LINEAR GROUP PRECODING FOR MASSIVE MIMO SYSTEMS UNDER EXPONENTIAL SPATIAL CORRELATION

Dinh Van Khoi[1*], Le Minh Tuan[2], Ngo Vu Duc[3], Ta Chi Hieu[1]

[1]Le Quy Don Technical University; [2]MobiFone R&D Center, MobiFone Corporation; [3]Hanoi University of Science and Technology

Abstract

In this paper, a low-complexity linear group precoding algorithm in the exponential correlation channel model is proposed for massive MIMO systems. The proposed precoder consists of two components: The first one minimizes the interferences among neighboring user groups; The second one improves the system performance by utilizing the ELR-SLB technique. Numerical and simulation results show that the proposed precoder has remarkably lower computational complexity than its LC-RBD-LR-ZF counterpart, while its bit error rate (BER) performance is asymptotic to that of the LC-RBD-LR-ZF precoder as the number of groups increases.

Keywords: *MU-MIMO system; massive MIMO system; linear precoding algorithms; nonlinear precoding algorithms; lattice reduction algorithm in MIMO system.*

1. Introduction

Multiple-Input Multiple-Output (MIMO) technology has been widely studied for years and already implemented in 4G mobile communication systems [1]. The initial research focuses on point-to-point MIMO systems. In recent years, more and more researchers are interested in Multiuser MIMO (MU-MIMO) scenarios. However, a limitation of MU-MIMO system is that BS is usually equipped with small numbers of antenna elements (normally fewer than 10) [2]. Therefore, the spectrum efficiency and system capacity are still relatively modest.

To solve this problem, massive MIMO systems have recently been proposed [1, 3, 4, 5]. In the Massive MIMO, the number of antennas at the base station (BS) can be up to hundreds of antennas (or even thousands) to simultaneously serve dozens of users using the same frequency resource. The Massive MIMO system can significantly improve the channel capacity, enhance the spectrum utilization efficiency and quality of the system [4]. It is expected that Massive MIMO will be a key and a potential candidate for the next generation wireless network (e.g., 5G network) [1, 4, 6].

Although the massive MIMO systems have numerous advantages, they face a number of challenges such as hardware complexity, power consumption, and system cost due to the large number of antennas equipped at the BS. Therefore, reducing the

[*] Email: dinhvankhoi.tcu@gmail.com

complexity of the signal processing algorithms for both uplink and downlink in massive MIMO systems is essential.

In massive MIMO systems, the dimensions of transmit/receive signal vectors are normally very large due to large numbers of antennas and users. Therefore, the precoding algorithms with low-complexity, e.g. Zero Forcing (ZF), Minimum Mean Square Error (MMSE) and Maximum Ratio Transmission (MRT) are considered as suitable solutions [7, 8, 9]. The Dirty Paper Coding (DPC) proposed in [10] can achieve the capacity region for multiuser precoding. However, its complexity becomes significantly large as the system dimensions grow due to the implementation of random nonlinear encoding and decoding [11, 12].

The combination of lattice reduction algorithms and precoding techniques for the downlinks of massive MIMO systems is an important solution to improve the system performance. In [13], the authors adopted the Seysen's lattice reduction algorithm (SA) to create a LR-aided precoding technique for the MU-MIMO system. The simulation results show that the proposed algorithm gives better performance than the precoding algorithm adopting the Lenstra-Lenstra-Lovasz (LLL) method. In [14], a Block Diagonalization (BD) aided precoding algorithm was proposed based on the Pseudo-Inverse Block Diagonalization (PINVBD) presented in [15] and the QR decomposition of the channel matrix. Furthermore, in each block, the Lattice Reduction and Tomlinson-Halashima precoder (THP) algorithms are applied to improve the quality of the system. In [16], the authors proposed the low-complexity Lattice Reduction (LR)-aided BD algorithms for the MU-MIMO, referred to as LC-RBD-LRZF and LC-RBD-LR-MMSE. In the authors' proposal, the first precoding matrix is obtained using the QR decomposition instead of the Singular Value Decomposition (SVD). The second precoding matrix is computed based on either the ZF or MMSE algorithm to provide the corresponding LC-RBD-LR-ZF or LC-RBD-LR-MMSE precoder. It was shown in [16] that the precoders significantly improved the system performance, while reducing the computational complexity compared to the original BD one. However, the computational complexities of the precoders presented in [14] and [16] are still very high due to the adoptions of the QR decomposition and THP algorithms.

In this paper, we propose a low complexity precoding algorithm for massive MIMO systems using the exponential correlation channel model. Based on the linear precoding algorithms and the lattice reduction technique, we propose the Zero Forcing group precoder combining with the low-complexity lattice reduction technique (or ZF-GP-LR precoder for short). In our proposal, the channel matrix from the BS to all users is divided into L groups (i.e., sub-matrices), each of which contains a number of rows of the channel matrix. The sizes of the sub-matrices are all the same. The proposed precoding matrix is designed to have two components: The first one minimizes the interferences from neighboring user groups by using QR decomposition of the sub-matrices; The second one enhances the system performance thanks to the combination of the Zero Forcing precoding and the ELR-SLB lattice reduction algorithms. Numerical and simulation results show that the ZF-GP-LR precoder has remarkably lower computational complexity than the LC-RBD-LR-ZF in [16], whereas its BER performance is asymptotic to that of the LC-RBD-LR-ZF as L increases. Besides, the complexity of the proposed algorithm grows proportionally to the number of groups. Simulation results also show that the spatial correlation adversely affects the system performance no matter which precoder is adopted. Fortunately, the proposed precoder still works well as compared with the LC-RBD-LR-ZF under such circumstances.

The rest of this paper is organized as follows. In Section 2, we present massive MIMO system model. The LC-RBD-LR-ZF and element-based lattice reduction (ELR) algorithms are reviewed in Section 3. The linear group precoding algorithm in combination with ELR-SLB technique is presented in Section 4. Simulation results are evaluated in Section 5. Finally, conclusions are drawn in Section 6.

Notation: The notations are defined as follows: Matrices and vectors are represented by symbols in bold; $(.)^T$ and $(.)^H$ denote the transpose and conjugate transpose, respectively. We denote $|a|$ for the absolute value of scalar a and det(**B**) for the determinant of **B**. $\lfloor \alpha \rceil$ rounds the real and imaginary parts of the complex number α to the nearest integers. $Tr\{.\}$ is the trace of a square matrix.

2. The downlink channel model in massive MIMO system

Let us consider a massive MIMO system, where the BS is equipped with N_T antennas to simultaneously serve K users, each user has N_u antennas. Thus, the total number of antennas of K users is $N_R = KN_u$. In addition, the Channel State Information (CSI) is assumed to be perfectly known at the BS. In reality, although the theoretical distance is guaranteed, there till exist certain amounts of correlation among the antennas. These correlation can be modeled based on the actual measurements. Therefore, spatial correlations always exist among transmit and receive antennas, thereby degrading the system performance. In order to take into account the effect of the spatial correlation, the channel model is given by the following equation [17]:

$$\mathbf{H}_{corr} = \mathbf{R}_R^{1/2}\mathbf{H}_{\omega}\mathbf{R}_T^{1/2}, \tag{1}$$

where $\mathbf{H}_{corr} = \left[(\mathbf{H}_{corr_1})^T\ (\mathbf{H}_{corr_2})^T ... (\mathbf{H}_{corr_K})^T\right]^T \in \mathbb{C}^{N_R \times N_T}$ is the channel matrix with antenna correlations, $\mathbf{R}_T$ is the $N_T \times N_T$ transmit correlation matrix and $\mathbf{R}_R$ is the

$N_R \times N_R$ receive correlation matrix. $\mathbf{H}_\omega$ is the uncorrelated channel matrix, whose entries, h_{ij}^{ω}, are complex Gaussian random variables with zero mean and unit variance. In this paper, we investigate the massive MIMO system in correlated channels using the exponential correlation matrix model [18]. In this model, the components of $\mathbf{R}_T$ and $\mathbf{R}_R$ are determined as follows:

$$r_{sv} = \begin{cases} r^{v-s}, & s \le v \\ r_{vs}^{*}, & s > v \end{cases}, |r| \le 1, \qquad (2)$$

where $r \geq 0$ is the correlation coefficient between any two neighboring transmit or receive antennas. Let $\mathbf{x}_u \in \mathbb{C}^{N_u \times 1}$ represents the transmitted signal vector of the *u*th user. The received signal vector for the *u*th user, (u = 1, 2,…, K), $\mathbf{y}_u \in \mathbb{C}^{N_u \times 1}$ is given by

$$\mathbf{y} = \mathbf{H}_{corr,u} \sum_{k=1}^{K} \mathbf{W}_{corr,u} \mathbf{x}_k + \mathbf{n}_u = \mathbf{H}_{corr,u} \mathbf{W}_{corr,u} \mathbf{x}_u + \sum_{k=1,k\neq u}^{K} \mathbf{H}_{corr,u} \mathbf{W}_{corr,u} \mathbf{x}_k + \mathbf{n}_u \qquad (3)$$

where $\mathbf{H}_{corr,u} \in \mathbb{C}^{N_u \times N_T}$ is channel matrix from the BS to the *u*th user; $\mathbf{W}_{corr,u} \in \mathbb{C}^{N_T \times N_u}$ denotes the precoding matrix for the *u*th user; $\mathbf{n}_u \in \mathbb{C}^{N_u \times 1}$ is noise vector at the *u*th user. Note that, in (3), $\mathbf{H}_{corr,u} \mathbf{W}_{corr,u} \mathbf{x}_u$ is the desired signal component of the *u*th user, $\sum_{k=1,k\neq u}^{K} \mathbf{H}_{corr,u} \mathbf{W}_{corr,k} \mathbf{x}_k$ represents unwanted signals at the *u*th user.

Let $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_1^T & \cdots & \mathbf{y}_K^T \end{bmatrix}^T \in \mathbb{C}^{N_R \times 1}$ be the overall received signal vector for all users. Then, the relationship between the transmitted signal vector, $\mathbf{x} \in \mathbb{C}^{N_R \times 1}$ and the received signal vector $\mathbf{y}$ can be expressed as

$$\mathbf{y} = (\mathbf{H}_{corr} \mathbf{W}_{corr} \mathbf{x} + \mathbf{n}), \qquad (4)$$

where $\mathbf{H}_{corr}$ is channel matrix from BS to all K users, defined in (1); $\mathbf{W}_{corr} \in \mathbb{C}^{N_T \times N_R}$ is the precoding matrix for all users; $\mathbf{n} \in \mathbb{C}^{N_R \times 1}$ is noise vector at the K users, whose entries are assumed to be identical independent distributed (i.i.d) random variables with zero mean and variance σ_n^2.

3. Review of LC-RBD-LR-ZF and element-base lattice reduction (ELR) algorithms

A. LC-RBD-LR-ZF algorithm

The LC-RBD-LR-ZF algorithm is proposed for Multiuser MIMO (MU-MIMO) system using the uncorrelated channel model [16]. This means that the channel matrix from BS to all users is $\mathbf{H}_\omega = \left[(\mathbf{H}_{\omega_1})^T (\mathbf{H}_{\omega_2})^T ... (\mathbf{H}_{\omega_K})^T \right]^T \in \mathbb{C}^{N_R \times N_T}$. The precoding matrix of the LC-RBD-LR-ZF algorithm is expressed as follows:

$$\mathbf{W} = \mathbf{W}^a \mathbf{W}^b, \tag{5}$$

where $\mathbf{W}^a = \left[\mathbf{W}_1^a, \mathbf{W}_2^a, ..., \mathbf{W}_K^a\right] \in \mathbb{C}^{N_T \times KN_T}$; $\mathbf{W}_u^a$ is the precoding matrix for the *u*th user, created by applying QR decomposition to the channel matrix $\bar{\mathbf{H}}_{\omega_u} = \{\rho \mathbf{I}_{\bar{N}_u}, \tilde{\mathbf{H}}_{\omega_u}\}$; the matrix $\tilde{\mathbf{H}}_{\omega_u} = \left[(\mathbf{H}_{\omega_1})^T ... (\mathbf{H}_{\omega_{u-1}})^T (\mathbf{H}_{\omega_{u+1}})^T ... (\mathbf{H}_{\omega_K})^T\right]^T$ is obtained by removing $(\mathbf{H}_{\omega_u})^T$ from $\mathbf{H}_\omega$; $\rho = \sqrt{\frac{N_R \sigma_n^2}{E_s}}$; $\bar{N}_u = N_R - N_u$; and E_s is the energy of each transmitted signal symbol. The QR decomposition of $\bar{\mathbf{H}}_{\omega_u}$ is given by

$$\bar{\mathbf{H}}_{\omega_u} = \mathbf{Q}_u \mathbf{R}_u. \tag{6}$$

Then, the precoding matrix $\mathbf{W}_u^a$ for the *u*th user is obtained as

$$\mathbf{W}_u^a = \mathbf{Q}_u(\bar{N}_u + 1 : \bar{N}_u + N_T, \bar{N}_u + 1 : \bar{N}_u + N_T). \tag{7}$$

After getting $\mathbf{W}_u^a$, the effective channel matrix for the *u*th user is expressed as

$$\hat{\mathbf{H}}_{\omega_u} = \mathbf{H}_{\omega_u} \mathbf{W}_u^a, \tag{8}$$

which is subsequently converted into the LR domain by using the LLL algorithm in [19] as

$$\hat{\mathbf{H}}_{\omega_u}^{LR} = \mathbf{U}_{\omega_u}^T \hat{\mathbf{H}}_{\omega_u}, \tag{9}$$

where $\mathbf{U}_{\omega_u}^T$ is a unimodular matrix with integer elements ($\det|\mathbf{U}_{\omega_u}^T| = 1$); $\hat{\mathbf{H}}_{\omega_u}^{LR}$ is the channel matrix in the LR domain.

The precoding matrix $\mathbf{W}_u^b$ for the *u*th user is created by applying the ZF algorithm on $\hat{\mathbf{H}}_{\omega_u}^{LR}$. Finally, the precoding matrix $\mathbf{W}^b$ for all users is expressed as follows:

$$\mathbf{W}^b = \begin{bmatrix} \mathbf{W}_1^b & 0 & \cdots 0 \\ 0 & \mathbf{W}_2^b & \cdots 0 \\ \vdots & \vdots & \ddots \vdots \\ 0 & 0 & \mathbf{W}_K^b \end{bmatrix} \in \mathbb{C}^{KN_T \times N_R}. \tag{10}$$

It can be seen that the LC-RBD-LR-ZF precoder involves numerous QR decomposition operations. Besides, the size of the matrices $\mathbf{W}^a$ and $\mathbf{W}^b$ increases linearly with the number of users. Therefore, this precoder is suitable for small size MU-MIMO systems. For massive MIMO systems with large number of antennas at the BS to serve dozens of users, the complexity of the LC-RBD-LR-ZF precoder becomes so high that it could hardly be applicable.

B. Element-based Lattice Reduction (ELR) Algorithm

The ELR algorithm was proposed by Qi Zhou and Xiaoli Ma in [20]. The algorithm aims at minimizing the elements on the main diagonal of the error covariance matrix, which is defined as [20]:

$$\tilde{\mathbf{C}} = (\mathbf{H}^H\mathbf{H})^{-1}, \quad (11)$$

where $\mathbf{H} \in \mathbb{C}^{N_A \times N_B}$. As shown in [20], the ELR algorithm gives better performance than the SA and LLL lattice reduction algorithms. Moreover, the computational complexity of the ELR algorithm is significantly reduced compared to those of the SA and LLL ones. Therefore, the ELR algorithm is a suitable candidate for large MIMO systems.

The ELR algorithm has two versions: 1) element-base lattice reduction shortest longest basis (ELR-SLB); and 2) the element-base lattice reduction shortest longest vector (ELR-SLV). Among the two, the ELR-SLB algorithm minimizes all elements on the diagonal of $\tilde{\mathbf{C}}$. The algorithm completes when all the diagonal elements of $\tilde{\mathbf{C}}$ are irreducible. On the contrary, the ELR-SLV algorithm selects the largest element on the diagonal of $\tilde{\mathbf{C}}$ to reduce. The algorithm is finished when the largest element on the diagonal of $\tilde{\mathbf{C}}$ is irreducible. To balance the computational complexity and system performance, in this paper, we adopt the ELR-SLB algorithm as a part of our proposed precoder. For convenience, the ELR-SLB algorithm is summarized in Algorithm 1.

Algorithm 1 The ELR-SLB algorithm

1. **Input** $N_A, N_B, \mathbf{H} \in \mathbb{C}^{N_A \times N_B}$
2. Compute $\tilde{\mathbf{C}} = (\mathbf{H}^H\mathbf{H})^{-1}$ and set $\mathbf{T}' = \mathbf{I}_{N_B}$
3. **Do:**
4. Find the largest element $\tilde{C}_{k,k}$.
5. Compute $\lambda_{i,k} = \left\lfloor \frac{\tilde{C}_{i,k}}{\tilde{C}_{i,i}} \right\rceil, \forall i \neq k$
6. Compute $\Delta_{i,k} = -|\lambda_{i,k}|^2 \tilde{C}_{i,i} - \lambda^*_{i,k}\tilde{C}_{i,k} - \lambda_{i,k}\tilde{C}^*_{i,k}$ and chooses index $i = \arg\max_{i=1,N_R, i\neq k} \Delta_{i,k}$.
7. **If**: $\Delta_{i,k} = 0 \, \forall i,k \in [1, N_A]$ go to step 12
8. $t'_k = t'_k + \lambda_{i,k} t'_i$
9. $\tilde{c}_k = \tilde{c}_k + \lambda_{i,k}\tilde{c}_i$
10. $\tilde{c}^k = \tilde{c}^k + \lambda^*_{i,k}\tilde{c}^i$
11. **While (true)**:
12. **Output:** $\mathbf{T} = (\mathbf{T}'^{-1})^H$, and $\mathbf{H}^{LR} = \mathbf{HT}$

Algorithm 2 The ZF-GP-LR precoding algorithm

1. **Input** $N_T, N_R, \mathbf{H}_{corr}$
2. Decide the number of user groups L and compute the size of the sub-matrices.
3. Generate the matrix $\tilde{\mathbf{H}}^1_{corr}$
4. Apply QR decomposition to $\tilde{\mathbf{H}}^1_{ext}$
5. Generate the matrix $\mathbf{W}^a_{GP_1}$
6. Repeat Step 3 to Step 5 for the next user group until the precoding matrices $\mathbf{W}^a_{GP_l}$ are obtained for all user groups.
7. Generate the matrix $\mathbf{W}^a_{GP}$ as in (14).
8. Generate the matrix $\tilde{\mathbf{H}}_1 = \mathbf{H}^1_{corr}\mathbf{W}^a_{GP_1}$
9. Convert $(\tilde{\mathbf{H}}_1)^T$ into $\tilde{\mathbf{H}}^{LR}_1$ by utilizing **Algorithm 1**.
10. Create the matrix $\mathbf{W}^b_{ZF_1}$
11. Repeat Step 8 to Step 10 for the next user group until the precoding matrices $\mathbf{W}^b_{GP_l}$ are obtained for all user groups.
12. Generate the matrix $\mathbf{W}^b_{GP}$ as in (22).
13. **Output:** β_{GP}, $\mathbf{W}_{corr}$

4. Proposed ZF-GP-LR precoder

A. Proposed ZF-GP-LR precoder

In this section, based on the method in [16], we present a linear group precoding method in combination with the low complexity ELR-SLB technique for massive MIMO systems using the exponential correlation channel model. Block diagram of the proposed ZF-GP-LR precoder is described in Fig. 1.

The overall precoding matrix for all users is defined as follows:

$$\mathbf{W}_{corr} = \beta_{GP}\mathbf{W}_{GP}^{a}\mathbf{W}_{GP}^{b}, \tag{12}$$

where $\mathbf{W}_{GP}^{a} \in \mathbb{C}^{N_T \times (LN_T)}$ is designed to minimize the interferences from other user groups and $\mathbf{W}_{GP}^{b} \in \mathbb{C}^{(LN_T) \times N_R}$ is designed to enhance the system performance.

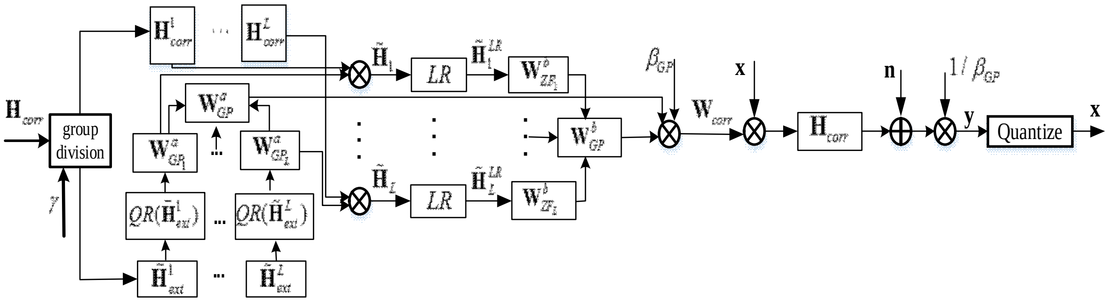

Fig. 1. Block diagram of the proposed ZF-GP-LR precoder

In the first step, the correlation channel matrix, $\mathbf{H}_{corr}$ is divided into $L(L = N_R / \gamma)$ groups (i.e., sub-matrices) $\mathbf{H}_{corr}^{l} \in \mathbb{C}^{\gamma \times N_T} (l = 1, 2, ..., L)$ where γ is an integer greater than one. The first group, $\mathbf{H}_{corr}^{1}$, consists of the first row to the γth row of the channel matrix $\mathbf{H}_{corr}$; the second group, $\mathbf{H}_{corr}^{2}$, is from the (γ + 1)th row to the 2γth row; and the last group, $\mathbf{H}_{corr}^{L}$, is from the $(N_R - \gamma)$th row to the N_Rth row. Specifically, the correlation channel matrix from BS to all users can be represented as follows:

$$\mathbf{H}_{corr} = \begin{bmatrix} \left.\begin{matrix} \mathbf{h}_{11} & \mathbf{h}_{12} & \cdots \mathbf{h}_{1N_T} \\ \vdots & \vdots & \vdots\vdots \\ \mathbf{h}_{\gamma 1} & \mathbf{h}_{\gamma 2} & \cdots \mathbf{h}_{\gamma N_T} \end{matrix}\right\} \mathbf{H}_{corr}^{1} \\ \begin{matrix} \mathbf{h}_{(\gamma+1)1} & \mathbf{h}_{(\gamma+1)2} & \cdots \mathbf{h}_{(\gamma+1)N_T} \\ \vdots & \vdots & \vdots\vdots \end{matrix} \\ \left.\begin{matrix} \mathbf{h}_{(N_R-\gamma)1} & \mathbf{h}_{(N_R-\gamma)2} & \cdots \mathbf{h}_{(N_R-\gamma)N_T} \\ \vdots & \vdots & \vdots\vdots \\ \mathbf{h}_{N_R 1} & \mathbf{h}_{N_R 2} & \cdots \mathbf{h}_{N_R N_T} \end{matrix}\right\} \mathbf{H}_{corr}^{L} \end{bmatrix}. \tag{13}$$

In the second step, the precoding matrix $\mathbf{W}_{GP}^{a}$ is designed to have the following form

$$\mathbf{W}_{GP}^{a} = \left[\mathbf{W}_{GP_1}^{a} \ \mathbf{W}_{GP_2}^{a} \ldots \mathbf{W}_{GP_L}^{a} \right], \tag{14}$$

where $\mathbf{W}_{GP_l}^{a}$ is the precoding matrix for the *l*th group.

To obtain $\mathbf{W}_{GP_l}^{a}$, let us first construct the channel matrix $\tilde{\mathbf{H}}_{corr}^{l} \in \mathbb{C}^{(N_R-\gamma)\times(N_T)}$ consisting of the channel coefficients for all groups except those for the *l*th group as the following

$$\tilde{\mathbf{H}}_{corr}^{l} = \left[(\mathbf{H}_{corr}^{1})^{T} \ldots (\mathbf{H}_{corr}^{l-1})^{T} (\mathbf{H}_{corr}^{l+1})^{T} \ldots (\mathbf{H}_{corr}^{L})^{T} \right]^{T}. \tag{15}$$

After that an extension of $\tilde{\mathbf{H}}_{corr}^{l}$ is constructed as follows:

$$\tilde{\mathbf{H}}_{ext}^{l} = \{ \rho \mathbf{I}_{N_l}, \tilde{\mathbf{H}}_{corr}^{l} \}, \tag{16}$$

where $\tilde{\mathbf{H}}_{ext}^{l} \in \mathbb{C}^{(N_R-\gamma)\times(N_R+N_T-\gamma)}, N_l = N_R - \gamma$ and $\rho = \sqrt{\frac{N_R \sigma_n^2}{E_s}}$.

Applying QRD to $\tilde{\mathbf{H}}_{ext}^{l}$, we get

$$\tilde{\mathbf{H}}_{ext}^{l} = \mathbf{Q}_l \mathbf{R}_l, \tag{17}$$

where $\mathbf{Q}_l \in \mathbb{C}^{(N_l+N_T)\times(N_l+N_T)}$ is an unitary matrix and $\mathbf{R}_l$ is an upper triangular matrix. From $\mathbf{Q}_l$ the precoding matrix $\mathbf{W}_{GPl}^{a}$ for the *l*th group is constructed as

$$\mathbf{W}_{GP_l}^{a} = \mathbf{Q}_l (N_l + 1 : N_l + N_T, N_l + 1 : N_l + N_T). \tag{18}$$

After getting all the weight matrices $\mathbf{W}_{GP_l}^{a}, (l = 1, \cdots, L)$, we define the effective channel matrix for the *l*th group as follows:

$$\tilde{\mathbf{H}}_l = \mathbf{H}_{corr}^{l} \mathbf{W}_{GP_l}^{a}. \tag{19}$$

The channel matrix $(\tilde{\mathbf{H}}_l)$ in (19) is then transposed and converted into the matrix $\tilde{\mathbf{H}}_l^{LR}$ in the LR domain by using the ELR-SLB algorithm to give

$$\tilde{\mathbf{H}}_l^{LR} = \mathbf{U}_l^{T} \tilde{\mathbf{H}}_l, \tag{20}$$

where $\tilde{\mathbf{H}}_l^{LR} \in \mathbb{C}^{\gamma \times N_T}$. The weight matrix $\mathbf{W}_{ZF_l}^{b}$ for the *l*th group is created by applying the ZF procedure to $\tilde{\mathbf{H}}_l^{LR}$ as follows:

$$\mathbf{W}_{ZF_l}^{b} = \left(\tilde{\mathbf{H}}_l^{LR} \right)^{H} \left[\left(\tilde{\mathbf{H}}_l^{LR} \right) \left(\tilde{\mathbf{H}}_l^{LR} \right)^{H} \right]^{-1}. \tag{21}$$

Finally, the precoding matrix $\mathbf{W}_{GP}^{b}$ and unimodular matrix $\mathbf{U}_{GP}^{b}$ for all groups can be obtained as follows:

$$\mathbf{W}_{GP}^{b}=\begin{bmatrix}\mathbf{W}_{ZF_1}^{b} & 0 & \cdots 0\\ 0 & \mathbf{W}_{ZF_2}^{b} & \cdots 0\\ \vdots & \vdots & \ddots \vdots\\ 0 & 0 & \cdots \mathbf{W}_{ZF_L}^{b}\end{bmatrix},\ \mathbf{U}_{GP}^{b}=\begin{bmatrix}\mathbf{U}_1^{T} & 0 & \cdots 0\\ 0 & \mathbf{U}_2^{T} & \cdots 0\\ \vdots & \vdots & \ddots \vdots\\ 0 & 0 & \cdots \mathbf{U}_L^{T}\end{bmatrix}. \quad (22)$$

In order to make sure that the transmit power is unchanged after the transmit signals are precoded, the normalized power factor β_{GP} is computed to be

$$\beta_{GP}=\sqrt{\frac{N_R}{Tr\left[\left(\mathbf{W}_{GP}^{a}\mathbf{W}_{GP}^{b}\right)\left(\mathbf{W}_{GP}^{a}\mathbf{W}_{GP}^{b}\right)^{H}\right]}}. \quad (23)$$

The proposed algorithm ZF-GP-LR is summarized in Algorithm 2. At the user side, the received signal vector for all groups can be expressed as

$$\mathbf{y}=(\mathbf{H}_{corr}\mathbf{W}_{corr}\mathbf{x}+\mathbf{n})/\beta_{GP}. \quad (24)$$

Using $\mathbf{y}$ in (25), the estimated signal vector is given by

$$\tilde{\mathbf{x}}=\mathbf{U}_{GP}^{b}\frac{1}{\alpha}\left(\left\lfloor\alpha\mathbf{y}+\beta_z(\mathbf{U}_{GP}^{b})^{-1}\mathbf{1}_L\right\rceil-\beta_z(\mathbf{U}_{GP}^{b})^{-1}\mathbf{1}_L\right)=\mathbf{x}+2\mathbf{U}_{GP}^{b}Q_z\left[\frac{1}{2}\frac{\mathbf{n}}{\beta_{GP}}\right] \quad (25)$$

where $\alpha=1/2$, $\beta_z=\frac{m-1}{2}(1+j)$, $\mathbf{1}_L\in R^{N_R\times 1}$ is a column vector with N_R ones, $Q_z[a]$ denotes the operation that rounds a to the nearest integer, m is the number of bits in a transmitted symbol.

From (25) it follows that $\mathbf{x}$ is decoded correctly if $Q_z\left[\frac{1}{2}\frac{\mathbf{n}}{\beta_{GP}}\right]=\mathbf{0}$. This means that for a given noise power, the component $1/\beta_{GP}$ will be the factor that determines the system performance. In Fig. 2 and Fig. 3, the empirical cumulative distribution functions (ECDFs) of $1/\beta_{GP}$ are shown for the LC-RBD-LR-ZF and ZF-GP-LR precoders in the case of exponential correlation channel at the BS side (i.e., $\mathbf{H}_{corr}=\mathbf{H}_{\omega}\mathbf{R}_T^{1/2}$). The simulation results show that $1/\beta_{GP}$ increases as the correlation coefficient increases. For the same system configuration, the LC-RBD-LR-ZF precoder generates smaller $1/\beta_{GP}$ than the ZF-GP-LR precoder. In addition, the more sub-groups are generated, the smaller $1/\beta_{GP}$ becomes. This means that the system performance will be degraded as the spatial correlation increases. Besides, the LC-RBD-LR-ZF precoder will probably outperform the ZF-GP-LR precoder in the aforementioned scenarios.

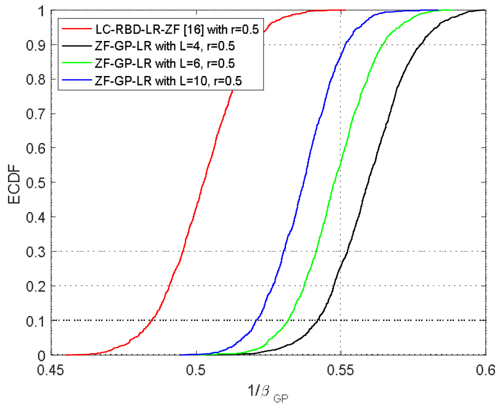

Fig. 2. Empirical CDF of $1/\beta_{GP}$ *for the LC-RBD-LR-ZF and ZF-GP-LR precoders with* $N_T = 60, N_u = 1, K = 60, L = 4, 6$ *and* 10, $r = 0.5$

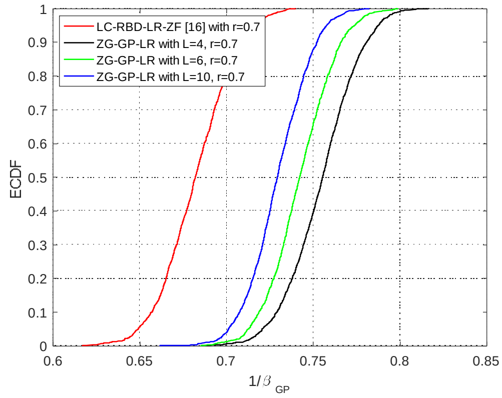

Fig. 3. Empirical CDF of $1/\beta_{GP}$ *for the LC-RBD-LR-ZF and ZF-GP-LR precoders with* $N_T = 60, N_u = 1, K = 60, L = 4, 6$ *and* 10, $r = 0.7$

B. Computational Complexity Analysis

In this sub-section, we evaluate the computational complexity of the proposed algorithm and compare it with that of the LC-RBD-LR-ZF algorithm in [16]. The complexities are evaluated by counting the necessary floating point operations (flops). We assume that each real operation (such as an addition, a multiplication or a division) is counted as a flop. Hence, a complex multiplication and a division require 6 flops and 11 flops, respectively. It is worth noting that the QR decomposition of an $m \times n$ complex matrix requires $6mn^2 + 4mn - n^2 - n$ flops. Based on the above assumptions, the computational complexities of the proposed algorithms ZF-GP-LR is given by:

$$F = F_a + F_b + F_c \quad (flops). \tag{26}$$

where F_a and F_b are the number of flops needed to calculate $\mathbf{W}_{GP}^a$ and $\mathbf{W}_{GP}^b$, respectively; F_c is the total complexity of the multiplication two matrices $\mathbf{W}_{GP}^a$ and $\mathbf{W}_{GP}^a$.

In the proposed algorithm, to find the precoding matrix $\mathbf{W}_{GP_1}^a$ for the first user group, we have to perform the QR decomposition to the correlation channel matrix $\tilde{\mathbf{H}}_{ext}^l \in \mathbb{C}^{(N_R-\gamma)\times(N_R+N_T-\gamma)}$. So the complexity of this work is given by:

$$F_1 = 6(N_R-\gamma)(N_R+N_T-\gamma)^2 + 4(N_R-\gamma)(N_R+N_T-\gamma) - (N_R+N_T-\gamma)^2 - (N_R+N_T-\gamma) \qquad (flops). \tag{27}$$

The QR operation must be carried out *L* times. Hence, the total number of flops to find the precoding matrix $\mathbf{W}_{GP}^a$ is calculated as follows:

$$F_1 = L\times F_1 = L\Big[6(N_R-\gamma)(N_R+N_T-\gamma)^2+4(N_R-\gamma)(N_R+N_T-\gamma)-(N_R+N_T-\gamma)^2 -(N_R+N_T-\gamma) \quad (flops).\Big] \tag{28}$$

The number of flops to calculated $\mathbf{W}_{GP}^b$ is represented as follows:

$$F_b = F_2+F_3+F_4 \quad (flops), \tag{29}$$

where F_2 is the number of flops to find $\tilde{\mathbf{H}}_l$, F_3 is the computational cost for all groups when the ELR-SLB algorithm is adopted to find $\tilde{\mathbf{H}}_l^{LR}$, and F_4 is the total number of flops to find the precoding matrix $\mathbf{W}_{ZF_l}^b$, respectively. Based on the above definitions, F_2 is calculated as follows:

$$F_2 = L(8N_T^2\gamma-2N_T\gamma) \quad (flops). \tag{30}$$

In this paper, we apply the ELR-SLB algorithm to convert the channel matrix $(\tilde{\mathbf{H}}_l)^T$ into the matrix $\tilde{\mathbf{H}}_l^{LR}$. Therefore, F_3 is given by

$$F_3 = F_5+F_6+F_{update-SLB} \quad (flops), \tag{31}$$

where F_5 and F_6 are the number of flops to calculate $\tilde{\mathbf{C}}=\left[\left(\tilde{\mathbf{H}}_l{}^T\right)^H\left(\tilde{\mathbf{H}}_l{}^T\right)\right]^{-1}$ and $\tilde{\mathbf{H}}_l^{LR}=\mathbf{U}_l^T\tilde{\mathbf{H}}_l$, respectively. $F_{update-SLB}$ is computational cost of the ELR-SLB algorithm's update operation, which can only be obtained from the computer simulation. Note that every update operation in the ELR-SLB algorithm requires (16γ + 8) flops. The computations of λ_{ik} and $\Delta_{i,k}$ in Step 5 and Step 6 in Algorithm 1 need 4 flops and 10 flops, respectively. Therefore, $F_{update-SLB}$ is calculated as follows:

$$F_{update-SLB} = CUpdate\times(16\gamma+8)++CLamda\times 4+CDelta\times 10 \quad (flops). \tag{32}$$

where $CLamda$ is the number of updates λ_{ik}, $CDelta$ is the number of updates $\Delta_{i,k}$, $CUpdate$ is the number of updates $t_k^{'}$, $\tilde{c}_k$ and $\tilde{c}^k$ from Step 8 to Step 10 in Algorithm 1. Hence, the total number of flops to convert the channel matrix $(\tilde{\mathbf{H}}_l)^T$ into the matrix $\tilde{\mathbf{H}}_l^{LR}$ is calculated as follows:

$$F_3 = L(8\gamma^3+16\gamma^2N_T-2\gamma^2-2\gamma N_T+F_{update-SLB}) \quad (flops). \tag{33}$$

The number of flops to find the precoding matrix $\mathbf{W}_{ZF_l}^b$ for all group is given by

$$F_4 = L(8\gamma^3+16\gamma^2N_T-2\gamma^2-2\gamma N_T) \quad (flops). \tag{34}$$

Therefore, the total number of flops to find the precoding matrix $\mathbf{W}_{GP}^b$ is calculated as follows:

$$F_b = F_2 + F_3 + F_4 = L(8N_T^2\gamma - 2N_T\gamma) + L(8\gamma^3 + 16\gamma^2 N_T - 2\gamma^2 - 2\gamma N_T + F_{update-SLB}) \\ + L(8\gamma^3 + 16\gamma^2 N_T - 2\gamma^2 - 2\gamma N_T) \qquad (flops). \tag{35}$$

The number of flops for F_c is calculated by

$$F_c = 8LN_T^3 - 2N_T^2 \quad (flops). \tag{36}$$

From the above analysis results, the total number of flops for the ZF-GP-LR algorithm is given by

$$\begin{aligned} F &= F_a + F_b + F_c \\ &= L\Big[6(N_R - \gamma)(N_R + N_T - \gamma)^2 + 4(N_R - \gamma)(N_R + N_T - \gamma) - (N_R + N_T - \gamma)^2 \\ &\quad -(N_R + N_T - \gamma)\Big] + L(8N_T^2\gamma - 2N_T\gamma) + L(8\gamma^3 + 16\gamma^2 N_T - 2\gamma^2 - 2\gamma N_T + F_{update-SLB}) \\ &\quad + L(8\gamma^3 + 16\gamma^2 N_T - 2\gamma^2 - 2\gamma N_T) + 8LN_T^3 - 2N_T^2 \qquad (flops). \end{aligned} \tag{37}$$

The complexities of the precoding algorithms ZF-GP-LR and LC-RBD-LR-ZF are summarized in Tab. 1. From Tab. 1, we can see that the computational complexity of the ZF-GP-LR proposed algorithm is a third-order function of N_T. In contrast, the computational complexity of algorithm LC-RBD-LR-ZF is a fourth-order function of N_T.

5. Simulation results

In this section, we compare both the computational complexity and the system performance of the proposed algorithm with those of the LC-RBD-LR-ZF algorithm in [16].

Figure 4 demonstrates the computational complexities of the ZF-GP-LR and LC-RBD-LR-ZF precoders. In this scenario, N_T is varied from 40 to 100 transmit antennas. It can be seen from the figure that the complexities of the ZF-GP-LR precoder are significantly lower than those of the LC-RBD-LR-ZF. For example, at $N_R = N_T = 60$ antennas, the complexities of the ZF-GP-LR algorithm with L = 2; 4 and L = 10 are approximately equal to 3.04%, 5.52% and 15.21% of the LC-RBD-LR-ZF precoder's complexity, respectively. The computational complexity of the proposed algorithm increases as the number of groups L increases. However, the reduction in complexity is obtained at the cost of performance degradation as illustrated in the figures 5, 6 and 7.

BER performances of the proposed algorithms ZF-GP-LR and the LC-RBD-LR-ZF precoders are illustrated in Fig. 5 to Fig. 7. In Fig. 5, the system is assumed to work in an uncorrelated massive MIMO channel with the following parameters: $N_R = N_T = 60$, and 4-QAM modulation. The channels between the BS and all users are assumed to be semi-static Rayleigh Fading channel, the entries are i.i.d with zero mean and unit variance. The numbers of user groups for the ZF-GP-LR precoder are L = 4; 6

and 10. As can be seen from Fig. 5, in the low and medium SNR regions, the BER curves of the proposed ZF-GP-LR precoder get closer to the LC-RBD-LR-ZF precoder as L increases. Specifically, at BER = 10^{-3}, the proposed algorithm suffers from performance degradations of around 0.6 dB, 0.7 dB and 0.9 dB in SNR respectively for L = 10; 6 and 4 as compared to the LC-RBD-LR-ZF. However, at sufficiently high SNRs, the proposed algorithm provides better system performance than the LC-RBD-LR-ZF algorithm.

In Fig. 6 and Fig. 7, we simulate the system performance in the case exponential correlation channel at the BS side (i.e., $\mathbf{H}_{corr} = \mathbf{H}_{\omega}\mathbf{R}_T^{1/2}$) with the correlation coefficient r = 0.5 and r = 0.7. Other parameters are the same as those used to generate Fig. 5. Similar to the results in Fig. 5, the results in Fig. 6 and Fig. 7 show that, at low SNR, the performance of the proposed ZF-GP-LR precoder approaches that of the LC-RBD-LR-ZF algorithm when L increases. Besides, at high SNR, the proposed algorithm outperforms its LC-RBD-LR-ZF counterpart. From Fig. 6 and Fig. 7, it can also be observed that the spatial correlation has an adverse effect on the system performance no matter which precoder is employed.

Tab. 1. Computational complexity comparison

Precoding algorithms	Complexity (flops)	Computational complexity level
LC-RBD-LR-ZF algorithm [16]	$K\left[6(N_R-N_u)(N_R+N_T-N_u)^2+4(N_R-N_u)(N_R+N_T-N_u) -(N_R+N_T-N_u)^2-(N_R+N_T-N_u)\right]+K(8N_T^2N_u-2N_TN_u) +K(N_u^2N_T+F_{update-LLL})+K(8N_u^3+16N_u^2N_T-2N_u^2-2N_uN_T) +8KN_T^3-2N_T^2$	$O(KN_T^2N_R)$
ZF-GP-LR algorithm	$L\left[6(N_R-\gamma)(N_R+N_T-\gamma)^2+4(N_R-\gamma)(N_R+N_T-\gamma) -(N_R+N_T-\gamma)^2-(N_R+N_T-\gamma)\right]+L(8N_T^2\gamma-2N_T\gamma) +L(8\gamma^3+16\gamma^2N_T-2\gamma^2-2\gamma N_T+F_{update-SLB}) +L(8\gamma^3+16\gamma^2N_T-2\gamma^2-2\gamma N_T)+8LN_T^3-2N_T^2$	$O(LN_T^2N_R)$

It is worth emphasizing that as the number of antennas at the user side is greater than 1, i.e., $N_u > 1$, the correlation channel matrix becomes $\mathbf{H}_{corr} = \mathbf{R}_R^{1/2}\mathbf{H}_{\omega}\mathbf{R}_T^{1/2}$. In such a case, performances of all the precoders under consideration are further degraded. However, the behaviors of the BER curves are still the same as those illustrated in Fig. 5 to Fig. 7. To balance between the computational complexity and system performance, L should be selected by $N_R/2N_u$ when K is an even number. Conversely, K is an odd number, L should be selected by the adjacent divisor to the greatest divisor of K.

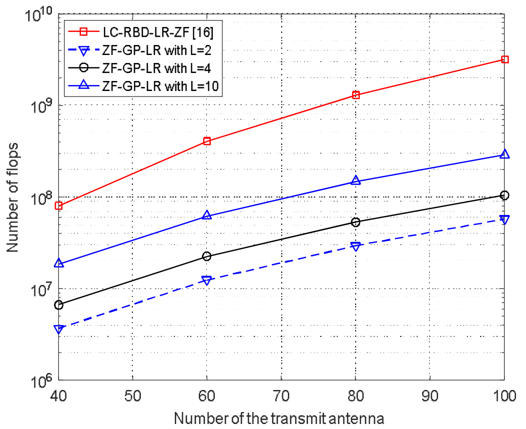

Fig. 4. Compare the complexity of the proposed algorithm and the LC-RBD-LR-ZF algorithm in [16].

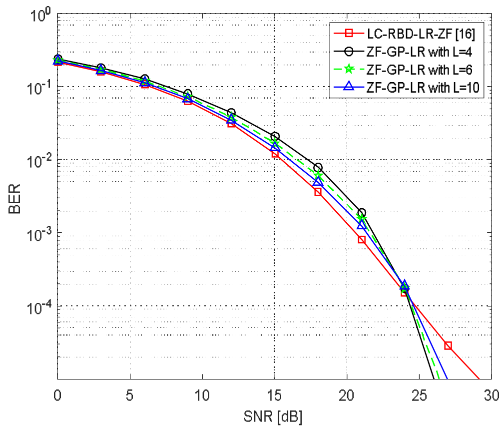

Fig. 5. The system performance with $N_T = 60,\ N_u = 1,\ K = 60,\ L = 4, 6$ *and* 10 *in the case of uncorrelated channel.*

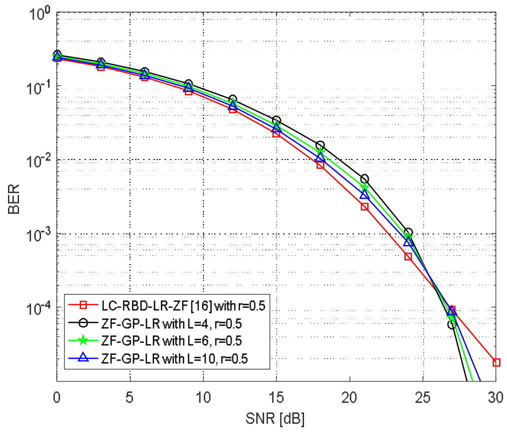

Fig. 6. The system performance with $N_T = 60,$ $N_u = 1,\ K = 60,\ L = 4, 6$ *and* 10 *in the case of correlated channel use the exponential correlation chanel model,* $r = 0.5$.

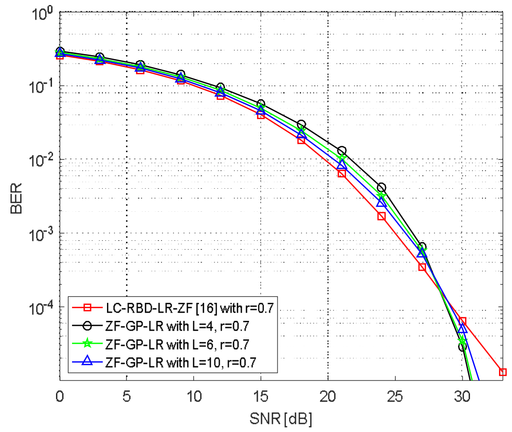

Fig. 7. The system performance with $N_T = 60,$ $N_u = 1,\ K = 60,\ L = 4, 6$ *and* 10 *in the case of correlated channel use the exponential correlation chanel model,* $r = 0.7$.

6. Conclusions

In this paper, we propose the ZF-GP-LR precoder which is a ZF-based group precoding algorithm in combination with the low-complexity ELR-SLB technique to improve the BER performance of massive MIMO systems. Performance and complexity of the proposed precoder are then investigated in massive MIMO systems using the exponential correlation channel model at the BS side. It is shown that the ZF-GP-LR precoder has remarkably lower complexity than its LC-RBDLR-ZF counterpart. In addition, the BER performance of the proposed ZF-GP-LR approaches those of the LC-RBD-LR-ZF algorithm when L increases in the low and medium SNR regions. The proposed precoder even outperforms the LC-RBD-LR-ZF in the high SNR region in both correlated and uncorrelated channels. As a consequence, the proposed ZF-GP-LR precoder can be a potential digital beamforming technique at the base stations of massive MIMO systems.

References

1. H. Q. Ngo (2015). Massive MIMO: Fundamentals and system designs. *Linköping University Electronic Press, 1642*.
2. L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang (Oct 2014). An overview of massive MIMO: Benefits and challenges. *IEEE Journal of Selected Topics in Signal Processing*, *8*(5), pp. 742-758.
3. T. L. Marzetta (November 2010). Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Transactions on Wireless Communications*, *9*(11), pp. 3590-3600.
4. T. L. Marzetta (2015). Massive MIMO: An introduction. *Bell Labs Technical Journal, 20*, pp. 11-22.
5. T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo (2016). *Fundamentals of Massive MIMO*. Cambridge University Press.
6. E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta (February 2014). Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, *52*(2), pp. 186-195.
7. V. P. Selvan, M. S. Iqbal, and H. S. Al-Raweshidy (Aug 2014). Performance analysis of linear precoding schemes for very large multi-user MIMO downlink system. *Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014)*, pp. 219-224.
8. H. Q. Ngo, E. G. Larsson, and T. L. Marzetta (2013). Massive MU-MIMO downlink tdd systems with linear precoding and downlink pilots. *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton),* pp. 293-298.
9. Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang (2010). *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons.
10. Costa (1983). Writing on dirty paper. *IEEE Transactions on Signal Processing*, *29*(3).
11. O. Bai, H. Gao, T. Lv, and C. Yuen (Oct 2014). Low-complexity user scheduling in the downlink massive MU-MIMO system with linear precoding. in *2014 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 380-384.

12. D. H. N. Nguyen, H. Nguyen-Le, and T. Le-Ngoc (February 2014). Block-diagonalization precoding in a multiuser multicell MIMO system: Competition and coordination. *IEEE Transactions on Wireless Communications, 13*(2), pp. 968-981.
13. H. An, M. Mohaisen, and K. Chang (Sept 2009). Lattice reduction aided precoding for multiuser MIMO using Seysen's algorithm. *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 2479-2483.
14. M. Simarro, F. Domene, F. J. Martínez-Zaldívar, and A. Gonzalez (June 2017). Block diagonalization aided precoding algorithm for large MU-MIMO systems. *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 576-581.
15. W. Li and M. Latva-aho (March 2011). An efficient channel block diagonalization method for generalized zero forcing assisted MIMO broadcasting systems. *IEEE Transactions on Wireless Communications*, *10*(3), pp. 739-744.
16. K. Zu and R. C. de Lamare (June 2012). Low-complexity lattice reduction-aided regularized block diagonalization for MU-MIMO systems. *IEEE Communications Letters*, *16*(6), pp. 925-928.
17. R. N. A. Paulraj and D. Gore (2003). *Introduction to space-time wireless communications*. New York: Cambridge University Press.
18. S. L. Loyka (Sep. 2001). Channel capacity of MIMO architecture using the exponential correlation matrix. *IEEE Communications Letters*, *5*(9), pp. 369-371.
19. C. Windpassinger and R. F. H. Fischer (March 2003). Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction. in *Proceedings 2003 IEEE Information Theory Workshop (Cat. No. 03EX674)*, pp. 345-348.
20. Q. Zhou and X. Ma (February 2013). Element-based lattice reduction algorithms for large MIMO detection. *IEEE Journal on Selected Areas in Communications, 31*(2), pp. 274-286.

TIỀN MÃ HÓA TUYẾN TÍNH THEO NHÓM CHO CÁC HỆ THỐNG MASSIVE MIMO DƯỚI ĐIỀU KIỆN TƯƠNG QUAN KHÔNG GIAN HÀM MŨ

Tóm tắt: *Trong bài báo này, thuật toán tiền mã hóa tuyến tính theo nhóm trong mô hình kênh tương quan hàm mũ được đề xuất cho các hệ thống massive MIMO. Bộ tiền mã hóa đề xuất gồm hai thành phần: Thành phần thứ nhất được thiết kế để giảm thiểu can nhiễu từ những nhóm người dùng lân cận; Thành phần thứ hai được thiết kế để cải thiện hiệu suất của hệ thống bằng cách áp dụng kỹ thuật rút gọn giàn ELR-SLB. Kết quả tính toán và mô phỏng cho thấy rằng, bộ tiền mã hóa đề xuất có độ phức tạp tính toán thấp hơn đáng kể so với bộ tiền mã hóa LC-RBD-LR-ZF trong khi tỷ lệ lỗi bít (BER) gần tiệm cận với bộ tiền mã hóa LC-RBD-LR-ZF khi số lượng nhóm tăng lên.*

Từ khóa: Hệ thống MU-MIMO; hệ thống massive MIMO; thuật toán tiền mã hóa tuyến tính; thuật toán tiền mã hóa phi tuyến; thuật toán rút gọn giàn trong hệ thống MIMO.

Received: 28/6/2019; *Revised:* 03/4/2020; *Accepted for publication:* 06/4/2020

❑