A Survey on AI-Enhanced CSI Feedback in Cell-Free Massive MIMO

Seongjin Choi, Donghyun Lee, Junsuk Oh, Thanh Phung Truong, Seungchan Lee, Juyoung Kim and Sungrae Cho

School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea

Emails: {sjchoi, dhlee, jsoh, tptruong, sclee, jykim}@uclab.re.kr, srcho@cau.ac.kr

Abstract— **Massive MIMO has become a cornerstone of modern wireless communication, yet its performance in traditional cell-centric architectures remains constrained by inter-cell interference and feedback overhead. To address these limitations, Cell-Free Massive MIMO (CF-mMIMO) eliminates cell boundaries by allowing a large number of distributed access points (APs) to cooperatively serve users. However, the system's efficiency still depends critically on accurate and efficient Channel State Information (CSI) acquisition—particularly in Frequency Division Duplexing (FDD) systems where feedback overhead is substantial.**

This paper presents a comprehensive survey of recent advancements in AI-enhanced CSI feedback techniques tailored for CF-mMIMO systems. Specifically, we review five key research directions: *1)* deep convolutional compression, *2)* end-to-end learning for feedback and precoding, *3)* implicit feedback, *4)* predictive feedback using Transformers and meta-learning, and *5)* variable-rate and variable-length adaptive encoding. By leveraging deep learning, these approaches achieve significant reductions in feedback overhead while maintaining reconstruction accuracy and temporal adaptability. Furthermore, they open promising opportunities for dynamic feedback scheduling and distributed cooperation in CF-mMIMO networks, offering a scalable path toward the realization of AI-native 6G wireless systems.

Keywords—CSI Feedback, Cell-Free Massive MIMO, AI-Enhanced

I. INTRODUCTION

Massive MIMO technology has established itself as a core component of 5G and 6G networks, maximizing spatial multiplexing gains through large antenna arrays. However, in traditional cell-bound architectures, inter-cell interference and load imbalance continue to impose fundamental performance limitations.

To overcome these challenges, Cell-Free Massive MIMO (CF-mMIMO) has emerged, in which a large number of distributed Access Points (APs) cooperatively and simultaneously serve all users without predefined cell boundaries [1][2]. Unlike conventional architectures, CF-mMIMO adopts a user-centric topology wherein each user is dynamically surrounded by multiple APs. This mitigates cell-edge degradation and enhances macro-diversity and connectivity through coordinated AP transmission [3].While CF-mMIMO demonstrates remarkable advantages in terms of spatial diversity and uniform Quality of Service (QoS), its overall performance heavily depends on the accuracy of Channel State Information (CSI) acquisition. In particular, under the Frequency Division Duplexing (FDD) operation mode, the downlink CSI must be estimated at the User Equipment (UE) and fed back through the uplink channel, resulting in significant feedback overhead [4]. A promising direction to alleviate this issue is the use of AI-enhanced CSI feedback, which leverages Deep Learning (DL) and data-driven learning mechanisms to model the spatial and temporal structures of the channel, thereby enabling compression, prediction, and variable-rate feedback optimization in an adaptive manner.

This survey paper analyzes such AI-based CSI feedback technologies in the context of CF-mMIMO system requirements, providing a comprehensive overview of recent advancements—particularly focusing on Variable-Rate and Variable-Length learning-based feedback schemes that have gained significant traction in recent years.

II. BACKGROUND AND MOTIVATION

A. Evolution from mMIMO to Cell-Free Architectures

Cell-Free Massive MIMO extends the concept of distributed MIMO by deploying a large number of APs across a wide area, all connected to a Central Processing Unit (CPU) that coordinates their operations.

In this structure, each user is simultaneously served by multiple nearby APs, and their transmission signals are coherently managed by the CPU. As a result, performance disparities caused by cell boundaries are effectively eliminated.

Previous studies have demonstrated that CF-mMIMO can improve the average per-user throughput by more than an order of magnitude compared to small-cell networks, while maintaining robust service quality even under highly correlated shadowing conditions. These findings confirm that CF-mMIMO is a fundamental technology for realizing ultra-dense 6G wireless networks. This gain is largely attributed to the architectural shift from cell-centric to user-centric design, where spatially sparse APs cooperate to jointly serve all users across overlapping coverage zones. Additionally, the scalability and flexibility of CF-mMIMO are enhanced by clustering mechanisms and multi-CPU coordination schemes [5].

B. The Importance and Limitations of CSI Feedback

The theoretical gains of Massive MIMO rely on the assumption of perfect CSI. However, in practical systems, CSI must be periodically estimated and fed back, which imposes a heavy overhead in high-dimensional antenna and wideband environments.

In FDD systems, since channel reciprocity does not hold between downlink and uplink, the UE must estimate the downlink CSI and transmit it to the network via the uplink control channel. Conventional approaches such as codebook-based quantization or compressive sensing have been proposed to mitigate this burden by encoding the channel into a limited number of bits.

However, these methods assume fixed codebooks and static bit rates, making them poorly adaptive to environmental dynamics or varying channel complexity. Similar limitations have been identified in adaptive streaming and vehicular communication systems, where deep reinforcement learning (DRL)-based algorithms have been proposed to dynamically adjust bitrate and content delivery according to real-time channel states [5]. Consequently, under fast-varying or frequency-selective channels, substantial performance degradation becomes inevitable.

C. The Need for AI-Based CSI Feedback

To overcome these limitations, AI-based methods have recently gained significant attention. For instance, deep reinforcement learning approaches such as the Deep Deterministic Policy Gradient (DDPG) framework have shown that communication parameters can be optimized adaptively in dynamic wireless environments [6]. DL models can autonomously learn the spatial and frequency-domain correlations of CSI matrices, enabling more efficient compression and reconstruction than traditional analytical techniques. Moreover, within neural architectures, parameters such as bit rate, quantization precision, and feedback length can be dynamically adjusted in response to channel conditions, thereby eliminating the inefficiencies of fixed feedback schemes.

III. AI-BASED CSI FEEDBACK TECHNIQUES

Recent research on AI-based CSI feedback has focused on overcoming the limitations of fixed, codebook-based approaches by introducing learning-driven, adaptive frameworks. These studies can be broadly classified into five directions: *1)* deep convolutional compression, which employs autoencoders to efficiently encode high-dimensional CSI; *2)* end-to-end optimization, integrating feedback and precoding into a unified learning process; *3)* implicit feedback, directly inferring precoders without explicit CSI reconstruction; *4)* predictive feedback, leveraging temporal models such as Transformers and meta-learning to anticipate future CSI; and *5)* variable-rate and variable-length feedback, dynamically adjusting bit allocation based on channel complexity. Together, these methods demonstrate how AI can significantly reduce feedback overhead while maintaining reconstruction accuracy, thereby enhancing scalability and adaptability in Cell-Free Massive MIMO systems.

A. Deep Convolutional Compression

[7] proposed a framework called Distributed Deep Convolutional Compression (DeepCMC). This model employs a convolutional autoencoder to compress the spatial structure of the CSI and optimizes the rate–distortion balance in the reconstruction stage through end-to-end learning. By removing fully connected layers, the model achieves flexibility with respect to input size, and by incorporating arithmetic entropy coding, it further enhances bit efficiency. In its distributed extension, multiple users independently encode their local CSI, while the base station performs joint decoding by exploiting the spatial correlations among neighboring users. This cooperative strategy results in approximately 30% reduction in feedback bits and significant Normalized Mean Square Error (NMSE) improvement compared to conventional methods. Such a distributed compression and joint reconstruction structure aligns well with the cooperative multi-user nature of CF-mMIMO systems.

B. End-to-End Learning for Feedback and Precoding

[8] proposed an end-to-end optimization framework that integrates channel estimation, quantization, feedback, and precoding into a unified learning process. Unlike conventional methods where CSI compression and precoding optimization are performed separately, the proposed network jointly optimizes both stages by directly maximizing the sum-rate during training. This approach achieves 10–15% higher throughput under the same feedback bit budget compared to traditional techniques. In CF-mMIMO, such end-to-end learning allows multiple APs to share learned latent features, enabling the central CPU to efficiently perform collaborative beamforming and resource allocation.

C. Implicit CSI Feedback

[9] proposed a DL–based implicit CSI feedback framework that replaces conventional codebook-based explicit feedback with a neural mapping between uplink and downlink channels. Instead of transmitting the downlink channel matrix or its quantized representation, the proposed model learns to directly infer the precoding matrix indicator (PMI) from the uplink channel features, thereby bypassing explicit CSI transmission. By exploiting spatial and frequency correlations across subcarriers, the network achieves 25–40% feedback overhead reduction compared to 3GPP Type-I/II codebook methods, while maintaining comparable or higher sum-rate performance. The study further demonstrates that this implicit mapping generalizes across antenna configurations and user mobility scenarios, highlighting the robustness of data-driven inference over handcrafted quantization. For CF-mMIMO, where each distributed AP possesses localized CSI, such implicit feedback allows APs to directly generate user-specific precoders with minimal coordination latency, offering a highly efficient and scalable approach to cooperative beamforming.

D. Predictive CSI Feedback

To alleviate the performance degradation caused by rapid channel variations, [10] developed a Transformer-based channel prediction framework that leverages the self-attention mechanism to capture long-term temporal dependencies. Unlike recurrent networks (Recurrent Neural Network, RNN or Long Short-Term Memory, LSTM), which are limited by sequential processing, the Transformer can model global time correlations in parallel, achieving more than 20% NMSE improvement in high-mobility scenarios. The study demonstrates that accurate temporal prediction can effectively extend the channel coherence time, thereby enabling longer feedback intervals without compromising reliability in densely deployed CF-mMIMO systems.

[11] introduced a Meta-Learning and Deep Denoising framework for CSI prediction, addressing the challenge of limited training data in realistic deployments.

By combining Model-Agnostic Meta-Learning (MAML) with denoising autoencoders, the model rapidly adapts to new propagation environments using only a few fine-tuning samples. This approach is particularly well suited for CF-mMIMO networks, where heterogeneous AP placements and diverse channel statistics make global retraining impractical. Together, these predictive feedback methods enable time-adaptive CSI acquisition, significantly reducing the feedback frequency and computational burden in large-scale distributed systems.

E. Variable-Rate and Variable-Length CSI Feedback

To overcome the inefficiency of fixed-rate feedback encoders, [12] proposed the Learning Variable-Rate Codes for CSI Feedback framework, which dynamically adjusts feedback bit allocation according to channel complexity. The method employs a rate-adaptive quantizer trained to minimize distortion under a variable feedback budget, allowing each sample to use a different number of bits depending on its reconstruction difficulty. This adaptive encoding strategy achieves over 30% average bit reduction while maintaining the same NMSE as fixed-rate baselines. By integrating reinforcement-guided bit selection and latent-space regularization, the model provides smooth rate–distortion trade-offs that are essential for dynamic feedback scheduling.

Building on this idea, [13] introduced a machine learning–based variable-length feedback mechanism that combines Principal Component Analysis (PCA) with bit-allocation optimization. The scheme assigns quantization bits based on the variance of principal components, allowing for per-user or per-link adaptation within the total feedback budget. Experimental results show up to 40% bit savings and over 1 dB NMSE improvement, demonstrating the effectiveness of data-driven bit allocation in high-dimensional MIMO environments. In CF-mMIMO networks, such variable-rate and variable-length feedback can be extended to the fronthaul and backhaul domains, where the central CPU dynamically redistributes feedback resources among APs according to instantaneous channel conditions and cooperation levels. This adaptability makes AI-based rate control a crucial component for scalable and efficient cell-free network operation.

IV. Conclusion

This survey summarized the recent progress of AI-based CSI feedback techniques for Cell-Free Massive MIMO systems, emphasizing how data-driven learning overcomes the inefficiencies of traditional fixed codebook approaches. Deep convolutional compression enables compact CSI representation, end-to-end learning jointly optimizes feedback and precoding, and implicit feedback removes explicit CSI transmission through uplink–downlink neural mapping. Moreover, predictive feedback using Transformers and meta-learning effectively handles channel variation and mobility, while variable-rate and variable-length encoding dynamically allocate feedback bits according to channel complexity. Overall, these advancements demonstrate that integrating AI into CSI feedback design can greatly improve spectral efficiency, scalability, and adaptability in CF-mMIMO networks. The combination of distributed cooperation and intelligent feedback establishes CF-mMIMO as a key enabler for 6G network intelligence. Future research will likely focus on federated and reinforcement learning–based feedback adaptation and the development of lightweight neural architectures suitable for large-scale, edge-deployed CF-mMIMO systems.

Acknowledgment

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2026-RS-2022-00156353, 50% / IITP-2026-RS-2023-00258639, 50%) grants funded by the Korea government (Ministry of Science and ICT).

References

[1] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai and L. Hanzo, "Cell-Free Massive MIMO: A New Next-Generation Paradigm," *IEEE Access*, vol. 7, pp. 99878-99888, 2019.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson and T. L. Marzetta, "Cell-Free Massive MIMO Versus Small Cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834-1850, March 2017.

[3] M. Ajmal, A. Siddiqa, B. Jeong, J. Seo, and D. Kim, "Cell-free massive multiple-input multiple-output challenges and opportunities: A survey, " *ICT Express*, vol. 10, no. 1, pp. 194-212, February 2024.

[4] 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 Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341-1365, October 2008.

[5] B. Tamiru, J. Jee, and H. Park, "Channel estimation and pilot reduction for mmWave massive MIMO systems using deep neural networks, " *ICT Express*, vol. 10, no. 4, pp. 798-803, August 2024.

[6] D. Kwon, J. Jeon, S. Park, J. Kim and S. Cho, "Multiagent DDPG-Based Deep Learning for Smart Ocean Federated Learning IoT Networks," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9895-9903, Oct. 2020.

[7] M. B. Mashhadi, Q. Yang and D. Gündüz, "Distributed Deep Convolutional Compression for Massive MIMO CSI Feedback," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2621-2633, April 2021.

[8] F. Sohrabi, K. M. Attiah and W. Yu, "Deep Learning for Distributed Channel Feedback and Multiuser Precoding in FDD Massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4044-4057, July 2021.

[9] M. Chen, J. Guo, C. -K. Wen, S. Jin, G. Y. Li and A. Yang, "Deep Learning-Based Implicit CSI Feedback in Massive MIMO," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 935-950, Feb. 2022.

[10] H. Jiang, M. Cui, D. W. K. Ng and L. Dai, "Accurate Channel Prediction Based on Transformer: Making Mobility Negligible," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2717-2732, Sept 2022.

[11] H. Kim, J. Choi and D. J. Love, "Massive MIMO Channel Prediction Via Meta-Learning and Deep Denoising: Is a Small Dataset Enough?," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9278-9290, Dec 2023.

[12] H. Kim, H. Kim and G. De Veciana, "Learning Variable-Rate Codes for CSI Feedback," in *Proc. of GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Rio de Janeiro, Brazil, pp. 1435-1441, 2022.

[13] M. Nerini, V. Rizzello, M. Joham, W. Utschick and B. Clerckx, "Machine Learning-Based CSI Feedback With Variable Length in FDD Massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 2886-2900, May 2023.