

Multi-Scale and Multi-Channel Networks for CSI Feedback in Massive MIMO System

Bingyang Cheng, Jianing Zhao, Yu Hu

National Mobile Communications Research Laboratory, South University, Nanjing, China

Email: 220190736@seu.edu.cn, jnzha@seu.edu.cn, 220190737@seu.edu.cn

How to cite this paper: Cheng, B.Y., Zhao, J.N. and Hu, Y. (2021) Multi-Scale and Multi-Channel Networks for CSI Feedback in Massive MIMO System. *Journal of Computer and Communications*, 9, 132-141. <https://doi.org/10.4236/jcc.2021.910009>

Received: October 10, 2021

Accepted: October 26, 2021

Published: October 29, 2021

Abstract

In the frequency division duplex (FDD) mode of the massive MIMO system, the system needs to perform coding through channel state information (CSI) to obtain performance gains. However, the number of antennas of the base station has been greatly increased, resulting in a rapid increase in the overhead for the user terminal to feedback CSI to the base station. In this article, we propose a method based on multi-task CNN to achieve compression and reconstruction of channel state information through a multi-scale and multi-channel convolutional neural network. We also introduce a dynamic learning rate model to improve the accuracy of channel state information reconstruction. The simulation results show that compared with the original CsiNet and other work, the proposed CSI feedback network has better reconstruction performance.

Keywords

Deep Learning, Multi-Scale, Multi-Channel, Massive MIMO

1. Introduction

Massive multiple-input multiple-output (MIMO) technology was proposed in the 20th century and has become particularly important in the latest 5G wireless communication systems [1]. Massive MIMO system refers to the simultaneous provision of services to multiple users through base stations equipped with a large number of transmitting antennas on the same time-frequency resources [2]. In MIMO systems, increasing the number of antennas can improve channel capacity and transmission efficiency. However, the above benefits are based on the base station's ability to obtain accurate channel state information (CSI). In the uplink, the base station can accurately estimate the channel state information through the pilots sent by the user equipment [3]. For the downlink, the us-

er equipment needs to estimate the CSI and feed it back to the base station for precoding. However, due to the increase in the number of antennas, the size of the CSI matrix is also greatly increased [4]. If the traditional feedback method is used, the cost of the system will increase.

With the rise of deep learning, great progress has been made in computer vision, natural language processing, etc. Deep learning technology was introduced into the field of wireless communication. It is used to compress and reconstruct the CSI matrix. An auto-encoder network called CsiNet proposed by Wen and *et al.* [5] compresses and reconstructs CSI through a convolutional neural network, whose reconstruction performance is better than that of traditional compressed sensing methods (e.g. LASSO and TVAL3) [6] [7] [8]. In addition, they also proposed LSTM-CsiNet, a recurrent neural network that uses channel time correlation and has good reconstruction performance in the face of high compression ratios [9]. They also proposed a multi-rate CSI network in [10], which reduces the amount of parameters and a novel quantitative CSI feedback network is adopted. Based on the channel reciprocity, the uplink CSI information is used to reconstruct the downlink CSI. The work of literature [11] [12] reduces the impact of transmission delay in feedback. The above researches roughly show that the convolutional neural network has a better effect on CSI feedback processing. However, in these methods, the number of parameters is huge and computational complexity is relatively high. So we consider to use convolutional neural network as the infrastructure and use multi-scale and multi-channel convolutional neural network to improve the quality of CSI reconstruction. In addition, we introduce the dynamic learning rate model to obtain optimization. Our main work is listed below.

- We propose a new CSI compression recovery mechanism in FDD massive MIMO systems, which is called multi-scale and multi-channel convolutional CsiNet (MSMCNet). In this multi-task network, we set the number of the channels to 3 because of the balance of complexity and accuracy.
- We introduce a dynamic learning rate model to improve the robustness of the automatic encoder especially in case of high compression ratio. At the same time, we adopted block convolution and hole convolution to the convolution layer which can improve robustness in case of different sparsity matrix.

The rest of the article is arranged as follows. The Section 2 introduces the CSI feedback system model; the Section 3 presents the specific architecture of MSMCNet; the Section 4 shows the simulation results and analysis. Finally, the conclusion is drawn in Section 5.

2. System Model

We consider a simple single-cell downlink massive MIMO system. The system has $N_t \gg 1$ transmitting antennas at the base station and a single receiving antenna at the user equipment. Orthogonal frequency division multiplexing (OFDM)

with \tilde{N}_C orthogonal sub-carriers is adopted where the received signal on the n_{th} subcarrier can be derived as follows:

$$y_n = \tilde{h}_n^H v_n x_n + z_n \tag{1}$$

where $\tilde{h}_n \in \mathbb{C}^{N_r \times 1}$, $v_n \in \mathbb{C}^{N_r \times 1}$, $x_n \in \mathbb{C}$, $z_n \in \mathbb{C}$ respectively represent the channel vector, precoding vector, data bearing symbol and additive noise of the n th subcarrier. The downlink CSI matrix is a stack of subcarrier channel vectors [13]:

$$\tilde{\mathbf{H}} = [\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_{\tilde{N}_C}]^H \tag{2}$$

After that, two pre-processes are implemented on CSI:

1) Using two dimension (2D) discrete Fourier transform (DFT) to transform $\tilde{\mathbf{H}}$ into \mathbf{H} , \mathbf{H} is sparse in the angular delay domain.

2) In the delay domain, except for the first few non-zero columns, most of the elements in $\tilde{\mathbf{H}}$ are zero. Because the time delay between multipath arrivals lies within a limited period of time. Therefore, only the first N_c non-zero columns are retained for the processed \mathbf{H} . The size of the new CSI matrix obtained is $N_c \times N_r$, that is:

$$\mathbf{H} = \mathbf{F}_d \tilde{\mathbf{H}} \mathbf{F}_a^H \tag{3}$$

where F_d and F_a are the DFT matrices with dimension $N_c \times N_c$ and $N_r \times N_r$ [14]. For angular-delay domain channel matrix \mathbf{H} , only the first N_c row contains large values. We use \mathbf{H}_a to denote the first N_c rows of \mathbf{H} .

Although the number of the elements \mathbf{H}_a in is much lower than the initial $\tilde{\mathbf{H}}$, it is still too large for feedback. So we have to further compress \mathbf{H}_a before feedback. Traditional LASSO and AMP methods rely on the priori of the channel structure and cannot guarantee the recovery performance. Therefore, scholars introduces deep learning into feedback compression and use the black box model to obtain better reconstruction performance.

In this article, we use the encoder-decoder network for downlink CSI feedback to perform further compression. MSMCNet will compress \mathbf{H}_a into a feature vector \mathbf{V} according to a certain compression ratio. The user side then feeds back \mathbf{V} to the base station, and the decoder of the base station performs \mathbf{H}_a reconstruction through decompression. Finally, the original CSI matrix is restored through zero column padding and inverse DFT (IDFT).

$\tilde{\mathbf{H}}_a$ is reconstructed matrix:

$$\tilde{\mathbf{H}}_a = D(E(\mathbf{H}_a, \theta_e), \theta_d), \tag{4}$$

where E and D denote the encoder and the decoder of MSMCNet. θ_e and θ_d represent their network parameters. We propose to design and train θ_e and θ_d so that the distance between reconstructed matrix $\tilde{\mathbf{H}}_a$ and \mathbf{H}_a is minimized.

The work in the article only focuses on the feedback scheme, so we assume that the uplink feedback and downlink channel estimation are ideal. The article uses the COST2100 model to simulate the channel matrix \mathbf{H} for the massive

MIMO FDD system [15].

3. Design of MSMCNet

In the literature [5], CsiNet adopts the residual structure of RefineNet which has been proved to be effective in calculating CSI. CsiNet applies a fixed resolution which has a fixed convolution kernel to extract the features of the CSI matrix. However, the degree of CSI matrix sparsity shows different adaptability to different resolutions.

For example, if a CSI matrix has poorer sparsity, we should use a convolution kernel with a smaller kernel size to extract finer features. However, when the sparsity of the CSI matrix is very high, if we continue to use a smaller kernel size, it may result in a large blank area and cannot effectively extract its features. Therefore, in case of different CSI matrices, the size of the convolution kernel adopted should be different in order to adapt to different sparsity.

We introduce a multi-scale and multi-channel convolution kernel and propose a network called multi-scale multi-channel-Net (MSMCNet). The structure of MSMCNet is shown in **Figure 1** and **Figure 2**. MSMCNet consists of two parts: the encoder at the user and the decoder at the base station. The \mathbf{H}_a is $2 \times N_c \times N_r$, where 2 represents the real part and the imaginary part of the matrix. We set three channels which is a trade-off between complexity and results.

Firstly, for the encoder, the input image passes through three parallel channels. The size of the convolution kernel of each channel is 3×3 , but the three channels use different degrees of hole convolution, and the dilation is 1, 2 and 3. The dilation 1 represents the ordinary 3×3 resolution, and the 2 corresponds to the 5×5 , and the 3 corresponds to the 7×7 convolution kernels. Therefore, convolution kernels of different scales are concatenated to the outputs and

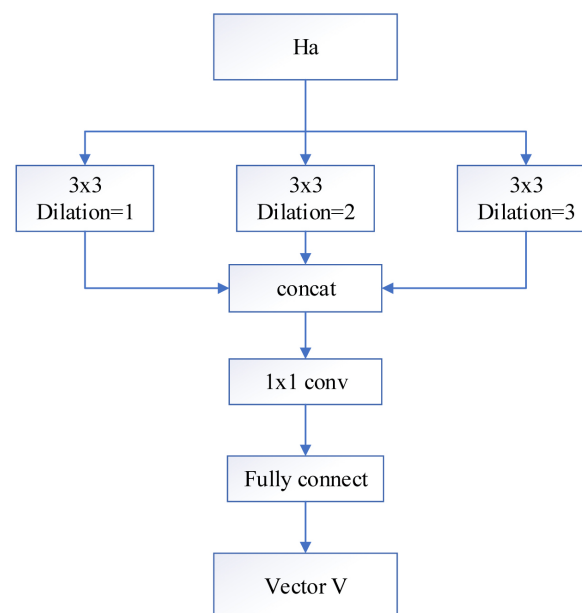


Figure 1. The structure of encoder with multi-scale and multi-channel.

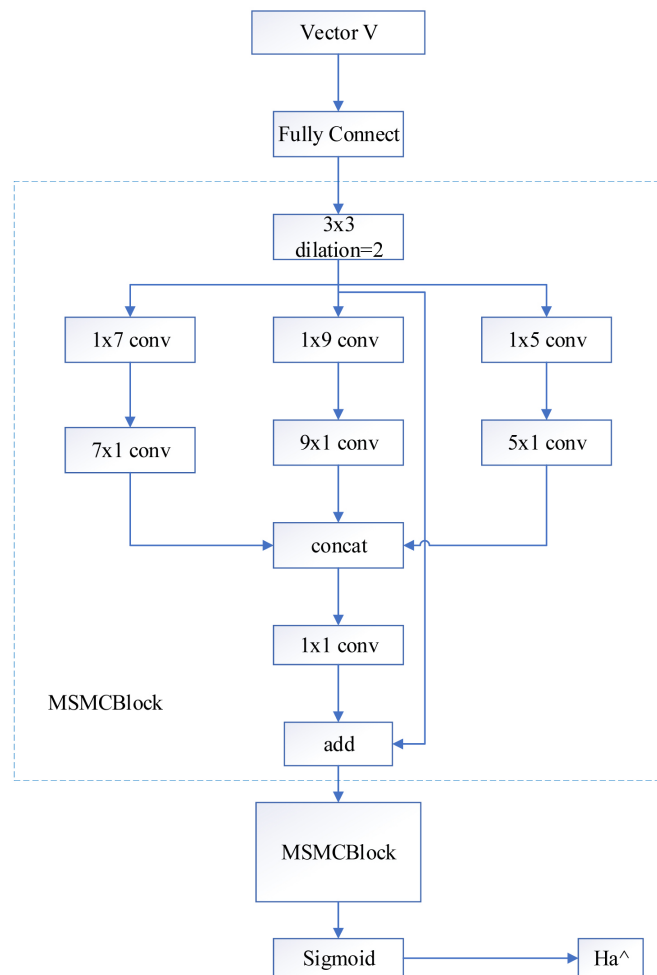


Figure 2. The structure of decoder with multi-scale and multi-channel.

merged them through a 1×1 convolution layer. Then we adopt the fully connected layer to obtain the desired compression rate.

For the decoder, firstly, it amplifies the received feature vector V to a specific size and roughly extract features through a 3×3 , dilation = 2 convolution kernel. After that, the output characteristics are obtained through two MSMCBlock modules and finally the CSI matrix is reconstructed through the Sigmoid layer. As shown in **Figure 2**, MSMCBlock has three parallel channels, one passes through 1×9 and 9×1 convolution kernels, and the second passes directly through 1×5 and 5×1 convolution kernels, and the third passes through 1×7 and 7×1 convolution kernels. The results are concatenated through 1×1 convolution layer. In this way, it can perform well in the case of different sparsity CSI matrices. MSMCBlock also directly merges the original data to the subsequent layers and adds the original data and the multi-channel convolution results. Multichannel refers that three channels we used and multiscale refers that different convolution kernels.

We use hole convolution and 1×9 , 9×1 serial convolution to replace the huge convolution kernel of 9×9 . They can retain the resolution effect of 9×9 in

the resolution area and reduce the complexity of the calculation and the number of parameters. The MSMCNet proposed in this paper has higher accuracy than CsiNet.

For each convolutional layer, the activation function LeakRelu is adopted instead of Relu, because the negative part of Relu has a slope of 0. When a large gradient flow passes through the Relu neuron, this neuron will be assigned a value of 0 after the parameters are updated. This neuron will not activate the subsequent data. The LeakRelu has a negative slope, and the above problem will not occur. And when we increase the slope appropriately, the performance of MSMCNet will be improved.

In addition, we also use a dynamic learning rate learning scheme. If we use 0.001 as the value of learning rate in 1000 learning epochs, we can indeed get good results. If we want to increase the efficiency of deep learning, we set the learning rate in the initial state to be high so that they can quickly enter the correct range. After that, the learning rate is reduced to ensure that over-fitting does not occur and it can get better results. The learning rate used follows a cosine function and the network performance can be improved through this dynamic learning method. Under the reference of literature [2] [5] and tested in advance, the maximum learning rate we set is 0.0025 and the minimum learning rate is 0.0005. The initial value of function is maximum and the last value is minimum, the whole process is monotonously decreasing. Such a learning rate setting can not only accelerate the learning speed of the first half of the network, but also prevent overfitting in the second half and further approach the true value.

4. Simulation Results and Analysis

The experiment is completed in an indoor scene at 5.3 GHz band. We use the COST2100 model to generate training samples. In order to facilitate comparison, the MSMCNet model has the same basic settings as the CsiNet model. The base station is a uniform linear array (ULA) model with $N_t = 32$. For the FDD system, we take $N_c = 1024$ in the frequency domain and $N_a = 32$ in the angular domain. 150,000 independently generated channels are divided into three parts: training, verification and test sets, including 100,000, 30,000 and 20,000 channel matrices, respectively. The Batchsize used in this training is 200. Reconstruction results with compression ratios (CR) of 4, 8, 16, and 32 were obtained. We also test the NMSE in an outdoor environment with compression ratio of 4. It is used to compare with indoor data.

The entire experimental system is implemented in Pytorch. Both the convolutional layer and the fully connected layer are initialized with Xavier. Typically, the default settings we use are $b1 = 0.9$, $b2 = 0.999$, $e = 1e-8$ and we use the Adam optimizer with mean square error (MSE) as the loss function. We also use a dynamic learning rate in the system, the maximum learning rate is 0.0025 and the minimum learning rate is 0.0005.

In order to evaluate the performance of MSMCNet, we use the normalized

mean square error (NMSE) to measure the distance between the original matrix \mathbf{H}_a and the reconstructed matrix $\tilde{\mathbf{H}}_a$.

The training epochs chosen in our test is 1000. During the test, we find that the training results usually get better with the growth of the epochs. However, when the epochs increase to 5000, the improvement will be very slight. So we use 1000 epochs in our experiment. In [5] CsiNet is better than traditional compressed sensing methods. From **Table 1** that in case of 4 compression ratios, MSMCNet with dynamic learning rate is better than const-MSMC and both they are better than CsiNet.

Figure 3 shows that whether the dynamic learning rate is used, the MSMCNet is better than the CsiNet and its loss is lower. When we compare the red and green curves in **Figure 3**, it is obvious that the loss of the red is lower, which shows that the usage of dynamic learning rate is obviously helpful for results.

Figure 4 shows the changes of NMSE under the three network frameworks. The data is recorded every 10 epochs and there are a total of 100 data in 1000 epochs. MSMCNet in the article has a better reconstruction rate with both dynamic and non-dynamic learning rate. Since the dynamic learning rate in the middle period is higher than the fixed learning rate, the red line in **Figure 4** fluctuates more greatly. At the end of the period, the red curve becomes more stable compared to the green. The final result shows that the dynamic learning rate has a better NMSE.

From **Figure 5**, we can see that MSMCNet's effect on outdoor channel matrix restoration is far from ideal and far lower than the indoor effect. When experiment at a high compression ratio, the reconstruction efficiency of the channel matrix will also decrease. So the further task is how to obtain a more ideal NMSE with high compression ratio and outdoor conditions.

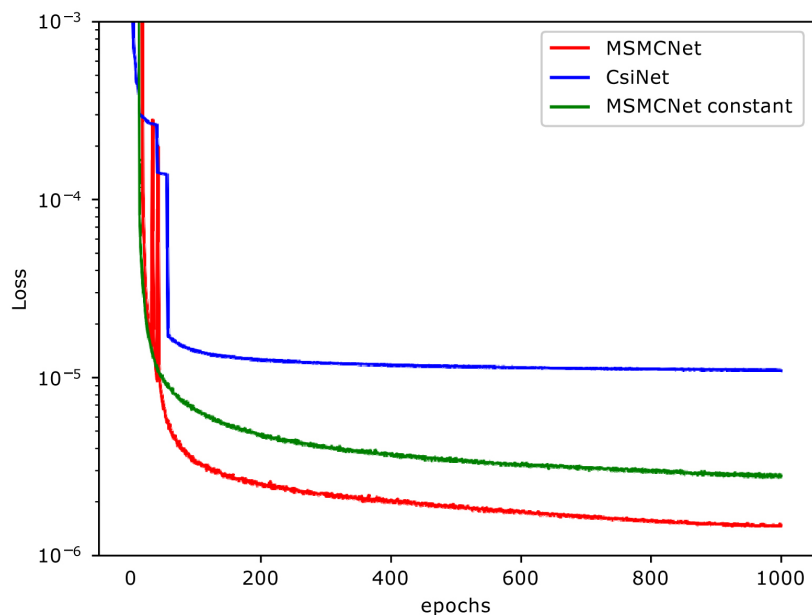


Figure 3. MSE (as loss function) between CsiNet and MSMCNet.

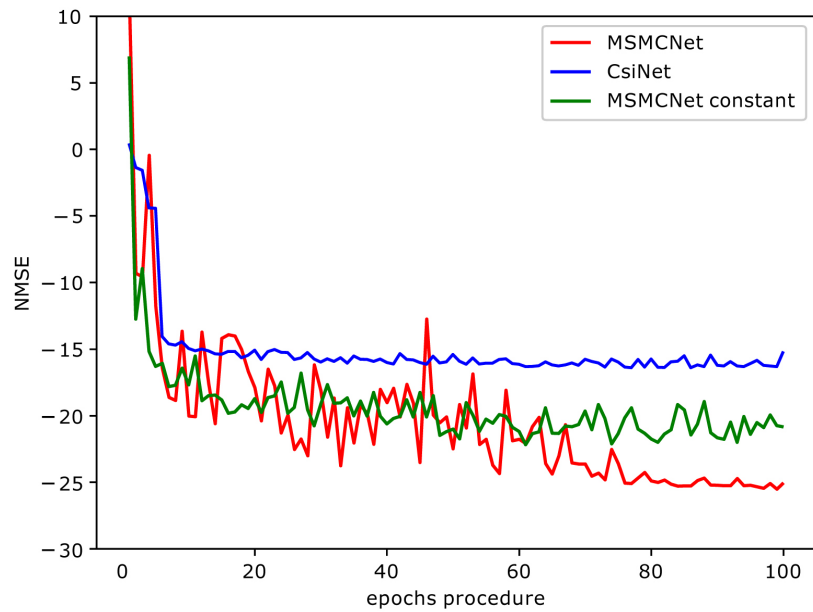


Figure 4. NMSE (dB) of CsiNet and MSMCNet in 1000 epochs with compression ratio 4.

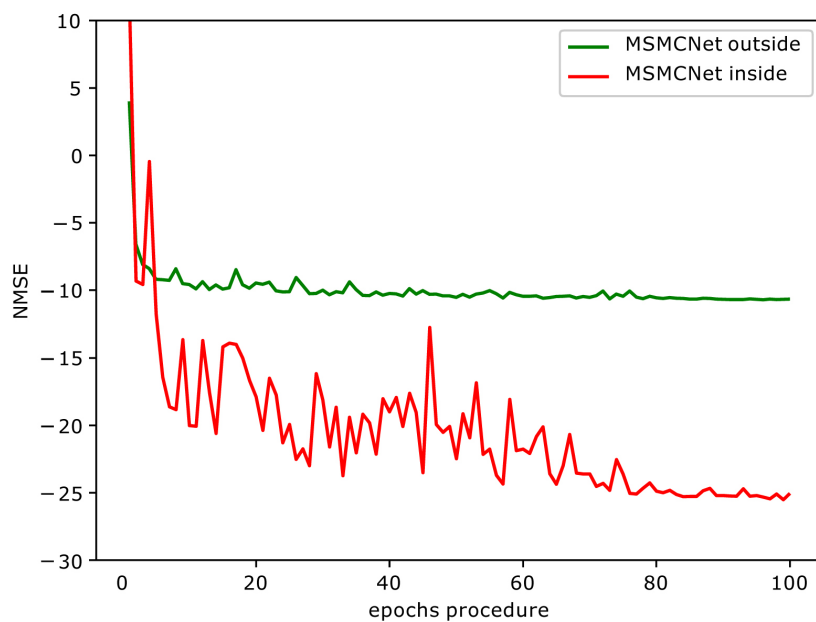


Figure 5. NMSE (dB) of MSMCNet in inside and outside with compression ratio 4.

Table 1. NMSE of different methods for 4 ratios.

Compress ratio	NMSE (dB) of different methods		
	Const-MSMC	MSMC	CSINet
4	-21.98	-25.52	-17.36
8	-13.69	-14.75	-12.70
16	-9.86	-10.42	-8.65
32	-8.23	-8.66	-6.24

We find that the CsiNet network almost reached the optimal value when it was trained for 100 epochs and the subsequent 900 epochs did not improve the results. MSMCNet can be significantly improved because it is more capable of exploiting subtle changes among adjacent elements than CsiNet methods.

5. Conclusion

For the downlink CSI feedback in the massive MIMO FDD system, we proposed MSMCNet on the basis of CsiNet. The multi-channel and multi-scale convolution was introduced into the CSI feedback task and it is proved to be effective. At the same time, the concept of dynamic learning rate was adopted to further improve the efficiency of CSI reconstruction. Experiments have shown that our scheme has higher reconstruction efficiency than CsiNet. But experiments also presented that deep learning methods have poor reconstruction efficiency in case of outdoor environments and high compression ratios. We hope this paper will encourage future research in this direction.

Acknowledgements

This work has been supported by the Research Fund of National Mobile Communications Research Laboratory, Southeast University (No. 2021C01).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Marzetta, T.L. (2010) Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Trans. Wireless Commun.*, **9**, 3590-3600. <https://doi.org/10.1109/TWC.2010.092810.091092>
- [2] Zhang, J., Wen, C.-K., Jin, S., Gao, X. and Wong, K.-K. (2013) On Capacity of Large-Scale MIMO Multiple Access Channels with Distributed Sets of Correlated Antennas. *IEEE J. Sel. Areas Commun.*, **31**, 133-148. <https://doi.org/10.1109/JSAC.2013.130203>
- [3] Rusek, F., *et al.* (2013) Scaling up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Process. Mag.*, **30**, 40-60. <https://doi.org/10.1109/MSP.2011.2178495>
- [4] Hoydis, J., Brink, S.T. and Debbah, M. (2013) Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need? *IEEE Journal on Selected Areas in Communication*, **31**, 160-171. <https://doi.org/10.1109/JSAC.2013.130205>
- [5] Wen, C., Shih, W. and Jin, S. (2018) Deep Learning for Massive MIMO CSI Feedback. *IEEE Wireless Communications Letters*, **7**, 748-751. <https://doi.org/10.1109/LWC.2018.2818160>
- [6] Kuo, P.H., Kung, H.T. and Ting, P.A. (2012) Compressive Sensing Based Channel Feedback Protocols for Spatially-Correlated Massive Antenna Arrays. 2012 *IEEE Wireless Communications and Networking Conference (WCNC)*, 492-497. <https://doi.org/10.1109/WCNC.2012.6214417>

-
- [7] Rao, X. and Lau, V.K.N. (2014) Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems. *IEEE Transactions on Signal Processing*, **62**, 2361-3271. <https://doi.org/10.1109/TSP.2014.2324991>
- [8] Daubechies, I., Defrise, M. and De Mol, C. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics. A Journal Issued by the Courant Institute of Mathematical Sciences*, **57**, 1413-1457. <https://doi.org/10.1002/cpa.20042>
- [9] Wang, T., et al. (2019) Deep Learning Based CSI Feedback Approach for Time-Varying Massive MIMO Channels. *IEEE Wireless Commun. Lett.*, **8**, 416-419. <https://doi.org/10.1109/LWC.2018.2874264>
- [10] Guo, J., Wen, C.-K., Jin, S. and Li, G.Y. (2019) Convolutional Neural Network Based Multiple-Rate Compressive Sensing for Massive MIMO CSI Feedback: Design, Simulation, and Analysis. arXiv preprint arXiv: 1906.06007.
- [11] Yang, Y., Gao, F., Li, G.Y. and Jian, M. Deep Learning Based Downlink Channel Prediction for FDD Massive MIMO System. *IEEE Communications Letters*, in Press.
- [12] Liu, Z., Zhang, L. and Ding, Z. (2019) Exploiting Bi-Directional Channel Reciprocity in Deep Learning for Low Rate Massive MIMO CSI Feedback. *IEEE Wireless Communications Letters*, **8**, 889-892. <https://doi.org/10.1109/LWC.2019.2898662>
- [13] Wen, C.K., Jin, S., Wong, K.K., et al. (2015) Channel Estimation for Massive MIMO Using Gaussian-Mixture Bayesian Learning. *IEEE Transactions on Wireless Communications*, **14**, 1356-1368. <https://doi.org/10.1109/TWC.2014.2365813>
- [14] Choi, J., Love, D.J. and Bidigare, P. (2014) Downlink Training Techniques for FDD Massive MIMO Systems: Open-Loop and Closed-Loop Training with Memory. *IEEE J. Sel. Topics Signal Process*, **8**, 802-814. <https://doi.org/10.1109/JSTSP.2014.2313020>
- [15] Liu, L., et al. (2012) The COST 2100 MIMO Channel Model. *IEEE Wireless Commun.*, **19**, 92-99. <https://doi.org/10.1109/MWC.2012.6393523>