This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2008 proceedings.

Generalization of Channel Inversion Algorithms for Multiuser MIMO Downlink Systems

Hakjea Sung, Sang-Rim Lee and Inkyu Lee
School of Electrical Eng., Korea University, Seoul, Korea
Email: {jaysung, inkyu}@korea.ac.kr and srlee@wireless.korea.ac.kr

***Abstract*— Recently a number of transmission schemes have been introduced to achieve sum capacity for multiuser multi-input multi-output (MIMO) broadcast channels (BC). A block diagonalization (BD) is an attractive method which operates only a few dB away from the sum capacity. This scheme is a generalization of the zero-forcing channel inversion to the case where each receiver is equipped with multiple antennas. One of the limitation of the BD is that the sum rate does not grow linearly with the number of users due to the noise enhancement. In this paper, we propose a generalized minimum mean-squared error (MMSE) channel inversion algorithm for users with multiple antennas to overcome the drawbacks of the BD for multiuser MIMO systems. Simulation results confirm that the proposed scheme achieves performance improvement over the conventional BD scheme. Also, we present a precoding method for systems with channel estimation errors and show that the proposed algorithm is robust to the channel estimation errors.**

I. Introduction

Multi-input multi-output (MIMO) systems have drawn a lot of attention in the past few years due to their great potential to achieve high throughput in wireless communication systems [1]. It is well known that in a single user case the capacity scales linearly with the minimum number of transmit and receive antennas in Rayleigh fading channels [2]. More recently, the investigation of the capacity region has been of concern in multiuser MIMO broadcast channels (BC), where each user has possibly multiple receive antennas [3][4][5].

In [6], it was shown that the maximum sum rate in multiuser MIMO BC can be achieved by using dirty paper coding (DPC), the DPC is difficult to implement in practical systems.

For linear processing systems where the base station has multiple antennas but all users employ a single antenna, several practical precoding techniques have been proposed [7]. A zero-forcing channel inversion (ZF-CI) [7] is one of the simplest precoding techniques for this case. However its performance is rather poor for all signal-to-noise-ratios (SNRs) due to a transmit power boost issue. Although a minimum mean-squared error channel inversion (MMSE-CI) method [7] overcomes the drawback of the ZF-CI, this is still confined to a single receive antenna case.

This work was supported in part by the Ministry of Information and Communication (MIC), Korea, under the Information Technology Research Center (ITRC) support program, supervised by the Institute of Information Technology Assessment (IITA) and in part by grant No. R01-2006-000-11112-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

For the case where the users in the network have multiple antennas, a block diagonalization (BD) algorithm, which can be considered as a generalization of the ZF-CI, is a well-known precoding algorithm for this case [8]. The key idea of the BD is to eliminate the multi-user interference (MUI) by placing all the unintended users at nullspace of the intended user's channels. As it attempts to completely eliminate the MUI without any consideration on the noise, the BD is inferior to the DPC in terms of sum capacity especially for the case of a large number of users.

In this paper, we propose a generalized MMSE-CI (GMI) algorithm which supports multiple data stream transmission to each user in multiuser MIMO BC based on a noniterative method. Unlike the conventional BD algorithm, our GMI algorithm takes the noise into account for finding each user's precoding matrix to increase the signal-to-interference-plus-noise ratio (SINR) at each user's receiver. In addition, we introduce a design method when the transmitter has incomplete channel state information (CSI).

The following notations are used throughout the paper. We employ uppercase boldface letters for matrices and lowercase boldface for vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ and $\mathbf{A}^H$ denote the transpose and the conjugate transpose, respectively. $\mathrm{Tr}(\mathbf{A})$ indicates the trace and the Frobenius norm of matrix $\mathbf{A}$ is $\|\mathbf{A}\|_F^2 = \mathrm{Tr}(\mathbf{A}\mathbf{A}^H)$. For $m \times m$ matrices $\mathbf{A}_j$, $\mathbf{A} =$ diag $\{\mathbf{A}_1 \cdots \mathbf{A}_n\}$ denotes an $mn \times mn$ block diagonal matrix.

II. System Model

We consider multiuser MIMO downlink systems where the base station is transmitting to K independent users simultaneously. In this system, the base station is equipped with N_t transmit antennas and user j has $n_j \geq 1$ receive antennas, referred to in the following as $\{n_1, \cdots, n_K\} \times N_t$. The total number of receive antennas at all users is defined as $N_r = \sum_{j=1}^{K} n_j$. In the discrete-time complex baseband MIMO case, the channel from the base station to the jth user is modeled by the $n_j \times N_t$ channel matrix $\mathbf{H}_j$. We assume that $\mathbf{H}_j$ has full row rank and independently and identically distributed (i.i.d.) entries according to $N_c(0, 1)$. Also, we assume that the full rank channel matrix $\mathbf{H}_s = [\mathbf{H}_1^T\ \mathbf{H}_2^T \cdots \mathbf{H}_K^T]^T$ is known perfectly at the base station.

We define the transmitted data symbol vector $\mathbf{s}_s$, the noise vector $\mathbf{w}_s$ and the precoding matrix $\mathbf{P}_s$ for all users as

$$\begin{aligned}\mathbf{s}_s &= [\mathbf{s}_1^T \quad \mathbf{s}_2^T \quad \cdots \quad \mathbf{s}_K^T]^T,\\ \mathbf{w}_s &= [\mathbf{w}_1^T \quad \mathbf{w}_2^T \quad \cdots \quad \mathbf{w}_K^T]^T,\\ \mathbf{P}_s &= [\mathbf{P}_1 \quad \mathbf{P}_2 \quad \cdots \quad \mathbf{P}_K]\end{aligned}$$

where $\mathbf{s}_j = [s_{j,1} \cdots s_{j,n_j}]^T \in C^{n_j}$ and $\mathbf{w}_j = [w_{j,1} \cdots w_{j,n_j}]^T \in C^{n_j}$ are the jth user's data and noise vectors, respectively and $\mathbf{P}_j$ represents the associated precoding matrix. Here the symbols $s_{j,i}$ are assumed to be independently generated with unit variance and the components $w_{j,i}$ of the noise vector $\mathbf{w}_j$ are i.i.d. with zero mean and variance σ_w^2. Then, the total received signal can be expressed as

$$\mathbf{y}_s = \mathbf{H}_s\mathbf{P}_s\mathbf{s}_s + \mathbf{w}_s$$

and the received signal vector at the jth user is given by

$$\mathbf{y}_j = \mathbf{H}_j\mathbf{P}_j\mathbf{s}_j + \mathbf{H}_j\sum_{k\neq j}^{K}\mathbf{P}_k\mathbf{s}_k + \mathbf{w}_j \qquad (1)$$

where $\mathbf{y}_j$ denotes the corresponding received signal vector as $\mathbf{y}_j = [y_{j,1} \cdots y_{j,n_j}]^T \in C^{n_j}$. Let $\tilde{\mathbf{s}}_s = \mathbf{P}_s\mathbf{s}_s$ denote the signal vector actually transmitted at the base station, it satisfies the total power constraint $\mathbb{E}[\|\tilde{\mathbf{s}}_s\|^2] \leq P_{total}$.

Denoting the overall receive filter $\mathbf{M}_s$ as

$$\mathbf{M}_s = \text{diag}\{\mathbf{M}_1 \quad \mathbf{M}_2 \quad \cdots \quad \mathbf{M}_K\}$$

where $\mathbf{M}_j$ represents the jth user's receive filter, the receive filter output vector of the jth user $\mathbf{x}_j$ can be written as

$$\mathbf{x}_j = \mathbf{M}_j\mathbf{H}_j\mathbf{P}_j\mathbf{s}_j + \mathbf{M}_j\mathbf{H}_j\sum_{k\neq j}^{K}\mathbf{P}_k\mathbf{s}_k + \mathbf{M}_j\mathbf{w}_j \qquad (2)$$

where $\mathbf{x}_j = [x_{j,1} \cdots x_{j,n_j}]^T \in C^{n_j}$.

III. Generalization of Zero-Forcing Channel Inversion

In this section, we represent an alternative way of representing the conventional BD [8] by extending the ZF-CI method [7] for the case where each user has more than a single antenna. This new ZF-CI method achieves the performance equivalent to the conventional BD scheme [8] with reduced complexity. This will be referred to as the generalized zero-forcing CI (GZI).

A. Generalized Zero-Forcing Channel Inversion

The key idea of the GZI algorithm is to identify the precoding matrix $\mathbf{P}_s$ which make all MUI zero. To eliminate all the MUI, we impose a constraint such that

$$\mathbf{H}_j\mathbf{P}_k = \mathbf{0} \quad \text{for all} \quad j \neq k \quad \text{and} \quad 1 \leq j,\, k \leq K. \qquad (3)$$

In order to compute the orthogonal vectors of $\tilde{\mathbf{H}}_j$, we define the pseudo-invserse of the channel matrix $\mathbf{H}_s$ as

$$\hat{\mathbf{H}}_s = \mathbf{H}_s^H(\mathbf{H}_s\mathbf{H}_s^H)^{-1} = [\hat{\mathbf{H}}_1\ \hat{\mathbf{H}}_2 \cdots \hat{\mathbf{H}}_K]. \qquad (4)$$

Consider the QR decomposition of $\hat{\mathbf{H}}_j$ as

$$\hat{\mathbf{H}}_j = [\hat{\mathbf{Q}}_j^{(1)}\ \ \hat{\mathbf{Q}}_j^{(0)}]\begin{bmatrix}\hat{\mathbf{R}}_j^{(1)}\\ \mathbf{0}\end{bmatrix} = \hat{\mathbf{Q}}_j^{(1)}\hat{\mathbf{R}}_j^{(1)} \qquad (5)$$

where $\hat{\mathbf{R}}_j^{(1)}$ is an $n_j \times n_j$ upper triangular matrix and $\hat{\mathbf{Q}}_j^{(1)}$ is an $N_t \times n_j$ matrix of which columns form an orthonormal basis for $\hat{\mathbf{H}}_j$. Since we obtain $\hat{\mathbf{H}}_j$ from the inverse operation, we have $\tilde{\mathbf{H}}_j\hat{\mathbf{Q}}_j^{(1)}\hat{\mathbf{R}}_j^{(1)} = \mathbf{0}$ from (5). Thus, it follows $\tilde{\mathbf{H}}_j\hat{\mathbf{Q}}_j^{(1)} = \mathbf{0}$. Here, we can see that the columns of $\hat{\mathbf{Q}}_j^{(1)}$ form an orthonormal basis for nullspace of $\tilde{\mathbf{H}}_j$ so that the jth user's precoder of the GZI which is constructed by a linear combination of $\hat{\mathbf{Q}}_j^{(1)}$ is also satisfy the zero MUI constraint in (3). If each user has a single receive antenna, this reduces to the ZF-CI solution with unit norm precoding vectors.

By multiplying the matrix $\hat{\mathbf{Q}}_j^{(1)}$ to the channel matrix $\mathbf{H}_s$, the jth user has the non-interfering block channel $\mathbf{H}_j\hat{\mathbf{Q}}_j^{(1)}$ for $j = 1, \cdots, K$. In order to decompose the block channel $\mathbf{H}_j\hat{\mathbf{Q}}_j^{(1)}$ into parallel subchannels, we now apply the SVD operation of $\mathbf{H}_j\hat{\mathbf{Q}}_j^{(1)}$ as

$$\mathbf{H}_j\hat{\mathbf{Q}}_j^{(1)} = \mathbf{U}_j^{(\mathrm{q})}\mathbf{\Lambda}_j^{(\mathrm{q})}[\mathbf{V}_j^{(\mathrm{q}(1))}\ \ \mathbf{V}_j^{(\mathrm{q}(0))}]^H$$

where $\mathbf{V}_j^{(\mathrm{q}(1))}$ denotes the set of right singular vectors corresponding to non-zero singular values and $\mathbf{U}_j^{(\mathrm{q})}$ is the left singular matrix. Defining $\mathbf{P}_j$ and $\mathbf{M}_j$ as $\mathbf{P}_j = \hat{\mathbf{Q}}_j^{(1)}\mathbf{V}_j^{(\mathrm{q}(1))}$ and $\mathbf{M}_j = \mathbf{U}_j^{(\mathrm{q})\mathrm{H}}$, the jth user's receive filter output vector of the GZI scheme $\mathbf{x}_j$ in (2) is given as

$$\mathbf{x}_j = \mathbf{\Lambda}_j^{(\mathrm{q})}\mathbf{s}_j + \mathbf{U}_j^{(\mathrm{q})\mathrm{H}}\mathbf{w}_j. \qquad (6)$$

Let us denote $\mathbf{\Phi}_j$ as the power allocation matrix for the jth user, the achievable sum rate of the GZI can be expressed as

$$\begin{aligned}R_{\mathrm{GZI}} &= \max_{\mathbf{\Phi}_j}\sum_{j=1}^{K}\log_2\det\left(\mathbf{I} + \frac{(\mathbf{\Lambda}_j^{(\mathrm{q})})^2\mathbf{\Phi}_j}{\sigma_w^2}\right)\\ &\qquad \text{subject to}\sum_{j=1}^{K}\mathrm{Tr}(\mathbf{\Phi}_j) \leq P_{total} \end{aligned}\qquad (7)$$

and the optimal power loading matrix $\mathbf{\Phi}_j$ can be calculated by using the water-filling (WF) method [9]. Finally, the precoding matrix and the overall receive filter are defined as

$$\begin{aligned}\mathbf{P}_s^{\mathrm{GZI}} &= [\hat{\mathbf{Q}}_1^{(1)}\mathbf{V}_1^{(\mathrm{q}(1))}\ \hat{\mathbf{Q}}_2^{(1)}\mathbf{V}_2^{(\mathrm{q}(1))}\cdots\hat{\mathbf{Q}}_K^{(1)}\mathbf{V}_K^{(\mathrm{q}(1))}]\mathbf{\Phi}^{\frac{1}{2}},\\ \mathbf{M}_s^{\mathrm{GZI}} &= \text{diag}\{\mathbf{U}_1^{(\mathrm{q})\mathrm{H}} \quad \mathbf{U}_2^{(\mathrm{q})\mathrm{H}} \quad \cdots \quad \mathbf{U}_K^{(\mathrm{q})\mathrm{H}}\}\end{aligned}$$

where $\mathbf{\Phi}$ = diag $\{\mathbf{\Phi}_1\ \mathbf{\Phi}_2 \cdots \mathbf{\Phi}_K\}$.

B. Complexity Analysis

As for the computational complexity of the GZI algorithms with K users, the proposed GZI needs to compute an inverse operation in (4) and the QR decomposition of $N_t \times n_j$ matrix $\hat{\mathbf{H}}_j$ in (5) K times, while the conventional BD requires to perform the SVD operation of $(N_r - n_j) \times N_t$ other user's channel matrix K times. The complexity of the SVD of $(N_r - n_j) \times N_t$ matrix is $\mathcal{O}(N_t^2(N_r - n_j))$ [10]. The QR-decomposition of $N_t \times n_j$ matrix $\hat{\mathbf{H}}_j$ has $\mathcal{O}(N_t n_j^2)$ complexity [10] and the complexity of the Moor-Penrose pseudo-inverse $\mathbf{H}_s^H(\mathbf{H}_s\mathbf{H}_s^H)^{-1}$ follows $\mathcal{O}({N_r}^{\omega})$, where $2 < \omega < 3$ [11]. Consequently, the proposed GZI algorithm has lower computational complexity than the conventional BD in [8].

IV. Generalization of MMSE Channel Inversion

In this section, we propose a generalized MMSE-CI (GMI) algorithm. Based on the new interpretation of the GZI made in the previous section, we outline a procedure for identifying the GMI precoding matrix which could balance the MUI and the noise for each user.

A. Generalized MMSE channel inversion

In the GMI scheme, the precoding matrix can be determined by applying the MMSE-CI introduced in [7]. We denote $\bar{\mathbf{H}}_s$ as

$$\bar{\mathbf{H}}_s = (\mathbf{H}_s^H\mathbf{H}_s + \alpha\mathbf{I})^{-1}\mathbf{H}_s^H = [\,\bar{\mathbf{H}}_1 \;\; \bar{\mathbf{H}}_2 \quad \cdots \bar{\mathbf{H}}_K\,] \tag{8}$$

where α represents the ratio of the total noise variance to the total transmit power [7]. We assume the unit variance data symbol and $N_t = N_r$, α can be denoted by σ_w^2. For orthogonalization of $\bar{\mathbf{H}}_j$, we can employ

$$\bar{\mathbf{H}}_j = [\bar{\mathbf{Q}}_j^{(1)} \;\; \bar{\mathbf{Q}}_j^{(0)}]\begin{bmatrix}\bar{\mathbf{R}}_j^{(1)} \\ \mathbf{0}\end{bmatrix} = \bar{\mathbf{Q}}_j^{(1)}\bar{\mathbf{R}}_j^{(1)} \tag{9}$$

where $\bar{\mathbf{R}}_j^{(1)}$ is an $n_j \times n_j$ upper triangular matrix and the matrix $\bar{\mathbf{Q}}_j^{(1)}$ is composed of n_j orthonormal basis vectors of $\bar{\mathbf{H}}_j$.

Similar to the GZI in the previous section, we can construct the precoding matrix $\mathbf{P}_j$ for the jth user using a linear combination of columns of $\bar{\mathbf{Q}}_j^{(1)}$. In comparison with the GZI, the columns of $\bar{\mathbf{Q}}_j^{(1)}$ in the GMI span the nullspace of other users' effective channel matrix while taking the noise into account. This leads to an increase of SINR at each user's receiver. However, the jth user's precoder of the GMI generates the residual interference. Thus, a proper whitening or interference-suppression process is needed. In the following, we introduce the noise whiting process.

In order to compute the whiting matrix which is employed to the precoding matrix, we define the transmit combining matrix $\bar{\mathbf{T}}_s$ as

$$\bar{\mathbf{T}}_s \;\; = \;\; \text{diag}\{\bar{\mathbf{T}}_1 \quad \bar{\mathbf{T}}_2 \quad \cdots \quad \bar{\mathbf{T}}_K\}$$

where $\bar{\mathbf{T}}_j$ is an $n_j \times n_j$ square matrix. Denoting $\bar{\mathbf{P}}_s$ as

$$\bar{\mathbf{P}}_s = [\bar{\mathbf{P}}_1\,, \bar{\mathbf{P}}_2 \cdots \bar{\mathbf{P}}_K] = [\bar{\mathbf{Q}}_1^{(1)} \;\; \bar{\mathbf{Q}}_2^{(1)} \cdots \bar{\mathbf{Q}}_K^{(1)}]\bar{\mathbf{T}}_s, \tag{10}$$

the corresponding received signal vector of the jth user can be written from (1) as

$$\bar{\mathbf{y}}_j = \mathbf{H}_j\bar{\mathbf{Q}}_j^{(1)}\bar{\mathbf{T}}_j\mathbf{s}_j + \mathbf{H}_j\sum_{k\neq j}\bar{\mathbf{Q}}_k^{(1)}\bar{\mathbf{T}}_k\mathbf{s}_k + \mathbf{w}_j. \tag{11}$$

From (11), the transmit combining matrix affects other users' interference, thus we define the power of other users' interference induced by the jth user's precoder plus the total noise power of its receiver as

$$\begin{aligned}\sigma^2_{\text{OIN},j} &= \|\tilde{\mathbf{H}}_j\bar{\mathbf{Q}}_j^{(1)}\bar{\mathbf{T}}_j\|_F^2 + n_j\sigma_w^2 \\ &= \text{Tr}\,(\bar{\mathbf{T}}_j^H(\bar{\mathbf{Q}}_j^{(1)H}\tilde{\mathbf{H}}_j^H\tilde{\mathbf{H}}_j\bar{\mathbf{Q}}_j^{(1)} + \sigma_w^2\mathbf{I}_{n_j})\bar{\mathbf{T}}_j).\end{aligned}$$

Since the matrix $\bar{\mathbf{Q}}_j^{(1)H}\tilde{\mathbf{H}}_j^H\tilde{\mathbf{H}}_j\bar{\mathbf{Q}}_j^{(1)} + \sigma_w^2\mathbf{I}_{n_j}$ is Hermitian and positive definite, we can decompose this matrix using Cholesky factorization as

$$\bar{\mathbf{Q}}_j^{(1)H}\tilde{\mathbf{H}}_j^H\tilde{\mathbf{H}}_j\bar{\mathbf{Q}}_j^{(1)} + \sigma_w^2\mathbf{I}_{n_j} = \bar{\mathbf{L}}_j^H\bar{\mathbf{L}}_j. \tag{12}$$

Then, $\sigma^2_{\text{OIN},j}$ can be expressed by $\sigma^2_{\text{OIN},j} = \text{Tr}\,(\bar{\mathbf{T}}_j^H\bar{\mathbf{L}}_j^H\bar{\mathbf{L}}_j\bar{\mathbf{T}}_j)$. We can obtain the jth user's transmit combining matrix which minimizes $\sigma^2_{\text{OIN},j}$ as $\bar{\mathbf{T}}_j = \bar{\mathbf{L}}_j^{-1}$.

Multiplying the $\bar{\mathbf{P}}_s$ in (10) to the network channel matrix $\mathbf{H}_s$, each user has the interference suppressed block channel. Before we decouple the block channel, we find the whiting matrix which is employed to each user's receive filter. Let us define the receive combining matrix $\bar{\mathbf{R}}_s$ as

$$\bar{\mathbf{R}}_s = \text{diag}\{\bar{\mathbf{R}}_1 \quad \bar{\mathbf{R}}_2 \quad \cdots \quad \bar{\mathbf{R}}_K\}$$

and $\bar{\mathbf{M}}_s$ as $\bar{\mathbf{M}}_s = \bar{\mathbf{R}}_s$, the corresponding receive filter output vector of the jth user $\bar{\mathbf{x}}_j$ is written from (2) and (10) as

$$\bar{\mathbf{x}}_j = \bar{\mathbf{R}}_j\mathbf{H}_j\bar{\mathbf{P}}_j\mathbf{s}_j + \bar{\mathbf{R}}_j\mathbf{H}_j\sum_{k\neq j}\bar{\mathbf{P}}_k\mathbf{s}_k + \bar{\mathbf{R}}_j\mathbf{w}_j. \tag{13}$$

Then, from (13), the SINR of the jth user is given by

$$\text{SINR}_j = \frac{\text{Tr}(\bar{\mathbf{R}}_j\mathbf{H}_j\bar{\mathbf{P}}_j\bar{\mathbf{P}}_j^H\mathbf{H}_j^H\bar{\mathbf{R}}_j^H)}{\sigma_w^2\text{Tr}(\bar{\mathbf{R}}_j\bar{\mathbf{R}}_j^H) + \text{Tr}(\bar{\mathbf{R}}_j\mathbf{H}_j\tilde{\mathbf{P}}_j\tilde{\mathbf{P}}_j^H\mathbf{H}_j^H\bar{\mathbf{R}}_j^H)} \tag{14}$$

where $\tilde{\mathbf{P}}_j = [\bar{\mathbf{P}}_1 \cdots \bar{\mathbf{P}}_{j-1}\bar{\mathbf{P}}_{j+1} \cdots \bar{\mathbf{P}}_K]$. As the denominator of (14) represents the total power of the interference plus noise of the jth user, this can be expressed as

$$\sigma^2_{\text{IN},j} = \text{Tr}\,(\bar{\mathbf{R}}_j(\mathbf{H}_j\tilde{\mathbf{P}}_j\tilde{\mathbf{P}}_j^H\mathbf{H}_j^H + \sigma_w^2\mathbf{I}_{n_j})\bar{\mathbf{R}}_j^H). \tag{15}$$

Since the receive combining matrix $\bar{\mathbf{R}}_j$ which minimizes $\sigma^2_{\text{IN},j}$ in (15) can maximize the SINR_j in (14), we decompose $\mathbf{H}_j\tilde{\mathbf{P}}_j\tilde{\mathbf{P}}_j^H\mathbf{H}_j^H + \sigma_w^2\mathbf{I}_{n_j}$ as

$$\mathbf{H}_j\tilde{\mathbf{P}}_j\tilde{\mathbf{P}}_j^H\mathbf{H}_j^H + \sigma_w^2\mathbf{I}_{n_j} = \tilde{\mathbf{L}}_j^H\tilde{\mathbf{L}}_j. \tag{16}$$

Then, we can obtain the jth user's receive combining matrix $\bar{\mathbf{R}}_j = \tilde{\mathbf{L}}_j^{-H}$.

Note that, when SNR increases, the transmit and receive combining matrices converge to $\beta\mathbf{I}_{N_r}$ where β is a constant. Also, $\bar{\mathbf{P}}_s$ can be scaled to satisfy the power constraint, i.e., $\text{Tr}\,(\bar{\mathbf{P}}_s^H\bar{\mathbf{P}}_s) \le N_t$.

In order to decompose the jth user's block channel $\bar{\mathbf{R}}_j\mathbf{H}_j\bar{\mathbf{P}}_j$ in (13) into the parallel subchannels, we denote the SVD of $\bar{\mathbf{R}}_j\mathbf{H}_j\bar{\mathbf{P}}_j$ as

$$\bar{\mathbf{R}}_j\mathbf{H}_j\bar{\mathbf{P}}_j = \mathbf{U}_j^{(\text{r})}\mathbf{\Lambda}_j^{(\text{r})}[\mathbf{V}_j^{(\text{r}(1))} \;\; \mathbf{V}_j^{(\text{r}(0))}]^H$$

where $\mathbf{V}_j^{(\mathrm{r}(1))}$ denotes the set of right singular vectors and $\mathbf{U}_j^{(\mathrm{r})}$ is the left singular matrix. Thus, the precoding matrix and the overall receive filter of the GMI scheme are obtained as

$$\begin{aligned}\mathbf{P}_s^{\mathrm{GMI}} &= [\bar{\mathbf{P}}_1\mathbf{V}_1^{(\mathrm{r}(1))}\ \bar{\mathbf{P}}_2\mathbf{V}_2^{(\mathrm{r}(1))}\cdots\bar{\mathbf{P}}_K\mathbf{V}_K^{(\mathrm{r}(1))}],\\ \mathbf{M}_s^{\mathrm{GMI}} &= \mathrm{diag}\{\mathbf{U}_1^{(\mathrm{r})\mathrm{H}}\bar{\mathbf{R}}_1\ \ \mathbf{U}_2^{(\mathrm{r})\mathrm{H}}\bar{\mathbf{R}}_2\ \ \cdots\ \ \mathbf{U}_K^{(\mathrm{r})\mathrm{H}}\bar{\mathbf{R}}_K\}.\end{aligned}$$

Finally, the receive filter output signal vector at the jth user can be written as

$$\mathbf{x}_j = \mathbf{\Lambda}_j^{(\mathrm{r})}\mathbf{s}_j + \mathbf{M}_j^{\mathrm{GMI}}\mathbf{H}_j\sum_{k\neq j}\mathbf{P}_k^{\mathrm{GMI}}\mathbf{s}_k + \mathbf{M}_j^{\mathrm{GMI}}\mathbf{w}_j. \qquad (17)$$

Let $\lambda_{j,i}^{(\mathrm{r})}$ denote the ith diagonal element of $\mathbf{\Lambda}_j^{(\mathrm{r})}$. Each received signal in (17) contains in part the signal of interest with the channel gain $(\lambda_{j,i}^{(\mathrm{r})})^2$ and in part the interference from the other users plus Gaussian noise. The SINR of each stream can be expressed as

$$\mathrm{SINR}_{j,i} = \frac{\left(\lambda_{j,i}^{(\mathrm{r})}\right)^2}{\sigma_w^2\|\mathbf{m}_{j,i}\|^2 + \sum_{k\neq j}\|\mathbf{m}_{j,i}\mathbf{H}_j\mathbf{P}_k^{\mathrm{GMI}}\|^2} \qquad (18)$$

where $\mathbf{m}_{j,i}$ is the ith row vector of $\mathbf{M}_j^{\mathrm{GMI}}$. Then, the sum rate of the proposed GMI scheme is given by

$$R_{\mathrm{GMI}} = \sum_{j=1}^{K}\sum_{i=1}^{n_j}\log_2\left(1+\mathrm{SINR}_{j,i}\right). \qquad (19)$$

B. Design with imperfect channel information

So far, we have assumed that the base station has knowledge of full CSI. However, in practical downlink systems the CSI available at the transmitter is generally imperfect. In this section, we illustrate how the GMI algorithm can overcome such cases where the CSI at the base station is inaccurate.

We consider the channel estimation error model introduced in [12]

$$\mathbf{H}_s = \mathbf{H}_{est} + \mathbf{H}_{err} \qquad (20)$$

where $\mathbf{H}_s$, $\mathbf{H}_{est}$ and $\mathbf{H}_{err}$ represent the true channel matrix, the estimated channel matrix and the estimation error matrix, respectively. We assume that $\mathbf{H}_{est}$ and $\mathbf{H}_{err}$ are uncorrelated, and that $\mathbf{H}_{err}$ in (20) has i.i.d. elements with zero mean and the estimation error variance $\sigma_{e,h}^2$. The entries of $\mathbf{H}_s$ are also i.i.d. with zero mean and unit variance. We also assume that $\mathbf{H}_{err}$ is independent of the data vector $\mathbf{s}_s$, and that $\sigma_{e,h}^2$ is known to the base station.

In this system model, the received signal vector is given by

$$\mathbf{y}_s = \mathbf{H}_{est}\mathbf{P}_s\mathbf{s}_s + \mathbf{H}_{err}\mathbf{P}_s\mathbf{s}_s + \mathbf{w}_s \qquad (21)$$

where $\mathbf{H}_{err}\mathbf{P}_s\mathbf{s}_s$ in (21) results from the estimation error. Defining the error term as $\mathbf{e} = \mathbf{H}_{err}\mathbf{P}_s\mathbf{s}_s + \mathbf{w}_s$, the mean square error (MSE) can be computed as

$$\sigma_e^2 = \mathbb{E}[\|\mathbf{e}\|^2] = N_r\,\sigma_{e,h}^2\,\mathrm{Tr}\,(\mathbf{P}_s^H\mathbf{P}_s) + N_r\,\sigma_w^2. \qquad (22)$$

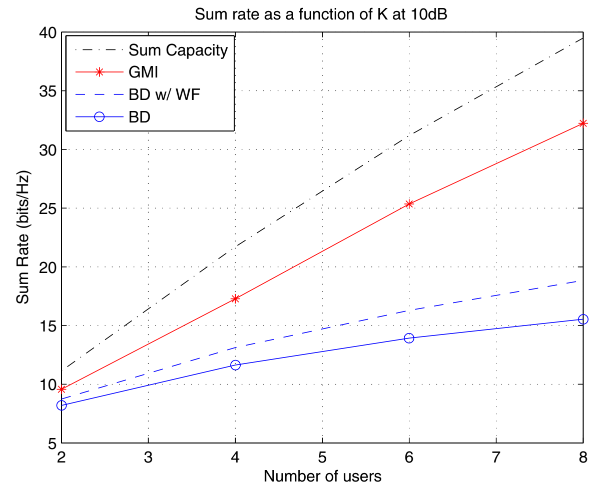

Fig. 1. Comparison of the sum rate as a function of K for SNR = 10dB

From (8) and (22), the precoding matrix can be obtained by applying the MMSE-CI of the estimated channel matrix with the above MSE. We denote $\bar{\mathbf{H}}_{est}$ as

$$\begin{aligned}\bar{\mathbf{H}}_{est} &= (\mathbf{H}_{est}^H\mathbf{H}_{est} + \alpha_e\mathbf{I})^{-1}\mathbf{H}_{est}^H\\ &= [\,\bar{\mathbf{H}}_{est,1}\quad \bar{\mathbf{H}}_{est,2}\quad \cdots\quad \bar{\mathbf{H}}_{est,K}\,]\end{aligned} \qquad (23)$$

where α_e is given by $\alpha_e = \sigma_e^2/P_{total} = N_r\sigma_{e,h}^2 + \sigma_w^2$ and employ the QR-decomposition to $\bar{\mathbf{H}}_{est,j}$ as $\bar{\mathbf{H}}_{est,j} = \bar{\mathbf{Q}}_{est,j}^{(1)}\bar{\mathbf{R}}_{est,j}^{(1)}$. Then, the Cholesky factorizations in (12) and (16) are given as

$$\begin{aligned}\bar{\mathbf{Q}}_{est,j}^{(1)H}\,\tilde{\mathbf{H}}_{est,j}^H\,\tilde{\mathbf{H}}_{est,j}\,\bar{\mathbf{Q}}_{est,j}^{(1)} + \alpha_e\mathbf{I}_{n_j} &= \bar{\mathbf{L}}_{est,j}^H\bar{\mathbf{L}}_{est,j},\\ \bar{\mathbf{H}}_{est,j}\,\tilde{\mathbf{P}}_{est,j}^{(1)}\tilde{\mathbf{P}}_{est,j}^{(1)H}\,\bar{\mathbf{H}}_{est,j}^H + \alpha_e\mathbf{I}_{n_j} &= \tilde{\mathbf{L}}_{est,j}^H\,\tilde{\mathbf{L}}_{est,j}.\end{aligned}$$

Finally, the transmit and receive combining matrices are computed as $\bar{\mathbf{T}}_{est,j} = \bar{\mathbf{L}}_{est,j}^{-1}$ and $\bar{\mathbf{R}}_{est,j} = \tilde{\mathbf{L}}_{est,j}^{-H}$, respectively. A performance gain over the conventional BD in the presence of channel estimation errors will be verified in the following simulation section.

V. Numerical Results

In this section, we present the performance of the proposed GMI scheme comparing with the BD scheme in [8] through Monte carlo simulations.

In Fig. 1, we compare the sum capacity and sum rates for the proposed GMI and BD schemes in terms of the number of users K for the case where each user has two receive antennas and the base station has $N_t = 2K$ antennas. The sum capacity is obtained by calculating the sum power iterative water-filling (SP-IWF) algorithm in [13], and the sum rate of the GMI scheme is computed using (18) and (19). Unlike the conventional BD scheme, the sum rate slope of the GMI is much steeper than the BD scheme and exhibits a linear growth with K. It is clear from the plot that the capacity gain of the GMI over the BD grows as the number of users increases.

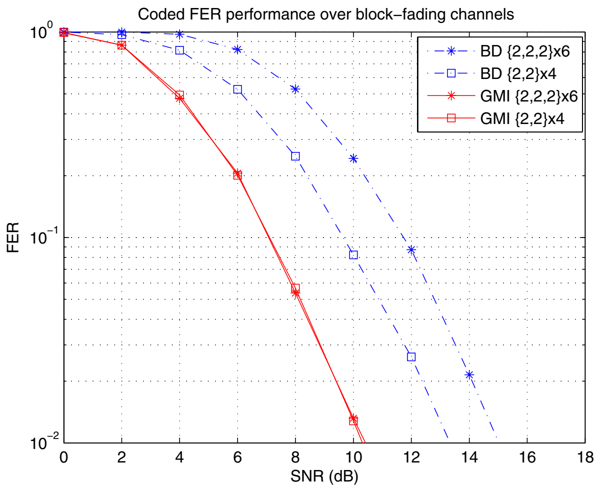

Fig. 2. FER with turbo codes for $\{2,2\}\times 4$ and $\{2,2,2\}\times 6$ in block-fading MIMO channels

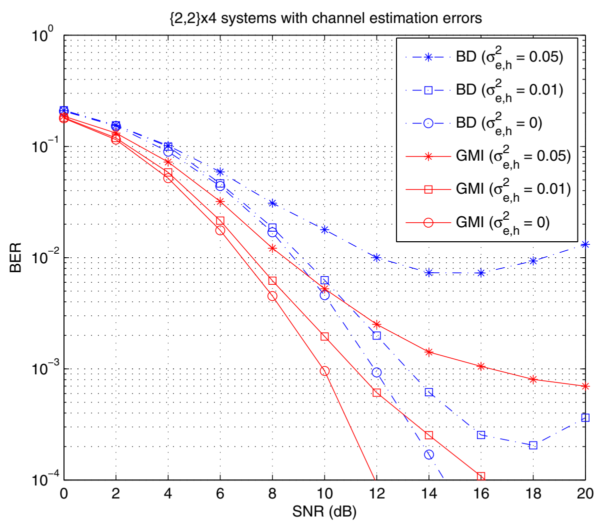

Fig. 3. BER with turbo codes for $\{2,2\}\times 4$ in block-fading MIMO channels with various $\sigma^2_{e,h}$

In Fig. 3, we show the simulation results of coded systems for various channel models in terms of FER with respect to SNR in dB for the $\{2,2\}\times 4$ and $\{2,2,2\}\times 6$ cases. For all FER simulations, a rate-1/2 turbo code based on parallel concatenated code with polynomial (15,13) in octal notation is employed. The number of decoding iterations is set to 6 for the turbo code. We employ 4-QAM with Gray mapping in both cases. The network channel $\mathbf{H}_s$ is generated by an ergodic random process at each frame and is fixed during the transmission of the frame. A 3dB SNR gain at 1% FER for the $\{2,2\}\times 4$ case is observed in Fig. 3. Also, the figure shows that the proposed GMI algorithm outperforms the BD by more than 5dB at 1% FER for the $\{2,2,2\}\times 6$ case. Note that, as shown in the sum rate comparison, the FER performance gap increases as the number of users grows.

Finally, in Fig. 3, we evaluate the performance in the block fading channels with various channel estimation error variance values $\sigma^2_{e,h}$. We plot the bit error rate (BER) performance of the proposed GMI and the BD schemes with turbo codes. In the presence of the channel estimation error, we can see from (22) that the performance becomes limited by the error variance as SNR increases. Nevertheless, it is clear that the proposed GMI scheme is much more robust to the estimation errors compared to the conventional BD scheme.

VI. Conclusion

In this paper, we have proposed a generalized MMSE channel inversion algorithm for multiuser MIMO downlink systems where each user has more than or equal to one antenna. An alternative approach of the conventional BD has been presented by using the ZF channel inversion and the orthogonalization process. The proposed GMI precoder and receive filter are obtained by employing the MMSE channel inversion and some decomposition methods and as a result, the SINR is increased at each user's receiver. Through the simulations, we have showed that the proposed GMI outperforms the conventional BD and demonstrated that the proposed scheme is robust to the estimation error.

References

[1] G. J. Foschini and M. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, March 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, November 1999.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, July 2003.

[4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, October 2003.

[5] W. Yu and J. M. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 1875–1892, September 2004.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, May 1983.

[7] C. B. Peel, B. M. Hochwald, and A. L. Swindelhurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, pp. 195–202, January 2005.

[8] Q. H. Spencer, A. L. Swindelhurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 461–471, February 2004.

[9] J. M. Cioffi, *EE379A Class note: Signal Processing and detection*. Stanford Univ.

[10] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Second Edition, The Johns Hopkins University Press, Baltimore and London, 1989.

[11] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *J. Symbolic Computation*, vol. 9, pp. 251–280, August 1990.

[12] T. Yoo and A. J. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, pp. 2203–2214, May 2006.

[13] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 1570–1580, April 2005.