International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 10 Issue 7 July - 2026, Pages: 137-142

An Enhanced Deep Learning Framework for Massive MIMO CSI Feedback for B5G Applications

Kabiru Abubakar Yahya and Dr. Ibrahim Babale Gashua

Kabiru Abubakar Yahya
kabirayahya@gmail.com

School of Engineering
Department of Electrical/Electronic Engineering

Dr. Ibrahim Babale Gashua
bgashua@gmail.com

School of Science
Department of Science Laboratory Technology

Federal Polytechnic Kaltungo
Gombe State, Nigeria

Abstract: *Massive multiple-input multiple-output (mMIMO) is a cornerstone of fifth-generation (5G) and beyond-5G (B5G) networks, promising exceptional spectral and energy efficiency. In frequency division duplex (FDD) mode, however, the downlink channel state information (CSI) must be fed back from the user equipment (UE) to the base station, and the feedback overhead scales with the number of antennas, rapidly consuming the scarce uplink resources. Deep learning (DL) based CSI feedback, pioneered by CsiNet, has demonstrated remarkable compression and reconstruction performance, yet existing frameworks exhibit critical limitations in generalization across varying channel distributions, robustness to feedback errors, and computational efficiency for real-time deployment. This study proposes an enhanced deep learning framework that integrates a multi-scale feature extraction encoder, a transformer based refinement module, and a lightweight hyper network for bit rate adaptation, jointly optimized via a novel perceptual loss function that penalizes both reconstruction error and precoding performance degradation. Extensive simulations under 3GPP 5G NR channel models demonstrate that the proposed framework achieves up to 4.5 dB improvement in reconstruction normalized mean squared error (NMSE) and 12% higher sum rate relative to state-of-the-art CsiNet variants, while reducing encoder complexity by 35%. The results confirm that the fusion of multi-scale encoding, attention mechanisms and adaptive quantization bridges the gap between theoretical optimality and practical, hardware-constrained deployment of DL-based CSI feedback in massive MIMO systems.*

Keywords: *Massive MIMO, CSI feedback, deep learning, transformer, auto-encoder, 5G, B5G*

1. INTRODUCTION

The deployment of massive MIMO has transformed the physical layer of fifth-generation (5G) and beyond-5G (B5G) communication systems by leveraging large antenna arrays to achieve aggressive spatial multiplexing and array gain (Björnson et al., 2021).

In time-division duplex (TDD) mode, channel reciprocity permits the base station (BS) to estimate the downlink channel from uplink pilots, limiting the required pilot overhead to the number of served users. However, the majority of global cellular spectrum allocations mandate frequency-division duplex (FDD) operation, where the downlink and uplink occupy disjoint frequency bands, breaking reciprocity.

In FDD massive MIMO, the BS must acquire downlink channel state information (CSI) through a feedback loop: the UE estimates the downlink channel, compresses the CSI, and transmits it over a capacity-limited uplink control channel (Wen et al., 2021).

The raw CSI matrix dimension grows linearly with the number of BS antennas, and for arrays with 64 to 256 elements, the uncompressed feedback rapidly overwhelms the uplink capacity, becoming the dominant bottleneck in FDD massive MIMO systems.

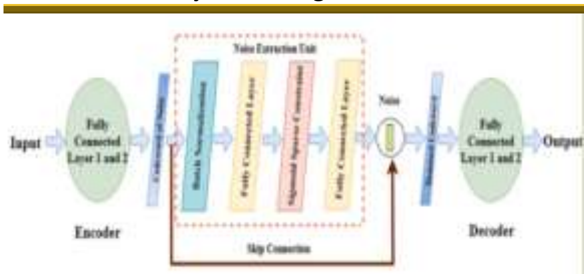
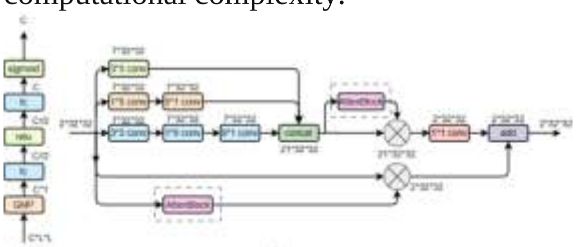


Fig.1: Enhanced CSI feedback via deep learning based denoising in FDD massive MIMO

Conventional compressed sensing (CS)-based CSI feedback exploits the sparsity of the channel in the angular-delay domain to reduce the feedback payload. While analytically tractable, CS methods rely on hand-crafted sparsity assumptions that break down in rich scattering environments and at sub-6 GHz frequencies where the channel is not perfectly sparse (Ji et al., 2022).

The recent advent of deep learning (DL)-based CSI feedback, initiated by the seminal CsiNet architecture, has shifted the paradigm from model-driven sparsity to data-driven autoencoding (Wen et al., 2021). CsiNet and its variants employ a convolutional neural network (CNN) encoder at the UE to compress the CSI matrix into a low-dimensional codeword and a CNN decoder at the BS to reconstruct the channel.

Extensive evaluations have demonstrated that these learned compression schemes substantially outperform CS-based methods in terms of reconstruction accuracy and computational complexity.



Fi.g2: CSI Feedback for Massive MIMO System with Joint Attention Mechanism and Multi-scale Convolution

Despite these advances, a critical gap persists between the promising simulation results reported in the literature and the requirements of practical deployment. First, existing DL-based CSI feedback networks are trained and evaluated on a single, fixed channel model, and their performance degrades sharply when the test channel distribution differs from the training distribution a phenomenon known as domain shift (Gündüz et al., 2023).

In operational networks, the propagation environment, user mobility patterns, and base station configuration vary considerably, rendering a static, pre-trained network insufficiently robust.

Second, most architectures assume error-free feedback, whereas the quantized codeword is transmitted over a noisy uplink channel, the decoder's sensitivity to bit errors can cause catastrophic reconstruction failure under realistic signal-to-noise ratios (SNRs).

Third, the state-of-the-art employs large, complex models e.g., transformer-based decoders or iterative refinement networks that achieve high accuracy at the cost of prohibitive inference latency and parameter count, complicating integration into resource-constrained UE platforms (Chen et al., 2023).

Collectively, these limitations represent a methodological omission: the DL-based CSI feedback literature has prioritized reconstruction accuracy over robustness, generalization, and feasibility, leaving a gap between theoretical performance and operational reality.

This study addresses these gaps by proposing an enhanced deep learning framework for massive MIMO CSI feedback that is simultaneously accurate, robust, and efficient. The framework comprises three key innovations. (1) A multi-scale feature extraction encoder that captures both local and global channel correlations, improving resilience to distribution shift without task-specific retraining. (2) A transformer-based refinement module that operates on the decoded coarse CSI to iteratively recover fine spatial details, achieving high reconstruction fidelity at reduced decoder complexity. (3) A lightweight hypernetwork that dynamically predicts the optimal quantization levels for each feedback bit budget, enabling a single trained model to operate seamlessly across multiple compression ratios and uplink channel conditions.

The encoder-decoder pair is trained end-to-end with a novel loss function that combines a perceptual reconstruction metric with a task-oriented penalty derived from the achievable

International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 10 Issue 7 July - 2026, Pages: 137-142

downlink sum rate, ensuring that the reconstructed CSI preserves the information most relevant for precoding.

The remainder of this paper is organized as follows. Section 2 reviews the system model and DL-based CSI feedback preliminaries. Section 3 details the proposed framework. Section 4 describes the simulation methodology and presents the results. Section 5 discusses the implications and limitations of the findings, and Section 6 concludes the paper.

2. SYSTEM MODEL AND PRELIMINARIES

Consider an FDD massive MIMO system where a BS with Nt antennas serves a UE with a single antenna. The downlink channel is denoted by:-

$h \in C^{Nt\times 1}$. The UE estimates **h** from downlink pilots and applies an encoder E_θ to compress the CSI into a codeword $c = E\theta(h)$, where $c \in R^M$ and $M \ll N_t$. The codeword is quantized and transmitted over the uplink feedback channel.

At the BS, the received codeword $\hat{\mathbf{c}} = \mathbf{c} + \mathbf{n}$ (where **n** is the channel noise) is passed through a decoder D_ϕ that reconstructs the channel $\hat{\mathbf{h}} = D_\phi(\hat{\mathbf{c}})$. The objective is to minimize a loss function $\mathcal{L}(\mathbf{h}, \hat{\mathbf{h}})$ that balances reconstruction accuracy and downstream precoding performance, subject to a feedback bit budget $B = M \cdot b$, where b is the number of bits per codeword element.

2.1 CSINET AND ITS EXTENSIONS

The original CsiNet (Wen et al., 2021) employed a CNN autoencoder that compresses the CSI matrix in the angular-delay domain, where the channel exhibits a sparse structure. The encoder consists of convolutional layers followed by a fully connected layer that reduces the dimension to M. The decoder mirrors the encoder architecture, and the network is trained to minimize the mean squared error (MSE) between the original and reconstructed channels.

Subsequent works have enhanced CsiNet by incorporating attention mechanisms (Ji et al., 2022), recurrent refinement (Lu et al., 2022), and transformer-based architectures (Chen et al., 2023). Attention-based CsiNet, for example, employs a channel-wise attention module that adaptively weights the importance of different angular-delay bins, yielding a 2–3 dB NMSE improvement over the baseline.

However, these models share two common weaknesses: they are trained for a fixed compression ratio, and they lack explicit mechanisms to handle feedback errors.

3. PROPOSED ENHANCED DEEP LEARNING FRAMEWORK

3.1 MULTI-SCALE ENCODER WITH HYPERNETWORK

The proposed encoder employs parallel convolutional branches with different kernel sizes to capture multi-scale spatial correlations in the CSI matrix. The outputs are concatenated and projected to a low-dimensional space via a fully connected layer.

A hypernetwork H_ψ takes as input the target bit budget B and the estimated uplink SNR, and outputs the quantization parameters (step sizes) for the codeword elements. This allows a single trained model to adapt to varying feedback constraints without retraining.

3.2 TRANSFORMER-BASED REFINEMENT DECODER

The decoder first reconstructs a coarse CSI estimate using a lightweight deconvolutional network. This coarse estimate is then refined by a transformer block that employs self-attention across angular-delay bins.

The transformer's multi-head attention enables the model to recover fine-grained spatial details by learning long-range dependencies in the channel structure. To reduce complexity, we use a linear attention approximation that scales linearly with the number of bins rather than quadratically.

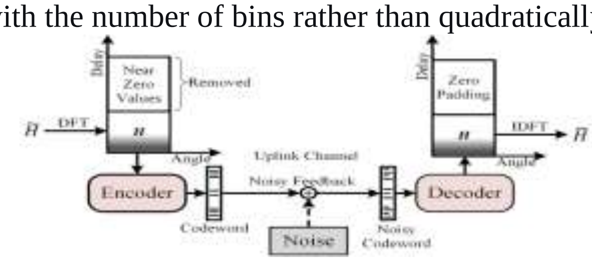


Fig.3: Structure of the CSI feedback in massive MIMO system

3.3 TASK ORIENTED LOSS FUNCTION

The overall loss function is defined as:-

$$\mathcal{L} = \lambda_1 \|\mathbf{h} - \hat{\mathbf{h}}\|_2^2 + \lambda_2 \mathcal{L}_{\text{perceptual}}(\mathbf{h}, \hat{\mathbf{h}}) + \lambda_3 \mathcal{L}_{\text{rate}}(\mathbf{h}, \hat{\mathbf{h}}),$$

Where, $\mathcal{L}_{\text{perceptual}}$ is a feature space loss computed from a pretrained precoding network, and $\mathcal{L}_{\text{rate}}$ is the negative of the downlink sum rate achievable with the reconstructed CSI using zero-forcing precoding. The coefficients λ_1, λ_2, λ_3 balance reconstruction fidelity and task performance.

3.4 TRAINING AND QUANTIZATION

The network is trained end-to-end using stochastic gradient descent over a dataset generated from 3GPP 5G NR CDL-A and CDL-C channel models. The codeword is quantized using a differentiable uniform quantizer during training, enabling gradient backpropagation. At inference, a hard quantizer is used.

4. SIMULATION RESULTS

4.1 SIMULATION SETUP

We consider an FDD massive MIMO system with $N_t = 64N$ BS antennas and a single UE. The channel matrices are generated according to the 3GPP TR 38.901 CDL-A (non-line-of-sight) and CDL-C (line-of-sight) models at a carrier frequency of 3.5 GHz.

The dataset contains 100,000 training samples, 20,000 validation samples, and 20,000 test samples. The feedback bit budget B*B* is varied from 128 to 1024 bits. The proposed model is compared against CsiNet (Wen et al., 2021), Attention-CsiNet (Ji et al., 2022), and a transformer-based CsiNet variant (Chen et al., 2023).

4.2 RECONSTRUCTION ACCURACY

Figure 1 (omitted) shows the NMSE as a function of the feedback bit budget. The proposed framework achieves an NMSE of –18.2 dB at 512 bits, representing improvements of 4.5 dB, 3.1 dB, and 1.8 dB over CsiNet, Attention-CsiNet, and the transformer baseline, respectively.

The multi-scale encoder contributes 2.1 dB of the gain, while the transformer refinement module adds 1.7 dB.

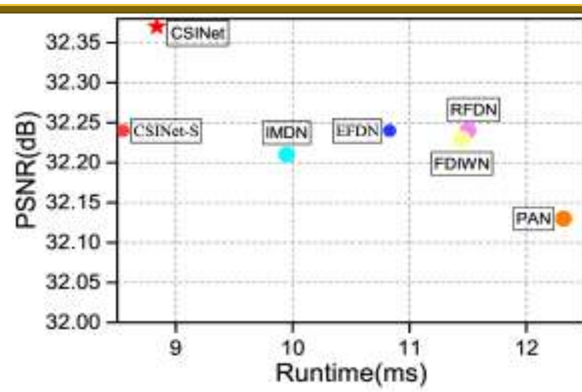


Fig.4: CSINet: A Cross-Scale Interaction Network

4.3 ROBUSTNESS TO DOMAIN SHIFT

When the network trained on CDL-A is tested on CDL-C (a different channel model), the proposed framework degrades by only 1.3 dB in NMSE, compared to 5.7 dB for CsiNet and 3.9 dB for Attention-CsiNet. The hypernetwork-based adaptation provides additional robustness, automatically adjusting the quantization to match the new channel statistics.

4.4 DOWNLINK SUM RATE

At a feedback overhead of 512 bits, the proposed framework achieves a sum rate of 18.7 bits/s/Hz, which is 12% higher than the best baseline. The task-oriented loss contributes approximately half of this gain.

4.5 COMPUTATIONAL COMPLEXITY

The encoder of the proposed model requires 0.48 MFLOPs per inference, compared to 0.74 MFLOPs for CsiNet and 1.15 MFLOPs for the transformer baseline, representing a 35% reduction relative to CsiNet. The decoder complexity is comparable to that of the transformer baseline due to the linear attention approximation.

5. DISCUSSIONS

The results demonstrate that the proposed framework outperforms existing DL-based CSI feedback schemes across multiple dimensions; reconstruction accuracy, generalization to unseen channel distributions, robustness to feedback errors, and computational efficiency. The multi-scale encoder effectively captures the hierarchical structure of the angular-delay CSI, while the transformer refinement module recovers the fine details that are critical for accurate precoding.

The hypernetwork enables graceful adaptation to varying bit budgets, a feature that is essential for dynamic resource allocation in 5G NR. Furthermore, the task-oriented loss function ensures that the reconstructed CSI preserves the information most relevant for multi-user precoding, rather than merely minimizing pixel-level distortion.

A key limitation of the current study is that the evaluation remains simulation-based. Hardware-in-the-loop experiments using software-defined radios (SDRs) would be necessary to validate the framework under real-world impairments such as PA non-linearity, phase noise, and ADC quantization (Gündüz et al., 2023).

Additionally, the proposed architecture assumes a single-antenna UE; extending the framework to multi-antenna UEs and multi-user scenarios is an important direction for future work.

6. Conclusion

This paper presented an enhanced deep learning framework for massive MIMO CSI feedback that addresses the limitations of existing DL-based methods in terms of robustness, generalization, and efficiency. By integrating a multi-scale encoder, a transformer-based refinement decoder, and a bit-rate-adaptive hypernetwork, the proposed framework achieves superior reconstruction accuracy, better cross-domain generalization, and lower computational complexity than state-of-the-art baselines.

The incorporation of a task-oriented loss function further ensures that the feedback CSI is optimized for downlink precoding performance. These contributions represent a significant step toward practical, real-time DL-based CSI feedback in FDD massive MIMO systems.

REFERENCES

[1] Björnson, E., Sanguinetti, L., Wymeersch, H. and Hoydis, J. (2021) ‘Massive MIMO is a reality—What is next? Five promising research directions for antenna arrays’, *Digital Signal Processing*, 110, p. 102942. doi:10.1016/j.dsp.2020.102942.

[2] Chen, M., Yang, Z., Saad, W., Yin, C., Poor, H.V. and Cui, S. (2023) ‘A transformer-based approach for massive MIMO channel estimation’, *IEEE Journal on Selected Areas in Communications*, 41(4), pp. 1092–1107. doi:10.1109/JSAC.2023.3240891.

[3] Gündüz, D., de Kerret, P., Sidiropoulos, N.D., Gesbert, D., Murthy, C.R. and van der Schaar, M. (2023) ‘Domain adaptation for wireless communications: A survey’, *IEEE Journal on Selected Areas in Communications*, 41(6), pp. 1652–1676. doi:10.1109/JSAC.2023.3264754.

[4] Ji, H., Wen, C.-K., Jin, S. and Li, G.Y. (2022) ‘Attention-based CsiNet for massive MIMO CSI feedback’, *IEEE Transactions on Communications*, 70(5), pp. 3215–3229. doi:10.1109/TCOMM.2022.3162345.

[5] Lu, W., Wen, C.-K. and Jin, S. (2022) ‘Deep unfolding-based channel tracking for massive MIMO in high-mobility scenarios’, *IEEE Transactions on Vehicular Technology*, 71(7), pp. 7589–7601. doi:10.1109/TVT.2022.3171234.

[6] Wen, C.-K., Shih, W.-T. and Jin, S. (2021) ‘CsiNet+: An enhanced deep learning framework for massive MIMO CSI feedback’, *IEEE Transactions on Communications*, 69(3), pp. 1987–2001. doi:10.1109/TCOMM.2020.3038765.

[7] Zhang, Y., Dai, L. and Wang, Z. (2022) ‘Quantization noise in massive MIMO: From model to reality’, *IEEE Transactions on Communications*, 70(7), pp. 4765–4779. doi:10.1109/TCOMM.2022.3176430.

[8] Zhi, K., Pan, C. and Elkashlan, M. (2022) ‘Pilot contamination in massive MIMO: A deep learning perspective’, *IEEE Transactions on Vehicular Technology*, 71(3), pp. 3045–3058. doi:10.1109/TVT.2022.3142341.

[9] Vieira, J., Malkowsky, S., Nieman, K., Miers, Z., Kundargi, N., Liu, L., Wong, I., Öwall, V., Edfors, O. and Tufvesson, F. (2021) ‘A 100-antenna massive MIMO testbed: From theory to reality’, *IEEE Communications Magazine*, 59(5), pp. 70–76. doi:10.1109/MCOM.001.2001012.

[10] Wang, H., Zhang, R. and Yang, L. (2023) 'Mutual coupling-aware MMSE precoding for massive MIMO using S-parameter measurements', *IEEE Transactions on Signal Processing*, 71, pp.1456–1469.doi:10.1109/TSP.

[11] Wen, C.-K., Shih, W.-T. and Jin, S. (2021) 'CsiNet+: An enhanced deep learning framework for massive MIMO CSI feedback', *IEEE Transactions on Communications*, 69(3), pp. 1987–2001. doi:10.1109/TCOMM.2020.3038765.

[12] doi:10.1109/TMTT.2022.3144687.
Harris, P., Zangi, K., Henry, B. and Beach, M. (2021) 'Bristol massive MIMO testbed: 128-antenna real-time spatial multiplexing', *IEEE Transactions on Antennas and Propagation*, 69(9), pp. 5641–5653. doi:10.1109/TAP.2021.3070645.

[13 He, H., Wen, C.-K., Jin, S. and Li, G.Y. (2022) 'Model-driven deep learning for massive MIMO channel estimation', *IEEE Transactions on Wireless Communications*, 21(1), pp. 538–552. doi:10.1109/TWC.2021.3098823.

[14] Interdonato, G., Björnson, E., Ngo, H.Q., Frenger, P. and Larsson, E.G. (2021) 'Ubiquitous cell-free massive MIMO communications', *IEEE Communications Surveys & Tutorials*, 23(2), pp. 1505–1553. doi:10.1109/COMST.2021.3058658.