

Performance Analysis of a Novel Lightweight Transformer-Based CSI Feedback Framework for 5G-Advanced Massive MIMO Systems

RAHUL PAL

rahulpal@manit.ac.in

Maulana Azad National Institute of Technology Bhopal <https://orcid.org/0000-0002-4384-5156>

Research Article

Keywords: Massive MIMO, CSI Feedback, Precoder, Deep Learning

Posted Date: July 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10270318/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Performance Analysis of a Novel Lightweight Transformer-Based CSI Feedback Framework for 5G-Advanced Massive MIMO Systems

Rahul Pal^{1*}

^{1*}Electronics & Communication Engineering, Maulana Azad National Institute of Technology Bhopal, Street, Bhopal, 462003, M. P., India.

Corresponding author(s). E-mail(s): rahulpal@manit.ac.in;

Abstract

This paper introduces TPCNet, a novel lightweight transformer-based framework for channel state information (CSI) feedback that directly estimates the precoder, eliminating the need for explicit CSI reconstruction in massive MIMO-OFDM systems. TPCNet adopts an autoencoder–autodecoder architecture in which the encoder integrates, for robust feature learning, a multi-scale residual feature extraction network (MRFEN), an attention-based transformer block (TB), and a spatial resolution network (SRN) to generate compact codewords (CWs), while the decoder employs a TB, an SRN, and a feature refinement network (FRN) to reconstruct the precoder corresponding to the dominant transmission layers. By generating a single CW irrespective of the number of receive antennas, TPCNet achieves an inherent reduction in feedback overhead while eliminating the scalability issue associated with increasing receive antennas. Extensive simulations using the CDL-C MIMO channel model compliant with the 3GPP standards demonstrate that TPCNet consistently outperforms well-established CSI feedback approaches, providing performance gains of approximately 0.5–2 dB and reducing computational complexity by nearly 25% in terms of FLOPs compared with the leading transformer-based benchmark. Moreover, TPCNet achieves a superior accuracy–complexity trade-off across a wide range of compression ratios, providing an efficient, low-complexity, and scalable CSI feedback approach suitable for practical 5G-Advanced and future 6G deployments.

Keywords: Massive MIMO, CSI Feedback, Precoder, Deep Learning

1 Introduction

Massive multi-input multi-output (MIMO) enables 5G-Advanced wireless communication systems to achieve peak data rates up to 10 Gbps [1–3]. To enable enhanced beamforming and improved multiuser interference mitigation, such systems deploy a massive antenna array, typically consisting of 32 to 1024 radiating elements, at the access point (AP) [2]. To fully exploit these gains, accurate CSI is required at the AP. In Rel-18, downlink CSI in frequency-division duplex (FDD) mode is estimated using pilot sequences at the user terminal (UT) and shared with the AP via a feedback channel, resulting in significant feedback overhead [4–8]. Feedback schemes Type-I, -II, and enhanced Type-II are recommended by 3GPP to reduce this overhead by reporting compressed CSI components. However, their reliance on predefined structures limits system performance.

To address these limitations, Deep learning (DL) has recently emerged as an effective solution for CSI feedback. These schemes employ an autoencoder–autodecoder framework, where the UT generates a CW using the encoder, while the AP regenerates the CSI from the CW using the decoder for precoder design. Both networks share the same inputs and outputs, and they are jointly trained. The primary DL-based CSI feedback scheme is CsiNet [10]. First, the channel is translated into the angle-delay (AD) domain. This translation enables a sparse representation of the CSI. As a result, only a small fraction of channel entries pose significant energy. This sparsity enables to effectively exploit spatial correlations for feature extraction and accurate CSI reconstruction. To further leverage temporal correlation, CsiNet has been extended to CsiNet-LSTM [11] and RecCsiNet [12]. In CsiNet-LSTM [11], a long short-term memory (LSTM) network identifies the temporal features in the channel. On the other hand, a recurrent neural network (RNN) is integrated into the encoder as well as the decoder in RecCsiNet [12], which performs comparably to CsiNet-LSTM with a smaller number of training parameters. Next, both the encoder and decoder in CRNet [13] are deployed with hierarchical feature extraction blocks based on a multi-resolution architecture, which achieves higher gain than CsiNet [10].

DCRNet [14] and DeConvD-CRNet [15], motivated by CsiNet [10] and CRNet [13], obtain a broader receptive field while keeping the kernel dimensions unchanged by exploiting the concept of dilated convolutions. This allows for capturing wider spatial dependencies by integrating information from non-adjacent channel dimensions. Specifically, DCRNet [14] emphasizes receptive field expansion to enhance reconstruction performance at high CR, whereas DeConvD-CRNet [15] integrates dilated reconstruction with residual and inception modules to improve multi-scale feature learning. Due to their ability to extract spatially distant correlations available in the input tensor, transformer architectures [16–20] have become increasingly popular in DL-driven approaches. Two transformer-assisted CSI feedback frameworks [16], namely CsiTransformer and MixedCsiNet, have been proposed. By leveraging the strong feature extraction and long-range dependency, these schemes achieve improved reconstruction performance compared with the CsiNet [10]. Specifically, CsiTransformer provides superior CSI recovery accuracy, whereas MixedCsiNet balances the trade-off between computational efficiency and reconstruction accuracy more effectively. Motivated by Google’s transformer architecture, TransNet [17] deploys a

dual-layer transformer, where the decoder output is incorporated into the encoder to enhance attention-guided reconstruction for refined CSI. Whereas CsiFormer [18] employs a convolution-assisted transformer architecture that effectively models global contextual relationships using residual connections to effectively fuse self-attention features and a convolution-based window transformer. Furthermore, an efficient CSI feedback framework build upon a parallel self-attention transformer architecture was proposed in [19] to strengthen feature representation while preserving low computational complexity, thus improving CSI reconstruction accuracy. Along similar lines, SCANet [20] incorporates both spatial and channel attention mechanisms into a lightweight CSI feedback architecture. Despite its significantly reduced complexity, SCANet achieves CSI reconstruction performance comparable to that of TransNet [17], while lowering the computational complexity by 85.76%.

We propose TPCNet, a novel transformer-based architecture that enables a lightweight CSI feedback framework. It is specifically designed to estimate the precoder for massive MIMO-OFDM systems instead of reconstructing the explicit CSI. The framework comprises an autoencoder and an encoder, both of which comprise TB. Apart from TBs, there are a few additional components. The encoder additionally integrates an MRFEN to capture directional and non-uniform correlations and an SRN to enhance spatial representation, while the decoder employs both an SRN and an FRN to reconstruct the precoder from the received CW accurately. Compared with the baseline schemes, a single CW is generated in this framework irrespective of the number of radiati elements at the UT. Thus, regardless of the number of UT radiating elements. The proposed framework maintains the same feedback overhead. Numerical results validate the proposed framework through the 3GPP-standardized CDL-C channel model for MIMO systems across a wide range of CRs.

2 System Model

We assume FDD transmission in a massive MIMO-OFDM system. The setup comprises an AP and a UT. The AP employs a two-dimensional antenna array consisting of $N_t = N_{tx} \times N_{ty}$ elements, where each radiating element in the array is dual-polarized with vertical and horizontal polarizations. On the other hand, the UT is configured with $N_r = N_{rx} \times N_{ry}$ dual-polarised radiating elements. The resources are structured as a three-dimensional resource grid across the time-frequency and spatial domains for the OFDM system, comprising N_l data streams, N_{sc} subcarriers, and 14 OFDM symbols (i.e., one slot). The available bandwidth is partitioned into N_{RB} resource blocks (RBs), each comprising 12 subcarriers [6]. For data transmission, the total number of subcarriers are expressed as $N_{sc} = 12N_{RB}$. Based on this resource allocation, the received signal to the k^{th} subcarrier and n^{th} OFDM symbol is represented as

$$\mathbf{y}_{n,k} = \mathbf{H}_{n,k} \mathbf{V}_{n,k} \mathbf{s}_{n,k} + \mathbf{w}_{n,k}, \quad (1)$$

where $\mathbf{H}_{n,k} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix in time-frequency domain for the k^{th} subcarrier of the n^{th} OFDM symbol. The precoder matrix $\mathbf{V}_{n,k} \in \mathbb{C}^{N_t \times N_l}$ is used to linearly precode the information vector $\mathbf{s}_{n,k} \in \mathbb{C}^{N_l \times 1}$ is such a way that it satisfies

the normalized transmit power constraint $\mathbb{E}\{\|\mathbf{V}_{n,k}\mathbf{s}_{n,k}\|^2\} = 1$. Here, N_l denotes the number of spatial data streams. The noise vector $\mathbf{w}_{n,k} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{N_r})$ models additive white Gaussian noise (AWGN) at the receiver.

3 DL-based CSI Feedback

To obtain optimal system performance, the AP computes the precoder. On the other hand, the UT computes the combiner. In this work, to derive the precoding and combining matrices, singular value decomposition (SVD) is applied to the channel matrix. Specifically, precoder uses the right singular vectors and the combiner uses the left singular vectors [6]. To enable precoder computation at the AP, the UT feeds back the right singular matrix instead of the full CSI. However, direct transmission of this matrix is impractical due to limited uplink capacity. Therefore, an efficient DL-based feedback mechanism is required to achieve optimal system performance.

Let $\mathbf{H}_n \in \mathbb{C}^{N_{sc} \times N_r \times N_t}$ represent the frequency-domain channel tensor formed by concatenating the per-subcarrier channel matrices $\mathbf{H}_{n,k}$, for $k = 0, 1, \dots, N_{sc}-1$. Based on $\mathbf{H}_n$, the UT computes the SVD on each subcarrier and feeds back the corresponding precoding matrix $\mathbf{V}_n \in \mathbb{C}^{N_{sc} \times N_t \times N_l}$, where $\mathbf{V}_n^H$ contains the right singular vectors. The resulting feedback overhead is equal to the total number of elements borne by the precoder, i.e., $N = N_{sc}N_tN_l$, which becomes prohibitively large in massive MIMO systems. Throughout this work, perfect CSI acquired through pilot-driven estimation is considered.

To leverage the sparse representation of the precoding matrix, a two-dimensional discrete Fourier transform (DFT) is performed along the subcarrier and layer dimensions [10]. Owing to the energy compaction property of the transformed representation, most of the substantial energy is, along the subcarrier dimension, within the first $\tilde{N}_{sc}$ rows. Consequently, only these rows are retained to reduce the input dimensionality. The truncated matrix having complex-valued is subsequently expressed into its in-phase and quadrature-phase representations and normalized to construct a four-dimensional tensor $\bar{\mathbf{V}}_n \in \mathbb{R}^{\tilde{N}_{sc} \times N_t \times N_l \times 2}$, where the 2 denotes the in-phase and quadrature components of the precoder. To facilitate efficient processing by the neural network, the precoder is converted into a three-dimensional representation $\bar{\mathbf{V}}_n \in \mathbb{R}^{\tilde{N}_{sc} \times (N_t N_l) \times 2}$. Finally, the resulting tensor is encoded by a deep learning-based encoder to generate the compressed latent representation, expressed as

$$f_{\text{enc}}(\bar{\mathbf{V}}_n) = \mathbf{q}_n. \quad (2)$$

The encoder maps the input tensor $\bar{\mathbf{V}}_n$ into a compact latent representation, referred to as the CW, denoted by $\mathbf{q}_n \in \mathbb{R}^M$. The corresponding CR is defined as M/N , where N denotes the total number of input elements. The original tensor, at the receiver, is reconstructed from the compressed codeword using a DL-based decoder, given by

$$f_{\text{dec}}(\mathbf{q}_n) = \bar{\mathbf{V}}_n. \quad (3)$$

The output of the decoder is converted back to its original form by performing denormalization and combining its real and imaginary components, resulting in $\hat{\mathbf{V}}_n \in \mathbb{C}^{\tilde{N}_{\text{sc}} \times N_t \times N_l}$.

The reconstruction accuracy is assessed using the normalized mean square error (NMSE). Let $\bar{\mathbf{V}}_n$ and $\hat{\mathbf{V}}_n$ denote the true and reconstructed precoder tensors, respectively. The NMSE is computed as

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\|\bar{\mathbf{V}}_n - \hat{\mathbf{V}}_n\|_F^2}{\|\bar{\mathbf{V}}_n\|_F^2} \right\}. \quad (4)$$

Although NMSE is commonly used to evaluate CSI feedback performance, Subspace Cosine Similarity (SCS) is more suitable when the precoder is reconstructed directly. Let $\bar{\mathbf{V}}_n$ and $\hat{\mathbf{V}}_n$ denote the true and reconstructed precoder tensors, respectively. Since SCS measures the similarity between the subspaces spanned by the precoder matrices, the precoder corresponding to each retained subcarrier is first orthonormalized using QR decomposition to obtain $\bar{\mathbf{Q}}_{n,k}$ and $\hat{\mathbf{Q}}_{n,k}$. The SCS is then computed as

$$\text{SCS} = \frac{1}{\tilde{N}_{\text{sc}} N_l} \sum_{k=1}^{\tilde{N}_{\text{sc}}} \|\bar{\mathbf{Q}}_{n,k}^H \hat{\mathbf{Q}}_{n,k}\|_F^2. \quad (5)$$

Subsequently, the reconstructed tensor is first expanded to the original subcarrier dimension by padding $N_{\text{sc}} - \tilde{N}_{\text{sc}}$ zero rows along the subcarrier axis. Finally, the precoding matrix $\mathbf{V}_n \in \mathbb{C}^{N_{\text{sc}} \times N_t \times N_l}$ is recovered after applying a two-dimensional inverse discrete Fourier transform (IDFT).

4 Architecture of the Proposed Model

To compress $\bar{\mathbf{V}}_n \in \mathbb{R}^{\tilde{N}_{\text{sc}} \times (N_t N_l) \times 2}$, TPCNet employs an autoencoder, as shown in Fig. 1, which begins by extracting initial spatial features with a 5×5 convolutional layer, followed by the MRFEN. Residual connections are then used to fuse shallow and deep representations, preserving directional information. The fused features undergo patch embedding with positional encoding, followed by a TB to capture global dependencies. Subsequently, an SRN progressively restores the spatial structure, followed by a 3×3 convolutional layer. Finally, a dense layer generates the compact CW $\mathbf{q}_n$.

At the AP, the autodecoder, as shown in Fig. 2, reconstructs the precoder from the received CW $\mathbf{q}_n$. The compressed vector is first expanded using a dense layer and reshaped into a spatial structure, followed by a 5×5 convolutional layer to enhance feature representation before transformer processing. The SRN then restores the spatial resolution, while residual connections reuse earlier features to preserve spatial information during upsampling. To further exploit spatial correlation, a feature FRN is employed. Finally, a 3×3 convolutional layer with sigmoid activation generates the reconstructed precoder. The details of the intermediate networks are provided in the following subsections.

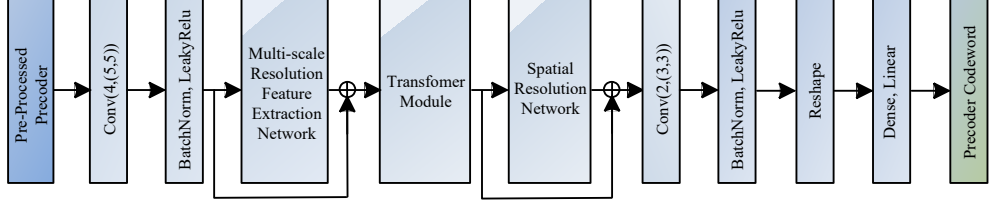


Fig. 1 TPC Encoder

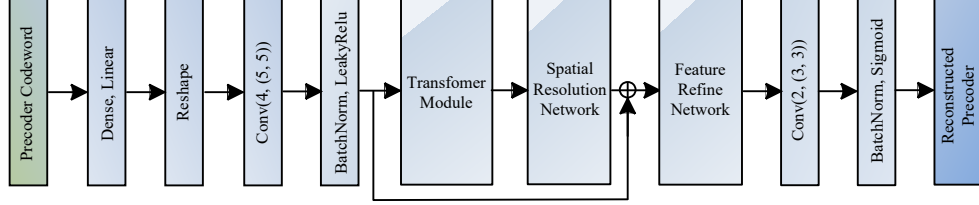


Fig. 2 TPC Decoder

4.1 Multi-scale Resolution Feature Extraction Network

To capture directional and non-uniform correlations along the subcarrier and layer dimensions, an MRFEN, as illustrated in Fig. 3, is introduced at the encoder front end. To extract local and axis-oriented features, convolutional layers with kernel sizes of 3×3 , 1×9 , and 9×1 are employed, while a parallel 3×3 branch preserves the low-level spatial structure. Furthermore, to produce a compact and effective multi-scale representation for subsequent processing, the extracted feature maps are first fused along the channel dimension, followed by a 1×1 convolutional layer.

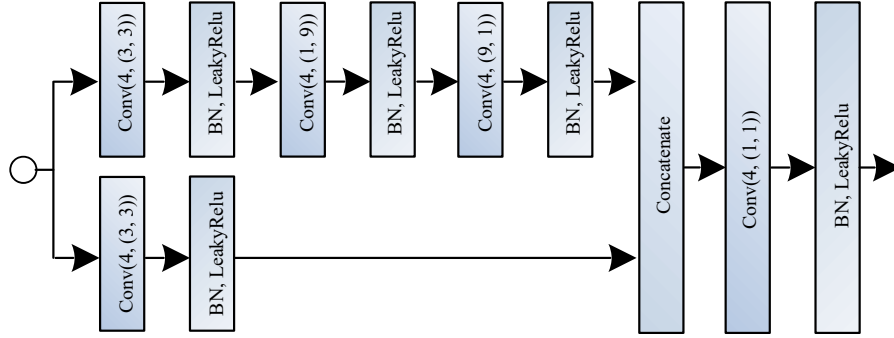


Fig. 3 Multi-scale Resolution Feature Extraction Network

4.2 Transformer Block

To exploit sparsity in the AD domain and focus on the most informative components, the feature map $\bar{\mathbf{V}}_n \in \mathbb{R}^{\tilde{N}_{\text{sc}} \times (N_t N_l) \times 2}$ is, as illustrated in Fig. 4, partitioned into $N = \frac{\tilde{N}_{\text{sc}} N_t N_l}{P^2}$ non overlapping $P \times P$ patches, where P denotes the patch size. Each patch $\mathbf{x}_p \in \mathbb{R}^{2P^2}$ is flattened and linearly projected into a 1D token representation as

$$\mathbf{E}_p = \mathbf{x}_p \mathbf{W}_e, \quad \mathbf{W}_e \in \mathbb{R}^{2P^2 \times D}.$$

To maintain the structural ordering, learnable separable 2D positional embeddings are introduced along the delay and angle axes. The positional encoding (PE) at location (i, j) is defined as

$$\text{PE}_{i,j} = \text{PE}_{\text{delay}}[i] + \text{PE}_{\text{angle}}[j],$$

where PE_{delay} and PE_{angle} denote trainable embedding vectors indexed by delay and angle positions. The final l th token in the sequence is obtained as $z_l = E_p^l + \text{PE}_{i,j}$.

The embedded patch sequence $\mathbf{z} \in \mathbb{R}^{N \times D}$ is processed using a network to capture global dependencies across the AD representation. This network consists of a layer normalization layer, followed by a multi-head self-attention (MHSA) module and a multi-layer perceptron (MLP) module. Residual connections and dropout are applied to both the MHSA and MLP modules. The MHSA mechanism models interactions among distant subcarriers and transmission layer indices beyond the capability of local convolutional receptive fields, while the MLP further refines token representations after global attention through non-linear transformation. Dropout enhances generalization, and residual connections stabilize feature learning by preserving intermediate representations. Finally, the resulting sequence is reshaped into a 2D feature map of size $\mathbb{R}^{H \times W \times D}$ for subsequent spatial processing.

4.3 Spatial Resolution Network

After transformer processing and reshaping to $\mathbb{R}^{H \times W \times D}$, the SRN restores the spatial resolution of the precoder representation through staged transposed convolutions, as illustrated in Fig. 5. To enable smooth spatial reconstruction from patch-level features, the resolution is progressively recovered through two stride-controlled steps of $\frac{P}{2}$, instead of directly upsampling the feature map by a factor of P . A subsequent convolutional layer further refines local details and produces a spatially consistent feature map for the subsequent refinement stages.

4.4 Feature Refinement Network

The FRN, as illustrated in Fig. 6, improves reconstruction fidelity by refining the spatial characteristics of the precoder after spatial resolution recovery. To capture localized patterns, subcarrier dependencies, and layer correlations in the AD domain, sequential directional convolutions with kernel sizes 3×3 , 1×9 , and 9×1 are employed. In parallel, a 5×5 convolution extracts complementary multi-scale contextual features from the input. Finally, to enable iterative correction while preserving the precoder structure, the resulting feature maps are fused with the original representation through

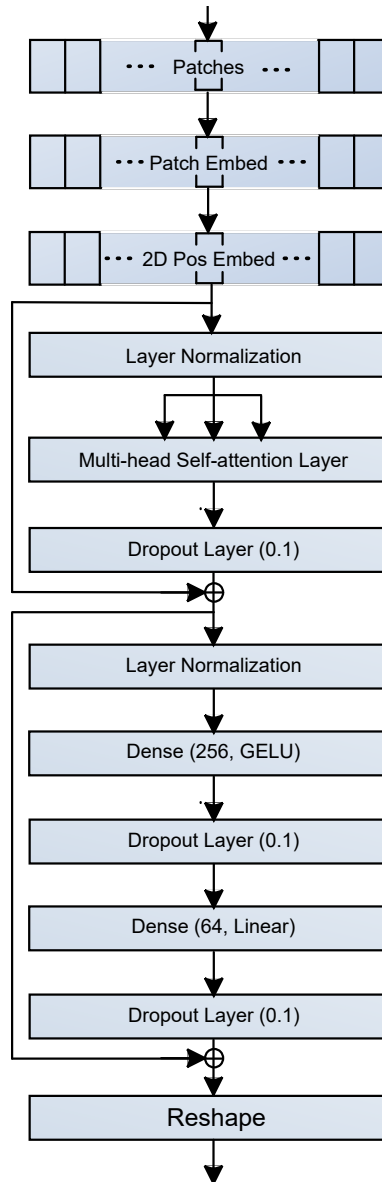


Fig. 4 Transformer Block

residual aggregation. This refinement stage enhances spatial consistency before the model output.

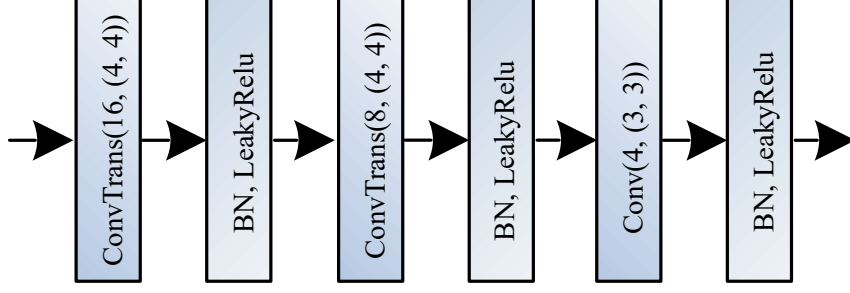


Fig. 5 Spatial Resolution Network

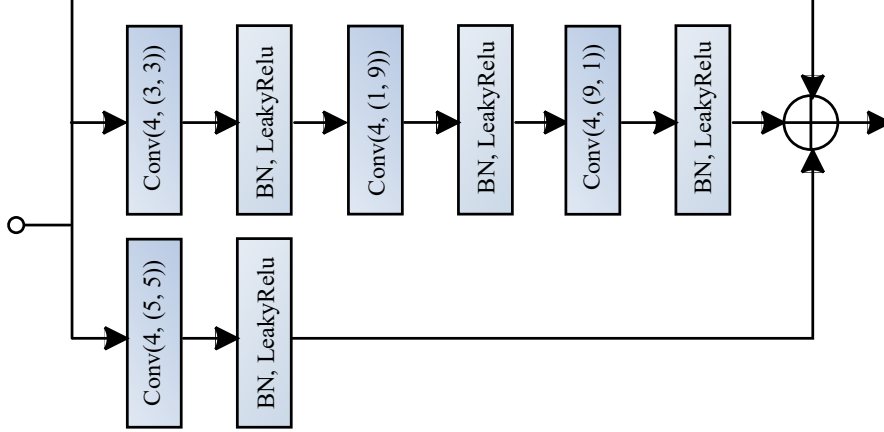


Fig. 6 Features Refine Network

5 Numerical Results

The numerical results of the proposed framework are evaluated through extensive simulations for a massive MIMO-OFDM system operating in FDD mode. The AP employs 32 dual-polarized radiating elements arranged in an $8 \times 2 \times 2$ planar array, whereas each UT employs four dual-polarized omnidirectional radiating elements configured as a $2 \times 1 \times 2$ array. The AP is positioned at the center of a hexagonal cell, while the UTs are uniformly distributed throughout the coverage area with an inter-site distance of 500 m. The OFDM configuration consists of 1024 subcarriers, over which both data and pilot symbols are allocated according to the resource grid. Table 1 illustrates the rest of the simulation parameters.

The performance of various feedback schemes is computed through a comprehensive simulation. The proposed scheme, TPCNet, is compared with baseline schemes, namely CsiNet [10], CRNet [13], DeConvD-CRNet [15], and CsiFormer [18]. The proposed TPCNet employs a patch size of 4 with a single attention head. For a fair performance evaluation, all benchmark schemes are trained and tested using an identical dataset generated according to the 3GPP-compliant CDL-C channel model. The

Table 1 Simulation parameters

Parameter	Value
Propagation environment	“CDL-C”
Time Dispersion	30 ns
Transmission Mode, Signaling Waveform	FDD, OFDM
Center Frequency (f_c)	2 GHz
Subcarrier Frequency Separation (Δf)	30 kHz
System Bandwidth	30 MHz
MCS Index	7
Number of RBs (N_{RB})	80
FFT size (N_{FFT})	1024
Number of Tx radiating elements (N_t)	32
Number of Rx radiating elements (N_r)	4
Channel Acquisition using CSI-RS	Perfect
Number of spatial layers (N_L)	1, 2
SB size (N_{RB}^{sb})	8 RBs
No. Of Beams (L)	4
Phase Alphabet Size (N_{PSK})	8
Amplitude Subset Restriction	Supported
SB Amplitude	supported
ParamCombination-r16	5
No. of PMI-SBs Per CQI-SB (R)	2

dataset contains 100,000 channel realizations and is split into 80% training, 10% validation, and 10% testing dataset. Adam optimizer is used for training of all discussed models, including TPCNet, with 1000 epochs for a batch size of 200 across CRs of $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{32}$, and $\frac{1}{64}$. A cosine annealing learning rate scheduler is deployed, inspired by CRNet [13], for the following baseline [13, 15, 18] schemes and TPCNet. The first 20 epochs are used for linear warm-up, and the remaining epochs are used with a cosine annealing scheduler

Table 2 NMSE Comparison of Various CSI Feedback Methods

Methods	NMSE (dB)			
	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{64}$
CsiNet [10]	-30.73	-28.21	-24.85	-23.51
CRNet [13]	-31.74	-28.76	-24.90	-23.82
DeConvD-CRNet [15]	-31.80	-28.79	-24.92	-23.92
CsiFormer [18]	-32.29	-29.17	-25.06	-24.26
TPCNet	-32.72	-29.67	-25.42	-24.44

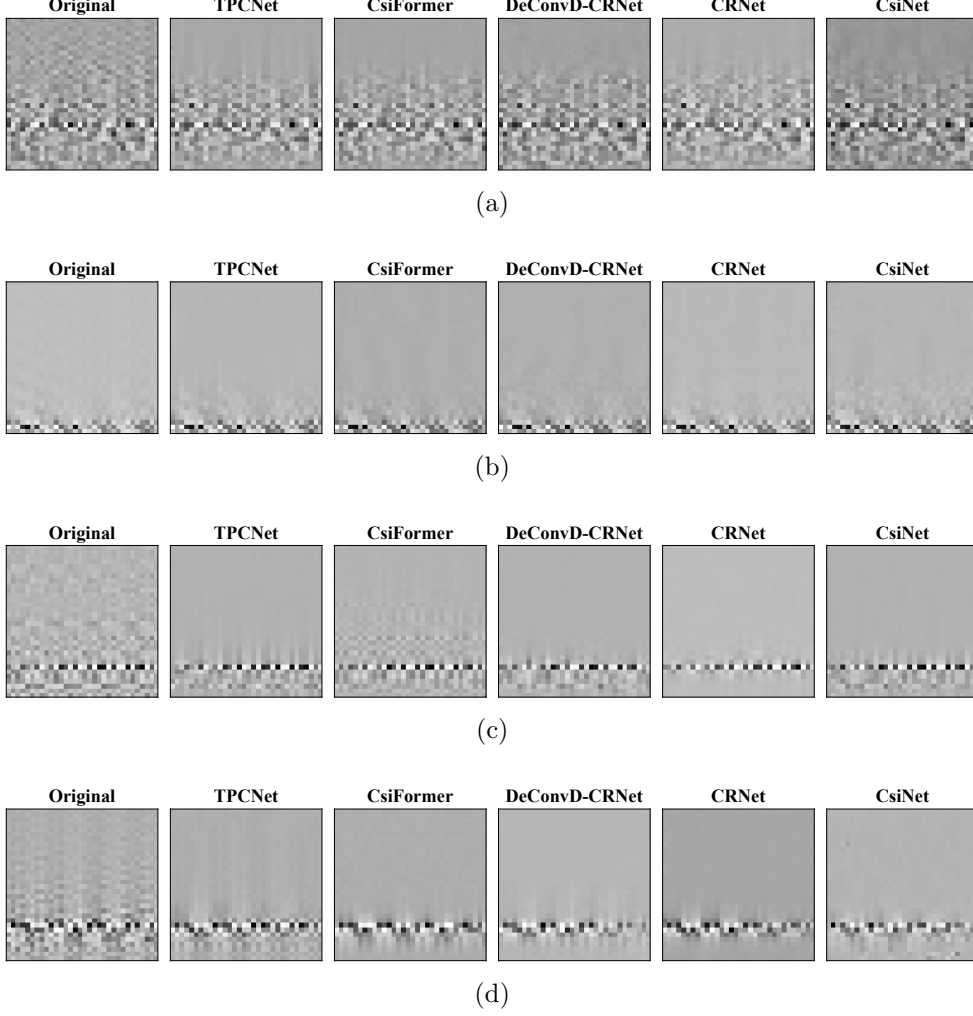


Fig. 7 (a) $CR = \frac{1}{4}$, (c) $CR = \frac{1}{8}$, (b) $CR = \frac{1}{32}$, and (d) $CR = \frac{1}{64}$.

Fig. 7 and Table 2 compare the NMSE performance of CsiNet [10], CRNet [13], DeConvD-CRNet [15], and CsiFormer [18] with TPCNet. As observed, TPCNet delivers the lowest NMSE values across all evaluated CRs. Table 3 compares the SCS performance achieved by TPCNet against those of CsiNet [10], CRNet [13], DeConvD-CRNet [15], and CsiFormer [18]. The results indicate that TPCNet consistently delivers the highest SCS performance among all competing schemes over the entire range of CRs. In addition, Table 4 compares the computational parameters of various CSI feedback methods in terms of floating-point operations (FLOPs). Despite its lower computational cost, TPCNet outperforms CsiFormer [18] in NMSE performance, achieving an approximate 25% reduction in FLOPs. Moreover, TPCNet surpasses

lightweight architectures, including CsiNet, CRNet, and DeConvD-CRNet, by achieving a more favorable balance between reconstruction performance and computational complexity. These results confirm the effectiveness of the proposed architectural design in enhancing channel reconstruction performance.

Table 3 SCS Comparison of Various CSI Feedback Methods

Methods	SCS			
	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{64}$
CsiNet [10]	0.740	0.652	0.561	0.541
CRNet [13]	0.741	0.668	0.564	0.545
DeConvD-CRNet [15]	0.742	0.669	0.565	0.547
CsiFormer [18]	0.766	0.680	0.573	0.551
TPCNet	0.774	0.697	0.596	0.557

Table 4 FLOPs Comparison of Various CSI Feedback Methods

Methods	No. of FLOPs			
	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{64}$
CsiNet [10]	29.92M	21.54M	15.24M	14.20M
CRNet [13]	28.70M	20.31M	14.02M	12.97M
DeConvD-CRNet [15]	29.51M	21.12M	14.83M	13.78M
CsiFormer [18]	169.52M	161.14M	154.89M	153.80M
TPCNet	127.49M	119.36M	113.07M	112.02M

Table 5 Ablation Study Results of TPCNet

CR	SRN		MRFEN	FRN	NMSE (dB)
	(K_1^u, F_1^u)	(K_2^u, F_2^u)			
$\frac{1}{8}$	(4,2)	(4,2)	—	—	-29.07
	(4,2)	(4,2)	✓	—	-29.26
	(4,2)	(4,2)	✓	✓	-29.41
	(8,4)	(0,0)	✓	✓	-29.44
	(8,2)	(8,2)	✓	✓	-29.54
	(8,2)	(16,2)	✓	✓	-29.60
	(16,2)	(8,2)	✓	✓	-29.67

Note: K_1^u, F_1^u and K_2^u, F_2^u denote the kernel size and upsampling factors of the first and second transposed convolution layers, respectively.

5.1 Ablation Study

To quantify the contribution of individual components in TPCNet, an ablation study is conducted while retaining the TB as the backbone. As shown in Table 5, the baseline configuration, comprising the TB and SRN, achieves an NMSE of -29.07 dB at $CR = \frac{1}{8}$. Incorporating MRFEN improves the NMSE to -29.26 dB by enhancing directional feature representation in the AD domain, while the addition of FRN further reduces the NMSE to -29.41 dB through local feature refinement. Furthermore, SRN design variations show that progressive two-stage upsampling provides more stable spatial reconstruction than single-stage expansion, with the optimal configurations $(16, 2)$ and $(8, 2)$ achieving -29.67 dB.

6 Conclusion

In this paper, we introduce TPCNet, a novel transformer-based architecture for a lightweight CSI feedback framework in massive MIMO-OFDM systems. The network integrates an MRFEN and an MHSA layer to effectively capture local and long-range spatial dependencies, followed by an FRN for enhanced reconstruction. The proposed framework is trained and evaluated on a dataset generated through the 3GPP-standardized CDL-C MIMO channel model. In contrast to conventional CSI feedback approaches that reconstruct the channel matrix, TPCNet directly estimates the downlink precoding matrix corresponding to the two dominant transmission layers. The proposed TPCNet consistently outperforms state-of-the-art CSI feedback schemes, providing performance gains of approximately 0.5–2 dB while reducing the computational complexity by nearly 25% in terms of FLOPs compared with the leading transformer-based benchmark. In addition, TPCNet inherently achieves a reduction in feedback overhead by a factor of N_r as it reconstructs the precoding matrix from a single CW, irrespective of the number of radiating elements at the UT. Moreover, for $N_r < N_t$, the input tensor dimension remains independent of the number of receive antennas, thereby eliminating the scalability limitations associated with increasing N_r . Extensive simulations demonstrate that TPCNet outperforms existing CSI feedback methods over a wide range of CRs, including $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{32}$, and $\frac{1}{64}$. Its low feedback overhead and computational complexity enable scalable deployment in 5G-Advanced and beyond.

Acknowledgements. The authors would like to express their gratitude towards Maulana Azad National Institute of Technology Bhopal for providing the necessary infrastructure to carry out this work successfully.

Funding. This research work is not funded from any organisation.

Declarations

Competing interests. The authors declare no competing interests.

Conflict of interest. The authors declare that there is no conflict of interest regarding the publication of this article.

References

- [1] D.A.U.V. *et al.*, “Spectral Efficiency of Precoded 5G-NR in Single and Multi-User Scenarios under Imperfect Channel Knowledge: A Comprehensive Guide for Implementation,” *Electronics* 2022, 11, 4237.
- [2] Ziao Qin *et al.*, “A review of codebooks for CSI feedback in 5G new radio and beyond,” arXiv 2023.
- [3] J. Kumar *et al.*, “A review on 5G and beyond wireless communication channel models: Applications and challenges,” *Physical Communication*, vol. 67, p. 102488, 2024.
- [4] “R1-1709232 WF on Type-I and II CSI codebooks RAN1 89, Tech. Rep., 2017.”
- [5] V. Ramireddy *et al.*, “Enhancements on Type-II 5G New Radio Codebooks for UE Mobility Scenarios,” in *IEEE Communications Standards Magazine*, vol. 6, no. 1, pp. 35-40, March 2022.
- [6] R. Pal, *et al.*, “Performance evaluation of AI-based CSI feedback schemes compliant with 3GPP standards,” *Physical Communication*, vol. 72, p. 102803, 2025.
- [7] B. Hu, H. Xu, Z. Wang, L. Zhao, and A. Zhou, “Multiple-compression-rate CSI feedback for mmWave massive MIMO based on federated learning,” in *Physical Communication*, vol. 63, pp. 102305, 2024, ISSN: 1874-4907.
- [8] R. Pal, V. Singh, and S. Maurya, “PrecoderNet: A feedback framework for massive MIMO-OFDM systems,” in *Proc. 18th Int. Conf. Communication Systems and Networks (COMSNETS)*, Bengaluru, India, pp. 305–310, 2026.
- [9] X. Fu *et al.*, “A Tutorial on Downlink Precoder Selection Strategies for 3GPP MIMO Codebooks,” in *IEEE Access*, vol. 11, pp. 897-922, 2023.
- [10] C. -K. Wen, *et al.*, “Deep Learning for Massive MIMO CSI Feedback,” in *IEEE Wirel. Commun. Lett.*, vol. 7, no. 5, pp. 748-751, 2018.
- [11] T. Wang *et al.*, “Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels,” in *IEEE Wirel. Commun. Lett.*, vol. 8, no. 2, pp. 416-419, April 2019.
- [12] C. Lu, *et al.*, “MIMO Channel Information Feedback Using Deep Recurrent Network,” in *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 188-191, Jan. 2019.
- [13] Z. Lu, *et al.*, “Multi-resolution CSI Feedback with Deep Learning in Massive MIMO System,” *ICC 2020, Dublin, Ireland, 2020*, pp. 1-6.

- [14] S. Tang, *et al.*, “Dilated Convolution Based CSI Feedback Compression for Massive MIMO Systems,” in *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11216-11221, Oct. 2022.
- [15] A. Swain, *et al.*, “Low-Complexity Deep Learning Framework for CSI Feedback in Massive MIMO System,” in *IEEE Wirel. Commun. Lett.*, vol. 13, no. 3, pp. 864-868, March 2024.
- [16] Y. Xu, M. Yuan, and M.-O. Pun, “Transformer empowered CSI feedback for massive MIMO systems,” in *Proc. 30th Wireless Opt. Commun. Conf. (WOCC)*, Taipei, Taiwan, 2021, pp. 157–161.
- [17] Y. Cui, *et al.*, “TransNet: Full Attention Network for CSI Feedback in FDD Massive MIMO System,” in *IEEE Wirel. Commun. Lett.*, vol. 11, no. 5, pp. 903-907, May 2022.
- [18] X. Bi, *et al.*, “A Novel Approach Using Convolutional Transformer for Massive MIMO CSI Feedback,” in *IEEE Wirel. Commun. Lett.*, vol. 11, no. 5, pp. 1017-1021, May 2022.
- [19] Z. Liu, T. Song, R. Zhao, J. Jin, and G. Jin, “An efficient parallel self-attention transformer for CSI feedback,” in *Physical Communication*, vol. 66, pp. 102483, 2024, ISSN: 1874-4907.
- [20] H. Chen, W. Tan, J. Guo, and F. Yang, “SCANet: A lightweight deep learning network for massive MIMO CSI feedback based on spatial and channel attention mechanism,” in *Physical Communication*, vol. 67, pp. 102516, 2024, ISSN: 1874-4907.