM MODEL

We consider the downlink of a multiuser MIMO communication system. The base station is equipped with n_T transmit antennas and each user terminal is equipped with n_R receive antennas. It is assumed that $n_T \geq n_R$ and there are U users being served by the base station. The base station will schedule K out of U users together and simultaneously communicate with them, that is, in an SDMA mode. The signal intended for the kth user is precoded with $\mathbf{W}_k \in \mathcal{C}^{n_T \times M}$, $M \leq \min(n_T, n_R)$, thus the base station will simultaneously transmit M independent data streams to this user. The overall transmit signal at the base station can be expressed as

$$\mathbf{S} = \sum_{k=1}^{K} \mathbf{W}_k \mathbf{s}_k, \tag{1}$$

where $\mathbf{s}_k$ denotes the transmit signal vector for the kth user, and we assume that $E[\mathbf{s}_k \mathbf{s}_k^H] = E_s \mathbf{I}_M$. The received signal at the kth user can be written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{W}_k \mathbf{s}_k + \mathbf{H}_k \sum_{i \neq k, i=1}^{K} \mathbf{W}_i \mathbf{s}_i + \mathbf{n}_k, \tag{2}$$

where $\mathbf{H}_k$ denotes an $n_R \times n_T$ flat channel matrix whose entry $h_{nm,k}$ represents the channel response from the kth scheduled user's transmit antenna m to the receive antenna n, and $\mathbf{n}_k$ denotes the noise vector whose entry is the complex white noise with zero mean and N_0 variance.

In order to reduce the amount of feedback information, we predefine an SDMA codebook which includes all the precoding matrices $\{\mathbf{W}_k\}$. Once the base station gathers the limited feedback information from all the users, the scheduler will rapidly select multiple simultaneously active users together with their preferred precoders. Different from the precoder selection in point-to-point communication system, the opportunistic SDMA scheduler should tackle the extra issue of interuser interference suppression.

3. SDMA CODEBOOK

3.1. IEEE 802.20 SDMA codebook

An SDMA codebook is presented in IEEE 802.20 standard [16], which is defined as a set of precoders for individual active users in SDMA. This codebook consists of two clusters, each defined as a set of precoders which column span a same subspace. This section provides a quick review of this SDMA codebook construction and shows some of its advantages.

The IEEE 802.20 SDMA codebook focuses on the configuration of 4 transmit antennas at the base station and 2 receive antennas at the terminal, thus the codewords in the SDMA codebook belong to $\mathcal{C}^{4\times 2}$. As stated in [16], the subspaces represented by two clusters are constructed from a preset DFT-based matrix $\mathbf{B}$, which is given as

$$\mathbf{B} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ e^{j\pi/4} & e^{j3\pi/4} & e^{j5\pi/4} & e^{j7\pi/4} \\ e^{j\pi/2} & e^{j3\pi/2} & e^{j5\pi/2} & e^{j7\pi/2} \\ e^{j3\pi/4} & e^{j9\pi/4} & e^{j15\pi/4} & e^{j21\pi/4} \end{bmatrix}. \tag{3}$$

Next, we will introduce the detailed construction of two clusters, each including 14 precoders.

(i) Two clusters of IEEE 802.20 SDMA codebook

We denote the first cluster of IEEE 802.20 SDMA codebook as $\mathcal{G}_1$, and denote the precoders included in $\mathcal{G}_1$ as $\{\mathbf{F}_i,\ i = 1,\dots,14\}$. The ith precoder is generated as $\mathbf{F}_i = \mathbf{B}(:,1:2)\mathbf{\Lambda}_i\mathbf{D}$, where $\mathbf{B}(:,1:2)$ denotes the matrix formed by the first and the second columns of $\mathbf{B}$, $\mathbf{\Lambda}_i = \operatorname{diag}(e^{j\phi_{i1}}, e^{j\phi_{i2}})$, with ϕ_{im} being a random variable uniformly distributed in $[0,2\pi]$; and $\mathbf{D}$ is the normalized 2×2 DFT matrix, that is, $\mathbf{D} = \{D_{mn}\}$, with $D_{mn} = (1/\sqrt{2})e^{j2\pi(m-1)(n-1)/2}$, $m,n = 1,2$. Actually, the matrix $\mathbf{\Lambda}_i\mathbf{D}$ can be viewed as a random unitary matrix distributed on $\mathcal{U}(2,2)$ [23].

Similarly, the second cluster, denoted as $\mathcal{G}_2$, has its precoders $\mathbf{F}_i$, $i = 15,\dots,28$. The ith precoder is generated as $\mathbf{F}_i = \mathbf{B}(:,3:4)\mathbf{\Lambda}_i\mathbf{D}$, where $\mathbf{B}(:,3:4)$ denotes the matrix formed by the third and fourth columns of $\mathbf{B}$, and $\mathbf{\Lambda}_i$, $\mathbf{D}$ are the same as those in the first cluster $\mathcal{G}_1$.

(ii) Advantages of the two-cluster construction

From the Grassmannian subspace definition [18, 24], the subspace spanned by $\mathbf{B}(:,1:2)$ is equivalent to the one spanned by $\mathbf{B}(:,1:2)\mathbf{A}$, $\forall \mathbf{A} \in \mathcal{U}(2,2)$. This is because the distance between these two subspaces always equals zero, no matter what type of distance definition is used. Therefore, the different precoders in the same cluster span the same subspace, which implies that the SDMA codebook is packed with two subspaces spanned by $\mathbf{B}(:,1:2)$ and $\mathbf{B}(:,3:4)$. We can also find that any two precoders from different clusters are orthogonal to each other, that is,

$$\mathbf{F}_i^H\mathbf{F}_j = \mathbf{F}_j^H\mathbf{F}_i = \mathbf{0}, \quad \forall (i = 1,\dots,14,\ j = 15,\dots,28). \tag{4}$$

Assuming that each individual user in SDMA supports $\min(n_T, n_R)$ multiplexed substreams, the 4-2 antenna configuration implies that the base station can schedule up to two users on the same time-frequency resource. To suppress the interuser interference, it is required that the two selected precoders be from different clusters, meaning that the resultant multiuser precoder, that is, the collection of the precoders for all the active users, has to be a unitary matrix.

For a given precoder $\mathbf{F}_i$, we define its unitary perturbation as $\mathbf{T}_i = \mathbf{F}_i\mathbf{A}_i$, where $\mathbf{A}_i \in \mathcal{U}(2,2)$ is called the perturbation matrix. Accordingly, the construction of the clusters $\mathcal{G}_1$ and $\mathcal{G}_2$ can be viewed as a finite number of unitary perturbations to $\mathbf{B}(:,1:2)$ and $\mathbf{B}(:,3:4)$, using the perturbation matrices $\mathbf{A}_i = \mathbf{\Lambda}_i\mathbf{D}$. It is found that a unitary perturbation will not affect the original precoder from the subspace perspective, since the subspace spanned by $\mathbf{T}_i$ is identical to that spanned by $\mathbf{F}_i$. However, a different perturbation may have a distinct impact on the sum throughput achieved by the SDMA system, which will be described in detail in Section 4. This property reveals the significance of including different perturbations into the codebook construction. In the SDMA implementation, the scheduled user k will receive not only its own signal but also the signal from the other scheduled user (denoted as the interfering user). As shown in Section 4, when $\mathbf{F}_i$ is used as the precoder of the interfering user, any unitary perturbation to $\mathbf{F}_i$ will not change the interference imposed on the user k, which means the use of any precoder in the same cluster/subspace by the interfering user will impose the same interference on the user k. Since the codebook only consists of two clusters/subspaces, the user can easily estimate the possible interference from the other scheduled users. As a result, the scheduling algorithm can easily take into account the suppression of the interuser interference.

3.2. Proposed SDMA codebook

Although the SDMA codebook adopted in IEEE 802.20 is able to simplify the design of the scheduling algorithm and provides a reasonable performance even in sparse networks, the following issues have not been addressed.

(i) The IEEE 802.20 SDMA codebook actually restricts the resultant multiuser precoder into a unitary matrix, which is optimal in the case of a large number of users [14]. However, it has been shown that in sparse networks, especially when the number of users is equal to that of the supported simultaneously active users, the optimal linear multiuser precoder designed with full CSI is not a unitary matrix in most cases [5–9]. How can we design the SDMA codebook while taking into account both the above two scenarios?

(ii) The SDMA codebook is packed with two subspaces/clusters, which leads to a low density of subspace packing. Since the design of the SDMA codebook should consider both the optimization of the interference-free performance of the individual users and the suppression of interuser interferences, the criterion of maximizing the minimum distance [18] is not suitable to increase the density of subspace packing. We need to find a new method to extend the SDMA codebook for any arbitrary number of subspaces/clusters.

(iii) The feedback overhead mainly depends on the size of the SDMA codebook, and is as such determined by the density of subspace packing as well as the cluster size. Is it possible to find a good tradeoff between the number of packed subspaces and the cluster size such that the SDMA performance will be improved with a feedback overhead comparable to that required by IEEE 802.20 SDMA codebook?

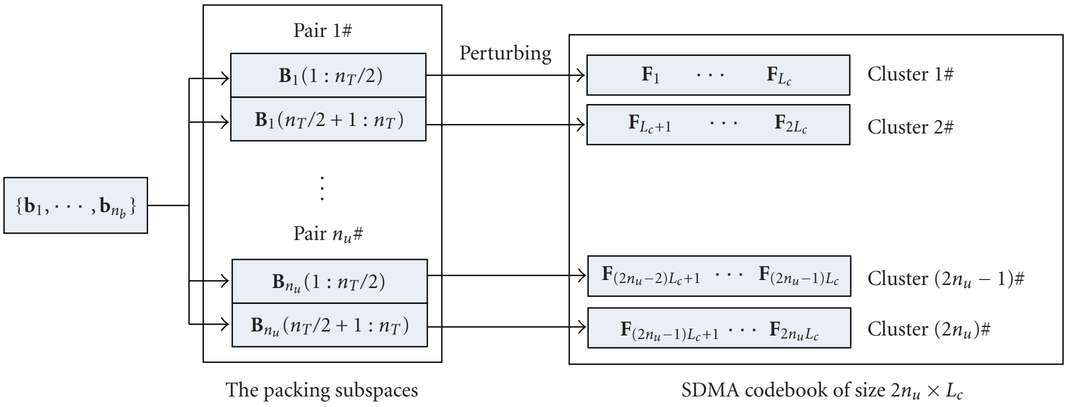

Pair 1#
Perturbing
$\mathbf{B}_1(1 : n_T/2)$
$\mathbf{B}_1(n_T/2+1 : n_T)$
$\{\mathbf{b}_1, \cdots, \mathbf{b}_{n_b}\}$
Pair n_u#
$\mathbf{B}_{n_u}(1 : n_T/2)$
$\mathbf{B}_{n_u}(n_T/2+1 : n_T)$
$\mathbf{F}_1 \cdots \mathbf{F}_{L_c}$
Cluster 1#
$\mathbf{F}_{L_c+1} \cdots \mathbf{F}_{2L_c}$
Cluster 2#
$\mathbf{F}_{(2n_u-2)L_c+1} \cdots \mathbf{F}_{(2n_u-1)L_c}$
Cluster $(2n_u - 1)$#
$\mathbf{F}_{(2n_u-1)L_c+1} \cdots \mathbf{F}_{2n_uL_c}$
Cluster $(2n_u)$#
The packing subspaces
SDMA codebook of size $2n_u \times L_c$

FIGURE 1: The SDMA codebook construction.

With an aim to address the above questions, we now design a new SDMA codebook with precoder in size of $n_T \times n_T/2$, for a multiuser network in which each terminal is equipped with no more than $n_T/2$ antennas. Different from the conventional Grassmannian subspace packing that uses the criterion of maximizing the minimum distance, we would like to construct a general number of packed subspaces from the beam direction perspective. In our scheme, each column of a precoder is regarded as a beam, and the codebook construction is completed with two phases.

In the first phase, we design a set of beams and then group them into several subsets. The beams in each subset constitute a precoding matrix. Let each beam be given by

$$\mathbf{b}_l = \left(\frac{1}{\sqrt{n_T}}\right)\left[1 \quad e^{j\theta_l} \quad \cdots \quad e^{j(n_T-1)\theta_l}\right]^T, \tag{5}$$

where θ_l is the phase of the beam, which indicates the beam direction. Obviously, this type of beam is in period of 2π with respect to θ_l. By equally separating the phase value between 0 and 2π, that is, $\theta_l = 2\pi(l-1)/n_b$, $l = 1, \dots, n_b$, a general number of n_b beams are generated and written as

$$\mathbf{b}_l = \frac{1}{\sqrt{n_T}}\left[1 \quad e^{j2\pi(l-1)/n_b} \quad \cdots \quad e^{j2\pi(n_T-1)(l-1)/n_b}\right]^T, \qquad l = 1, \dots, n_b. \tag{6}$$

From the viewpoint of array processing, the resultant n_b beams convey uniform power at intervals of $\pi/n_b/d_T$ in angular domain [25, 26], where d_T denotes the spacing between two neighboring transmit antennas in wavelength. Assuming that the selected number n_b satisfies $n_b = n_U n_T$, where n_U is an integer number, the following unitary matrices can be constructed by subset grouping, and are given by

$$\mathbf{B}_s = \left[\mathbf{b}_s, \mathbf{b}_{s+n_u}, \dots, \mathbf{b}_{s+(n_T-1)n_u}\right], \quad s = 1, \dots, n_u. \tag{7}$$

In the second phase of codebook construction, the above unitary matrices are used to generate multiple clusters. Each $\mathbf{B}_s$ is first partitioned into a pair of submatrices denoted as $\mathbf{B}_s(1 : n_T/2)$, $\mathbf{B}_s(n_T/2+1 : n_T)$. It is seen that the submatrices $\{\mathbf{B}_s(1 : n_T/2), \mathbf{B}_s(n_T/2 + 1 : n_T)\}$ actually follow a DFT-based structure. However, it should be mentioned that these submatrices are different from those generated by the DFT-based subspace packing, where the criterion of maximizing the minimum distance is used. The key difference lies in that the subspaces spanned by each pair of submatrices are orthogonal to each other, due to the fact that each pair of submatrices is a partition of a unitary matrix. Actually, this orthogonality characteristic provides a potential of achieving the asymptotical optimal SDMA performance [14]. To construct a codebook comprising multiple clusters, each submatrix is used to generate a particular cluster, by right multiplying a finite number of unitary perturbation matrices denoted as $\{\mathbf{\Lambda}_i\mathbf{D}, i = 1, \dots, L_c\}$ (an alternative type of perturbation matrices is available too, the interested readers please refer to [22]), where L_c denotes the cluster size and can be chosen as any integer number.

The two-phase codebook construction process is illustrated in Figure 1, it is seen that the constructed SDMA codebook consists of $N_c = 2n_U$ clusters/subspaces, with each cluster comprising L_c precoders. Essentially, the IEEE 802.20 SDMA codebook is a special case of the proposed SDMA codebook, with the configuration of $n_U = 1$, $n_T = 4$. It is worth mentioning that the proposed codebook would not only extend the number of clusters, but it would also extend the pattern of the resultant multiuser precoder. Since the clusters from different pairs are commonly nonorthogonal, the resultant multiuser precoder is not always a unitary matrix now.

The results in [14] have shown that the use of an arbitrary unitary matrix as the multiuser precoder is capable of achieving the optimal capacity scaling in the case of very large number of users. In our proposed SDMA codebook, if the active users are scheduled on a pair of orthogonal clusters/subspaces, the resultant multiuser precoder has to be a unitary matrix, thus the asymptotical optimal performance would be preserved. In sparse networks, the optimal multiuser precoder is commonly not a unitary matrix, thus it is highly possible that the case of scheduling users on two nonorthogonal clusters brings lower interference than that on a pair of orthogonal clusters. In this sense, the proposed SDMA codebook provides a better performance than the IEEE 802.20 codebook does in terms of interference suppression. This advantage can also be confirmed from the viewpoint of array processing, since multiple clusters are constructed from a set of beams with a uniform separation of beam direction and the simultaneous transmission with these clusters has a good potential of interuser interference suppression. Moreover, the increased density of subspace packing in the proposed SDMA codebook will considerably improve the performance of the individual user. By far the first two problems mentioned above have been addressed to some degree in the extension of SDMA codebook. As for the third problem, we use computational simulations to seek a reasonable tradeoff between the number of clusters and the cluster size. The simulation results given in Section 5 show that a four-cluster SDMA codebook provides a better tradeoff between the cluster number and the cluster size than the IEEE 802.20 SDMA codebook does.

4. SCHEDULING ALGORITHM

The goal of an SDMA scheduling algorithm is to fully exploit the multiuser diversity [27–29] and spatial multiplexing gain. The extension of SDMA codebook disables the scheduling algorithm adopted in IEEE 802.20. In this section, we propose two schemes of opportunistic scheduling which only require limited feedback information. The proposed algorithms would carefully schedule multiple users on a fixed number of n_T active substreams, aiming at a maximum sum throughput. Although a fixed number of active substreams are not always optimal, the adaptation of the number of active substreams will dramatically increase the feedback overhead and scheduling complexity. This section first focuses on the configuration of $n_T = 2n_R$, in which each individual user is multiplexed with n_R substreams. Thus, the scheduling algorithm needs to select two MIMO-transmission users each time. The extension of the scheduling algorithms to a more general configuration is introduced in Section 4.3.

4.1. Markov opportunistic scheduling

To implement SDMA scheduling, each user should first derive the feedback information based on the proposed codebook and send it to the base station. After collecting the feedback information from all the users, the base station should then select several simultaneously active users and assign them specific precoders. The selection scheme should guarantee that the users in good channel conditions be selected and the matched precoders be assigned, such that the multiuser diversity is fully exploited. Also, the selection scheme should guarantee some degree of orthogonality among the effective channels of the simultaneously active users. Unfortunately, these two requirements usually conflict with each other, thus we need to find a good tradeoff between them.

In this section, we propose a novel opportunistic scheduling algorithm which is implemented in a Markov-like way. In the proposed algorithm, each feedback information from a user falls into two parts denoted as part I and part II. For the kth scheduling, part II information from the current feedback along with part I information from the previous feedback, that is, the $(k-1)$th feedback, will be used to select active users and assign precoders. In particular, as depicted in Figure 2, before the actual SDMA transmission, three scheduling stages are employed to complete user selection and precoder assignment. In the first stage, part I information from the previous feedback is utilized to select the first active user, henceforth denoted as the main user, and assigns its preferred precoder. At the same time, the expected interfering cluster is determined to restrict the range of precoder selection for the second active user (henceforth denoted as the secondary user), so that the interference from the main user on the secondary user is controlled. In order that all the users respond to these selection results, the indices of the preferred precoder for the main user and the expected interfering cluster, termed as preschedule information, are broadcast immediately, which finishes the job of the first stage. In the second stage, each user derives the feedback information (two parts) based on the preschedule information and transmits it to the base station. In the last stage, the base station utilizes part II information from the current feedback to select the second active user and its preferred precoder included in the expected interfering cluster. Note that in this stage the interference from the secondary user to the main user can be controlled. If the channels keep quasistatic between these two continuous scheduling intervals, the above Markov OS scheme succeeds in suppressing the mutual interference between two successively scheduled users, that is, the main and secondary users, and the multiuser diversity is achieved as well through the above successive user selection.

To present the above scheduling algorithm in more detail, we will first introduce the derivation for the two parts of feedback information, and then present the related selection/assignment scheme.

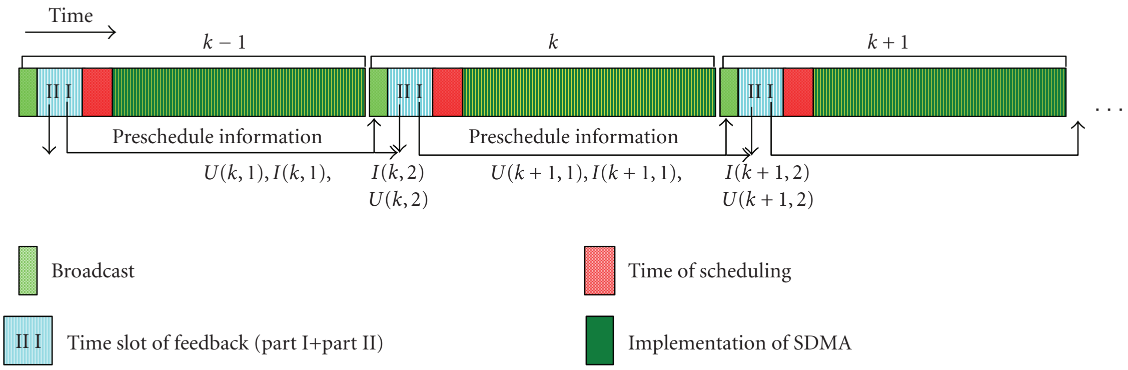
4.1.1. Feedback information part I

This part of feedback information includes the user's maximum supported rate in SDMA mode, the indices of its preferred precoder, and the expected interfering cluster.

We assume perfect channel state information (CSI) is available at the user terminal. For the kth user, in order to obtain its maximum supported rate in SDMA mode, we assume the base station will transmit signal $\mathbf{s}_1$ to this user using a precoder $\mathbf{U}_1$, and the base station will simultaneously transmit $\mathbf{s}_2$ with equal power to the other user using a precoder $\mathbf{U}_2$, where $\mathbf{U}_1$ and $\mathbf{U}_2$ can be any different precoders selected from the SDMA codebook, and $[\mathbf{U}_1, \mathbf{U}_2]$ form the resultant multiuser precoder. Thus, a virtual receive model for the user k can be expressed as

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{U}_1\mathbf{s}_1 + \mathbf{H}_k\mathbf{U}_2\mathbf{s}_2 + \mathbf{n}_k. \tag{8}$$

Then, the maximum supported rate of the user k in SDMA mode will be obtained by searching all the possible receive models with respect to different $\mathbf{U}_1$ and $\mathbf{U}_2$.



* $U(k,i)$ indicates the scheduled ith user of the kth implementation of SDMA
$I(k,i)$ indicates the scheduled ith precoding matrix for the kth implementation of SDMA

FIGURE 2: The conceptual model of Markov opportunistic scheduling.

If an MMSE linear receiver is employed, the output of its linear filtering for the virtual receive signal is written as

$$\mathbf{z}_k = \mathbf{G}_k^H \mathbf{y}_k = \mathbf{G}_k^H \mathbf{H}_k \mathbf{U}_1 \mathbf{s}_1 + \mathbf{G}_k^H \mathbf{H}_k \mathbf{U}_2 \mathbf{s}_2 + \mathbf{G}_k^H \mathbf{n}_k, \tag{9}$$

where

$$\mathbf{G}_k = \left(\mathbf{H}_k \mathbf{U}_1 \mathbf{U}_1^H \mathbf{H}_k^H + \mathbf{H}_k \mathbf{U}_2 \mathbf{U}_2^H \mathbf{H}_k^H + \frac{N_0}{E_s} \mathbf{I}_N \right)^{-1} \mathbf{H}_k \mathbf{U}_1. \tag{10}$$

For the ith possible value of $(\mathbf{U}_1, \mathbf{U}_2)$, denoted as $(\mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)})$, the corresponding $\mathbf{G}_k$ is denoted as $\mathbf{G}_k^i$, the signal to interference plus noise ratio (SINR) for the mth data stream can be represented as

$$\mu_{k,m,i} = \frac{Q}{R+S+Z}, \quad m \in [1, M], \tag{11}$$

where Q denotes $|((\mathbf{G}_k^{(i)})^H \mathbf{H}_k \mathbf{U}_1^{(i)})_{mm}|^2$, R denotes $\sum_{n \neq m} |((\mathbf{G}_k^{(i)})^H \mathbf{H}_k \mathbf{U}_1^{(i)})_{mn}|^2$, S denotes $\|((\mathbf{G}_k^{(i)})^H \mathbf{H}_k \mathbf{U}_2^{(i)})_m\|^2$, and Z denotes $(N_0/E_s)\|((\mathbf{G}_k^{(i)})^H)_m\|^2$, among which $(\mathbf{A})_{nm}$ denotes the (n,m) entry of the matrix $\mathbf{A}$, and $(\mathbf{A})_m$ denotes the mth row of the matrix $\mathbf{A}$. The expression of SINR can be further simplified as

$$\mu_{k,m,i} = \frac{1}{1 - \left[(\mathbf{U}_1^{(i)})^H \mathbf{H}_k^H \mathbf{G}_k^{(i)} \right]_{kk}} - 1, \quad m \in [1, M]. \tag{12}$$

The ideal supported rate associated with $(\mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)})$ is then written as

$$C_{k,i} = \sum_{m=1}^{M} \log(1 + \mu_{k,m,i}). \tag{13}$$

To seek the maximum supported rate of the user k, we need to search all the possible $(\mathbf{U}_1, \mathbf{U}_2)$. Mathematically, the maximum supported rate is expressed as

$$C_k^1 = \underbrace{\max}_{i \in [1 \quad L(L-1)]} \{C_{k,i}\}, \tag{14}$$

where $L = N_c L_c$ denotes the size of codebook, and $L(L-1)$ represents the search space for $(\mathbf{U}_1, \mathbf{U}_2)$. The specific values of $\mathbf{U}_1$ and $\mathbf{U}_2$ which achieve C_k^1 are viewed as the user's preferred precoder (henceforth denoted as I_k^1) and the expected interfering precoder in SDMA mode, respectively.

As for the computational complexity, obviously the exhaustive elementwise search of both $\mathbf{U}_1$ and $\mathbf{U}_2$ over the SDMA codebook requires $L(L-1)$ times of rate calculation using (12)-(13). It is to be noted that the cluster-based structure of the proposed codebook can be exploited to greatly reduce the search complexity. To see this, we first investigate the impact of a unitary perturbation to $\mathbf{U}_1$ and $\mathbf{U}_2$ on the sum rate supported by the SDMA system with a linear receiver. We define the unitary perturbed precoders as $\widetilde{\mathbf{U}}_1 = \mathbf{U}_1 \mathbf{A}_1$ and $\widetilde{\mathbf{U}}_2 = \mathbf{U}_2 \mathbf{A}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ denote the unitary matrices. It follows from (12) that the SINR for the mth stream in the SDMA system using the precoders $(\widetilde{\mathbf{U}}_1, \widetilde{\mathbf{U}}_2)$ is expressed as

$$\widetilde{\mu}_{k,m,i} = \frac{1}{1 - [V]_{kk}} - 1, \quad m \in [1, M], \tag{15}$$

where V denotes $(\mathbf{A}_1^H \mathbf{U}_1^H \mathbf{H}_k^H (\mathbf{H}_k \mathbf{U}_1 \mathbf{U}_1^H \mathbf{H}_k^H + \mathbf{H}_k \mathbf{U}_2 \mathbf{U}_2^H \mathbf{H}_k^H + (N_0/E_s)\mathbf{I}_N)^{-1} \mathbf{H}_k \mathbf{U}_1 \mathbf{A}_1)$. It is seen that a perturbation to the precoder $\mathbf{U}_1$ affects the value of SINR, which in turn affects the system performance, actually, the unitary perturbation has a potential of compensating for the performance loss caused by a linear receiver, more details can be found in [22], while a perturbation to the interfering precoder $\mathbf{U}_2$ has no impact on the SINR value, which means that different precoders included in the same cluster produce the same influence on the rate performance of the receive model expressed as (8). Thus, in order to search for the maximum supported rate, we need only a clusterwise search of $\mathbf{U}_2$ together with an elementwise search of $\mathbf{U}_1$, instead of an exhaustive elementwise search of both $\mathbf{U}_1$ and $\mathbf{U}_2$. Moreover, the search case that $\mathbf{U}_1$ and $\mathbf{U}_2$ fall into the same cluster can be omitted due to the excessive interuser interference. Hence, we finally only need $L(N_c - 1)$ times of rate calculations. Also, the specific interfering cluster agreeing with the maximum

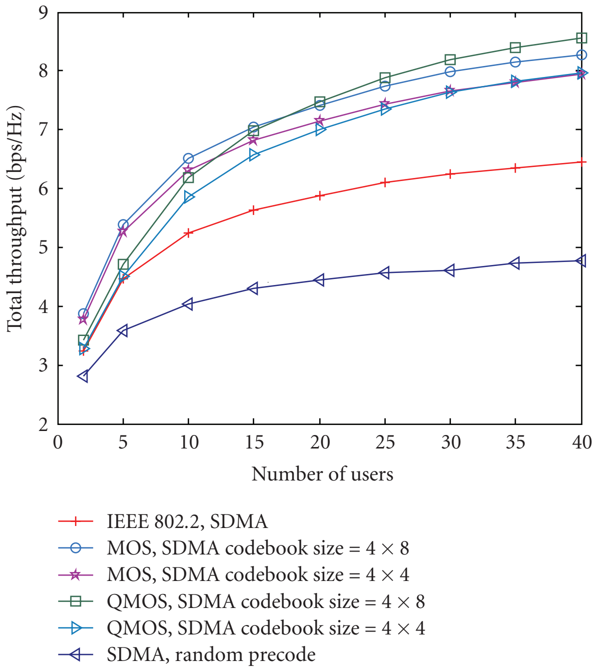


Figure 3: System throughput versus number of users, proposed SDMA codebook combining Markov opportunistic scheduling (MOS) or quick matching opportunistic scheduling (QMOS), SNR = 10 dB, mobile velocity of 3 km/h.

rate is denoted as the expected interfering cluster (henceforth expressed as J_k^2).

4.1.2. *Feedback information part II*

Part II feedback information is derived based on the broadcast preschedule information, and includes the user's maximum supported rate (denoted as C_k^2) and the index of its preferred precoder (denoted as I_k^2), both in prescheduled mode, where the prescheduled mode means that the user is viewed as a candidate for the second active user, with the precoder for the main user determined and its possible precoder restricted in the expected interference cluster.

The virtual model of (8) will be used again to find C_k^2 and I_k^2 by searching the expected cluster. Note that I_k^2 actually denotes the relative index of the preferred precoder in the restricting cluster. As for the computational complexity, since now $\mathbf{U}_2$ in (8) has been determined in the preschedule information, and the search with respect to $\mathbf{U}_1$ is restricted in a given cluster, we only need L_c times of rate calculations to obtain C_k^2 and I_k^2.

4.1.3. *Selection scheme*

As stated above, two parts of feedback information are denoted as C_k^1, I_k^1, and J_k^2 (part I); C_k^2 and I_k^2 (part II). We have clarified in the second paragraph of Section 4.1 that in the first stage, a main active user and its preferred precoder are selected by using part I feedback $\{C_k^1, I_k^1, J_k^2\}$ from the previous feedback, and in the third stage, the main task is to select the second active users and its preferred precoder.

If we define

$$\hat{k} = \underbrace{\arg\max}_{k}\{C_k^1\}, \tag{16}$$

then $\hat{k}$ is the index of the main active user, $I_{\hat{k}}^1$ is the index of its preferred precoder, $J_{\hat{k}}^2$ is the index of its expected interfering cluster. Noting that some individual users may terminate traffic request at the current scheduling, the C_k^1 associated with these users should be first removed from the set $\{C_k^1\}$ in the above process. In order to control the interference from the secondary active user to the main active user, we restrict the precoder of the second active user in the cluster indexed by $J_{\hat{k}}^2$.

Note that the third stage is implemented until all the users have responded to the broadcast preschedule information and complete feedback at the current scheduling. The set $\{C_k^2, I_k^2\}$ from the current feedback will be utilized to select the second active users and its preferred precoder.

If we define

$$\tilde{k} = \underbrace{\arg\max}_{k}\{C_k^2\}, \tag{17}$$

then $\tilde{k}$ is the index of the secondary user, $I_{\tilde{k}}^2$ indicates the relative position of its preferred precoder in the given cluster indexed by $J_{\hat{k}}^2$.

4.1.4. *Feedback overhead and scheduling complexity*

We assume the SDMA codebook consists of $N_c = 2n_U$ clusters and each cluster has L_c codewords, namely, the size of codebook is $L = N_c \cdot L_c$. The total number of feedback bits for $\{C_k^1, I_k^1, J_k^2, C_k^2, I_k^2\}$ is $Q(C_k^1) + Q(C_k^2) + 2\log(N_c \cdot L_c)$, where $Q(x)$ denotes the number of bits used to quantize the scalar quantity x. This amount is about twice the feedback amount acquired by the IEEE 802.20 SDMA. In addition, the number of the broadcast bits used in preschedule information $\{I_{\hat{k}}^1, J_{\hat{k}}^2\}$ is $\log(N_c \cdot L_c) + \log(N_c)$.

The majority of the scheduling complexity lies in the calculation of feedback information and the comparison of the supported rate among all the users. Since the comparison operation aims to find the maximum one, its complexity linearly increases with the number of users. The complexity of feedback information calculation mainly depends on the times of rate calculation using (12)-(13). In the proposed MOS algorithm, each user needs $N_c L_c(N_c - 1) + L_c$ times of rate calculations.

4.2. ***Quick matching opportunistic scheduling***

The above MOS requires each user to feed back two parts of information. In this section, we propose another opportunistic scheduling algorithm which requires less feedback information, where the cluster structure of the proposed SDMA codebook is fully exploited to reduce the feedback information and simplify the scheduling algorithm.

4.2.1. Feedback information

In every schedule, each user k feeds back the following information: (1) the maximum supported rate in SDMA mode, denoted as C_k; (2) the index of its preferred precoder, denoted as I_k; (3) and the index of its expected interfering cluster, denoted as J_k. The way of obtaining the above feedback information is similar to that introduced in Section 4.1, which will not be repeated.

4.2.2. Selection scheme

After gathering feedback information from all the users, the base station will first classify the users according to the feedback indices, every two classes with good orthogonality between their effective channels are paired together. Through a pairwise comparison, the users in favorable channels and with good interuser interference suppression can be easily found. The detailed processes are provided as follows.

(1) We define the cluster including the preferred precoder as the preferred cluster. For the user k, we denote the index of its preferred cluster as $\widetilde{J}_k$. Based on this information plus the index of the expected interfering cluster $\{J_k\}$, the users can be classified as follows: the users with the same preferred cluster m and the same expected interfering cluster n are classified into one class, denoted by (m, n), namely,

$$\text{Class}(m,n) \triangleq \{k \mid \widetilde{J}_k = m,\ J_k = n\}. \tag{18}$$

If we assume that two active users must be scheduled on different clusters, there will be $N_g = 2n_U \cdot (2n_U - 1)$ different classes.

(2) We regard the class (m, n) and class (n, m) as one pair, denoted as $n \sim m$. Since the users in the paired classes have their preferred cluster and expected interfering cluster pointed to each other, the effective channels between these two classes of users have good orthogonality characteristic, which means that the mutual interference between these two classes can be suppressed.

(3) For the pair $n \sim m$, we find one user in class (m, n) and one user in class (n, m) such that the sum of their supported rate is maximum for that pair. In this way, the multiuser diversity will be exploited.

(4) Among all the possible pairs, we find the pair with the maximum sum rate, the two users which achieve the maximum rate will be scheduled as the current active users, and their preferred precoders are scheduled as well.

In essence, this algorithm aims at finding two users with matching channel conditions and both in favorable conditions; hence we call it quick matching opportunistic scheduling. Note that it is possible for the user in good channel condition to be unable to find its matching user while its supported rate is bigger than the sum rate of any two matched users. In this case, the algorithm will schedule this single user to transmit signal, and no other simultaneously active user will be scheduled together. Compared with MOS, this QMOS only depends on the feedback information from the current scheduling, and requires no broadcast of preschedule information at the very beginning of scheduling.

It is worth mentioning that the proposed algorithm reduces to the one adopted in IEEE 802.20 when the proposed SDMA codebook only consists of two clusters, that is, $n_U = 1$. Since now the number of the available classes is $N_g = 2n_U \cdot (2n_U - 1) = 2$, only a unique pair exists in the scheduling algorithm and thus the fourth step can be omitted. Also, the feedback of the index of the expected interfering cluster is redundant now, since the expected interfering cluster must be the opposite one to the preferred cluster.

4.2.3. Feedback overhead and scheduling complexity

The feedback information consists of $\{C_k, I_k, J_k\}$, the number of the total feedback bits is $Q(C_k) + \log(N_c \cdot L_c)$, which is half of that acquired by MOS.

We focus on the computational complexity of feedback information calculation for each user. In order to obtain the maximum rate, we need $N_c L_c (N_c - 1)$ times of rate calculations using (12)-(13), which is slightly smaller than that required in MOS.

4.3. Extension of proposed scheduling algorithms

The above scheduling algorithms only work in the case of two simultaneously active users. Here, we extend the proposed algorithms into a more general case. We assume that the antenna configuration satisfies $n_R \in \{n_T/2, n_T/4, n_T/8, \ldots, 1\}$, such that the SDMA may include more than two simultaneously active users. In order to extend the proposed scheduling algorithms for this case, we still employ the SDMA codebook with the precoders in size of $n_T \times n_T/2$, with one precoder possibly carrying signal intended for more than one user. Instead of selecting two preferred active users and their preferred precoders, the extended OS algorithms each time will select two preferred user groups and two preferred precoders, each user group transmitting signal with one preferred precoder. If we view one user group as one virtual user and provide it with the equivalent feedback information, the extension of the proposed OS algorithms introduced in Sections 4.1 and 4.2 is straightforward. Thus, the key to the extension lies in the definition of the virtual user.

If the number of multiplexing substreams of user k is less than $n_T/2$, that is, $M < n_T/2$, only a submatrix in one precoder will be used by such a user. Thus, the virtual receive model of (8) should be modified to find its maximum supported rate, together with both its preferred precoder and preferred submatrix. To this end, $\mathbf{U}_1$ is first partitioned as $\mathbf{U}_1 = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_P]$, where $\mathbf{\Phi}_1 \in \mathcal{C}^{n_T \times M}$ and $P = n_T/M$. The virtual receive model of user k is modified as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{\Phi}_p \mathbf{s}_1 + \mathbf{H}_k \overline{\mathbf{\Phi}}_p \overline{\mathbf{s}}_1 + \mathbf{H}_k \mathbf{U}_2 \mathbf{s}_2 + \mathbf{n}_k, \tag{19}$$

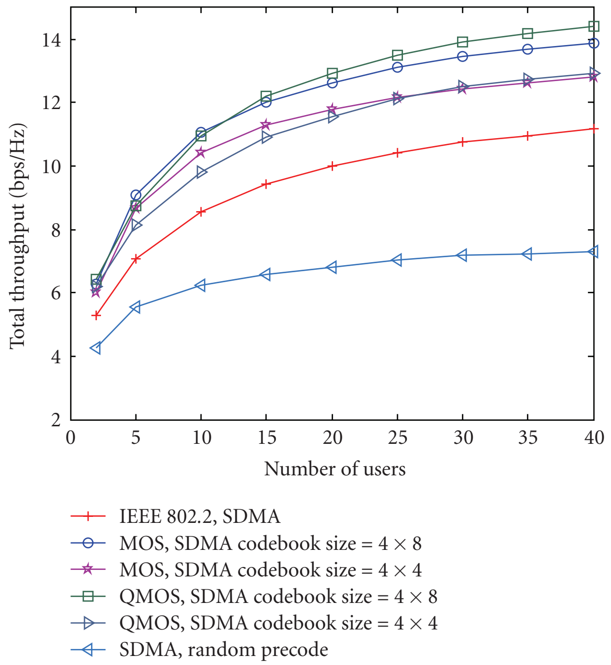


Figure 4: System throughput versus number of users, proposed SDMA codebook combining Markov opportunistic scheduling (MOS) or quick matching opportunistic scheduling (QMOS), SNR = 20 dB, mobile velocity of 3 km/h.

where $\mathbf{s}_1$ denotes the transmit signal of the user k, and $\mathbf{\Phi}_p$ is its transmit matrix; $\overline{\mathbf{\Phi}}_p$ denotes the matrix constructed by subtracting $\mathbf{\Phi}_p$ from $\mathbf{U}_1$; $\bar{\mathbf{s}}_1$ denotes the signal transmitted to the other users with the beams included in $\overline{\mathbf{\Phi}}_p$. This model assumes that each time two precoders $\mathbf{U}_1$ and $\mathbf{U}_2$ are selected, and the users associated with the transmit signal of $\mathbf{s}_1$ and $\bar{\mathbf{s}}_1$ are viewed as a user group scheduled on the same precoder $\mathbf{U}_1$. Similar to the method introduced in Section 4.1, the SINR of the mth substream of user k can be easily calculated. In order to obtain the maximum supported rate in SDMA mode, in the modified model we need to search not only all the possible values of $(\mathbf{U}_1, \mathbf{U}_2)$ but also those of $\mathbf{\Phi}_p$. If we express one specific SINR value of the mth substream as $\mu_{k,m,i,p}$, where the indices i and p indicate the specific selection of $(\mathbf{U}_1, \mathbf{U}_2)$ and that of $\mathbf{\Phi}_p$, respectively. The maximum supported rate can then be calculated as

$$C_k = \underbrace{\max}_{i,p} \sum_{m=1}^{M} \log\left(1 + \mu_{k,m,i,p}\right). \tag{20}$$

Similarly, we define the specific selection of $\mathbf{U}_1$ achieving C_k as the preferred precoder, denoted by I_k (with a slight notation abuse). Also, the definition of the expected interfering cluster follows that introduced in Section 4.1. Since user k now only uses a submatrix of its preferred precoder as the transmit matrix, a new index P_k is required to feed back to indicate the position of the preferred submatrix. We classify the users reporting the same preferred precoder into one set, which is expressed as

$$\mathcal{S}_i = \{k \mid I_k = i\}, \tag{21}$$

where i denotes the index of a specific precoder, the elements in the set denote the user index. In order to define the virtual user, we further divide the set into multiple subsets $\mathcal{S}_{i,j}$, where j denotes the subset index. The criterion is to make the collection of all the preferred transmit matrices of the user subset exactly constitute the preferred precoder, namely, $\mathcal{S}_{i,j}$ satisfies the following property:

$$\bigcup_{k \in \mathcal{S}_{i,j}} \{P_k\} = \{1, 2, \ldots, P\}, \tag{22}$$

Such a user subset is then defined as a virtual user. Based on the feedback information, the maximum supported sum capacity of a virtual user can be calculated as $\sum_{k \in \mathcal{S}_{i,j}} C_k$. Provided with the definitions of all the possible virtual users, their corresponding maximum supported sum rate, their preferred precoder, and expected interfering cluster, the extension of MOS and QMOS remains to replace the selection of a preferred user with that of a preferred virtual user.

5. SIMULATION RESULTS

We assume a multiuser network where multiple users are randomly distributed around the base station and have the same distance to the base station. The channels between the users and the base station experience time-varying flat fading, and the 3GPP spatial channel model is used to simulate the outdoor multiuser MIMO channels, in which it is assumed that the time variation of the channels depends on the mobile speed. We consider the configuration of $(n_T, n_R) = (4, 2)$ and a scheduling interval of 5 milliseconds. At the beginning of the interval, the users calculate their feedback information and transmit them to the base station, after collecting the feedback information the base station schedules two simultaneously active users and implement the SDMA transmission at the end of the interval. Therefore, a delay error of 5 milliseconds arises between the channels of SDMA scheduling and implementation. The Monte Carlo simulations will take this channel delay error into consideration but assume no channel estimation error and feedback error. We assume all the receivers have the same noise variance, and define SNR $= n_T E_s/N_0$. The system throughput is calculated as the sum of the throughputs of all the simultaneously active users, where the throughput for each active user can be calculated with (12)–(13), in which $\mathbf{H}_k, \mathbf{U}_1$, and $\mathbf{U}_2$ should be replaced with the actual ones in the simulation.

Figures 3 and 4 illustrate the system throughput of the proposed SDMA schemes under channels with 3 km/h mobile speed. For comparison, the IEEE 802.20 SDMA [16] and the opportunistic SDMA with an arbitrary unitary precoder [14] are simulated as well. Note that the IEEE 802.20 SDMA is also based on a codebook, and the required feedback information includes a CQI (such as the maximum supported rate) and an index of the preferred precoder.

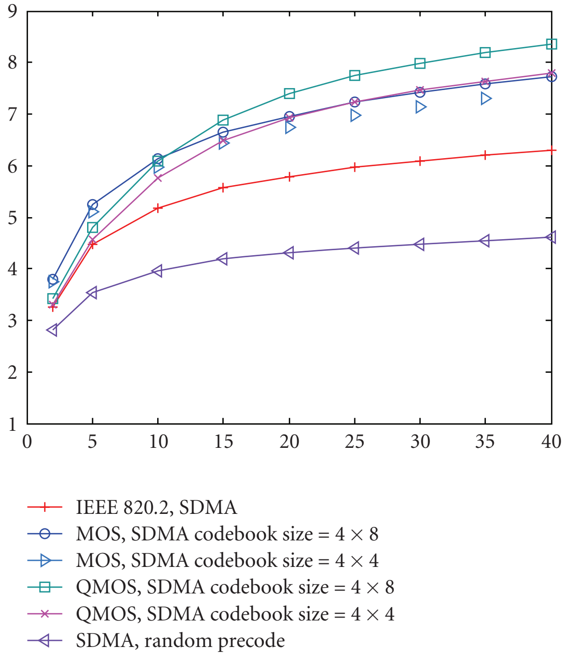


Figure 5: System throughput versus number of users, proposed SDMA codebook combining Markov opportunistic scheduling (MOS) or quick matching opportunistic scheduling (QMOS), SNR = 10 dB, mobile velocity of 30 km/h.

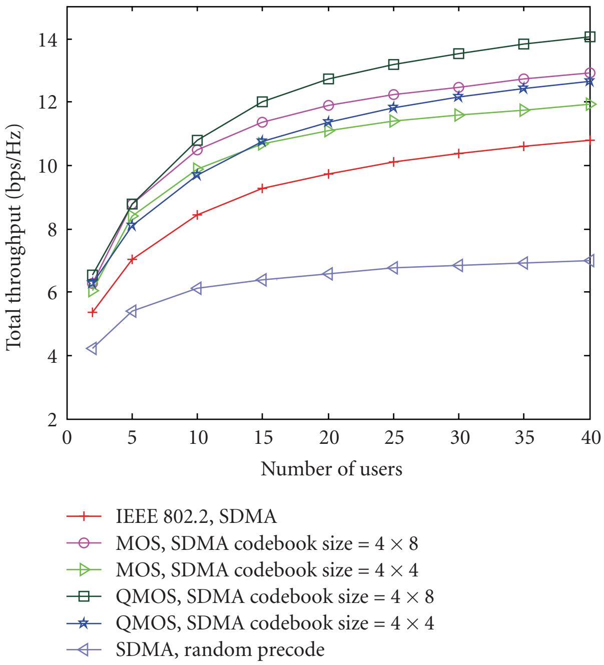


Figure 6: System throughput versus number of users, proposed SDMA codebook combining Markov opportunistic scheduling (MOS) or quick matching opportunistic scheduling (QMOS), SNR = 20 dB, mobile velocity of 30 km/h.

Thus the feedback overhead is comparable to that in the proposed QMOS SDMA. The simulated IEEE 802.20 SDMA employs a codebook of size 28 (2 clusters, 14 precoders in each cluster). Results show that both of the two proposed SDMAs outperform the IEEE 802.20 SDMA in terms of system throughput, even with the codebook of size 4 × 4 ($m \times n$ means $N_c = m$ and $L_c = n$). In particular, the proposed SDMAs exhibit 1 ~ 1.5 bps/Hz gain at the SNR of 10 dB and 1.5 ~ 2 bps/Hz gain at the SNR of 20 dB, with the number of users ranging from 15 to 40. With the increase of the codebook size, more gain is observed, especially when the user number is large. This gain is obtained by exploiting the potential of performance enhancement offered by the proposed SDMA codebook. Simulation results also show that in sparse networks the SDMA with an arbitrary unitary precoder [14] has a much poorer performance than the SDMA employing SDMA codebook.

In addition, it is seen that the MOS SDMA outperforms the QMOS SDMA in the case of small number of users, while the opposite result (the QMOS SDMA outperforms the MOS SDMA) is observed in the case of large number of users. Since the MOS selects active users in a successive way, it always succeeds in scheduling together two active users for simultaneous transmission, while the QMOS may fail in scheduling two simultaneously active users, especially when the number of users is insufficient. On the other hand, the successive scheduling way may lose some multiuser diversity at the second step where the selection of the secondary scheduled user is restricted by the already scheduled main user. These characteristics can explain why an intersection exists between the performances of the proposed two SDMAs.

Figures 5 and 6 illustrate the system throughput of the SDMA schemes under channels with 30 km/h mobile speed. Simulation results show that the proposed SDMAs still outperform the IEEE 802.20 SDMA under this channel environment. In contrast to the case of 3 km/h mobile speed, the performance of all the SDMAs degrades slightly, which is caused by the more severe delay error of feedback information [30]. When the number of users ranges in 20 ~ 40, the MOS SDMA has a throughput loss of 6 ~ 7%, while the other two SDMAs have a loss of 2 ~ 3%. Note that different from the other two SDMAs, the MOS SDMA depends on not only the feedback information at the current scheduling, but also that at the previous scheduling. Thus, its performance loss is determined by the time variation of channels between two continuous scheduling intervals, which is more severe than the other SDMAs.

In the previous simulations of MOS SDMA, we have assumed that the traffic request of all the users between two continuous scheduling intervals keep static. In practical scenarios, some users may terminate traffic request at a certain scheduling interval. At the same time some new users may be added with traffic request. Figure 7 provides the system throughput performance of the MOS SDMA in such a practical scenario; in each schedule time some percent of users are newly added while some users terminate traffic request, where we assumed the number of newly added users is equal to that of the users terminating traffic request. Simulation results show that the performance of the

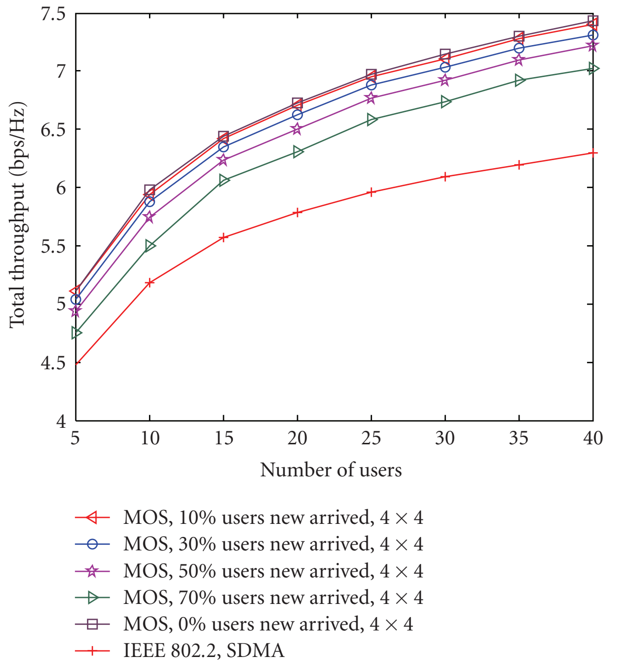


Figure 7: System throughput versus number of users, proposed SDMA codebook combining Markov opportunistic scheduling (MOS), SNR = 10 dB, mobile velocity of 30 km/h.

MOS SDMA degrades with the increasing of the percent of the newly added users. That is because MOS SDMA cannot utilize the feedback information of the recently added users in the process of scheduling a main user and thereby results in a multiuser diversity loss. However, it is seen that, under channel condition with 30 km/h mobile speed, the performance loss with 10% newly users is very small. Even with 70% users newly added, the MOS SDMA employing 4 × 4 codebook still outperforms the IEEE 802.20 SDMA.

Figure 8 illustrates the average rate distortion of the proposed QMOS SDMA, in comparison with that of the IEEE 802.20 SDMA. We have assumed no channel delay error in this simulation. If we use $\mathrm{Th}^{\mathrm{QMOS}}(\mathbf{H})$ to denote the throughput obtained by the QMOS SDMA for one time of channel realization, and use $\mathrm{Th}^{\mathrm{BD}}(\mathbf{H})$ to denote the throughput obtained by the block diagonalized precoding [6] with full channel knowledge of all the active users, the average rate distortion is defined as $E[\mathrm{Th}^{\mathrm{BD}}(\mathbf{H}) - \mathrm{Th}^{\mathrm{QMOS}}(\mathbf{H})]$, where $E[\cdot]$ denotes the expectation over the channel realization. The simulation results show that the average distortion increases with the number of the users, which implies that the SDMA schemes employing the SDMA codebook are more effective in sparse networks. It is also seen that the average distortion decreases with the increasing of the cluster number N_c and the cluster size L_c. It is worth mentioning that, even with a smaller codebook size ($L = N_c \cdot L_c$), the SDMA codebook packed with 4 clusters exhibits a lower distortion than the codebook packed with 2 clusters.

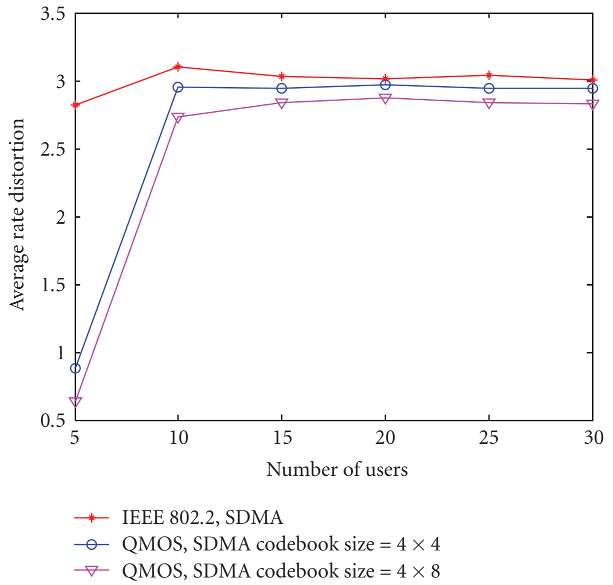


Figure 8: Average rate distortion versus number of users, proposed QMOS SDMA and IEEE 802.20 SDMA, SNR = 10 dB, static channels.

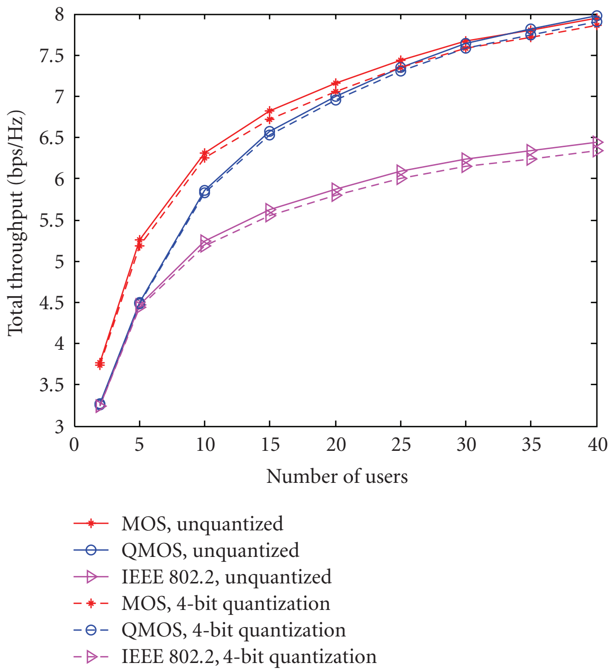


Figure 9: System throughput versus number of users, proposed SDMA codebook combining Markov opportunistic scheduling (MOS) or quick matching opportunistic scheduling (QMOS), quantized feedback, SNR = 10 dB, mobile velocity of 3 km/h.

Finally, the effect on the throughput performance of quantizing the feedback scalar quantity, that is, the maximum supported rate, in the proposed opportunistic scheduling algorithms is investigated by simulation. As shown in Figure 9, when a simple linear quantization with 4 bits is employed, the performance loss of the proposed two scheduling algorithms is minor and can be ignored in the case of small number of users. It is also found that, compared with the IEEE 802.20 SDMA scheme, the superiority of the proposed schemes in terms of the throughput is obviously preserved in the quantization case.

6. CONCLUSION

In this paper, we have presented a novel design method for limited feedback SDMA by combining the codebook-based multiuser precoding and the opportunistic scheduling. We have first proposed an SDMA codebook construction from the perspective of array processing and unitary perturbation. The proposed codebook provides a good property of interuser interference suppression, and its cluster-based structure is helpful for simplifying the scheduling algorithm. Then, we proposed two codebook related opportunistic scheduling algorithms, that is, a Markov OS and a quick matching OS. The MOS schedules active users and their precoders in a successive way, while the QMOS schedules by a way of classifying plus matching. Simulation results have shown that in sparse networks, the proposed SDMAs outperform the IEEE 802.20 SDMA in terms of throughput, while both having a comparable feedback overhead.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Bin Li, Dr. Yinggang Du, and Professor Wei-Ping Zhu for their helpful discussion. The authors would also like to thank all the anonymous reviewers for their valuable comments that have helped improve the presentation of this paper. This work was supported by National Basic Research Program of China (2007CB310603), National Natural Science Foundation of China (60672093, 60496310, and 60572105), High Technology Project of China (2007AA01Z262), the Program for New Century Excellent Talents (NCET-05-0582) in University of China, and the Grant of Huawei Technologies Corporation.

REFERENCES

[1] M. H. M. Costa, “Writing on dirty paper,” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[2] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, “Precoding in multiantenna and multiuser communications,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1305–1316, 2004.

[3] K. Kusume, M. Joham, W. Utschick, and G. Bauch, “Efficient Tomlinson-Harashima precoding for spatial multiplexing on flat MIMO channel,” in *Proceedings of IEEE International Conference on Communications (ICC ’05)*, vol. 3, pp. 2021–2025, Seoul, Korea, May 2005.

[4] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, “Space-time transmission using Tomlinson-Harashima precoding,” in *Proceedings of the 4th ITG Conference on Source and Channel Coding*, pp. 139–147, Berlin, Germany, January 2002.

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

[6] K.-K. Wong, R. D. Murch, and K. B. Letaief, “A joint-channel diagonalization for multiuser MIMO antenna systems,” *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 773–786, 2003.

[7] S. Serbetli and A. Yener, “Transceiver optimization for multiuser MIMO systems,” *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 214–226, 2004.

[8] A. Tarighat, M. Sadek, and A. H. Sayed, “A multiuser beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’05)*, vol. 3, pp. 1129–1132, Philadelphia, Pa, USA, March 2005.

[9] M. Schubert and H. Boche, “Solution of the multiuser downlink beamforming problem with individual SINR constraints,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, 2004.

[10] A. Morell, A. Pascual-Iserte, A. I. Pérez-Neira, and M. A. Lagunas, “Robust scheduling in MIMO-OFDM multiuser systems based on convex optimization,” in *Proceedings of the 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP ’05)*, pp. 113–116, Puerto Vallarta, Mexico, December 2005.

[11] N. Jindal, “MIMO broadcast channels with finite-rate feedback,” *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[12] P. Ding, D. J. Love, and M. D. Zoltowski, “Multiple antenna broadcast channels with limited feedback,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’06)*, vol. 4, pp. 25–28, Toulouse, France, May 2006.

[13] K. Huang, J. G. Andrews, and R. W. Heath Jr., “Orthogonal beamforming for SDMA downlink with limited feedback,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’07)*, pp. 97–100, Honolulu, Hawaii, USA, April 2007.

[14] M. Sharif and B. Hassibi, “On the capacity of MIMO broadcast channels with partial side information,” *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 506–522, 2005.

[15] M. Kountouris and D. Gesbert, “Robust multiuser opportunistic beamforming for sparse networks,” in *Proceedings of the 6th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC ’05)*, pp. 975–979, New York, NY, USA, June 2005.

[16] IEEE 802.20 C802.20-06-04, “Part 12: precoding and SDMA codebooks,” January 2006.

[17] D. J. Love, R. W. Heath Jr., and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[18] D. J. Love and R. W. Heath Jr., “Limited feedback unitary precoding for spatial multiplexing systems,” *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, 2005.

[19] S. Zhou and B. Li, “BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers,” *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1653–1665, 2006.

[20] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, “From single user to multiuser communications: shifting the MIMO paradigm,” *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007.

[21] R. de Francisco, M. Kountouris, D. Slock, and D. Gesbert, “Orthogonal linear beamforming in MIMO broadcast channels,” in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC ’07)*, pp. 1210–1215, Kowloon, China, March 2007.

[22] Y. Huang, D. Xu, L. Yang, and W.-P. Zhu, "A limited feedback precoding system with hierarchical codebook and linear receiver," to appear in *IEEE Transactions on Wireless Communications*.

[23] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1962–1973, 2000.

[24] A. Barg and D. Yu. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2450–2454, 2002.

[25] R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Boston, Mass, USA, 1993.

[26] C. Mun, J.-K. Han, and D.-H. Kim, "Quantized principal component selection precoding for limited feedback spatial multiplexing," in *Proceedings of IEEE International Conference on Communications (ICC '06)*, vol. 9, pp. 4149–4154, Istanbul, Turkey, June 2006.

[27] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.

[28] J. Chung, C.-S. Hwang, K. Kim, and Y. K. Kim, "A random beamforming technique in MIMO systems exploiting multiuser diversity," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 848–855, 2003.

[29] S.-S. Huang and Y.-H. Lee, "Multi-beam multiplexing using multiuser diversity and random beams in wireless systems," in *Proceedings of IEEE International Conference on Communications (ICC '05)*, vol. 4, pp. 2717–2721, Seoul, Korea, May 2005.

[30] V. Hassel, M.-S. Alouini, G. E. Øien, and D. Gesbert, "Rate-optimal multiuser scheduling with reduced feedback load and analysis of delay effects," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, Article ID 36424, 7 pages, 2006.