Performance Evaluation of Deep Learning-Based CSI Feedback in FDD Massive MIMO

Bora Yoon, Junghyun Kim
Department of Artificial Intelligence and Data Science
Sejong University
Seoul, Republic of Korea
boray0614@sju.ac.kr, j.kim@sejong.ac.kr

Dukhyun You, JunHwan Lee
Satellite Communication Research Division
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
dhyou@etri.re.kr, junhwanlee@etri.re.kr

***Abstract*—In frequency division duplex (FDD) systems, user equipment (UE) must feed back the estimated downlink channel state information (CSI) to the base station (BS), which leads to significant overhead in massive MIMO environments. To mitigate this issue, deep learning-based CSI feedback schemes have been proposed; however, most of them focus primarily on reconstruction error metrics such as normalized mean squared error (NMSE), which do not sufficiently reflect actual transmission performance. In this paper, we compare the 5G NR codebook-based scheme, CsiNet (autoencoder-based), and CsiFormer (transformer-based) under realistic channel conditions. Using throughput and bit error rate (BER) as performance metrics, the results demonstrate that CsiFormer achieves the best performance, but at the cost of substantially higher complexity.**

***Index Terms*—5G, Transformer, Deep Learning, CSI feedback.**

I. Introduction

Massive multiple-input multiple-output (MIMO) is a key technology for next-generation wireless communications (5G and 6G), where large antenna arrays at the base station enhance spectral efficiency and reliability [1]. Depending on deployment, antenna arrays may follow centralized or distributed architectures, but in all cases, accurate channel state information (CSI) is essential.

In frequency division duplex (FDD) systems, the user equipment (UE) estimates downlink CSI and feeds it back to the base station (BS). However, as antenna numbers scale to hundreds, the feedback overhead grows significantly, causing latency and performance degradation. Thus, efficient CSI feedback is indispensable for FDD-based massive MIMO.

Conventional approaches such as codebook-based feedback (RI, PMI, CQI) and vector quantization [2] face scalability issues in large antenna systems. To overcome this, deep learning-based CSI compression and reconstruction methods have emerged, including autoencoder-based CsiNet [3] and transformer-based CsiFormer [4]. Yet, most works emphasize reconstruction metrics like normalized mean squared error (NMSE), with less focus on transmission-level metrics such as throughput and bit error rate (BER).

In this paper, we evaluate codebook-based feedback [2], CsiNet [3], and CsiFormer [4] under a clustered delay line (CDL) channel model with multiple receive antennas. Results show that CsiFormer achieves the best performance in terms of throughput and BER among the considered schemes, although this comes at the cost of significantly higher complexity.

II. Transformer based CSI Feedback for 5G NR

A Transformer-based CSI feedback scheme, CsiFormer [4], effectively learns the correlation between antennas and subcarriers, and incorporates long-range dependency information to simultaneously improve compression efficiency and reconstruction performance. The received signal $y_{m,c}$ at the m-th received antenna through the c-th subcarrier can be expressed as follows:

$$y_{m,c} = \mathbf{h}_{m,c}^H \mathbf{v}_c x_c + z_{m,c}, \tag{1}$$

where $\mathbf{h}_{m,c}$ denotes the complex channel vector corresponding to the m-th received antenna and the c-th subcarrier, $\mathbf{v}_c$ represents the precoding vector, x_c is the transmitted data symbol, and $z_{m,c}$ denotes noise affecting the received signal. The notion $(\cdot)^H$ indicates the Hermitian transpose.

As shown in Fig. 1, the structure of CsiFormer [4] consists of an encoder that takes a CSI matrix as input and generates a compressed codeword, and a decoder that reconstructs the original CSI matrix from the codeword.

The encoder of CsiFormer is centered around the Convolutional Window Transformer Block (CWTB). The CSI input is first processed by a 5×5 convolutional layer to extract initial features. It is then divided into small patches, flattened into sequences, and enriched with position embeddings. The Window Transformer Block (WTB) performs self-attention within fixed-size windows to capture both efficiency and local dependencies. The output is transformed back into 2D through convolutional blocks and combined with residual connections from earlier features. Finally, a 3×3 convolution and a fully connected layer generate the compressed codeword.

The decoder receives the codeword generated by the encoder and transforms it into a feature map through a fully connected layer. This is followed by a 5×5 convolutional layer and a CWTB for refinement. The CWTB output is residually connected with the output of the preceding 5×5 convolution. Lastly, the features pass through a 3×3 convolutional layer, and a sigmoid activation function is applied to produce the reconstructed CSI matrix, normalized to the range $[0, 1]$.

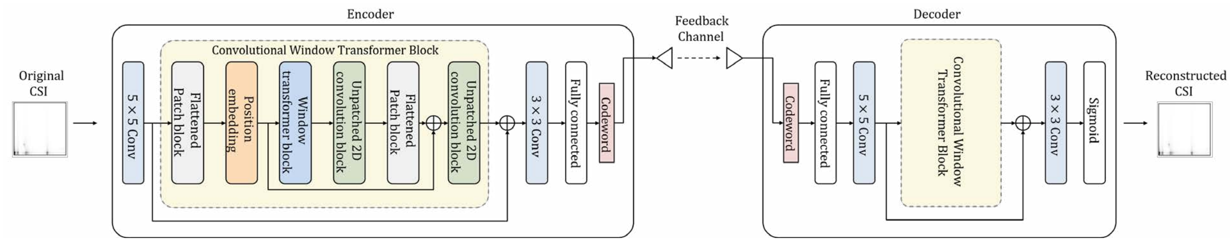
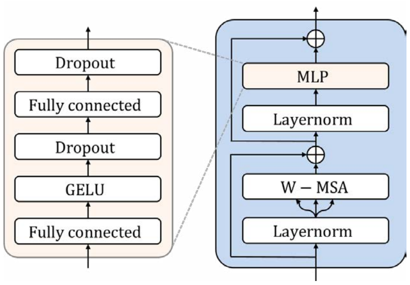
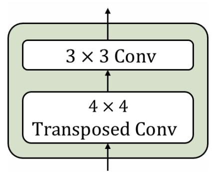


Fig. 1: Overall architecture of CsiFormer.



(a) Window Transformer Block.



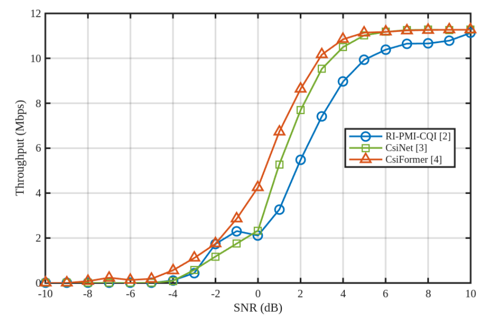
(b) Unpatched Convolution Block.

Fig. 2: Window Transformer Block and Unpatched Convolution Block of CsiFormer.

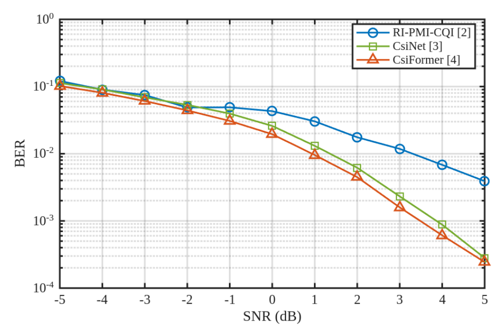
III. Evaluation Results

All the experiments were conducted under practical wireless channel conditions by applying the 5G NR CDL-A channel model, a MIMO configuration, and the MCS table. The modulation scheme was set to 16QAM, the code rate to 0.643, and the number of transmission layers to one. Performance evaluation was carried out over an SNR range from –5 dB to 5 dB, using throughput and BER as the performance metrics.

For training, the Adam optimizer was employed with an initial learning rate of 8×10^{-4}, and a polynomial decay scheduler was applied for up to 1000 epochs. Fig. 3 presents the throughput and BER performance of the 5G NR codebook-based CSI feedback scheme RI-PMI-CQI [2] and the deep learning-based CSI feedback schemes CsiNet [3] and CsiFormer [4], as a function of SNR. The results showed that CsiFormer achieved the best performance in both metrics, although there were regions where the performance gap was not significant.



(a) Throughput.



(b) BER.

Fig. 3: Throughput and BER performance comparison of CSI feedback schemes.

IV. Conclusion

In this paper, we compared the performance of the 5G NR codebook-based CSI feedback scheme (RI-PMI-CQI) [2] and the deep learning-based schemes CsiNet [3] and CsiFormer [4] under realistic wireless channel conditions. The evaluation was based on throughput and BER, and the results showed that CsiFormer achieved the best performance, although at the cost of significantly higher complexity. This highlighted the need for future research on CSI feedback schemes that could reduce complexity while maintaining high performance.

Acknowledgment

This study was supported by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-00794, Development of 3D Spatial Mobile Communication Technology.

References

[1] H. Lee, B. Lee, H. Yang, J. Kim, S. Kim, W. Shin, B. Shim, and H. V. Poor, "Towards 6G hyper-connectivity: Vision, challenges, and key enabling technologies," *J. Commun. Netw.*, vol. 25, no. 3, pp. 344–354, Jun. 2023.

[2] 3GPP, "NR; Physical layer procedures for data," Technical Specification (TS) 38.214, 3rd Generation Partnership Project (3GPP), 06 2025. Version 19.0.0.

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

[4] X. Bi, S. Li, C. Yu, and Y. Zhang, "A novel approach using convolutional transformer for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 1017–1021, May 2022.