Received May 29, 2019, accepted June 8, 2019, date of publication June 24, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924673

CSI Feedback Based on Deep Learning for Massive MIMO Systems

YONG LIAO[1], (Member, IEEE), HAIMEI YAO[1], YUANXIAO HUA[1], AND CHUNGUO LI[2], (Senior Member, IEEE)

[1]Center of Communication and TT&C, Chongqing University, Chongqing 400044, China
[2]School of Information Science and Engineering, Southeast University, Nanjing 210096, China

Corresponding authors: Yong Liao (liaoy@cqu.edu.cn) and Chunguo Li (chunguoli@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501066, in part by the Chongqing Frontier and Applied Basic Research Project under Grant cstc2015jcyjA40003, and in part by the Fundamental Research Funds for the Central Universities under Grant 106112017CDJXY500001.

ABSTRACT Aiming at the problem of high complexity and low feedback accuracy of existing channel state information (CSI) feedback algorithms for frequency-division duplexing (FDD) massive multiple-input multiple-output (MIMO) systems, this paper proposes a CSI compression feedback algorithm based on deep learning (DL), which is suitable for single-user and multi-user scenarios in massive MIMO systems. This algorithm considers the spatial correlation of massive MIMO channel and uses bidirectional long short-term memory (Bi-LSTM) and bidirectional convolutional long short-term memory (Bi-ConvLSTM) network to decompress and recover the CSI for single-user and multi-user, respectively. The proposed DL-based CSI feedback network is trained offline by massive MIMO channel data and could learn the structural characteristics of the massive MIMO channel by fully exploiting the channel information in the training samples. The simulation results show that compared with several classical CSI compression feedback algorithms, the proposed CSI compression feedback algorithm has lower computational complexity, higher feedback accuracy, and better system performance in massive MIMO systems.

INDEX TERMS FDD, massive MIMO, channel state information, compressed feedback, deep learning.

I. INTRODUCTION

As a key technology of the fifth generation (5G) communication system, massive multiple-input multiple-output (MIMO) technology has many advantages, such as high spectrum efficiency [1], large system capacity, strong system robustness [2]–[4]. So, massive MIMO has attracted more and more attention from industry and academia. However, the significant advantages of massive MIMO technology can be largely determined by the fact that the transmitter can obtain the channel state information (CSI) of the downlink. In the frequency-division duplexing (FDD) massive MIMO systems, the base station (BS) needs to obtain the downlink CSI through the feedback of the receiver [5], [6]. However, the use of massive antenna arrays causes a sharp increase in channel feedback overhead. Thus, how to reduce the feedback overhead of CSI in practical applications has become an urgent problem to be solved [7].

In massive MIMO systems, the channel has a strong spatial correlation because of the use of the massive antenna arrays and the dense deployment of antennas. Considering this channel characteristics, compression sensing (CS) or dimensionality reduction technology have been proposed to solve the CSI compressed feedback problems recently. In [8], the CS is introduced into the finite feedback of massive MIMO for the first time. Based on the spatial correlation characteristics of massive MIMO channels, two adaptive compression algorithms, Karhunen-Loeve transform (KLT) and discrete cosine transform (DCT) are proposed to improve the feedback efficiency. In [9], the DCT and KLT are also used as sparse basis. Based on the strong spatial correlation of the channel, when the BS adopts a uniform square matrix, the channel is compressed after independent principal component transform, which reduces the feedback overhead and codebook search complexity. However, the algorithms proposed in [8] and [9] need to use the correlation matrix of the downlink channel to generate compression matrix and recovery matrix, and users need to periodically feed back the

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

information to BS according to the change of the channel. When the channel changes quickly, it is difficult for BS to obtain the channel correlation information accurately and timely, which causes a decrease in feedback accuracy. And also, when the number of antennas increases, the dimensions and the number of the elements of the channel correlation matrix will increase greatly, resulting in an increase in feedback overhead. Considering the system performance and complexity during compression, principal component analysis (PCA) is an effective compromise solution. The authors in [10]–[12] proposed a PCA-based dimensionality reduction compression feedback algorithm for multi-user scenarios. The proposed algorithm can reduce the feedback overhead and achieve a tradeoff between system capacity and computational complexity, but the compression matrix is estimated by the previous long-term phase, so it is difficult for BS to obtain accurate CSI.

Deep learning (DL) has attracted great attention in large data processing recently, it has been successfully applied in physical layer of wireless communication systems [13]–[16]. In [14], the authors proposed a DL-based method, which combines two convolutional neural networks (CNNs) trained on different datasets, to achieve higher accuracy of automatic modulation recognition (AMR). A fast beamforming design method using unsupervised learning is proposed in [15], this method trains the deep neural network (DNN) offline and provides real-time service online only with simple neural network operations, reducing computational complexity significantly. The authors in [16] proposed a novel and effective DL-aided non-orthogonal multiple access (NOMA) system, a long short-term memory (LSTM) network based on DL is exploited to learn a completely unknown channel environment, the proposed scheme is more robust and efficient compared with conventional approaches. In [17], a DL-based joint channel equalization and decoding algorithm is proposed. This algorithm uses CNN and DNN to achieve channel equalization and decoding in non-linear channel respectively. Compared with some existing equalization and decoding algorithms, this algorithm can obtain better results. In [18], a DL-based channel estimation and direction-of-arrival (DOA) estimation algorithm for massive MIMO systems is proposed. This algorithm uses DNN to effectively learn the statistical characteristics of wireless channels and the spatial structure in the angle domain. Compared with conventional algorithms, this algorithm can obtain better performance of DOA estimation and channel estimation. Additionally, a DL-based hybrid precoding algorithm in millimeter-wave massive MIMO systems is proposed in [19], in which each selection of the precoders for obtaining the optimized decoder is regarded as a mapping relation in DNN. Specifically, the hybrid precoder is selected through training based on DNN for optimizing precoding process of the millimeter-wave massive MIMO. Compared with conventional precoding, the DNN-based hybrid precoding can minimize bit error ratio (BER) and enhance spectral efficiency, which achieves better performance in hybrid precoding while substantially reducing the computational complexity. In [20], the authors proposed a DL-based CSI feedback algorithm for massive MIMO system. The CSI feedback is realized by using CsiNet network which is composed of fully connected network and residual network. Compared with conventional CS algorithms, CsiNet network has higher recovery accuracy and better performance. However, the network has many training parameters, and only convolutional layers and fully connected layers are used to extract the features of the data to complete CSI compression and recovery, the spatial correlation between antennas is not fully utilized in massive MIMO system.

Therefore, this paper studies the CSI acquisition problems for FDD massive MIMO systems. Aiming at the problems of high computational complexity, low feedback accuracy in conventional algorithms and a lack of consideration of spatial correlation between antennas in CsiNet network, this paper proposes a DL-based CSI compression feedback algorithm with low feedback overhead and high feedback accuracy for FDD massive MIMO systems, which considers the spatial correlation of massive MIMO channel data. The main contributions are as follows:

1) Aiming at the shortcomings of existing CSI feedback algorithms in massive MIMO systems, this paper proposes a DL-based CSI feedback algorithm for FDD massive MIMO systems. The network structure used in this algorithm is suitable for single-user and multi-user scenarios, and signal processing includes compression process and decompression process.
2) In compression process, the proposed algorithm uses two two-dimensional (2D) CNN and two three-dimensional (3D) CNN networks to learn the non-linear structural characteristics of the channel and extract the channel feature vectors accurately for single-user and multi-user respectively. Then, the outputs of 2D CNN and 3D CNN can be reduced by using 2D Maxpooling network and 3D Maxpooling network respectively, which can effectively compress the data.
3) In decompression process, firstly, a fully connected layer is used to increase the compressed CSI data dimension to the dimension before compression, and then making full use of the strong ability of bidirectional long short-term memory (Bi-LSTM) network and bidirectional convolutional long short-term memory (Bi-ConvLSTM) for processing sequence data, the Bi-LSTM and Bi-ConvLSTM network are used to recover the original CSI for single-user and multi-user respectively by using the structure characteristics of forward and backward antenna data, so as to improve the reconstruction quality and feedback accuracy.
4) The proposed network can be seen as a black box to realize end-to-end CSI feedback design and it is trained offline by massive MIMO channel data. The complexity of the proposed algorithm and several classical CSI feedback algorithms are analyzed. In addition, their

feedback performance under different massive antenna numbers is also compared.

The rest of this paper is organized as follows. In Section II, the system model of massive MIMO is described. In Section III, the core ideas of and the specific steps of the proposed algorithm are presented. In Section IV, the complexity of various algorithms is analyzed. Simulation results are provided in Section V. Finally, the conclusion is discussed in Section VI.

II. SYSTEM MODEL

This paper consider the multi-user massive MIMO wireless communication system in FDD mode with $N_t >> 1$ transmitting antennas and $N_r \geq 1$ receiving antennas and K mobile users in the cell, the channel is Rayleigh flat fading channel. The system model is showed in Fig. 1.

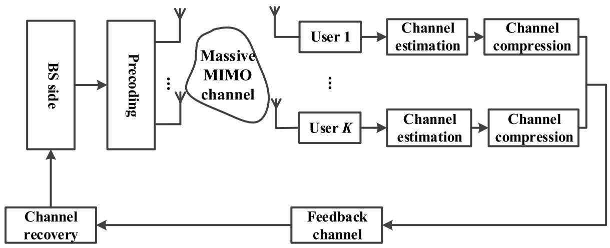


FIGURE 1. Diagram of channel feedback of multi-user MIMO system.

In the FDD system, the downlink direction is first considered. The BS side pre-codes the data stream and then transmits the signal through the wireless channel. After receiving the signal, the user side first obtains the CSI through the channel estimation algorithm, then processes the CSI, and finally feeds back the CSI to the BS through the uplink channel. The BS side recovers the CSI and performs precoding and the like by using the obtained CSI. Assuming that there are K mobile users in the cell, the channel matrix from BS to kth user is $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$. $\mathbf{U}$ is the precoding matrix of $N_t \times KN_r$, s is the transmitting signal vector of $KN_r \times 1$. The receiving signal of the kth user is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{U}\mathbf{s} + \mathbf{n}_k \tag{1}$$

where $\mathbf{n}_k$ is the additive Gaussian white noise of the kth user for which the mean is 0 and the variance is 1. The signals received by all users are

$$\mathbf{y} = \mathbf{H}\mathbf{U}\mathbf{s} + \mathbf{n} \tag{2}$$

where $\mathbf{y} = \left(\mathbf{y}_1^T, \cdots, \mathbf{y}_K^T\right)^T$, $\mathbf{H} = \left(\mathbf{H}_1^H, \cdots, \mathbf{H}_K^H\right)^H$, $\mathbf{n} = \left(\mathbf{n}_1^T, \cdots, \mathbf{n}_K^T\right)^T$, and $(\cdot)^T$ is the transposition of a matrix. Assuming that $E\left[\mathbf{s}\mathbf{s}^H\right] = 1$, $E\left[\mathbf{n}\mathbf{n}^H\right] = 1$ and the transmitting power of the BS is limited to P, so $tr\left(\mathbf{U}\mathbf{U}^H\right) \leq P$, where $tr(\cdot)$ is the trace of the matrix. In the massive MIMO system, there is a certain correlation between antennas because of the large number of antennas at transmitter and receiver, the dense arrangement of antennas makes the channel correlated highly. The channel matrix $\mathbf{H}$ with spatial correlation can be modeled as [8]

$$\mathbf{H} = \frac{1}{\sqrt{tr\left(\mathbf{R}_r\right)}} \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}_{iid} \mathbf{R}_t^{\frac{1}{2}} \tag{3}$$

where $\mathbf{R}_r \in \mathbb{R}^{N_r \times N_r}$ is the receiving correlation matrix, $\mathbf{R}_t \in \mathbb{R}^{N_t \times N_t}$ is the transmitting correlation matrix, and $\mathbf{H}_{iid} \in \mathbb{C}^{N_r \times N_t}$ is the independent and identically distributed complex Gaussian random vector with the mean value of 0 and the variance of 1. Assuming that a uniform linear array antenna are deployed at both the transmitter and receiver of the wireless link, the spatial correlation matrix can be obtained by the Clarke [21] model. All elements in $\mathbf{R}_t$ or $\mathbf{R}_r$ are r_{ij}, where r_{ij} is the correlation coefficient between the ith antenna and the jth antenna of the transmitter or the receiver, it can be expressed as

$$r_{ij} = J_0\left(\frac{2\pi d_{ij}}{\lambda}\right) = J_0\left(\frac{2\pi d}{\lambda}|i-j|\right) \tag{4}$$

where $J_0(\cdot)$ is the zero-order Bessel function of first kind, i and j denote the array element number of the antenna, d_{ij} is the distance between the ith transmitting antennas and the jth receiving antennas, d represents the distance between adjacent transmitting or receiving antennas, and λ is the carrier wavelength.

III. DL-BASED MASSIVE MIMO CSI FEEDBACK ALGORITHM

A. PROPOSED DL-BASED NETWORK FRAMEWORK

Aiming at the problems of large CSI feedback overhead in massive MIMO systems, this paper proposes a DL-based CSI feedback network which is suitable for single-user and multi-user scenarios. The proposed CSI feedback network is divide into two parts: compression process and decompression process, and its structure is shown in Fig. 2.

In the compression process, this network extracts the feature vectors through two 2D CNN or 3D CNN, and then compresses the parameters to be estimated through the 2D Maxpooling or 3D Maxpooling layer. In the decompression process, the dimension of the compressed data is first transformed to the same dimension before compressing by the fully connected network, and then the transformed data is used to predict the CSI through Bi-LSTM or Bi-ConvLSTM network. Finally, the output dimension of Bi-LSTM or Bi-ConvLSTM network is reduced through the fully connected network, and the fully connected network outputs the final recovered CSI. In particular, for single-user case, the 2D CNN, 2D Maxpooling and Bi-LSTM network are used, for multi-user case, the 3D CNN, 3D Maxpooling, and Bi-ConvLSTM network are used.

The original CSI is fed back through the feedback network, and the data flow is shown in Fig. 3. As shown in Fig. 3, the compression processing and decompression processing are described in detail below, respectively.

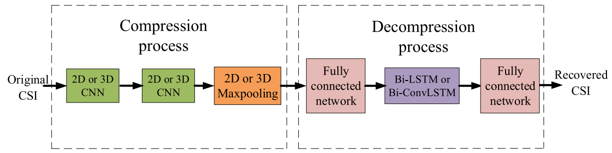


FIGURE 2. The structure of proposed CSI feedback network.

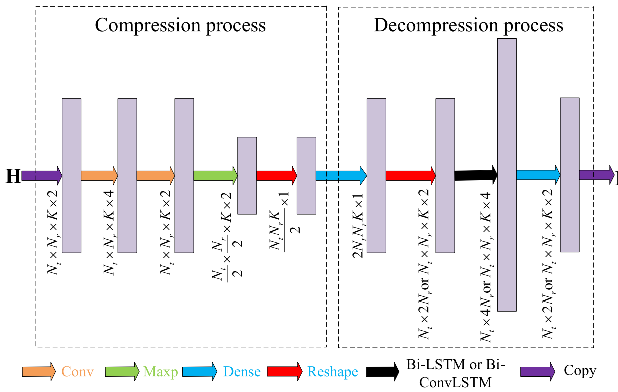


FIGURE 3. Data flow in the proposed massive MIMO CSI feedback network.

1) COMPRESSION PROCESS

In this paper, the input data of the learning network is a channel matrix of $N_t \times N_r \times K$ size, that is, $\mathbf{H} \in \mathbb{C}^{N_t \times N_r \times K}$, for single-user case, K is set as 1. Since the channel data is a complex signal, the input data needs to be pre-processed before it is input into the proposed learning network. The real part and the imaginary part of the input data are extracted separately and then used as the number of channels, the input data becomes $\mathbf{H}'_s \in \mathbb{R}^{N_t \times N_r \times 2}$ for single-user case and $\mathbf{H}'_m \in \mathbb{R}^{N_t \times N_r \times K \times 2}$ for multi-users case.

The input data is pre-processed and sent to two 2D CNN or 3D CNN network. The main task of the 2D CNN or 3D CNN network is to extract the data feature vectors. Unlike conventional machine learning algorithms that perform feature extraction manually, CNN can automatically extract feature vectors from the data to obtain representative vectors from the input data. In this paper, the two 2D CNN network is mainly used to extract feature vectors of the original channel $\mathbf{H}$.

For single-user case, the number of convolution kernel of the first and second 2D CNN are 4 and 2, respectively, and all the convolution kernel size are 3×3. The output of first two 2D CNN networks is

$$\mathbf{H}''_s = f\left(f\left(\mathbf{H}'_s * \mathbf{W}_{1,s} + \mathbf{b}_{1,s}\right) * \mathbf{W}_{2,s} + \mathbf{b}_{2,s}\right) \tag{5}$$

where $\mathbf{H}''_s \in \mathbb{R}^{N_t \times N_r \times 2}$, $\mathbf{W}_{1,s}$ and $\mathbf{W}_{2,s}$ are the weights of the first and second 2D CNN, respectively, $\mathbf{b}_{1,s}$ and $\mathbf{b}_{2,s}$ are the baises of the first and second 2D CNN, respectively. f is the ReLU activation function and $*$ denotes the convolution operation.

For multi-user case, the number of convolution kernel of the first and second 3D CNN also are 4 and 2, respectively, and the convolution kernel size both are $3 \times 3 \times 3$. The output of first two 3D CNN networks is

$$\mathbf{H}''_m = f\left(f\left(\mathbf{H}'_m * \mathbf{W}_{1,m} + \mathbf{b}_{1,m}\right) * \mathbf{W}_{2,m} + \mathbf{b}_{2,m}\right) \tag{6}$$

where $\mathbf{H}''_m \in \mathbb{R}^{N_t \times N_r \times K \times 2}$, $\mathbf{W}_{1,m}$ and $\mathbf{W}_{2,m}$ are the weights of the first and second 3D CNN, respectively. $\mathbf{b}_{1,m}$ and $\mathbf{b}_{2,m}$ are the baises of the first and second 3D CNN, respectively. Like the single-user, f is also the ReLU activation function and $*$ denotes the convolution operation. The expression of ReLU activation function is

$$ReLU\,(x) = \max\,(0, x) \tag{7}$$

The Maxpooling network is mainly used to reduce the dimension of the feature vector after convolution, and a pooling window is used to find the maximum value of the filter output. The output of the second 2D CNN and the second 3D CNN network will be compressed through the 2D Maxpooling and 3D Maxpooling networks respectively, and the data will be compressed to 1/4 of the original.

The pooling window sizes of 2D and 3D Maxpooling networks are 2×2 and $2 \times 2 \times 1$, respectively, so the output of Maxpoling are $\widehat{\mathbf{H}}_s \in \mathbb{R}^{\frac{N_t}{2} \times \frac{N_r}{2} \times 2}$ and $\widehat{\mathbf{H}}_m \in \mathbb{R}^{\frac{N_t}{2} \times \frac{N_r}{2} \times K \times 2}$, respectively. Then they are reshaped into a one-dimensional vector $\widehat{\mathbf{H}}'_s \in \mathbb{R}^{\frac{N_t N_r}{2}}$ and $\widehat{\mathbf{H}}'_m \in \mathbb{R}^{\frac{N_t N_r K}{2}}$.

2) DECOMPRESSION PROCESS

In the decompression part, the fully connected network is first used to increase the dimension of $\widehat{\mathbf{H}}'_s \in \mathbb{R}^{\frac{N_t N_r}{2}}$ and $\widehat{\mathbf{H}}'_m \in \mathbb{R}^{\frac{N_t N_r K}{2}}$, and the dimension is increased by 4 times to get $\widehat{\mathbf{H}}''_s \in \mathbb{R}^{2N_t N_r}$ and $\widehat{\mathbf{H}}''_m \in \mathbb{R}^{2N_t N_r K}$, respectively. They are then reshaped into the form which is suitable for Bi-LSTM and Bi-ConvLSTM network inputs, $\widehat{\mathbf{H}}''_s \in \mathbb{R}^{2N_t N_r}$ is reshaped to $\breve{\mathbf{H}}_s \in \mathbb{R}^{N_t \times 2N_r}$, and $\widehat{\mathbf{H}}''_m \in \mathbb{R}^{2N_t N_r K}$ is reshaped to $\breve{\mathbf{H}}_m \in \mathbb{R}^{N_t \times N_r \times K \times 2}$. The reshaped CSI is then predicted by the Bi-LSTM or Bi-ConvLSTM network. For CSI feedback, each time step of the Bi- LSTM and Bi-ConvLSTM network has an output. Let $\breve{\mathbf{H}}_s = \left[\breve{\mathbf{h}}_s^{(1)}, \cdots, \breve{\mathbf{h}}_s^{(t)}, \cdots, \breve{\mathbf{h}}_s^{(N_t)}\right]^{\mathbf{H}}$ and $\breve{\mathbf{H}}_m = \left[\breve{\mathbf{h}}_m^{(1)}, \cdots, \breve{\mathbf{h}}_m^{(t)}, \cdots, \breve{\mathbf{h}}_m^{(N_t)}\right]^H$, where $\breve{\mathbf{h}}_s^{(t)} \in \mathbb{R}^{2N_r}$ and $\breve{\mathbf{h}}_m^{(t)} \in \mathbb{R}^{N_r \times K \times 2}$.

Considering the correlation between antennas in massive MIMO systems, Bi-LSTM and Bi-ConvLSTM networks are designed to predict CSI using correlation between antennas in massive MIMO systems. Because Bi-LSTM and Bi-ConvLSTM networks features well performance in sequence task learning, this paper uses Bi-LSTM and Bi-ConvLSTM networks for CSI prediction. For forward prediction, the CSI of the latter antenna is predicted by the CSI of the previous antenna. For backward prediction, the CSI of the previous antenna is predicted by the latter antenna, and the CSI of the forward and backward antennas is fully utilized to further improve the recovery accuracy of the CSI.

The Bi-LSTM network is used to predict the data for single-user, which is a combination of two LSTM networks. One of the LSTM networks is used for forward data prediction and the other LSTM network is used for backward prediction. Each LSTM network is composed of several LSTM units. Each unit is composed of the input gate, forget gate, output gate and memory unit. This LSTM structure can also be called feedforward LSTM because its internal departments are calculated based on a feedforward-like neural network.

The hidden layer and output update transformation formula of each time step LSTM network can be simplified to equations (8) and (9)

$$\mathbf{l}_s^{(t)} = LSTM\left(\mathbf{l}_s^{(t-1)}, \breve{\mathbf{h}}_s^{(t)}, \theta_s\right) \tag{8}$$

$$\mathbf{o}_s^{(t)} = \tanh\left(\mathbf{W}_{so}\mathbf{l}_s^{(t)} + b_{so}\right) \tag{9}$$

where, $\mathbf{l}_s^{(t)}$ and $\mathbf{o}_s^{(t)}$ denote the hidden layer and the final output vector of the LSTM network at tth antenna, respectively, $\mathbf{l}_s^{(t-1)}$ denotes the hidden layer vector of LSTM network at t-1th antenna, θ_s denotes all the parameters of the LSTM network. $\mathbf{W}_{so}$ and $\mathbf{b}_{so}$ denote hidden-to-output weight and bais, respectively, $LSTM(\cdot)$ denotes the update transformation formula of the hidden layer from the t-1th antenna to the tth antenna. The tanh activation function is used in LSTM, and its expression is

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{10}$$

Bi-LSTM network is a combination of two LSTM networks, so the output transformation formula of the Bi-LSTM network is

$$\tilde{\mathbf{h}}_s^{(t)} = Concat\left(\mathbf{o}_{sf}^{(t)}, \mathbf{o}_{sr}^{(t)}\right) \tag{11}$$

where $\mathbf{o}_{sf}^{(t)}$ and $\mathbf{o}_{sr}^{(t)}$ denote the forward and backward output of Bi-LSTM at tth antenna, respectively. $\tilde{\mathbf{h}}_s^{(t)}$ denotes the total output of Bi-LSTM at tth antenna, $Concat(\cdot)$ denotes the function that concatenates two vectors in a specified dimension. Similarly, Bi-ConvLSTM, which is the same as Bi-LSTM, is used to predicts CSI, but Bi-ConvLSTM is used for multi-user, and Bi-ConvLSTM is somewhat different from Bi-LSTM. The feedforward LSTM can handle the timing data well, but for spatial data, it will bring redundancy because the spatial data has strong local features, and the feedforward LSTM cannot describe this local feature. ConvLSTM attempts to solve this problem, which replaces the feedforward connection with convolution between the input and each gate in the feedforward LSTM network, and the convolution operation is also replaced between states.

So the output of Bi-LSTM at each time step can be expressed as

$$\tilde{\mathbf{h}}_s^{(t)} = Bi - LSTM\left(\mathbf{l}_{sf}^{(t-1)}, \mathbf{l}_{sr}^{(t-1)}, \breve{\mathbf{h}}_{sf}^{(t)}, \breve{\mathbf{h}}_{sr}^{(t)}, \Theta_s\right) \tag{12}$$

$$\tilde{\mathbf{h}}_m^{(t)} = Bi - ConvLSTM\left(\mathbf{l}_{mf}^{(t-1)}, \mathbf{l}_{mr}^{(t-1)}, \breve{\mathbf{h}}_{mf}^{(t)}, \breve{\mathbf{h}}_{mr}^{(t)}, \Theta_m\right) \tag{13}$$

where, $\mathbf{l}_{sf}^{(t-1)}$ and $\mathbf{l}_{sr}^{(t-1)}$ denote the forward and backward hidden layer vectors of the Bi-LSTM network at t-1th antenna, respectively, $\breve{\mathbf{h}}_{sf}^{(t)}$ and $\breve{\mathbf{h}}_{sr}^{(t)}$ denote forward and backward input of the Bi-LSTM network at tth antenna, respectively, Θ_s denotes all the parameters of the Bi-LSTM network, $\tilde{\mathbf{h}}_s^{(t)}$ denotes the final output of the Bi-LSTM network at tth antenna. $\mathbf{l}_{mf}^{(t-1)}$ and $\mathbf{l}_{mr}^{(t-1)}$ denote the forward and backward hidden layer vectors of the Bi-ConvLSTM network at t-1th antenna, respectively, $\breve{\mathbf{h}}_{mf}^{(t)}$ and $\breve{\mathbf{h}}_{mr}^{(t)}$ denote forward and backward input of the Bi-ConvLSTM network at tth antenna, respectively, Θ_m denotes all the parameters of the Bi-ConvLSTM network, $\tilde{\mathbf{h}}_m^{(t)}$ denotes the final output of the Bi-ConvLSTM network at tth antenna.

The output dimensions of the Bi-LSTM and Bi-ConvLSTM network are twice their input, respectively, i.e., $\tilde{\mathbf{h}}_s^{(t)} \in \mathbb{R}^{4N_r}$ and $\tilde{\mathbf{h}}_m^{(t)} \in \mathbb{R}^{N_r \times K \times 4}$. Finally, the output of each time step of the Bi-LSTM and Bi-ConvLSTM network is dimensionally transformed through the fully connected network, so that the final output dimension is consistent with the input dimension. After the data passes through the fully connected network, the outputs $\bar{\mathbf{h}}_s^{(t)} \in \mathbb{R}^{2N_r}$ and $\bar{\mathbf{h}}_m^{(t)} \in \mathbb{R}^{N_r \times K \times 2}$ are obtained. Finally, add their real and imaginary parts together to get the final output

$$\hat{\mathbf{H}}_s = \left[\hat{\mathbf{h}}_s^{(0)}, \cdots, \hat{\mathbf{h}}_s^{(t)}, \ldots, \hat{\mathbf{h}}_s^{(N_t)}\right]^{\mathbf{H}} \tag{14}$$

$$\hat{\mathbf{H}}_m = \left[\hat{\mathbf{h}}_m^{(0)}, \cdots, \hat{\mathbf{h}}_m^{(t)}, \ldots, \hat{\mathbf{h}}_m^{(N_t)}\right]^{\mathbf{H}} \tag{15}$$

where $\hat{\mathbf{H}}_s \in \mathbb{C}^{N_t \times N_r}$, $\hat{\mathbf{H}}_m \in \mathbb{C}^{N_t \times N_r \times K}$, $\hat{\mathbf{h}}_s^{(t)} \in \mathbb{C}^{N_r}$, $\hat{\mathbf{h}}_m^{(t)} \in \mathbb{C}^{N_r \times K}$. So far, the signal processing flow of the proposed DL-based CSI feedback algorithm for massive MIMO system is sorted out as shown in Algorithm 1.

B. OFFLINE MODEL TRAINING AND ONLINE FEEDBACK

Considering the two sides of the existing wireless communication system, especially the user side is limited by energy and computing power, the algorithm proposed in this paper adopts offline model training and online feedback for signal processing. This paper collects data through MATLAB and uses Python for the offline model training. For the well-trained model, we can use it directly when we need feedback instead of repetitive training.

The training device used in this article is configured as a GeForce GTX 1060 GPU and an Intel Xeon® E3-1231 V3 CPU. The channel model we use is a massive MIMO Rayleigh flat fading channel [8]. Using this channel model, our training data can be obtained through the MATLAB simulation platform. For the offline training, the CSI of massive MIMO channel is used as input data and label data to train the learning network. The number of samples of the training sets, the verification sets and the test sets are 20000, 5000 and 1000 respectively. The initial learning rate of the proposed CSI feedback network is set to 0.01, the training epochs is 100, and the bacth size is 100. In this paper, we use the end-to-end method to obtain all the weights

Algorithm 1 DL-Based CSI Feedback Algorithm for Massive MIMO System

Input: Original channel matrix $\mathbf{H}_s$ and $\mathbf{H}_m$

Output: The estimated values of CSI feedback $\hat{\mathbf{H}}_s$ and $\hat{\mathbf{H}}_m$

Step 1: The input data is preprocessed, and the real and imaginary parts of the channel matrix are separated by the Copy function, and the real part and the imaginary part are taken as a new dimension, and obtain real values $\mathbf{H}'_s$ and $\mathbf{H}'_m$, respectively;

Step 2: After the data is preprocessed, input $\mathbf{H}'_s$ to two 2D CNN networks to extract frequency feature vectors, the output of the first two 2D CNN networks is $\mathbf{H}''_s$; input $\mathbf{H}'_m$ to two 3D CNN networks to extract frequency feature vectors, and the output of the first two 3D CNN networks is $\mathbf{H}''_m$;

Step 3: The outputs parameters to be estimated of the second 2D CNN and the second 3D CNN network are compressed by a 2D Maxpooling and 3D Maxpooling network, respectively, and the compressed outputs are $\widehat{\mathbf{H}}_s$ and $\widehat{\mathbf{H}}_m$, respectively;

Step 4: Reshape $\widehat{\mathbf{H}}_s$ and $\widehat{\mathbf{H}}_m$ into a one-dimensional vector, respectively, to obtain $\widehat{\mathbf{H}}'_s$ and $\widehat{\mathbf{H}}'_m$;

Step 5: Increase the dimension of $\widehat{\mathbf{H}}'_s$ by a fully connected layer with $2N_tN_r$ neurons to obtain $\widehat{\mathbf{H}}''_s$, and the dimension of $\widehat{\mathbf{H}}'_m$ is increased by a fully connected layer with $2N_tN_rK$ neurons to obtain $\widehat{\mathbf{H}}''_m$;

Step 6:$\widehat{\mathbf{H}}''_s$ is reshaped into $\breve{\mathbf{H}}_s$, which is suitable for Bi-LSTM network input form, and $\widehat{\mathbf{H}}''_m$ is reshaped into $\breve{\mathbf{H}}_m$, which is suitable for Bi-ConvLSTM network input form;

Step 7: Input $\breve{\mathbf{H}}_s$ into the Bi-LSTM network to obtain the predicted $\tilde{\mathbf{H}}_s$, and input $\breve{\mathbf{H}}_m$ into the Bi-ConvLSTM network to obtain the predicted $\tilde{\mathbf{H}}_m$;

Step 8: Input $\tilde{\mathbf{H}}_s$ and $\tilde{\mathbf{H}}_m$ to the fully connected layer to reduce their dimensions and obtain $\bar{\mathbf{H}}_s$ and $\bar{\mathbf{H}}_m$, respectively;

Step 9: The real part and imaginary part of $\bar{\mathbf{H}}_s$ and $\bar{\mathbf{H}}_m$ are added respectively and get the final output $\hat{\mathbf{H}}_s$ and $\hat{\mathbf{H}}_m$.

and biases in the network, and use the adaptive moment estimation (ADAM) optimization algorithm to train the network. The ADAM algorithm is different from the conventional gradient descent algorithm with fixed learning rate, it can update the learning rate adaptively through training. The loss function used in this paper is mean square error (MSE), all parameters in the network are updated by minimizing the MSE between the output of the learning network and the label data. The calculation method of the MSE loss function is as follows:

$$L(\Theta_{csi}) = \frac{1}{N}\sum_{i=1}^{N}(f(\mathbf{H}_i;\Theta_{csi}) - \mathbf{H}_i)^2 \tag{16}$$

where $\mathbf{H}_i$ is the original input data of the ith sample, Θ_{csi} is all the parameters in the network, f is the transformation formula of the entire network, $f(\mathbf{H}_i;\Theta_{csi})$ is the recovered matrix of the ith sample, and N is the total number of samples.

TABLE 1. Comparison of the complexity of different algorithms.

Algorithm	Complexity
DCT	$2O\left(N_r^2N_t^2\right)$
KLT	$2O\left(N_r^3N_t^3\right)+2O\left(N_r^2N_t^2\right)$
PCA	$O\left(N_t^3\right)+O\left(N_t^2\right)$
CsiNet	$2O\left(N_rN_t\right)$
Proposed algorithm	$O\left(N_rN_t\right)+O\left(N_t\right)$

IV. ALGORITHMIC COMPLEXITY ANALYSIS

In the following, the complexity of the DL-based CSI feedback algorithm proposed in this paper is compared with the existing algorithms. The complexity of the compression process and the inverse recovery channel matrix process are mainly considered, as shown in Table 1.

For compression process, the DCT algorithm directly performs discrete cosine transform on the channel vector, the real part and imaginary part need to be compressed separately. The KLT algorithm needs to first vectorize the channel matrix and then obtain the KLT compressed sparse basis by calculating the covariance matrix of channel vector. The PCA algorithm needs to calculate a small number of main components. As for the decompression process, the DCT algorithm needs to decompress both real part and imaginary part respectively, KLT algorithm needs to use sparse matrix with large dimensions to decompress the original signal, the PCA algorithm needs to use the compression matrix with small dimensions to decompress the compressed CSI. While in the DL algorithm, the DL network only needs simple matrix multiplication to complete the CSI feedback, so it is less computationally complex than the conventional CSI feedback algorithms. In particular, compared with CsiNet network, the proposed network structure is simpler and has fewer parameters. Theoretical analysis shows that the proposed algorithm has lowest computational complexity.

V. SIMULATION ANALYSIS

In this section, in order to verify the performance of the proposed CSI compression feedback algorithm for FDD massive MIMO systems, the proposed algorithm and other CSI compression feedback algorithms are compared through MATLAB simulation platform. Assuming that the ideal downlink CSI has been obtained on the user side, the main parameters of the simulation system are shown in Table 2. According to formula (4), the spatial correlation can be obtained by the antenna number, working frequency and antenna spacing.

A. COMPARISON OF NORMALIZED MEAN SQUARE ERRORS

For the easy-to-handle analysis of communication performance, the compression feedback error is only considered.

TABLE 2. Simulation parameters.

Simulation parameters	Setting
Transmitting antennas	N_t=32/64
Receiving antennas	N_r=2
Duplex mode	FDD
Modulation	QPSK
Number of users	1/3
Channel model	Flat Rayleigh fading
Frequency	2.6GHz
Power allocation	Equal power division
Antenna spacing	0.5λ

In this paper, the normalized MSE (NMSE) performance of proposed DL-based CSI feedback algorithm and the DCT, PCA, KLT and CsiNet algorithms are compared under the data compression rate is 1/4. The NMSE is the difference between the recovered $\hat{\mathbf{H}}$ and the original $\mathbf{H}$, and it is defined as

$$NMSE = E\left\{\| \mathbf{H} - \hat{\mathbf{H}} \|_2^2 / \| \mathbf{H} \|_2^2\right\} \tag{17}$$

where the smaller the NMSE, the smaller the channel compression feedback error and the better the performance. In this paper, the NMSE is represented in logarithmic form and the results are shown in Table 3.

As can be seen from Table 3, for the case of single-user, when the number of antennas at BS side are 32 and 64, respectively, the NMSE of the proposed DL-based CSI feedback is about 3 and 5 dB smaller than conventional PCA and DCT algorithms, and is about 1.4 and 2 dB smaller than the CsiNet algorithm. For the case of multi-user, the NMSE of the proposed DL-based algorithm is also smaller than that of other algorithms. Table 3 shows that whether it is single-user or multi-user and the number of antennas at the BS side is 32 or 64, the proposed DL-based CSI feedback algorithm has excellent NMSE performance. This algorithm can recover CSI more accurately and improve the quality of the recovered CS significantly. This is mainly because the proposed DL-based CSI feedback network does not need to know the channel distribution, it can be trained by using training samples. The network makes full use of the spatial correlation of massive MIMO channels and can learn the channel structure characteristics well. Therefore, the recovered CSI of the network is closer to the original data.

It is worth mentioning that we do not mention KLT algorithm in this section. This is because KLT algorithm could vectorize the channel matrix and then calculate the covariance matrix of channel vectors to get KLT compression sparse base. When the base station and mobile users all get instantaneous correlation matrix, KLT algorithm can get the perfect sparse performance. So it can recover the information accurately using few measured data and the error is almost 0 all the time. However, when the base station does not get the instantaneous correlation matrix, receiver needs extra feedback overhead to feedback the KLT sparse base to the base station to recover the original signal. As the amount of information is large, it is almost equal to complete feedback. Therefore, when the BS does not get instantaneous correlation matrix, KLT algorithm cannot have sparse compression effect.

B. BER

In general, the BER performance is a macro measure of the impact of CSI recovery algorithms on overall system performance. In this paper, the feedback compression ratio is set to 1/4. Under the condition of ensuring the same feedback compression ratio, the BER performance of single-user and multi-user massive MIMO system are simulated respectively.

Fig. 4 shows the variation curves of BER performance with signal to noise ratio (SNR) of different algorithms for a single-user system with 32 and 64 antennas at BS and 2 antennas at receiver. It can be seen from Fig. 4 that the BER performance of the proposed DL-based CSI feedback algorithm is better than the DCT, PCA, KLT and CsiNet algorithms. For conventional DCT, PCA and KLT algorithms, the BER performance of DCT is the worst, KLT is the best, and CsiNet algorithm is better than the conventional algorithms.

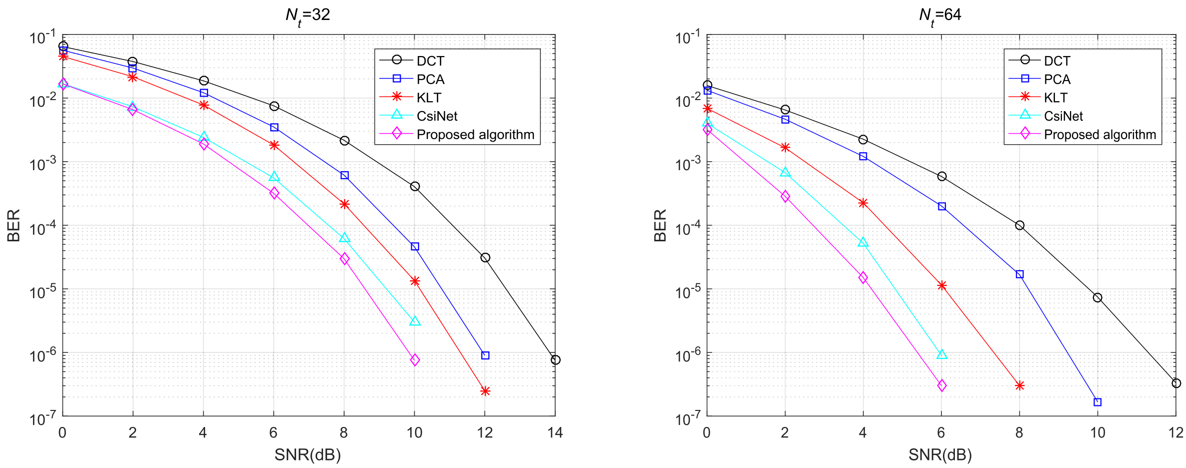
When the BS antenna is 32, the proposed algorithm has a SNR gain of about 2-4 dB compared with the conventional DCT, PCA, and KLT algorithms, and the SNR gain is about 0.5 dB compared with the CsiNet algorithm. When the number of antennas at BS is 64, the proposed algorithm has about 2-6 dB SNR gain compared with the conventional DCT, PCA, and KLT algorithms, and has about 0.7 dB SNR gain compared with the CsiNet algorithm. As can be seen from Fig. 4, when the BER performance is 10^{-6} and the number of antennas at BS is 32, the SNR required for the proposed algorithm is about 10 dB, while is about 5.5 dB required for 64 antennas. This shows that the more the number of antennas at BS in massive MIMO system based on spatial correlation, the better the BER performance.

Fig. 5 shows the curves in which BER performance varies with signal to noise ratio (SNR) for different algorithms of a multi-user system when the BS has 32 or 64 antennas, the receiver has 2 antennas and the number of users is 3. The BER performance of multi-user system is consistent with that of single-user system, regardless of 32 or 64 antennas at BS. The BER performance of proposed algorithms is better than that of DCT, PCA, KLT and CsiNet algorithms, and the more antennas at BS side, the better the BER performance.

As can be seen from Fig. 4 and Fig. 5, the proposed algorithm has better BER performance than the DCT, PCA, KLT, and CsiNet algorithms under the conditions of 32 and 64 antennas at BS, single-user and multi-user scenarios. This is because the DCT needs to use the correlation matrix of the downlink channel to generate compression matrix and recovery matrix, and terminals need to periodically feedback the information to the BS according to the change of the channel. When the channel changes rapidly, it is difficult for the BS to acquire the channel-relative information accurately

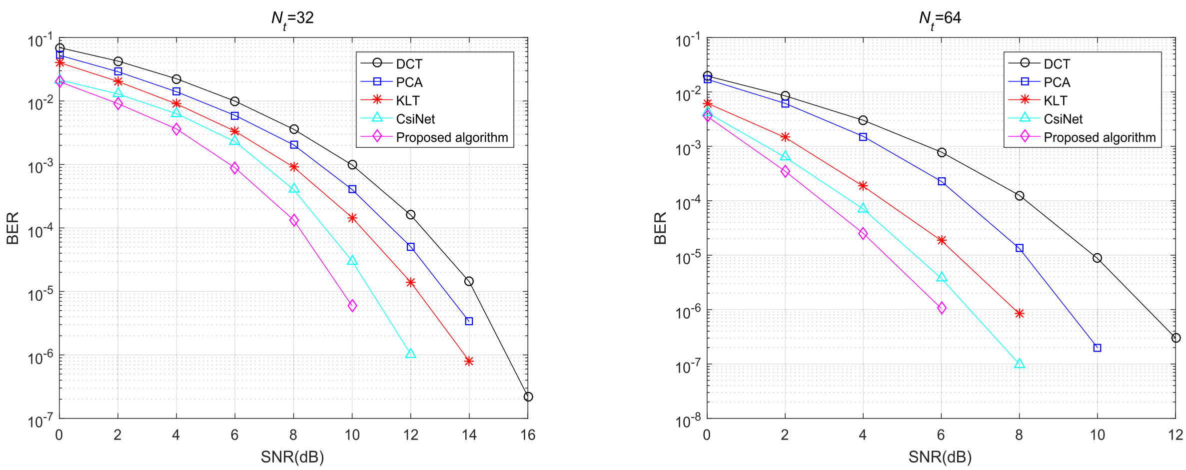
TABLE 3. NMSE (dB) comparison of various CSI feedback algorithms.

Number of user	Number of transmitting antennas	DCT	PCA	CsiNet	Proposed algorithm
1	32	-14.32	-14.85	-16.12	**-17.58**
	64	-15.01	-16.28	-18.53	**-20.48**
3	32	-13.89	-14.48	-15.81	**-17.16**
	64	-14.57	-15.92	-18.23	**-20.19**



(a) N_t=32 (b) N_t=64

FIGURE 4. Comparison of BER of various feedback algorithms for single-user system.

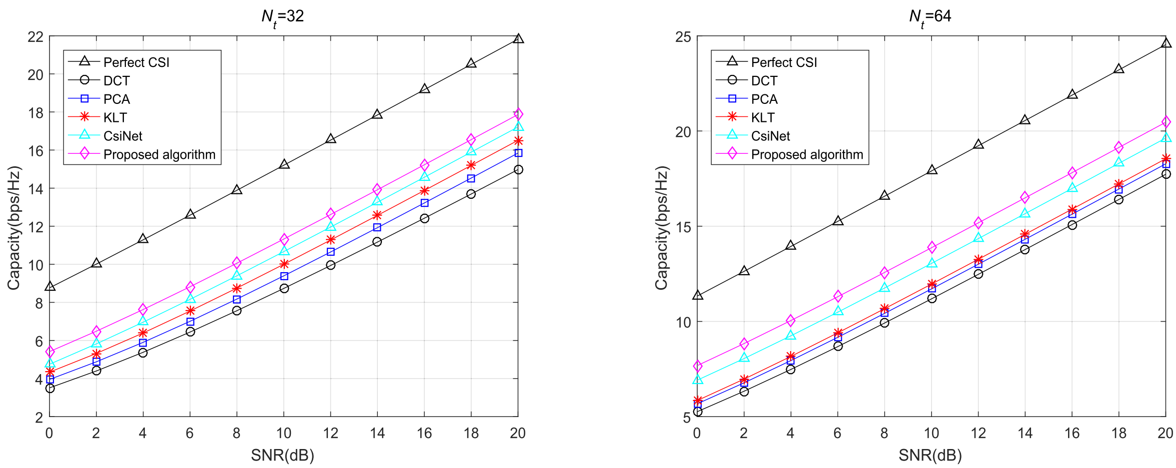


(a) N_t=32 (b) N_t=64

FIGURE 5. Comparison of BER of various feedback algorithms for multi-user system.

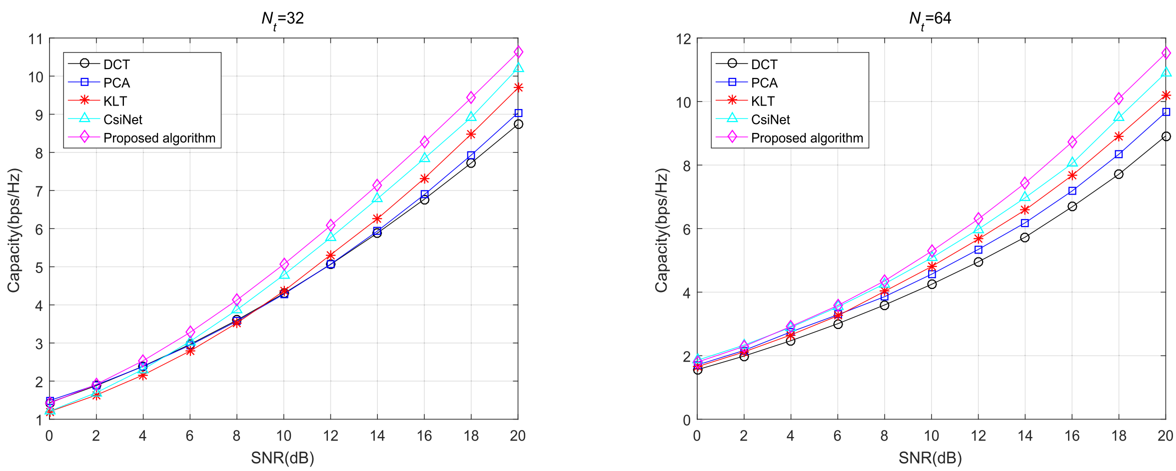
and timely, which causes a decrease in feedback accuracy. The PCA compresses and recovers the channel matrix by extracting the main components, but the compression matrix is estimated by the previous long period, so it is difficult for the BS to obtain accurate CSI. The KLT algorithm can recover channel matrix better than DCT and PCA algorithm. This is because KLT base has the best sparse representation when both BS and mobile terminal know instantaneous correlation matrix. It only needs very little measurement data to recover channel matrix accurately. However, KLT base has signal dependence and needs to feed back the channel instantaneous matrix, which increases the amount of feedback, and the computational complexity of the KLT algorithm is very high.

In addition, CsiNet algorithm does not consider the antenna correlation in massive MIMO system, but the proposed algorithm makes full use of the antenna correlation in massive MIMO systems to accurately recover CSI. The network used can better mine the non-linear characteristics of data, and the



(a) N_t=32 (b) N_t=64

FIGURE 6. **Comparison of BER of various feedback algorithms for single-user system.**



(a) N_t=32 (b) N_t=64

FIGURE 7. **Comparison of BER of various feedback algorithms for multi-user system.**

features of the channel can be learned well. Therefore, for all of above case, the proposed algorithm shows excellent BER performance, and the error of the recovered matrix is very small, which effectively improves the feedback accuracy of CSI.

C. SYSTEM CAPACITY

Fig. 6 shows the system capacity variation with SNR for a single-user system with 32 antennas and 64 antennas at the BS and 2 antennas at the receiver. It can be seen from Fig. 6 that the system capacity of the proposed algorithm increases with the increase of SNR, and the system capacity of the proposed algorithm is higher than other algorithms. When the number of antennas at BS side is 32, the system capacity of the proposed algorithm is increased by 1.6-2.8 bps/Hz than the conventional algorithms. Compared with the CsiNet algorithm, the system capacity is increased by about 0.3 bps/Hz. When the BS end is 64 antennas, the system capacity of the proposed algorithm is increased by about 2-3 bps/Hz than the conventional algorithms. Compared with the CsiNet algorithm, the system capacity is increased by about 0.5 bps/Hz. The proposed algorithm maintains a high system performance gain under the different antenna configurations at the BS.

Fig. 7 shows the curves of system capacity with SNR for different algorithms of multi-user system with 32 and 64 antennas at BS and 2 antennas at receiver and 3 users. From Fig. 7, it can be seen that for massive multi-user MIMO systems, the proposed algorithm has obvious performance improvement and better channel capacity under high SNR conditions. When the SNR is 20 dB and the number of antennas at BS side is 32, the system capacity of the proposed algorithm is about 1-2 bps/Hz higher than the conventional algorithms, about 0.4 bps/Hz higher than the CsiNet algorithm. When the number of antennas at BS side is 64, the system capacity of the proposed algorithm is about 1.5-2.5 bps/Hz higher than the conventional algorithms and about 0.6 bps/Hz higher than the CsiNet algorithm. However, compared with single-user, under the same number of antennas at BS and SNR, the capacity of multi-user system is smaller than that of single-user system because of user interference between multiple users.

VI. CONCLUSION

In this paper, the CSI compression feedback algorithm based on spatial correlation for FDD massive MIMO systems is studied. Considering the high complexity and inaccurate feedback of the conventional CSI compression feedback algorithms, and the CsiNet algorithm does not consider the spatial correlation of the antenna in the massive MIMO systems and its feedback accuracy is not high, this paper proposes a novel DL-based CSI compression feedback algorithm and apply it to single-user and multi-user scenarios. The algorithm considers the spatial correlation of massive MIMO uniform linear antenna arrays, and makes full use of the channel information in the training samples, the channel structural characteristics learned by the model can better represent the channel, thus improving the accuracy of CSI compression feedback. The theoretical analysis and simulation results show that the proposed algorithm has lower complexity, better BER and system capacity performance than the conventional CSI compression feedback algorithms and CsiNet algorithm.

REFERENCES

[1] C. Li, P. Liu, C. Zou, F. Sun, J. M. Cioffi, and L. Yang, "Spectral-efficient cellular communications with coexistent one- and two-hop transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6765–6772, Aug. 2016.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.*, vol. 59, Aug. 2016, Art. no. 081301.

[4] C. Li, Y. Li, K. Song, and L. Yang, "Energy efficient design for multiuser downlink energy and uplink information transfer in 5G," *Sci. China Inf. Sci.*, vol. 59, no. 2, pp. 1–8, Feb. 2016.

[5] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.

[6] C. Li, L. Yang, and W.-P. Zhu, "Two-way MIMO relay precoder design with channel state information," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3358–3363, Dec. 2010.

[7] C.-C. Tseng, J.-Y. Wu, and T.-S. Lee, "Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block ℓ_1-minimization," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1055–1067, Mar. 2016.

[8] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Jun. 2012, pp. 492–497.

[9] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, "Compressed channel feedback for correlated massive MIMO systems," *J. Commun. Netw.*, vol. 18, pp. 95–104, Feb. 2016.

[10] T. Zhang, A. Ge, N. C. Beaulieu, Z. Hu, and J. Loo, "A limited feedback scheme for massive MIMO systems based on principal component analysis," *EURASIP J. Adv. Signal Process.*, vol. 1, p. 64, Dec. 2016.

[11] A. Ge, T. Zhang, Z. Hu, Z. Zeng, "Principal component analysis based limited feedback scheme for massive MIMO systems," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Dec. 2015, pp. 326–331.

[12] A. Ge, T. Zhang, Z. Zeng, and Y. Sun, "PCA based limited feedback scheme for massive MIMO with Kalman filter enhancing performance," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Apr. 2015, pp. 1–6.

[13] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions," 2019, *arXiv:1904.09673*. [Online]. Available: https://arxiv.org/abs/1904.09673

[14] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019. doi: 10.1109/TVT.2019.2900460.

[15] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2018.

[16] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[17] W. Xu, Z. Zhong, Y. Be'ery, X. You, and C. Zhang, "Ery, "Joint neural network equalizer and decoder," in *Proc. Int. Symp. Wireless Commun. Syst.*, Oct. 2018, pp. 326–331.

[18] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Sep. 2019.

[19] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[20] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[21] R. Iqbal, T. D. Abhayapala, and T. A. damahewa, "Generalised clarke model for mobile-radio reception," *IET Commun.*, vol. 3, no. 4, pp. 644–654, Apr. 2009.

YONG LIAO (M'14) received the Ph.D. degree from Chongqing University, Chongqing, China, in 2014, where he is currently a Research Associate and the Deputy Director with the Key Laboratory of Aircraft TT&C and Communication, Ministry of Education. His research interests include high-speed mobile communication, 5G and future communication, and aerocraft TT&C and communication.

HAIMEI YAO received the B.S. degree from the Jiangxi University of Science and Technology, Ganzhou, China, 2016. She is currently pursing the M.S. degree with Chongqing University, Chongqing, China. Her current research interest includes AI and its application in wireless communication.

YUANXIAO HUA received the B.S. degree from the North University of China, Taiyuan, China, 2016. He is currently pursuing the M.S. degree with Chongqing University, Chongqing, China. His current research interests include wireless communications channel estimation and deep learning in communication applications.

CHUNGUO LI (SM'16) received the bachelor's degree in wireless communications from Shandong University, in 2005, and the Ph.D. degree in wireless communications from Southeast University, in 2010.

In 2010, he joined the Faculty of Southeast University, Nanjing, where he became an Associate Professor, in 2012, and a Full Professor, in 2017. From 2012 to 2013, he was a Postdoctoral Researcher with Concordia University, Montreal, Canada. From 2013 to 2014, he was with the DSL Laboratory as a Visiting Associate Professor, supervised by Prof. J. M. Cioffi. He has been the Supervisor of Ph.D. candidate, since 2016. His research interests include 5G cellular transmission, underwater communications, machine learning for video signal processing, and next generation of Wi-Fi.

Dr. Li is a Senior Member of the Chinese Institute of Electronics, a Senior Member of the Chinese Computer Foundation (CCF). He is also a Fellow of IET. He was a recipient of the Best Ph.D. Thesis Award of Southeast University, in 2010, the Excellent Foreign Postdoc Award of Canada, in 2012, the Science and Technology Progress Award of the National Education Ministry of China, in 2014, the Excellent Visiting Associate Professor at Stanford, in 2014, the Southeast University Excellent Young Professor Award, in 2015, the Excellent Teaching Award of Southeast University, in 2016, the Cyrus Tang Foundation Endowed Professorship, in 2018, the Scientific Research Achievement Award of the Jiangsu Province Education Department, in 2018, and several conference best paper awards. He has served for many IEEE conferences, including the IEEE 16th International Symposium on Communications and Information Technologies as the Track Chair of Wireless Communications, Special Session Signal processing and air interface design solutions for Beyond 5G wireless communications on the IEEE PIMRC-2018 as the Chair, the International Conference on Communications, the International Conference on Acoustics, Speech and Signal Processing as the TPC Member. He is a regular Reviewer for many IEEE Journals. He is currently an Associate Editor for *IET Communications*, an Area Editor of the *AEU-International Journal of Electronics and Communications* (Elsevier), an Editor of *Telecommunications Systems*, an Associate Editor of *Circuits, Systems and Signal Processing*, an Editor of the *KSII Transactions on Internet and Information Systems*, the Leading Guest Editor of the Special Issue Ultra-Reliable-and-Available Low-Latency Communications for 5G/B5G-enabled IoT on *EURASIP Journal on Wireless Communications and Networking*, and the Guest Editor of the Special Issue Unmanned Aerial Vehicle Assisted Communications and Networking on *Physical Communications*.