DL-based CSI Feedback and Cooperative Recovery in Massive MIMO

Jiajia Guo, Xi Yang, Chao-Kai Wen, *Member, IEEE*, Shi Jin, *Senior Member, IEEE*, and Geoffrey Ye Li, *Fellow, IEEE*

Abstract

In this paper, we exploit the correlation between nearby user equipment (UE) and develop a deep learning-based channel state information (CSI) feedback and cooperative recovery framework, CoCsiNet, to reduce the feedback overhead. The CSI can be divided into two parts: shared by nearby UE and owned by individual UE. The key idea of exploiting the correlation is to reduce the overhead used to repeatedly feedback shared information. Unlike in the general autoencoder framework, an extra decoder is added at the base station to recover shared information from the feedback CSI of two nearby UE, but no modification is performed at the UE. For a UE with multiple antennas, we also introduce a baseline neural network architecture with long short-term memory modules to extract the correlation of nearby antennas. We propose two representative methods for converting CSI into bitstreams at the UE through quantization and binarization. Given that the CSI phase is not sparse, we propose a magnitude-dependent phase feedback strategy that introduces CSI magnitude information to the loss function of the phase feedback. Simulation results show the effectiveness of the proposed CoCsiNet and explain the mechanism of the CoCsiNet via parameter visualization.

Index Terms

Massive MIMO, CSI feedback, cooperation, deep learning.

J. Guo, X Yang, and S. Jin are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210096, P. R. China (email: {jiajiaguo, yangxi, jinshi}@seu.edu.cn).

C.-K. Wen is with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan (e-mail: chaokai.wen@mail.nsysu.edu.tw).

G. Y. Li is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: liye@ece.gatech.edu).

I. Introduction

Massive multiple-input multiple-output (MIMO) is regarded as a critical technology in future wireless communications [1]–[4]. Spectral efficiency can be dramatically improved in massive MIMO systems, which use numerous antennas at the base station (BS) [5]. The performance gain achieved by this technology primarily comes from the spatial multiplexing in addition to diversity, which relies on the knowledge of the instantaneous uplink and downlink channel state information (CSI) at the BS. Therefore, the quality of available CSI at the BS directly affects the performance of massive MIMO systems .

The uplink CSI can be estimated at the BS by the pilots sent from the user equipment (UE). In time-division duplexing systems, the downlink CSI can be inferred from the uplink CSI by using channel reciprocity. In frequency-division duplexing (FDD) systems, which are widely employed by cellular systems, little reciprocity exists between uplink and downlink channels. Consequently, inferring the downlink CSI using only the uplink CSI is impossible. To obtain the downlink CSI, the BS should first transmit pilot symbols for the UE to estimate CSI locally. Then, the UE feeds the estimated CSI back to the BS. The feedback overhead linearly scales with the product of the numbers of antennas at the BS and the UE and is extremely large in massive MIMO systems, which occupy precious bandwidth source and are sometimes unaffordable [6], [7]. Therefore, a technique must be developed to efficiently reduce the feedback overhead while exerting negligible negative effects on system performance in FDD systems.

One well-known and promising technique is compressive sensing (CS) theory via exploiting the CSI sparsity in certain domains [8]. In [9], the sparsity in the spatial-frequency domain due to the short distance between antennas is exploited to compress CSI and reduce the feedback overhead. In [10], a distributed compressive downlink channel estimation and feedback scheme uses the hidden joint sparsity structure of CSI resulting from shared local scatterers.

Although existing CS-based methods can greatly reduce the feedback overhead, two main problems remain. The reconstruction based on the compressed CSI at the BS can be regarded as an optimization problem that is solved by iterative algorithms, which demand substantial time and computing sources. By contrast, during the reconstruction phase, the correlation of CSI to the nearby UE is usually ignored.

Since the success achieved by AlexNet [11] in the ImageNet competition, deep learning (DL) has made tremendous progress in computer vision and nature language processing [12]. The

communication community has recently shown great interest in applying DL to improve the performance of communication systems [13]–[16]. A totally data-driven end-to-end communication system based on the autoencoder architecture is proposed in [17], [18] where DL-based encoder and decoder play the role of transmitter and receiver, respectively. DL is also widely used to enhance certain blocks of conventional communication systems, such as, channel estimation [19], [20] and joint channel estimation and signal detection [21], [22].

DL is first applied to the CSI feedback in [23] based on the autoencoder architecture, where the encoder at the UE and the decoder at the BS perform compression and reconstruction, respectively. The compression of the hidden layers forces the encoder to capture the most dominant features of the input data, that is, compressing the CSI. Given its numerous potentials, such as feedback accuracy and reconstruction speed, DL-based CSI feedback has become an active research area in communications, and related works can be divided into the following four main categories:

- Exploiting expert knowledge and extra correlation in massive MIMO systems. The long short-term memory (LSTM) architecture adopted in [24], [25] extracts the temporal correlation of time-varying channels. In [26], CSI magnitude and phase are compressed, and the uplink CSI magnitude is used to help recover downlink CSI magnitude during the reconstruction phase at the BS.
- Designing novel neural network (NN) architectures. This kind of work focuses on improving feedback performance by introducing novel NN architectures in computer vision to the CSI feedback problem, including the attention model [27], the joint convolutional residual network [28], depth-wise separable convolution [29], and multi-resolution convolutional architecture [30]. In [31], we change the convolutional kernel size from 3×3 to 7×7 to enlarge the receptive field, thereby exhibiting the CSI sparsity characteristic well.
- Quantization and entropy encoding. Given that the UE must feed the bitstream back rather than the float-point number in practical digital systems, the encoder module should be followed by a quantization module. After comparing the performance of uniform and non-uniform quantization in DL-based CSI feedback, an offset NN is introduced to minimize the quantization distortion in [31] . Given that the quantization operation is non differentiable, its gradient is set to unit in [28]. Moreover, to further reduce the feedback overhead, an entropy coding module following a quantization module is used to further reduce the feedback bits in [32]. Unlike other quantization strategies, binary representation is exploited in [33] to

generate bitstreams.

- Tackling the deployment problem. Some new problems appear when deploying the DL-based CSI feedback method. A multiple-rate CSI feedback framework is proposed in [31] to address different compression ratios by using a single network. Fully connected (FC) layers are replaced by fully convolutional layers in [32], enabling the same NNs to work with different antenna numbers. In [34], we apply network pruning and the quantization technique to CSI feedback, greatly reducing the storage and computational requirement of NNs. In addition, we analyze the security of DL-based analog CSI feedback under adversarial attack [35].

Many studies [36]–[39] observe that the CSI of nearby users may share deterministic multipath components and have similar sparsity structures due to some shared local scattering clusters, which have been widely exploited in CS-based channel estimation. The above observation can become increasingly common as the device density grows to hundred(s) per cubic meter in the future [3]. Therefore, the correlation of nearby UE should be exploited to further reduce the feedback overhead and improve the feedback performance.

In this paper, we develop a cooperative CSI feedback neural network, called CoCsiNet. Enabled by device-to-device (D2D) communications, the conventional cooperation mechanism is developed in [40], [41]. CSI magnitude and phase are separately fed back, rather than together as complex channel gains. However, CSI phase is not sparse and is difficult to greatly compress. We notice that a highly accurate phase feedback is unnecessary for CSI with a small magnitude. Inspired by [26], we propose a magnitude-dependent phase feedback (MDPF) NN by introducing the corresponding magnitude information to the NN loss function for the phase feedback. In the CSI magnitude feedback phase, we divide the CSI magnitude information into two parts: shared by nearby UE and owned by individual UE. These two kinds of information are automatedly learned from the data and combined at the BS by using an end-to-end approach and the CSI need not be exchanged between UE. For UE with multiple antennas, an LSTM architecture is introduced to exploit the correlation among the antennas at the same UE. Our contributions in this paper are summarized as follows:

- We introduce the cooperation mechanism to DL-based CSI feedback. We divide the CSI magnitude information into two parts as in [39]: shared by nearby UE and owned by individual UE. An extra shared decoder is introduced to recover the shared information

from the feedback bits of two nearby UE.

- We develop an autoencoder-based framework. For UE with multiple antennas, the LSTM architecture is added to the decoder at the BS to exploit the correlation of CSI to different antennas at the same UE.
- To generate the bitstreams at the UE, we investigate two representative methods (i.e., quantization and binarization), and adopt them to the DL-based CSI feedback.
- To compress the CSI phase, which is not sparse and is usually difficult to compress, an MDPF framework is proposed, where the magnitude information is added to the NN loss function of the feedback but no modification on the NN architecture is required.

The rest of this paper is organized as follows. Section II introduces the massive MIMO channel model and the CSI feedback process. Section III presents the cooperative CSI feedback strategy and the NN framework and then describes the baseline NN architecture, the bitstream generation, and the CSI phase feedback strategy. Section IV provides the numerical results of the proposed methods and demonstrates the mechanism of the CoCsiNet via parameter visualization. Section V finally concludes our paper.

II. System Model

After introducing the massive MIMO channel model in this section, we describe the DL-based CSI feedback process.

A. Massive MIMO Channel Model

Considering a massive MIMO system, the BS and the UE are equipped with N_t ($\gg 1$) antennas and N_r (≥ 1) antennas, respectively. According to the spatial multipath channel model [42], the downlink channel matrix, $\tilde{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$, can be written as

$$\tilde{\mathbf{H}} = \sqrt{\frac{N_r N_t}{N_c}} \sum_{l=1}^{N_c} g_l \mathbf{a}_r(\varphi_{r,l}) \mathbf{a}_t^H(\varphi_{t,l}), \tag{1}$$

where N_c is the total path number; $g_l \sim \mathcal{CN}(0,1)$ denotes the complex gain of the l-th path; $\mathbf{a}_r(\cdot)$ and $\mathbf{a}_t(\cdot)$ are the steering vectors of the receiver and transmitter; $\varphi_{r,l}$ and $\varphi_{t,l}$ represent the angle-of-arrival (AoA) and the angle-of-departure (AoD), respectively. If the BS and the UE are all equipped with uniform linear antenna arrays (ULAs), then their steering vectors can be written as

$$\mathbf{a}_r(\varphi_{r,l}) = \sqrt{\frac{1}{N_r}} [1, e^{\frac{-j2\pi d \sin(\varphi_{r,l})}{\lambda}}, \quad \cdots \quad , e^{\frac{-j2\pi d (N_r-1) \sin(\varphi_{r,l})}{\lambda}}]^T, \tag{2}$$

$$\mathbf{a}_r(\varphi_{t,l}) = \sqrt{\frac{1}{N_t}}[1, e^{\frac{-j2\pi d \sin(\varphi_{t,l})}{\lambda}}, \quad \cdots \quad , e^{\frac{-j2\pi d(N_t-1)\sin(\varphi_{t,l})}{\lambda}}]^T, \tag{3}$$

where λ and d denote the carrier wavelength and antenna element spacing, respectively. The BS is usually located at a high building, and channel path number N_c is limited due to the few scatterers. Therefore, CSI matrix $\mathbf{H}$ in the angle domain, which can be obtained through discrete Fourier transform, is usually sparse [43].

B. Conventional DL-based Digital CSI Feedback Process

Here, we assume that the UE knows the perfect CSI and focuses on the CSI feedback. For most DL-based CSI feedback methods, once the UE obtains channel matrix $\mathbf{H}$, an NN is first used to compress the CSI, which directly outputs the measurement vectors with 32-bit floating point numbers. Then, the following quantization module generates the bitstreams, which are fed back to the BS immediately. Subsequently, the BS reconstructs the CSI from the measurement vectors disturbed by the quantization noise. The entire process can be expressed as

$$\hat{\mathbf{H}} = f_{\text{de}}(\mathcal{D}(\mathcal{Q}(f_{\text{en}}(\mathbf{H}, \Theta_{\text{en}}))), \Theta_{\text{de}}), \tag{4}$$

where $f_{\text{en}}(\cdot)$ and $f_{\text{de}}(\cdot)$ represent the DL-based compression (encoder) and decompression (decoder) operations at the UE and the BS, respectively; $\mathcal{Q}$ and $\mathcal{D}$ denote the quantization and the dequantization operations, respectively; Θ_{en} and Θ_{de} are the parameters at the NN-based encoder at the UE and the decoder at the BS, respectively. The final feedback overhead, that is, the bitstream length, is derived through the compression ratio and the number of quantization bits as

$$N_{\text{bits}} = L \times \gamma \times B, \tag{5}$$

where L is the original dimension of the CSI, γ is the compression ratio, and B denotes the number of quantization bits.

The entire process is optimized by an end-to-end approach to minimize the loss function via updating the NN parameters. The most widely used loss function in CSI feedback is the mean-squared error (MSE):

$$(\hat{\Theta}_1, \hat{\Theta}_2) = \underset{\Theta_1, \Theta_2}{\arg\min} \quad \|\mathbf{H} - f_{\text{de}}(\mathcal{D}(\mathcal{Q}(f_{\text{en}}(\mathbf{H}, \Theta_1))), \Theta_2)\|_2^2, \tag{6}$$

where $\|\cdot\|_2$ represents the Euclidean norm.

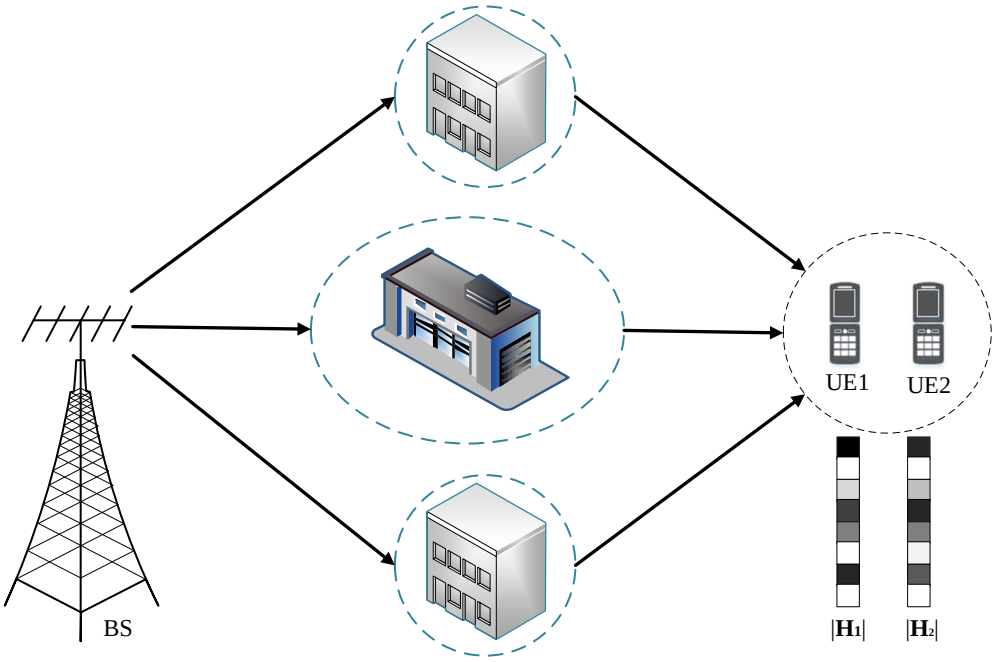


Fig. 1. Illustration of the system model. The BS and the two nearby UE are equipped with ULAs with N_t transmit and N_r receive antennas, respectively.

To evaluate the difference between the original and the reconstructed CSI at the BS, the normalized MSE (NMSE) is used as follows:

$$\text{NMSE} = \text{E}\left\{\frac{\|\hat{\mathbf{H}} - \mathbf{H}\|_2^2}{\|\mathbf{H}\|_2^2}\right\}. \tag{7}$$

III. DL-based CSI Feedback and Cooperative Recovery

In this section, we consider the scenario with two nearby UE that share some reflection or diffraction media and are correlated in the CSI magnitude. Unlike the conventional cooperation strategy, which is enabled by D2D communications, no extra operations are performed at the UE in our DL-based cooperation strategy. The key idea is to reduce the feedback overhead for the shared information by nearby UE. After introducing our cooperation strategy, we present the NN modules in the CoCsiNet.

A. Cooperation Strategy

As shown in Fig. 1, UE can share the same scatterers and similar deterministic multipath components if they are nearby. According to [36], [39], [44], [45], the channel parameters of

these two nearby UE [1] are similar and the magnitude of the CSI in the angular domain is correlated. However, no correlation exists in the CSI phase. To exploit the shared common sparsity structures in the CSI, the sparse vector of CSI is divided into two parts (as in [39]): the commonly shared and the individual sparse representation vectors. Inspired by this work, we partition the CSI magnitude $|\mathbf{H}_i|$ information of the i-th UE into two parts: shared by nearby UE and owned by individual UE.

To the best of our knowledge, the exploitation of the correlation of the CSI magnitude in the angular domain has not been investigated in the DL-based CSI feedback. To exploit this characteristic, the autoencoder-based NN framework should be modified. However, we make no modifications at the encoder framework of the UE, that is, we maintain the NN architectures and the parameter number because the UE has limited computation power and storage capability. We only modify the NN framework at the BS. The decoder at the BS has two modules: shared and individual modules that recover the vectors containing shared and individual information, respectively.

For the case of two nearby UE, the BS has three decoder modules, as illustrated in Fig. 2. Given that the shared information of nearby UE is the same, the first decoder module can be used for the two nearby UE. Specifically, $|\hat{\mathbf{H}}_1|_{\mathrm{v}}$ and $|\hat{\mathbf{H}}_2|_{\mathrm{v}}$ are the outputs of two individual decoders, and $|\hat{\mathbf{H}}|_{\mathrm{s}}$ is the output of the shared decoder

$$|\hat{\mathbf{H}}_1|_{\mathrm{v}} = f_{\mathrm{de1}}\Big(\mathcal{D}(\mathcal{Q}(f_{\mathrm{en1}}(|\mathbf{H}_1|, \Theta_{\mathrm{en1}}))), \Theta_{\mathrm{de1}}\Big) \tag{8}$$

$$|\hat{\mathbf{H}}_2|_{\mathrm{v}} = f_{\mathrm{de2}}\Big(\mathcal{D}(\mathcal{Q}(f_{\mathrm{en2}}(|\mathbf{H}_2|, \Theta_{\mathrm{en2}}))), \Theta_{\mathrm{de2}}\Big) \tag{9}$$

$$|\hat{\mathbf{H}}|_{\mathrm{s}} = f_{\mathrm{co-de}}\Big(\mathcal{D}(\mathcal{Q}(f_{\mathrm{en1}}(|\mathbf{H}_1|, \Theta_{\mathrm{en1}}))), \mathcal{D}(\mathcal{Q}(f_{\mathrm{en2}}(|\mathbf{H}_2|, \Theta_{\mathrm{en2}}))), \Theta_{\mathrm{co-de}}\Big) \tag{10}$$

where $f_{\mathrm{en1}}(\cdot)$ and $f_{\mathrm{en2}}(\cdot)$ are the encoder modules at the two UE; Θ_{en1} and Θ_{en2} are the NN parameters of the two encoder modules; Θ_{de1} and Θ_{de2} represent the parameters of the two individual decoders, that is., $f_{\mathrm{de1}}(\cdot)$ and $f_{\mathrm{de2}}(\cdot)$;, and $\Theta_{\mathrm{co-de}}$ is the parameter of shared decoder $f_{\mathrm{co-de}}(\cdot)$. Subsequently, two modules combine the information shared by nearby UE and owned by individual UE by using FC layers as

$$|\hat{\mathbf{H}}_1| = f_{\mathrm{com1}}(|\hat{\mathbf{H}}_1|_{\mathrm{v}}, |\hat{\mathbf{H}}|_{\mathrm{s}}, \Theta_{\mathrm{com1}}), \tag{11}$$

[1]In this paper, we just consider the scenario of two nearby UE. There will be more performance gain if more nearby UE are considered. Our strategy developed there can be also extended to the scenario of more nearby UE.

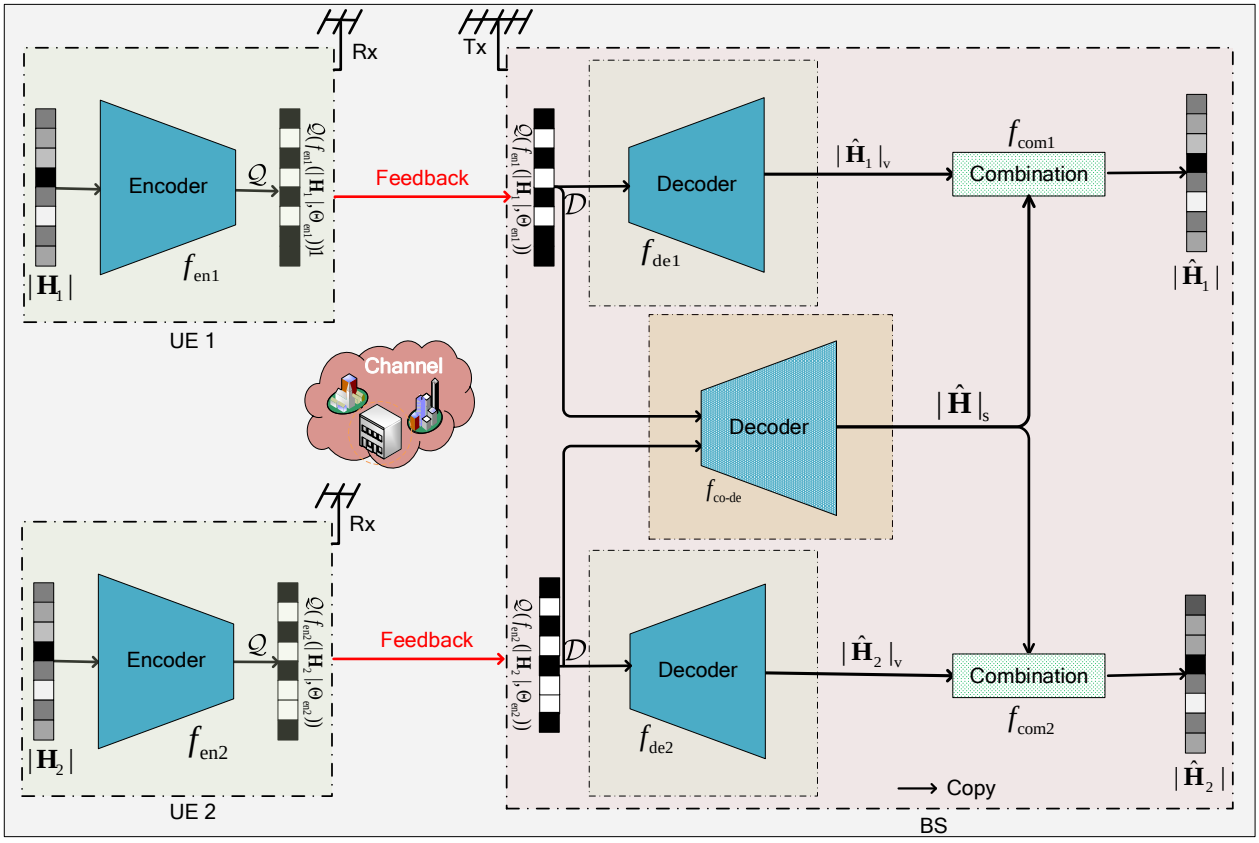


Fig. 2. Illustration of the cooperative NN framework for two nearby UE that includes two encoder and three decoders, that is, two individual decoders and one shared decoder. The encoders at the UE generate the bitstream and then feedback the bitstreams via the uplink channel. The decoders at the BS recover the shared CSI information and individual information from the feedback bitstreams by using the FC layers.

$$|\hat{\mathbf{H}}_2| = f_{\text{com2}}(|\hat{\mathbf{H}}_2|_{\text{v}}, |\hat{\mathbf{H}}|_{\text{s}}, \Theta_{\text{com2}}), \tag{12}$$

where $f_{\text{com1}}(\cdot)$ and $f_{\text{com2}}(\cdot)$ are the combination modules at the two nearby UE; Θ_{com1} and Θ_{com2} are the NN parameters of these two modules; and $|\hat{\mathbf{H}}_1|$ and $|\hat{\mathbf{H}}_2|$ are the final reconstructed magnitude matrices of the CSI.

Fig. 2 shows the entire cooperative CSI feedback framework, which consists of the encoders at the UE and the decoders at the BS. Details of the encoder and decoder modules are provided subsequently. At the UE, the two encoders generate the bits via the FC layers and the binary (or quantization) layer. Once the bitstreams are generated, they feed back the bits via the feedback transmission. Then, the shared decoder at the BS recovers the shared CSI information from the feedback of the two nearby UE. At the same time, the individual decoders recover the individual

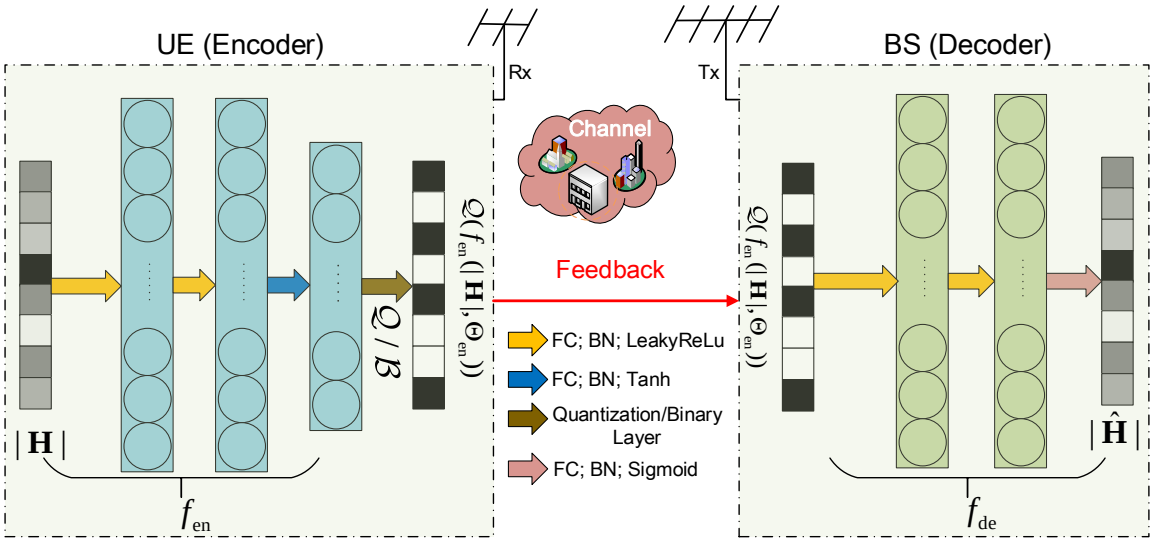


Fig. 3. Illustration of the NN architecture, which includes the encoder and the decoder. The encoder at the UE generates the bitstream by using the FC layers and the quantization or binary layer and then feeds back the bitstreams via the uplink channel. The decoder at the BS recovers the CSI from the feedback bitstreams by using the FC layers.

CSI information for the two UE. The final CSI magnitude matrix is the combination of the recovered shared and individual CSI magnitude information.

The training process of the two nearby UE should be performed together, thereby enabling CoCsiNet to extract and recover the shared CSI information. Therefore, we directly train the proposed CoCsiNet by using a one-step end-to-end learning approach. Training loss function $\mathrm{MSE}_{\mathrm{CoCsiNet}}$ is the overall MSE of the two UE as

$$\mathrm{MSE}_{\mathrm{CoCsiNet}} = \left\| |\mathbf{H}_1| - |\hat{\mathbf{H}}_1| \right\|_2^2 + \left\| |\mathbf{H}_2| - |\hat{\mathbf{H}}_2| \right\|_2^2 . \tag{13}$$

Given that the encoder of the CoCsiNet at the UE is the same as many existing feedback frameworks without UE cooperation, the NN parameter number and floating-point operations do not increase. Although an extra shared decoder exists at the BS, which increases the NN complexity, little effect can be exerted on the running speed of the CSI reconstruction process because the computation power is not limited at the BS, and three decoders can run parallelly.

B. NN Modules in the CoCsiNet

In this part, we first introduce the baseline NN architecture used in this work. Then, we describe the binarization operation, which generates the bitstreams and can be regarded as an extension of the one-bit quantization. Finally, we propose a phase strategy called MDPF.

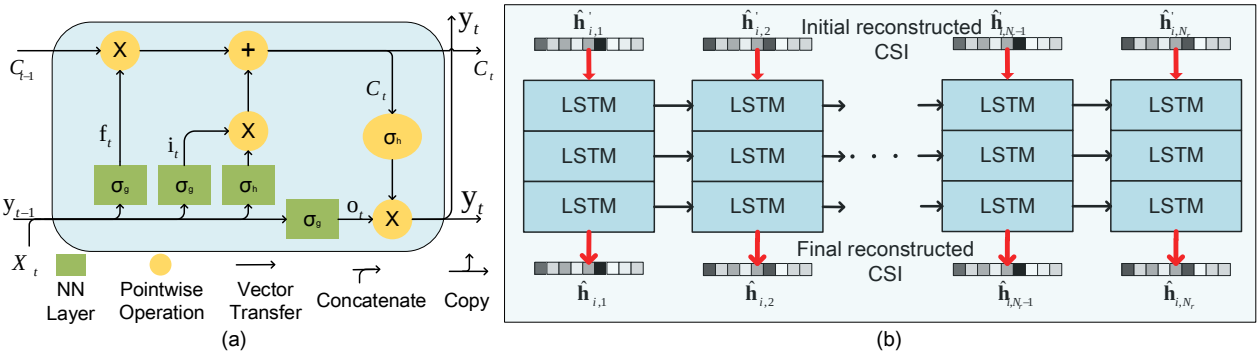


Fig. 4. (a) Architecture of the LSTM module. (b) Process refinement of the initial reconstruction. Once the FC layers output the initial reconstructed CSI $\hat{\mathbf{H}}_i^{'} = [\hat{\mathbf{h}}_{i,1}^{'}, \hat{\mathbf{h}}_{i,2}^{'}, \cdots, \hat{\mathbf{h}}_{i,N_r}^{'}]^{\mathrm{T}} \in \mathbb{R}^{N_r \times N_t}$ of the i-th UE is divided into N_r vectors, and each vector is then input to three stacked LSTM layers. The outputs of N_r LSTM modules are concatenated to generate the final reconstructed $\hat{\mathbf{H}}_i$.

1) Baseline NN Architecture: Fig. 3 shows the NN architecture, including an encoder and a decoder. The encoder at the UE is composed of three FC layers followed by three batch normalization (BN) layers. The activation function of the first two BN layers is the LeakyReLu function, which assigns a non-zero slope to the negative part compared with the standard ReLu activation function, as ,

$$\text{LeakyReLu}(x) = \begin{cases} x & x \geq 0 \\ ax & x < 0 \end{cases}, \tag{14}$$

where a is a small number. The activation function of the last BN layers is a tanh function, which aims to generate the numbers in the continuous interval $[-1, 1]$ as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{15}$$

The last layer at the encoder is the quantization or the previously mentioned binary layer. The decoder at the BS also contains three FC layers [2]. Unlike the NN layers at the encoder, the neuron number of the last FC layer here is $N_r \times N_t$, and the last BN layer employs the Sigmoid function to normalize the outputs into $[0, 1]$.

If the UE is equipped with multiple antennas, then an LSTM architecture is added to the last layers at the decoder to exploit the correlation among different antennas at the same UE,

[2]If we adopt the quantization layer, the dequantizer is necessary but we here do not draw this part.

as in Fig. 4 [3]. The extra LSTM at the decoder can well reconstruct the CSI because this NN architecture has inherent memory cells and can keep the previously extracted information for a long period for later reconstruction. The LSTM architecture is illustrated in Fig. 4(a). Further details about the LSTM module can be found in [46].

In [23], after the FC layers at the decoder, several convolutional layer-based RefineNet modules are added to refine the initial reconstructed CSI. Here, we can also regard the LSTM layers as a refinement module that refines the initial output of the baseline decoder by using the antenna correlation. Specifically, once the baseline decoder outputs the initial reconstructed CSI $\hat{\mathbf{H}}_i^{'} = [\hat{\mathbf{h}}_{i,1}^{'}, \hat{\mathbf{h}}_{i,2}^{'}, \cdots, \hat{\mathbf{h}}_{i,N_r}^{'}]^{\mathrm{T}} \in \mathbb{R}^{N_r \times N_t}$ of the i-th UE is divided into N_r vectors, and each vector is then used as an input to the three stacked LSTM layers. The outputs of N_r LSTM modules are concatenated to generate the final reconstructed $\hat{\mathbf{H}}_i$. The entire NN is trained via an end-to-end approach, and the loss function used in the CSI magnitude feedback is the MSE function.

2) Binarization Operation: Conventional DL-based methods, which can be regarded as DL-based CS algorithms, have two main operations[4], namely, compression and quantization. From the perspective of the machine learning domain, the first operation aims to generate the latent represent vectors of the original CSI or extract the features of the original CSI. The quantization operation quantizes the features to facilitate feedback overhead.

Furthermore, the binary representation can be used to generate the bitstreams [33]. To an extent, this binary presentation-based CSI feedback can be regarded as a codebook-based data-driven feedback strategy [47]. Unlike the previous methods on codebook design, we can regard the NN-based encoder, including the binary layer, as the module that generates the codebook indices, which are composed of a set of $\{-1, 1\}$. Correspondingly, the NN-based decoder recovers the CSI by using the codebook indices. According to [48], the binarization operation contains two main steps.

(1): The neuron number of the last FC layer at the encoder is set the same as the desired length of the feedback bits. To generate the numbers in the continuous interval $[-1, 1]$, the activation function of the BN [49] following the last FC layer is $\tanh(\cdot)$.

[3]Obviously, if a LSTM module is added to the encoder part, the encoder will be much more efficient and extract more features. However, considering the complexity of the LSTM module and the limited computation power at the UE, we here do not add the LSTM module at the encoder.

[4]Similar to [31], we here do not consider the entropy encoding because this operation is lossless.

(2): The real-valued representation is then taken as the input to generate a discrete output in the $\{-1, 1\}$ set by binarization function $b(\cdot)$ of $x \in [-1, 1]$, which is defined as,

$$b(x) = x + \epsilon, \tag{16}$$

$$\epsilon = \begin{cases} 1 - x & \text{with probability } \frac{1+x}{2}, \\ -x - 1 & \text{with probability } \frac{1-x}{2}, \end{cases} \tag{17}$$

where ϵ denotes the quantization noise.

Given that the second step is nondifferentiable, the gradient of this step is set as one to back-propagate gradients through this operation, that is, binary layer $\mathcal{B}$.

3) NN for Phase Feedback: Unlike the CSI magnitude, which is sparse due to limited paths, the phase matrix of the CSI is not sparse at all. If we feed back the phase information by using the same method as magnitude feedback, then too much useless information must be fed back. That is, the feedback bits corresponding to the CSI with a small magnitude have little useful information and waste resource. If we can drop this information during the feedback process, the feedback overhead will be efficiently exploited, and the feedback performance can be greatly improved.

In [26], a magnitude dependent phase quantization strategy is proposed, where CSI coefficients with large magnitudes adopt finer phase quantization and vice versa. That is, the more important the CSI phase is, the more bits can be allocated. In [50], an extra NN is introduced to allocate the quantization bits. Inspired by the above work, we adopt a magnitude dependent strategy for the CSI phase feedback. Unlike the previous work focusing on the allocation of quantization bits, we make the NN focus on coefficients with large magnitudes by modifying the loss function.

In the above-mentioned magnitude feedback, the loss function is the naive MSE function. The MSE loss function equally treats each coefficient in the matrix and narrows the gap between the ground truth and the prediction. That is, this naive loss function does not consider the importance of the phase coefficient. Therefore, as illustrated in Fig. 5, we modify the MSE loss function by introducing important information, that is, the magnitude information, as

$$MSE_{\text{phase}} = \|(\measuredangle\mathbf{H} - \measuredangle\hat{\mathbf{H}}) \odot |\mathbf{H}|\|_2^2, \tag{18}$$

where $\measuredangle\mathbf{H}$ denotes the original phase of the CSI, $\measuredangle\hat{\mathbf{H}}$ is the predicted phase, $|\mathbf{H}|$ is the corresponding CSI magnitude, and $\odot$ represents the Hadamard product.

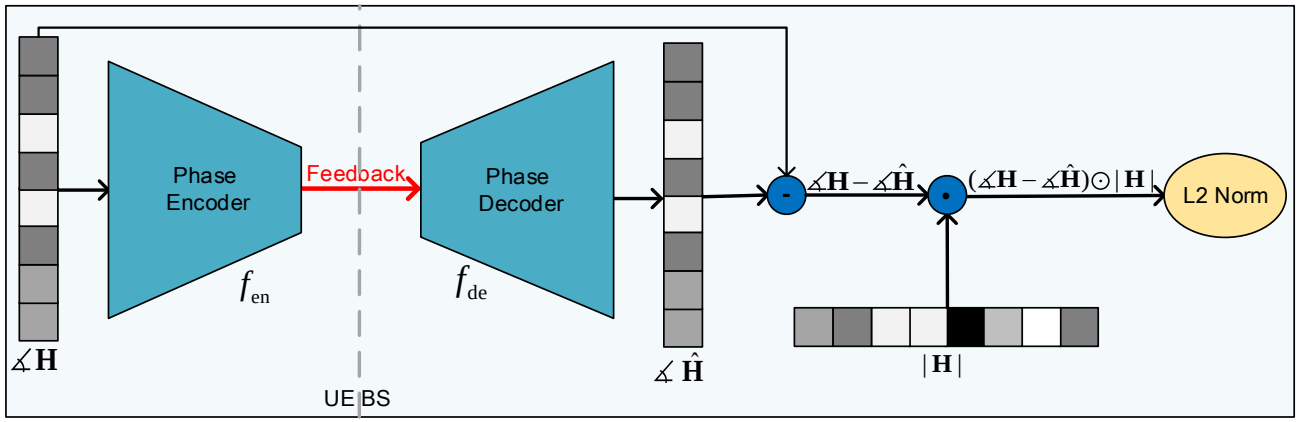


Fig. 5. Illustration of MDPF. The final loss function of the CSI phase feedback NNs is the weighted MSE function dependent on the CSI magnitude.

During the parameter update, large errors have much influence on MSE loss because MSE squares the errors before they are averaged [51]. Given that the importance of the phase is decided by the corresponding CSI magnitude, the larger the magnitude, the larger the MSE. By adding the magnitude information to the loss function of the phase feedback, the parameter update can favor the phase coefficients with large magnitudes. The encoder, including the binary layer and the decoder, are the same as those used in the magnitude feedback.

IV. Simulation Results and Discussions

In this section, we conduct a numerical simulation to evaluate the feedback performance of our proposed methods. We first introduce the simulation setting, including the channel parameters, the NN architecture details, and the NN training details. Then, we compare the quantization and binarization performance. Numerical simulation is conducted to demonstrate the superiority of the proposed methods in Section III, namely, the extra LSTM and the MDPF strategy. Finally, we compare the proposed cooperation-based NNs, that is, CoCsiNet, with the NNs, ignoring the CSI correlation between the nearby UE. According to [31], we also attempt to obtain insight into the cooperative CSI feedback via NN parameter visualization.

A. Parameter Setting

1) Channel: We generate channels of the urban microcell by using the 3GPP spatial channel model [42]. Following the setting in [39], the frequency of the downlink is 2.17 GHz. The inter-antenna spacing is half wavelength, that is, $d = c/(2f_0)$, where c is the light speed and $f_0 = 2$

GHz in carrier frequency. $N_c = 3, 4, 5, 6$ random scattering clusters (paths) ranging from $-\pi/2$ to $\pi/2$ exist[5]. The number of BS antennas is 64 or 256, and the number of UE antennas is 1 or 4. Given that we assume the UE are nearby, such as the stadium scenario, the nearby UE's CSI shares the most scattering clusters and has similar path loss. Specifically, we first randomly set the channel parameters of the first UE, namely, AoA, AoD, and gain. Then, the channel parameters of the second UE are generated by adding small random value to those of the first UE.

We generate 100,000 pairs of CSI samples by using MATLAB and randomly divide them into training, validation, and test datasets, with 70%, 10%, and 20% samples, respectively. During the training, the first dataset is used to update the NN parameters and the second is used to avoid over-fitting. Furthermore, we here bit per dimension (BPD) to describe the feedback overhead, which is widely used in the image compression domain and defined as

$$\text{BPD} = \frac{N_{\text{bits}}}{N_{\text{r}} N_{\text{t}}}. \tag{19}$$

2) NN Architecture: In Section III-B1, we briefly introduce the NN architecture. In this part, we provide the details of the NN architecture and the training. Given that our main contribution is not designing NN architectures for CSI feedback, we only use the standard FC, BN, and LSTM layers. The UE and the BS have three FC layers. Considering sufficient computation power at the BS, the FC layers at the BS are twice wider than those at the UE. Table I shows the parameters of the proposed NNs [6], including the neuron numbers of the FC layers and the activation function.

Simulation runs on TensorLayer 1.11.0 and is conducted on one NVIDIA DGX-1 station. The batch size is set to 200, the learning rate is 0.001, and the epoch is 1,000. During the training, the adaptive moment estimation optimizer is used to optimize the loss function.

B. Performance of NN-based CSI Magnitude Feedback

In this part, we evaluate the performance of the NN-based CSI magnitude feedback. We first compare the two main methods of generating bitstreams, that is, quantization and binary

[5]Different from the channel estimation in [39], which considers the sub-paths of each scattering cluster, we here do not take it into consideration, because the feedback CSI is eatimated at the UE, whose path resolution is limited.

[6]If the BPD is changed, only the last FC layer has been changed for the encoder at the UE. Therefore, the multiple-rate CSI feedback strategy can be easily extended to this work.

TABLE I

THE DETAIL ARCHITECTURE OF THE PROPOSED NNS

	Layer name	Output size	Activation function
Encoder (UE)	Input Layer	$N_r \times N_t$	None
	Reshape1	$N_r N_t \times 1$	None
	FC1+BN1	$2N_r N_t \times 1$	LeakyReLu
	FC2+BN2	$2N_r N_t \times 1$	LeakyReLu
	FC3+BN3	N_{bits}/B [*] or $N_{\text{bits}} \times 1$	Tanh
	Quantization or Binary layer	$N_{\text{bits}} \times 1$	None
Decoder (BS)	FC4+BN4	$4N_r N_t \times 1$	LeakyReLu
	FC5+BN5	$4N_r N_t \times 1$	LeakyReLu
	FC6+BN6	$N_r N_t \times 1$	Sigmoid / Tanh [**]
	Reshape2	$N_r \times N_t$	None
	LSTM1	$N_r \times N_t$	-
	LSTM2	$N_r \times N_t$	-
	LSTM3	$N_r \times N_t$	-

[*] B represents the number of quantization bits.

[**] If the decoder is used to recover CSI magnitude, we choose the Sigmoid function. If the decoder is used to recover CSI phase, we choose the Tanh function.

representation. We consider the single-user scenario where the BS is equipped with 256 antennas, and the UE has a single antenna. We set the number of the quantization bits $B = 4$, which is proved the best through simulation in [28]. The BPD-NMSE tradeoffs achieved by the quantization and binarization are plotted in Fig. 6.

The figure shows that with the same BPD, the feedback accuracy decreases as the channel path number increases. For example, when the BPD is 0.1, and quantization is adopted rather than binarization, the NMSEs of the 3-path, 4-path, 5-path, and 6-path cases are -12.18, -6.90, -4.82, and -4.02 dB, respectively. The more paths exist, the more information is provided in the CSI matrix. Therefore, the larger the channel path number is, the more the feedback bits must be to maintain the CSI accuracy. Therefore, the feedback bit number must be adjusted according to the deployment environments.

To an extent, the binarization operation can be regarded as a special case of quantization with $B = 1$. Unlike the general quantization, as in (17), the binarization operation introduces

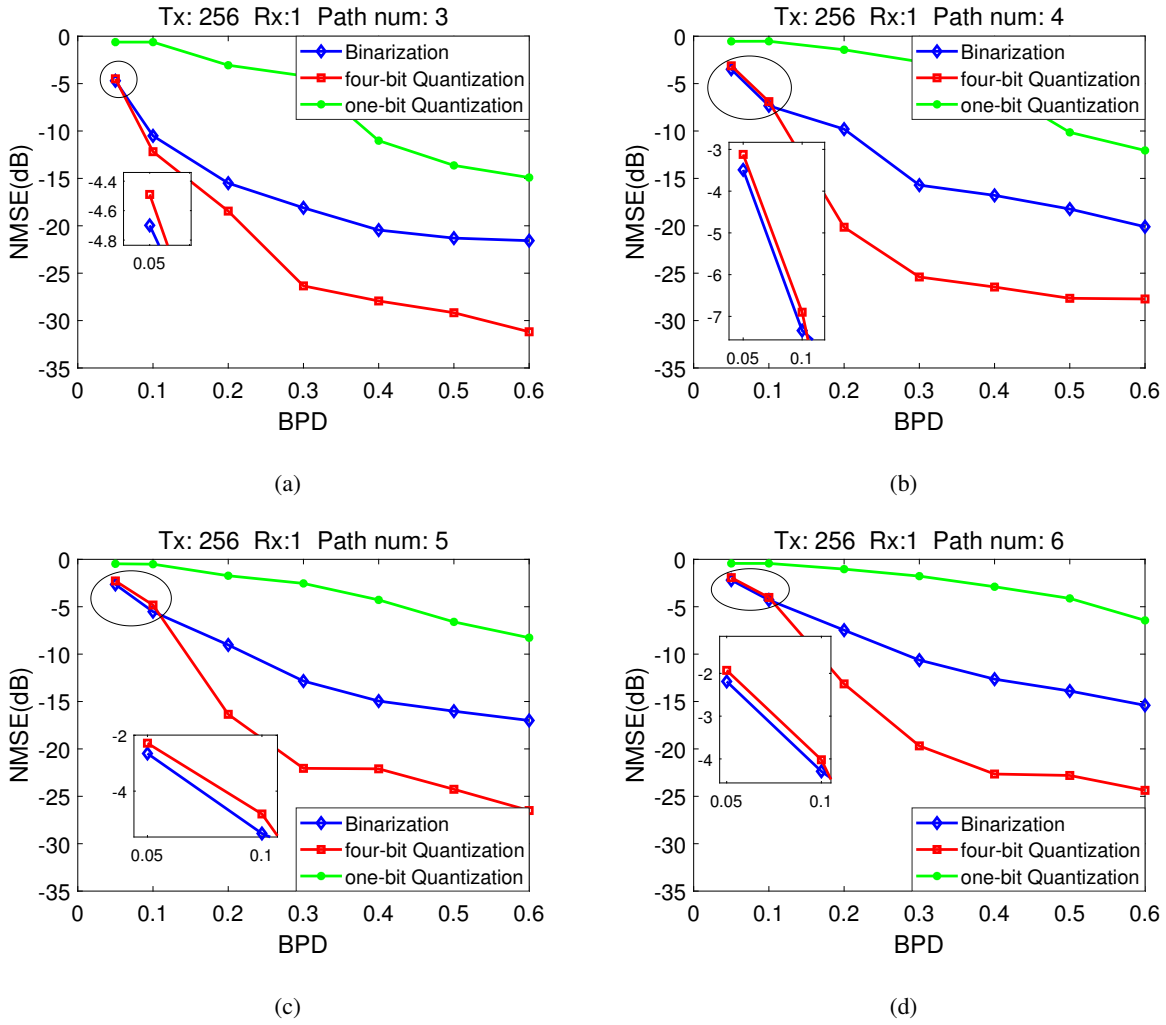


Fig. 6. NMSE (dB) performance comparison of CSI magnitude feedback between four and one-bit quantization, binarization operation under different BPD. The BS is equipped with 256 antennas and the UE has a single antenna. The channel path number is $N_c = 3, 4, 5, 6$.

quantization noise ϵ with a zero-mean property, thereby making $b(x)$ an unbiased estimator for soft value x, that is, $\mathbb{E}_\epsilon[b(x)|x] = x$ [33]. We also compare the binarization operation and the common one-bit quantization in Fig. 6. The figure shows that the former outperforms the latter by a large margin, indicating the importance of zero-mean ϵ.

The performance of the quantization operation is better than that of the binarization operation for most scenarios, except when the BPD is very low, that is, the feedback overhead is extremely limited. As pointed out in [31], if the feedback bit number is fixed, then feedback accuracy is more sensitive to the quantization distortion then compression ratio $\gamma = \text{BPD}/\text{B}$. Therefore, in most scenarios, the quantization operation performs better than the binarization operation,

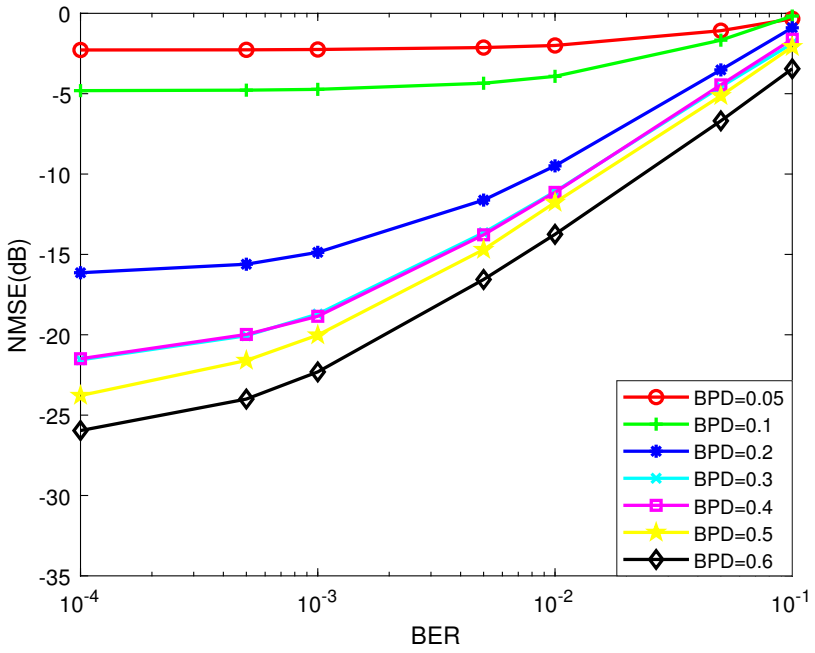


Fig. 7. NMSE (dB) performance of CSI feedback with imperfect uplink transmission. We consider the scenario where the BS is equipped with 256 antennas, the UE is with a single antenna, and the channel path number is $N_c = 5$. The BER is from 0.1 to 0.0001.

which is an improved one-bit quantization. For extremely low overhead, such as BPD = 0.05 [7], feedback accuracy is less sensitive to quantization distortion but more sensitive to compression ratio γ.

Fig. 7 illustrates the robustness of the proposed NNs by using the quantization operation to the imperfect uplink transmission. We consider the scenario where the BS is equipped with 256 antennas, the UE has a single antenna, and the channel path number is $N_c = 5$. Unlike most works that directly add noise to the codewords and use signal-to-noise ratio to describe the uplink condition [31], [52], we use the bit-error ratio (BER) to denote the feedback condition because we are now focusing on the digital CSI feedback, rather than the analog one [53]. During the test phase, the errors are introduced by the bit-wise exclusive-or between feedback bits and 1 at a predefined probability, that is, BER. Obviously, the NN performance worsens as the BER increases. However, an interesting phenomenon is that the feedback with a high BPD is more sensitive to the BER, that is, the uplink transmission condition, than that with

[7]Here, the compression ratio γ is $1/80$ for the quantization operation with B = 4.

a low BPD. Fig. 7 indicates that for the feedback with a low BPD, namely, 0.05 and 0.1, no degradation in feedback accuracy is observed if the BER is lower than 0.005. For the feedback with a high BPD, namely, 0.4, 0.5, and 0.6, even the BER is very low (e.g., 0.0001), performance degradation is still observed. Therefore, the feedback with a high BPD is more sensitive to the uplink condition than that with a low BPD.

If the UE is equipped with multiple antennas, the extra LSTM module is added after the FC layers of the decoder to well extract the correlation between the nearby antennas at the UE. Here, we consider a scenario where the BS and the UE are equipped with 64 and 4 antennas, respectively. Table II compares the feedback performance of the fully FC NNs and the NNs with LSTM modules when the BPD is set as 0.3. The table displays the improvement of the feedback accuracy for different path numbers due to the extra LSTM modules.

C. Performance of MDPF NNs

In this part, we evaluate the performance of the proposed MDPF NNs, where the magnitude information is added to the loss function of the phase feedback. We consider the scenario that the BS is equipped with 256 antennas, the UE is equipped with a single antenna, and the channel path number is $N_c = 3$. We directly evaluate the accuracy of the recovered CSI magnitude. However, we cannot directly evaluate the phase accuracy because we drop certain secondary information. Assuming that the CSI magnitude has been perfectly fed back, we obtain the complex CSI by combining the magnitude and the phase together, and the complex CSI is compared with the original complex one at the UE.

Fig. 8 plots the BPD-NMSE tradeoffs achieved by the MDPF and the naive phase feedback method. When the BPD is very low, the NMSE of the original phase feedback method exceeds

TABLE II

NMSE (dB) PERFORMANCE COMPARISON BETWEEN THE FCNN AND THE LSTM.

N_c	3	4	5	6
LSTM *	-24.84	-20.98	-20.82	-18.94
FCNN **	-21.95	-18.75	-18.31	-16.52

* LSTM represents the NNs with LSTM modules.

** FCNN represents the fully FC NNs.

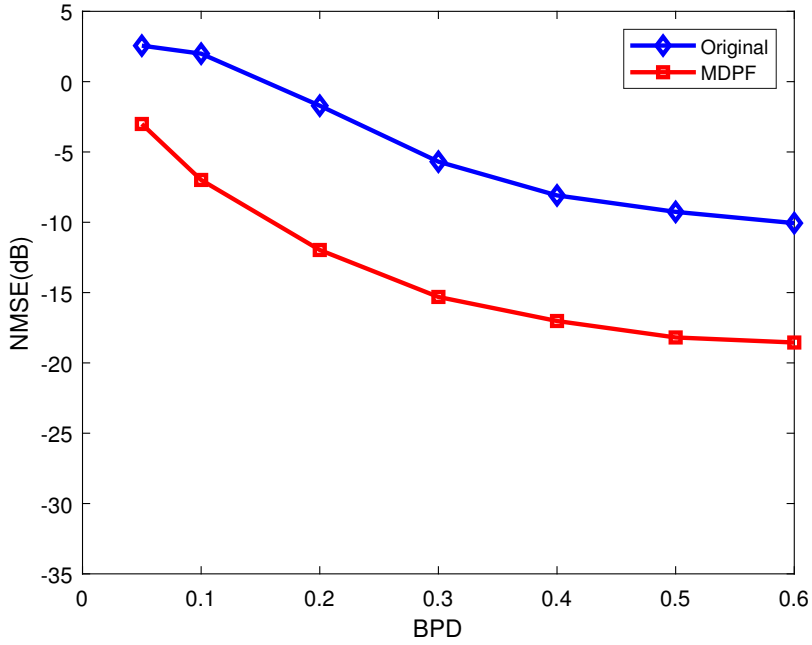


Fig. 8. NMSE (dB) performance of CSI phase feedback. The BS is equipped with 256 antennas, the UE is equipped with a single antenna, and the channel path number is $N_c = 3$.

0 dB, which means that little useful information about the CSI phase is fed back because the NNs do not know which information is significant and only attempt to feed back all the phase information and require many feedback bits. A 8.93 dB improvement is achieved when the BPD is 0.5. Through the weighted loss function proposed in Section III-B3, the secondary information is dropped, and the NNs focus on feeding back significant information, that is, the phase whose corresponding magnitude is large.

D. Performance of Cooperative CSI Feedback

In this part, we evaluate the performance of the proposed cooperative CSI feedback NNs, CoCsiNet, and explain the mechanism via NN parameter visualization. Fig. 9 compares the performance of the CoCsiNet and the NNs without cooperation when the BS has 64 antennas, the UE has 4 antennas, and the channel path number is $N_c = 3, 4, 5, 6$.

When the BPD is low, the proposed CoCsiNet outperforms the original one by a large margin because the feedback information of the single UE is extremely few to accurately recover the CSI for the decoder at the BS when the BPD is low, that is, the feedback overhead is extremely

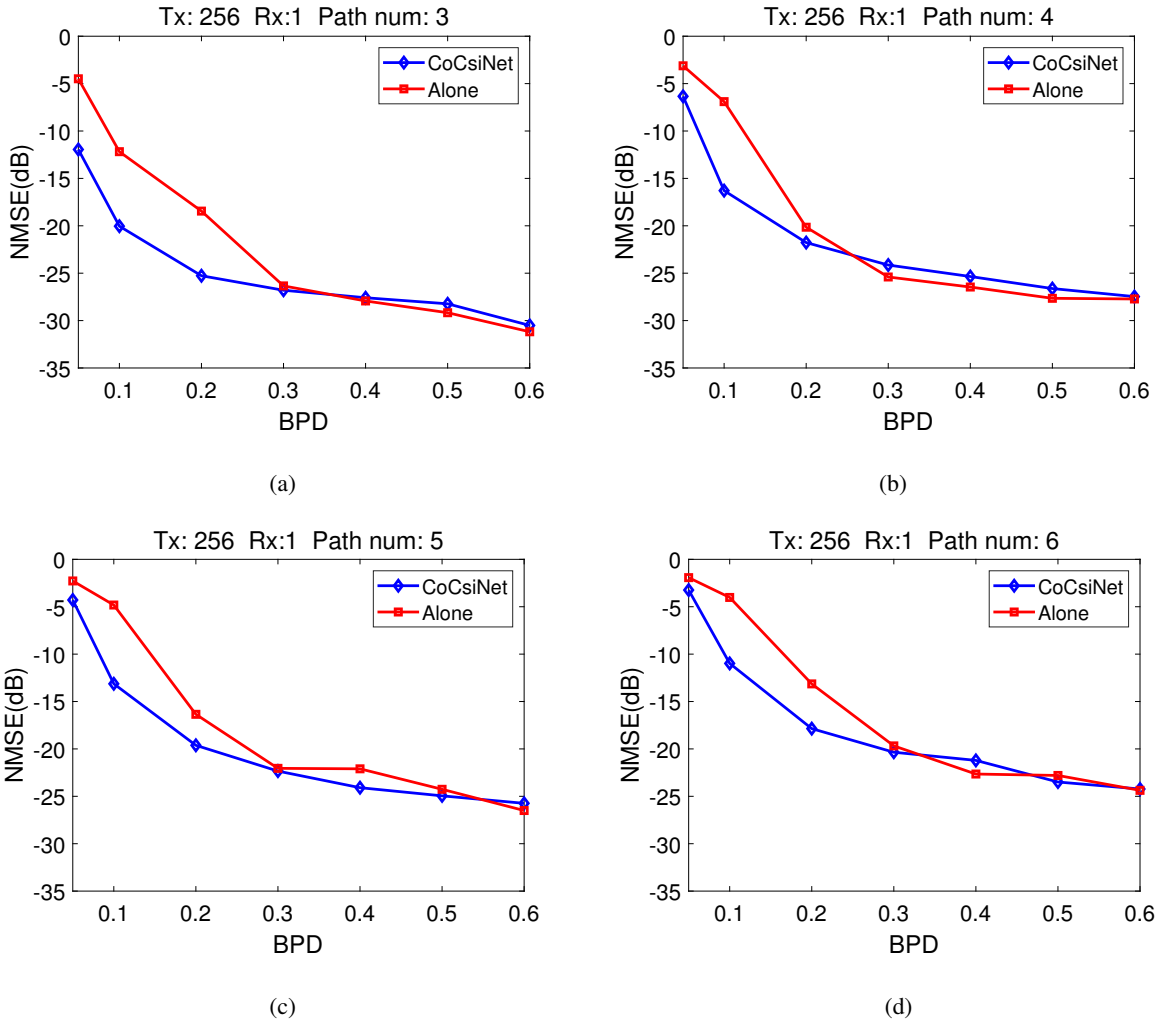


Fig. 9. NMSE (dB) performance comparison between the CoCsiNet and the NNs without cooperation when there are 256 antennas at the BS and a single antenna at the UE, and the channel path number is $N_c = 3, 4, 5, 6$.

constrained. Therefore, in this scenario, the shared information does not need to be fed back repeatedly if the two UE can feed back the CSI via cooperation. However, when the BPD exceeds 0.3, a small gap exists between the two methods, and the original one sometimes even performs a little better than CoCsiNet because the feedback information of the single UE is enough to accurately recover the CSI for the decoder at the BS.

Fig. 10 compares the feedback performance of the CoCsiNet with LSTM modules and the NNs without cooperation when the BS has 64 antennas and the UE has 4 antennas at the UE, respectively. The BPD is 0.05 or 0.10, and the channel path number is $N_c = 3, 4, 5, 6$. For the scenario with low BPD, the CoCsiNet outperforms the NNs without cooperation by a large

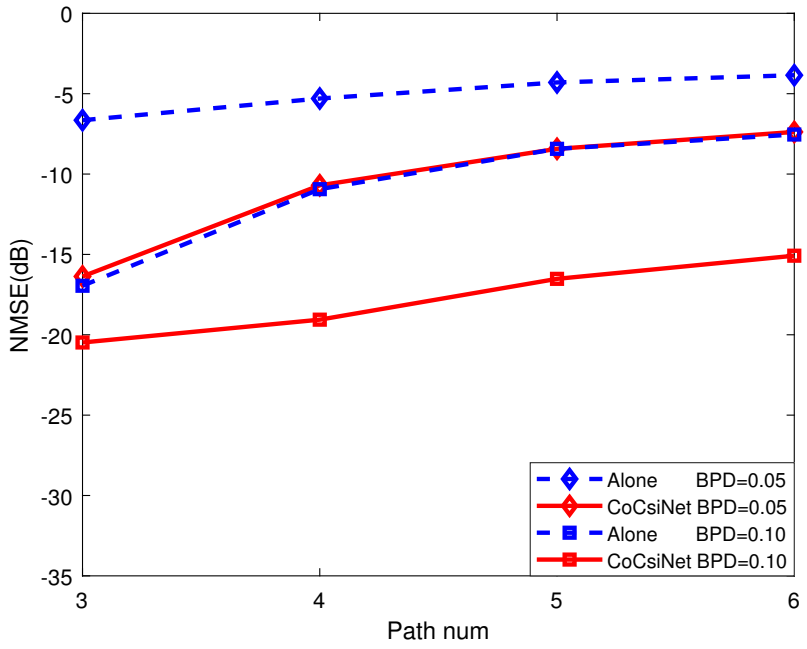
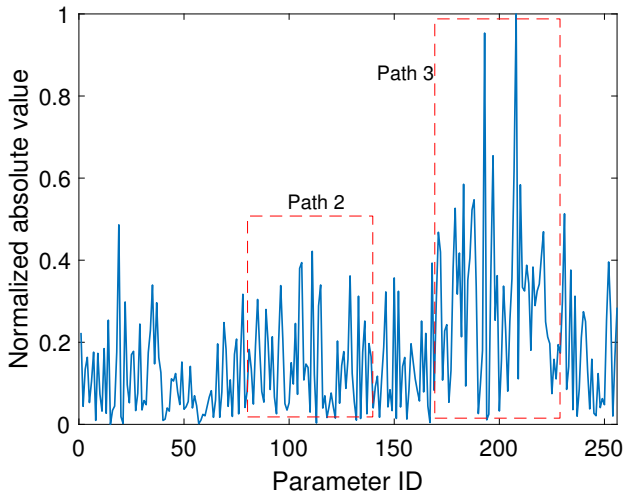


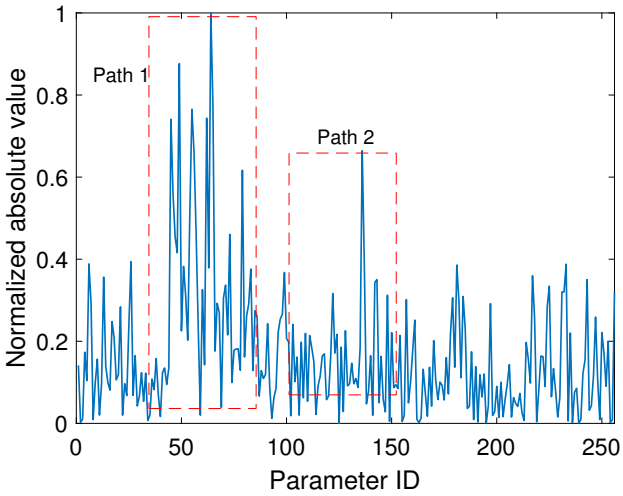
Fig. 10. NMSE (dB) performance comparison between the CoCsiNet with LSTM modules and the NNs without cooperation when the BS has 64 antennas and the UE has 4 antennas, respectively; the BPD is 0.05 or 0.10; and the channel path number is $N_c = 3, 4, 5, 6$.

margin. The figure shows when the BPD is 0.05, the gap between the CoCsiNet and the NNs without cooperation decreases as the channel path number increases. By contrast, when the BPD is 0.10, the gap increases as the channel path number increases. For extremely low BPDs, such as 0.05, the feedback bits are insufficient for the CoCsiNet. Therefore, if many channel paths exist and further information is needed to be fed back, the CoCsiNet cannot handle this scenario, thereby narrowing the performance gap.

Fig. 11 is the visualization of the normalized absolute value of the first FC layer at the encoders for the CoCsiNet. The BS is equipped with 256 antennas, and the UE is equipped with a single antenna. The BPD and the channel path number are 0.05 and 3, respectively. The normalized absolute value can represent the attention of the NNs. If this value is large, then the corresponding CSI coefficient can have much effect on the generated codewords, that is, its information can be fed back. In this figure, dotted rectangles circle the areas, including the channel paths. The first encoder pays the most attention to the third path, whereas the second encoder pays the most attention to the first path. Therefore, they feed back the information of



(a) Encoder parameter visualization for UE 1



(b) Encoder parameter visualization for UE 2

Fig. 11. Visualization of the normalized absolute value of the first FC layer at the encoders for the CoCsiNet when the BPD and channel path number N_c are 0.05 and 3, respectively. The higher the absolute value, the greater the attention that the NNs must pay.

the different paths. At the BS, they share their feedback information via the proposed shared decoder and the feedback bits are not wasted to feed back the duplicate information. This situation explains why cooperation-based NNs can outperform NNs without cooperation when the BPD is low.

V. Conclusion

In this paper, we propose a DL-based cooperative CSI feedback and recovery framework, namely, CoCsiNet, to exploit the CSI correlation between the nearby UE in massive MIMO systems. The nearby UE's CSI information is divided into two parts: shared by nearby UE and owned by individual UE. Specifically, an extra decoder is added at the BS to recover shared information from the feedback bits of two nearby UE, whereas the individual information is recovered by the individual decoders of two UE. Then, these two kinds of information are combined at the BS by FC layers. Unlike existing cooperation mechanisms, the proposed CoCsiNet need not CSI sharing, that is, D2D communication. Moreover, we investigate two representative bit generation methods (i.e., quantization and binarization) and introduce an MDPF framework that introduces CSI magnitude information to the loss function of CSI phase feedback. The proposed CoCsiNet framework outperforms the NNs without cooperation by a large margin when the feedback overhead is extremely constrained. NN parameter visualization explains the cooperation mechanism. In the CoCsiNet, the encoders of two nearby UE cooperatively extracts the information of different channel paths, thereby reducing the overhead used to repeatedly feedback shared information.

References

[1] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key technologies for 5G wireless systems*. Cambridge university press, 2017.

[2] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[3] M. Latva-aho and K. Leppänen, "Key drivers and research challenges for 6G ubiquitous wireless intelligence," *University of Oulu, While Paper*, 2019, Available: http://urn.fi/urn:isbn:9789526223544.

[4] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[5] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, Mar. 2015.

[6] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[7] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[8] Z. Gao, L. Dai, S. Han, C. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.

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

[10] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.(NIPS)*, 2012, pp. 1097–1105.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[13] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, Nov. 2017.

[14] Z. Qin, H. Ye, G. Y. Li, and B. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.

[15] H. He, S. Jin, C. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 77–83, Oct. 2019.

[16] Z. Qin, X. Zhou, L. Zhang, Y. Gao, Y. Liang, and G. Y. Li, "20 years of evolution from cognitive to intelligent communications," *IEEE Trans. on Cogn. Commun. Netw.*, 2019, Early Access.

[17] T. OâĂŹShea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[18] H. Ye, L. Liang, G. Y. Li, and B. Juang, "Deep learning based end-to-end wireless communication systems with conditional GAN as unknown channel," *IEEE Trans. Wireless Commun.*, 2020, Early Access.

[19] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, Apr. 2019.

[20] Z. Liu, L. Zhang, and Z. Ding, "Overcoming the channel estimation barrier in massive MIMO communication systems," *arXiv preprint arXiv:1912.10573*, 2019.

[21] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[22] P. Jiang, T. Wang, B. Han, X. Gao, J. Zhang, C.-K. Wen, S. Jin, and G. Y. Li, "Artificial intelligence-aided OFDM receiver: Design and experimental results," *arXiv preprint arXiv:1812.06638*, 2018.

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

[24] T. Wang, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, Apr. 2019.

[25] C. Lu, W. Xu, H. Shen, J. Zhu, and K. Wang, "MIMO channel information feedback using deep recurrent network," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 188–191, Jan. 2019.

[26] Z. Liu, L. Zhang, and Z. Ding, "Exploiting bi-directional channel reciprocity in deep learning for low rate massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 889–892, Jun. 2019.

[27] Q. Cai, C. Dong, and K. Niu, "Attention model for massive MIMO CSI compression feedback and recovery," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–5.

[28] C. Lu, W. Xu, S. Jin, and K. Wang, "Bit-level optimized neural network for multi-antenna channel quantization," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2019.

[29] X. Li and H. Wu, "Spatio-temporal representation with deep neural recurrent network in MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2020, Early Access.

[30] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," *arXiv preprint arXiv:1910.14322*, 2019.

[31] J. Guo, C. Wen, S. Jin, and G. Y. Li, "Convolutional neural network based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," *IEEE Trans. Wireless Commun.*, 2020, Early Access.

[32] Q. Yang, M. B. Mashhadi, and D. GÃijndÃijz, "Deep convolutional compression for massive MIMO CSI feedback," in *Proc. IEEE Int. Conf. Mach. Learn. Signal Process. (MLSP)*, Oct. 2019, pp. 1–6.

[33] J. Jang, H. Lee, S. Hwang, H. Ren, and I. Lee, "Deep learning-based limited feedback designs for MIMO systems," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2019, Early Access.

[34] J. Guo, J. Wang, C. Wen, S. Jin, and G. Y. Li, "Compression and acceleration of neural networks for communications," *arXiv preprint arXiv:1907.13269*, 2019.

[35] Q. Liu, J. Guo, C. Wen, and S. Jin, "Adversarial attack on DL-based massive MIMO CSI feedback," *arXiv preprint arXiv:2002.09896*, 2020.

[36] F. Kaltenberger, D. Gesbert, R. Knopp, and M. Kountouris, "Correlation and capacity of measured multi-user MIMO channels," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2008, pp. 1–5.

[37] F. Kaltenberger, M. Kountouris, D. Gesbert, and R. Knopp, "On the trade-off between feedback and capacity in measured MU-MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4866–4875, Sep. 2009.

[38] S. Aghaeinezhadfirouzja, H. Liu, B. Xia, Q. Luo, and W. Guo, "Third dimension for measurement of multi user massive MIMO channels based on LTE advanced downlink," in *Proc. IEEE Int. Conf. Global Signal Process. (GlobalSIP)*, Dec. 2016, pp. 758–762.

[39] J. Dai, A. Liu, and V. K. N. Lau, "Joint channel estimation and user grouping for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 622–637, Feb. 2019.

[40] J. Chen, H. Yin, L. Cottatellucci, and D. Gesbert, "Feedback mechanisms for FDD massive mimo with D2D-based limited CSI sharing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5162–5175, Aug. 2017.

[41] J. Song, B. Lee, S. Noh, and J. Lee, "Limited feedback designs for machine-type communications exploiting user cooperation," *IEEE Access*, vol. 7, pp. 95 154–95 169, Jul. 2019.

[42] A. F. Molisch, A. Kuchar, J. Laurila, K. Hugl, and R. Schmalenberger, "Geometry-based directional model for mobile radio channelsâĂŤprinciples and implementation," *Eur. Trans. Telecommun.*, vol. 14, no. 4, pp. 351–359, 2003.

[43] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 40–58, May 2018.

[44] A. Algans, K. I. Pedersen, and P. E. Mogensen, "Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 523–531, Apr. 2002.

[45] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.

[46] K. Greff, R. K. Srivastava, J. KoutnÃ■k, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[47] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[48] T. Raiko, M. Berglund, G. Alain, and L. Dinh, "Techniques for learning binary stochastic feedforward neural networks," *arXiv preprint arXiv:1406.2989*, 2014.

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[50] Z. Liu, L. Zhang, and Z. Ding, "An efficient deep learning framework for low rate massive MIMO CSI reporting," *arXiv preprint arXiv:1912.10608*, 2019.

[51] J. Guo, H. Du, J. Zhu, T. Yan, and B. Qiu, "Relative location prediction in CT scan images using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 160, pp. 43–49, Jul. 2018.

[52] Y. Jang, G. Kong, M. Jung, S. Choi, and I. Kim, "Deep autoencoder based CSI feedback with feedback errors and feedback delay in FDD massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 833–836, Jun. 2019.

[53] M. B. Mashhadi, Q. Yang, and D. Gunduz, "CNN-based analog CSI feedback in FDD MIMO-OFDM systems," *arXiv preprint arXiv:1910.10428*, 2019.