User Restriction Scheme for Feedback Reduction of Unitary Matrix Based SDMA

Woohyun Seo*, Sungsoo Park, Dongyoung Kwon, and Daesik Hong**
Information and Telecommunication Lab. School of Electrical and Electronic Eng., Yonsei Univ.
134, Shinchon-dong, Seodaemun-gu, Seoul, Korea, 120-749
Hompage: http://mirinae.yonsei.ac.kr, E-mail : wfyswh@yonsei.ac.kr*, daesikh@yonsei.ac.kr**

Abstract—With the quantized channel state information (CSI), space division multiple access (SDMA) can extract the multiplexing gain under the limited feedback burden. However, huge signaling burden of feedback can still suffer SDMA system because the total feedback data of SDMA is linearly dependent on the number of users. Hence, we propose the user restriction scheme to control the feedback load. In this scheme, the cut-off level, which restricts the feedbacks of poor conditioned users, is suggested for the reduction of the feedback burden without the performance loss. From simulation results, then, we show that the proposed feedback scheme can achieve not only the sum-rate gain but also the reasonable feedback reduction.

Index Terms—Space division multiple access, Quantized CSI, Feedback restriction

I. INTRODUCTION

Space division multiple access (SDMA) represents a promising strategy to increase the spectral efficiency because of the advantage to transmit the packets of multiple users simultaneously [1]. Dirty paper coding is known as the optimal transmission scheme for SDMA [2], but the computational complexity makes it impossible to be implemented in the real system [3]. By this reason, several techniques such as Tomlinson-Harashima precoding were proposed through the efforts for the practical SDMA [4]. However, the performances of these schemes are totally sensitive on the accuracy of channel state information (CSI). Unfortunately, high accurate CSIs are not available for the real-time system [5].

For more practical SDMA, the quantized CSI feedback is adopted as a limited feedback policy, where the presdesigned codebook are already aware to base station (BS) and users [6]-[9]. In this scheme, users feed back only the index of the most favorable code-vector to the channel vector and the signal-to-interference and noise ratio (SINR) information of the corresponding code-vector. At the same time, by combining beamforming and user scheduling, SDMA with the quantized CSI feedback is able to achieve the multiuser diversity, and its performance is approaching to the sum capacity of DPC under infinite users [10].

Thus, SDMA with user scheduling generally considers CSIs from a large number of users to increase multiuser diversity (MUDiv) gain. However, this consideration can cause to enormously magnify the feedback burden because the total feedback data linearly increases to the number of users. In other words, even if the quantized CSI feedback scheme reduces the CSI data per user, the feedback burden for SDMA can be still too huge to be processed by the uplink resource, which is strictly restricted. Furthermore, CSIs of poor conditioned users are not helpful to achieve MUDiv in spite of their signaling burdens. Therefore, for a higher spectral efficiency, it is required to control the feedback data of unavailable users.

In this paper, we propose the user restriction scheme, whose objectives are to reduce the feedback burden without the degradation of the performance. To do so, the proposing feedback scheme introduces the cut-off level for the feedback restriction of users with the large distortion between the channel vectors and the code-vectors. Then, the reasonable cut-off level is suggested from the key factor to increase the achievable ergodic sum-rate of SDMA.

II. SYSTEM MODEL AND USER SELECTION ALGORITHM

A. System Model

Fig. 1 describes the downlink beamforming transmission of frequency division duplex (FDD) system in a macro BS with $N$ transmit antennas. Let us consider $K$ active users with single antenna and assume that all the active users are basically adopting the scheduling for frequency and time resources such as the proportional fair scheduling [11]. Then, active users can utilize the additional uplink feedback channel of CSIs for the joint scheme of beamforming and user scheduling under $K > N$, and we focus on the reduction of this additional feedback.

This work was supported by Korea Science and Engineering Foundation through the NRL Program (Grant R0A-2007-000-20043-0). This research also was supported by the MKE(Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2009-(C1090-0902-0005)).
To manage uplink resource efficiently, the feedback for a joint SDMA scheme is assumed as the dedicated channel, and each user should request the dedicated channel for CSI feedback to BS. If $B$ is the feedback load per user, to construct a robust SDMA scheme with smaller $B$ is one of the most important managements for the efficiency of the system. Thus, the quantized CSI feedback based on unitary matrix can be the solution of this management [12]. In this feedback scheme, since BS with $N$ antennas can assign spatial channels to $N$ users using orthogonal beam weights at most, the user scheduling scheme selects $N$ users with the best CSIs among $K$. For the simple and powerful quantization of the channel vectors, we apply random vector quantization (RVQ) as a codebook generation method [6].

If $x$ and $K$ are the sum of the transmit signals for users and the set of the scheduled users, respectively, the received signal of the scheduled user $k$ is

$$y_k = h_k \cdot x + z_k, \quad k \in K,$$

where $z_k$ represents the additive white Gaussian noise (AWGN) with zero mean and $N_o$ variance. $h_k \in \mathbb{C}^{1 \times N}$ is defined to the channel vector between BS and user $k$, and it is assumed to be i.i.d. unitary complex Gaussians for users. In addition, the expected value of the channel gain is

$$E[\alpha_k] = N,$$

where $\alpha_k = \|h_k\|^2$.

Under user scheduling schemes, the capacity of equal power allocation approaches to the optimal capacity of water-filling method [13][14]. By this reason, we consider equal power allocation for SDMA scheme. When the transmit power and the transmit data symbol of user $k$ is $P$ and $s_k$, respectively, the transmit signal is

$$x = \sqrt{P/N_t} \sum_{u \in K} w_u s_u,$$

where $N_t$ is the number of users in $K$ and $w_k \in \mathbb{C}^{N \times 1}$ is the precoding weight vector of user $k$. In (3), the data symbol and the precoding weight vectors are in the constraints of $E[|s_u|^2] = 1$ and

$$|w_u \cdot w_k| = \begin{cases} 1, & k = u, \\ 0, & k \neq u, \end{cases}$$

respectively, where $\dagger$ is a conjugate transpose vector or matrix operation. Then, the received signal is

$$y_k = \sqrt{P/N_t} h_k w_k s_k + \sqrt{P/N_t} h_k \sum_{u \in K, u \neq k} w_u s_u + z_k, \quad k \in K.$$

### B. User Scheduling Algorithm of Unitary Matrix Based SDMA

In this subsection, we briefly review the user scheduling algorithm for unitary matrix based SDMA with the quantized CSI feedback, which was describe in [7]-[9]. For the simplicity of the review, three assumptions are required. First, users estimate the channel with perfection. This subsection concentrates on the user scheduling algorithm of SDMA rather than the effect of the channel estimation error. Additionally, CSI data is fed back to BS without error. The third assumption is that both BS and users know the same codebook.

Without loss of generality, it can be defined that the codebook is composed of $M$ unitary code-vector sets as follows:

$$C = [C_1, C_2, ..., C_M].$$

In (6), the $m$-th unitary code-vector set is made of $N$ orthonormal vectors and written by

$$C_m = [c_1^m, c_2^m, ..., c_N^m],$$

where $c_n^m \in \mathbb{C}^{N \times 1}$ is the $n$-th code-vector in the $m$-th unitary set. From (6) and (7), the codebook size is $|C| = MN$. By the definition of $h_k$, it can be decomposed into

$$h_k = \sqrt{\alpha_k} \cdot \bar{h}_k,$$

where the channel gain and the channel shape of $h_k$ are $\alpha_k = \|h_k\|^2$ and $\bar{h}_k = h_k/\|h_k\|$, respectively.

In the quantized CSI feedback, each user feeds back only one index of the code-vector to BS, and BS uses the designated code-vector as the information to generate transmit weight vectors instead of the complex values for the channel vector. Thus, if the distortion between $\bar{h}_k$ and $c_n^m$ is denoted to the quantization error, $\Phi_{\bar{h}_k, c_n^m}$, it is [9]

$$\Phi_{\bar{h}_k, c_n^m} = 1 - \| h_k \cdot c_n^m \|^2 = 1 - \Gamma_{\bar{h}_k, c_n^m},$$

where $\Gamma_{\bar{h}_k, c_n^m}$ is the correlation value of $\bar{h}_k$ and $c_n^m$. If the quantization error is occurred, user $k$ receives it as the interference due to the characteristic of unitary matrix based weight vectors, and we denote this to the non-coherent interference. Hence, users should feed back the index of the closest code-vector to $\bar{h}_k$ for the minimization of $\Phi_{\bar{h}_k, c_n^m}$ as follows:

$$\forall k, \Phi_{\bar{h}_k, c_{n}^{\hat{m}}} = \min_{1 \leq m \leq M \atop 1 \leq n \leq N} \Phi_{\bar{h}_k, c_n^m},$$

$$i_k \leftarrow \arg \min_{1 \leq m \leq M \atop 1 \leq n \leq N} \Phi_{\bar{h}_k, c_n^m},$$

where $i_k$ is the index of $c_{n}^{\hat{m}}$ which is the selected code-vector by user $k$.

With $i_k$, the feedback data includes additional information such as SINR and the channel gain. Among them, SINR is one of the most helpful for the user scheduling algorithm to increase sum-rate under the same feedback amount [8]. If the signal-to-noise ratio (SNR) per antenna is $\eta = P/ (N \cdot N_o)$, SINR of user $k$ is

$$\text{SINR}_k = \frac{\eta \cdot \alpha_k \cdot (1 - \Phi_{\bar{h}_k, c_{n}^{\hat{m}}})}{1 + \eta \cdot \alpha_k \cdot \Phi_{\bar{h}_k, c_{n}^{\hat{m}}}}, \hat{n}, \hat{m} \in i_k,$$

from (10). To transmit $i_k$ and SINR$_k$, the required feedback load per user is

$$B = \log_2 (MN) + b,$$

where $b$ is the bit resolution for SINR quantization.

Once $i_k$ and SINR$_k$ are fed back from all users, BS picks up $N_t$ scheduled users among $K$ users based on those information. From the feedback data of user $k$, BS recovers
SINR of the indicated code-vector by $i_k$ to $\hat{\gamma}_{\tilde{m},k}$. Using the user scheduling algorithm, BS finds the user with the maximum SINR on each code-vector such as
\[
\forall \tilde{m}, \forall n, \hat{\gamma}_{\tilde{m},n} = \max_{1 \leq k \leq K} \hat{\gamma}_{\tilde{m},k},
\]
where $\kappa_{\tilde{m}}$ is the index of the user with the maximum SINR on the $\tilde{m}$-th code-vector of the $\tilde{m}$-th unitary. Since the orthogonality in the simultaneous packet transmission of multiuser is kept on only one unitary code-vector set, BS should choose one code-vector set for the maximization of the achievable sum-rate, i.e.,
\[
\tilde{m} = \arg \max_{1 \leq \tilde{m} \leq M} \left[ \sum_{n=1}^{N} \log \left( 1 + \hat{\gamma}_{\tilde{m},n} \right) \right].
\]
Finally, BS allocates the transmission power to all the code-vectors with user $\kappa_{\tilde{m}}$ in the $\tilde{m}$-th unitary set.

In multi-antenna broadcasting system, $N_t$ can grow to $N$ at most, but $N_t$ is actually determined by the size of the user pool, where the user pool is denoted to the number of users with the request of CSI feedback. In a small user pool, some of the code-vectors of the $\tilde{m}$-th unitary set may have no CSIs. At this time, no power allocation to those code-vectors is more sophisticated for the sum-rate maximization [9], and $N_t$ is naturally reduced. In a large user pool, as the user scheduler can have enough CSIs of users on each code-vector, it is guaranteed that $N_t$ becomes $N$ to maximize the achievable sum-rate.

III. EFFICIENT FEEDBACK SCHEME FOR SDMA

As for SDMA with the quantized CSI feedback, to reduce the feedback burden is an important work to increase the spectral efficiency. Hence, in this section, we propose the user restriction scheme to reduce the feedback burden. To do so, firstly, the achievable sum-rate of SDMA is derived in a large non-coherent interference in SDMA. The first condition is that $\mu$ and $K$ is approaching to zero, SDMA can achieve the maximum achievable ergodic sum-rate, which is same as the sum-capacity of the multiuser system with $N$ orthogonal multichannel.

As shown in (20), the achievable ergodic sum-rate is decided by $\mu$, $K$, and $\Phi_{h_k,c}$, $\kappa$. Among these parameters, $\mu$ and $K$ are commonly included into all the parts of the achievable ergodic sum-rate. Hence, under a good channel gain and high SNR, the interference due to the quantization error as well as multiuser diversity gain are increased with the same ratio. By this reason, if user $k$ has a large $\Phi_{h_k,c}$, it is difficult to expect SINR of user $k$ even in a good channel condition.

A. Achievable ergodic sum-rate of SDMA with the quantized CSI feedback

In this subsection, we derive the achievable ergodic sum-rate of SDMA with the quantized CSI feedback. For the simplicity of the derivation, we assume that BS has a large user pool, and then $N_t = N$. Additionally, it is assumed that $\alpha_k$ and $\Phi_{h_k,c}$ are uncorrelated. It is because that $\Phi_{h_k}$ is independently decomposed with $\alpha_k$ from $h_k$ in (8) and $\Phi_{h_k,c}$ is evaluated from $\tilde{h}_k$ and $c_n$ in (9).

For the scheduled user $\kappa_{\tilde{m}}$ in (14), if SINR on the $n$-th code-vector of the $\tilde{m}$-th unitary set is
\[
\text{SINR}_{\kappa_{\tilde{m}}} = \frac{\eta \cdot \alpha_{\kappa_{\tilde{m}}} \left( 1 - \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}} \right)}{1 + \eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}},
\]
the achievable sum-rate is given to
\[
R = \sum_{n=1}^{N} \log \left( 1 + \frac{\eta \cdot \alpha_{\kappa_{\tilde{m}}} \left( 1 - \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}} \right)}{1 + \eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}} \right).
\]
Based on (16), the achievable ergodic sum-rate is
\[
R = E \left[ \sum_{n=1}^{N} \log \left( 1 + \frac{\eta \cdot \alpha_{\kappa_{\tilde{m}}} \left( 1 - \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}} \right)}{1 + \eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}} \right) \right] = N \cdot E \left[ \log \left( 1 + \frac{\eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}} {1 + \eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}} \right) \right],
\]
(17) can be easily changed to
\[
R = N \cdot E \left[ \log \left( 1 + \frac{\eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}} {1 + \eta \cdot \alpha_{\kappa_{\tilde{m}}} \Phi_{h_{\kappa_{\tilde{m}}},c_{\kappa_{\tilde{m}}}}} \right) \right].
\]
From the above assumption about the uncorrelated relation of $\alpha_k$ and $\Phi_{h_k,c}$, (18) can be sequentially calculated by the two expectations for $\alpha_k$ and $\Phi_{h_k,c}$. Firstly, using the asymptotic analysis of the selection diversity [15][16], the expectation for $\alpha_k$ in (18) is
\[
\alpha_k \approx N \cdot \log \left( 1 + \mu \ln(K) \right) - \mu \cdot K \cdot \Phi_{h_k,c} \cdot \kappa
\]
where $\mu = N \eta$. Then, by Jensen’s inequality, the expectation for $\Phi_{h_k,c} \kappa$ in (19) is
\[
\Phi_{h_k,c} \kappa \geq N \cdot \log \left( 1 + \mu \ln(K) \right) - \mu \cdot K \cdot \Phi_{h_k,c} \kappa,
\]
(20) where $\Phi_{h_k,c} \kappa$ is the expectation value of $\Phi_{h_k,c} \kappa$. If $\Phi_{h_k,c} \kappa$ is approaching to zero, SDMA can achieve the maximum achievable ergodic sum-rate, which is the same as the multiuser system with $N$ orthogonal multichannel.

B. Feedback restriction by the cut-off level

From (20), we can find that users with a large quantization error is not able to achieve the high achievable sum-rate, and then to exclude the feedback of user $k$ with a large $\Phi_{h_k,c} \kappa$ from user pool can be a good approach to reduce the feedback burden even without any performance loss. Therefore, we propose the feedback restriction scheme to increase the spectral efficiency. In this case, all users have the cut-off level of the quantization error, which is defined to $\Phi_c$. If $\Phi_c < \Phi_{h_k,c} \kappa$ in (10), user $k$ does not request the dedicated feedback channel to BS. In other words, the system denies the CSI feedback of user $k$, who has the probability to experience a large non-coherent interference in SDMA.

In order to decide $\Phi_c$, we introduce two conceptual and basic conditions to prevent a poor performance from the non-coherent interference of SDMA. The first condition is that the
received power of the desired signal should be larger than that of the interference signal at least. If this condition is not satisfied, SINRs of users are limited to less than 0dB regardless of SNR. The other condition is that the achievable sum-rate of SDMA should be larger those of other orthogonal multiple access schemes (OMASs) for a single user transmission. If the non-coherent interference increases due to a large $\Phi_{h_k,c_{\hat{n}}}^c$, SDMA may have the worse performance than OMASs, and then the advantage of the multiuser transmission is disappeared. If user $k$ is out of these conditions, a single user transmission can be the better choice for the spectral efficiency. Especially, when user pool is small, this method is able to improve the performance even with the reduced feedback burden.

Hence, we design $\Phi_c$ to apply the aforementioned conditions for the better performance of SDMA. Instinctively, to satisfy the first condition, $\tilde{\Phi}_c$ should be less than $1/2$ because $\tilde{\Phi}_c$ determines the received power ratio of the interference signal to the desired signal. In addition, for the satisfaction of the second condition, the achievable ergodic sum-rate of SDMA is compared to those of OMASs with a multiuser scheduling scheme. Using (19), $\Phi_c$ can be decided by

$$\log (1 + \mu \ln (K)) < \sum_{n=1}^{N} \log (1 + \text{SINR}_k)$$

$$= N \{ \log (1 + \mu \ln (K)) - \log (1 + \mu \ln (K) \cdot \tilde{\Phi}_c) \}.$$  \hfill (21)

In (21), $\tilde{\Phi}_c$ should be the largest value to restrict the user with a large $\Phi_{h_k,c_{\hat{n}}}^c$ as follows:

$$\tilde{\Phi}_c = \left[ \frac{(1 + \mu \ln (K))^{N-1} - 1}{(1 + \mu \ln (K))} \cdot \mu \ln (K), \mu > 0. \right]$$ \hfill (22)

Finally, to combine the above discussed results, $\Phi_c$ for SDMA is determined as the smallest value between (22) and $1/2$ as follows:

$$\Phi_c = \min \left( \frac{\left[ (1 + \mu \ln (K))^{N-1} - 1 \right]}{\mu \ln (K)}, \frac{1}{2} \right), \mu > 0. \quad (23)$$

Based on $\tilde{\Phi}_c$ of (23), the feedback of user $k$ with $\Phi_{h_k,c_{\hat{n}}}^c$ is cut off if

$$\tilde{\Phi}_c < \Phi_{h_k,c_{\hat{n}}}^c = 1 - \Gamma_{h_k,c_{\hat{n}}}^c, \quad \hat{n}, \hat{m} \in i_k,$$ \hfill (24)

where $i_k$ is decided in (11). If the feedback requests of all users are cut off by (24), BS transmits through OMASs rather than SDMA.

IV. Simulation Results

Simulation results are presented to evaluate the performance of the proposed feedback scheme compared to that of the feedback scheme without the cut-off level. In this simulation, we assume that BS has 4 antennas, and users feed back SINR with 3 bit quantization; $b = 3$. For the convenience of the simulation, all users are assumed to have the same SNR. To verify the performance of the proposed feedback scheme, we compare two different feedback methods; SDMA with the cut-off level (WCL) and SDMA without the cut-off level (WoCL).

Fig. 2 shows the performances of WCLs with $|C| = 4$ and 16 over $K$ in 5dB SNR compared to those of WoCLs. For all the results, the solid and the dot lines stand for SDMAs with $|C| = 4$ and 16, respectively. In WCL of Fig. 2, though the feedback requests of users are restricted by $\tilde{\Phi}_c$, it has the improved or similar performance than WoCL. This is because WCL can effectively reduce the feedbacks of users with a large non-coherent interference through $\tilde{\Phi}_c$. In the previous section, $\Phi_c$ is derived with the assumption of a large user pool, but the achievable ergodic sum-rates of WCLs are better than those of WoCLs in small $K$ as well.

In Fig. 3, the feedback rates of WCLs with $|C| = 4$ and 16 are displayed over $K$ in 5dB SNR, where the feedback rate is denoted to the ratio of users with the feedback request to $K$; feedback users/$K$. Considering (23), as $K$ increases, $\Phi_c$ decreases by the factor of $\ln(K)$. Then, the quantization errors
of users are more strictly cut off to achieve higher SINR in larger $K$, and the feedback rate also decreases. In addition, since the quantization errors of users are generally reduced in larger $|C|$, the feedback rate in $|C| = 16$ is larger than $|C| = 4$.

In Fig. 4, the achievable ergodic sum-rates of WCLs with $|C| = 4$ and 16 are compared to those of WoCLs over SNR in $K = 40$. As SNR increases, the system gets closer to the interference limited environment, and the performance of SDMA is sensitive to the non-coherent interference rather than noise. From (23), since $\Phi_c$ of WCL becomes the stricter criterion at the larger SNR, the non-coherent interference can be reduced by $\Phi_c$. By this reason, the achievable ergodic sum-rates of WCLs increase without the limitation as SNR increases, while those of WoCLs are bounded in a large SNR.

Fig. 5 plots the feedback rates of WCLs with $|C| = 4$ and 16 over SNR, where the feedback rate is denoted to the ratio of users with the feedback request to $K$. As shown in this figure, when SNR increases, the system more tightly limits the feedback requests of users by (23). Thus, when SNR is larger than 5dB, the feedback ratio of the proposed feedback is steeply reduced. Specifically, only less than 10% of users have the feedback requests in larger than 25dB SNR.

In summary, based on these simulation results, it is found that WCL can improve the performance as well as reduce the feedback burden.

V. CONCLUSIONS

This paper proposed the efficient feedback with the goal of increasing the spectral efficiency. From the derivation of the achievable ergodic sum-rate for SDMA, WCL was suggested to control the non-coherent interference caused by the quantization error with the reduced feedback burden. In simulation results, we showed that WCL has the better performance than WoCL with the small portion of the feedback load.

REFERENCES