

Generalized Design of Low-Complexity Block Diagonalization Type Precoding Algorithms for Multiuser MIMO Systems

Keke Zu, Rodrigo C. de Lamare, *Senior Member, IEEE*, and Martin Haardt, *Senior Member, IEEE*

Abstract—Block diagonalization (BD) based precoding techniques are well-known linear transmit strategies for multiuser MIMO (MU-MIMO) systems. By employing BD-type precoding algorithms at the transmit side, the MU-MIMO broadcast channel is decomposed into multiple independent parallel single user MIMO (SU-MIMO) channels and achieves the maximum diversity order at high data rates. The main computational complexity of BD-type precoding algorithms comes from two singular value decomposition (SVD) operations, which depend on the number of users and the dimensions of each user's channel matrix. In this work, low-complexity precoding algorithms are proposed to reduce the computational complexity and improve the performance of BD-type precoding algorithms. We devise a strategy based on a common channel inversion technique, QR decompositions, and lattice reductions to decouple the MU-MIMO channel into equivalent SU-MIMO channels. Analytical and simulation results show that the proposed precoding algorithms can achieve a comparable sum-rate performance as BD-type precoding algorithms, substantial bit error rate (BER) performance gains, and a simplified receiver structure, while requiring a much lower complexity.

Index Terms—Multiuser MIMO (MU-MIMO), block diagonalization (BD), regularized block diagonalization (RBD), low-complexity, lattice reduction (LR).

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) systems have drawn a considerable research effort in the past years due to the fact that they can greatly increase the spectrum efficiency of wireless communications [2], [3]. In order to meet the continuous growing data traffic, a downlink peak spectrum efficiency of 30 bps/Hz and an uplink peak spectrum efficiency of 15 bps/Hz is proposed in LTE-Advanced [4], and a configuration of up to 8 transmit antennas for the downlink is suggested. A new amendment for the WLAN standard IEEE 802.11ac [5] also recommends up to 8 MIMO spatial streams. Configurations with dozens of antennas are now being considered [6]. High-dimensional MIMO systems or large MIMO systems are very promising for the next generation of wireless communication systems due to their potential to improve rate and reliability dramatically [7]. However, it is a

challenge to design a suitable precoding algorithm with good overall performance and low computational complexity at the same time for high-dimensional MIMO systems.

Unlike the received signal in single user MIMO (SU-MIMO) systems, the received signals of different users in multiuser MIMO (MU-MIMO) systems not only suffer from the noise and the inter-antenna interference but are also affected by the multiuser interference (MUI). Channel inversion based precoding or linear precoding algorithms such as zero forcing (ZF) and minimum mean squared error (MMSE) precoding [8] can still be used to cancel the MUI, but they result in a reduced throughput or require a higher power at the transmitter in the MU-MIMO scenarios. As a generalization of the ZF precoding algorithm, block diagonalization (BD) based precoding algorithms have been proposed in [9], [10] for MU-MIMO systems. However, BD based precoding algorithms only take the MUI into account and thus suffer a performance loss at low signal to noise ratios (SNRs) when the noise is the dominant factor. Therefore, a regularized block diagonalization (RBD) precoding algorithm which introduces a regularization factor to take the noise term into account has been proposed in [11]. We term the BD and RBD based precoding schemes as BD-type precoding algorithms in this work for convenience.

The main steps of the BD-type precoding algorithms are two SVD operations, which need to be implemented for each user. Therefore, the computational complexity of the BD-type precoding algorithms depends on the number of users and the dimensions of each user's channel matrix. For MU-MIMO systems with a large number of users and multiple receive antennas, this could result in a considerable computational cost. Another distinctive aspect of the BD-type precoding algorithms is that they need a decoding matrix obtained from the second SVD operation to orthogonalize each user's streams. The requirement of this decoding matrix brings extra control overhead or computational complexity [12].

Recent work on BD-type precoding algorithms has focused on how to equivalently implement the BD-type precoding algorithms with less computational complexity. A low-complexity generalized ZF channel inversion (GZI) method has been proposed in [13] to equivalently implement the first SVD operation of the original BD precoding, and a generalized MMSE channel inversion (GMI) method is also developed in [13] for the original RBD precoding. In [14] the first SVD operation of the RBD precoding is replaced with a less complex QR decomposition [15]. We term the work in

Manuscript received January 16, 2013; revised June 22, 2013. The editor coordinating the review of this paper and approving it for publication was B. Clerckx.

The authors are with the University of York, York, North Yorkshire, United Kingdom (e-mail: zukeke@gmail.com, {kz511, rcd15}@york.ac.uk, martin.haardt@tu-ilmenau.de).

Part of this work has been submitted to IEEE Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, Oct. 2012 [1].

Digital Object Identifier 10.1109/TCOMM.2013.090513.130038

[13] as GMI-type precoding and the work in [14] as QR/SVD-type precoding. For the second SVD operation, however, both the GMI-type and QR/SVD-type precoding schemes employ it in a similar way as the conventional BD-type precoding algorithms to parallelize each user's streams. Therefore, the second SVD operation needs to be implemented multiple times and the decoding matrix for the effective channel still needs to be known or estimated at the receiver of each user for the GMI-type or QR/SVD-type precoding algorithms.

The GMI-type and QR/SVD-type techniques are solely low complexity equivalent implementations of the BD-type precoding algorithms. As an improvement of the BD-type precoding algorithms, a low-complexity lattice reduction-aided RBD (LC-RBD-LR) type precoding algorithms has been proposed in [17], [18] based on the QR decomposition scheme. Not only much less complexity but also considerable BER gains are achieved by the LC-RBD-LR-type precoding algorithms. However, the QR decomposition in LC-RBD-LR-type precoding algorithms still needs to be implemented for each user, which could result in a high complexity for large MIMO systems.

A new category of low-complexity high performance precoding algorithms based on the channel inversion scheme is proposed in this work. A simplified GMI (S-GMI) precoding scheme which employs a common channel inversion for all users is developed first. Equivalent parallel SU-MIMO channels are obtained from the S-GMI precoding process. Then, these effective channels are transformed into the lattice space by utilizing the lattice reduction (LR) technique [16], whose complexity is dictated by a QR decomposition. Linear precoding strategies are applied in the lattice space to parallelize each user's streams. Finally, the proposed lattice reduction-aided simplified GMI (LR-S-GMI) precoding algorithms are obtained. According to the specific linear precoding constraint used, the proposed LR-S-GMI-type precoding algorithms are categorized as LR-S-GMI-ZF and LR-S-GMI-MMSE, respectively.

The algorithm structure of the proposed LR-S-GMI-type precoding is different from the LC-RBD-LR-type precoding since the channel inversion is only implemented once for all users, while the QR decomposition needs to be implemented multiple times in LC-RBD-LR-type precoding. Therefore, the computational complexity can be reduced considerably by the proposed LR-S-GMI-type precoding. A comprehensive mathematical analysis is developed to analyze and predict the performance of the proposed LR-S-GMI-type precoding algorithms. The simulation results verified that the proposed LR-S-GMI-type precoding algorithms have the lowest computational complexity compared to BD-type [9], [11], GMI-type [13], QR/SVD-type [14] and LC-RBD-LR-type [17] precoding algorithms, a comparable sum-rate performance as BD-type precoding algorithms, and substantial BER performance gains over prior art.

The main contributions of the work are summarized below:

- 1) A simplified GMI (S-GMI) precoding is developed in this work as an improvement of the original RBD in [11]. A mathematical analysis is given to show that the S-GMI has a better BER performance and much less complexity than that of RBD, which is a clear difference

compared to the GMI in [13] which only provides an equivalent implementation of RBD.

- 2) A new category of low-complexity high-performance LR-S-GMI-type precoding algorithms is proposed for MU-MIMO systems based on a channel inversion technique, QR decompositions, and lattice reductions.
- 3) The BD-type precoding algorithms are systematically analyzed and summarized. We show that the computational complexity of the BD-type precoding depends on the number of users and the system dimensions.
- 4) A comprehensive performance analysis is carried out in terms of BER performance, achievable sum-rate, and computational complexity.
- 5) A simulation study of the proposed algorithms under imperfect channel situations is also conducted, which completes this paper.

The proposed and existing precoding techniques are all performed with the help of downlink channel state information (CSI). The assumption that full CSI is available at the transmit side is valid in time-division duplex (TDD) systems because the uplink and downlink share the same frequency band. For frequency-division duplex (FDD) systems, however, the CSI needs to be estimated at the receiver and fed back to the transmitter.

This paper is organized as follows. The system model is given in Section II. A brief review of the BD-type precoding algorithms is presented in Section III. The proposed LR-S-GMI-type precoding algorithms are described in detail in Section IV and the performance analysis is developed in Section V. Simulation results and conclusions are displayed in Section VI and Section VII, respectively.

Notation: Matrices and vectors are denoted by upper and lowercase boldface letters, and the transpose, Hermitian transpose, inverse, pseudo-inverse of a matrix \mathbf{B} are described by \mathbf{B}^T , \mathbf{B}^H , \mathbf{B}^{-1} , \mathbf{B}^\dagger , respectively. The trace, determinant, Frobenius norm, round function are denoted as $Tr(\cdot)$, $det(\cdot)$, $\|\cdot\|_F$, $\lceil \cdot \rceil$. With $\text{diag}\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_K\}$ creates a block diagonal matrix with the matrices \mathbf{B}_k on the main diagonal.

II. SYSTEM MODEL

We consider an uncoded MU-MIMO downlink channel, with N_T transmit antennas at the base station (BS) and N_i receive antennas at the i th user equipment (UE). With K users in the system, the total number of receive antennas is $N_R = \sum_{i=1}^K N_i$. A block diagram of such a system is illustrated in Fig. 1.

From the system model, the combined channel matrix \mathbf{H} and the joint precoding matrix \mathbf{P} are given by

$$\mathbf{H} = [\mathbf{H}_1^T \ \mathbf{H}_2^T \ \dots \ \mathbf{H}_K^T]^T \in \mathbb{C}^{N_R \times N_T}, \quad (1)$$

$$\mathbf{P} = [\mathbf{P}_1 \ \mathbf{P}_2 \ \dots \ \mathbf{P}_K] \in \mathbb{C}^{N_T \times N_R}, \quad (2)$$

where $\mathbf{H}_i \in \mathbb{C}^{N_i \times N_T}$ is the i th user's channel matrix. The quantity $\mathbf{P}_i \in \mathbb{C}^{N_T \times N_i}$ is the i th user's precoding matrix. We assume a flat fading MIMO channel and the received signal $\mathbf{y}_i \in \mathbb{C}^{N_i}$ at the i th user is given by

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{H}_i \sum_{j=1, j \neq i}^K \mathbf{x}_j + \mathbf{n}_i, \quad (3)$$

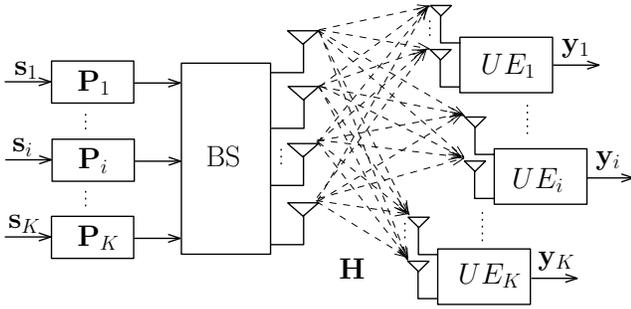


Fig. 1. MU-MIMO System Model

where the quantity $\mathbf{x}_i \in \mathbb{C}^{N_i}$ is the i th user's transmitted signal, and $\mathbf{n}_i \in \mathbb{C}^{N_i}$ is the i th user's Gaussian noise with independent and identically distributed (i.i.d.) entries of zero mean and variance σ_n^2 . Assuming that the average transmit power for each user is ξ_i , then, the power constraint $\mathbb{E}\|\mathbf{x}_i\|^2 = \xi_i$ is imposed. We construct an unnormalized signal \mathbf{s}_i such that

$$\mathbf{x}_i = \frac{\mathbf{s}_i}{\sqrt{E_{\gamma_i}}}, \quad (4)$$

where $\mathbf{s}_i = \mathbf{P}_i \mathbf{d}_i$ with \mathbf{d}_i being the transmit data vector and E_{γ_i} is the average energy of γ_i with $\gamma_i = \|\mathbf{s}_i\|^2 / \xi_i$. The physical meaning of dividing \mathbf{s}_i by the scalar $\sqrt{E_{\gamma_i}}$ is to make sure the average transmit power ξ_i is still the same after the precoding process. With this normalization, \mathbf{x}_i obeys $\mathbb{E}\|\mathbf{x}_i\|^2 = \xi_i$.

The received signal \mathbf{y}_i is weighted by the scalar $\sqrt{E_{\gamma_i}}$ to form the estimate

$$\hat{\mathbf{d}}_i = \sqrt{E_{\gamma_i}} \mathbf{y}_i. \quad (5)$$

Note that it is necessary to cancel $\sqrt{E_{\gamma_i}}$ out at the receiver to get the correct amplitude of the desired signal part. The average energy E_{γ_i} is independent from the channel and the data, which means the receivers do not need to know the instantaneous CSI for the precoding techniques to work [19]. As analyzed and illustrated in [20], however, the performance difference between the average E_{γ_i} and the instantaneous γ_i is very small. Therefore, we follow the strategy developed in [19] and [20] to assume the receivers need to know only $\sqrt{E_{\gamma_i}}$ but use γ_i instead of E_{γ_i} for simulation convenience as γ_i is simpler to compute. The simulation results represent the performance of either normalization method. In this case, we can replace (4) and (5) with the instantaneous γ_i in the simulations and employ

$$\mathbf{x}_i = \frac{\mathbf{s}_i}{\sqrt{\gamma_i}} \text{ and } \hat{\mathbf{d}}_i = \sqrt{\gamma_i} \mathbf{y}_i. \quad (6)$$

III. REVIEW OF BD-TYPE PRECODING ALGORITHMS

The design of BD-type precoding algorithms is performed in two steps [9], [11]. The first precoding filter is used to completely eliminate (by BD) or balance the MUI with noise (by RBD), then exact (by BD) or approximate (by RBD) parallel SU-MIMO channels are obtained. The second precoding filter is implemented to parallelize each user's streams.

Correspondingly, the precoding matrix \mathbf{P}_i for the i th user can be rewritten in two parts as

$$\mathbf{P}_i = \mathbf{P}_i^a \mathbf{P}_i^b, \quad (7)$$

where $\mathbf{P}_i^a \in \mathbb{C}^{N_T \times M_i}$ and $\mathbf{P}_i^b \in \mathbb{C}^{M_i \times N_i}$. The parameter M_i is dependent on the specific choice of the precoding algorithm. We exclude the i th user's channel matrix and define $\overline{\mathbf{H}}_i$ as

$$\overline{\mathbf{H}}_i = [\mathbf{H}_1^T \dots \mathbf{H}_{i-1}^T \mathbf{H}_{i+1}^T \dots \mathbf{H}_K^T]^T \in \mathbb{C}^{\overline{N}_i \times N_T}, \quad (8)$$

where $\overline{N}_i = N_R - N_i$. Then, the interference generated to the other users is determined by $\overline{\mathbf{H}}_i \mathbf{P}_i^a$.

In order to eliminate all the MUI, we impose the constraint that

$$\forall i \in (1, \dots, K) \quad \overline{\mathbf{H}}_i \mathbf{P}_i^a = \mathbf{0} \text{ s.t. } \mathbb{E}\|\mathbf{x}_i\|^2 = \xi_i. \quad (9)$$

We term (9) as the BD constraint. Note that the BD constraint is actually an extension of the ZF constraint in [8] for MU-MIMO with multiple receive antennas. In order to take the noise term into account as well, an RBD constraint is developed in [11] and given by

$$\begin{aligned} \mathbf{P}_i^a &= \min_{\mathbf{P}_i^a} \mathbb{E}\{\|\overline{\mathbf{H}}_i \mathbf{P}_i^a\|^2 + \gamma_i \|\mathbf{n}_i\|^2\} \\ &\text{s.t. } \mathbb{E}\|\mathbf{x}_i\|^2 = \xi_i. \end{aligned} \quad (10)$$

Assuming that the rank of $\overline{\mathbf{H}}_i$ is \overline{L}_i , define the SVD of $\overline{\mathbf{H}}_i$

$$\overline{\mathbf{H}}_i = \overline{\mathbf{U}}_i \overline{\Sigma}_i \overline{\mathbf{V}}_i^H = \overline{\mathbf{U}}_i \overline{\Sigma}_i [\overline{\mathbf{V}}_i^{(1)} \quad \overline{\mathbf{V}}_i^{(0)}]^H, \quad (11)$$

where $\overline{\mathbf{U}}_i \in \mathbb{C}^{\overline{N}_i \times \overline{N}_i}$ and $\overline{\mathbf{V}}_i \in \mathbb{C}^{N_T \times N_T}$ are unitary matrices. The diagonal matrix $\overline{\Sigma}_i \in \mathbb{C}^{\overline{N}_i \times N_T}$ contains the singular values of the matrix $\overline{\mathbf{H}}_i$. Factorizing $\overline{\mathbf{V}}_i$ into two parts, $\overline{\mathbf{V}}_i^{(1)} \in \mathbb{C}^{N_T \times \overline{L}_i}$ consists of the first \overline{L}_i non-zero singular vectors and $\overline{\mathbf{V}}_i^{(0)} \in \mathbb{C}^{N_T \times (N_T - \overline{L}_i)}$ holds the last $N_T - \overline{L}_i$ zero singular vectors. Thus, $\overline{\mathbf{V}}_i^{(0)}$ forms an orthogonal basis for the null space of $\overline{\mathbf{H}}_i$. The solution for the BD constraint (9) is given by

$$\mathbf{P}_i^{a(\text{BD})} = \overline{\mathbf{V}}_i^{(0)}. \quad (12)$$

As shown in [11], the solution for the RBD constraint can be obtained as

$$\mathbf{P}_i^{a(\text{RBD})} = \overline{\mathbf{V}}_i (\overline{\Sigma}_i^T \overline{\Sigma}_i + \alpha \mathbf{I}_{N_T})^{-1/2}, \quad (13)$$

where $\alpha = \frac{N_R \sigma_n^2}{\xi}$ is the regularization factor with ξ is the whole average transmit power.

After the first precoding process, the MU-MIMO channel is decoupled into a set of K parallel independent SU-MIMO channels by the BD precoding. For the RBD precoding, there are residual interferences between these channels due to the regularization process, but, these interferences tend to zero at high SNRs. Therefore, the effective channel matrix for the i th user can be expressed as

$$\mathbf{H}_{\text{eff}_i} = \mathbf{H}_i \mathbf{P}_i^a. \quad (14)$$

Define $L_{\text{eff}} = \text{rank}(\mathbf{H}_{\text{eff}_i})$ and consider the second SVD operation on the effective channel matrix

$$\mathbf{H}_{\text{eff}_i} = \mathbf{U}_i \Sigma_i \mathbf{V}_i^H = \mathbf{U}_i \begin{bmatrix} \Sigma_i & 0 \\ 0 & 0 \end{bmatrix} [\mathbf{V}_i^{(1)} \quad \mathbf{V}_i^{(0)}]^H, \quad (15)$$

using the unitary matrix $\mathbf{U}_i \in \mathbb{C}^{L_{\text{eff}} \times L_{\text{eff}}}$ and $\mathbf{V}_i^{(1)}$ contains the first L_{eff} singular vectors. The second precoding filters for BD and RBD precoding can be respectively obtained as

$$\mathbf{P}_i^{b(\text{BD})} = \mathbf{V}_i^{(1)} \mathbf{\Lambda}^{(\text{BD})}, \quad (16)$$

$$\mathbf{P}_i^{b(\text{RBD})} = \mathbf{V}_i \mathbf{\Lambda}^{(\text{RBD})}, \quad (17)$$

where $\mathbf{\Lambda}$ is the power loading matrix that depends on the optimization criterion. An example power loading is the water filling (WF) [2]. The i th user's decoding matrix is obtained as

$$\mathbf{G}_i = \mathbf{U}_i^H, \quad (18)$$

which needs to be known by each user's receiver.

Note that for the conventional BD-type precoding algorithms, there is a dimensionality constraint below to be satisfied

$$N_T > \max\{\text{rank}(\overline{\mathbf{H}}_1), \text{rank}(\overline{\mathbf{H}}_2), \dots, \text{rank}(\overline{\mathbf{H}}_K)\}. \quad (19)$$

Then, we can get the matrix dimension relationship as $N_T \geq N_R > \overline{N}_i \geq \overline{L}_i > N_i \geq L_{\text{eff}}$. Note that the first SVD operation in (11) needs to be implemented K times on $\overline{\mathbf{H}}_i$ with dimension $\overline{N}_i \times N_T$ and the second SVD operation in (15) needs to be implemented K times on $\mathbf{H}_{\text{eff}_i}$ with dimensions $L_{\text{eff}} \times (N_T - \overline{L}_i)$ for the BD precoding and $L_{\text{eff}} \times N_T$ for the RBD precoding. From the above analysis, most of the computational complexity of the BD-type precoding algorithms comes from the two SVD operations which make the computational complexity of the BD-type precoding algorithms increase with the number of users K and the system dimensions.

IV. PROPOSED S-GMI BASED PRECODING ALGORITHMS

In this section, we describe the proposed LR-S-GMI-type precoding algorithms based on a strategy that employs a channel inversion method, QR decompositions, and lattice reductions. Similar to the BD-type precoding algorithms, the design of the proposed LR-S-GMI-type precoding algorithms is computed in two steps.

First, we obtain the first precoding filter \mathbf{P}_i^a for the LR-S-GMI-type precoding algorithms by using one channel inversion and K QR decompositions each implemented on individual users with matrix dimension $N_i \times N_i$. By applying the MMSE inversion to the combined channel matrix, we have

$$\begin{aligned} \mathbf{H}_{\text{mse}}^\dagger &= \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + \alpha \mathbf{I})^{-1} \\ &= [\mathbf{H}_{1,\text{mse}}, \mathbf{H}_{2,\text{mse}}, \dots, \mathbf{H}_{K,\text{mse}}]. \end{aligned} \quad (20)$$

where $\mathbf{H}_{i,\text{mse}} \in \mathbb{C}^{N_T \times N_i}$ is the sub-matrix of $\mathbf{H}_{\text{mse}}^\dagger$. Considering a high SNR case, it can be shown that the regularization factor α approaches zero and thus we have $\mathbf{H}\mathbf{H}^\dagger_{\text{mse}} \approx \mathbf{I}_{N_T}$ [8]. This means the off-diagonal block matrices of $\mathbf{H}\mathbf{H}^\dagger_{\text{mse}}$ converge to zeros with the increase of SNR. Hence, the matrix $\mathbf{H}_{i,\text{mse}}$ is approximately in the null space of $\overline{\mathbf{H}}_i$ defined in (8), that is,

$$\overline{\mathbf{H}}_i \mathbf{H}_{i,\text{mse}} \approx \mathbf{0}. \quad (21)$$

Considering the QR decomposition of $\mathbf{H}_{i,\text{mse}} = \mathbf{Q}_{i,\text{mse}} \mathbf{R}_{i,\text{mse}}$, we have

$$\overline{\mathbf{H}}_i \mathbf{H}_{i,\text{mse}} = \overline{\mathbf{H}}_i \mathbf{Q}_{i,\text{mse}} \mathbf{R}_{i,\text{mse}} \approx \mathbf{0} \text{ for } i = 1, \dots, K, \quad (22)$$

TABLE I
THE S-GMI PRECODING ALGORITHM

Steps	Operations
Applying the MMSE Channel Inversion	
(1)	$\mathbf{H}_{\text{mse}}^\dagger = (\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I})^{-1} \mathbf{H}^H$
(2)	for $i = 1 : K$
(3)	$[\mathbf{Q}_{i,\text{mse}}^\dagger, \mathbf{R}_{i,\text{mse}}^\dagger] = \text{QR}(\mathbf{H}_{i,\text{mse}}^\dagger, 0)$
(4)	$\mathbf{P}_i^a = \mathbf{Q}_{i,\text{mse}}^\dagger$
(5)	$\mathbf{H}_{\text{eff}_i} = \mathbf{H}_i \mathbf{P}_i^a = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{V}_i^H$
(6)	$\mathbf{P}_i^b = \mathbf{V}_i$
(7)	$\mathbf{G}_i = \mathbf{U}_i^H$
(8)	end
Compute the overall precoding and decoding matrix	
(9)	$\mathbf{P}^a = [\mathbf{P}_1^a, \mathbf{P}_2^a, \dots, \mathbf{P}_K^a]$
(10)	$\mathbf{P}^b = \text{diag}\{\mathbf{P}_1^b, \mathbf{P}_2^b, \dots, \mathbf{P}_K^b\}$
(11)	$\mathbf{P} = \mathbf{P}^a \mathbf{P}^b \quad \mathbf{G} = \text{diag}\{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_K\}$
Calculate the scaling factor γ	
(12)	$\gamma = (\ \mathbf{P}\mathbf{d}\ _F^2 / E_s)$
Get the received signal	
(13)	$\mathbf{y} = \mathbf{G}(\mathbf{H}\mathbf{P}\mathbf{d} + \sqrt{\gamma}\mathbf{n})$

where $\mathbf{Q}_{i,\text{mse}} \in \mathbb{C}^{N_T \times N_i}$ is an orthogonal matrix and $\mathbf{R}_{i,\text{mse}} \in \mathbb{C}^{N_i \times N_i}$ is an upper triangular matrix. Since $\mathbf{R}_{i,\text{mse}}$ is invertible, we have

$$\overline{\mathbf{H}}_i \mathbf{Q}_{i,\text{mse}} \approx \mathbf{0}. \quad (23)$$

Thus, $\mathbf{Q}_{i,\text{mse}}$ satisfies the RBD constraint (10) to balance the MUI and the noise term.

We have simplified the design of the first precoding filter \mathbf{P}_i^a here as compared to [13] where a residual interference suppression filter \mathbf{T}_i is applied after the first precoding process \mathbf{P}_i^a . The filter \mathbf{T}_i increases the complexity and cannot completely cancel the MUI. Therefore, we omit the residual interference suppression part since it is not necessary for the RBD constraint based precoding. We term the simplified GMI as S-GMI in this work. Then, the first precoding filter for S-GMI can be obtained as

$$\mathbf{P}_i^a = \mathbf{Q}_{i,\text{mse}}, \quad (24)$$

where $\mathbf{P}_i^a \in \mathbb{C}^{N_T \times N_i}$. By implementing the QR decomposition in (22) K times on $\mathbf{H}_{i,\text{mse}}$ with dimension $N_i \times N_i$, the first combined precoding matrix for S-GMI is

$$\mathbf{P}^a = [\mathbf{P}_1^a, \mathbf{P}_2^a, \dots, \mathbf{P}_K^a]. \quad (25)$$

Note the K QR decompositions of the LC-RBD-LR-type precoding in [17], [18] are implemented on $\overline{\mathbf{H}}_i$ with dimension $\overline{N}_i \times N_i$. The S-GMI algorithm can be completed by applying the SVD operation to the effective channel matrix $\mathbf{H}_{\text{eff}_i} = \mathbf{H}_i \mathbf{P}_i^a = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{V}_i^H$. Then, the second precoding filter of S-GMI is obtained as $\mathbf{P}_i^b = \mathbf{V}_i$. The S-GMI algorithm is summarized in Table I.

Similarly, the extension of the channel inversion method from the RBD constraint based precoding to the BD constraint based precoding is straightforward on

$$\mathbf{H}_{\text{zf}}^\dagger = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} = [\mathbf{H}_{1,\text{zf}}, \mathbf{H}_{2,\text{zf}}, \dots, \mathbf{H}_{K,\text{zf}}]. \quad (26)$$

Moreover, the obtained MUI is strictly zero as $\overline{\mathbf{H}}_i \mathbf{H}_{i,\text{zf}} = \mathbf{0}$. Assuming the QR decomposition of $\mathbf{H}_{i,\text{zf}}$ is $\mathbf{H}_{i,\text{zf}} = \mathbf{Q}_{i,\text{zf}} \mathbf{R}_{i,\text{zf}}$, then, we have

$$\overline{\mathbf{H}}_i \mathbf{Q}_{i,\text{zf}} = \mathbf{0}. \quad (27)$$

Thus, $\mathbf{Q}_{i,zf}$ satisfies the BD constraint (9). The first precoding matrix for the BD constraint based precoding can be equivalently obtained as

$$\mathbf{P}_i^a = \mathbf{Q}_{i,zf}. \quad (28)$$

This equivalent method is termed as GZI in [13].

For the proposed LR-S-GMI-type precoding algorithms, we get the first precoding filter as S-GMI in (24), while we employ the LR-aided linear precoding technique instead of the SVD operation in S-GMI to obtain the second precoding filter $\tilde{\mathbf{P}}_i^b$. The aim of the LR transformation is to find a new basis $\tilde{\mathbf{H}}$ which is nearly orthogonal compared to the original matrix \mathbf{H} for a given lattice $L(\mathbf{H})$. The most commonly used LR algorithm has been first proposed by Lenstra, Lenstra and L. Lovász (LLL) in [21] with polynomial time complexity. In order to reduce the computational complexity, a complex LLL (CLLL) algorithm was proposed in [22], which reduces the overall complexity of the LLL algorithm by nearly half without sacrificing any performance. We employ the CLLL algorithm to implement the LR transformation in this work.

After the first precoding, we transform the MU-MIMO channel into parallel or approximately parallel SU-MIMO channels and the effective channel matrix for the i th user is

$$\mathbf{H}_{\text{eff}_i} = \mathbf{H}_i \mathbf{P}_i^a. \quad (29)$$

We perform the LR transformation on $\mathbf{H}_{\text{eff}_i}^T$ in the precoding scenario [23], that is

$$\tilde{\mathbf{H}}_{\text{eff}_i} = \mathbf{T}_i \mathbf{H}_{\text{eff}_i} \text{ and } \mathbf{H}_{\text{eff}_i} = \mathbf{T}_i^{-1} \tilde{\mathbf{H}}_{\text{eff}_i}, \quad (30)$$

where \mathbf{T}_i is a unimodular matrix with $\det|\mathbf{T}_i| = 1$ and all elements of \mathbf{T}_i are complex integers, i.e. $t_{l,k} \in \mathbb{Z} + j\mathbb{Z}$. The physical meaning of the constraint $\det|\mathbf{T}_i| = 1$ is that the channel energy is unchanged after the LR transformation.

Following the LR transformation, we employ the linear precoding constraint to get the second precoding filter to parallelize each user's streams. The ZF precoding constraint is implemented for user i as

$$\tilde{\mathbf{P}}_{\text{ZF}_i}^b = \tilde{\mathbf{H}}_{\text{eff}_i}^H (\tilde{\mathbf{H}}_{\text{eff}_i} \tilde{\mathbf{H}}_{\text{eff}_i}^H)^{-1}. \quad (31)$$

It is well-known that the performance of MMSE precoding is always better than that of ZF precoding. We can get the second precoding filter by employing an MMSE precoding constraint. The MMSE precoding is actually equivalent to the ZF precoding with respect to an extended system model [24], [25]. The extended channel matrix $\underline{\mathbf{H}}$ for the MMSE precoding scheme is defined as

$$\underline{\mathbf{H}} = [\mathbf{H}, \sqrt{\alpha} \mathbf{I}_{N_R}]. \quad (32)$$

By introducing the regularization factor α , a trade-off between the level of MUI and the noise is introduced [8]. Then, the MMSE precoding filter is obtained as

$$\mathbf{P}_{\text{MMSE}} = \mathbf{A} \underline{\mathbf{H}}^H (\underline{\mathbf{H}} \underline{\mathbf{H}}^H)^{-1}, \quad (33)$$

where $\mathbf{A} = [\mathbf{I}_{N_T}, \mathbf{0}_{N_T \times N_R}]$, and the multiplication by \mathbf{A} will not result in transmit power amplification since $\mathbf{A} \mathbf{A}^H = \mathbf{I}_{N_T}$. From the mathematical expression in (33), the rows of $\underline{\mathbf{H}}$ determine the effective transmit power amplification of the MMSE precoding. Correspondingly, the LR transformation for

the MMSE precoding should be applied to the transpose of the extended channel matrix $\underline{\mathbf{H}}_{\text{eff}_i}^T = [\mathbf{H}_{\text{eff}_i}, \sqrt{\alpha} \mathbf{I}_{N_i}]^T$, and the LR transformed channel matrix $\tilde{\mathbf{H}}_{\text{eff}_i}$ is obtained as

$$\tilde{\mathbf{H}}_{\text{eff}_i} = \mathbf{T}_i \underline{\mathbf{H}}_{\text{eff}_i}^T, \quad (34)$$

where \mathbf{T}_i is the unimodular matrix for $\underline{\mathbf{H}}_{\text{eff}_i}$. Then, the LR-aided MMSE precoding filter is given by

$$\tilde{\mathbf{P}}_{\text{MMSE}_i}^b = \mathbf{A}_i \tilde{\mathbf{H}}_{\text{eff}_i}^H (\tilde{\mathbf{H}}_{\text{eff}_i} \tilde{\mathbf{H}}_{\text{eff}_i}^H)^{-1}, \quad (35)$$

where the matrix $\mathbf{A}_i = [\mathbf{I}_{M_i}, \mathbf{0}_{M_i \times N_i}]$. Finally, the combined second precoding matrix $\tilde{\mathbf{P}}^b$ for all users is

$$\tilde{\mathbf{P}}^b = \text{diag}\{\tilde{\mathbf{P}}_1^b, \tilde{\mathbf{P}}_2^b, \dots, \tilde{\mathbf{P}}_K^b\}. \quad (36)$$

The overall precoding matrix is $\tilde{\mathbf{P}} = \mathbf{P}^a \tilde{\mathbf{P}}^b$. Since the lattice reduced precoding matrix $\tilde{\mathbf{P}}^b$ has near orthogonal columns, the required transmit power will be reduced compared to the BD-type precoding algorithms. Thus, a better BER performance than that of the BD-type precoding algorithms can be achieved by the proposed LR-S-GMI-type precoding algorithms.

The received signal is finally obtained as

$$\mathbf{y} = \mathbf{H} \tilde{\mathbf{P}} \mathbf{d} + \sqrt{\gamma} \mathbf{n}, \quad (37)$$

where $\gamma = \|\tilde{\mathbf{P}} \mathbf{d}\|^2$. The main processing work left for the receiver is to quantize the received signal \mathbf{y} to the nearest data vector and the decoding matrix \mathbf{G} described in the BD-type [9], [11], QR/SVD-type [14], and GMI-type [13] precoding algorithms is not needed anymore. The receiver structure is thus simplified, and a significant amount of transmit power can be saved which is very important considering the mobility of the distributed users.

The proposed precoding algorithms are called LR-S-GMI-ZF and LR-S-GMI-MMSE depending on the choice of the second precoding filter as given in (31) and (35), respectively. We will focus on the LR-S-GMI-MMSE precoding since a better performance is achieved. The implementing steps of the LR-S-GMI-MMSE precoding algorithm are summarized in Table II. By replacing the steps (8) and (9) in Table II with the formulation in (31), the LR-S-GMI-ZF precoding algorithm can be obtained. Similarly, the first precoding matrix can also be computed according to the GZI method in (28), and combined with (31) or (35) to get the second precoding matrix. Then, the LR-GZI-ZF or LR-GZI-MMSE precoding algorithms can be obtained, respectively.

V. PERFORMANCE ANALYSIS

In this section, we carry out an analysis of the performance of the proposed LR-S-GMI-type precoding algorithms. We consider a performance analysis in terms of BER, sum-rate and computational complexity. In the BER analysis part, we show that the residual interference matrix of the RBD precoding actually converges to an identity matrix, which is a new result in the literature so far. We also mathematically demonstrate that the residual interference of the proposed LR-S-GMI-type precoding algorithms converges to a zero matrix. Finally, we illustrate the quality of the effective channel matrices of the proposed and existing precoding algorithms by

TABLE II
 THE LR-S-GMI-MMSE PRECODING ALGORITHM

Steps	Operations
Applying the MMSE Channel Inversion	
(1)	$\mathbf{H}_{\text{mse}}^\dagger = (\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I})^{-1} \mathbf{H}^H$
(2)	for $i = 1 : K$
(3)	$[\mathbf{Q}_{i,\text{mse}}^\dagger \ \mathbf{R}_{i,\text{mse}}^\dagger] = \text{QR}(\mathbf{H}_{i,\text{mse}}^\dagger, 0)$
(4)	$\mathbf{P}_i^a = \mathbf{Q}_{i,\text{mse}}^\dagger$
(5)	$\mathbf{H}_{\text{eff}_i} = \mathbf{H}_i \mathbf{P}_i^a$
(6)	$\underline{\mathbf{H}}_{\text{eff}_i} = [\mathbf{H}_{\text{eff}_i} \ \sqrt{\alpha} \mathbf{I}_{N_i}]$
(7)	$[\underline{\mathbf{T}}_i^T \ \underline{\mathbf{H}}_{\text{eff}_i}^T] = \text{CLLL}(\underline{\mathbf{H}}_{\text{eff}_i}^T)$
(8)	$\mathbf{A}_i = [\mathbf{I}_{M_i} \ \mathbf{0}_{M_i \times N_i}]$
(9)	$\tilde{\mathbf{P}}_{\text{MMSE}_i}^b = \mathbf{A}_i \underline{\tilde{\mathbf{H}}}_{\text{eff}_i} (\underline{\tilde{\mathbf{H}}}_{\text{eff}_i} \underline{\tilde{\mathbf{H}}}_{\text{eff}_i}^H)^{-1}$
(10)	end
Compute the overall precoding matrix	
(11)	$\mathbf{P}^a = [\mathbf{P}_1^a, \mathbf{P}_2^a, \dots, \mathbf{P}_K^a]$
(12)	$\tilde{\mathbf{P}}^b = \text{diag}\{\tilde{\mathbf{P}}_1^b, \tilde{\mathbf{P}}_2^b, \dots, \tilde{\mathbf{P}}_K^b\}$
(13)	$\tilde{\mathbf{P}} = \mathbf{P}^a \tilde{\mathbf{P}}^b$
Calculate the scaling factor γ	
(14)	$\gamma = (\ \tilde{\mathbf{P}} \mathbf{d}\ _F^2 / E_s)$
Get the received signal	
(15)	$\mathbf{y} = \mathbf{H} \tilde{\mathbf{P}} \mathbf{d} + \sqrt{\gamma} \mathbf{n}$
Transform back from lattice space	
(16)	$\hat{\mathbf{d}} = \underline{\mathbf{T}}[\mathbf{y}]$

measuring their condition numbers. The maximum achievable sum-rate of the proposed LR-S-GMI-type precoding algorithms is given in the sum-rate analysis part. The computational complexity of the proposed and existing precoding algorithms is summarized in tables in the complexity analysis part. The trend of the computational complexity with the increase of the dimensions is also given and an analysis is developed.

A. BER Performance Analysis

For the BD precoding, the effective SU-MIMO channels are strictly parallel between each other after the first precoding filtering. For the RBD precoding, however, the residual interference $\overline{\mathbf{H}}_i \mathbf{P}_i^a (\text{RBD})$ is not zero between the users. We use \mathbf{J}_f to denote $\overline{\mathbf{H}}_i \mathbf{P}_i^a (\text{RBD})$ for convenience. From (13), the following formula is obtained

$$\mathbf{J}_f \mathbf{J}_f^H = \overline{\mathbf{H}}_i \overline{\mathbf{V}}_i (\overline{\boldsymbol{\Sigma}}_i^T \overline{\boldsymbol{\Sigma}}_i + \alpha \mathbf{I}_{N_T})^{-1} \overline{\mathbf{V}}_i^H \overline{\mathbf{H}}_i^H. \quad (38)$$

Mathematically, the quantity $\overline{\mathbf{V}}_i (\overline{\boldsymbol{\Sigma}}_i^T \overline{\boldsymbol{\Sigma}}_i + \alpha \mathbf{I}_{N_T})^{-1} \overline{\mathbf{V}}_i^H$ can be expressed as $(\overline{\mathbf{H}}_i^H \overline{\mathbf{H}}_i + \alpha \mathbf{I}_{N_T})^{-1}$. Substituting this into (38), the formula can be rewritten as

$$\mathbf{J}_f \mathbf{J}_f^H = \overline{\mathbf{H}}_i (\overline{\mathbf{H}}_i^H \overline{\mathbf{H}}_i + \alpha \mathbf{I}_{N_T})^{-1} \overline{\mathbf{H}}_i^H. \quad (39)$$

With the increase of the SNR, α approaches zero and then we have

$$\mathbf{J}_f \mathbf{J}_f^H \approx \overline{\mathbf{H}}_i (\overline{\mathbf{H}}_i^H \overline{\mathbf{H}}_i)^{-1} \overline{\mathbf{H}}_i^H. \quad (40)$$

By further manipulating the expression in (40), we obtain

$$\mathbf{J}_f \mathbf{J}_f^H \overline{\mathbf{H}}_i \approx \overline{\mathbf{H}}_i (\overline{\mathbf{H}}_i^H \overline{\mathbf{H}}_i)^{-1} \overline{\mathbf{H}}_i^H \overline{\mathbf{H}}_i = \overline{\mathbf{H}}_i$$

$$\text{Thus } \mathbf{J}_f \mathbf{J}_f^H \approx \mathbf{I}_{N_T}, \quad (41)$$

that is, the residual interference matrix \mathbf{J}_f of the RBD precoding converges to an identity matrix at high SNRs. While,

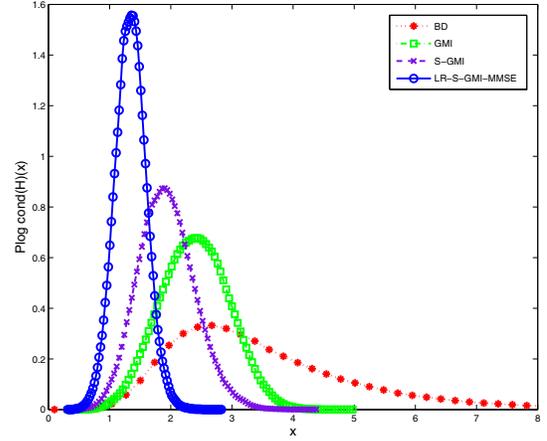


Fig. 2. PDFs of the natural logarithm of $\text{cond}(\mathbf{H})$ for 6×6 matrices

for the S-GMI precoding algorithm developed in Section IV with the SNR increase we have

$$\overline{\mathbf{H}}_i \mathbf{P}_i^a = \overline{\mathbf{H}}_i \mathbf{Q}_{i,\text{mse}}^\dagger \approx \mathbf{0}. \quad (42)$$

By comparing (41) and (42), we can see that the impact of the residual interference of S-GMI precoding would be smaller than that of the conventional RBD precoding algorithm. Thus, we expect that a better BER performance is achieved by the S-GMI precoding algorithm over the conventional RBD precoding algorithm.

As pointed out in [19], the BER performance for a MIMO precoding system is actually determined by the energy of the transmitted signal γ . In order to reduce γ and improve the BER performance further, we transform the effective channel \mathbf{H}_{eff} into the lattice space. By doing this, an improved basis $\tilde{\mathbf{H}}_{\text{eff}}$ is computed. Actually, the LR transformed channel matrix $\tilde{\mathbf{H}}_{\text{eff}}$ is quasi-orthogonal rather than strictly orthogonal. We can employ the condition number which is defined as [15]

$$\text{cond}(\mathbf{H}) = \|\mathbf{H}\|_F \|\mathbf{H}^{-1}\|_F \quad (43)$$

to measure the orthogonality of the channel matrix. From the above definition of the condition number in (43), we get that $\text{cond}(\mathbf{H}) = 1$ with equality for an orthogonal basis while matrices which are nearly singular have large condition numbers. In Fig. 2, the probability density functions (PDFs) of the condition numbers for the effective channel matrices are illustrated. For the effective channel matrix of the proposed LR-S-GMI-MMSE precoding algorithm, not only the spread in the condition numbers but also their average value is much smaller compared to the effective channel matrices achieved by the existing precoding algorithms. Therefore, a significant reduction in the required transmit power γ is achieved and a better BER performance can be obtained by the proposed LR-S-GMI-MMSE precoding algorithm. Note for the special case of each user with a single receive antenna, the proposed LR-S-GMI-type precoding will not converge to GMI or S-GMI because the second precoding filter is designed in the lattice space.

B. Achievable Sum-Rate Analysis

Recall that at high SNRs, the MU-MIMO channel is approximately decoupled into equivalent SU-MIMO channels by applying the first precoding filtering in (23). Then, we can transform the MU-MIMO sum-rate analysis [27] to a set of SU-MIMO sum-rate analysis tasks. For the second precoding filter, the LR-aided MMSE precoding is actually equal to the LR-aided ZF precoding under the high SNR scenario. Therefore, the i th user's received signal is

$$\mathbf{y}_i = \mathbf{z}_i + \sqrt{\gamma_i} \mathbf{n}_i, \quad (44)$$

where $\mathbf{z}_i = \mathbf{T}_i^{-1} \mathbf{d}_i$. By assuming that the average transmit power is $\xi_i = 1$, and because of the fact that $\mathbf{H}_{\text{eff}_i} = \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{V}_i^H$, we get the normalization factor γ_i as

$$\begin{aligned} \gamma_i &= \|\mathbf{H}_{\text{eff}_i}^{-1} \mathbf{z}_i\|_{\text{F}}^2 = \text{Tr}(\boldsymbol{\Sigma}_i^{-2} \mathbf{z}_i \mathbf{z}_i^H) \\ &= \sum_{l=1}^{L_{\text{eff}}} \frac{\xi_l^2}{\lambda_l^2}, \end{aligned} \quad (45)$$

where the quantity λ_l is the l th singular value of $\boldsymbol{\Sigma}_i$, and ξ_l denotes the energy of the l th stream of \mathbf{z}_i .

From (45), the received SNR for the l th stream of user i is obtained as

$$\text{SNR}_l = \frac{\xi_l^2}{\sigma_n^2 \sum_{m=1}^{L_{\text{eff}}} \frac{\xi_m^2}{\lambda_m^2}}. \quad (46)$$

Then, the achievable sum-rate for user i is given by

$$C_i = \sum_{l=1}^{L_{\text{eff}}} \log \left(1 + \frac{\xi_l^2}{\sigma_n^2 \sum_{m=1}^{L_{\text{eff}}} \frac{\xi_m^2}{\lambda_m^2}} \right) = \sum_{l=1}^{L_{\text{eff}}} \log \left(1 + \frac{\xi_l^2}{\sigma_n^2 \gamma_i} \right). \quad (47)$$

Note that the achievable sum-rate C_i is degraded by the normalization factor γ_i . The value of C_i approaches its maximum when $\frac{\xi_1^2}{\lambda_1^2} = \frac{\xi_2^2}{\lambda_2^2} = \dots = \frac{\xi_{L_{\text{eff}}}^2}{\lambda_{L_{\text{eff}}}^2}$, thus we have

$$C_i \leq \sum_{l=1}^{L_{\text{eff}}} \log \left(1 + \frac{\lambda_l^2}{\sigma_n^2 L_{\text{eff}}} \right). \quad (48)$$

Finally, the maximum achievable sum-rate of the proposed LR-S-GMI-type precoding algorithms at high SNRs can be expressed as

$$C = \sum_{i=1}^K \sum_{l=1}^{L_{\text{eff}}} \log_2 \left(1 + \frac{\lambda_l^2}{\sigma_n^2 L_{\text{eff}}} \right). \quad (49)$$

For the BD precoding, we multiply the decoding matrix $\mathbf{G}_i = \mathbf{U}_i^H$ at the i th user's receiver and the received signal is given by

$$\mathbf{y}_i = \boldsymbol{\Sigma}_i \bar{\mathbf{V}}_i^{(0)} \mathbf{V}_i^{(1)} \mathbf{d}_i + \mathbf{U}_i^H \mathbf{n}_i. \quad (50)$$

Due to the fact that the two precoding matrices $\bar{\mathbf{V}}_i^{(0)}$ and $\mathbf{V}_i^{(1)}$ are semi-unitary matrices, we get $\bar{\mathbf{V}}_i^{(0)H} \bar{\mathbf{V}}_i^{(0)} = \mathbf{I}$ and $\mathbf{V}_i^{(1)H} \mathbf{V}_i^{(1)} = \mathbf{I}$. Then, by applying the equivalence $\text{Tr}(\mathbf{ABC}) = \text{Tr}(\mathbf{CAB})$, the normalization factor $\gamma_i^{(\text{BD})}$ for BD can be expressed as

$$\gamma_i^{(\text{BD})} = \|\bar{\mathbf{V}}_i^{(0)} \mathbf{V}_i^{(1)} \mathbf{d}_i\|^2 = \|\mathbf{d}_i\|^2. \quad (51)$$

Since the statistical property of \mathbf{n}_i is not changed by the multiplication with the unitary matrix \mathbf{U}_i^H , we get the l th received SNR as

$$\text{SNR}_l = \frac{\lambda_l^2}{\sigma_n^2}. \quad (52)$$

For simplicity, we do not consider the power loading between users and streams in the following derivation and term this strategy as no power loading (NPL). Then, the achievable sum-rate for the BD precoding algorithm is given by

$$C^{(\text{BD})} = \sum_{i=1}^K \sum_{l=1}^{L_{\text{eff}}} \log_2 \left(1 + \frac{\lambda_l^2}{\sigma_n^2} \right). \quad (53)$$

By comparing the maximum achievable sum-rate of the proposed LR-S-GMI-type precoding algorithms in (49), we conclude that the sum-rate of the proposed LR-S-GMI-type precoding algorithms will be slightly inferior to that of the BD precoding algorithm at high SNRs. At low SNRs, however, we expect that the achieved sum-rate of the proposed LR-S-GMI-type precoding algorithms will be better than that of the BD precoding since a regularization factor is employed to mitigate the degradation by the noise term.

The sum-rate performance of the BD precoding is actually dependent on the power loading scheme being used. Hence, the BD precoding algorithm can achieve its maximum sum-rate performance by allocating the power between streams according to a WF power loading scheme. As pointed out in [13], we do not consider the power loading strategy for the RBD or the proposed LR-S-GMI-type precoding algorithms for two reasons. One is that it is not easy to identify the optimal power allocation coefficients because of the existence of residual interference. The second reason is that the MMSE condition (10) is already satisfied. Therefore, an allocation of different powers between streams is not needed.

C. Computational Complexity Analysis

In this section, we use the total number of floating point operations (FLOPs) to measure the computational complexity of the precoding algorithms discussed above. It is worth noting that the lattice reduction algorithm has variable complexity, and the average complexity of the CLLL algorithm has been given in FLOPs by [22]. A reduced and fixed complexity lattice reduction structure is proposed in [28], however, we employ the conventional CLLL algorithm for the reason that the lattice reduction algorithm is not the main focus in this work. The number of FLOPs for the complex QR decomposition and the real SVD operation are given in [15]. As shown in [17], the number of FLOPs required by a $m \times n$ complex SVD operation is equivalent to its extended $2m \times 2n$ real matrix. The total number of FLOPs needed by the matrix operations is summarized below:

- Multiplication of $m \times n$ and $n \times p$ complex matrices: $8mnp - 2mp$;
- QR decomposition of an $m \times n$ ($m \leq n$) complex matrix: $16(n^2m - nm^2 + \frac{1}{3}m^3)$;
- SVD of an $m \times n$ ($m \leq n$) complex matrix where only $\boldsymbol{\Sigma}$ and \mathbf{V} are obtained: $32(nm^2 + 2m^3)$;

TABLE III
 COMPUTATIONAL COMPLEXITY OF CONVENTIONAL RBD

Steps	Operations	Flops	Case (2, 2, 2) × 6
1	$\bar{\mathbf{U}}_i \bar{\Sigma}_i \bar{\mathbf{V}}_i^H$	$32K(N_T \bar{N}_i^2 + 2\bar{N}_i^3)$	21504
2	$(\bar{\Sigma}_i^T \bar{\Sigma}_i + \alpha \mathbf{I}_{N_T})^{-\frac{1}{2}}$	$K(18N_T + \bar{N}_i)$	336
3	$\bar{\mathbf{V}}_i \bar{\mathbf{D}}_i, (\bar{\mathbf{D}}_i \leftarrow 2)$	$8K N_T^3$	5184
4	$\mathbf{H}_i \mathbf{P}_i^a$	$K(8N_i^2 N_T - 2N_i^2)$	552
5	$\mathbf{U}_i \Sigma_i \mathbf{V}_i^H$	$64K(\frac{8}{3}N_i^3 + N_T N_i^2 + \frac{1}{2}N_T^2 N_i)$	13248 Total 40824

 TABLE IV
 COMPUTATIONAL COMPLEXITY OF S-GMI

Steps	Operations	Flops	Case (2, 2, 2) × 6
1	$\mathbf{H}_{i,\text{mse}}^\dagger$	$(\frac{4}{3}N_R^3 + 12N_R^2 N_T - 2N_R^2 - 2N_R N_T)$	2736
2	$\mathbf{Q}_i^\dagger \mathbf{R}_i^\dagger$	$16K(N_T^2 N_i - N_T N_i^2 + \frac{1}{3}N_i^3)$	2432
3	$\mathbf{H}_i \mathbf{P}_i^a$	$K(8N_i^2 N_T - 2N_i^2)$	552
4	$\mathbf{U}_i \Sigma_i \mathbf{V}_i^H$	$64K(\frac{8}{3}N_i^3 + N_T N_i^2 + \frac{1}{2}N_T^2 N_i)$	13248 Total 18968

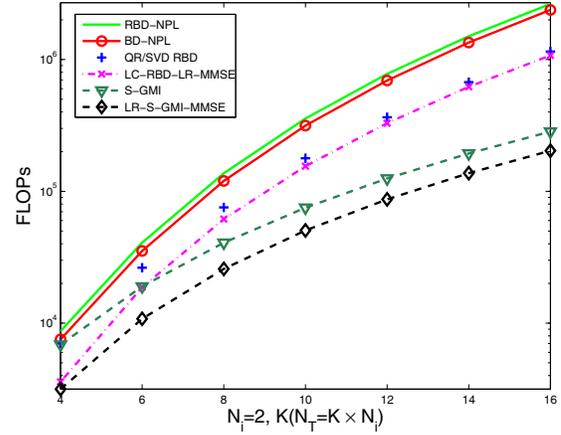
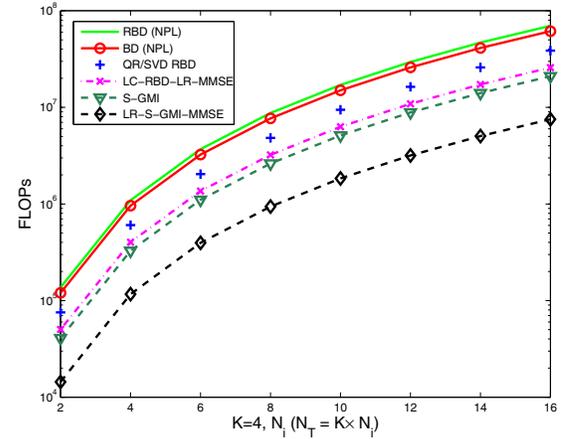
 TABLE V
 COMPUTATIONAL COMPLEXITY OF LR-S-GMI-MMSE

Steps	Operations	Flops	Case (2, 2, 2) × 6
1	$\mathbf{H}_{i,\text{mse}}^\dagger$	$(\frac{4}{3}N_R^3 + 12N_R^2 N_T - 2N_R^2 - 2N_R N_T)$	2736
2	$\mathbf{Q}_i^\dagger \mathbf{R}_i^\dagger$	$16K(N_T^2 N_i - N_T N_i^2 + \frac{1}{3}N_i^3)$	2432
3	$\mathbf{H}_i \mathbf{P}_i^a$	$K(8N_i^2 N_T - 2N_i^2)$	552
4	$\underline{\mathbf{H}}_{\text{eff},i}$	CLLL	4787.58
5	$\underline{\mathbf{H}}_{\text{eff},i}^\dagger$	$K(\frac{4}{3}N_i^3 + 12N_i^3 - 4N_i^2)$	272 Total 10780

- SVD of an $m \times n$ ($m \leq n$) complex matrix where \mathbf{U} , Σ and \mathbf{V} are obtained: $8(4n^2m + 8nm^2 + 9m^3)$;
- Inversion of an $m \times m$ real matrix using Gauss-Jordan elimination: $4m^3/3$.

We illustrate the required FLOPs for the conventional RBD, S-GMI and LR-S-GMI-MMSE precoding algorithms in Table III, Table IV and Table V, respectively. The complexity of the QR/SVD RBD [14] and LC-RBD-LR-MMSE precoding algorithms is already given in [17]. A system with $N_T = 6$ transmit antennas and $K = 3$ users each equipped with $N_i = 2$ receive antennas is considered; this scenario is denoted as the (2, 2, 2) × 6 case. For the (2, 2, 2) × 6 case, the reduction in the number of FLOPs obtained by the proposed LR-S-GMI-MMSE precoding is 73.6%, 69.5%, 59.1% and 49.9% as compared to the RBD, BD, QR/SVD RBD and LC-RBD-LR-MMSE precoding algorithms, respectively. Clearly, the proposed LR-S-GMI-MMSE precoding algorithm requires the lowest complexity.

In order to further reveal the relationship between the computational complexity and the system dimensions, we first fix the receive antenna configuration and assume that each user is equipped with $N_i = 2$ antennas. Fig. 3 shows that with N_i fixed, the computational complexity of the BD-type precoding algorithms grows relatively faster than the other precoding algorithms with the increase of K . The reason is that, the first


 Fig. 3. Computational Complexity - I Fixed N_i

 Fig. 4. Computational Complexity - II Fixed K

SVD operation of the RBD precoding is implemented K times on $\bar{\mathbf{H}}_i$ with dimension $\bar{N}_i \times N_T$. The QR decomposition of the LC-RBD-LR-type precoding also needs to be implemented K times on $\bar{\mathbf{H}}_i$, but it requires less FLOPs because the QR decomposition is much simpler than the SVD operation in the case of same matrix dimensions [15]. While, the S-GMI only needs one common channel inversion and the QR decomposition is computed on $\mathbf{H}_{i,\text{mse}}$ with a lower dimension $N_i \times N_i$.

Similarly, if we fix the number of users to $K = 4$, the system dimensions will change with the variable N_i . From Fig. 4, the S-GMI precoding algorithm can offer a much lower complexity than that of the BD, RBD and QR/SVD RBD precoding algorithms. The reason is that, with K fixed, the first K SVD operations have a higher cost than the common channel inversion in (20) plus the QR decompositions in (24).

The complexity of the proposed LR-S-GMI-MMSE precoding algorithm, however, shows the lowest computational complexity in Fig. 3 and Fig. 4. The reason is that we use a less complex LR transformation to replace the second SVD operation in S-GMI precoding algorithm. It is worth noting that with the increase of the system dimensions, the complex-

ity reduction by the proposed LR-S-GMI-MMSE precoding algorithm becomes more considerable.

VI. SIMULATION RESULTS

A system with $N_T = 8$ transmit antennas and $K = 4$ users each equipped with $N_i = 2$ receive antennas is considered; this scenario is denoted as the $(2, 2, 2, 2) \times 8$ case. The vector \mathbf{d}_i of the i th user represents the data transmitted with QPSK modulation.

A. Perfect Channel Scenario

The channel matrix \mathbf{H}_i of the i th user is modeled as a complex Gaussian channel matrix with zero mean and unit variance. We assume an uncorrelated block fading channel, that is, the channel is static during each transmit packet and there is no correlation between the antennas. We also assume that the channel estimation is perfect at the receive side and the feedback channel is error free. The number of simulation trials is 10^6 and the packet length is 10^2 symbols. The E_b/N_0 is defined as $E_b/N_0 = \frac{N_R \xi}{N_T M \sigma_n^2}$ with M being the number of transmitted information bits per channel symbol.

Fig. 5 shows the BER performance of the proposed and existing precoding algorithms. The GMI, QR/SVD RBD and RBD precoding algorithms share the same BER performance. A better BER performance is achieved by the proposed S-GMI precoding scheme compared to the BD, GMI, QR/SVD RBD, and RBD precoding algorithms. The reason is that the residual interference between the users can be suppressed further by the S-GMI precoding scheme as we analyzed in Section V.A. The proposed LR-S-GMI-MMSE precoding algorithm shows the best BER performance. At the BER of 10^{-2} , the LR-S-GMI-MMSE precoding has a gain of more than 5.5 dB compared to the RBD precoding. It is worth noting that the BER performance of the RBD precoding is outperformed by the proposed LR-S-GMI-MMSE precoding in the whole E_b/N_0 range and gains become more significant with the increase of E_b/N_0 . From the analysis developed in Section V.A, a better channel quality as measured by the condition number of the effective channel is achieved by the proposed LR-S-GMI-MMSE precoding. Therefore, the required transmit power γ is reduced and a better BER performance is obtained.

Fig. 6 illustrates the sum-rate of the proposed and existing precoding algorithms. The sum-rate is calculated using [27]:

$$C = \log(\det(\mathbf{I} + \sigma_n^{-2} \mathbf{H} \mathbf{P} \mathbf{P}^H \mathbf{H}^H)) \quad (\text{bits/Hz}). \quad (54)$$

The BD precoding with WF power loading shows a better sum-rate performance than the BD precoding without power loading as we mentioned in Section V.B. However, as shown in Fig. 5, the BER performance is degraded by applying this WF scheme. Similar to the BER figure, the GMI, QR/SVD RBD, and RBD precoding algorithms show a comparable sum-rate performance. The S-GMI precoding also achieves the sum-rate performance of the RBD precoding. The proposed LR-S-GMI-MMSE precoding algorithm illustrates almost the same sum-rate performance as the RBD precoding at low E_b/N_0 s. At high E_b/N_0 s, however, the sum-rate of LR-S-GMI-MMSE precoding is slightly inferior to that of the RBD precoding

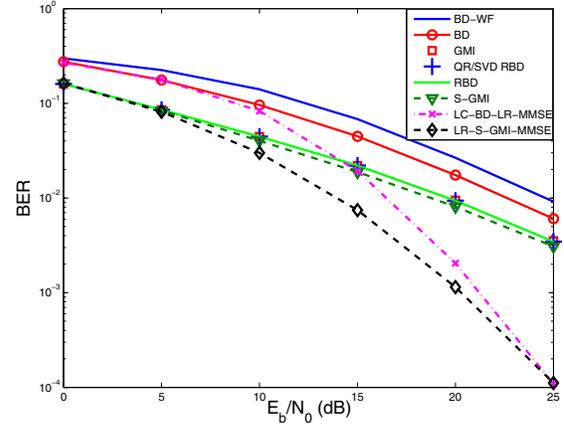


Fig. 5. BER performance, $(2, 2, 2, 2) \times 8$ MU-MIMO

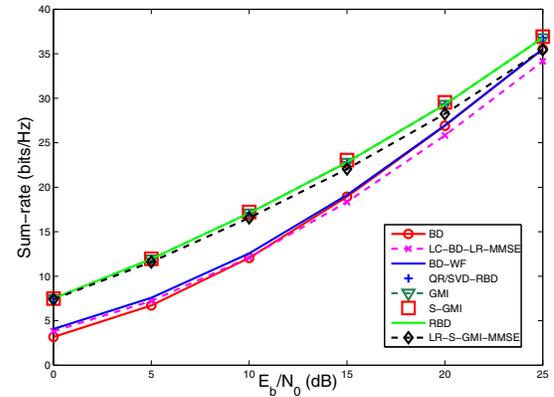


Fig. 6. Sum-rate performance, $(2, 2, 2, 2) \times 8$ MU-MIMO

and approaches the performance of the BD precoding as we analyzed in Section V.B.

B. The impact of imperfect channels

The use of perfect CSI is impractical in wireless systems due to the often inaccurate channel estimation and the CSI feedback errors. From [18], [29], the estimation errors or feedback errors can be modeled as a complex random Gaussian noise \mathbf{E} with i.i.d. entries of zero mean and variance σ_e^2 . Another factor that usually needs to be considered in the MU-MIMO systems is the antenna correlation at the transmit side [30]. In this work, we simulate the correlated channel based on the exponential correlation model in [31]. The imperfect channel matrix \mathbf{H}_e is defined as

$$\mathbf{H}_e = \mathbf{H} \mathbf{R}_T^{\frac{1}{2}} + \mathbf{E}, \quad (55)$$

where the quantity \mathbf{R}_T is a transmit covariance matrix with the elements defined below

$$R_{ij} = \begin{cases} r^{j-i}, & i \leq j \\ r_{ji}^*, & i > j \end{cases}, |r| \leq 1 \quad (56)$$

where r is the correlation coefficient between any two neighboring antennas. The precoding matrix \mathbf{P} has to be designed based on the imperfect channel \mathbf{H}_e while the physical channel is \mathbf{H} during each transmission.

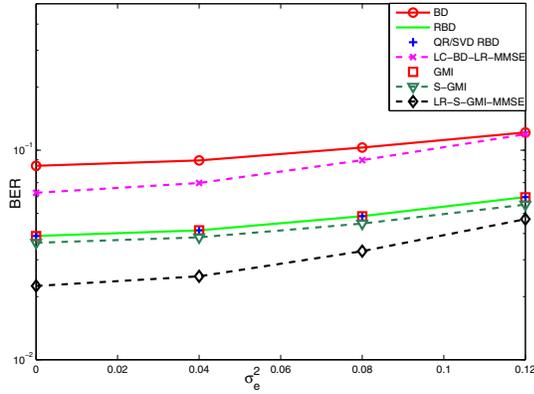


Fig. 7. BER as a function of σ_e^2 for a fixed $E_b/N_0=15$ dB

Fig. 7 gives the BER performance of the above precoding algorithms under H_e with $|r| = 0.2$ at $E_b/N_0 = 15$ dB. All the above precoding algorithms are affected by the imperfect channel H_e . The proposed LR-S-GMI-MMSE precoding algorithm outperforms the RBD precoding algorithm when σ_e^2 is below 0.12, however, the performance of the RBD precoding algorithm decays more slowly.

VII. CONCLUSION

Based on a channel inversion technique, low-complexity high-performance LR-S-GMI-type precoding algorithms have been proposed for MU-MIMO systems in this paper. Compared to the BD-type precoding algorithms, the complexity of the precoding process is reduced and a considerable BER gain is achieved by the proposed algorithms at a cost of a slight sum-rate loss at high SNRs. The BER performance, the achievable sum-rate and the computational complexity of the LR-S-GMI-type precoding algorithms have been analyzed and compared to existing precoding algorithms. Since the LR-S-GMI-type precoding algorithms are only implemented at the transmit side, the decoding matrix is not needed any more at the receive side compared to the BD-type precoding algorithms. Then, the structure of the receiver can be simplified, which is an additional benefit from the proposed LR-S-GMI-type precoding algorithms. The proposed algorithms also show a robust performance in the presence of imperfect CSI and spatial correlation, which emphasizes their value for practical applications.

REFERENCES

[1] K. Zu, R. C. de Lamare, and M. Haardt, "Low-complexity lattice reduction-aided channel inversion methods for large multi-user MIMO systems," in *Proc. 2012 IEEE Asilomar Conf. Signals, Syst. Comput.*
 [2] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.
 [3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.
 [4] Requirements for Further Advancements for E-UTRA (LTE-Advanced), 3GPP TR 36.913 Standard, 2011.
 [5] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Enhancements for Very High Throughput for Operation in Bands Below 6GHz, IEEE P802.11ac/D1.0 Standard, Jan. 2011.

[6] C. Chiu, J. Yan, and R. Murch, "24-port and 36-port antenna cubes suitable for MIMO wireless communications," *IEEE Trans. Antennas Propag.*, vol. 56, no. 4, pp. 1170–1176, Apr. 2008.
 [7] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, to be published.
 [8] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53 no. 8, pp. 2700–2712, Aug. 2005.
 [9] 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, no. 2, pp. 461–471, Feb. 2004.
 [10] L. U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, Jan. 2004.
 [11] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 953–961, Mar. 2008.
 [12] C. B. Chae, S. Shim, and R. W. Heath, "Block diagonalized vector perturbation for multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4051–4057, Nov. 2008.
 [13] H. Sung, S. Lee, and I. Lee, "Generalized channel inversion methods for multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3489–3499, Nov. 2009.
 [14] H. Wang, L. Li, L. Song, and X. Gao, "A linear precoding scheme for downlink multiuser MIMO precoding systems," *IEEE Commun. Lett.*, vol. 15, no. 6, pp. 653–655, June 2011.
 [15] G. Golub and C. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.
 [16] D. Wübben, D. Seethaler, J. Jaldén, and G. Matz, "Lattice reduction: a survey with applications in wireless communications," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, May 2011.
 [17] K. Zu and R. C. de Lamare, "Low-complexity lattice reduction-aided regularized block diagonalization for MU-MIMO systems," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 925–928, June 2012.
 [18] K. Zu, R. C. de Lamare, and M. Haardt, "Lattice reduction-aided regularized block diagonalization for multiuser MIMO systems," in *Proc. 2012 IEEE Wireless Commun. Netw. Conf.*, pp. 131–135.
 [19] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near capacity multiantenna multiuser communication—part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 195–202, Jan. 2005.
 [20] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near capacity multiantenna multiuser communication—part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.
 [21] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.
 [22] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2701–2710, July 2009.
 [23] C. Windpassinger and R. Fischer, "Low-complexity near-maximum likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. 2003 IEEE Inf. Theory Workshop*, pp. 345–348.
 [24] D. Wübben, R. Böhnke, V. Kühn, and K. Kammeyer, "Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice-reduction," in *Proc. 2004 IEEE International Conf. Commun.*, pp. 798–802.
 [25] R. Habendorf and G. Fettweis, "On ordering optimization for MIMO systems with decentralized receivers," in *Proc. 2006 IEEE Veh. Technol. Conf. - Spring*, pp. 1844–1848.
 [26] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4801–4805, Dec. 2007.
 [27] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. 2002 IEEE International Conf. Commun.*, pp. 1444–1450.
 [28] H. Vetter, V. Ponnampalam, M. Sandell, and P. A. Hoeher, "Fixed complexity LLL algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1634–1637, Apr. 2009.
 [29] C. Windpassinger, "Detection and precoding for multiple input multiple output channels," Ph.D. dissertation, Univ. Erlangen-Nurnberg, Erlangen, Germany, 2004.
 [30] M. T. Ivrlac, W. Utschick, and J. A. Nossek, "Fading correlations in wireless MIMO communications systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 819–828, June 2003.

- [31] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sept. 2001.



Keke Zu received the BSc degree in Communications Engineering from Southwest Jiaotong University, China in 2006 and the MSc degree in Communication & Information Systems from National Mobile Communications Research Laboratory, Southeast University, China in 2009. He is currently pursuing the Ph.D. degree at The University of York, UK. His research interests include MIMO precoding, MIMO detection, lattice reduction, and security in computing & communications.



Rodrigo C. de Lamare (S'99 - M'04 - SM'10) received the Diploma in electronic engineering from the Federal University of Rio de Janeiro (UFRJ) in 1998 and the M.Sc. and Ph.D. degrees, both in electrical engineering, from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 2001 and 2004, respectively. Since January 2006, he has been with the Communications Research Group, Department of Electronics, University of York, where he is currently a Reader. Since April 2012, he has also been a Professor at PUC-RIO. His research interests

lie in communications and signal processing, areas in which he has published about 300 papers in refereed journals and conferences. Dr. de Lamare serves as an associate editor for the *EURASIP Journal on Wireless Communications and Networking* and for *IEEE SIGNAL PROCESSING LETTERS*. He is a Senior Member of the IEEE and has served as the general chair of the 7th IEEE International Symposium on Wireless Communications Systems (ISWCS), held in York, UK in September 2010, and as the technical programme chair of ISWCS 2013 held in Ilmenau, Germany in August 2013.



Martin Haardt (S'90 - M'98 - SM'99) has been a Full Professor in the Department of Electrical Engineering and Information Technology and Head of the Communications Research Laboratory at Ilmenau University of Technology, Germany, since 2001. Since 2012, he has also served as an Honorary Visiting Professor in the Department of Electronics at the University of York, UK.

After studying electrical engineering at the Ruhr-University Bochum, Germany, and at Purdue University, USA, he received his Diplom-Ingenieur (M.S.) degree from the Ruhr-University Bochum in 1991 and his Doktor-Ingenieur (Ph.D.) degree from Munich University of Technology in 1996. In 1997 he joined Siemens Mobile Networks in Munich, Germany, where he was responsible for strategic research for third generation mobile radio systems. From 1998 to 2001 he was the Director for International Projects and University Cooperations in the mobile infrastructure business of Siemens in Munich, where his work focused on mobile communications beyond the third generation. During his time at Siemens, he also taught in the international Master of Science in Communications Engineering program at Munich University of Technology.

Martin Haardt has received the 2009 Best Paper Award from the IEEE Signal Processing Society, the Vodafone (formerly Mannesmann Mobilfunk) Innovations-Award for outstanding research in mobile communications, the ITG best paper award from the Association of Electrical Engineering, Electronics, and Information Technology (VDE), and the Rohde & Schwarz Outstanding Dissertation Award. In the fall of 2006 and the fall of 2007 he was a visiting professor at the University of Nice in Sophia-Antipolis, France, and at the University of York, UK, respectively. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, as well as numerical linear and multi-linear algebra.

Prof. Haardt has served as an Associate Editor for the *IEEE TRANSACTIONS ON SIGNAL PROCESSING* (2002-2006 and since 2011), the *IEEE SIGNAL PROCESSING LETTERS* (2006-2010), the *Research Letters in Signal Processing* (2007-2009), the *Hindawi Journal of Electrical and Computer Engineering* (since 2009), the *EURASIP Signal Processing Journal* (since 2011), and as a guest editor for the *EURASIP Journal on Wireless Communications and Networking*. He has also served as an elected member of the Sensor Array and Multichannel (SAM) technical committee of the IEEE Signal Processing Society (since 2011), as the technical co-chair of the IEEE International Symposiums on Personal Indoor and Mobile Radio Communications (PIMRC) 2005 in Berlin, Germany, as the technical program chair of the IEEE International Symposium on Wireless Communication Systems (ISWCS) 2010 in York, UK, as the general chair of ISWCS 2013 in Ilmenau, Germany, and as the general-co chair of the 5-th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) 2013 in Saint Martin, French Caribbean.