Channel Quantization for Block Diagonalization with Limited Feedback in Multiuser MIMO Downlink Channels

Sung-Hyun Moon, Sang-Rim Lee, Jin-Sung Kim, and Inkyu Lee

Abstract: **Block diagonalization (BD) has been proposed as a simple and effective technique in multiuser multiple-input multiple-output (MU-MIMO) broadcast channels. However, when channel state information (CSI) knowledge is limited at the transmitter, the performance of the BD may be degraded because inter-user interference cannot be completely eliminated. In this paper, we propose an efficient CSI quantization technique for BD precoded systems with limited feedback where users supported by a base station are selected by dynamic scheduling. First, we express the received signal-to-interference-plus-noise ratio (SINR) when multiple data streams are transmitted to the user, and derive a lower bound expression of the expected received SINR at each user. Then, based on this measure, each user determines its quantized CSI feedback information which maximizes the derived expected SINR, which comprises both the channel direction and the amplitude information. From simulations, we confirm that the proposed SINR-based channel quantization scheme achieves a significant sum rate gain over the conventional method in practical MU-MIMO systems.**

Index Terms: **Block diagonalization (BD), limited feedback, multiple-input multiple-output (MIMO) broadcast channel, signal-to-interference-plus-noise ratio (SINR).**

I. INTRODUCTION

Over the past decade, wireless communication systems have succeeded in significantly improving spectral efficiency through the aid of multiuser multiple-input multiple-output (MU-MIMO) techniques. Especially, transmission techniques for downlink broadcast channels (BCs) have been extensively studied in line with Costa's dirty paper coding (DPC) results [1]. From an information theoretic viewpoint, the capacity region of MIMO Gaussian BCs has been characterized based on the DPC approach [2], [3].

In parallel with theoretical developments, practical MU-MIMO schemes have been proposed to approach the maximum capacity gains of MIMO BCs. The most popular ones are zero-forcing beamforming (ZFBF) [4] and block diagonalization (BD) [5] which support multiple users at the same time and frequency resource by pre-canceling inter-user interference using a simple inverse or inverse-like operation. Also, in order to resolve an issue of power boost at the transmitter, regularized type ZFBF and BD precoders have been introduced [6], [7]. Assuming perfect knowledge of channel state information (CSI) at the transmitter, these schemes achieve fairly good sum rate performance gains when combined with efficient scheduling algorithms [4], [8].

However, in most practical cellular systems, especially with the frequency division duplex (FDD) mode, it is a great burden to obtain the accurate CSI of users over the whole operating bandwidth at the base station (BS). In practice, the CSI is quantized based on a predetermined codebook and only the index of the selected codeword is fed back to the BS. In [9] and [10], the performance of a ZFBF based on limited feedback with or without scheduling was analyzed for vector channels. For the MIMO case, it was shown that receive combining provides a large gain by reducing the amount of channel quantization errors [11], [12]. Although the channel feedback strategies for the BD was also discussed in [13], issues regarding the user scheduling were not taken into account.

Recently, a channel quantization technique has been presented in [14] for the FDD cellular systems where the BD precoding and user scheduling are employed to support a large number of users in a cell. In order to fully utilize multiuser diversity gains, the authors proposed a useful metric for channel quality measurements based on the derivation of the expected received signal-to-interference-plus-noise ratio (SINR) at each user. Based on this prior work, in this paper, we generalize the channel quality indicator (CQI) feedback scheme provided in [14] to systems with minimum mean-square error (MMSE) receivers and an arbitrary number of data streams (not larger than the number of receive antennas) per each user.

Also, we discuss how to choose the best codeword index in the codebook together with the CQI based on the proposed expected SINR. Then, an analysis for the outage probability of the proposed expected SINR metric is newly provided. From simulations, we confirm that the proposed SINR-based channel quantization scheme achieves a significant sum rate performance gain over the conventional feedback method in practical MU-MIMO systems.

The remainder of this paper is organized as follows. In Section II, we introduce the system model with limited CSI feedback and review the BD in Section III. Section IV describes the

Manuscript received February 23, 2013; approved for publication by Kim, Il-Min, Division I Editor, December 16, 2013.

This work was supported in part by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0017909), and in part by the IT R&D program of MSIP/KEIT (10044933, Development of Multi-tier Multi-RAT Cooperation Technology for Local Access Efficiency Improvement).

S. -H. Moon was with the School of Electrical Eng., Korea University, Seoul, Korea. He is now with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, email: sh.moon@etri.re.kr.

S. -R. Lee and I. Lee are with the School of Electrical Eng., Korea University, Seoul, Korea, email: {sangrim78, inkyu}@korea.ac.kr.

J. -S. Kim is with Bell Labs Seoul, Seoul, Korea, email: jin_sung.kim@alcatel-lucent.com.

Digital object identifier 10.1109/JCN.2014.000003

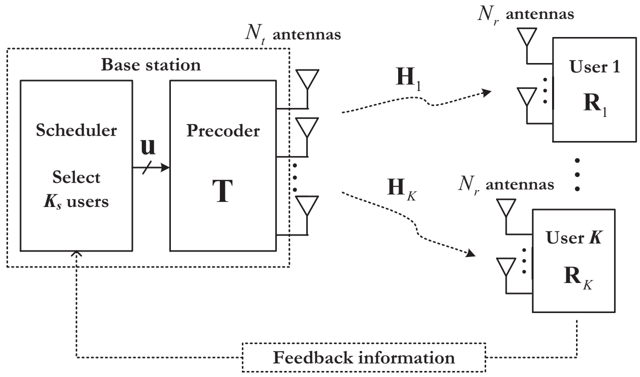


Fig. 1. Block diagram of downlink multiuser MIMO systems with limited feedback.

proposed limited feedback scheme for the BD based on the estimation of the received SINR, and the outage probability analysis is presented in Section V. In Section VI, the sum rate performance is evaluated through simulations. Finally, the paper ends with conclusions in Section VII.

Throughout this paper, we use the following notations. Normal letters represent scalar quantities, bold face lowercase letters indicate vectors, and boldface uppercase letters designate matrices. The superscripts $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and Hermitian transpose, respectively. The two-norm of a vector is represented by $\|\cdot\|$ and the Frobenius norm of a matrix is denoted by $\|\cdot\|_F$. The determinant and the trace of a matrix are given by $\det(\cdot)$ and Tr$(\cdot)$, respectively, and $\mathbf{I}_d$ indicates an identity matrix of size d.

II. SYSTEM MODEL

We consider a MIMO BC where K users exist in a cell. We assume that the BS has N_t transmit antennas and each user is equipped with N_r receive antennas as shown in Fig. 1. In each time slot, the BS performs scheduling to select K_s users among K users and transmits N_s data streams equally to all the selected K_s users, where K_s is limited to $K_s \le N_t/N_s$ such that the condition for orthogonal transmission among users is satisfied [15].

First, we define the data symbol vector $\mathbf{u}_k \in \mathbb{C}^{N_s\times 1}$ and the precoding matrix $\mathbf{T}_k \in \mathbb{C}^{N_t\times N_s}$ for the kth user ($k = 1,\cdots,K_s$). Then, by denoting $\mathbf{T} = [\mathbf{T}_1\cdots\mathbf{T}_{K_s}]$ and $\mathbf{u} = [\mathbf{u}_1^T\cdots\mathbf{u}_{K_s}^T]^T$, the precoded signal vector $\mathbf{x} \in \mathbb{C}^{N_t\times 1}$ is given by $\mathbf{x} = \mathbf{T}\mathbf{u} = \sum_{k=1}^{K_s}\mathbf{T}_k\mathbf{u}_k$. Here, we assume that $\mathbf{x}$ satisfies the total transmit power constraint Tr$(\mathbb{E}[\mathbf{x}\mathbf{x}^H]) \le P$, where P represents the total transmit power at the BS. Assuming that all symbols of $\mathbf{u}$ are independently generated with unit variance, $\mathbf{T}$ is constrained to satisfy Tr$(\mathbf{T}\mathbf{T}^H) \le P$.

Then, the received signal vector $\mathbf{y}_k \in \mathbb{C}^{N_r\times 1}$ for the kth user is determined as

$$\mathbf{y}_k = \mathbf{H}_k^H\mathbf{T}_k\mathbf{u}_k + \sum_{j=1,j\neq k}^{K_s}\mathbf{H}_k^H\mathbf{T}_j\mathbf{u}_j + \mathbf{n}_k \quad (1)$$

where $\mathbf{H}_k \in \mathbb{C}^{N_t\times N_r}$ is the channel matrix whose entries are independent and identically distributed (i.i.d.) complex Gaussian with $\mathcal{CN}(0,1)$ and $\mathbf{n}_k \in \mathbb{C}^{N_r}$ denotes the additive white Gaussian noise (AWGN) vector with zero mean and covariance matrix $\mathbb{E}[\mathbf{n}_k\mathbf{n}_k^H] = \mathbf{I}_{N_r}$. It is assumed that each user has perfect knowledge of its own channel $\mathbf{H}_k$, which has full column rank.

At the receiver, the kth user estimates the transmitted signal $\hat{\mathbf{u}}_k \in \mathbb{C}^{N_s\times 1}$ from the receive filter output as $\hat{\mathbf{u}}_k = \mathbf{R}_k^H\mathbf{y}_k$, where $\mathbf{R}_k \in \mathbb{C}^{N_r\times N_s}$ indicates the receive combining filter at user k. We adopt the linear MMSE receiver for $\mathbf{R}_k$

$$\mathbf{R}_k = \left(\mathbf{I}_{N_r} + \rho\sum_{j=1,j\neq k}^{K_s}\mathbf{H}_k^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{H}_k\right)^{-1}\mathbf{H}_k^H\mathbf{T}_k \quad (2)$$

under the assumption that $\{\mathbf{H}_k^H\mathbf{T}_j\}_{j=1}^{K_s}$ can be estimated at the receiver using dedicated pilot symbols where $\rho = P/N_t$ denotes the transmit power applied to each symbol.

Then, the kth user's received SINR corresponds to the desired signal vector $\mathbf{u}_k$ is defined from (1) as

$$\text{SINR}_k = \frac{\rho\left\|\mathbf{H}_{e,k}^H\mathbf{T}_k\right\|_F^2}{\left\|\mathbf{R}_k\right\|_F^2 + \sum_{j=1,j\neq k}^{K_s}\rho\left\|\mathbf{H}_{e,k}^H\mathbf{T}_j\right\|_F^2} \quad (3)$$

where $\mathbf{H}_{e,k} = \mathbf{H}_k\mathbf{R}_k \in \mathbb{C}^{N_t\times N_s}$ represents the effective channel combined with $\mathbf{R}_k$ and $\mathbf{T}_k$ satisfies Tr$(\mathbf{T}_k\mathbf{T}_k^H) = N_s$. The Frobenius norm expression in (3) is employed since each data signal $\mathbf{u}_k$ may consist of more than one symbol ($N_s > 1$) [16], [17], which reduces to the conventional SINR definition for $N_s = 1$.

Since the feedback link is bandwidth-constrained, users perform quantization on their effective CSI $\mathbf{H}_{e,k}$ before feeding back to the BS. The quantized CSI is composed of channel directional information (CDI) and CQI, each of which are denoted by $\hat{\mathbf{H}}_{e,k} \in \mathbb{C}^{N_t\times N_s}$ and γ_k. The CDI accounts for the spatial direction of the MIMO channel, i.e., the subspace spanned by the column vectors of $\mathbf{H}_{e,k}$, and the CQI indicates the level of the fading amplitudes. The most well-known and widely used approach to handle the CDI is the use of a codebook, which is an application of the quantization problem on the Grassmann manifold [18], denoted as $\mathcal{G}(N_t, N_s)$[1]. We consider a codebook $\mathcal{C}$ composed of 2^B codeword matrices $\mathbf{W}_1,\cdots,\mathbf{W}_{2^B} \in \mathbb{C}^{N_t\times N_s}$ whose column vectors are unit norm and orthogonal to each other. Due to its simplicity and analytical tractability, we employ random vector quantization (RVQ) for $\mathcal{C}$ where $\mathbf{W}_1,\cdots,\mathbf{W}_{2^B}$ are chosen independently and isotropically over $\mathbb{C}^{N_t}$ [19]. On the other hand, the type of the CQI is relatively not well specified, especially for the case of multi-stream transmission ($N_s > 1$) where the effective channels $\mathbf{H}_{e,k}$ are matrices. This issue will be investigated in detail in Section IV.

III. REVIEW OF BLOCK DIAGONALIZATION

The BD transforms a multiuser MIMO channel into parallel single-user MIMO channels by suppressing inter-user interference [5]. Many papers regarding the BD assume that users have

[1] The column space of a matrix $\mathbf{F} \in \mathbb{C}^{N_t\times N_s}$ is contained in $\mathcal{G}(N_t, N_s)$.

multiple receive antennas ($N_r > 1$) and each selected user receives full $N_s = N_r$ data streams. In this work, we consider a more general setting where an arbitrary number of N_s ($\leq N_r$) is supported for each user by adopting the receive combining filter $\mathbf{R}_k$. If N_s becomes one, the precoder reduces to the ZF beamforming.

In the BD procedure, we first seek a precoding matrix which eliminates inter-user interference, and then the channels for each user are decoupled into N_s parallel subchannels to achieve the maximum information rate. To suppress the interference term $\sum_{j\neq k} \mathbf{H}_{e,k}^H \mathbf{T}_j \mathbf{u}_j$ in (1) after the receive combining, the precoding matrix $\mathbf{T}_k$ should satisfy

$$\mathbf{H}_{e,j}^H \mathbf{T}_k = \mathbf{0} \text{ for all } k \neq j. \tag{4}$$

In other words, $\mathbf{T}_k$ should lie in the left nullspace of $\tilde{\mathbf{H}}_{e,k}$ where $\tilde{\mathbf{H}}_{e,k}$ is defined as

$$\tilde{\mathbf{H}}_{e,k} = [\mathbf{H}_{e,1} \cdots \mathbf{H}_{e,k-1} \;\; \mathbf{H}_{e,k+1} \cdots \mathbf{H}_{e,K_s}] \,.$$

Denote $\tilde{\mathbf{V}}_k^{(0)} \in \mathbb{C}^{N_t \times (N_t - L_k)}$ as the matrix which consists of orthonormal basis vectors of the left nullspace of $\tilde{\mathbf{H}}_{e,k}$ where $L_k = \text{rank}(\tilde{\mathbf{H}}_{e,k})$. Then, precoding with $\tilde{\mathbf{V}}_k^{(0)}$ can nullify the interference as $\mathbf{H}_{e,j}^H \tilde{\mathbf{V}}_k^{(0)} = \mathbf{0}$ for $j \neq k$ and forms the kth user's non-interfering block channel $\mathbf{H}_{e,k}^H \tilde{\mathbf{V}}_k^{(0)}$.

In order to decouple this block channel into N_s parallel subchannels, singular value decomposition (SVD) is performed on $\mathbf{H}_{e,k}^H \tilde{\mathbf{V}}_k^{(0)}$ as $\mathbf{H}_{e,k}^H \tilde{\mathbf{V}}_k^{(0)} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H$. Then, the precoder $\tilde{\mathbf{T}}_k$ and the receiver $\tilde{\mathbf{R}}_k \in \mathbb{C}^{N_r \times N_s}$ are determined as $\tilde{\mathbf{T}}_k = \tilde{\mathbf{V}}_k^{(0)} \mathbf{V}_k$ and $\tilde{\mathbf{R}}_k = \mathbf{U}_k$, respectively. By assuming Gaussian codebook and combining with $N_s \times N_s$ diagonal power loading matrices $\mathbf{\Phi}_k$ for $k = 1, \cdots, K_s$, the sum rate for the BD can be written as

$$R_{\text{BD}} = \max_{\mathbf{\Phi}_k} \sum_{k=1}^{K_s} \log\det\left(\mathbf{I}_{N_s} + \mathbf{\Sigma}_k^2 \mathbf{\Phi}_k\right)$$

subject to $\sum_{k=1}^{K_s} \text{Tr}(\mathbf{\Phi}_k) \leq P$. This is a well-known convex problem and the optimal solution $\mathbf{\Phi}_k$ can be obtained using the water-filling method. Finally, we have the precoding matrix $\mathbf{T}$ as $\mathbf{T} = [\tilde{\mathbf{T}}_1 \cdots \tilde{\mathbf{T}}_{K_s}] \mathbf{\Phi}^{\frac{1}{2}}$ where $\mathbf{\Phi} = \text{diag}(\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_{K_s})$.

IV. SINR-BASED CHANNEL QUANTIZATION FOR BLOCK DIAGONALIZATION

When true channels are not perfectly known at the BS, the null constraint (4) is broken since the BD chooses the precoder orthogonal to the quantized channels, not the actual channels. Then, there remain residual interference signals among users, which degrades the sum rate performance especially in the high signal-to-noise ratio (SNR) regime [9]. Therefore, it is important to design the CDI and the CQI carefully in order to minimize the inter-user interference. In this section, motivated by this, we propose an efficient channel quantization strategy for the BD which decides a proper CDI and CQI feedback based on the received SINR defined in (3). For simplicity, we assume uniform power allocation across data streams, i.e., $\mathbf{\Phi} = \mathbf{I}_{K_s N_s}$, which is known to be asymptotically optimal at the high SNR region [20].

First, we recall a useful matrix decomposition provided in [13] for our effective channel $\mathbf{H}_{e,k}$. Let the eigenvalue decomposition (EVD) of $\mathbf{H}_{e,k}\mathbf{H}_{e,k}^H$ be given as

$$\mathbf{H}_{e,k}\mathbf{H}_{e,k}^H = \overline{\mathbf{H}}_{e,k} \mathbf{\Lambda}_{e,k} \overline{\mathbf{H}}_{e,k}^H \tag{5}$$

where $\mathbf{\Lambda}_{e,k} = \text{diag}(\lambda_{k,1}, \cdots, \lambda_{k,N_s})$ consists of N_s non-zero eigenvalues of $\mathbf{H}_{e,k}\mathbf{H}_{e,k}^H$ and $\overline{\mathbf{H}}_{e,k} \in \mathbb{C}^{N_t \times N_s}$ is the matrix of the corresponding orthonormal basis vectors which span the column space of $\mathbf{H}_{e,k}$. Then, it was shown in [13] that $\overline{\mathbf{H}}_{e,k}$ can be decomposed as

$$\overline{\mathbf{H}}_{e,k} = \hat{\mathbf{H}}_{e,k} \mathbf{A}_k \mathbf{B}_k + \mathbf{S}_k \mathbf{C}_k \tag{6}$$

where $\mathbf{S}_k \in \mathbb{C}^{N_t \times N_s}$ indicates an orthonormal basis matrix which spans an isotropically distributed N_s-dimensional plane in the $(N_t - N_s)$-dimensional left nullspace of $\hat{\mathbf{H}}_{e,k}$, $\mathbf{C}_k \in \mathbb{C}^{N_s \times N_s}$ denotes an upper triangular matrix with positive diagonal entries satisfying $\text{Tr}(\mathbf{C}_k^H \mathbf{C}_k) = d^2(\overline{\mathbf{H}}_{e,k}, \hat{\mathbf{H}}_{e,k})$ with $d(\cdot,\cdot)$ being a distance function on $\mathcal{G}(N_t, N_s)$, $\mathbf{A}_k \in \mathbb{C}^{N_s \times N_s}$ is unitary and $\mathbf{B}_k \in \mathbb{C}^{N_s \times N_s}$ is upper triangular which satisfies $\mathbf{B}_k^H \mathbf{B}_k = \mathbf{I}_{N_s} - \mathbf{C}_k^H \mathbf{C}_k$. Here, $\mathbf{S}_k$ and $\mathbf{C}_k$ are independent.

A. Conventional Feedback Scheme

For the CDI, a well-known quantization method is to choose a codeword which is the closest to the column space of $\mathbf{H}_k$, or the effective channel $\mathbf{H}_{e,k} = \mathbf{H}_k \mathbf{R}_k$ in our case, which can be expressed as [13]

$$\hat{\mathbf{H}}_{e,k} = \arg\min_{\mathbf{W}_i \in \mathcal{C}} d(\overline{\mathbf{H}}_{e,k}, \mathbf{W}_i). \tag{7}$$

In [13], the authors proposed to adopt the chordal distance for $d(\cdot,\cdot)$, which is one of common distance metrics between two subspaces. When the two subspaces are represented by $N_t \times N_s$ complex matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ whose columns form an orthonormal basis in $\mathcal{G}(N_t, N_s)$, the chordal distance $d\left(\mathbf{F}_1, \mathbf{F}_2\right)$ is defined by

$$\begin{aligned} d\left(\mathbf{F}_1, \mathbf{F}_2\right) &= \frac{1}{\sqrt{2}} \|\mathbf{F}_1\mathbf{F}_1^H - \mathbf{F}_2\mathbf{F}_2^H\|_F \\ &= \sqrt{N_s - \text{Tr}\big(\mathbf{F}_1^H \mathbf{F}_2 \mathbf{F}_2^H \mathbf{F}_1\big)}. \end{aligned} \tag{8}$$

On the other hand, to the best of authors' knowledge, no proper definition on the CQI and its quantization exists for the BD precoded systems with $N_s \geq 2$. Even the information size of the CQI should be determined in this case. In this paper, we consider the CQI as a single real value, which we denote as γ_k, in terms of minimizing the feedback overhead. In that case, one natural choice for γ_k can be the Frobenius norm of the channel matrix given by

$$\gamma_k^{\text{norm}} = \|\mathbf{H}_{e,k}\|_F^2 = \text{Tr}(\overline{\mathbf{H}}_{e,k} \mathbf{\Lambda}_{e,k} \overline{\mathbf{H}}_{e,k}^H) = \text{Tr}(\mathbf{\Lambda}_{e,k}) \tag{9}$$

which indicates the sum of the effective channel gains $\sum_{j=1}^{N_s} \lambda_{k,j}$ of user k. These channel gains are mainly achievable when the CDI feedback is ideal and the interference term in (1) is completely eliminated. Therefore, the metric (9) can be simply obtained, but may not be suitable with the quantized CDI. If (9) is adopted as the CQI feedback, the receive filter $\mathbf{R}_k$ can be computed as N_s dominant right singular vectors of $\mathbf{H}_k$ from the SVD operation.

B. Proposed SINR-Based Feedback Scheme

Now, we describe the proposed SINR-based channel quantization technique. Our original goal is to identify the CDI and the CQI which maximize the received SINR in (3). However, as pointed out in [14], the true values of the SINR cannot be estimated at the receiver, since users do not know the precoders $\{\mathbf{T}_k\}_{k=1}^{K_s}$ in the feedback stage which occurs before the user selection process. Therefore, we develop an expression of the expected SINR where the expectation is taken in terms of $\{\mathbf{T}_k\}_{k=1}^{K_s}$, which can be a useful performance measure in such MU-MIMO situations [10], [12].

By taking the expectation on (3) and applying Jensen's inequality with respect to its interference term, the expected SINR for the kth user, denoted by $\overline{\text{SINR}}_k$, is lower bounded as

$$\begin{aligned}\overline{\text{SINR}}_k &= \mathbb{E}_{\mathbf{T}_1,\cdots,\mathbf{T}_{K_s}}[\text{SINR}_k] \\ &\geq \mathbb{E}_{\mathbf{T}_k}\left[\frac{\rho\text{Tr}\left(\mathbf{H}_{e,k}^H\mathbf{T}_k\mathbf{T}_k^H\mathbf{H}_{e,k}\right)}{\|\mathbf{R}_k\|_F^2+\sum_{j\neq k}\rho\text{Tr}\left(\mathbb{E}_{\mathbf{T}_j}\left[\mathbf{H}_{e,k}^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{H}_{e,k}\right]\right)}\right]. \quad (10)\end{aligned}$$

When greedy type user scheduling is employed, it is assumed that K_s chosen users are nearly orthogonal to each other as K grows large enough [4]. In that case, the subspace of $\mathbf{T}_k$ is closely aligned to that of $\hat{\mathbf{H}}_{e,k}$ and is almost orthogonal to that of $\mathbf{S}_k$, which yields

$$\hat{\mathbf{H}}_{e,k}^H\mathbf{T}_k\mathbf{T}_k^H\hat{\mathbf{H}}_{e,k}\simeq\mathbf{I}_{N_s}\quad\text{and}\quad\mathbf{S}_k^H\mathbf{T}_k\simeq\mathbf{0}. \quad (11)$$

By substituting (5) and (6) and applying (11), the numerator in (10) can be approximated by

$$\begin{aligned}&\rho\text{Tr}\big(\mathbf{H}_{e,k}^H\mathbf{T}_k\mathbf{T}_k^H\mathbf{H}_{e,k}\big)\\ &\approx\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}(\hat{\mathbf{H}}_{e,k}\mathbf{A}_k\mathbf{B}_k)^H\mathbf{T}_k\mathbf{T}_k^H(\hat{\mathbf{H}}_{e,k}\mathbf{A}_k\mathbf{B}_k)\right)\\ &\approx\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}\mathbf{B}_k^H\mathbf{A}_k^H\mathbf{A}_k\mathbf{B}_k\right)\\ &=\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}(\mathbf{I}_{N_s}-\mathbf{C}_k^H\mathbf{C}_k)\right) \quad (12)\\ &=\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}\overline{\mathbf{H}}_{e,k}^H\hat{\mathbf{H}}_{e,k}\hat{\mathbf{H}}_{e,k}^H\overline{\mathbf{H}}_{e,k}\right) \quad (13)\end{aligned}$$

where (12) comes from the definitions of $\mathbf{A}_k$ and $\mathbf{B}_k$ and (13) follows from $\mathbf{C}_k^H\mathbf{C}_k=\mathbf{I}_{N_s}-\overline{\mathbf{H}}_{e,k}^H\hat{\mathbf{H}}_{e,k}\hat{\mathbf{H}}_{e,k}^H\overline{\mathbf{H}}_{e,k}$, which is a sufficient condition for $\mathbf{C}_k$ to satisfy its definition $\text{Tr}(\mathbf{C}_k^H\mathbf{C}_k)=d^2(\overline{\mathbf{H}}_{e,k},\hat{\mathbf{H}}_{e,k})$.

Next, we consider the interference term. By applying (5) and (6), the jth interference term of (10) is given as

$$\begin{aligned}&\rho\text{Tr}\big(\mathbb{E}\big[\mathbf{H}_{e,k}^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{H}_{e,k}\big]\big)\\ &\quad=\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}\mathbb{E}\big[\overline{\mathbf{H}}_{e,k}^H\mathbf{T}_j\mathbf{T}_j^H\overline{\mathbf{H}}_{e,k}\big]\right)\\ &\quad=\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}\mathbf{C}_k^H\mathbb{E}\big[\mathbf{S}_k^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{S}_k\big]\mathbf{C}_k\right) \quad (14)\\ &\quad=\frac{\rho N_r}{N_t-N_r}\text{Tr}\left(\mathbf{\Lambda}_{e,k}\mathbf{C}_k^H\mathbf{C}_k\right) \quad (15)\end{aligned}$$

where we have (14) since the BD makes $\hat{\mathbf{H}}_{e,k}^H\mathbf{T}_j=\mathbf{0}$ for $j\neq k$ and (15) comes from the fact that both $\mathbf{S}_k$ and $\mathbf{T}_j$ are isotropically distributed in the left nullspace of $\hat{\mathbf{H}}_{e,k}$ and independent to each other. Hence, $\mathbf{S}_k^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{S}_k$ is matrix variate beta distributed with parameters N_s and N_t-2N_s, which means $\mathbb{E}\big[\mathbf{S}_k^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{S}_k\big]=N_s/(N_t-N_s)$ [13]. Therefore, the total interference term in (10) becomes

$$\begin{aligned}&\sum_{j\neq k}\rho\text{Tr}\big(\mathbb{E}_{\mathbf{T}_j}\big[\mathbf{H}_{e,k}^H\mathbf{T}_j\mathbf{T}_j^H\mathbf{H}_{e,k}\big]\big)\\ &\quad=\rho\text{Tr}\left(\mathbf{\Lambda}_{e,k}\big(\mathbf{I}_{N_s}-\overline{\mathbf{H}}_{e,k}^H\hat{\mathbf{H}}_{e,k}\hat{\mathbf{H}}_{e,k}^H\overline{\mathbf{H}}_{e,k}\big)\right). \quad (16)\end{aligned}$$

Finally, by inserting (13) and (16) into (10), the lower bound of the expected SINR for the kth user can be expressed as

$$\begin{aligned}\overline{\text{SINR}}_k &\gtrsim \frac{\rho\|\mathbf{H}_{e,k}^H\hat{\mathbf{H}}_{e,k}\|_F^2}{\|\mathbf{R}_k\|_F^2+\rho\big(\|\mathbf{H}_{e,k}\|_F^2-\|\mathbf{H}_{e,k}^H\hat{\mathbf{H}}_{e,k}\|_F^2\big)} \quad (17)\\ &=\frac{\rho\text{Tr}(\mathbf{\Lambda}_{e,k})-\rho\text{Tr}(\mathbf{\Lambda}_{e,k}\mathbf{Z}_k)}{\|\mathbf{R}_k\|_F^2+\rho\text{Tr}(\mathbf{\Lambda}_{e,k}\mathbf{Z}_k)}\triangleq\gamma_k^{\text{SINR}} \quad (18)\end{aligned}$$

where $\mathbf{Z}_k$ denotes $\mathbf{Z}_k=\mathbf{C}_k^H\mathbf{C}_k$ and (18) comes from (12) and (15). The outer expectation in (10) vanishes since the signal term does not depend on $\mathbf{T}_k$ any more. It is straightforward to show that (17) or (18) approaches the true SINR in (3) when the CDIs of the selected users are fully orthogonal. This is because in this case, the precoding matrices $\mathbf{T}_k$ simply become the CDIs $\hat{\mathbf{H}}_{e,k}$ fed back by the selected users. More importantly, equation (17) is independent of $\{\mathbf{T}_k\}_{k=1}^{K_s}$ and is only a function of a given CDI $\hat{\mathbf{H}}_{e,k}$ and a receive filter $\mathbf{R}_k$.

Based on our expected SINR metric (17), we now explain how to obtain the CDI and CQI feedback.

First, we choose the CDI matrix which maximizes (17) according to

$$\hat{\mathbf{H}}_{e,k}=\arg\max_{\mathbf{W}_i\in\mathcal{C}}\frac{\rho\|\mathbf{R}_k^H\mathbf{H}_k^H\mathbf{W}_i\|_F^2}{\|\mathbf{R}_k\|_F^2+\rho\big(\|\mathbf{H}_k\mathbf{R}_k\|_F^2-\|\mathbf{R}_k^H\mathbf{H}_k^H\mathbf{W}_i\|_F^2\big)}. \quad (19)$$

Here, for computing $\mathbf{R}_k$, the MMSE receiver (2) is not available since $\{\mathbf{T}_k\}_{k=1}^{K_s}$ is not known yet as discussed before. On the other hand, we know that if K is large, the precoders $\{\mathbf{T}_k\}_{k=1}^{K_s}$ are almost orthogonal to each other, i.e., $\sum_{j=1}^{K_s}\mathbf{T}_j\mathbf{T}_j^H\simeq\mathbf{I}_{N_t}$, and then $\mathbf{T}_k$ can be simply chosen as the CDI. Accordingly, for a given codeword $\mathbf{W}_i$ in (19), we can alternatively compute the MMSE receiver $\mathbf{R}_k$ without the knowledge of $\{\mathbf{T}_k\}_{k=1}^{K_s}$ as

$$\mathbf{R}_k=\left(\mathbf{I}_{N_r}+\rho\mathbf{H}_k^H\big(\mathbf{I}_{N_t}-\mathbf{W}_i\mathbf{W}_i^H\big)\mathbf{H}_k^H\right)^{-1}\mathbf{H}_k^H\mathbf{W}_i. \quad (20)$$

Once the CDI $\hat{\mathbf{H}}_{e,k}$ and the corresponding $\mathbf{R}_k$ are determined by (19) and (20), respectively, the proposed CQI γ_k^{SINR} is directly obtained from (17).

Comparing with the conventional metric γ_k^{norm} in (9), the proposed CQI (17) is expected to closely reflect each user's actual channel quality, since it takes into account the effect of both the BD precoder and the MMSE receiver. Furthermore, it can be seen that the term $\|\mathbf{H}_{e,k}^H\hat{\mathbf{H}}_{e,k}\|_F^2 = \text{Tr}(\mathbf{\Lambda}_{e,k}\overline{\mathbf{H}}_{e,k}^H\hat{\mathbf{H}}_{e,k}\hat{\mathbf{H}}_{e,k}^H\overline{\mathbf{H}}_{e,k})$ in (17) gets larger when $\overline{\mathbf{H}}_{e,k}$ and $\hat{\mathbf{H}}_{e,k}$ are closely aligned. Therefore, γ_k^{SINR} also accounts for the quantization accuracy as well as the channel magnitude. Thus, we expect that the proposed scheme provides good performance also in the interference regime of high P.

Based on the effective channels $\{\gamma_k^{\text{SINR}}\hat{\mathbf{H}}_{e,k}\}_{k=1}^K$ constructed by the CDI and the CQI feedback from users, the BS selects the best K_s users by scheduling and performs the BD precoding as described in Section III. Then, after the transmission, each active user decodes its signal vector $\mathbf{u}_k$ by applying the MMSE receiver of (2) which is now available at the receiver.

V. ASYMPTOTIC ANALYSIS

In this section, we analyze the distribution of the proposed expected SINR-based CQI metric (18). The exact distribution of (18) depends on the Grassmannian packing problem related to $\mathbf{Z}_k$. However, this analysis is difficult because for finite N_t and N_s, a quantization bound of the subspace packing is not known for $N_s \geq 2$. Alternatively, we seek an asymptotic distribution where both $N_t = \tau N$ and $N_s = N$ simultaneously grow large with a fixed ratio $\tau = N_t/N_s$, which is a useful approach when the exact analysis is mathematically intractable [21], [22]. Later on, it will be shown that our analysis matches well even with the small number of antennas.

In the large antenna regime, γ_k^{SINR} reduces to

$$\lim_{N\to\infty}\gamma_k^{\text{SINR}} = \frac{\text{Tr}\big(\mathbf{\Lambda}_{e,k}\big)}{\text{Tr}\big(\mathbf{\Lambda}_{e,k}\mathbf{Z}_k\big)} - 1 \tag{21}$$

even if the transmit power level P is finite. We first look at the probability density function (PDF) of (21) and after that the outage probability will be evaluated. The central limit theorem is useful to sketch the context of the work here.

A. Approximate Density Function

Denote the two terms $\text{Tr}(\mathbf{\Lambda}_{e,k})$ and $\text{Tr}(\mathbf{\Lambda}_{e,k}\mathbf{Z}_k)$ in (21) by X and Y, respectively. Then, we present the following two lemmas.

Lemma 1: As N goes to infinity, X follows

$$X \sim \mathcal{N}\left(\mu_{\mathcal{X}}, \sigma_{\mathcal{X}}^{\in}\right) \tag{22}$$

where $\mu_X = \sigma_X^2 = \tau N^2$.

Proof: Since $\text{Tr}(\mathbf{\Lambda}_{e,k}) = \sum_{i=1}^{N_t}\sum_{j=1}^{N_s}|h_{ij}|^2$ is a sum of independent and identically distributed random variables, the central limit theorem says that X is asymptotically normal [23]. The mean and the variance of $\text{Tr}(\mathbf{\Lambda}_{e,k})$, a sum of unordered eigenvalues of the complex Wishart matrix $\mathbf{H}_{e,k}\mathbf{H}_{e,k}^H$, are commonly known, for example, in [21]. □

Lemma 2: When the random codeword selection is assumed for the decision of $\hat{\mathbf{H}}_{e,k}$, Y is given approximately with large N by

$$Y \sim \mathcal{N}\left(\mu_{\mathcal{Y}}, \sigma_{\mathcal{Y}}^{\in}\right) \tag{23}$$

where we have $\mu_Y = (\tau-1)N^2$ and $\sigma_Y^2 \approx \xi N^2$, and ξ is defined as $\xi \triangleq (\tau^2+1)(\tau-1)/\tau^2$.

Proof: See Appendix. □

In deriving Lemma 2, we have assumed that an arbitrary codeword is selected in the codebook $\mathcal{C}$ for the ease of analysis, which corresponds to the worst case of channel quantization. This provides a lower bound on the actual system performance. If N_t is sufficiently large compared to the number of bits B, our results hold with practical codebook selection methods such as (7). Under the same assumption, the cross-correlation of X and Y can be computed as $\mathbb{E}[XY] = (\tau-1)(\tau N^2+1)N^2$ by adopting the same approach used in the Appendix. Then, from the above two lemmas, the correlation coefficient ρ_{XY} between X and Y is obtained as

$$\rho_{XY} = \frac{\mathbb{E}[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y} \approx \frac{\tau-1}{\sqrt{\tau\xi}}. \tag{24}$$

Now, we have two Gaussian random variables X and Y with the correlation coefficient ρ_{XY}. In this case, the distribution of $V = X/Y$, called as the Gaussian ratio distribution, is known as a closed-form in [24]. We use this result to get the approximate PDF of (21) in the limit of large N, which yields

$$\begin{aligned} f_\gamma(x) & \\ \approx \quad & \frac{\sqrt{\omega}}{\pi g(x+1)}e^{-\frac{\tau N^2}{2}} \\ - \; & \frac{\omega N(x+1)}{\pi g(x+1)^{1.5}}e^{-\frac{N^2(1-(\tau-1)x)^2}{2g(x+1)}}\,\Theta\left(-\frac{\sqrt{\omega}N(x+1)}{g(x+1)^{0.5}}\right) \end{aligned} \tag{25}$$

where we have

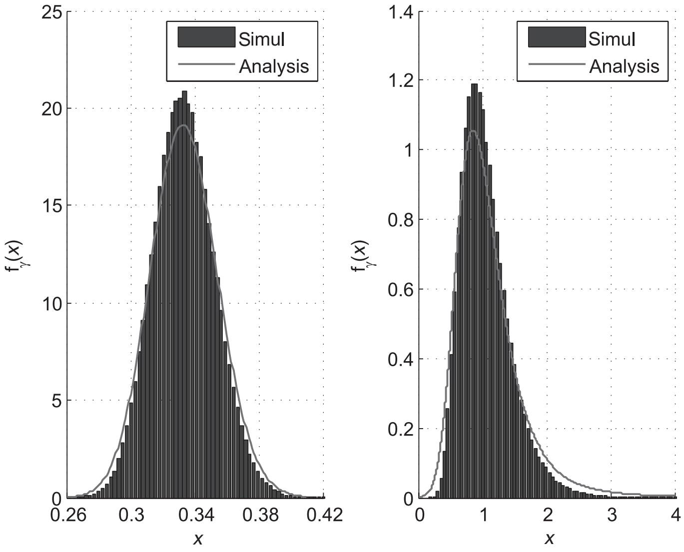
$$\begin{aligned} g(x) &= \xi x^2 - 2(\tau-1)x + \tau, \quad \omega = \tau - \tau^{-1}, \\ \Theta(y) &= \int_0^y e^{-u^2/2}du. \end{aligned}$$

In Fig. 2, we compare the asymptotic approximation for $f_\gamma(x)$ with the simulated results where $100,000$ Monte Carlo runs were carried out. From this figure, it can be seen that the analytical and empirical results match closely even for the small number of antennas $N_t = 8$ and $N_s = 4$. Thus, the validity of the asymptotic analysis can be confirmed.

B. Outage Probability

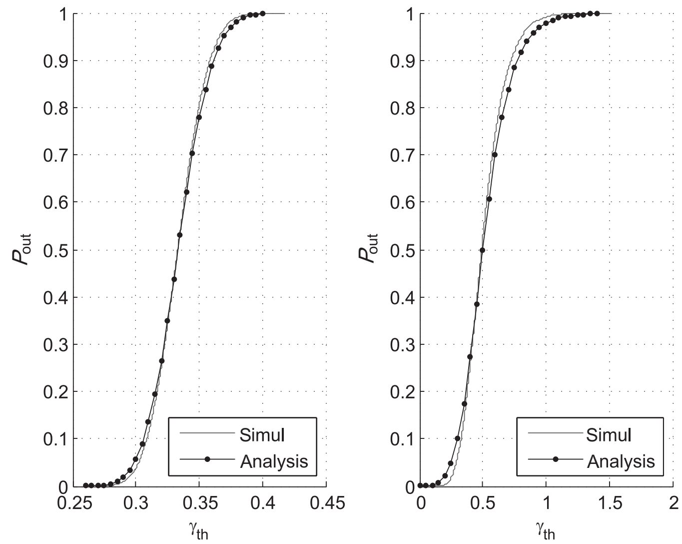
From (25), the CDF of γ_k^{SINR} is found by direct calculation to be [25]

$$\begin{aligned} F_\gamma(x) \approx L & \left\{\frac{N-(\tau-1)Nx}{g(x+1)^{0.5}}, -\frac{(\tau-1)N}{\sqrt{\omega}}; \frac{\xi x + \xi - \tau + 1}{\sqrt{\omega}g(x+1)^{0.5}}\right\} \\ & + L\left\{\frac{(\tau-1)Nx - N}{g(x+1)^{0.5}}, \frac{(\tau-1)N}{\sqrt{\omega}}; \frac{\xi x + \xi - \tau + 1}{\sqrt{\omega}g(x+1)^{0.5}}\right\} \end{aligned} \tag{26}$$



(a) $N_t = 80$ and $N_s = 20$ (b) $N_t = 8$ and $N_s = 4$

Fig. 2. Comparison of the simulated and asymptotic theoretical PDF of γ_k^{SINR}.



(a) $N_t = 80$ and $N_s = 20$ (b) $N_t = 12$ and $N_s = 4$

Fig. 3. Outage probability of γ_k^{SINR}.

where $L\{m, n; r\}$ is the standard bivariate normal integral given by [26]

$$L\{m, n; r\} = \frac{1}{2\pi\sqrt{1-r^2}} \int_m^\infty \int_n^\infty e^{-\frac{x^2+y^2-2rxy}{2(1-r^2)}} dxdy.$$

The expression (26) is quite complicated. However, in the asymptotic regime of $N \to \infty$, (26) is simplified as [25]

$$F_\gamma(x) \xrightarrow{\text{a.s.}} 1 - Q\left(\frac{(\tau-1)Nx - N}{g(x+1)^{0.5}}\right)$$

where $\xrightarrow{\text{a.s.}}$ denotes the asymptotic convergence and $Q(\cdot)$ is the Gaussian Q-function.

Therefore, the outage probability that γ_k^{SINR} is below a certain

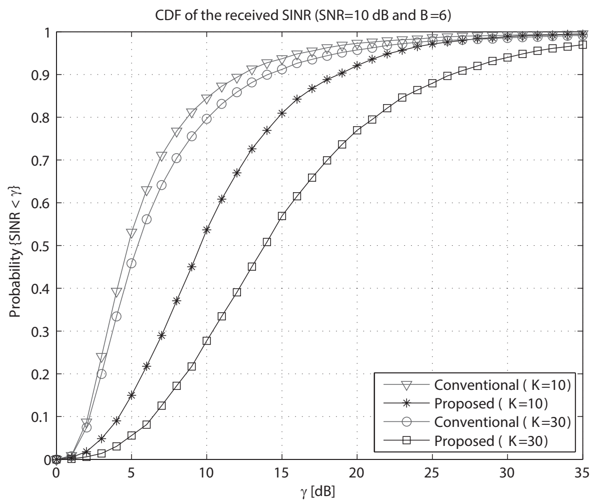


Fig. 4. CDF of the received SINR for $N_t = 4$ and $N_r = 2$ with $K = 10$ and 30 users.

SINR threshold γ_{th}, $P_{\text{out}} \triangleq P\{\gamma_k^{\text{SINR}} \leq \gamma_{\text{th}}\}$ can be evaluated as

$$P_{\text{out}} \approx 1 - Q\left(\frac{(\tau-1)N\gamma_{\text{th}} - N}{g(\gamma_{\text{th}}+1)^{0.5}}\right). \quad (27)$$

For a given desired outage probability P_{out}, γ_{th} can be obtained by solving equation (27). Fig. 3 plots the outage probability (27) for different numbers of antennas. Also the simulated cumulative distributions are displayed for comparison. By comparing the results, we note that our asymptotic approximation is accurate even for the moderate number of N_t and N_s as well as the large system case.

VI. SIMULATION RESULTS

In this section, we investigate the sum rate performance of multiuser MIMO downlink systems employing the BD precoding and the proposed limited feedback scheme. For all simulations, we use spatially uncorrelated MIMO Rayleigh fading channels which are independently generated for each transmission. We assume that each user adopts different codebooks to prevent a case where more than two users choose the same codeword. For user scheduling, we apply the SUS algorithm proposed in [4] by slightly modifying the operations so as to be applicable to the multi-stream case of $N_s > 1$ with the threshold parameter $\alpha = 0.5$ and 0.4 for $N_t = 4$ and 6, respectively.

In Fig. 4, we plot the cumulative distribution function (CDF) of the average received SINR for the scheduled users under two different feedback schemes for $(N_t, N_r) = (4, 2)$ and $N_s = 2$ at the SNR of 10 dB. The average SINR is defined as $1/K_s \sum_{k=1}^{K_s} \text{SINR}_k$ from (3). In this plot, we observe that the proposed SINR feedback outperforms the magnitude feedback in (9) for both $K = 10$ and 30 users. We emphasize that when the magnitude feedback scheme is applied, a multiuser diversity gain is marginal even with a large number of users, as it fails to exploit multiuser diversity. In contrast, our feedback scheme significantly improves the SINR with increasing users,

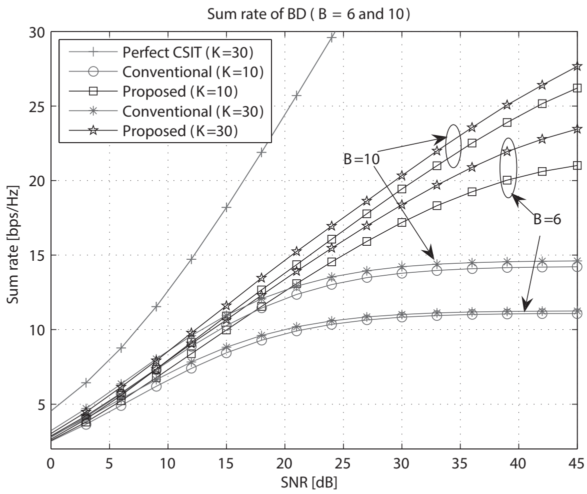


Fig. 5. Sum rates for $(N_t, N_r) = (4, 2)$ and $N_s = 2$ with $K = 10$ and 30 users.

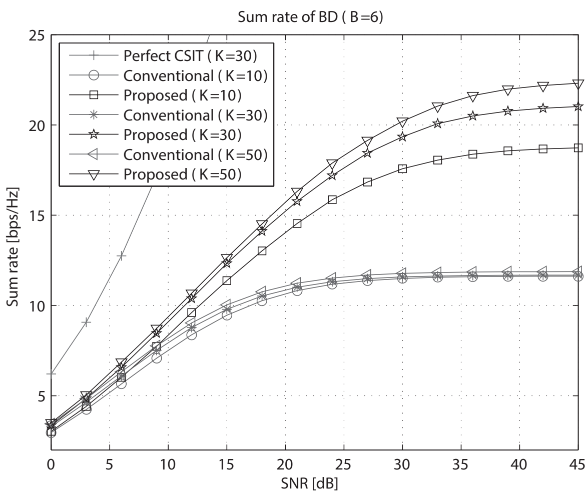


Fig. 6. Sum rates for $(N_t, N_r) = (6, 3)$ and $N_s = 2$ with $K = 10, 30$ and 50 users.

which implies that the proposed expected SINR metric (17) is effective in utilizing the multiuser diversity gain.

Figs. 5 and 6 show the sum rate comparison between the conventional and the proposed channel quantization method for $(N_t, N_r) = (4, 2)$ and $(N_t, N_r) = (6, 3)$, respectively, with $N_s = 2$. Different from the perfect CSI case[2], the sum rates with limited feedback are saturated as $P \to \infty$ due to unavoidable residual interference which comes from the CDI quantization error. We confirm that as expected from Fig. 4, our SINR-based feedback strategy outperforms the conventional scheme with the practical codebook size $B = 6$ and 10, especially at high SNR. The sum rate gains of the proposed SINR feedback scheme are roughly 100% over the magnitude feedback method at the high SNR region for $K = 10, 30,$ and 50 users. This is because the magnitude feedback does not reflect the amount of the quantization error which dominates the performance of multiuser systems for large P. Note that a scheduling gain of the proposed scheme becomes larger as K increases.

[2]For the antenna setting of Fig. 6, the best known technique with full CSI is the so-called coordinated beamforming [15]. However, in this simulation, we have simply performed the BD at the transmitter with the feedback of dominant channel eigenmodes.

VII. CONCLUSIONS

In this paper, we have investigated the issue of limited feedback for BD precoded MIMO BCs along with the user selection process. In order to appreciate a multiuser diversity gain, we have proposed an efficient CDI and CQI quantization technique based on a derivation of the received SINR for each user by evaluating the average received SINR for each user. Also, we have performed an asymptotic analysis on the distribution of the proposed CQI metric. From simulations, we have verified that by reflecting both the channel gain and the quantization error, the proposed SINR-based channel quantization outperforms the conventional feedback in practical BD system environments.

APPENDIX

A. *Proof of Lemma 2*

First, we examine the asymptotic behavior of the distribution of Y under the assumption of the random codeword selection. In that case, $\mathbf{Z}_k$ follows the matrix variate beta distribution $\beta(N_t - N_r, N_r)$ [13]. Then, each diagonal entry of $\mathbf{Z}_k$, $z_{k,i}$ $(i = 1, \cdots, N_r)$ is also beta distributed with parameters $N_t - N_r$ and N_r [27]. Considering that the unordered eigenvalues $\lambda_{k,1}, \cdots, \lambda_{k,N_r}$ are also homogeneous in distribution, we confirm that $\{\lambda_{k,i} z_{k,i}\}_{i=1}^{N_r}$ are identically distributed.

The correlation coefficient ρ_{ij} between $\lambda_{k,i} z_{k,i}$ and $\lambda_{k,j} z_{k,j}$ for $i \neq j$ can be shown to be approximated by

$$\rho_{ij} \approx -\frac{(\tau - 1)(\tau N + 1)}{\tau((\tau - 1)N + 2)N^2}. \quad (28)$$

For brevity, the derivations are omitted here. We can easily see that (28) asymptotically converges to zero with $N \to \infty$, which means that $\lambda_{k,1} z_{k,1}, \cdots, \lambda_{k,N_r} z_{k,N_r}$ are uncorrelated. Therefore, the central limit theorem proves that $Y = \sum_{i=1}^{N_r} \lambda_{k,i} z_{k,i}$ follows an asymptotically normal distribution.

In the following, we evaluate the mean and the variance of Y, by using the fact that the first and the second moment of the beta random variable $z_{k,i}$ are given by $\mathbb{E}[z_{k,i}] = (N_t - N_r)/N_t$ and $\mathbb{E}[z_{k,i}^2] = (N_t - N_r)(N_t - N_r + 1)/\{N_t(N_t + 1)\}$, respectively. At first, we obtain μ_Y as

$$\begin{aligned} \mu_Y &= \mathbb{E}\left[\sum_{i=1}^{N_r} \lambda_{k,i} z_{k,i}\right] \\ &= N_t \sum_{i=1}^{N_r} \mathbb{E}\left[z_{k,i}\right] = N_r(N_t - N_r) = (\tau - 1)N^2 \quad (29) \end{aligned}$$

where the second equality holds since $\lambda_{k,i}$ and $z_{k,i}$ are independent.

Next, the second moment of Y can be expressed as

$$\begin{aligned}\mathbb{E}[Y^2] &= \mathbb{E}\left[\sum_{i=1}^{N_r}\lambda_{k,i}z_{k,i}\sum_{j=1}^{N_r}\lambda_{k,j}z_{k,j}\right] \\ &= \sum_{i=1}^{N_r}\mathbb{E}[\lambda_{k,i}^2]\mathbb{E}[z_{k,i}^2] \\ &+ \sum_{i=1}^{N_r}\sum_{j=1,j\neq i}^{N_r}\mathbb{E}[\lambda_{k,i}\lambda_{k,j}]\mathbb{E}[z_{k,i}z_{k,j}]. \end{aligned} \tag{30}$$

Here, $\mathbb{E}[\lambda_{k,i}^2]$, $\mathbb{E}[z_{k,i}^2]$, and $\mathbb{E}[\lambda_{k,i}\lambda_{k,j}]$ in (30) are well known. Also, noting that

$$\begin{aligned}\mathrm{cov}(z_{k,i}, z_{k,j}) &= \mathbb{E}[z_{k,i}z_{k,j}] - \mathbb{E}[z_{k,i}]\mathbb{E}[z_{k,j}] \\ &= \frac{N_r(N_t - N_r)}{N_t^2(N_t+1)} \xrightarrow{\text{a.s.}} 0\end{aligned}$$

we can replace the last term $\mathbb{E}[z_{k,i}z_{k,j}]$ by $\mathbb{E}[z_{k,i}]\mathbb{E}[z_{k,j}]$. Then, after some calculations, (30) reduces to a function of N and τ as

$$\begin{aligned}\mathbb{E}[Y^2] &= \frac{(\tau^2-1)((\tau-1)N+1)N^3}{\tau N+1} \\ &+ \frac{(\tau-1)^2(\tau N-1)(N-1)N^2}{\tau}.\end{aligned} \tag{31}$$

Now, from (29) and (31), the variance σ_Y^2 is given as

$$\begin{aligned}\sigma_Y^2 &= \mathbb{E}[Y^2] - \mu_Y^2 \\ &= N^2\left(\frac{(\tau^2+1)(\tau-1)}{\tau^2} - \frac{\tau^2-1}{\tau^2(\tau N+1)}\right) \\ &\approx \frac{(\tau^2+1)(\tau-1)}{\tau^2}N^2.\end{aligned} \tag{32}$$

Here, the second term of (32) is neglected for large N.

REFERENCES

[1] Max H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983.
[2] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, July 2003.
[3] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sept. 2006.
[4] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 528–541, Mar. 2006.
[5] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 461–471, Feb. 2004.
[6] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication -Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, pp. 195–202, Jan. 2005.
[7] H. Sung, S. -R. Lee, and I. Lee, "Generalized channel inversion methods for multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 57, pp. 3489–3499, Nov. 2009.
[8] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, pp. 3658–3663, Sept. 2006.
[9] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.
[10] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 1478–1491, Sept. 2007.
[11] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3834–3844, Oct. 2008.
[12] M. Trivellato, F. Boccardi, and H. Huang, "On transceiver design and channel quantization for downlink multiuser MIMO systems with limited feedback," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1494–1504, Oct. 2008.
[13] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1473–1482, Oct. 2008.
[14] S. -H. Moon, J. -S. Kim, and I. Lee, "Limited feedback design for block diagonalization MIMO broadcast channels with user scheduling," in *Proc. IEEE GLOBECOM*, Dec. 2010.
[15] Z. Pan, K. -K. Wong, and T. -S. Ng, "Generalized multiuser orthogonal space-division multiplexing," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1969–1973, Nov. 2004.
[16] 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, pp. 1711–1721, Mar. 2007.
[17] S.-H. Moon, J. Jung, H. Lee, and I. Lee, "Enhanced groupwise detection with a new receive combiner for spatial multiplexing MIMO systems," *IEEE Trans. Commun.*, vol. 58, pp. 2511–2515, Sept. 2010.
[18] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1108–1123, Mar. 2008.
[19] W. Santipach and M. Honig, "Asymptotic capacity of beamforming with limited feedback," *in Proc. IEEE ISIT*, July 2004.
[20] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
[21] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1893–1909, Sept. 2004.
[22] K. Kumar, G. Caire, and A. Moustakas, "Asymptotic performance of linear receivers in MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4398–4418, Oct. 2009.
[23] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed., McGraw-Hill, 2002.
[24] E. Fieller, "The distribution of the index in a normal bivariate population," *Biometrika*, 1932.
[25] D. Hinkley, "On the ratio of two correlated normal random variables," *Biometrika*, 1932.
[26] National Bureau of Std., "Tables of the bivariate normal distribution function and related functions," *National Bureau of Standards Appl. Mathematics Series*, no. 50, 1959.
[27] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions.* Chapman & Hall/CRC, 2000.

Sung-Hyun Moon received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Korea University, Seoul, Korea in 2005, 2007, and 2012, respectively. Since 2012, he has been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea as a Research Engineer. During spring 2009, he visited the University of Southern California (USC), Los Angeles, CA to conduct collaborative research under the Brain Korea 21. His current research interests include multiuser information theory and signal processing techniques for MIMO wireless cellular networks.

Sang-Rim Lee received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Korea University, Seoul, Korea in 2005, 2007, and 2013, respectively. From 2007 to 2010, he was a Research Engineer with Samsung Electronics, where he conducted research on WiMAX systems. Since September 2013, he has been a Postdoctoral Fellow at Korea University. He was awarded the Silver and Bronze Prizes in the Samsung Humantech Paper Contest in February 2012, and received the Silver Prize in the Inside Edge Paper Contest held by Samsung Electro-Mechanics in 2013. His research topics include communication theory and signal processing techniques for multi-user MIMO wireless networks and distributed antenna systems. He is currently interested in convex optimization, random matrix theory, and stochastic geometry.

Jin-Sung Kim received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Korea University, Seoul, Korea in 2006, 2008, and 2012, respectively. He is currently working as a Member of Technical Staff at Bell Labs Seoul since March 2012. During the summer in 2009, he visited University of Southern California, Los Angeles, CA to conduct collaborative research under the Brain Korea 21 (BK21) Program. In April 2011, he won the excellent graduate student award from BK21 Information Technology Center at Korea University. He was awarded the Silver Prize in the Samsung Humantech Paper Contest in February 2012. His research interests are signal processing techniques including transceiver filter designs and interference mitigation algorithms for wireless communication systems.

Inkyu Lee received the B.S. degree (Hon.) in Control and Instrumentation Engineering from Seoul National University, Seoul, Korea in 1990 and the M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA in 1992 and 1995, respectively. From 1995 to 2001, he was a Member of Technical Staff at Bell Laboratories, Lucent Technologies, where he conducted research on high-speed wireless system designs. He later worked for Agere Systems (formerly Microelectronics Group of Lucent Technologies), Murray Hill, NJ as a Distinguished Member of Technical Staff from 2001 to 2002. In September 2002, he joined the Faculty of Korea University, Seoul, Korea, where he is currently a Professor in the School of Electrical Engineering. During 2009, he visited University of Southern California, LA, USA as a Visiting Professor. He has published around 70 journal papers in IEEE and has 30 U.S. patents granted or pending. His research interests include digital communications and signal processing techniques applied for next generation wireless systems. Dr. Lee currently serves as an Associate Editor for IEEE Transactions on Communications and the IEEE Transactions on Wireless Communications. Also, he has been a Chief Guest Editor for the IEEE Journal on Selected Areas in Communications (Special Issue on 4G Wireless Systems). He was a recipient of the IEEE/IEEK IT Young Engineer Award in 2006, the APCC Best Paper Award in 2006, the IEEE VTC Best Paper Award in 2009, the Korea Institute of Communications and Information Sciences (KICS) Hae-Dong Best Research Award in 2011, the National Academy of Engineering in Korea (NAEK) Best Young Engineer Award in 2013, the ISPACS Best Paper Award in 2013, and the IEEE ComSc APB Outstanding Paper Award in 2013.