Quan-Transformer Based Channel Feedback for RIS-Aided Wireless Communication Systems

Wenwu Xie, Jian Zou, Jian Xiao, Min Li, and Xin Peng

***Abstract*— Reconfigurable intelligent surface (RIS) has become one of the key technologies for future 6G communications due to its characteristics of an intelligent and controllable wireless environment. In the frequency division duplex mode, the user equipment feeds back the downlink channel state information (CSI) to the base station through the feedback link, to obtain potential gains for the RIS-aided wireless communication system. A large number of RIS reflection elements greatly increase the feedback link cost of the system. In this letter, Quan-Transformer, a sample network framework for CSI compression and reconstruction, is proposed based on Transformer network. The proposed framework can not only compress CSI effectively but also recover CSI with high precision, which improves the performance of the RIS-aided wireless communication system. According to the experimental results, compared with the CsiNet scheme, the achievable sum-rate and normalized mean squared error of the proposed scheme are increased by 6.57% and 3.32%, respectively.**

***Index Terms*— Reconfigurable intelligent surface, channel feedback, transformer, deep learning, frequency division duplex.**

I. Introduction

RECONFIGURABLE intelligent surface (RIS), as one of the key 6G communications technologies, can break through the uncontrollable characteristics of traditional wireless channels and realize the active control of the wireless transmission environment. Moreover, RIS can increase the received signal intensity and improve the transmission performance between communication devices by adjusting the direction of signal propagation and superposition with phase [1]. Wireless communication system based on RIS technology mainly includes time division duplex (TDD) and frequency division duplex (FDD) modes. At present, research on wireless communication system based on RIS technology is mainly based on TDD mode, but Wireless communication system based on RIS technology under FDD mode is also of great significance. In FDD mode, due to the lack of channel reciprocity, downlink channel state information (CSI) needs additional links for feedback. Current research demonstrates that the more RIS elements there are, the greater the performance gain of the system will be [2] and [3]. A large number of RIS reflection elements will inevitably lead to a sharp increase in feedback information and greatly increase the cost of the feedback link. Therefore, it has become research challenge to effectively reduce the cost of CSI feedback on the basis of ensuring the performance of RIS-aided wireless communication system.

In [4], [5], [6], and [7], the channel feedback overhead of massive multiple-input multiple-output (MIMO) system is solved by introducing compressed sensing method. But traditional compressed sensing uses sparse structure as a prior information to recover the CSI matrix. In practical channels, channel state information is not fully sparse, and compressed sensing uses random projection to obtain low-dimensional CSI, which cannot fully extract channel feature information. In addition, the compressed sensing method of iterative solution can not meet the real-time performance of the communication system. In order to solve this problem effectively, channel feedback scheme based on deep learning is proposed, and its performance is far better than that of compressed sensing scheme. For the feedback problem of massive MIMO systems, reference [8] first proposed a channel feedback framework based on deep learning, named CsiNet. On this basis, the long-term Memory (LSTM) network named CsiNet-LSTM was introduced in [9], considering the time correlation of channels. In [10], the attention mechanism is applied to the network to obtain the global receptive field. In addition, literature [11], [12] proposed a multi-user angle domain channel sparse transformation scheme to solve RIS-aided wireless communication feedback problem based on codebook method. However, to the best of our knowledge, there has been little work on the employment of deep learning tools to handle CSI feedback in RIS-aided wireless communication systems.

Recently, Transformer network architecture has been applied in various fields of deep learning. This architecture abandons the traditional convolutional neural network (CNN) and recurrent neural network, and is completely composed of attention mechanism [13]. In traditional CNN networks, local feature information can only be proposed by changing the convolution kernel, but global feature information cannot be effectively extracted. Transformer network architecture can not only extract local feature information, but also consider global feature distribution.

In this letter, a simple CSI compression and reconstruction network framework, called Quan-Transformer, is proposed to

Manuscript received 20 July 2022; accepted 7 August 2022. Date of publication 16 August 2022; date of current version 11 November 2022. This work was supported by the Open Fund of Advanced Cryptography and System Security Key Laboratory of Sichuan Province under Grant SKLACSS-202115, the Natural Science Foundation of Hunan Province under Grant 2020JJ4341, the Key Projects of Hunan Provincial Department of Education Department under Grant 21A0408. The associate editor coordinating the review of this letter and approving it for publication was A.-A. Boulogeorgos. *(Corresponding author: Xin Peng.)*

Wenwu Xie is with Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China, and also with Advanced Cryptography and System Security Key Laboratory, Chengdu 610103, China (e-mail: gavinxie@hnist.edu.cn).

Jian Zou, Min Li, and Xin Peng are with the Department of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China (e-mail: zoujian250@gmail.com; 982743528@qq.com; pengxin@hnist.edu.cn).

Jian Xiao is with the College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China (e-mail: jianx@mails.ccnu.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2022.3198793

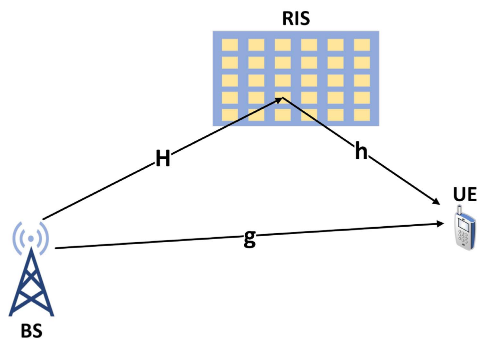


Fig. 1. The RIS-aided wireless communication system.

solve the CSI feedback problem caused by a large number of RIS reflection elements. The network is based on the autoencoder network, in which the encoder and decoder use Transformer network for CSI compression and recovery, respectively. Experimental results show that the performance of the Quan-Transformer is better than that of CsiNet, and the system achievable sum-rate and normalized mean square error (NMSE) are 6.57% and 3.32% higher than that of CsiNet, respectively.

II. System Model

As shown in Fig. 1, we consider the RIS-aided wireless communication system, where the BS is equipped with M antennas of uniform line array (ULA). The RIS is equipped with N reflection elements of uniform plane array (UPA).

The downlink signal received at the UE can be expressed as

$$\begin{aligned} y &= \left(\mathbf{g} + \mathbf{h}^{\mathrm{T}}\boldsymbol{\Theta}\mathbf{H}\right) s + z \\ &= \left(\mathbf{g} + \theta^{\mathrm{T}}\mathrm{diag}\left(\mathbf{h}^{\mathrm{T}}\right)\mathbf{H}\right) s + z, \end{aligned} \tag{1}$$

where $\mathbf{g} \in \mathbb{C}^{1\times M}$, $\mathbf{h}^{\mathrm{T}} \in \mathbb{C}^{1\times N}$, $\mathbf{H} \in \mathbb{C}^{N\times M}$ denote the BS-UE channel, the RIS-UE channel, and the BS-RIS channel, respectively. s is the signal sent by the transmitter, and z denotes additive white Gaussian noise. $\boldsymbol{\Theta} \in \mathbb{C}^{N\times N}$ represents the adjustable phase shift diagonal matrix of RIS as [14]

$$\boldsymbol{\Theta} = \mathrm{diag}\left(\theta^{\mathrm{T}}\right) = \mathrm{diag}\left(\omega e^{j\theta_1}, \ldots, \omega e^{j\theta_n}, \ldots, \omega e^{j\theta_N}\right), \tag{2}$$

where θ_n is the phase angle of the n-th RIS element, and $\omega \in [0, 1]$, denotes the amplitude coefficient, and in this work, ω is set to 1.

In this work, the cascaded (BS-RIS-UE) CSI is compressed and reconstructed. The cascaded CSI at UE can be expressed as

$$\mathbf{H}_{\mathrm{DL}} = \mathrm{diag}\left(\mathbf{h}^{\mathrm{T}}\right)\mathbf{H}. \tag{3}$$

According to [15], the BS-RIS channel can be written as

$$\mathbf{H} = \sum_{i=1}^{L_1} \rho_i \mathbf{m}\left(p_{1,i}, q_{1,i}\right) \mathbf{n}\left(p_i^{\mathrm{AOD}}\right), \tag{4}$$

where ρ_i denotes the complex gain of the i-th path, L_1 is the number of paths, the steering vector $\mathbf{n}(p_i^{\mathrm{AOD}})$ denotes the antenna array response of the i-th path. $\mathbf{m}(p_{1,i}, q_{1,i})$ is the steering vector of RIS between BS-RIS channel of the i-th path, which can be written as

$$\mathbf{m}\left(p_{1,i}, q_{1,i}\right) = \frac{1}{\sqrt{N}}[e^{j2\pi n_1 p_{1,i}}]^{\mathrm{T}} \otimes [e^{j2\pi n_2 q_{1,i}}]^{\mathrm{T}}, \tag{5}$$

where N_1 and N_2 represent the number of elements in the horizontal and vertical directions of RIS, respectively, $n_1 \in \{1, 2, \ldots, N_1\}$ and $n_2 \in \{1, 2, \ldots, N_2\}$. $p_{1,i} = \frac{d_1}{\lambda}\cos\beta_{\mathrm{BR},i}\sin\alpha_{\mathrm{BR},i}$, $q_{1,i} = \frac{d_1}{\lambda}\sin\beta_{\mathrm{BR},i}$, where λ is the wavelength, d_1 denotes the distance between RIS elements, $\alpha_{\mathrm{BR},i}, \beta_{\mathrm{BR},i}$ represents the azimuth and elevation in the arrival angle between BS and RIS, respectively.

$$\mathbf{n}\left(p_i^{\mathrm{AOD}}\right) = \frac{1}{\sqrt{M}}[e^{j2\pi m p_i^{\mathrm{AOD}}}]^{\mathrm{T}}, \tag{6}$$

where $p_i^{\mathrm{AOD}} = \frac{d_2}{\lambda}\sin\alpha_i^{\mathrm{AOD}}$, and α_i^{AOD} represents the angle of departure (AOD) the i-th path between BS and RIS, d_2 denotes the antenna spacing at the BS, and $m \in \{1, 2, \ldots, M\}$.

The RIS-UE channel can be written as

$$\mathbf{h} = \sum_{i=1}^{L_2} \xi_i m^{\mathrm{H}}\left(p_{2,i}, q_{2,i}\right), \tag{7}$$

where ξ_i denotes the complex gain of the i-th path.

$$\mathbf{m}^{\mathrm{H}}\left(p_{2,i}, q_{2,i}\right) = \frac{1}{\sqrt{N}}[e^{j2\pi n_1 p_{2,i}}]^{\mathrm{T}} \otimes [e^{j2\pi n_2 q_{2,i}}]^{\mathrm{T}}, \tag{8}$$

where $p_{2,i} = \frac{d_1}{\lambda}\cos\beta_{\mathrm{RU},i}\sin\alpha_{\mathrm{RU},i}$ and $q_{2,i} = \frac{d_1}{\lambda}\sin\beta_{\mathrm{RU},i}$, $\alpha_{\mathrm{BR},i}$, $\beta_{\mathrm{BR},i}$ represents the azimuth and elevation in the departure angle between RIS and UE, respectively.

According to (3), (4) and (7), $\mathbf{H}_{\mathrm{DL}}$ can be written as

$$\begin{aligned} \mathbf{H}_{\mathrm{DL}} = &\sum_{i=1}^{L_1}\sum_{j=1}^{L_2} \rho_i \xi_i \\ &\times \mathrm{diag}\left(\mathbf{m}^{\mathrm{H}}\left(p_{2,i}, q_{2,i}\right)\right)\mathbf{m}\left(\alpha_{BR,i}, \beta_{BR,i}\right)\mathbf{n}^{\mathrm{H}}\left(\alpha_i^{\mathrm{AOD}}\right). \end{aligned} \tag{9}$$

By 2D discrete fourier transform (DFT) transformation, CSI matrix will be sparse to facilitate effective compression recovery, and the transformed matrix can be written as

$$\mathbf{H}_{\mathrm{IN}} = \mathbf{F}_{\mathrm{d}}\mathbf{H}_{\mathrm{DL}}\mathbf{F}_{\mathrm{a}}{}^{\mathrm{H}}, \tag{10}$$

where $\mathbf{F}_{\mathrm{d}} \in \mathbb{C}^{N\times N}$ and $\mathbf{F}_{\mathrm{a}} \in \mathbb{C}^{M\times M}$ are DFT matrices.

III. Channel Feedback Process and Network Architecture

In this section, channel feedback process, channel feedback network architecture and quantitative methods will be introduced. In the network architecture section, each network layer will be introduced in detail, as well as the differences of network design.

A. Channel Feedback Process

When the BS-RIS-UE cascaded CSI matrix $\mathbf{H}_{\mathrm{IN}}$ is estimated at the UE, the CSI matrix is successively compressed and quantized. The processed matrix $\mathbf{H}_{\mathrm{bit}}$ can be expressed as

$$\mathbf{H}_{\mathrm{bit}} = Q\left(f_{\mathrm{encode}}\left(\mathbf{H}_{\mathrm{IN}}, \theta_1\right)\right), \tag{11}$$

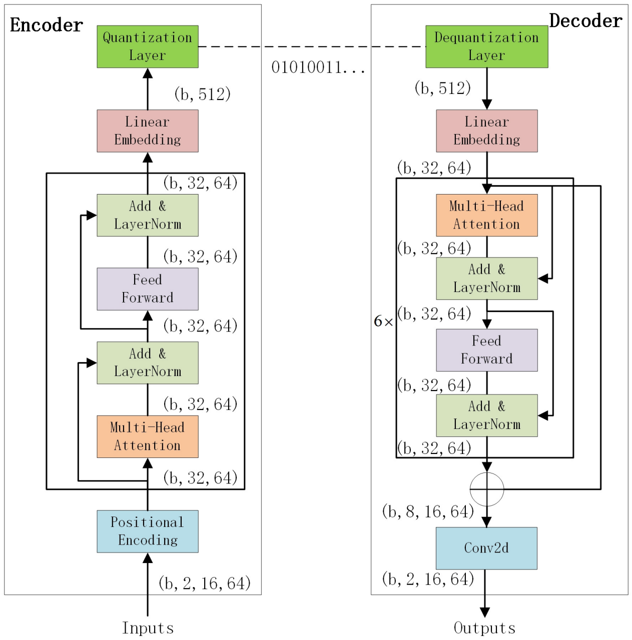
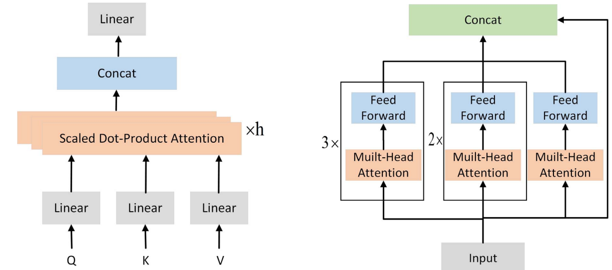


Fig. 2. Architecture of Quan-transformer.



(a) The architecture of Muilt-Head Attention

(b) Detailed structure of six attention in decoder

Fig. 3. Detailed structure in Quan-transformer.

where $f_{\text{encode}}(\bullet)$ and $Q(\bullet)$ represent compression and quantization, respectively, and θ_1 donotes the parameters of the compression module (Encoder).

After the compressed CSI matrix $\mathbf{H}_{\text{bit}}$ is received by BS, matrix $\mathbf{H}_{\text{bit}}$ is successively dequantized, decompressed and channel matrix recovered. The recovered CSI matrix $\hat{\mathbf{H}}$ can be expressed as

$$\hat{\mathbf{H}} = f_{\text{decode}}\left(P\left(\mathbf{H}_{\text{bit}}, \theta_2\right)\right), \tag{12}$$

where $f_{\text{decode}}(\bullet)$ and $P(\bullet)$ represent the decompression and dequantization, and θ_2 donotes the parameters of the compression module (Decoder), respectively.

Therefore, combining (11) and (12) with the mean square error (MSE) function, the expression of optimized compression and recovery can be obtained:

$$\left(\hat{\theta}_1, \hat{\theta}_2\right) = \underset{\theta_1, \theta_2}{\arg\min} ||\mathbf{H}_{\text{IN}} - \hat{\mathbf{H}}||_2^2. \tag{13}$$

B. Network Architecture for Channel Feedback

The network architecture mainly includes encoder and decoder. In the encoder and decoder, the backbone network is based on the Transformer network, and is named as Quan-Transformer network. The proposed network framework diagram is shown in Fig. 2, where b represents the training parameter batch size. In the encoder, CSI matrix goes through positional encoding, an attention module and linear embedding module. The attention module mainly includes multi-head attention and feed forward. Linear embedding is mainly used to compress and recover the dimensions of the features extracted by the Transformer network, so as to facilitate quantization and de quantization. It successively includes linear layer, batchnorm layer and sigmoid activation function. In the decoder, six attention modules are used, and the output of each module is spliced with the output of linear embedding, and finally passes through the convolution layer.

1) Network Layer Details: Positional Encoding is introduced in the network to encode each CSI so that CSI at different positions could be distinguished during feature extraction. In this work, sine and cosine functions of different frequencies were used for position encoding [16]:

$$\begin{aligned} \mathrm{P}\left(l, 2i\right) &= \sin\left(\frac{l}{10000^{2i/d_{\text{model}}}}\right) \\ \mathrm{P}\left(l, 2i+1\right) &= \cos\left(\frac{l}{10000^{2i/d_{\text{model}}}}\right), \end{aligned} \tag{14}$$

where i and l represent dimensions and positions, respectively, P_{l+k} can be represented anywhere by a linear function of P_l, d_{model} donate the length of information encoding.

The commonly used self attention includes dot product attention and additive attention [17]. This work uses scaled dot-product attention

$$\mathrm{A}\left(\mathrm{Q}', \mathrm{K}', \mathrm{V}'\right) = \mathrm{softmax}\left(\frac{\mathrm{Q}'\mathrm{K}'^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}\right)\mathrm{V}' \tag{15}$$

where Q', K', V' all represent outputs after linear embeedding raises dimension, and suitable $\mathrm{d_k}$ will be used to solve the gradient is too small.

According to [13], the feed-forward network (FFN) consists of two linear layers and a ReLu activation function, FFN can be expressed as

$$\mathrm{FFN}\left(x\right) = \max(0, xW_1 + b_1)W_2 + b_2, \tag{16}$$

where x is input, W_1 and W_2 are linear layer weights, b_1 and b_2 are linear layer intercepts.

2) Network Differences: Firstly, unlike traditional Transformer, we have abandoned positional encoding in the decoder. Secondly, in the FFN layer, each linear layer uses different hidden layers for feature extraction. Finally, in Decoder, the output of six attention modules is spliced and then passed through the convolution layer.

C. Quantization and Dequantization

Uniform quantization is essentially a process of rounding data, in which each sample value is rounded to the nearest value in a set of quantization of finite size [18]. It is assumed that the compression value of CSI feedback matrix is between $[a, b]$, we have

$$f\left(x\right) = \mu\left\lfloor\frac{x}{\mu}\right\rfloor, \tag{17}$$

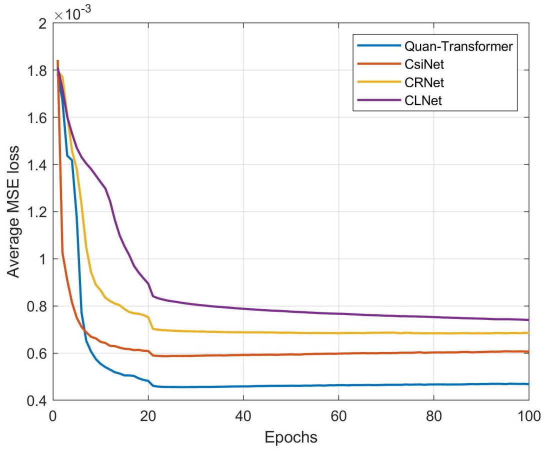


Fig. 4. Average MSE loss of different feedback schemes under different epochs.

where x is input, $\mu = \frac{b-a}{2^k-1}$. In this work, $k = 2$, and represents 2-bit quantization.

IV. Simulation Results

Four feedback simulation results of Quan-Transformer, CsiNet, CLNet and CRNet are given from two perspectives of network performance and system performance. In the experiment, $d_1 = d_2 = 0.5$, $M = 16$, $N = 64$, $L_1 = 4$ and $L_2 = 2$. The sample numbers of training set, test set and validation set were 60000, 2000 and 1000, respectively. Epochs, batch size and learning rate were set as 100, 16 and 0.001, respectively.

During the network training, we used NMSE and compression rate (CR) to evaluate the compression recovery performance of each network architecture. The NMSE can be expressed as

$$\text{NMSE} = \text{E}\left\{\frac{\left\|\mathbf{H}_{\text{IN}} - \hat{\mathbf{H}}\right\|_2^2}{\|\mathbf{H}_{\text{IN}}\|_2^2}\right\}, \tag{18}$$

where $\mathbf{H}_{\text{IN}}$ is the CSI matrix of network input, $\hat{\mathbf{H}}$ is the CSI recovery matrix output by the network.

According to [20], The CR can be written as

$$\text{CR} = \frac{2NM}{L}, \tag{19}$$

where $2NM$ represents dimension of CSI matrix, 2 denotes the real and imaginary part of the CSI matrix. In this work, N, M represent the number of RIS elements and antennas, respectively. L represents the dimension of the output of the last linear layer of the encoder.

Fig. 4 shows the MSE loss curve of different feedback schemes at different epochs. For the results in the figure, we set $CR = 4$, $N = 64$, $M = 16$. From the figure, we can see that when all schemes are trained to the optimal parameters, the MSE loss of Quan-Transformer is less than that of other schemes.

As Table I shown, the NMSE comparison results of Quan-Transformer and other three different deep learning schemes. Among them, the scheme corresponding to the bold values indicates that the NMSE performance of this method is the best compared with other methods under the same CR. According to the data analysis in the table, when CR is small, the NMSE performance of Quan-Transformer scheme is much better than the other three schemes. The results show that the NMSE performance of Quan-Transformer is improved by 3.32% compared with that of CsiNet. Fig. 5 shows the NMSE performance diagram of various feedback schemes under different compression ratios when $N = 64$ and $M = 16$. When $CR = 4$ and $M = 16$, the NMSE performance curves under different feedback schemes with the change of the number of reflection elements are described in Fig. 6. In Fig. 7, when $CR = 4$ and $N = 64$, NMSE performance curves under different feedback schemes change with the number of BS antennas.

TABLE I

NMSE(dB) Performance of Different Feedback Schemes Under Different Compression Rates

Schemes \ CR	32	16	8	4
CsiNet[8]	-22.512	-22.992	-24.660	-26.338
CLNet[10]	-22.293	-22.744	-24.170	-25.132
CRNet[19]	-22.348	-22.837	-24.195	-25.692
Quan-Transformer	**-22.542**	**-24.092**	**-25.865**	**-27.436**

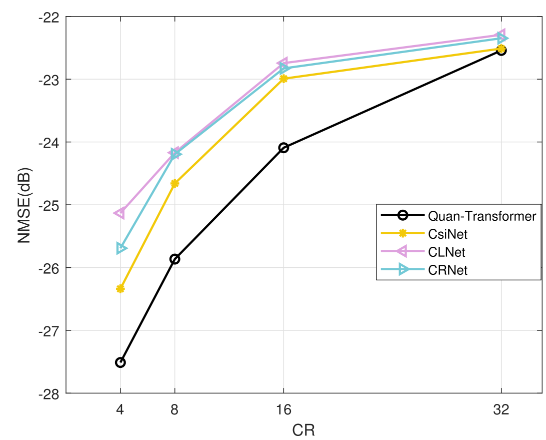


Fig. 5. The NMSE performance of different feedback schemes under different CR.

In order to further determine the effectiveness and reliability of the feedback scheme, we select achievable sum-rate (ASR) and to evaluate the feedback performance of the complete communication system. The system of ASR can be written as

$$\text{R} = \log_2\left(1 + \frac{p|(\mathbf{g} + \mathbf{h}^{\text{T}}\mathbf{\Theta}\mathbf{H})\hat{x}|^2}{\sigma^2}\right), \tag{20}$$

where, $\hat{x}$ and p are the precoding matrix and transmitted power at BS, respectively, σ is noise power.

Fig. 8 describes the variation curve of system ASR of two better feedback schemes (Quan-Transformer, CsiNet) under different SNR. The results show that with the increase of SNR, the system ASR of the two schemes also increase, and are far better than those without channel feedback. At the same CR, the system achievable sum-rate of Quan-Transformer scheme is much better than that of CsiNet. While SNR = 11dB, the

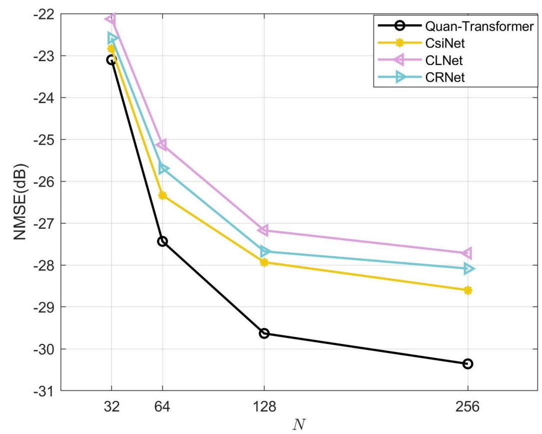


Fig. 6. The NMSE performance of different feedback schemes under different N.

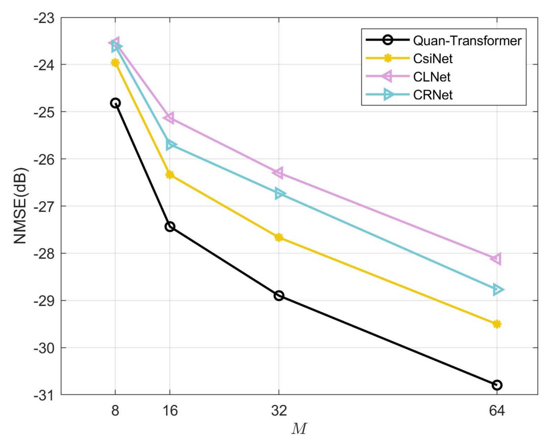


Fig. 7. The NMSE performance of different feedback schemes under different M.

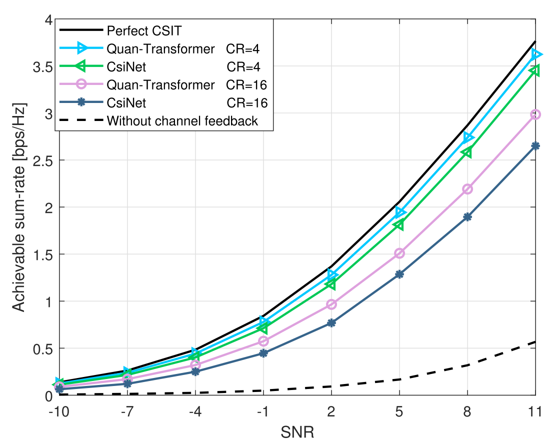


Fig. 8. The achievable sum-rate of CsiNet and Quan-Transformer schemes under different SNR.

system ASR of Quan-Transformer scheme is 6.57% higher than that of CsiNet scheme in different CR.

V. Conclusion

In this letter, a simple network framework for CSI compression and reconstruction, named Quan-Transformer, is proposed and applied for the first time in the channel feedback problem of RIS-aided wireless communication system. This proposed network is based on the autoencoder network, and the Transformer network is used for CSI compression and recovery in the encoder and decoder, respectively. On this basis, the quantization module is further introduced to improve the CSI reconstruction accuracy again. Experimental results show that Quan-Transformer has better performance than CsiNet.

References

[1] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[2] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, May 2020.

[3] E. Basar, M. D. Renzo, J. D. Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[4] 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)*, Apr. 2012, pp. 492–497.

[5] 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.

[6] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

[7] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Sep. 2009.

[8] 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.

[9] T. Wang, C.-K. 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.

[10] S. Ji and M. Li, "CLNet: Complex input lightweight neural network designed for massive MIMO CSI feedback," *IEEE Commun. Lett.*, vol. 10, no. 10, pp. 2318–2322, Oct. 2021.

[11] D. Shen and L. Dai, "Dimension reduced channel feedback for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7748–7760, Nov. 2021.

[12] X. Shi, J. Wang, and J. Song, "Triple-structured sparsity-based channel feedback for RIS-assisted MU-MIMO system," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1141–1145, May 2022, doi: 10.1109/LCOMM.2022.3147220.

[13] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[14] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, Nov. 2021.

[15] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.

[16] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *Proc. ICML*, 2017, pp. 1243–1252.

[17] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[18] M. I. Sezan, K. L. Yip, and S. J. Daly, "Uniform perceptual quantization: Applications to digital radiography," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-17, no. 4, pp. 622–634, Jul. 1987.

[19] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[20] J. Guo, C.-K. 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.*, vol. 19, no. 4, pp. 2827–2840, Apr. 2020.