Flexible Coordinated Beamforming with Lattice Reduction for Multi-User Massive MIMO Systems

Keke Zu\(^1\), Bin Song\(^1\), Martin Haardt\(^1\), and Rodrigo C. de Lamare\(^2\)

\(^1\) Communications Research Laboratory, Ilmenau University of Technology
PO Box 100565, D-98684 Ilmenau, Germany
\(^2\) Dept. of Electronics / CETUC, University of York / PUC-Rio
York YO10 5DD, U.K./ Marquêde S. Vicente, 225 Gávea, Rio de Janeiro, Brazil
Emails: zukeke@gmail.com, {bin.song, martin.haardt}@tu-ilmenau.de, rodrigo.delamare@york.ac.uk

Abstract—The application of precoding algorithms in multi-user massive multiple-input multiple-output (MU-Massive-MIMO) systems is restricted by the dimensionality constraint that the number of transmit antennas has to be greater than or equal to the total number of receive antennas. In this paper, a lattice reduction (LR)-aided flexible coordinated beamforming (LR-FlexCoBF) algorithm is proposed to overcome the dimensionality constraint in overloaded MU-Massive-MIMO systems. A random user selection scheme is integrated with the proposed LR-FlexCoBF to extend its application to MU-Massive-MIMO systems with arbitrary overloading levels. Simulation results show that significant improvements in terms of bit error rate (BER) and sum-rate performances can be achieved by the proposed LR-FlexCoBF precoding algorithm.

I. INTRODUCTION

The concept of multi-user massive multiple-input multiple-output (MU-Massive-MIMO) systems has been developed recently in [1]-[7] to bring huge improvements in throughput and radiated energy efficiency with inexpensive, low power components. This concept departs from the common understanding of multi-antenna systems and considers a framework in which certain nodes in the network are equipped with antenna arrays featuring a large number of closely spaced radiating elements (several tens to a few hundreds). By exploiting the potential of large spatial dimensions, MU-Massive-MIMO can increase the capacity 10 times or more and simultaneously improve the radiated energy-efficiency [2]. However, the base stations (BSs) are sometimes assumed to have an unlimited number of antennas [4] or the number of BS antennas per user is impractically large [5].

In this work, we focus on the design of downlink precoding algorithms which have dimensionality constraints, also known as overloaded MU-Massive-MIMO systems. As pointed in [8], [9], the application of most precoding algorithms is restricted by the dimensionality constraint that the number of transmit antennas has to be greater than or equal to the total number of receive antennas. Overloaded systems refer to those MIMO system scenarios in which the dimensionality constraint is violated, i.e., the total number of receive antennas exceeds the number of transmit antennas.

To overcome the dimensionality constraint, receive antenna selection and eigenmode selection have been proposed in [10]. In both cases, however, the transmitter and the receiver are not jointly optimized and some signaling techniques are required to indicate the selected receive antennas or eigenmodes. Alternatively, coordinated beamforming (CBF) techniques have been developed to jointly update the transmit-receive beamforming vectors [11]-[15]. However, a study of the convergence behavior in terms of the number of iterations is not considered in [11], and the number of iterations is set manually. The coordinated transmission strategy in [12] only supports a single data stream to each user. To support the transmission of multiple data streams to each user and to reduce the number of iterations, we have developed a flexible coordinated beamforming (FlexCoBF) algorithm in [13], [15].

In order to implement FlexCoBF algorithm, a receive beamforming matrix is employed at each user and is updated with the transmit beamforming matrices iteratively. The receive beamforming matrices may amplify the noise power at the receive side, resulting in a reduced throughput and a degraded bit error rate (BER) performance. To address this performance degradation, a lattice reduction (LR) [16] technique is employed and integrated with the FlexCoBF to further improve the system performances. We term the ratio between the total number of receive antennas and transmit antennas as loading coefficient (LC). In this work, a random user selection scheme is also developed for the proposed LR-FlexCoBF algorithm with arbitrary loading coefficients, which effectively extends the application of FlexCoBF type precoding algorithms to arbitrary overloading scenarios. Moreover, current algorithms for MU-Massive-MIMO systems assume that each user is equipped with a single antenna. However, distributed users equipped with multiple antennas are likely to be common since multiple antennas are already supported in LTE-Advanced [17] and in modern mobile devices [18].

The main contributions of this work are summarized below:

1) We study overloaded MU-Massive-MIMO systems. To the best of our knowledge, this is the first time this scenario is discussed for massive MIMO systems.
2) A LR-aided FlexCoBF (LR-FlexCoBF) algorithm is proposed to overcome the dimensionality constraint problem in overloaded MU-Massive-MIMO systems.
3) A user selection scheme is applied with the FlexCoBF...
algorithm to adjust the LC, resulting in a controlled sum-rate and BER performance.

This paper is organized as follows. The system model and the LR-aided precoding algorithm are described in Section II and Section III, respectively. The proposed LR-FlexCoBF with random user selection algorithm is described in detail in Section IV. Simulation results and conclusions are presented in Section V and Section VI, respectively.

II. SYSTEM MODEL

We consider an uncoded MU-Massive-MIMO broadcast system in a single cell environment as illustrated in Fig. 1, equipped with $M_T$ transmit antennas at the base station (BS), $K$ users in the system each equipped with $M_k$ receive antennas, and the total number of receive antennas is $M_R = \sum_{k=1}^{K} M_k$. The combined transmit data streams are denoted as $s = [s_1^T, s_2^T, \ldots, s_K^T]^T \in \mathbb{C}^{r \times 1}$ with $s_k \in \mathbb{C}^{r \times 1}$, where $r$ is the total number of transmit data streams and $r_k$ is the number of data streams for user $k$. The combined channel matrix is denoted as $H = [H_1^T, H_2^T, \ldots, H_K^T]^T \in \mathbb{C}^{M_R \times M_T}$ and $H_k \in \mathbb{C}^{M_k \times M_T}$ is the $k$th user’s channel matrix. When channel state information (CSI) is available at the transmitter, precoding techniques can be employed to pre-process the transmit data and to reduce the multi-user interference (MUI) [8]. Due to the large number of antennas at the BS, it is very challenging to acquire CSI in frequency division duplex (FDD) systems. For this reason, massive MIMO systems are likely to operate in the time-division duplex (TDD) model, where the reverse channel is used as an estimate of the forward channel [1]-[7]. We assume a flat fading MIMO channel and the received signal $y \in \mathbb{C}^{M_R \times 1}$ is given by

$$y = HF s + n, \quad (1)$$

where $F = [F_1 F_2 \ldots F_K] \in \mathbb{C}^{M_T \times r}$ is the combined precoding matrix and $n = [n_1^T, n_2^T, \ldots, n_K^T]^T \in \mathbb{C}^{M_R \times 1}$ is the combined Gaussian noise with independent and identically distributed (i.i.d.) entries of zero mean and variance $\sigma_n^2$.

III. LR-AIDED PRECODING ALGORITHM

Lattice reduction (LR) can be considered as a mathematical theory to find a basis with short, nearly orthogonal vectors for a given integer lattice basis [16]. Yao and Wornell [19] first applied the LR algorithm in conjunction with MIMO detection techniques. In prior work [20]-[25] LR-aided MIMO precoding algorithms have been devised and investigated. As shown and studied in these works, the symbol error rate curves of precoding and detection algorithms can approach the maximum diversity order at the receive or transmit side.

Linear precoding can be interpreted as the problem of designing the linear precoding matrix $F$ to satisfy an optimization criterion subject to a constraint associated with the transmit power $E\{|\beta Fs|\} \leq P_s$, where $P_s$ is the total transmit power and the factor $\beta$ is chosen to scale the transmitted power to $P_s$ [26]. Based on the optimization criterion, linear precoding algorithms can be categorized into zero forcing (ZF) and minimum mean square error (MMSE) based designs [26]. Linear precoding algorithms are attractive due to their simplicity. However, their transmit diversity order is limited as compared to non-linear precoding algorithms such as dirty paper coding (DPC) [27], Tomlinson-Harashima precoding (THP) [28], [29], and vector perturbation (VP) [30], [31]. Although better performance can be achieved by non-linear precoding algorithms, their computational complexity is relatively high certain scenarios [25].

With the aid of LR techniques, linear precoding techniques can achieve the maximum diversity order while maintaining their simplicity [20]. The most commonly used LR algorithm is the LLL algorithm which was first proposed by Lenstra, Lenstra, and L. Lovász in [16]. By using the LLL algorithm, only the real-valued matrix can be processed which means the channel matrix has to be transformed into equivalent double sized real-valued channel matrix. Thus extra unnecessary complexity could be introduced when the channel has large dimensions. In order to reduce the computational complexity, the complex LLL (CLLL) algorithm was proposed in [32]. The overall complexity of CLLL algorithm is nearly half of the LLL algorithm without sacrificing any performance. Therefore, we employ the CLLL algorithm in this work. A complex lattice is a set of points described by

$$L(H) = \{Hx|x_i \in \mathbb{Z} + j\mathbb{Z}\}, \quad (2)$$

where $H = [h_1, h_2, \ldots, h_M]^T$ contains the bases of the lattice $L(H)$. The aim of the CLLL algorithm is to find a new basis $\tilde{H}$ which is shorter and nearly orthogonal compared to the original matrix $H$. Let us calculate the QR decomposition, $H = QR$, where $Q$ is an orthogonal matrix and the upper-triangular matrix $R$ is a rotated and reflected representation of $H$. Thus, each column vector $h_m$ of $H$ is given by [33]

$$h_m = \sum_{l=1}^{m} r_{l,m} q_l, \quad (3)$$

where $q_l$ is the $l$th column of $Q$. If $|r_{1,m}|, \ldots, |r_{m-1,m}|$ are close to zero, we can say that $h_m$ is nearly orthogonal.
the space spanned by \( h_1, \ldots, h_{m-1} \). Similarly, the QR decomposition of \( \tilde{H} \) is \( \tilde{H} = QR \). Then, the basis for \( L(H) \) is CLLL reduced if both of the following conditions are satisfied

\[
|\tilde{r}_{l,m}| \leq \frac{1}{2} |\tilde{r}_{l,l}|, \quad 1 \leq l < m \leq M_T, \quad (4)
\]

\[
\delta |\tilde{r}_{m-1,m-1}|^2 \leq |\tilde{r}_{m,m}|^2 + |\tilde{r}_{m-1,m}|^2, \quad 2 \leq m \leq M_T, \quad (5)
\]

where \( \delta \in (\frac{1}{2}, 1] \) influences the quality of the reduced basis and the computational complexity. We usually set \( \delta = \frac{3}{4} \) to achieve a trade-off between performance and complexity [16].

We perform the LR transformation on the transpose of channel matrix \( H^T \) [20] as described by

\[
\tilde{H} = TH \quad \text{and} \quad H = T^{-1}\tilde{H}, \quad (6)
\]

where \( T \) is a unimodular matrix with \( \det(T) = 1 \) and all elements of \( T \) are complex integers, i.e., \( t_{l,m} \in \mathbb{Z} + j\mathbb{Z} \).

The physical meaning of the constraint \( \det(T) = 1 \) is that the channel energy is still the same after the LR transformation. Following the LR transformation, we employ the linear precoding constraint to get the precoding filter at the transmit side and to process the data streams. The ZF precoding is implemented as

\[
\tilde{F}_{ZF} = \tilde{H}^H(\tilde{H}\tilde{H}^H)^{-1} = H^H(HHH^{-1})^{-1}T^{-1}. \quad (7)
\]

IV. PROPOSED LR-AIDED FLEXCOBF ALGORITHM WITH RANDOM USER SELECTION

For linear or LR-aided precoding techniques to work, it is required that \( M_R \leq M_T \) [8]. When \( M_R > M_T \), however, precoding techniques cannot perform well simply because the requirement of \( M_R \) data streams is beyond the transmission ability of the systems. For the \( M_R > M_T \) case, we assume that the number of actually transmitted data streams is \( r \) and it should satisfy \( r \leq M_T \), which means that the maximum number of transmitted data streams cannot exceed the number of transmit antennas \( M_T \).

In [15], we have developed an iterative coordinated method named FlexCoBF to overcome the dimensionality constraint problem. The receive beamforming matrix \( W_k \in \mathbb{C}^{r \times M_k} \) is introduced at each user and initialized with random matrices.

Then, iterative computations are employed to update the receive beamforming matrix \( W_k \) and the \( k \)th user’s precoding matrix \( F_k \) jointly to enforce the zero MUI constraint. However, a random user selection scheme is developed in this work to extend the application of FlexCoBF for MU-Massive-MIMO systems with various loading coefficients.

Considering the case when \( M_T > K \), the system can afford one data stream for each user plus extra data streams for \( M_T - K \) users equipped with multiple receive antennas. Assume \( j = M_T - K \), a random user selection scheme is first implemented to select \( j \) users from a total set of \( K \) users.

Then, the remaining \( l = K - j \) users are allocated one data stream by implementing the iterative computations described above. We define the quantity \( H_T = [H_{L_1}^T, H_{L_2}^T, \ldots, H_{L_{K}}^T]^T \) is the combined channel matrix of the selected users, and \( H_T = [H_{L_1}^T, H_{L_2}^T, \ldots, H_{L_{K}}^T]^T \) is the combined channel matrix of the remaining users. Finally, the equivalent channel matrix \( H_e \in \mathbb{C}^{r \times M_T} \) is obtained as

\[
H_e = \begin{bmatrix}
I_1 & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & I_j & 0 & \ldots & 0 \\
0 & 0 & 0 & W_l^H & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \ldots & W_l^H \\
\end{bmatrix}
\begin{bmatrix}
H_{L_1} \\
\vdots \\
H_{L_j} \\
H_{L_{K}} \\
\end{bmatrix}. \quad (8)
\]

Since the standard lattice reduction algorithm is also restricted by the dimensionality constraint [32], we apply the LR transformation on the equivalent channel matrix \( H_e \) rather than the channel matrix \( H \) when \( M_R > M_T \), i.e.,

\[
\tilde{H}_e = T_e H_e, \quad (9)
\]

where the quantity \( \tilde{H}_e \) is the LR transformed matrix and the transformation matrix \( T_e \in \mathbb{C}^{r \times r} \) is unimodular (\( |\det(T_e)| = 1 \)) and all elements of \( T_e \) are complex integers. The LR transformed matrix \( \tilde{H}_e \) is a basis of \( H_e \) in the lattice space. Compared to the original equivalent channel matrix \( H_e \), the LR transformed channel matrix \( \tilde{H}_e \) is closer to orthogonality which can be measured by the orthogonality defect factor defined in [35]. Therefore, improved BER and sum-rate performances can be achieved by the proposed LR-FlexCoBF algorithm.

Assuming that the variable \( p \) represents the iteration index, the proposed LR-FlexCoBF algorithm is performed in the following seven steps:

1) Implement the random user selection scheme to select \( J \) users from \( K \), and get \( H_j \) and \( H_{L} \), respectively.

2) Initialize the iteration index \( p \) to zero and the combined receive beamforming matrix \( W^{(0)} = \text{diag}\{W_1^{(0)H}, W_2^{(0)H}, \ldots, W_l^{(0)H}\} \) to random matrices. Set the constant \( \epsilon \) as the threshold to iteratively enforce the zero MUI constraint for each receiver.

3) Set \( p = p+1 \) and compute the equivalent channel matrix \( H_{L_e}^{(p)} \) as

\[
H_{L_e}^{(p)} = \begin{bmatrix}
W_1^{(p-1)H} H_{L_1} \\
W_2^{(p-1)H} H_{L_2} \\
\vdots \\
W_l^{(p-1)H} H_{L_l} \\
\end{bmatrix}.
\]

4) Apply the ZF constraint based precoding design to the obtained equivalent channel matrix \( H_{L_e}^{(p)} \) to obtain the transmit beamforming matrices \( F_k^{(p)} \) for all single data stream users \( k = 1, \ldots, l \).

5) For the \( p \)th iteration, update \( W^{(p)} \) jointly with the newly obtained precoding matrix \( F^{(p)}_e \) as

\[
W^{(p)} = HF^{(p)}_e.
\]
where the quantity $H$ is the combined channel matrix defined in Section II.

6) Track the alterations of the residual MUI after the linear precoding as

$$\text{MUI}(H^{(p+1)}_{L_e} F^{(p)}_{L_e}) = \| \text{off}(H^{(p+1)}_{L_e} F^{(p)}_{L_e}) \|_2^2,$$

where the operation $\text{off}(B)$ denotes the selection of the off-diagonal elements of the matrix $B$. If the residual MUI is above the threshold $\epsilon$, go back to step 2. Otherwise, convergence is achieved and the iterative procedure stops. The final equivalent channel matrix $H_c$ and the receive beaforming matrix $W^{(p)}_e$ is respectively obtained as

$$H_c = [H^T_f, H^{(p)}_{L_e}]^T \text{ and } W^{(p)}_e = \text{diag}(I_p, W^{(p)}).$$

7) Perform the LR transformation on the obtained equivalent channel matrix $H_c$ to get $\tilde{H}_c$ and the transformation matrix $T_c$. Apply the transformation matrix $T_c$ to the obtained combined precoding matrix $F^{(p)}_c = [F^{(p)}_1, F^{(p)}_2, \ldots, F^{(p)}_K]$ , that is, $\tilde{F}^{(p)}_c = F^{(p)}_c T^{-1}_c$.

Finally, the received signal for the $M_R > M_T$ case can be expressed as

$$y = W^{(p)}_e H \tilde{F}^{(p)}_c s + n = W^{(p)}_e H F^{(p)}_c z + n,$$

where the quantity $z = T_c^{-1}s$. In order to get the estimation of $z$, a proper shifting and scaling work is needed to transform the signal back from the Lattice space [34]. Then, the estimated transmit signal $s$ is obtained as $\hat{s} = T\hat{z}$ with $\hat{z}$ is the estimation of $z$.

V. SIMULATION RESULTS

We consider an overloaded MU-Massive-MIMO system in this section. The quantity $E_b/N_0$ is defined as $E_b/N_0 = \frac{M_R P_s}{M_T N_0}$ with $n$ being the number of information bits transmitted per channel symbol, and the loading coefficient ($L_C$) is defined as $L_C = \frac{M_R}{M_T}$. An uncoded QPSK modulation scheme is employed in the simulations. The threshold $\epsilon$ is set to $10^{-5}$, and the maximum number of iterations is restricted to 20. The channel matrix $H$ is assumed to be a complex i.i.d. Gaussian matrix with zero mean and unit variance.

The case of $K = 20$ users each equipped with $M_k = 2$ receive antennas is first considered in Fig. 2 and Fig. 3. When the loading coefficient $L_C = 1.0$, FlexCoBF corresponds to conventional ZF precoding. It is shown in Fig. 2 that the BER performance becomes better with the increase of $L_C$. This is because that the inter-stream interference is reduced with the decrease of the equivalent channel matrix dimension. The proposed LR-FlexCoBF has a better performance as compared to the conventional FlexCoBF especially for the heavily overloaded case at high SNRs. The sum-rate performance, which is illustrated in Fig. 3, is inversely proportional to $L_C$ because more data streams are supported. For example, 40 data streams are supported with $L_C = 1.0$ while there are only 25 with $L_C = 1.6$. It worth noting that a much better sum-rate performance is achieved by the proposed LR-FlexCoBF algorithms.

![Fig. 2. BER performance with QPSK, $K = 20, M_k = 2$](image)

![Fig. 3. Sum-rate performance, $K = 20, M_k = 2$](image)

![Fig. 4. BER performance with QPSK, $K = 40, M_k = 2$](image)
VI. CONCLUSION

A LR-FlexCoBF algorithm has been proposed to support data transmission in overloaded MU-Massive-MIMO system. By employing a random user selection scheme in conjunction with the proposed LR-FlexCoBF algorithm, MU-Massive-MIMO systems with variety loading coefficients are studied. The proposed LR-FlexCoBF precoding algorithm can achieve a higher diversity order and a higher spatial multiplexing gain as compared to the standard FlexCoBF and other existing techniques.

REFERENCES