SCIENCE CHINA
Information Sciences

• LETTER •

Transformer-based downlink precoding design in massive MIMO systems for 5G-advanced and 6G

Hao JIANG[1,2] & Linglong DAI[1,2*]

[1]*Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China;*
[2]*Beijing National Research Center for Information Science and Technology (BNRist), Beijing* 100084, *China*

Citation

Massive multiple-input multiple-output (MIMO) with a larger number of base station (BS) antennas has been regarded as a key technique to support digital twins and the metaverse in future 5G-Advanced and 6G [1]. To realize the expected high-speed transmission, the channel information state (CSI) is required to design the precoding in BS. However, since the pilot and feedback overhead scales linearly with the number of BS antennas, the CSI acquisition overhead is prohibitively high for future frequency division duplexing (FDD) massive MIMO systems.

To reduce the CSI acquisition overhead for future 5G-Advanced and 6G, some solutions have been proposed to improve the downlink pilot training strategy in FDD mode. The compressed sensing (CS) algorithm is utilized to exploit the sparse structure of the channel, which could compress the high-dimensional channel to a low-dimension representation. However, iterative CS algorithms always bring high time consumption. Leveraging GPU, the CsiNet [2] composed of deep neural networks (DNN) could realize user equipment (UE) channel quantization and BS channel recovery in a short time. At the same time, with the help of joint optimization of these two modules, CsiNet could realize lower feedback overhead than CS algorithms. Moreover, [3] introduced an end-to-end (E2E) design shown in Fig. 1(a), which further takes BS downlink pilot design and UE downlink channel estimation into consideration and jointly optimized the four modules to realize lower pilot overhead.

Besides relying only on downlink pilot training, the uplink CSI could be used to reduce downlink CSI acquisition overhead by utilizing the *partial reciprocity* between the uplink and downlink channels in FDD mode. By using the CS algorithm, the frequency-independent channel parameters such as azimuth angles are first extracted from uplink CSI. Then, the pilot is customized according to the azimuth angles to obtain the frequency-dependent parameters. To shorten the running time, a CsiNet-based downlink CSI estimation network named DualNet was proposed in [4] by additionally considering uplink CSI. Furthermore, [5] developed the E2E design [3] to further reduce the CSI acquisition overhead with the help of the uplink CSI shown in Fig. 1(b). Unfortunately, the above uplink-aided E2E schemes do not include the downlink precoding design. Owing to the severe multi-user interference, the lack of downlink precoding in E2E models could result in a limited reduction CSI acquisition overhead.

To fill the gap, in this study, we consider the uplink-aided E2E design that takes precoding into consideration, and propose a transformer-based downlink precoding scheme to further reduce the CSI acquisition overhead. In the field of deep learning, the transformer model could be regarded as a multimodal fusion model [6], which could simultaneously process text and image data representing the same content and extract the fusion feature to serve downstream tasks. Inspired by this, we design a transformer model by considering the uplink CSI and feedback bits as multimodal data and utilize the transformer to fuse these two types of data to support subsequent precoding design.

Specifically, we consider an N-antennas BS which could simultaneously serve K single-antenna UEs. In the proposed scheme, the multi-user uplink channels $\mathbf{H} = [\mathbf{h}_1^{\mathrm{ul}}, \mathbf{h}_2^{\mathrm{ul}}, \cdots, \mathbf{h}_K^{\mathrm{ul}}] \in \mathbb{C}^{N\times K}$ are utilized to design the downlink pilot $\mathbf{P} \in \mathbb{C}^{N\times M}$ and downlink precoder $\mathbf{V} \in \mathbb{C}^{N\times K}$ at BS shown in Fig. 1(c), where M is the length of the pilot signal. First, the downlink pilot design could be denoted by a DNN with L_p layer, which could be written as

$$\mathbf{\Gamma}^{(i)} = \gamma_p^{(i)}(\mathbf{W}_p^{(i)}\mathbf{\Gamma}^{(i-1)} + \mathbf{b}_p^{(i)}), i = 1, 2, \cdots, L_p, \quad (1)$$

where $\mathbf{\Gamma}^{(0)} = \mathrm{vec}(\mathbf{H})$; $\mathbf{\Gamma}^{(L_p)} = \mathrm{vec}(\mathbf{P})$; $\gamma_p^{(i)}(\cdot)$, $\mathbf{W}_p^{(i)}$, and $\mathbf{b}_p^{(i)}$ are the activation function, weights, and bias of the i-th layer, respectively. Note that the $\gamma_p^{(L_p)}(\cdot)$ at the last layer is the normalized function which makes the downlink pilot $\mathbf{P}$ satisfy the power constraint. After receiving the pilot signals, each UE describes the downlink channel with information bits $\mathbf{q}_k \in \{\pm 1\}^{B\times 1}$ according to the observed pilot signals $\mathbf{y}_k \in \mathbb{C}^{M\times 1}$, which is also realized by another DNN as follows:

$$\mathbf{\Upsilon}_k^{(i)} = \gamma_q^{(i)}(\mathbf{W}_q^{(i)}\mathbf{\Upsilon}_k^{(i-1)} + \mathbf{b}_q^{(i)}), i = 1, 2, \cdots, L_q, \quad (2)$$

where $\mathbf{\Upsilon}_k^{(0)} = \mathbf{y}_k$, $\mathbf{\Upsilon}_k^{(L_q)} = \mathbf{q}_k$, and $\gamma_q^{(L_q)}(\cdot) = \mathrm{sign}(\cdot)$.

Next, as shown in Fig. 1(d), the downlink precoder $\mathbf{V}$ is computed by the proposed transformer-based precoding model according to the uplink CSI and feedback bits. Specifically, in the transformer model, self-attention mechanisms

* Corresponding author (email: daill@tsinghua.com)

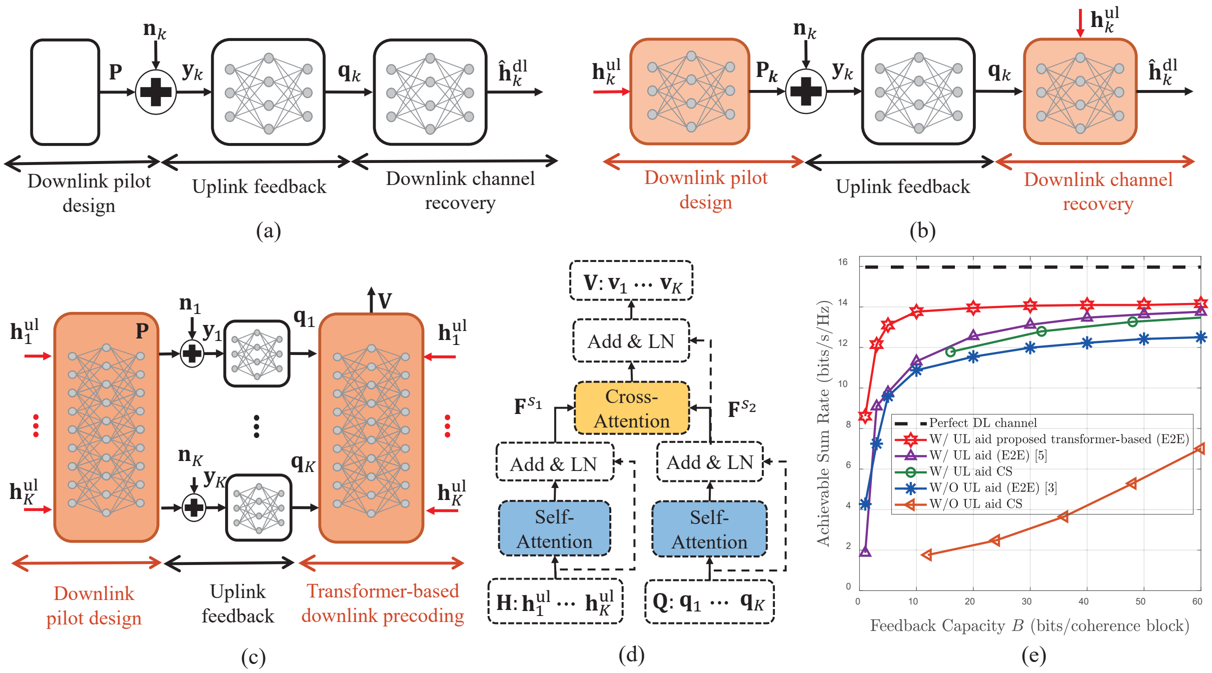

Figure 1 (a) The E2E scheme for downlink channel estimation without uplink aid [3], (b) the E2E scheme for downlink channel estimation with uplink aid [5], (c) the proposed transformer-based E2E scheme for downlink precoding with uplink aid, (d) the proposed transformer-based precoding model, and (e) the comparison of achievable sum rate performance.

are used to calculate the inter-user interference, which could be denoted by

$$\mathbf{Z}^{s_1} = \mathbf{W}_{val}^{s_1}\mathbf{H}\mathrm{Softmax}\big((\mathbf{W}_{key}^{s_1}\mathbf{H})^T\mathbf{W}_{que}^{s_1}\mathbf{H}/\sqrt{h}\big), \quad (3)$$

$$\mathbf{Z}^{s_2} = \mathbf{W}_{val}^{s_2}\mathbf{Q}\mathrm{Softmax}\big((\mathbf{W}_{key}^{s_2}\mathbf{Q})^T\mathbf{W}_{que}^{s_2}\mathbf{Q}/\sqrt{h}\big), \quad (4)$$

where $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_K] \in \{\pm 1\}^{B\times K}$ is the feedback bits from multi-user; $\mathbf{W}_{val}^{s_i}$, $\mathbf{W}_{key}^{s_i}$, and $\mathbf{W}_{que}^{s_i}$ are three trainable linear transformations. Then, the residual link and layer normalization (LN) are used to improve the training efficiency. The inter-user interference features extracted from the uplink channel and collected bits could be written as $\mathbf{F}^{s_1} = \mathrm{LN}(\mathbf{Z}^{s_1} + \mathbf{H})$ and $\mathbf{F}^{s_2} = \mathrm{LN}(\mathbf{Z}^{s_2} + \mathbf{Q})$, respectively.

Following, a cross-attention mechanism is applied to utilize uplink features $\mathbf{F}^{s_1}$ to assist downlink precoding. Since the uplink channel contains frequency-independent angle parameters, $\mathbf{F}^{s_1}$ can be used to mitigate inter-user angle interference. The output of the cross-attention is denoted by

$$\mathbf{Z}^{c} = \mathbf{W}_{val}^{c}\mathbf{F}^{s_2}\mathrm{Softmax}\big((\mathbf{W}_{key}^{c}\mathbf{F}^{s_1})^T\mathbf{W}_{que}^{c}\mathbf{F}^{s_2}/\sqrt{h}\big), \quad (5)$$

where the $\mathbf{W}_{val}^{c}$, $\mathbf{W}_{key}^{c}$, and $\mathbf{W}_{que}^{c}$ are corresponding trainable linear transformations of cross-attention. Then, the downlink precoding matrix could be given by $\mathbf{V} = \mathrm{LN}(\mathbf{Z}^c + \mathbf{F}^{s_2})$. Finally, by maximizing the achievable sum rate, the transformer-based downlink precoding scheme could be jointly optimized.

Simulation results. Fig. 1(e) shows the achievable sum rate performance comparison when $N = 64$, $K = 2$, and $M = 8$. The 3.5 GHz uplink channels are utilized to assist the downlink precoder design in 3.6 GHz. We could observe that the uplink-aided transformer-based downlink precoding scheme could achieve 20% gain over the existing uplink-aided E2E scheme when $B = 10$. In addition, compared with other uplink-aided schemes, the proposed scheme could reduce the feedback overhead by 60% when the achievable sum rate is 12 bits/s/Hz, which benefits from the joint optimization of multiple modules. Moreover, compared with the upper bound, the proposed uplink-aided transformer-based downlink precoding scheme suffers from 2 bits/s/Hz performance loss.

Conclusions. In this study, we have proposed an uplink-aided transformer-based downlink precoding scheme to further reduce the CSI acquisition overhead for massive MIMO system. Unlike the existing E2E schemes, the proposed scheme could realize the uplink-aided multi-user downlink precoding. This is achieved by using the multimodal fusion model named transformer in the deep learning field, which could simultaneously process the uplink CSI and feedback bits. Simulation results show that the proposed transformer-based downlink precoding scheme can achieve a much more overhead reduction than the existing E2E schemes.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China (Grant No. 2020YFB1805005) and in part by the National Natural Science Foundation of China (Grant No. 62031019).

References

1 Z. Zhang, Y. Xiao, Z. Ma, et al. 6G wireless networks: Vision, requirements, architecture, and key technologies. IEEE Veh. Technol. Mag., 2019, 14: 28-41

2 C.-K. Wen, W.-T. Shih, S. Jin. Deep learning for massive MIMO CSI feedback. IEEE Wireless Commun. Lett., 2018, 7: 748-751

3 F. Sohrabi, K. M. Attiah, W. Yu. Deep learning for distributed channel feedback and multiuser precoding in FDD massive MIMO. IEEE Trans. Wireless Commun., 2021, 20: 4044-4057

4 Z. Liu, L. Zhang, Z. Ding. Exploiting bi-directional channel reciprocity in deep learning for low rate massive MIMO CSI feedback. 2019, 8: 889-892

5 J. Guo, C.-K. Wen, S. Jin. CAnet: Uplink-aided downlink channel acquisition in FDD massive MIMO using deep learning. 2022, 70: 199-214

6 A. Jaegle, F. Gimeno, A. Brock, et al. Perceiver: General perception with iterative attention. In: Proceedings of Inter. Conf. Mach. Learn., 2021, 139: 4651-4664