A Vector Perturbation with User Selection for Multiuser MIMO Downlink

Byungju Lee, *Student Member, IEEE*, and Byonghyo Shim, *Senior Member, IEEE*

***Abstract*—Recent works on multiuser MIMO study have shown that the linear growth of capacity in single user MIMO system can be translated to the multiuser MIMO scenario as well. In this paper, we propose a method pursuing performance gain of vector perturbation in multiuser downlink systems. Instead of employing the maximum number of mobile users for communication, we use small part of them as virtual users for improving reliability of users participating communication. By controlling parameters of the virtual users including information and perturbation vector, we obtain considerable improvement in the effective SNR. Simulation results on the realistic multiuser MISO and MIMO downlink systems show that the proposed method brings substantial performance gain over the standard vector perturbation with marginal overhead in computations.**

***Index Terms*—Multiuser downlink, vector perturbation, non-integer least square, sphere encoding, transmit precoding, virtual users.**

I. Introduction

Recently, there has been growing interest in multiuser MIMO broadcast system where a basestation with multiple antennas transmits independent information to multiple mobile users, each with one or more receive antennas on the same frequency and time slots. Due to the fact that co-channel interference is hard to be managed by the receiver operation, and more importantly capacity of interference channel can be made equivalent to capacity without interference by judiciously controlling interference at the transmitter, pre-cancellation of the interference via the precoding at the transmitter has received much attention. It is now well known that dirty paper coding (DPC) [2] can achieve the capacity region of the multiuser MIMO downlink [3]–[5]. However, the shortcoming of the DPC is that system implementation leads to the high computational cost of successive encoding and decoding [6].

Lately, an approach referred to as vector perturbation technique [7] achieves the performance close to the capacity on multiantenna multiuser downlink channel. By adding a deliberately designed perturbation vector into the information vector, the vector perturbation provides substantial improvement in the performance over the linear precoding methods such as channel inversion and regularized inversion with no virtual overhead in the receiver. Although the vector perturbation is a promising scheme for maximizing the sum rate of the MISO type broadcasting channel, we might need a strong protection against fading to meet the quality of service (QoS) of next-generation wireless systems.

In this paper, we put forth an approach improving the bit error rate performance by sacrificing the goal of maximizing the sum rate. To be specific, when the channel of certain users is in outage, we exploit these users to improve the reliability of the rest participating in communication. Towards this end, the proposed approach selects virtual users among users in outage to improve the signal-to-interference-plus-noise ratio (SINR) of active users. It should be noted that, although the objective of proposed method and diversity technique [8] to resist the channel fading for improving reliability is the same, the means to achieve the goal is dearly distinct. While the diversity scheme accomplishes the goal by employing spatial or temporal replica of the transmitted signal, the proposed method attains the most out of the users whose contribution to the sum rate is trivial. Indeed, since parameters of sacrificing users including information data, perturbation vector, and even further the associated channel vector can be arbitrarily chosen, proper control of these parameters help improving the SINR of the mobile users participated in communication.

The rest of this paper is organized as follows. After a brief summary of system model and the vector perturbation technique in Section II, we present the proposed method in Section III. In section IV, we extend the proposed method to the multiuser MIMO downlink scenario. We discuss simulation results in Section V and provide conclusion in Section VI.

We briefly summarize notations used in this paper. We employ uppercase boldface letters for matrices and lowercase boldface for vectors. The superscripts $(\cdot)^H$ and $(\cdot)^T$ denote conjugate transpose and transpose, respectively. $\|\cdot\|$ indicates an L_2-norm of a vector. diag$(\cdot)$ is a diagonal matrix where nonzero elements exist only on the main diagonal of the matrix. $\mathcal{CN}(m, \sigma^2)$ denotes a complex Gaussian random variable with mean m and variance σ^2. $Q(x)$ denotes Q-function defined as $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx$.

II. System Model and Previous Works

In the multiuser downlink MISO system, a basestation composed of M antennas transmits information to K mobile users, each with a single receive antenna. Under the flat fading channel assumption, the received signal vector $\mathbf{y} = [y_1, \cdots, y_K]^T$ can be described by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where $\mathbf{H} = [\mathbf{h}_1^T, \cdots, \mathbf{h}_K^T]^T$ is the $K \times M$ channel matrix where $\mathbf{h}_i$ is the channel vector of the user i, $\mathbf{x} =$

This research was supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communication Agency) (KCA-2012-(12-911-01-110)).

The authors are with the School of Information and Communications, Korea University, Anam-dong, Seongbuk-gu, Seoul, Korea, 136-713 (e-mail: {bjlee, bshim}@isl.korea.ac.kr).

The part of this work appeared at Information Theory and Applications Workshop 2011 [1].

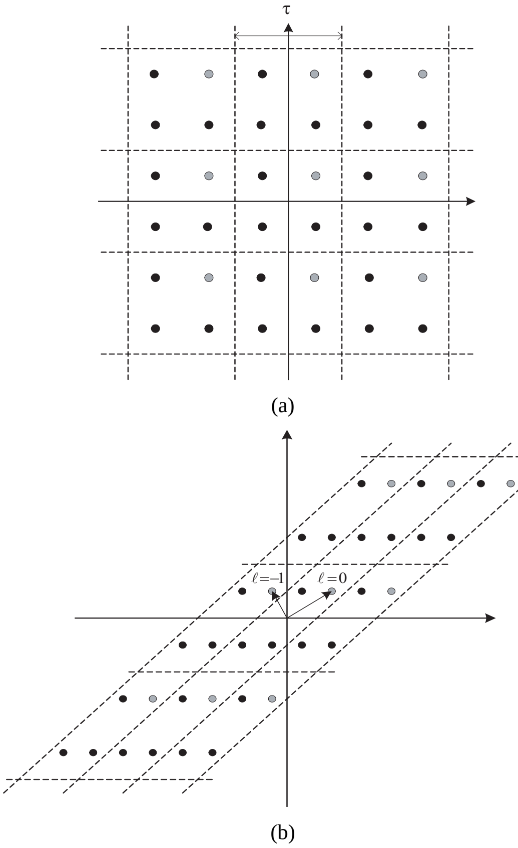


Fig. 1. Illustration of γ reduction using the vector perturbation. Lattice points with gray dots represent $\mathbf{s}+\tau\boldsymbol{\ell}'$ in (a) and $\mathbf{P}(\mathbf{s}+\tau\boldsymbol{\ell}')$ in (b). We can observe from (b) that $\boldsymbol{\ell}=-\mathbf{1}$ provides smaller γ than $\boldsymbol{\ell}=\mathbf{0}$.

$[x_1,\cdots,x_M]^T$ is the transmitted signal vector normalized with the unit power constraint ($\mathrm{E}\|\mathbf{x}\|^2=1$), and $\mathbf{n}=[n_1,\cdots,n_K]^T$ is a complex Gaussian noise vector ($\mathbf{n}\sim\mathcal{CN}(0,\sigma_n^2\mathbf{I})$).

In generating the transmitted signal, well-known channel inversion (zero forcing) scheme multiplies the pseudo inverse of the channel matrix $\mathbf{P}=\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$ to the information vector $\mathbf{s}$ to remove the interference at the transmitter. An enhanced scheme adding a multiple of the identity matrix $\alpha\mathbf{I}$ into the matrix to be inverted called regularized channel inversion was also proposed ($\mathbf{P}=\mathbf{H}^H(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I})^{-1}$) [9]. In this method, the transmitted signal vector at the basestation is

$$\mathbf{x}=\frac{\mathbf{P}\mathbf{s}}{\sqrt{\gamma}}=\frac{1}{\sqrt{\gamma}}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H+\alpha\mathbf{I})^{-1}\mathbf{s} \tag{2}$$

where $\gamma=\|\mathbf{H}^H(\mathbf{H}\mathbf{H}^H+\alpha\mathbf{I})^{-1}\mathbf{s}\|^2$ is employed to ensure the unit transmit power constraint and the desired information vector $\mathbf{s}=[s_1,\cdots,s_K]^T$ is chosen from q-ary quadrature amplitude modulation (QAM). The received signal vector is

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{n}=\frac{1}{\sqrt{\gamma}}\mathbf{H}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H+\alpha\mathbf{I})^{-1}\mathbf{s}+\mathbf{n} \tag{3}$$

where $\alpha>0$. When $\alpha=0$, regularized inversion returns to the channel inversion.

In the vector perturbation method, the information vector $\mathbf{s}$ is perturbed by an integer offset vector $\tau\boldsymbol{\ell}$ so that the modified information becomes $\tilde{\mathbf{s}}=\mathbf{s}+\tau\boldsymbol{\ell}$ where $\tau\in\mathbb{R}^+$ and $\boldsymbol{\ell}=\boldsymbol{\ell}_{\mathrm{re}}+j\boldsymbol{\ell}_{\mathrm{im}}$ ($\boldsymbol{\ell}_{\mathrm{re}},\boldsymbol{\ell}_{\mathrm{im}}\in\mathbb{Z}^K$) is a K-dimensional complex vector [7]. Similar to the regularized inversion technique, the transmitted signal of this approach is

$$\mathbf{x}=\frac{\mathbf{P}\tilde{\mathbf{s}}}{\sqrt{\gamma}}=\frac{1}{\sqrt{\gamma}}\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H+\alpha\mathbf{I}\right)^{-1}\tilde{\mathbf{s}} \tag{4}$$

where $\gamma=\|\mathbf{P}\tilde{\mathbf{s}}\|^2$. In this approach, $\boldsymbol{\ell}$ minimizing γ is chosen to boost the effective SNR and thus

$$\boldsymbol{\ell}=\arg\min_{\boldsymbol{\ell}'}\|\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H+\alpha\mathbf{I}\right)^{-1}(\mathbf{s}+\tau\boldsymbol{\ell}')\|^2. \tag{5}$$

This integer least squares problem can be solved by a closest lattice point search, referred to as *sphere encoder* [7]. Note that since the vector perturbation searches $\boldsymbol{\ell}$ from the set $\{\mathbf{s}+\tau\boldsymbol{\ell}\mid\boldsymbol{\ell}\in\mathbb{Z}^K+j\mathbb{Z}^K\}$ to generate minimum possible γ, γ of the standard vector perturbation is always smaller than or equal to the regularized channel inversion where $\boldsymbol{\ell}=\mathbf{0}$. Plugging (4) into (1), the received vector becomes $\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{n}\approx\frac{1}{\sqrt{\gamma}}\tilde{\mathbf{s}}+\mathbf{n}=\frac{1}{\sqrt{\gamma}}(\mathbf{s}+\tau\boldsymbol{\ell})+\mathbf{n}$. Since elements of $\boldsymbol{\ell}$ are integers, $\tau\boldsymbol{\ell}$ can be removed by the modulo operation [7] in the receiver and hence

$$\mathbf{y}'=\frac{1}{\sqrt{\gamma}}(\sqrt{\gamma}\mathbf{y}\bmod\tau)\approx\frac{\mathbf{s}}{\sqrt{\gamma}}+\mathbf{n}. \tag{6}$$

The salient feature of the vector perturbation lies on the use of integer based perturbation vector that can be easily removed by the modulo operation in the receiver, which provides considerable reduction of γ (see Fig. 1).

III. Vector Perturbation with Virtual Users

A. Motivation

Although the vector perturbation method achieves bit error rate performance close to the sum capacity [4], in many practical scenarios where some mobiles are under deep fading over multiple codeblocks, performance loss for each user will be considerable. Recalling that the received vector of jth user is $y_j\approx\frac{1}{\sqrt{\gamma}}\tilde{s}_j+n_j$, the received SNR of user j after the modulo operation becomes $\mathrm{SNR}_j=\frac{E\|s_j\|^2}{E\gamma\|n_j\|^2}$. In fact, since γ affects SNR of all users, whole mobile users will suffer performance degradation when the channel matrix is poorly conditioned due to the deep fading of certain users. An idea motivated by this observation is to exploit part of spatial degree of freedom as sacrificial lambs (virtual users) to enhance performance of the active users. Since the performance of virtual users is not our concern in this case, an aggressive control of their parameters is possible for maximizing performance gain of active users. In fact, the key distinctions of the proposed method over the standard vector perturbation are that 1) selection of virtual users when channels of certain users are in outage, 2) construction of well-conditioned $\mathbf{H}$ matrix by using channel vectors of virtual users, and finally 3) selection of the non-integer perturbed input $\mathbf{s}_v+\tau\boldsymbol{\ell}_v$ for virtual users using a modified sphere encoding. For the analytic simplicity, we do not consider the modulo loss at the receiver in our sum rate analysis.[1] Also, we consider γ without regularization ($\alpha = 0$). Thus, our result in this section might be slightly pessimistic.

B. Virtual User Selection

Our goal in this subsection is to select virtual users, only if at least one user is in outage, and then construct a channel matrix being used for the precoding. In the first stage, outage condition is tested for each user to select N_a active users ($N_{a,\min} \leq N_a \leq K$). Next, in case $N_a < K$ (i.e, at least one virtual user is selected), $K - N_a$ virtual channels are generated to replace the rows of sacrificing users with these.

To be specific, in the first stage, we check whether a user is in outage. For a given rate threshold ϵ, we say a user j is in ϵ-outage if the rate $R_j = \log(1 + \mathrm{SNR}_j)$ of the user is smaller than ϵ. If no user is in ϵ-outage, all users become the active users ($N_a = K$). Otherwise, we choose at most $K - N_{a,\min}$ virtual users among users in outage.

Once at least one virtual user is chosen, our concern in $\mathbf{H}$ matrix construction is only on the rows $\mathbf{h}_k$ of virtual users k. Instead of randomly generating $\mathbf{h}_k$, we select channels of mobile users not participating in real communication as candidates for virtual rows. That is, rows of sacrificing users in $\mathbf{H}$ are replaced by those of candidates. Since the generated channel matrix we use in the precoding is distinct from the original channel matrix $\mathbf{H}$, we henceforth denote it as $\tilde{\mathbf{H}}$. Without loss of generality, we assume that users are sequenced in the following order: active users and virtual users. Then the top N_a rows of $\tilde{\mathbf{H}}$ corresponding to active users are equivalent to those of $\mathbf{H}$ and hence

$$\mathbf{H}\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)^{-1} = \begin{bmatrix} \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{X}_0 & \mathbf{X}_1 \end{bmatrix} \tag{7}$$

where $\mathbf{X} = [\mathbf{X}_0 \ \mathbf{X}_1]$ is the rows of virtual users for which we do not care. Clearly, there is no impact of virtual channel on the received signal of active users y_j for $j = 1, \cdots, N_a$. After the precoding with $\tilde{\mathbf{H}}$, the received vector for active users is expressed as

$$\mathbf{y}_a = \frac{1}{\sqrt{\gamma}}\tilde{\mathbf{s}} + \mathbf{n}. \tag{8}$$

Let us now consider the virtual channel selection. Denoting γ of the proposed method as γ', $\gamma' = \|\tilde{\mathbf{H}}^H\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1}\tilde{\mathbf{s}}\|^2$ $= \tilde{\mathbf{s}}^H\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1}\tilde{\mathbf{s}}$. Also, using the eigenvalue decomposition of $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, we have $\gamma' = \tilde{\mathbf{s}}^H\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^H\tilde{\mathbf{s}}$. By letting $\tilde{\mathbf{v}} = \mathbf{Q}^H\tilde{\mathbf{s}}$, γ' is expressed as

$$\gamma' = \tilde{\mathbf{v}}^H\mathbf{\Lambda}^{-1}\tilde{\mathbf{v}} = \sum_{i=1}^{M}\frac{|\tilde{v}_i|^2}{\lambda_i} \tag{9}$$

where λ_i is ith eigenvalue of $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H$ and $\tilde{v}_i = \mathbf{q}_i^H\tilde{\mathbf{s}}$ is ith element of $\tilde{\mathbf{v}}$ vector ($\mathbf{q}_i$ is the ith column of $\mathbf{Q}$). Since $0 \leq$

[1] Since the modulo loss occurs only when the noise power is much larger than the signal power, it converges to 0 when the channel SNR increases (i.e., $((\mathbf{s} + \sqrt{\gamma}\mathbf{n} + \tau\mathbf{p}) \bmod \tau) = (\mathbf{s} + \sqrt{\gamma}\mathbf{n})$ when $\sigma_n^2 \to 0$). For example, the probability of modulo loss event for QPSK system is $1 - (Q\left(\frac{-3}{\sqrt{\gamma}\sigma_n}\right) - Q\left(\frac{1}{\sqrt{\gamma}\sigma_n}\right))^2$ so that this event rarely occurs for most mid and high SNR regime.

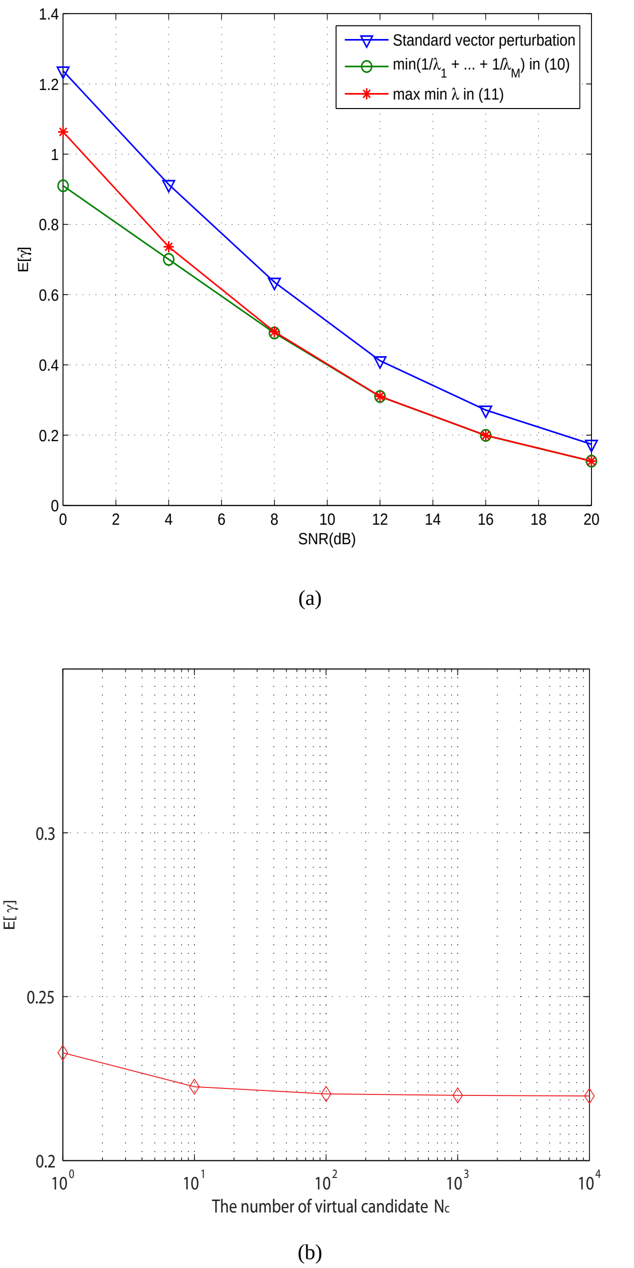


Fig. 2. (a) $E[\gamma]$ of the proposed method (using (10) and (11) and $N_{a,\min} = 3$ and $N_c = 10$) as well as standard vector perturbation for 4×4 multiuser MISO downlink system. (b) $E[\gamma]$ of the proposed method as a function of the number of virtual candidates N_c (SNR = 15 dB)

$\lambda_i \leq 8M$ [10] while $\Sigma_{i=1}^{M}|\tilde{v}_i|^2 = \|\tilde{\mathbf{s}}\|^2$, variations among $|\tilde{v}_i|^2$ is usually smaller than that of λ_i and thus the criterion to choose $\tilde{\mathbf{H}}$ among candidate matrix $\tilde{\mathbf{H}}_j$ can be approximated to

$$\tilde{\mathbf{H}} = \arg\min_{\tilde{\mathbf{H}}_j}\sum_i \frac{1}{\lambda_i^{(j)}} \tag{10}$$

where $\tilde{\mathbf{H}}_j$ is the jth candidate matrix including the rows $\mathbf{h}_j$ of virtual users j and $\lambda_i^{(j)}$ is the ith eigenvalue of $\tilde{\mathbf{H}}_j\tilde{\mathbf{H}}_j^H$. Note that in choosing $\tilde{\mathbf{H}}$ using (10), we need to find all the eigenvalues of $\tilde{\mathbf{H}}_j\tilde{\mathbf{H}}_j^H$, which can be cumbersome for a large dimensional matrix. Exploiting the approximation $\sum_i \frac{1}{\lambda_i} = \frac{1}{\lambda_1}(1 + \frac{\lambda_1}{\lambda_2} + \cdots + \frac{\lambda_1}{\lambda_M}) \approx \frac{1}{\lambda_1}$ where $\lambda_1 < \cdots < \lambda_M$, we obtain a simplified condition (max-min criterion) given by

$$\tilde{\mathbf{H}} = \arg\max_{\tilde{\mathbf{H}}_j} \lambda_{\min}^{(j)}. \tag{11}$$

By employing (11) instead of (10), computational overhead can be reduced substantially.[2]

Fig. 2(a) plots γ' as a function of SNR for the standard vector perturbation and proposed approaches employing (10) and (11). Although the average γ' using (10) is smaller than that obtained from (11) in low SNR regime, they become close when the SNR increases. This suggests that max-min criterion can provide cost effective solution with negligible performance loss for mid and high SNR regime. In order to observe how many virtual user candidates do we need for achieving best performance, we simulate γ' as a function of the number of candidates N_c. We see from Fig. 2(b) that with only small number of candidates ($N_c \approx 10$), $E[\gamma']$ becomes close to the achievable limit.

C. Modified Sphere Encoding for Searching Optimal γ'

In the standard vector perturbation, the sphere encoding solves

$$\min_{\boldsymbol{\ell}\in\mathbb{U}^M} \|\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H\right)^{-1}(\mathbf{s}+\tau\boldsymbol{\ell})\|^2 \tag{12}$$

where $\mathbb{U} = \mathbb{Z} + j\mathbb{Z}$. In practice, $\mathbb{U} = \{z + jz | z = 0, \pm1, \pm2\}$ is used since most of real and imaginary $\boldsymbol{\ell}$ values are either 0, ±1, or ±2 [13]. Since $\tilde{\mathbf{s}}$ of the proposed method can be represented as $\tilde{\mathbf{s}} = \mathbf{s} + \tau\boldsymbol{\ell} = \begin{bmatrix}\mathbf{s}_a \\ \mathbf{s}_v\end{bmatrix} + \tau\begin{bmatrix}\boldsymbol{\ell}_a \\ \boldsymbol{\ell}_v\end{bmatrix}$, it is clear that the precoding of the proposed scheme contains additional control variable $\mathbf{s}_v$ other than $\boldsymbol{\ell}$. Then the optimization problem of the proposed method becomes

$$\min_{\boldsymbol{\ell}\in\mathbb{U}^M, \mathbf{s}_v\in\mathbb{V}^{N_v}} \|\tilde{\mathbf{H}}^H\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1}(\mathbf{s}+\tau\boldsymbol{\ell})\|^2 \tag{13}$$

where $\mathbb{V}^{N_v}$ is the modulation set for virtual users. Before converting this into the standard least squares problem, it is desirable to rearrange $\mathbf{s} + \tau\boldsymbol{\ell}$ as

$$\mathbf{s}+\tau\boldsymbol{\ell} = \begin{bmatrix}\mathbf{s}_a \\ \mathbf{s}_v\end{bmatrix} + \tau\begin{bmatrix}\boldsymbol{\ell}_a \\ \boldsymbol{\ell}_v\end{bmatrix} = \begin{bmatrix}\mathbf{s}_a \\ \mathbf{0}\end{bmatrix} + \tau\begin{bmatrix}\boldsymbol{\ell}_a \\ \frac{\mathbf{s}_v}{\tau} + \boldsymbol{\ell}_v\end{bmatrix} \tag{14}$$

where $\mathbf{s}_a$ and $\boldsymbol{\ell}_a$ are information and perturbation vector of active users. Denoting $\tilde{\boldsymbol{\ell}}_v = \frac{\mathbf{s}_v}{\tau} + \boldsymbol{\ell}_v$, (14) can be rewritten as $\mathbf{s} + \tau\boldsymbol{\ell} = \mathbf{s}' + \tau\boldsymbol{\ell}'$ where $\mathbf{s}' = \begin{bmatrix}\mathbf{u}_a \\ \mathbf{0}\end{bmatrix}$ and $\boldsymbol{\ell}' = \begin{bmatrix}\boldsymbol{\ell}_a \\ \tilde{\boldsymbol{\ell}}_v\end{bmatrix}$. It is worth mentioning that even after this transform, the role that $\mathbf{s}'$ is given while $\boldsymbol{\ell}'$ needs to be optimized remains unchanged. By denoting $\mathbf{r} = \tilde{\mathbf{H}}^H\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1}\mathbf{s}'$ and $\mathbf{B} = -\tau\tilde{\mathbf{H}}^H\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1}$, (13) becomes

$$\gamma' = \min_{\boldsymbol{\ell}_a\in\mathbb{U}^{N_a}, \tilde{\boldsymbol{\ell}}_v\in\tilde{\mathbb{V}}^{N_v}} \|\mathbf{r} - \mathbf{B}\boldsymbol{\ell}'\|^2. \tag{15}$$

[2]Simple iteration algorithms to find the minimum eigenvalue of $\tilde{\mathbf{H}}_j\tilde{\mathbf{H}}_j^H$ include inverse power method and shifted power method. Refer to [11], [12], [17] for detailed description of these.

It is interesting to note that, while an integer constraint is imposed for $\boldsymbol{\ell}_a$, such is not necessarily for $\tilde{\boldsymbol{\ell}}_v$ since the symbols of virtual users are not our interest. Thus, the way to minimize γ' is to choose $\tilde{\boldsymbol{\ell}}_v$ from the N_v-dimensional complex plane ($\tilde{\mathbb{V}} = \mathbb{C}$) and to choose $\boldsymbol{\ell}_a$ from $\mathbb{U}^{N_a}$. In fact, the resulting problem becomes a mixture of integer and non-integer least squares problem given by

$$\gamma' = \min_{\boldsymbol{\ell}_a\in\mathbb{U}^{N_a}} \min_{\tilde{\boldsymbol{\ell}}_v\in\mathbb{C}^{N_v}} \left\|\mathbf{r} - \mathbf{B}_1\boldsymbol{\ell}_a - \mathbf{B}_2\tilde{\boldsymbol{\ell}}_v\right\|^2 \tag{16}$$

where $\mathbf{B}_1 \in \mathbb{C}^{M\times N_a}$ and $\mathbf{B}_2 \in \mathbb{C}^{M\times N_v}$ are the partitioned sub-matrices of $\mathbf{B} = [\mathbf{B}_1\ \mathbf{B}_2]$.

The way to solve this problem consists of two steps. In the first step, optimal $\tilde{\boldsymbol{\ell}}_v$ for each $\boldsymbol{\ell}_a$ is computed using the standard least squares technique. Then $\tilde{\boldsymbol{\ell}}_v$ becomes a function of $\boldsymbol{\ell}_a$ (i.e., $\tilde{\boldsymbol{\ell}}_v = f(\boldsymbol{\ell}_a)$). Denoting $g(\boldsymbol{\ell}_a) = \min_{\tilde{\boldsymbol{\ell}}_v\in\mathbb{C}^{N_v}} \|\mathbf{r} - \mathbf{B}_1\boldsymbol{\ell}_a - \mathbf{B}_2 f(\boldsymbol{\ell}_a)\|^2$, (16) can be expressed as

$$\gamma' = \min_{\boldsymbol{\ell}_a\in\mathbb{U}^{N_a}} g(\boldsymbol{\ell}_a). \tag{17}$$

In the second step, we solve this integer least squares problem (ILSP) using the sphere encoding. To be specific, for given $\boldsymbol{\ell}_a$ (i.e., when $\mathbf{r} - \mathbf{B}_1\boldsymbol{\ell}_a$ is fixed), the optimal solution for the non-integer least squares problem becomes minimum norm solution [14], [15]

$$\tilde{\boldsymbol{\ell}}_v = (\mathbf{B}_2^H\mathbf{B}_2)^{-1}\mathbf{B}_2^H(\mathbf{r} - \mathbf{B}_1\boldsymbol{\ell}_a). \tag{18}$$

Plugging (18) into (16) and after some manipulation, we have

$$\begin{aligned}\gamma' &= \min_{\boldsymbol{\ell}_a\in\mathbb{U}^{N_a}} \|(\mathbf{I} - \mathbf{B}_2(\mathbf{B}_2^H\mathbf{B}_2)^{-1}\mathbf{B}_2^H)\mathbf{r} \\ &\quad -(\mathbf{I} - \mathbf{B}_2(\mathbf{B}_2^H\mathbf{B}_2)^{-1}\mathbf{B}_2^H)\mathbf{B}_1\boldsymbol{\ell}_a\|^2. \end{aligned} \tag{19}$$

Let $\boldsymbol{\Gamma} = \mathbf{I} - \mathbf{B}_2(\mathbf{B}_2^H\mathbf{B}_2)^{-1}\mathbf{B}_2^H$, then (19) becomes

$$\gamma' = \min_{\boldsymbol{\ell}_a\in\mathbb{U}^{N_a}} \|\boldsymbol{\Gamma}\mathbf{r} - \boldsymbol{\Gamma}\mathbf{B}_1\boldsymbol{\ell}_a\|^2. \tag{20}$$

Note that $\boldsymbol{\Gamma}\mathbf{r}$ is the projection of $\mathbf{r}$ onto the complement subspace of $T = \text{span}\{\mathbf{b}_{2,i}\}$ where $\mathbf{b}_{2,i}$ is the ith column vector of $\mathbf{B}_2$. $\boldsymbol{\Gamma}\mathbf{B}_1\boldsymbol{\ell}_a$ can be interpreted in a similar way. Since we search $\boldsymbol{\ell}_a$ after projecting $\mathbf{r}$ and $\mathbf{B}_1\boldsymbol{\ell}_a$ vectors into the complement subspace of T, the search space of the proposed method is reduced to N_a-dimensional integer lattice from the K-dimensional integer lattice ($N_a < K$). By letting $\bar{\mathbf{r}} = \boldsymbol{\Gamma}\mathbf{r}$ and $\bar{\mathbf{B}}_1 = \boldsymbol{\Gamma}\mathbf{B}_1$ in (20), we have

$$\gamma' = \min_{\boldsymbol{\ell}_a\in\mathbb{U}^{N_a}} \|\bar{\mathbf{r}} - \bar{\mathbf{B}}_1\boldsymbol{\ell}_a\|^2. \tag{21}$$

Since this is a standard ILSP, γ' and $\boldsymbol{\ell}_a$ can be obtained by the sphere encoding. Once $\boldsymbol{\ell}_a$ is obtained, by plugging $\boldsymbol{\ell}_a$ back to (18), $\tilde{\boldsymbol{\ell}}_v$ is attained. Following simple example illustrates basic idea of the proposed approach.

Example : Suppose the channel of standard vector perturbation is given by

$$\mathbf{H} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}, \tag{22}$$

and the symbol vector chosen from 2-PAM is $\mathbf{u} = [1\ \ -1]^T$, then γ' value obtained from (12) becomes 0.4396. Whereas, if

TABLE I
OPERATION OF THE PROPOSED METHOD

Input:	$\mathbf{H}$, $\mathbf{s}$
Output:	$\boldsymbol{\ell}$
Variable:	j denotes the user index
	K denotes the number of users
	ϵ denotes the given threshold
Step 1:	Compute $\gamma = \|\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I})^{-1}\tilde{\mathbf{s}}\|^2$ of the standard vector perturbation
Step 2:	Compute sum rate as $\sum_{j=1}^{K} R_j = \sum_{j=1}^{K} \log(1 + \text{SNR}_j)$ where $\text{SNR}_j = \frac{E\|s_j\|^2}{E\gamma\|n_j\|^2}$
Step 3:	If users are in outage ($R_j < \epsilon$)
	Select the virtual users as $\tilde{\mathbf{H}} = \arg\max_{\tilde{\mathbf{H}}_j} \lambda_{\min}^{(j)}$
	Compute $\gamma' = \|\tilde{\mathbf{H}}^H \left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1} \tilde{\mathbf{s}}\|^2$ with the virtual users
	Transmit $\mathbf{x} = \frac{1}{\sqrt{\gamma'}}\tilde{\mathbf{H}}^H \left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1}(\mathbf{s}' + \tau\boldsymbol{\ell}')$
	Else
	Transmit $\mathbf{x} = \frac{1}{\sqrt{\gamma}}\mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1}(\mathbf{s} + \tau\boldsymbol{\ell})$
	End

a virtual user is introduced, then the second row of the virtual channel matrix can be designed to minimize γ' as

$$\tilde{\mathbf{H}} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}, \tag{23}$$

and the symbol vector chosen from 4-PAM (for ensuring 2 bits/channel use for both cases) is $\mathbf{s}' = [\frac{-3}{\sqrt{5}} \quad 0]^T$. In this case, γ' value obtained from (20) becomes 0.3663, resulting in 20% reduction of original value. When 2-PAM is employed, γ' value becomes 0.2731, but the information rate is reduced to half.

To analyze benefit of the proposed method over the standard vector perturbation, we consider the error probability using the QPSK modulation. Denoting the error probability of the proposed method and standard vector perturbation as P_e' and P_e, respectively, then the relationship between two can be described as $P_e' = \beta P_e$ where β is the attenuation factor indicating the improvement of the error performance (the decibel gain in performance is $-10\log\beta$).

Lemma 1: The attenuation factor β is lower bounded by

$$\beta > \sqrt{\frac{2}{\pi}}\frac{\gamma'}{c}\left(1 - \frac{\gamma'}{c}\right)\exp\left(-\frac{c}{2\gamma'}(1 - \frac{\gamma'}{\gamma})\right) \tag{24}$$

where $c = \frac{2}{\sigma_n^2}$.

Proof: Recalling that the error probability of QPSK modulation is $Q(\sqrt{2\text{SNR}})$, P_e and P_e' become $Q\left(\sqrt{\frac{c}{\gamma}}\right)$ and $Q\left(\sqrt{\frac{c}{\gamma'}}\right)$, where $c = \frac{2}{\sigma_n^2}$. Using the properties of Q-function [16], we have

$$\begin{aligned} P_e &\leq \frac{1}{2}\exp\left(-\frac{c}{2\gamma}\right) & (25)\\ P_e' &> \frac{1}{\sqrt{2\pi}}\frac{\gamma'}{c}\left(1 - \frac{\gamma'}{c}\right)\exp\left(-\frac{c}{2\gamma'}\right). & (26) \end{aligned}$$

Using (25) and (26) together with $P_e' = \beta P_e$, we get (24). ■

This result states that the gain of the proposed method strongly depends on γ'. The overall operation of the proposed method is summarized in Table I.

D. Lower Bound of the Sum Rate Loss

The sum rate achieved by the standard vector perturbation is

$$R_{\text{org}} = \sum_{j=1}^{K} R_j = \sum_{j=1}^{K} \log(1 + \text{SNR}_j) \tag{27}$$

where $\text{SNR}_j = \frac{E\|s_j\|^2}{E\gamma\|n_j\|^2}$. Without loss of generality, we assume that $R_1 > \cdots > R_K$. Then the virtual user selection strategy searches at most $K - N_{a,\min}$ users in outage ($R_j < \epsilon$).

Lemma 2: For given $\epsilon > 0$, the sum rate of the proposed method satisfies

$$R_{\text{prop}} > R_{\text{org}} - (K - N_{a,\min})\epsilon. \tag{28}$$

Proof: It is clear from (27) that $R_{\text{org}} = \sum_{j=1}^{N_{a,\min}} R_j + \sum_{j=N_{a,\min}}^{K} R_j$. Also, the sum rate of the proposed method is $R_{\text{prop}} = \sum_{j=1}^{N_a} \log(1 + \text{SNR}_j')$ where $\text{SNR}_j' = \frac{E\|s_j\|^2}{E\gamma'\|n_j\|^2}$. Since $N_a \geq N_{a,\min}$ and $\gamma > \gamma'$, we have

$$\begin{aligned} R_{\text{prop}} &\geq \sum_{j=1}^{N_{a,\min}} R_j = R_{\text{org}} - \sum_{j=N_{a,\min}+1}^{K} R_j \\ &> R_{\text{org}} - (K - N_{a,\min})\epsilon. & (29) \end{aligned}$$

In particular, if $N_{a,\min} = K - 1$, then $R_{\text{prop}} > R_{\text{org}} - \epsilon$. ■

Note that since (28) holds for any channel realization, the lemma will also hold in ergodic sense (average over all possible channel realizations). Note also that due to the reduction of γ and sporadic use of virtual users, the sum rate loss will be smaller than $(K - N_{a,\min})\epsilon$ in practice.

E. Comments on Complexity

In this subsection, we compare the complexity of the proposed method and the standard vector perturbation. Note that since the modulo operation at the receiver is common, we focus only on computations associated with the transmitter precoding. While the major operation of the conventional approach is the sphere encoding to compute perturbation vector $\boldsymbol{\ell}$, user selection step and modified sphere encoding

TABLE II
COMPUTATIONAL COMPLEXITY (FLOPS) OF THE PROPOSED METHOD AND STANDARD VECTOR PERTURBATION (AVERAGE OVER 10^3 CHANNEL REALIZATIONS)

Operations	$N_v = 1$	$N_v = 2$
Standard VP for outage check	3015	3015
Virtual user selection	343	686
Modified sphere encoding via FSE	307	83
$\tilde{\boldsymbol{\ell}}_v$ computation	16	33
Proposed method	3681	3817
Standard VP	3015	

are additionally required for the proposed method. Denoting the complexity associated with user selection, $\tilde{\boldsymbol{\ell}}_v$ computation in (18), and the sphere encoding as C_{u}, $C_{\tilde{\boldsymbol{\ell}}_v}$, and C_{SE}, respectively, then the number of required floating-point operations (flops) for each step is as follows [17]:

- C_{u} requires $\frac{4}{3}M^3$ flops for inverting of $\tilde{\mathbf{H}}_j\tilde{\mathbf{H}}_j^H$ matrix, $k \times 2M^2 + 2M^2 + 4M$ flops for computing the minimum eigenvalue using an inverse power method (k is the iteration number). These operations need to be repeated for $\binom{N_c}{N_v}$ times for choosing N_v virtual users among N_c candidates.
- $C_{\tilde{\boldsymbol{\ell}}_v}$ requires $\frac{4}{3}N_v^3 + 2MN_v^2$ for computing $(\mathbf{B}_2^H\mathbf{B}_2)^{-1}$ and $2MN_v^2 + 2MN_v + 2MN_a + M$ for the matrix multiplication in (18).

In contrast to the operations we just described, the complexity associated with the perturbation vector optimization is hard to be quantified mainly because the computational complexity of the sphere encoding is non-deterministic. As a popular way to quantify the complexity, the average number of nodes visited has been considered [18], [19]. The lower bound on the complexity of the sphere encoding C_{SE} is [18]

$$C_{\text{SE}} \geq \frac{\beta^{\nu N_s} - 1}{\sqrt{\beta} - 1} \tag{30}$$

where β is a modulation order, N_s is the number of search dimension, and ν is the complexity component given by $\nu = \frac{1}{2}\left(1 + \frac{4(\beta-1)}{3\omega^2}\text{SNR}\right)$ where ω is a constant. Noting that C_{SE} increases exponentially with the search dimension N_s and N_s of the modified sphere encoding is smaller than that of the standard sphere encoding ($K \geq N_a$), we can argue that the complexity associated with the modified sphere encoding is smaller than that associated with the sphere encoding of the standard vector perturbation. However, since the proposed method requires two sphere encoding operations (standard sphere encoding in (12) and modified sphere encoding in (21)), the overall complexity (sum of two sphere encoding) of the proposed method is larger than the standard vector perturbation. Note that due to the variation of complexity and latency, hardware implementation of the sphere encoding might be a difficult task. As a way to get around the latency and complexity issue, the fixed-complexity sphere encoding (FSE) might be considered [20], [21]. This together with the fact that extra operations are required only when virtual users are selected due to the ϵ-outage, we expect that the additional computations of the proposed method are fairly moderate. Table II summarizes flops of the proposed method and standard vector perturbation. We observe that the additional complexity of the proposed method over the standard vector perturbation are 22% ($N_v = 1$) and 26% ($N_v = 2$), respectively. We also observe that the complexity of the sphere encoding with $N_v = 2$ is smaller than that with $N_v = 1$ since the search dimension is reduced from 3 to 2. However, due to the increase in complexity of user selection part, overall complexity for $N_v = 1$ and 2 is more or less similar.

F. Scheduling for fairness

An important issue to be considered, when user channels are heterogeneous (i.e, there are large variations in SNR among users) or some users are under stringent delay constraint, is to guarantee fairness among users. Note that if the user selection threshold is set to $\epsilon = \frac{1}{K}R_{\text{org}}$, all symbols of worst users are sacrificed for the performance improvement of active users so that no fairness is guaranteed for the sacrifice user at all. Whereas, if $\epsilon = 0$, the proposed method returns to the standard vector perturbation so that strict fairness is ensured among users via a simple round robin scheduling. As an approach making the compromise between the fairness and the sum rate, proportional fair scheduling (PFS) has been popularly considered in practice [22]. In each scheduling instant, the PFS computes the ratio of the instantaneous to the average rate for each user and then selects users associated with the maximal ratio ($k^*(t) = \arg\max_k \frac{R_k}{\bar{R}_k}$). Even in this case, since the supportable rate R_k for the PFS decision is computed using channel vectors $\mathbf{h}_k$ only, we can combine the proposed method and the PFS to improve performance of active users. As mentioned, since all parameters of sacrificing users (information data, perturbation vector, and virtual channel vector) can be arbitrarily chosen to improve the SINR of the system, considerable performance gain can be obtained with negligible impact on the fairness.

We summarize the modified PFS algorithm in Table III. Due to the fact that the virtual user decision is performed after the PFS scheduling, the weight μ_k of the sacrificing user will increase, thereby providing better chance for this user in the next scheduling instance.[3]

IV. EXTENSION TO MULTIUSER MIMO DOWNLINK SCENARIO

So far, we have considered multiuser MISO scenario where the worst user under deep fading is sacrificed to improve the performance of other users. We can readily extend the proposed method to the multiuser MIMO downlink scenario. Note that, since each user has multiple streams in this scenario, we do sacrifice worst *stream* instead of *user*. In the multiuser MIMO downlink system, a basestation equipped with N_t antennas transmits information to K users. Each user has $N_{r,j}$ receive antennas and thus the total number of receive antennas is $N_r = \sum_{j=1}^{K} N_{r,j}$. In the sequel, we will use the notation $\{N_{r,1}, \cdots, N_{r,K}\} \times N_t$ for this multiuser MIMO

[3]For ensuring the strict sense fairness, one might add the condition for the minimal service requirement.

TABLE III
MODIFIED PFS ALGORITHM

Initialization step	U denotes the set of all users S denotes the set of users to be selected
Main step	1) At time t, initialize the set of selected vectors $S = \phi$ 2) Perform the PFS algorithm [22] to select K users maximizing $\mu_k R_k$ among $\|U\|$ users 3) Apply the standard vector perturbation and proposed method to compute $R_{\text{org}}(S)$ and $R_{\text{prop}}(S)$ 4) If $R_{\text{prop}}(S) - (R_{\text{org}}(S) - \min_j R_{j,\text{org}}) > \epsilon$ $S = S \backslash \{\kappa\}$ where $\kappa = \arg\min_j R_{j,\text{org}}$ Transmit $\mathbf{x} = \frac{1}{\sqrt{\gamma'}} \tilde{\mathbf{H}}^H \left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right)^{-1} (\mathbf{s}' + \tau \boldsymbol{\ell}')$ Else Transmit $\mathbf{x} = \frac{1}{\sqrt{\gamma}} \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1} (\mathbf{s} + \tau \boldsymbol{\ell})$ End 5) Update the weights which are the inverse of the time-averaged past throughput for all users $k \in U$ $\frac{1}{\mu_k(t+1)} = \left(1 - \frac{1}{N_s}\right) \frac{1}{\mu_k(t)} + \frac{R_k(S)}{N_s}$ for $k \in S$ $\frac{1}{\mu_k(t+1)} = \left(1 - \frac{1}{N_s}\right) \frac{1}{\mu_k(t)}$ for $k \notin S$

system. Denoting $\mathbf{T}_j$ as the precoding matrix, the received signal at the jth user becomes

$$\mathbf{y}_j = \mathbf{H}_j \sum_{j=1}^{K} \mathbf{T}_j \mathbf{x}_j + \mathbf{n}_j \tag{31}$$

where $\mathbf{H}_j$ is the $N_{r,j} \times N_t$ channel matrix from the transmitter to the jth user, $\mathbf{x}_j$ is the $N_{r,j} \times 1$ transmit signal vector of user j, and $\mathbf{n}_j$ is the normalized complex Gaussian noise vector with entries distributed according to $\mathcal{CN}(0,1)$.

As a well-known approach for converting the multiuser MIMO channel into the single user MIMO channel, the block diagonalization (BD) algorithm [23] can be employed. The main idea of the BD is to employ the beamforming vectors annihilating all multiuser interference. The precoding constraint to remove multiuser interference is

$$\mathbf{H}_i \mathbf{T}_j = \mathbf{0} \text{ for } i = 1, \cdots, j-1, j+1, \cdots, K. \tag{32}$$

Clearly, $\mathbf{T}_j$ should be a nullspace of $\tilde{\mathbf{H}}_j = [\mathbf{H}_1^T \cdots \mathbf{H}_{j-1}^T \; \mathbf{H}_{j+1}^T \cdots \mathbf{H}_K^T]^T$. If we denote the SVD of $\tilde{\mathbf{H}}_j$ as $\tilde{\mathbf{H}}_j = \tilde{\mathbf{U}}_j \tilde{\mathbf{\Sigma}}_j [\tilde{\mathbf{V}}_j^{(1)} \; \tilde{\mathbf{V}}_j^{(0)}]^H$, then the right singular vectors $\tilde{\mathbf{V}}_j^{(0)}$ form an orthogonal basis for the null space of $\tilde{\mathbf{H}}_j$. Thus, $\tilde{\mathbf{V}}_j^{(0)}$ becomes the right choice for the precoding matrix $\mathbf{T}_j$ and (31) becomes

$$\mathbf{y}_j = \mathbf{H}_{\text{eff},j} \mathbf{x}_j + \mathbf{n}_j = \mathbf{H}_j \mathbf{T}_j \mathbf{x}_j + \mathbf{n}_j = \mathbf{H}_j \tilde{\mathbf{V}}_j^{(0)} \mathbf{x}_j + \mathbf{n}_j \tag{33}$$

where $\mathbf{H}_{\text{eff},j}$ is the effective channel matrix for the jth user. By precoding with $\tilde{\mathbf{V}}_j^{(0)}$, the multiuser interference is completely removed and thus the jth user experiences a point-to-point MIMO link [23].

After the BD preprocessing, the vector perturbation technique can be applied to compensate for the transmit power suppression caused by the zero forcing operation of the BD. Since it is straightforward to apply the vector perturbation technique for the system in (33), we skip the details. The optimization function of the jth user is

$$\boldsymbol{\ell}_j = \arg \min_{\boldsymbol{\ell}'_j \in \mathbb{Z}^{2N_{r,j}}} \left\| \mathbf{H}_{\text{eff},j}^{-1} (\mathbf{s}_j + \tau \boldsymbol{\ell}'_j) \right\|^2 \tag{34}$$

where $\mathbb{Z}$ is the integer set and $N_{r,j}$ is the number of transmit streams for each user. The achievable rate of this MU-MIMO scheme is [24], [25]

$$\begin{aligned} R_{\text{org-MIMO}} &= \sum_{j=1}^{K} R_{j,\text{org}} = \sum_{j=1}^{K} \sum_{k=1}^{N_{r,j}} \log_2 (1 + \text{SNR}_{j,k}) \\ &= \sum_{j=1}^{K} \sum_{k=1}^{N_{r,j}} \left(1 + \frac{\rho \lambda_k^2}{N_{r,j}}\right) \end{aligned} \tag{35}$$

where $\rho = \frac{P_T}{\sigma_n^2}$ and λ_k $(k = 1, \cdots, N_{r,j})$ is the singular value of $\mathbf{H}_{\text{eff},j}$.

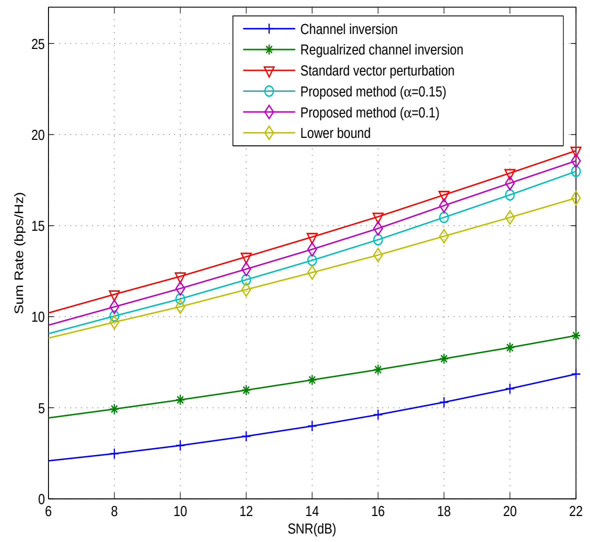
Same as the MISO scenario, we can employ the proposed scheme for the selection of virtual streams when effective channels of certain streams are in outage. Since the system is decomposed into K-independent single user MIMO systems depending on the QoS of each user, we might use separate outage threshold ϵ_j for each user's effective channel matrix $\tilde{\mathbf{H}}_{\text{eff},j}$. The achievable sum rate for the proposed MU-MIMO scheme is given by (see Appendix A)

$$R_{\text{prop-MIMO}} = \sum_{j=1}^{K} \sum_{l=1}^{N_a} \log \left(1 + \frac{\rho \xi_l^2}{\min_{\xi_m} \left(\sum_{m=1}^{N_{r,j}} \mu_m^2 \xi_m^2\right)}\right). \tag{36}$$

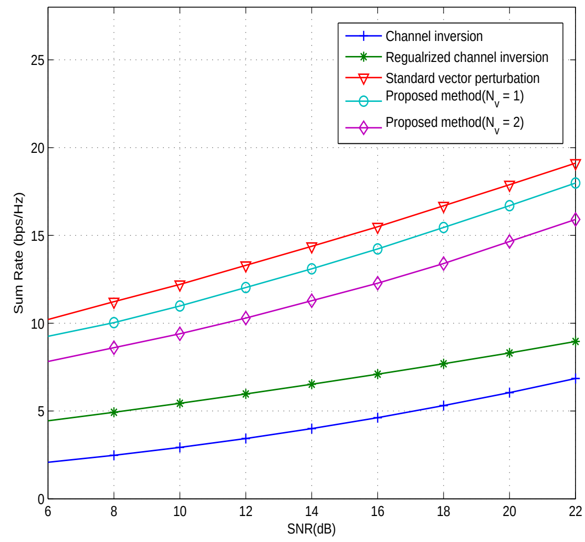
V. SIMULATION RESULTS AND DISCUSSIONS

In this section, we evaluate the sum rate and bit error rate (BER) performance of the multiuser downlink systems (both MISO and MIMO setup) through numerical simulations. We consider the multiuser downlink channel with Rayleigh fading ($h_{ij} \sim CN(0,1)$).

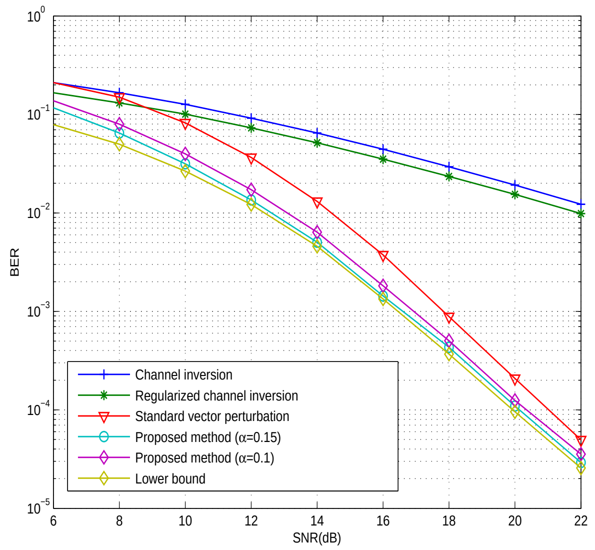
We first consider 4×4 MISO system with QPSK modulation where one out of four users can help improving performance of others ($K = 4$, $N_{a,\min} = 3$). For comparison, we employ channel inversion, regularized channel inversion, as well as standard vector perturbation. In Fig. 3(a), we compare the sum rate of the proposed method and standard vector perturbation, averaged over channel realizations, as a function of the SNR. In this simulation, we set the outage threshold of the proposed method to $\epsilon = 0.15 R_{\text{org}}$. As can be seen from



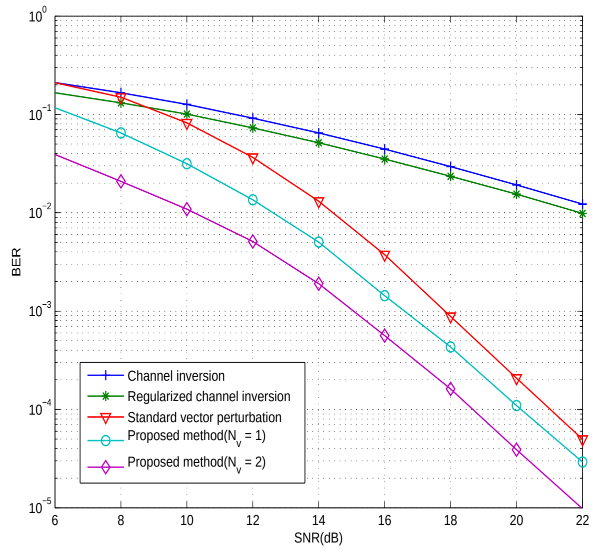
(a)



(a)



(b)



(b)

Fig. 3. Performance of 4×4 multiuser downlink system ($\epsilon = \alpha R_{\text{org}}$): (a) Sum rate and (b) BER.

Fig. 4. Performance of 4×4 multiuser downlink system with N_v variation ($\epsilon = 0.15 R_{\text{org}}$): (a) Sum rate and (b) BER.

the figure, the sum rate difference between two approaches is small and roughly constant. Therefore, as the SNR increases, percent loss ($\frac{R_{\text{org}} - R_{\text{prop}}}{R_{\text{org}}}$) of the sum rate decreases, resulting in about 12% loss at 20 dB SNR. In Fig. 3(b), we compare the BER performance of the proposed method along with standard vector perturbation and others. The performance gain, achieved at the expense of the sum rate loss, is noticeable so that the gain at 10^{-2} BER is around 2.5 dB. We also observe that the lower bound obtained by (24) is tight for mid and high SNR regime so that we can easily predict the achievable performance gain in the design stage. Overall, we observe that the proposed method with $\alpha = 0.1$ achieves higher sum rate than that with $\alpha = 0.15$, resulting in 7% gain at 20 dB SNR at the expense of 0.5 dB loss in performance.

In Fig. 4, we plot the sum rate and BER performance of the proposed schemes for $N_v = 1$ and 2. As expected, the performance of the proposed method with $N_v = 2$ outperforms that with $N_v = 1$ and standard vector perturbation, resulting in about 1.5 dB and 3 dB gain at BER $= 10^{-3}$, respectively. Since the performance trades off the sum rate, $N_v = 2$ scenario incurs additional 15% loss in the sum rate.

So far, we have assumed that the base station has knowledge of full channel state information (CSI). In Fig. 5, we investigate the performance of the proposed method along with the

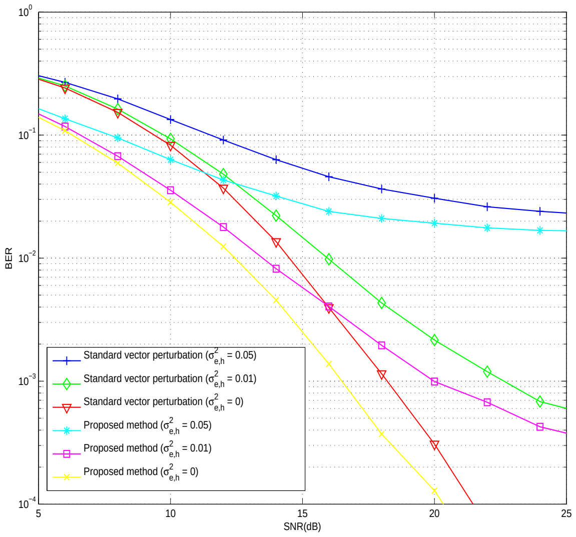


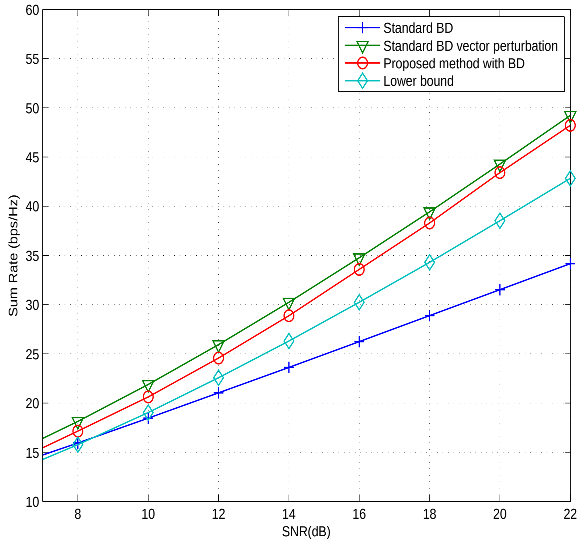
Fig. 5. Performance of 4×4 multiuser downlink system with various $\sigma^2_{e,h}$.

standard vector perturbation in the presence of channel estimation error. Since the mismatch between the actual CSI and the estimated CSI is unavoidable in a real communication, and this might result in degradation in performance, it is of importance to investigate the effect of channel estimation error. In our simulation, we use an additive channel estimation error model where $\mathbf{H} = \mathbf{H}_{\text{est}} + \mathbf{H}_{\text{err}}$ where $\mathbf{H}$, $\mathbf{H}_{\text{est}}$ and $\mathbf{H}_{\text{err}}$ represent the true channel matrix, the estimated channel matrix and the estimated error matrix, respectively. We assume that $\mathbf{H}_{\text{err}}$ is uncorrelated with $\mathbf{H}_{\text{est}}$ and $\mathbf{s}$, and $\mathbf{H}_{\text{err}}$ has i.i.d. elements with zero mean and the estimation error variance $\sigma^2_{e,h}$. In this setup, it is clear that the estimation error becomes dominant factor limiting performance improvement in high SNR. Although both schemes show error floors when $\sigma^2_{e,h} = 0.05$, we observe that the proposed scheme is more robust to the estimation errors than the standard vector perturbation. For example, the performance gain at 10^{-3} BER of the proposed scheme is about 3 dB over the standard vector perturbation when $\sigma^2_{e,h} = 0.01$, while the gain is around 2 dB for $\sigma^2_{e,h} = 0$.

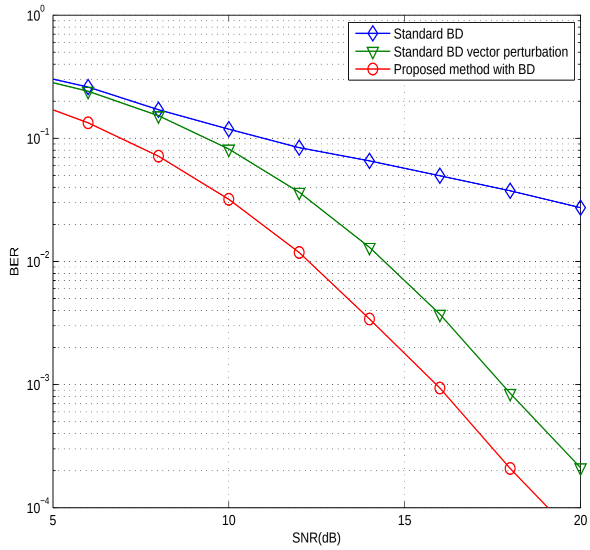
Finally, we consider the $\{4, 4\} \times 8$ MU-MIMO system with the QPSK modulation. As mentioned, the multiuser system is converted into the single user MIMO system by the BD operation so that we might need an individual outage thresholds ϵ_1 and ϵ_2. For simplicity, we set $\epsilon_1 = \epsilon_2 = 0.15R_{1,\text{org}}$ in our simulation and hence the lower bound of the sum rate becomes $\sum_{j=1}^{2}(R_{j,\text{org}} - (N_{r,j} - N_{a,\min})\epsilon_1)$. We observe form the Fig. 6(a) that the sum rate difference between two approaches is moderate and further decreases as SNR increases. Whereas, as shown in Fig. 6(b), the proposed method brings considerable gain over the BD and the BD with vector perturbation, resulting in 8 dB and 2.5 dB gain at 10^{-2} BER.

VI. Conclusion

In this work, we investigated an approach achieving robustness of multiuser downlink in the vector perturbation.



(a)



(b)

Fig. 6. Performance of $\{4, 4\} \times 8$ multiuser downlink system ($\epsilon_1 = \epsilon_2 = 0.15R_{1,\text{org}}$): (a) Sum rate and (b) BER.

Motivated by the fact that whole mobile users in a link suffer performance degradation since γ of the vector perturbation affects SNR of all users alike, we sacrificed small part of users in outage to improve the bit error performance of the rest. This can eventually strengthen the ability to meet the QoS of the practical mobile communication system. The benefit of the proposed method, on top of the enhanced reliability, lies in the point that the modification of transmit precoding does not affect the receiver operation at all and the additional complexity of the proposed method at the transmitter is fairly moderate. Future research direction may include the complexity reduction of the transmission precoding and application of the proposed method into the codebook based systems.

Appendix A
Achievable rate analysis of the proposed MU-MIMO scheme

Recall that the received signal $\mathbf{y}_j$ is given by $\mathbf{y}_j = \mathbf{H}_{\text{eff},j}\mathbf{x}_j + \mathbf{n}_j$ where $\mathbf{x}_j = \frac{1}{\sqrt{\gamma_j'}}\tilde{\mathbf{H}}_{\text{eff},j}^{-1}\tilde{\mathbf{s}}_j$. Define $\tilde{\mathbf{H}}_{\text{eff},j} = \mathbf{U}_j\mathbf{\Lambda}_j\mathbf{V}_j^H$ where $\mathbf{U}_j = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_{N_{r,j}}]$, $\mathbf{V}_j = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_{N_{r,j}}]$, and $\mathbf{\Lambda} = \text{diag}(\lambda_1 \ \cdots \ \lambda_{N_{r,j}})$. The normalized scalar factor γ_j' is given by

$$\begin{aligned} \gamma_j' &= \|\tilde{\mathbf{H}}_{\text{eff},j}^{-1}\tilde{\mathbf{s}}_j\|^2 = \tilde{\mathbf{s}}_j^H\left(\tilde{\mathbf{H}}_{\text{eff},j}\tilde{\mathbf{H}}_{\text{eff},j}^H\right)^{-1}\tilde{\mathbf{s}}_j \\ &= \tilde{\mathbf{s}}_j^H\mathbf{U}_j\Lambda_j^{-2}\mathbf{U}_j^H\tilde{\mathbf{s}}_j = \sum_{l=1}^{N_{r,j}}\mu_l^2\xi_l^2, \end{aligned} \tag{37}$$

where $\mu_l = \frac{1}{\lambda_l} > 0$ and $\xi_l = |\mathbf{u}_l^H\tilde{\mathbf{s}}_j|$.

The received signal-to-noise-ratio (SNR) of each stream SNR_l can be represented by $\text{SNR}_l = \frac{\rho\xi_l^2}{\gamma'} = \frac{\rho\xi_l^2}{\sum_{m=1}^{N_{r,j}}\mu_m^2\xi_m^2}$ where $\rho = \frac{P_T}{\sigma_n^2}$. Thus, the achievable rate of the jth user R_j is given by

$$R_j = \sum_{l=1}^{N_a}\log(1+\text{SNR}_l) = \sum_{l=1}^{N_a}\log\left(1+\frac{\rho\xi_l^2}{\sum_{m=1}^{N_{r,j}}\mu_m^2\xi_m^2}\right). \tag{38}$$

Note that the vector precoding is not applied in (38). The effect of vector perturbation is to force $\boldsymbol{\ell}_j$ to minimize γ_j' so that $\tilde{\mathbf{s}}_j$ can be (coarsely) oriented in the coordinate system defined by $\mathbf{u}_1, \cdots, \mathbf{u}_{N_{r,j}}$ [7].

By searching the proper precoding vector and controlling ξ_m to minimize γ_j', we obtain an upper bound on achievable rate of the proposed scheme for each user as

$$R_{j,\text{prop}} = \sum_{l=1}^{N_a}\log\left(1+\frac{\rho\xi_l^2}{\min_{\xi_m}\left(\sum_{m=1}^{N_{r,j}}\mu_m^2\xi_m^2\right)}\right). \tag{39}$$

From (39), we need to solve for $\xi_l = \arg\min_{\xi_m}\sum_{l=1}^{N_{r,j}}\mu_m^2\xi_m^2$.

By the arithmetic-geometric mean inequality, solution occurs when $\mu_1^2\xi_1^2 = \cdots = \mu_{N_{r,j}}^2\xi_{N_{r,j}}^2 = \omega_0^2$ for an arbitrary constant $\omega_0^2 > 0$. Note that $\xi_l = |\mathbf{u}_l^H\tilde{\mathbf{s}}_j|$ cannot be zero since τ is chosen large enough. Therefore, $R_{j,\text{prop}}$ can be represented by

$$\begin{aligned} R_{j,\text{prop}} &= \sum_{l=1}^{N_a}\log\left(1+\frac{\rho\frac{\omega_0^2}{\mu_l^2}}{N_{r,j}\omega_0^2}\right) = \sum_{l=1}^{N_a}\log\left(1+\frac{\rho}{\mu_l^2 N_{r,j}}\right) \\ &= \sum_{l=1}^{N_a}\log\left(1+\frac{\rho\lambda_l^2}{N_{r,j}}\right). \end{aligned} \tag{40}$$

We obtain (40) by substituting μ_l for $\frac{1}{\lambda_l}$. In summary, the upper bound of achievable sum rate for the proposed scheme becomes

$$R_{\text{prop-MIMO}} = \sum_{j=1}^{K} R_{j,\text{prop}}. \tag{41}$$

References

[1] B. Lee and B. Shim, “A vector perturbation based user selection for multi-antenna downlink channels,” Information Theory and Applications Workshop (ITA), 2011.

[2] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. 29, pp. 439-441, May 1983.

[3] H. Weingarten, Y. Steinberg, and S. Shamai, “The capacity region of the Gaussian MIMO broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936-3964, Sept. 2006.

[4] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658-2668, Oct. 2003.

[5] P. Viswanath and D. N. C. Tse, “Sum capacity of the vector gaussian broadcast channel and uplink downlink duality,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 1912-1921, Aug. 2003.

[6] 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.

[7] B. M. Hochwald, C. B. Peel, and A. Lee Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation,” *IEEE Trans. Commun.*, vol. 53, pp. 537-544, March 2005.

[8] L. Zheng and D. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073-1096, May 2003.

[9] C. B. Peel, B. M. Hochwald, and A. Lee 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.

[10] A. Edelman, “Eigenvalues and condition numbers of random matrices”, SIAM J. Matrix Anal. and Applic. 9, pp. 543-560, 1988.

[11] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms*, Prentice-Hall, 2000.

[12] I. Dimov, V. Alexandrov, and A. Karaivanova, “Resolvent monte carlo methods for linear algebra problems,” *Mathematics and Computers in Simulations*, vol. 55, pp. 25-36, 2001.

[13] C. Yuen and B. M. Hochwald, “How to gain 1.5dB in vector precoding,” in Proc. *GLOBECOM*, Nov. 2006.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[15] W. S. Chua, C. Yuen, and F. Chin, ”A continuous vector-perturbation for multi-antenna multi-user communication”, *VTC-Spring*, Dublin, Ireland, 2007.

[16] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

[17] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1989.

[18] J. Jalden and B. Ottersten, “On the complexity of sphere decoding in digital communication,” *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

[19] E. Zimmerman, W. Rave and G. Fettweis, “On the complexity of sphere decoding,” *Proc. of WPMC*, Sept. 2004, pp. 12-15.

[20] M. Mohaisen and K. Chang, “Fixed-complexity sphere encoder for multi-user MIMO systems,” *Journal of Communications and Networks (JCN)*, vol. 13, pp. 63-69, Feb. 2011.

[21] J. Choi, B. Shim, A. Singer, and N. Cho, “Low-complexity decoding via reduced dimension maximum-likelihood search,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1780-1793, Mar. 2010.

[22] A. Jalali, R. Padovani, and R. Pankai, “Data throughput of CDMA HDR a high efficiency-high data rate personal communication wireless system,” *Proc. of VTC*, May 2000, pp.1854-1858.

[23] 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.

[24] J. Park, B. Lee, and B. Shim, “A mmse vector precoding with block diagonalziation for multiuser MIMO downlink,” *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 569-577, Feb. 2012.

[25] C. B. Chae, S. Shim, and R. W. Heath Jr., “Block diagonalized vector perturbation for multiuser MIMO systems,” *IEEE Trans. Wireless Comm.*, vol. 7, pp. 4051-4057, Nov. 2008.

PLACE
PHOTO
HERE

Byungju Lee (S' 09) received the B.S. degree in the School of Information and Communication, Korea University, Seoul, Korea, in 2008, where he is currently working toward the Ph.D. degree. His research interests include wireless communications and network information theory.

PLACE
PHOTO
HERE

Byonghyo Shim (SM' 09) received the B.S. and M.S. degrees in control and instrumentation engineering (currently electrical engineering) from Seoul National University, Korea, in 1995 and 1997, respectively, and the M.S. degree in mathematics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, in 2004 and 2005, respectively.

From 1997 and 2000, he was with the Department of Electronics Engineering at the Korean Air Force Academy as an Officer (First Lieutenant) and an Academic Full-time Instructor. From 2005 to 2007, he was with Qualcomm Inc., San Diego, CA, as a Staff Member. Since September 2007, he has been with the School of Information and Communication, Korea University, where he is now an Associate Professor. His research interests include signal processing for wireless communications, compressive sensing, applied linear algebra, and information theory.

Dr. Shim was the recipient of 2005 M. E. Van Valkenburg research award from ECE department of University of Illinois and 2010 Haedong award from IEEK.