

# Deep Learning for Efficient CSI Feedback in Massive MIMO: Adapting to New Environments and Small Datasets

Zhenyu Liu<sup>id</sup>, Member, IEEE, Li Wang<sup>id</sup>, Senior Member, IEEE, Lianming Xu<sup>id</sup>, Member, IEEE, and Zhi Ding<sup>id</sup>, Fellow, IEEE

**Abstract**— Channel State Information (CSI) feedback, powered by Deep Learning (DL) methodologies, exhibits significant promise in enhancing spectrum efficiency within massive MIMO systems. However, DL-based approaches typically necessitate substantial CSI datasets for each specific scenario, and managing multiple learned models demands considerable storage and updating bandwidth. To overcome this costly barrier, we develop a solution for efficient training and deployment enhancement of DL-based CSI feedback, which involves a lightweight translation model to cope with new CSI environments and introduces a novel dataset augmentation based on domain knowledge. Specifically, we first develop a deep unfolding CSI feedback network, SPTM2-ISTANet+, which incorporates spherical normalization to mitigate the challenge of path loss variation. Additionally, SPTM2-ISTANet+ integrates a trainable measurement matrix and residual CSI recovery blocks to enhance efficiency and accuracy. Employing SPTM2-ISTANet+ as a foundational feedback model, we introduce an adaptive CSI feedback architecture termed CSI-TransNet. CSI-TransNet features a scenario-adaptive plug-in module for CSI translation, composed of a sparsity aligning function and a compact DL module, facilitating the reuse of pretrained models in unencountered environments. To accommodate the small datasets, we propose a lightweight and general augmentation strategy based on domain knowledge. Test results demonstrate the efficacy and efficiency of the proposed

Manuscript received 14 September 2023; revised 20 January 2024; accepted 6 April 2024. Date of publication 24 April 2024; date of current version 12 September 2024. The work of Zhenyu Liu was supported in part by the National Natural Science Foundation of China under Grant 62201089, Grant U2066201, and Grant 62171054. The work of Li Wang and Lianming Xu was supported in part by the National Natural Science Foundation of China under Grant 62201089, Grant U2066201, and Grant 62171054; in part by the Fundamental Research Funds for the Central Universities under Grant 24820232023YQTD01; and in part by the Double First-Class Interdisciplinary Team Project Funds under Grant 2023SYLTD06. The associate editor coordinating the review of this article and approving it for publication was S. Bi. (*Corresponding author: Li Wang*)

Zhenyu Liu was with the School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China. He is now with the 5GIC and 6GIC, Institute for Communication Systems, University of Surrey, GU2 7XH Guildford, U.K. (e-mail: zhenyu.liu@surrey.ac.uk).

Li Wang is with the School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China (e-mail: liwang@bupt.edu.cn).

Lianming Xu is with the School of Electronic Engineering, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China (e-mail: xulianming@bupt.edu.cn).

Zhi Ding is with the Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95616 USA (e-mail: zdong@ucdavis.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2024.3390583>.

Digital Object Identifier 10.1109/TWC.2024.3390583

solution for accurate CSI feedback given limited measurements for unseen CSI environments.

**Index Terms**— Massive MIMO, CSI feedback, CSI augmentation, deep learning, multi-scenario.

## I. INTRODUCTION

MODERN wireless communication systems have significantly advanced in harnessing spatial diversity through multiple-input multiple-output (MIMO) transceivers, enhancing radio link performance. Notably, massive MIMO systems have emerged as a pivotal technology for 5G and future wireless systems, offering substantial spectrum and energy efficiency. The effectiveness of massive MIMO in downlink operations hinges on precise channel state information (CSI) estimates at the gNodeB (gNB) for transmission precoding. However, in massive MIMO contexts, the volume of feedback data becomes considerable due to the extensive antenna arrays and broad bandwidth. Consequently, this situation has vigorously driven research towards achieving accurate downlink CSI feedback in frequency division duplex (FDD) systems.

Compressive sensing (CS) techniques have been leveraged to exploit intrinsic channel properties, such as low rank or sparsity in the spatial [1] or temporal domains [2], thereby reducing the bandwidth required for channel state information (CSI) feedback. However, the efficiency of CS approaches heavily relies on a strong sparsity condition. This condition may not always be strictly met, potentially limiting the effectiveness of these methods [3]. Consequently, recent research has shifted towards employing deep neural networks (DNNs) for CSI feedback.

Deep learning (DL)-based methods for CSI feedback have demonstrated remarkable success in recovery accuracy and time efficiency. These methods adeptly harness various channel characteristics, including spatial and spectral coherence [3], [4], [5], [6], [7], [8], [9], [10] and bi-directional reciprocity [11], [12]. Moreover, by optimizing the structure of the CSI feedback architecture, existing works have highlighted the advantages of incorporating temporal CSI correlation as side information in DL-based CSI feedback models [13], [14], [15], [16]. Specifically, [13] introduces a Recurrent Neural Network (RNN)-based massive MIMO channel predictor that shows promising results on real-world datasets. By compressing the difference between the predicted and estimated

channels, the studies in [14] and [15] demonstrate a significant reduction in feedback overhead. However, a notable challenge with DL models is their reliance on extensive channel measurements for training. The process of collecting such measurements is often laborious and costly. Furthermore, a DL model trained for one radio-frequency (RF) environment may not perform optimally in others due to model mismatch issues.

The challenge of limited real-time channel measurements necessitates effective data augmentation strategies to alleviate labor-intensive measurement costs and prevent model overfitting. Traditional augmentation techniques common in image processing, such as flipping, cropping, or rotation [17], are not suitable for physical radio channels. Besides, [18] multiplies the entire CSI matrix by a random value, neglecting the phase shift differences in various multipaths, which degrades the augmentation performance. This situation underscores the need for dedicated CSI data augmentation methods, which should integrate domain-specific knowledge of radio propagation physics. Another direction applies generative adversarial network (GAN) as a black box for CSI augmentation [19], [20], [21]. Ironically, GANs typically require a substantial amount of CSI measurements for effective training. Moreover, GAN designed for massive MIMO [21] can be computationally costly and demand billions of floating point operations (FLOPs).

In addition to data augmentation, to improve CSI recovery accuracy from limited available data samples, DNN should be capable of handling complex CSI features and variations. While existing DL-based CSI feedback methods [5], [6], [9], [10] show improved performance in indoor channel environments, their effectiveness often diminishes in more complex outdoor scenarios. Furthermore, CSI from wide bandwidth and massive antennas tend to exhibit a high degree of variation. For instance, path loss variations can significantly alter CSI magnitude, sometimes by several orders. This variability poses a notable challenge for DL-based CSI compression and feedback systems, particularly when trained on small datasets that do not adequately represent a diverse range of scenarios. Consequently, processing massive MIMO CSI from a wide variety of deployment scenarios becomes a formidable task, underscoring a critical hurdle in the practical implementation of DL-based CSI feedback solutions.

The principle of transfer learning, which utilizes correlations across different tasks and scenarios, offers a way to reduce training costs. Traditional approaches to transfer learning, however, often involve extensive network fine-tuning [22], [23]. This becomes a challenge when User Equipment (UE) transitions to a new environment, requiring updates to the encoder network. Given that CSI feedback encoder networks [3], [4], [5], [6], [7], [8], [9], [10] can encompass millions of parameters, this translates to substantial transmission bandwidth overhead. To mitigate the costs associated with encoder storage or updates, recent studies have proposed UE-friendly designs for one-sided encoder-based CSI feedback in multiple CSI scenarios [24], [25]. These designs employ a shared encoder alongside multiple task-specific decoders, aiming to significantly reduce UE storage requirements. However,

a shared encoder can degrade CSI recovery accuracy due to the lack of environment-specific features [25].

To develop a scenario-adaptive, high-accuracy CSI feedback solution with minimal CSI samples and deployment cost, we propose an efficient DL feedback architecture. This architecture promises enhanced training and communication-efficient deployment even with minimal measured data. Our approach aims to improve the recovery accuracy and robustness of the CSI feedback network across diverse environments, integrating an efficient UE-friendly design with a simplified, model-driven data augmentation approach based on physical insights and domain expertise. Our contributions are summarized as follows:

- Practical deployment of existing DL-based CSI feedback mechanism can be hindered by insufficient channel measurements needed to train large-scale DNNs. Training multiple models against various radio environments further elevate the difficulty. In this paper, we design a scenario-adaptive CSI compression framework which requires few channel measurements for training and can reuse pretrained DL model in a new and possibly unseen environment. Our proposed framework can seamlessly integrate with prevailing CSI feedback models, necessitating only the insertion and customization of a translation and retranslation module.
- We present a plug-in CSI feedback architecture featuring a lightweight translation module for efficient encoder updates in dynamic wireless environments. This module aligns new RF environment CSI data with pretrained data formats, enabling efficient reuse of the pretrained model. By incorporating a sparsity aligning function in the angular-delay domain and a lightweight CNN, our translation module achieves high recovery accuracy with significantly fewer parameters compared to existing methods.
- To meet the demands of intricate RF environments, we present the SPTM2-ISTANet+, a CS-inspired efficient CSI feedback framework. It features a spherical feedback structure for input distribution regulation and path loss mitigation, and a deep unfolding decoder network with a residual recovery structure for refined CSI accuracy. Simulation results show its exceptional performance, achieving an NMSE of  $-24.3$  dB at a compression ratio of 1/4 in a popular outdoor scenario [3].
- Instead of employing a black-box GAN for CSI sample augmentation, we propose a novel model-driven augmentation strategy, leveraging domain knowledge and physical features of CSI matrices. Acknowledging the geographical continuity and delay properties of MIMO channels, along with the circular nature of the discrete Fourier transform (DFT), we devise a method for circularly shifting CSI magnitudes in the angular-delay domain. Additionally, we introduce row-wise random phase variations to CSI elements, simulating the decoupled multipath channel phase shifts. Test results demonstrate significantly enhanced CSI recovery performance over black-box GAN, achieving an NMSE of

–15.8 dB with only 100 measurement samples at a compression ratio of 1/4 in the outdoor scenario.

## II. SYSTEM MODEL

Without loss of generality, we consider a massive MIMO gNB equipped with  $N_b$  antennas to serve a number of single-antenna UEs within its coverage. Orthogonal frequency division multiplexing (OFDM) is adopted in downlink transmission over  $N_f$  subcarriers. For subcarrier  $m$ , let  $\mathbf{h}_m \in \mathbb{C}^{N_b \times 1}$  denote the channel vector,  $\mathbf{w}_m \in \mathbb{C}^{N_b \times 1}$  denote transmit precoding vector,  $x_m \in \mathbb{C}$  denote the transmitted data symbol, and  $n_m \in \mathbb{C}$  denote the additive noise. Correspondingly, the received signal of the UE is

$$y_m = \mathbf{h}_m^H \mathbf{w}_m x_m + n_m, \quad (1)$$

where  $(\cdot)^H$  represents the conjugate transpose. The downlink CSI matrix in the spatial-frequency domain is denoted by  $\tilde{\mathbf{H}} = [\mathbf{h}_1, \dots, \mathbf{h}_{N_f}]^H \in \mathbb{C}^{N_f \times N_b}$ .

To reduce feedback overhead, we first exploit the sparsity of CSI in the delay domain. Applying 2D DFT, CSI matrix  $\mathbf{H}_{sf}$  in spatial-frequency domain can be transformed to be  $\mathbf{H}_{ad}$  in angular-delay domain using

$$\mathbf{F}_d^H \mathbf{H}_{sf} \mathbf{F}_a = \mathbf{H}_{ad}, \quad (2)$$

where  $\mathbf{F}_d$  and  $\mathbf{F}_a$  denote the  $N_f \times N_f$  and  $N_b \times N_b$  unitary DFT matrices, respectively. Owing to limited delay spread and scatters in the practical radio environment, most elements in the  $N_f \times N_b$  matrix  $\mathbf{H}_{ad}$  are negligibly small except for its first  $R_d$  rows [3]. Therefore, we can approximate the MIMO channel by keeping the first  $R_d$  rows of  $\mathbf{H}_{ad}$ , denoted by  $\mathbf{H}$ .

To enable DL-based CSI feedback systems to adapt to the spatially varying characteristics of wireless fading channels in different environments, mobile network operators are required to gather extensive channel measurements for each environment and train a corresponding DL network. Additionally, UEs moving into a new environment must adapt their encoder DL network, which may be substantial in size, often containing millions of parameters. Denoting  $T$  as the number of typical scenarios in a region, let  $\mathbf{H}^t$  represent the channel matrix for scenario  $t$ .

To reduce the deployment overhead and the need for extensive channel measurements in varying environments for CSI feedback, we propose a few-shot learning CSI framework. This framework builds upon prior training results from known channel environments. To utilize the knowledge gained from a previously known scenario, our DL feedback networks for CSI feedback under changing scenarios are developed in two phases: a) constructing a DL network in an anchor scenario without prior information ( $t = 1$ ); b) constructing DL networks for additional scenarios ( $t = 2, 3, \dots, T$ ) by leveraging the insights gained in the initial training phase.

Let  $\hat{\mathbf{H}}^t$  denote the recovered CSI matrix by the decoder in scenario  $t$ , corresponding to the ground-truth  $\mathbf{H}^t$ . Define the encoding and decoding functions as  $f_{en}(\cdot)$  and  $f_{de}(\cdot)$ , respectively. For the downlink CSI feedback architecture in the anchor scenario, the encoder and decoder networks are represented as follows:

$$\mathbf{s}_1 = f_{en,1}(\mathbf{H}^1; \Phi_1), \quad (3)$$

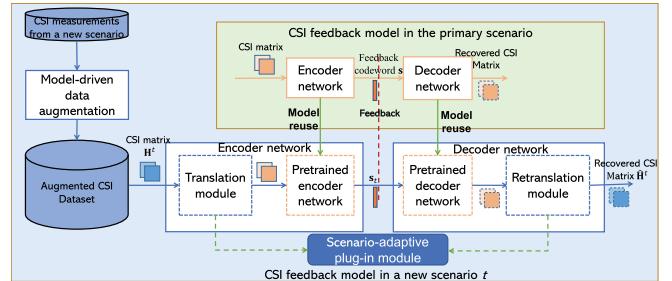


Fig. 1. Architecture of overall framework.

$$\hat{\mathbf{H}}^1 = f_{de,1}(\mathbf{s}_1; \Psi_1). \quad (4)$$

For subsequent scenarios  $t$  ( $t \geq 2$ ), the encoder and decoder networks are formulated as:

$$\mathbf{s}_t = f_{en,t}(\mathbf{H}^t; \Phi_t | \Phi_1, \Psi_1), \quad (5)$$

$$\hat{\mathbf{H}}^t = f_{de,t}(\mathbf{s}_t; \Psi_t | \Phi_1, \Psi_1). \quad (6)$$

## III. OVERALL FRAMEWORK

DL networks for CSI feedback are typically optimized through training with CSI data samples tailored for specific radio propagation scenarios. To adequately cover a city with diverse scenario types, a substantial number of channel measurements are often required for each scenario. Moreover, updating the encoder network as UE moves among various scenarios and sub-areas, especially those beyond spatial correlation distances, presents a significant challenge. This is due to the poor performance of a single CSI model when applied to different wireless environments, a result of model mismatch. Considering that CSI feedback encoder networks [5], [9], [10] may comprise millions of parameters, this implies considerable transmission bandwidth overhead. In response to these challenges, this work focuses on two key objectives: (1) designing a communication-efficient, scenario-adaptive DL framework for accurate CSI recovery; and (2) developing efficient utilization of limited CSI data measurements in DL training, leveraging domain knowledge.

To fulfill these objectives, we introduce a scenario-adaptive CSI feedback framework that reuses a pretrained high-accuracy CSI feedback model, supplemented with model-driven data augmentation in new environments. As depicted in Fig. 1, the framework comprises three primary modules:

- **Robust Backbone Model for an Anchor Scenario:** We propose a robust backbone model, suitable for deployment across diverse environments, thereby optimizing deployment costs. Recognizing the high degree of variation in CSI matrices of massive MIMO systems and the associated feedback performance degradation [24], we select SPTM2-ISTANet+ as our backbone network. This network is designed to handle complex CSI features and variations, combining a spherical feedback structure with a CS-inspired residual recovery DNN model to enhance feedback robustness and accuracy. The detailed design of SPTM2-ISTANet+ is presented in Section IV.
- **Scenario-Adaptive Plug-In Module:** For efficient reuse of the pretrained backbone model in new scenarios,

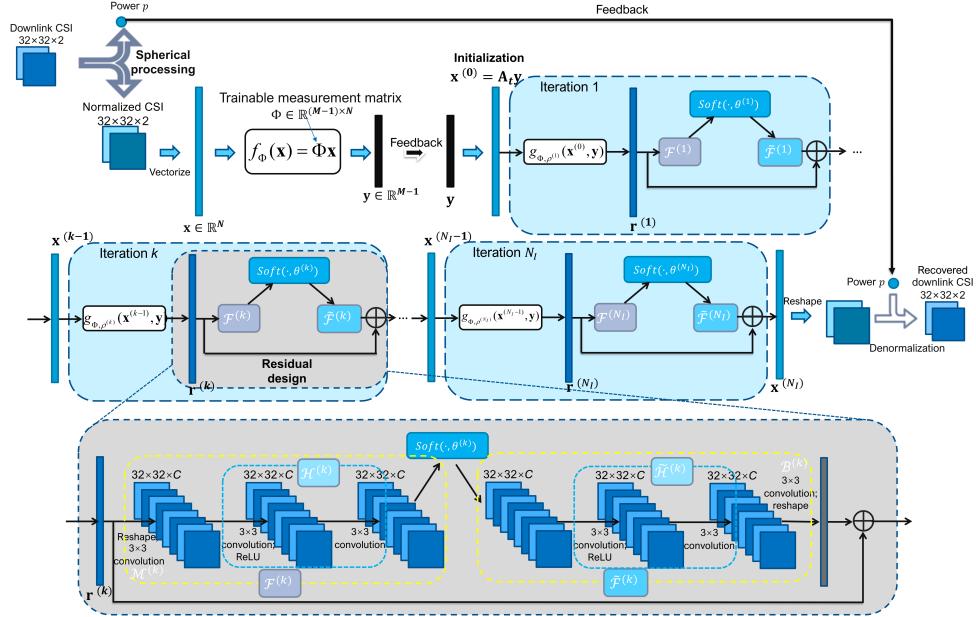


Fig. 2. Architecture of SPTM2-ISTANet+.

we propose a lightweight plug-in module that translates CSI matrices from new scenarios into a format compatible with the anchor scenario. This module, requiring only a few thousand trainable parameters, facilitates communication-efficient updates for new scenarios. It includes integrated translation and retranslation modules for adjusting and recovering the CSI style, further discussed in Section V.

- **Model-Driven Data Augmentation Module:** To mitigate the costs of labor-intensive channel measurement, we design a model-driven augmentation module that leverages the physical features of CSI matrices, informed by domain knowledge. This module separately augments CSI magnitudes and phases, considering their channel properties, with its detailed design elucidated in Section VI.

#### IV. SPTM2-ISTANET+

We now construct an efficient deep unfolding network as the backbone model for the anchor scenario to improve CSI feedback accuracy. Note that the deep unfolding modules are only used in the decoder network.

##### A. Encoding Network

To improve the robustness and CSI compression efficiency, we adopt a deep unfolding-based feedback network, and propose two key innovations to the encoder network in view of physical CSI features.

Our first innovation involves the construction of a spherical CSI feedback structure, tailored to the domain-specific characteristics of wireless channels. Unlike image data, the distribution of MIMO CSI coefficients exhibits a significantly wider dynamic range, primarily due to radio path loss. This can result in large discrepancies in CSI between different UEs, sometimes by orders of magnitude. A simplistic approach

to processing CSI may render the data for some UEs too minuscule, leading to substantial recovery errors. To mitigate this, we first separate the CSI matrix  $\mathbf{H}$  into a power value  $p$  and a spherical matrix  $\check{\mathbf{H}}$ . Here,  $p = \|\mathbf{H}\|$  represents the power, and  $\check{\mathbf{H}} = \mathbf{H}/\|\mathbf{H}\|$  is the unit norm spherical CSI.

As depicted in Fig. 2, following the spherical processing, we vectorize the normalized downlink CSI matrix  $\check{\mathbf{H}}$  as the DNN input for compression. The real and imaginary parts are separated for more straightforward processing, forming a vector  $\mathbf{x} \in \mathbb{R}^N$ , where  $N = 2 \times R_d \times N_b$ . To compress  $\mathbf{x}$ , we employ a measurement matrix  $\Phi \in \mathbb{R}^{(M-1) \times N}$ , with  $M = CR \times N - 1$ .

Our second innovation departs from the traditional approach of using randomly constructed measurement matrices in CS. We introduce a data-driven, trainable measurement matrix  $\Phi$ . This matrix is optimized to better capture the features of massive MIMO for CSI encoding, especially when the compression degree is high (i.e., CR is small). The use of a matrix multiplication at the encoder ensures that the computational cost for the UE remains modest.

##### B. Decoding Network

Assuming lossless feedback [3], [4], the low-dimensional vector  $\mathbf{y} \in \mathbb{R}^{M-1}$  received by the gNB is defined as  $\mathbf{y} = \Phi\mathbf{x}$ . From  $\mathbf{y}$ , the decoder network aims to reconstruct the original  $\mathbf{x}$  by solving the compressive sensing recovery problem given by:

$$\min_{\mathbf{x}} \frac{1}{2} \|\Phi\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathcal{F}(\mathbf{x})\|_1, \quad (7)$$

where  $\lambda$  is the regularization parameter and  $\|\cdot\|$  denotes the  $l_2$ -norm.  $\mathcal{F}(\cdot)$  represents the sparse transform function of  $\mathbf{x}$ .

Our decoder network in SPTM2-ISTANet+ leverages a deep unfolding structure. By adopting the configuration of ISTANet+ [26] to unfold the iterative shrinkage-thresholding algorithm (ISTA) [27], we iteratively recover CSI through two

primary steps:

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho \Phi^\top (\Phi \mathbf{x}^{(k-1)} - \mathbf{y}), \quad (8)$$

$$\mathbf{x}^{(k)} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{r}^{(k)}\|^2 + \lambda \|\mathcal{F}(\mathbf{x})\|_1, \quad (9)$$

where  $k$  denotes the iteration index and  $\rho$  the step size. We further expand Eq. (8) and Eq. (9) into deep unfolding modules corresponding to the  $k$ -th iteration, referred to as modules  $\mathbf{r}^{(k)}$  and  $\mathbf{x}^{(k)}$ , to solve the recovery problem.

Module  $\mathbf{r}^{(k)}$ , corresponding to Eq. (8), generates  $\mathbf{r}^{(k)}$  from the result of the  $(k-1)$ -th iteration. To enhance the flexibility of the recovery network, the step size  $\rho$  in Eq. (8) is adaptively adjusted per iteration, i.e.,  $\rho^{(k)}$  varies with each  $k$ . Hence, module  $\mathbf{r}^{(k)}$  functions based on  $\mathbf{x}^{(k-1)}$  and  $\mathbf{y}$ , as:

$$\mathbf{r}^{(k)} = g_{\Phi, \rho^{(k)}}(\mathbf{x}^{(k-1)}, \mathbf{y}) = \mathbf{x}^{(k-1)} - \rho^{(k)} \Phi^\top (\Phi \mathbf{x}^{(k-1)} - \mathbf{y}). \quad (10)$$

Module  $\mathbf{x}^{(k)}$ , aligning with Eq. (9), calculates  $\mathbf{x}^{(k)}$  from  $\mathbf{r}^{(k)}$  in the  $k$ -th iteration. The sparse transformation  $\mathcal{F}(\cdot)$  in Eq. (9) is constructed using a combination of two convolutional layers and a ReLU unit,  $\text{ReLU}(x) = \max(0, x)$ , i.e.,  $\mathcal{F}(\mathbf{x}) = \mathbf{B} \cdot \text{ReLU}(\mathbf{A}\mathbf{x})$ . Both  $\mathbf{A}$  and  $\mathbf{B}$  are convolutional layers without bias, simulating matrix operations. To address the vanishing gradient problem commonly encountered in deep unfolding, we incorporate a residual structure to boost recovery accuracy.

From Eq. (9), we assume  $\mathbf{x}^{(k)} = \mathbf{r}^{(k)} + \mathbf{w}^{(k)} + \mathbf{e}^{(k)}$ , where  $\mathbf{w}^{(k)}$  represents the high-frequency components missing in  $\mathbf{r}^{(k)}$ , and  $\mathbf{e}^{(k)}$  denotes noise. We apply a linear operation  $\mathcal{R}(\cdot)$  to isolate the missing component  $\mathbf{w}^{(k)}$  from  $\mathbf{x}^{(k)}$ , i.e.,  $\mathbf{w}^{(k)} = \mathcal{R}(\mathbf{x}^{(k)})$ . Here,  $\mathcal{R}(\cdot)$  is defined as  $\mathcal{R} = \mathcal{B} \circ \mathcal{M}$ , where both  $\mathcal{M}$  and  $\mathcal{B}$  correspond to a convolutional layer without bias, with a kernel size of  $3 \times 3$ . It's worth noting that when a sparse transformation satisfies  $\mathcal{F}(\mathbf{x}) = \mathbf{B} \text{ReLU}(\mathbf{A}\mathbf{x})$ , the approximation  $\|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{r}^{(k)})\|^2 \approx \alpha \|\mathbf{x} - \mathbf{r}^{(k)}\|^2$  holds [26], where  $\alpha$  is a scalar related only to the parameters of the transformation  $\mathcal{F}(\cdot)$ . We further decompose  $\mathcal{F}^{(k)}$  into  $\mathcal{F}^{(k)} = \mathcal{H}^{(k)} \circ \mathcal{M}^{(k)}$ , where  $\mathcal{H}^{(k)}$  consists of two convolutional layers without bias plus a ReLU activation function.

Consequently, Eq. (9) can be transformed into

$$\mathbf{x}^{(k)} = \arg \min_{\mathbf{x}} \frac{1}{2} \left\| \mathcal{H}^{(k)}(\mathcal{M}^{(k)}(\mathbf{x})) - \mathcal{H}^{(k)}\left(\mathcal{M}^{(k)}\left(\mathbf{r}^{(k)}\right)\right) \right\|^2 + \theta^{(k)} \|\mathcal{H}^{(k)}(\mathcal{M}^{(k)}(\mathbf{x}))\|_1. \quad (11)$$

Subsequently, we establish the left inverse function of  $\mathcal{H}^{(k)}(\cdot)$ , such that  $\tilde{\mathcal{H}}^{(k)} \circ \mathcal{H}^{(k)} = \mathcal{I}$ , where  $\mathcal{I}$  denotes the identity operator. A symmetric DNN structure of  $\tilde{\mathcal{H}}^{(k)}(\cdot)$  is constructed to mirror  $\mathcal{H}^{(k)}(\cdot)$ , and the constraint  $\tilde{\mathcal{H}}^{(k)} \circ \mathcal{H}^{(k)} = \mathcal{I}$  is incorporated into the loss function. The closed-form expression of  $\mathbf{x}^{(k)}$  is then given by

$$\mathbf{x}^{(k)} = \mathbf{r}^{(k)} + \mathcal{B}^{(k)} \left[ \tilde{\mathcal{H}}^{(k)} \left[ \text{soft} \left[ \mathcal{H}^{(k)} \left( \mathcal{M}^{(k)} \left( \mathbf{r}^{(k)} \right) \right), \theta^{(k)} \right] \right] \right], \quad (12)$$

where the soft threshold function is defined as  $\text{soft}(x, \theta) = \text{sgn}(x) \max(0, |x| - \theta)$ . The network structure corresponding to module  $\mathbf{x}^{(k)}$  is illustrated in the gray box in Fig. 2, with the default kernel number  $C$  set to 32.

The trainable parameters in SPTM2-ISTANet+ are thus  $\{\rho^{(k)}, \theta^{(k)}, \mathcal{H}^{(k)}, \tilde{\mathcal{H}}^{(k)}, \mathcal{M}^{(k)}, \mathcal{B}^{(k)}, \Phi\}$ .

To reconstruct the pre-compressed CSI vector  $\mathbf{x}$  over  $N_I$  iterative modules, an initial estimate  $\mathbf{x}^{(0)}$  is crucial. Unlike the zero-value initialization common in similar networks [8], we employ least squares estimation for initialization, i.e.,  $\mathbf{x}^{(0)} = \mathbf{A}_t \mathbf{y}$ , where  $\mathbf{A}_t$  represents the least square estimate of the transformation matrix  $\mathbf{A}$  in the linear transformation  $\mathbf{x} = \mathbf{A}\mathbf{y}$ . This estimate is computed as  $\mathbf{A}_t = \mathbf{XY}^\top (\mathbf{YY}^\top)^{-1}$ , with  $\mathbf{X} \in \mathbb{R}^{N \times N_T}$  and  $\mathbf{Y} \in \mathbb{R}^{M \times N_T}$  being matrices formed from column vectors  $\mathbf{x}$  and  $\mathbf{y}$  of the training set, respectively. Here,  $N_T$  denotes the total number of training samples.

The optimization of parameters in SPTM2-ISTANet+ necessitates an efficient loss function for training. Defining the size and the  $n$ -th CSI vector of the training set as  $N_T$  and  $\mathbf{x}_n \in \mathbb{R}^N$ , respectively, and the number of iteration modules as  $N_I$ , we construct the loss function as

$$\mathcal{L}_{\text{total}}(\Theta) = \mathcal{L}_{\text{MSE}} + \gamma \cdot \mathcal{L}_{\text{constraint}}, \quad (13)$$

where  $\mathcal{L}_{\text{MSE}} = \frac{1}{N_T N} \sum_{n=1}^{N_T} \|\mathbf{x}_n^{(N_I)} - \mathbf{x}_n\|^2$  is the mean square error, serving as the CSI reconstruction accuracy indicator. The term  $\mathcal{L}_{\text{constraint}} = \frac{1}{N_T N} \sum_{n=1}^{N_T} \sum_{k=1}^{N_I} \left\| \tilde{\mathcal{H}}^{(k)} \left( \mathcal{H}^{(k)} \left( \mathcal{M}^{(k)} \left( \mathbf{r}_n^{(k)} \right) \right) \right) - \mathcal{M}^{(k)} \left( \mathbf{r}_n^{(k)} \right) \right\|^2$  corresponds to the constraint  $\tilde{\mathcal{H}}^{(k)} \circ \mathcal{H}^{(k)} = \mathcal{I}$ , with  $\gamma$  as the regularization weight (set to 0.01 unless specified otherwise).

## V. CSI TRANSLATION ARCHITECTURE CSI-TRANSNET

Addressing the challenge of handling CSIs from various environments is crucial for the practical deployment of DL-based CSI feedback schemes. One solution involves training a specialized network for each scenario or region and dynamically switching to the appropriate feedback network based on the UE-detected channel environment. However, this solution requires UE to have sufficient memory to save many DL-based encoder networks, each of which may include millions of parameters. Alternatively, UE may frequently update its encoder by downloading new model parameters, thereby consuming high cost in terms of wireless bandwidth and UE energy.

In the field of computer vision, image-to-image translation [28] has been employed for diverse applications, including image stylization [29] and segmentation [30]. This technique aims to learn a mapping to convert an image from one domain to another while preserving the essential features of the input images. For instance, it can transform a horse image into one resembling the style of a zebra. Inspired by this approach, we propose a novel CSI feedback architecture, dubbed “CSI-TransNet.” This architecture introduces a lightweight module on the UE side to address the performance challenges of DL-based CSI models in varied environments. By utilizing CSI-to-CSI translation, CSI-TransNet enables the reuse of a pretrained CSI model with high recovery accuracy in new environments.

The comprehensive CSI-TransNet architecture is depicted in Fig. 3. It includes an encoder network on the UE side, equipped with a translation module and a shared pretrained

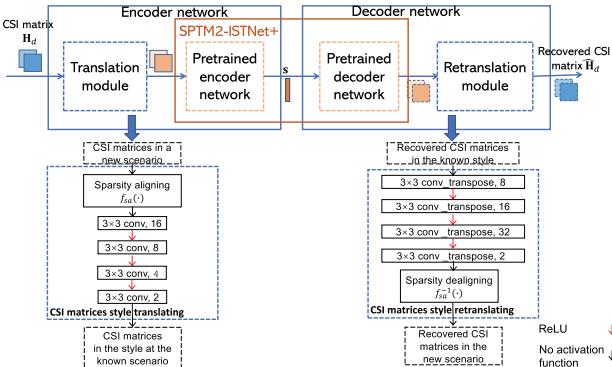


Fig. 3. Architecture of CSI-TransNet.

encoder network. The decoder network at the gNB comprises a shared pretrained decoder network and a customized retranslation module. Notably, the translation and retranslation modules are designed as plug-in modules requiring only a few thousand parameters, facilitating easy updates whenever a UE encounters a new channel environment.

CSI-to-CSI translation simplifies the process of efficiently compressing and accurately recovering CSI matrices using a pretrained model, without the need for additional tuning. Post-translation CSI matrices should possess similar properties to those in the anchor scenario. Unlike data-driven translation methods, as illustrated in Fig. 3, our translation module includes a specialized sparsity aligning function before the DL-based translation network, ensuring lightweight simplicity and effectiveness.

Given that CSI matrices exhibit sparsity in the angular-delay domain [3], we apply a circular shift in this domain to CSI matrices in a new scenario. This shift aims to achieve sparsity characteristics similar to those in the anchor scenario, which was used to pre-train the anchor model. We define the sparsity aligning function as  $f_{sa}(\cdot)$  and the circular shift function as  $f_{sh}(\cdot, i, j)$ , where  $i$  and  $j$  are the shift steps in rows and columns, respectively. For a given CSI matrix  $\mathbf{H}$ , the elements of the circularly shifted CSI matrix  $\mathbf{H}^{sh} = f_{sh}(\mathbf{H}, i, j)$  can be expressed as

$$\mathbf{H}_{m,n}^{sh} = \mathbf{H}_{(m-i) \bmod R_d, (n-j) \bmod N_b}, \forall m \in [R_d], n \in [N_b]. \quad (14)$$

The objective of the sparsity aligning function is to identify shift steps  $i$  and  $j$  that achieve the best similarity. One method is to calculate the shift steps that yield the highest circular cross-correlation [31] for CSI magnitude matrices from the two scenarios. Alternatively, a more direct approach is to calculate  $i$  and  $j$  that maximize the CSI recovery accuracy of the pretrained CSI feedback network, as follows:

$$\min_{i,j} \sum_n \|f_{sh}(\mathbf{H}_n, i, j) - f_{de,1}(f_{en,1}(f_{sh}(\mathbf{H}_n, i, j); \Phi_1); \Psi_1)\|^2 \quad (15)$$

where  $\mathbf{H}_n$  represents the  $n$ -th measured CSI matrix in the new channel scenario. Owing to the translation invariance of convolutional layers, we choose the second method (15) to determine shift steps  $i$  and  $j$ . With a selected pair of shift steps

$i$  and  $j$ , the CSI matrix after sparsity aligning in scenario  $t$  is given by

$$\mathbf{H}^{sa} = f_{sa}(\mathbf{H}^t) = f_{sh}(\mathbf{H}^t, i, j). \quad (16)$$

Following sparsity aligning, the DNN in the translation module comprises 4 convolutional layers for feature extraction and translation. These layers use  $3 \times 3$  kernels to generate 16, 8, 4, and 2 feature maps, respectively, with the first three layers utilizing the ReLU activation function.

The retranslation module is structured as a mirror of the translation module, where a DL network initially fine-tunes the CSI matrix, followed by a sparsity dealigning function  $f_{sa}^{-1}(\cdot)$  to restore sparsity in the new scenario. Transposed convolution layers [32] are used to reverse convolution in the retranslation module. This module contains 4 transposed convolutional layers, generating 32, 16, 8, and 2 feature maps, respectively, using ReLU activation in the first three layers. Utilizing the same shift steps  $i$  and  $j$  from the sparsity aligning, we recover the CSI matrix in scenario  $t$  as

$$\mathbf{H}^t = f_{sa}^{-1}(\hat{\mathbf{H}}^{sa}) = f_{sh}(\hat{\mathbf{H}}^{sa}, -i, -j). \quad (17)$$

Let  $\Theta_t$  and  $\Omega_t$  be DNN parameters in the translation module and retranslation module, respectively. The translation function is denoted as  $f_{tra}(\cdot; \Theta_t)$  and the retranslation function as  $f_{ret}(\cdot; \Omega_t)$ . For CSI-TransNet operating in scenario  $t$ , after determining the shift steps  $i$  and  $j$ , we focus on training only the parameters in the plug-in modules  $f_{tra}(\cdot; \Theta_t)$  and  $f_{ret}(\cdot; \Omega_t)$ . This training utilizes a loss function based on mean square error ( $\mathcal{L}_{MSE}$ ), expressed as:

$$\frac{1}{N_t} \sum_n \|\mathbf{H}_n^{sa} - f_{det,t}(f_{de,1}(f_{en,1}(f_{tra,t}(\mathbf{H}_n^{sa}; \Theta_t); \Phi_1); \Psi_1); \Omega_t)\|^2,$$

where  $N_t$  represents the number of training CSI data samples specific to the  $t$ -th scenario.

Upon detecting a new scenario, a UE within the CSI-TransNet framework reports this change to the gNB serving the area. The gNB then sends the weights of the translation module to the UE to maintain CSI feedback accuracy. In view of the lightweight design, only a few thousand parameters are required by the translation module, which significantly reduces the burden on wireless bandwidth for parameter updates in comparison with the total number of encoder parameters.

## VI. MODEL-DRIVEN DATA AUGMENTATION

In addition to lowering the deployment cost of CSI feedback models in variational environments, a significant challenge lies in training DL models for CSI feedback, which typically requires at least tens of thousands of real-time channel measurements in each scenario. Data augmentation has emerged as an effective solution in many DL models. We investigate how to leverage domain knowledge for data augmentation in compressive CSI feedback and recovery.

### A. GAN-Based Augmentation

Fig. 4(a) illustrates the conventional GAN-based approach for data augmentation. This method involves alternately solving maximization and minimization optimization problems

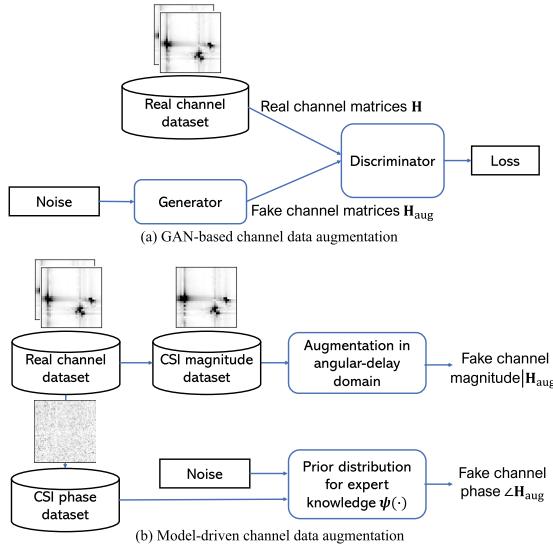


Fig. 4. Architecture comparison between proposed data augmentation and GAN-base data augmentation.

during training to minimize the discrepancy between the distributions of generated and real channels. The generator, employing a DNN, maps a Gaussian noise vector to produce imitations (fake channels). When the GAN successfully generates channels with a distribution matching that of real channels, it can serve as a stable source of augmented CSI data, thereby enlarging the training dataset for CSI-TransNet. However, training a GAN itself demands a significant amount of data samples. Moreover, when available CSI measurement samples only partially represent RF channel features in a coverage area, even a well-trained generator may become a biased channel model, potentially impairing the performance of compressive CSI models trained with such biased data. Therefore, it is crucial not only to obtain sufficient data samples in GAN-based training but also to ensure that augmented samples adequately represent features absent in existing measurement samples.

### B. Model-Driven Augmentation

To improve DNN training for CSI feedback, it is essential that augmented samples exhibit features that are either absent or under-represented in existing measurements. Utilizing domain knowledge, we propose a model-driven augmentation strategy that effectively decouples the characteristics in the magnitude and phase of CSI matrices. We start by separating the phase and magnitude of MIMO channel matrices before augmentation:

$$\mathbf{H} = |\mathbf{H}| \odot e^{j\angle \mathbf{H}}, \quad (18)$$

where  $\odot$  denotes the Hadamard product. The  $(m, n)$ -th element of  $\mathbf{H}$  is expressed as  $\mathbf{H}_{m,n} = |\mathbf{H}_{m,n}| e^{j\angle \mathbf{H}_{m,n}}$ . In this manner, the magnitude CSI matrix is  $|\mathbf{H}|$  with elements  $|\mathbf{H}_{m,n}|$ , and the phase matrix is  $\angle \mathbf{H}$  with elements  $\angle \mathbf{H}_{m,n}$ . Splitting the matrices in this way allows for the application of domain knowledge pertaining to CSI features, including multipath delay profiles and phase distribution.

Next, we utilize the geographical continuity of CSI variation to generate augmented magnitude matrices, which should exhibit similar characteristics to measured channels. These matrices are expected to exhibit characteristics similar to those of measured channels. In typical environments with fixed paths between the gNB and UE, geographical continuity in UE movement leads to smooth variations in the angular-delay domain [33]. This means that CSIs near a measurement spot are highly correlated in the angular-delay domain due to similar arrival/departure angles and delays of multipath propagation. Consequently, we can construct multiple angle-delay profiles by (circularly) shifting the CSI magnitude matrix in angular-delay domain to generate new CSI matrices that reflect features of nearby UE CSIs. Leveraging the circular characteristic of CSI matrices in the angular-delay domain based on the property of DFT, the shifts are circular. In short, the augmented magnitude CSI matrix  $|\mathbf{H}_{m,n}^{\text{aug}}|$  is generated as follows in the angular-delay domain:

$$|\mathbf{H}_{m,n}^{\text{aug}}| = \begin{cases} |\mathbf{H}_{m+i, (n+j) \bmod N_b}|, & 1 \leq m + i \leq R_d, \\ 0, & \text{else,} \end{cases} \quad (19)$$

$\forall m \in [R_d], n \in [N_b]$ . We select shift steps  $\lfloor -\frac{R_d}{2} \rfloor \leq i \leq \lfloor \frac{R_d}{2} \rfloor$  and  $\lfloor -\frac{N_b}{2} \rfloor \leq j \leq \lfloor \frac{N_b}{2} \rfloor$  in the angular and delay domains, respectively, and apply truncation in the delay domain by setting delay elements beyond  $R_d$  rows to zero.

Phase matrix augmentation is also conducted. The augmented phases should cover cases beyond the measured CSIs to enhance training and avoid overfitting. Considering that the path to antenna array arrived at the same delay can share frequency-independent phase shift that includes direct or reflected path [34], we select uniform distribution as augmented phase shifting distribution  $\psi_{\text{aug}}(\cdot) \sim \mathcal{U}(0, 2\pi)$  to elements in the same row (i.e., same delay). In other words, we construct a larger phase variation than the measured CSI phase, and we use recall to replace some precision to enhance CSI recovery accuracy in the practical deployment. Accordingly, for each row of a sampled CSI matrix, a random phase shift is applied:

$$\angle \mathbf{H}_{m,n}^{\text{aug}} = \angle \mathbf{H}_{m,n} + \angle e^{-j\theta_m}, \quad \forall m \in [R_d], n \in [N_b], \quad (20)$$

where  $\theta_m \sim \mathcal{U}(0, 2\pi)$ .

Finally, augmented magnitude and phase matrices are combined to generate CSI samples for training CSI feedback models.

The performance of CSI feedback can be adversely affected when the UE motion areas in the training and testing sets significantly differ within the same scenario. The study presented in [35], which references our arXiv version, optimizes magnitude augmentation to mitigate the delay gap between training and testing sets in the same scenario. This approach offers insights into reducing channel measurement biases when some information about the differences between training and testing sets is known in advance. In our future work, we plan to delve deeper into channel sampling optimization and explore the effects of diverse measurement biases on CSI feedback performance.

## VII. PERFORMANCE EVALUATION

### A. Experiment Setup

For performance evaluation, we utilize four distinct datasets, where the first two are commonly used in CSI estimation and feedback research [3], the third comprises actual field measurements, and the fourth is generated following the 3GPP standard model:

- 1) **Cost2100 Indoor.** This dataset is generated using the COST2100 model [36] with a 5.3 GHz downlink frequency for a gNB located in the center of a  $20\text{ m} \times 20\text{ m}$  area. The bandwidth is set to 20 MHz. It features  $N_b = 32$  antennas and  $N_f = 1024$  subcarriers at the gNB, serving single-antenna UEs randomly distributed within the area.
- 2) **Cost2100 Outdoor.** Generated from the COST2100 model [36] for a 300 MHz downlink frequency, this dataset has a gNB at the center of a  $400\text{ m} \times 400\text{ m}$  area. The bandwidth, number of antennas, and subcarriers are the same as in the Cost2100 Indoor dataset.
- 3) **Measured Indoor.** This dataset includes CSI samples from the KU Leuven Massive MIMO testbed in a  $9\text{ m}^2$  indoor area [37]. The gNB has a uniform linear array with 64 antennas at a 2.61 GHz frequency, of which the first 32 antennas are selected for this dataset.
- 4) **Quadriga 3GPP UMA.** This dataset is generated using the QuaDRiGa platform [38] based on the 3GPP TR 38.901 urban macrocell (UMa) scenario at 2.6 GHz over a 20 MHz bandwidth with  $N_b = 32$  antennas. The gNB is positioned in the center of a  $400\text{ m}$  square area, serving single-antenna UEs randomly located within the area.

For all four channel types, after transforming CSI matrices into the angular-delay domain, only the first 32 rows are retained due to sparsity. The training set size is set to 100,000 samples, and the testing set comprises 20,000 samples. The batch size used is 64. We train the SPTM2-ISTANet+ for 200 epochs, while the translation and retranslation modules in CSI-TransNet, due to their fewer parameters, are trained for 80 epochs.

To compare the recovery accuracy of different networks, we adopt the metric of normalized MSE, i.e.,  $\text{NMSE} = \frac{1}{N_k} \sum_{k=1}^{N_k} \|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2 / \|\mathbf{H}_k\|^2$ , where  $\hat{\mathbf{H}}$  denotes the recovered  $\mathbf{H}$ , and  $k$  and  $N_k$  represent the sample index and the total sample count in the test set, respectively. In addition to NMSE, we also measure the cosine similarity to gauge precoding performance. The cosine similarity metric,  $\rho$ , is expressed as:

$$\rho = \mathbb{E} \left\{ \frac{1}{N_c} \sum_{m=1}^{N_c} \frac{|\hat{\mathbf{h}}_m^H \mathbf{h}_m|}{\|\hat{\mathbf{h}}_m\|_2 \|\mathbf{h}_m\|_2} \right\}, \quad (21)$$

where  $\hat{\mathbf{h}}_m$  signifies the reconstructed channel vector for the  $m$ -th subcarrier. When the gNB uses  $\mathbf{v}_m = \hat{\mathbf{h}}_m / \|\hat{\mathbf{h}}_m\|$  as a beamforming vector (i.e., as in zero-forcing precoding), the cosine similarity can be indicative of the precoding gain.

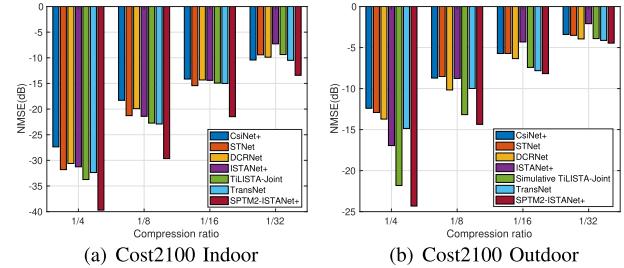


Fig. 5. NMSE comparison in different CRs.

### B. Feedback Accuracy Comparison of the Anchor Model

In our comparative analysis, SPTM2-ISTANet+ is evaluated against six established CSI feedback schemes known for their performance in massive MIMO systems. These schemes do not rely on additional auxiliary information such as uplink CSI or previous CSIs:

- **CsiNet+** [5]: This model enhances CSI feedback by employing larger convolution kernels ( $7 \times 7$ ) and optimizing the structure of residual units.
- **DCRNet** [6]: This model integrates multiple resolution convolution kernels with dilated convolutions to extract CSI features at different granularities. It also utilizes a warm-up process for optimizing learning rate adjustment.
- **ISTANet+** [39]: Inspired by compressive image processing [26], this CSI feedback design adopts an orthogonal random Gaussian measurement matrix  $\Phi \in \mathbb{R}^{M \times N}$ , with  $N = 2048$  and  $M$  dependent on the Compression Ratio (CR).
- **TiLISTA-Joint** [8]: A network inspired by compressive sensing, featuring a learnable sparse transformation module to enhance CSI recovery accuracy. Different from TiLISTA-Joint, SPTM2-ISTANet+ exploits the spherical structure to mitigate the effect of the channel variation, and residual recovery design and the least squares estimation for initialization instead of initializing with zero values are utilized to enhance the recovery accuracy and convergence. We use the 9-iteration TiLISTA-Joint results in [8] for the Cost 2100 indoor scenario, and we modify the itemized key differences in SPTM2-ISTANet+ to mimic the TiLISTA-Joint (termed as simulative TiLISTA-Joint) for the Cost 2100 outdoor scenario.
- **TransNet** [9] and **STNet** [10]: TransNet introduces an attention mechanism through a two-layer Transformer architecture, enabling CSI to recognize connections between its components during feedback. STNet is a lightweight transformer-based model employing a spatially separable attention mechanism, offering reduced complexity.

Fig. 5 presents a comparison of CSI feedback performance among seven schemes: CsiNet+, STNet, DCRNet, ISTANet+, TiLISTA-Joint, TransNet, and our SPTM2-ISTANet+, across different Compression Ratios (CRs) in indoor and outdoor scenarios using the Cost2100 model. These two datasets are commonly used in CSI feedback works [4], [5], [6], [7]. With the number of iteration blocks set to 9, Fig. 5(a) and Fig. 5(b) illustrate that SPTM2-ISTANet+ consistently outperforms the others at all tested CRs. Notably, in the

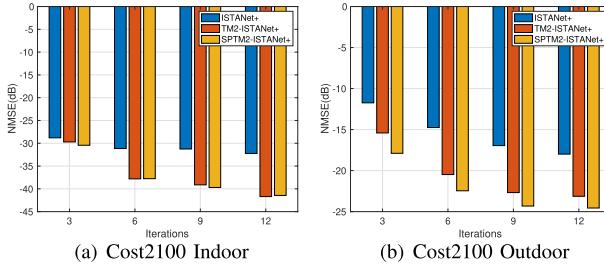


Fig. 6. NMSE at different iteration blocks for CR = 1/4.

outdoor scenario at a CR of 1/4, SPTM2-ISTANet+ offers a substantial improvement of approximately 7 dB in CSI reconstruction accuracy over traditional DL-based schemes like STNet, DCRNet, TransNet, and CsiNet+. TiLISTA-Joint exhibits better performance than these traditional schemes at larger CRs (e.g., 1/4), but its efficacy decreases at higher compression (smaller CRs), such as 1/32, while SPTM2-ISTANet+ maintains its superiority, underscoring the benefits of our spherical processing and residual design.

Fig. 6 displays the effect of the number of iteration blocks  $N_I$  on the CSI feedback performance of ISTANet+, TM2-ISTANet+ (a variant of SPTM2-ISTANet+ without spherical processing), and SPTM2-ISTANet+ at a CR of 1/4 in indoor and outdoor scenarios against the Cost2100 channel model. SPTM2-ISTANet+ consistently delivers the best performance across various iterations, achieving comparable results to ISTANet+ with 12 iterations in just 3 iterations. The performance of SPTM2-ISTANet+ stabilizes after reaching 9 iteration modules. For indoor channels at CR = 1/4, 3 iteration modules are sufficient, as the NMSE drops below -30 dB. However, in outdoor scenarios, due to the complexity of the channels and CSI variations, up to 9 iterations are needed.

### C. Data Augmentation Comparison

Since DL-based CSI feedback works have already achieved satisfactory performance for the indoor scenario [4], [5], [6], [7], we focus on the outdoor scenario where practical CSI measurement is even harder and CSI recovery is less accurate. For data augmentation in the Cost2100 Outdoor scenario, we randomly select samples from the limited measurements in the training set. The shift ranges are set from -15 to 15 in the angular domain and from -3 to 3 in the delay domain. The training dataset size after augmentation is maintained at the same level as the overall training set size, using repetition or phase randomization if the initial augmentation is not large enough, depending on the augmentation method used.

Fig. 7 presents a comparison of the performance of SPTM2-ISTANet+ using various augmentation strategies at a CR of 1/4. These strategies include the computationally expensive “ChannelGAN” [21], no augmentation (“No Aug”) which employs repetition for dataset enlargement, flipping, rotation, shift in the angular-delay domain (“ADS”) with dataset enlargement via repetition as needed, single-value random phase shift (“PRS-guo22”) [18] where all CSI matrix elements share the same phase shift, random phase shift (“PRS”), and a combination of ADS and PRS (ADS+PRS). The

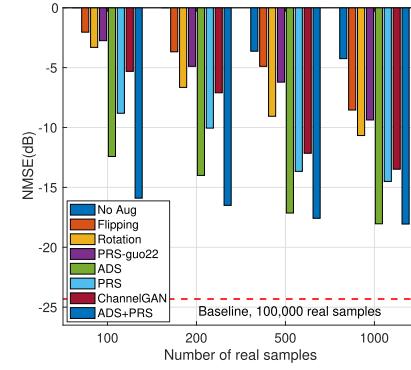


Fig. 7. Augmentation comparison at CR = 1/4.

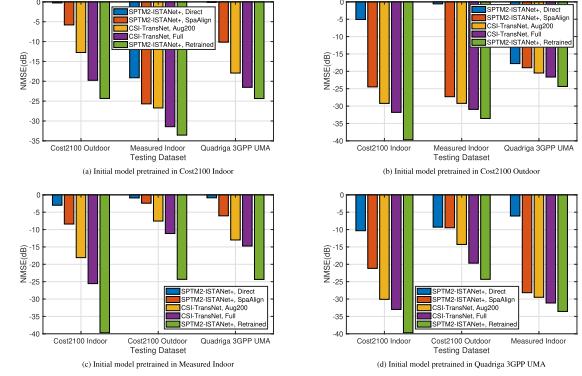


Fig. 8. NMSE comparison for the model pretrained in one scenario to the new scenarios when CR = 1/4.

number of CSI measurements before augmentation is limited to 100, 200, 500, and 1000. As illustrated in Fig. 7, our ADS and ADS+PRS methods significantly outperform ChannelGAN and other augmentation techniques. Notably, PRS alone already surpasses ChannelGAN. Additionally, our PRS method yields a gain of about 5 dB compared to PRS-guo22 by accounting for varying phase shifts among different paths. Particularly, ADS+PRS achieves an NMSE of -15.8 dB with only 100 CSI measurement samples, while ChannelGAN reaches only -5.3 dB. Furthermore, both ADS and PRS enhance CSI recovery accuracy. Overall, the proposed low-cost training enhancement using ADS consistently achieves higher gains than PRS, thanks to better exploitation of geographical correlation based on domain knowledge.

### D. Recovery Comparison Against New Scenarios

To evaluate recovery performance when a UE encounters a new scenario, we selected models pretrained on Cost2100 Indoor, Cost2100 Outdoor, Measured Indoor, and Quadriga 3GPP UMA datasets. For each model, we used the other three datasets, unseen by the pretrained model, as new scenarios.

Fig. 8 displays the CSI feedback performance of models pretrained in one scenario and tested in new scenarios at a CR of 1/4. “SPTM2-ISTANet+, Direct” shows the performance of the pretrained SPTM2-ISTANet+ in the new scenario without preprocessing, while “SPTM2-ISTANet+, SpaAlign” illustrates the performance with the aid of the sparsity aligning function  $f_{sa}(\cdot)$ . “CSI-TransNet, Aug200” indicates the performance of CSI-TransNet with the proposed ADS+PRS augmentation starting with 200 samples from the new scenario.

TABLE I

COSINE SIMILARITY COMPARISON FOR THE MODEL PRETRAINED IN ONE SCENARIO TO THE NEW SCENARIOS WHEN CR = 1/4

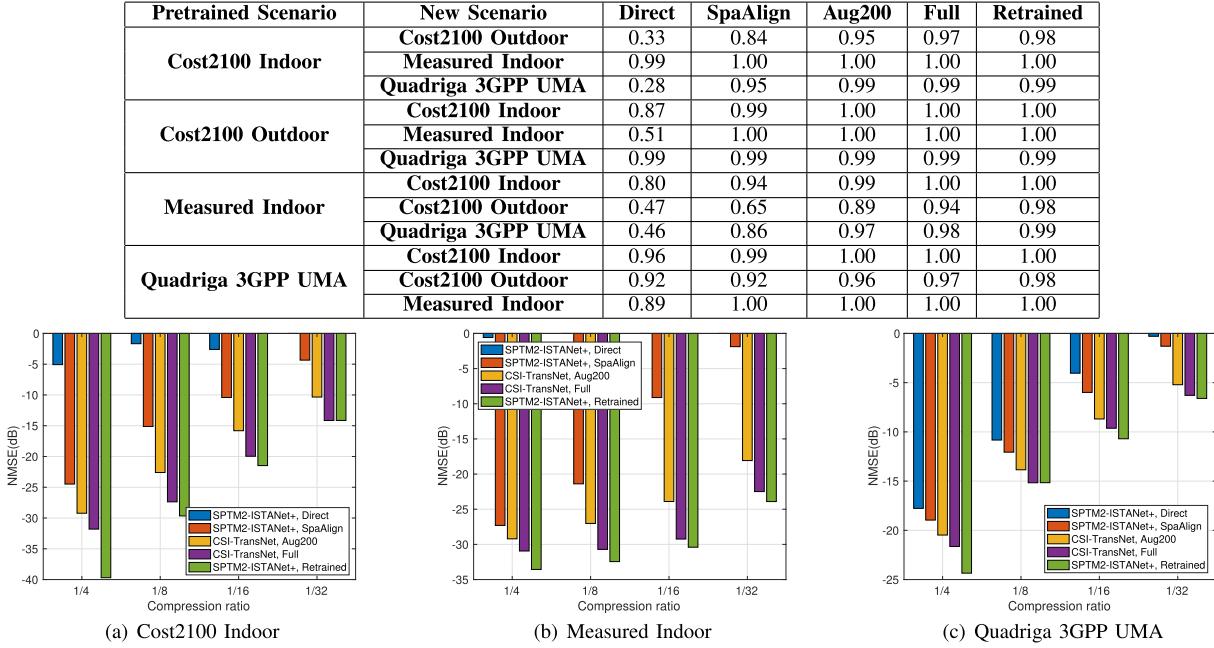


Fig. 9. NMSE comparison in different CRs for the model pretrained in the Cost2100 Outdoor dataset to the new scenarios.

“CSI-TransNet, Full” and “SPTM2-ISTANet+, Retrained” serve as baselines, representing CSI-TransNet trained with full training sets and SPTM2-ISTANet+ retrained from scratch in the new scenario, respectively.

As Fig. 8 shows, “SPTM2-ISTANet+, Direct” generally suffers from poor CSI recovery accuracy, which aligns with the known limitation of DL-based CSI feedback in handling general CSIs. With the help of sparsity aligning, the models pretrained on Cost2100 Outdoor and Quadriga 3GPP UMA demonstrate clear accuracy improvement in the new scenarios. Actually, compared with Cost2100 Indoor and Measured Indoor, CSI in Cost2100 Outdoor and Quadriga 3GPP UMA exhibit more diverse features in terms of multipath delay and angle of arrival/departure distributions. This allows models pretrained in more diverse environments to tackle more general channel scenarios.

With limited measurement data, “CSI-TransNet, Aug200” illustrates obvious performance improvement for each case, especially for the anchor model pretrained on Cost2100 Indoor and Measured Indoor. This observation confirms the effectiveness of our proposed plug-in translation module and model-driven channel data augmentation. “CSI-TransNet, Full” provides additional gains over “CSI-TransNet, Aug200”, which means that a more accurate CSI-to-CSI translation can benefit more from the pretrained model. “SPTM2-ISTANet+, Retrained” is selected as the performance bound for the CSI feedback accuracy. We observe that the models pretrained on Cost2100 Outdoor exhibit smaller gaps to the retraining bound. Consequently, we suggest using the model pretrained in a more diverse (complex) environment (such as the outdoor) to serve as the anchor network in CSI-TransNet.

Table I presents the cosine similarity of models pretrained in one scenario when applied to new scenarios at CR = 1/4.

For brevity, the terms “SPTM2-ISTANet+, Direct”, “SPTM2-ISTANet+, SpaAlign”, “CSI-TransNet, Aug200”, “CSI-TransNet, Full”, and “SPTM2-ISTANet+, Retrained” are abbreviated as “Direct”, “SpaAlign”, “Aug200”, “Full”, and “Retrained”, respectively. As depicted in Table I, and consistent with the NMSE results, our proposed plug-in translation module and model-driven channel data augmentation yield effective precoding gains compared to the direct reuse of the pretrained model in new scenarios, and anchor models pretrained on Cost2100 Outdoor overall register the best cosine similarity performance. Furthermore, the pairings of Cost2100 Indoor with Measured Indoor, and Cost2100 Outdoor with Quadriga 3GPP UMA demonstrate enhanced performance when reused, highlighting the potential to classify wireless environments based on similar features for model reuse to minimize model update overheads.

We extend our analysis to assess CSI recovery across different CRs. Fig. 9 and Table II illustrate the NMSE and cosine similarity for a model pretrained using the Cost2100 Outdoor dataset, applied to various scenarios. A significant finding is the comparable performance of “CSI-TransNet, Full” to “SPTM2-ISTANet+, Retrained” across different CRs. This observation underscores the effectiveness of our plug-in CSI-to-CSI translation design in reusing pretrained models and weights. Furthermore, “CSI-TransNet, Aug200” shows robust performance, achieving an NMSE below -20 dB at CR = 1/4 and below -8.6 dB at CR = 1/16. Correspondingly, it reaches a cosine similarity exceeding 0.99 at CR = 1/4 and above 0.93 at CR = 1/16. These results demonstrate the capability of “CSI-TransNet, Aug200” as a versatile anchor network, suggesting that backbone networks pretrained in diverse environments are advantageous for scenario-adaptive designs.

TABLE II  
COSINE SIMILARITY COMPARISON IN DIFFERENT CRs FOR THE MODEL PRETRAINED IN THE COST2100 OUTDOOR DATASET TO THE NEW SCENARIOS

Scenario		Direct	SpaAlign	Aug200	Full	Retrained
Cost2100 Indoor	$CR=\frac{1}{4}$	0.87	0.99	1.00	1.00	1.00
	$CR=\frac{1}{8}$	0.69	0.98	0.99	1.00	1.00
	$CR=\frac{1}{16}$	0.69	0.96	0.98	0.99	0.99
	$CR=\frac{1}{32}$	0.19	0.81	0.96	0.98	0.98
Measured Indoor	$CR=\frac{1}{4}$	0.51	1.00	1.00	1.00	1.00
	$CR=\frac{1}{8}$	0.24	1.00	1.00	1.00	1.00
	$CR=\frac{1}{16}$	0.18	0.94	1.00	1.00	1.00
	$CR=\frac{1}{32}$	0.17	0.62	0.99	1.00	1.00
Quadriga 3GPP UMA	$CR=\frac{1}{4}$	0.99	0.99	0.99	0.99	0.99
	$CR=\frac{1}{8}$	0.95	0.96	0.97	0.98	0.98
	$CR=\frac{1}{16}$	0.78	0.86	0.93	0.94	0.94
	$CR=\frac{1}{32}$	0.32	0.53	0.83	0.87	0.85

Interestingly, note that carrier frequencies of Cost2100 Indoor, Measured Indoor and Quadriga 3GPP UMA are different from those of Cost2100 Outdoor. In fact, the subcarrier spacing of Measured Indoor is also different from Cost2100 Outdoor. This flexibility in terms of carrier frequencies and subcarrier spacing indicates that the proposed CSI-TransNet architecture can effectively reuse pretrained networks across different scenarios. By requiring only a limited number of new measurements from a new environment, CSI-TransNet substantially reduces the cost of deployment in diverse and complex wireless environments.

In the context of collecting CSI samples for training the translation module and the backbone network in real-world scenarios, we suggest several strategies to construct robust and diverse datasets. Firstly, crowdsourcing data collection capitalizes on user participation to accumulate a wide range of downlink CSI data [40], providing a broad spectrum of channel conditions for training. Secondly, synthetic data generation can act as a viable substitute in the absence of sufficient real-world data. This approach employs ray-tracing tools such as Remcom's Wireless InSite [41] and integrates 3D maps from sources like OpenStreetMap [42] to generate realistic channel data. Additionally, uplink CSI-based model training, which uses uplink CSI data for downlink feedback models [43], [44], presents an effective method to reduce the signaling overhead of downlink CSI sampling while still maintaining efficient model performance. These strategies collectively contribute to the creation of comprehensive and varied datasets, crucial for the optimal functioning of the translation module across different scenarios.

#### E. Recovery Comparison to One-Sided CSI Feedback

The one-sided CSI reconstruction model, where only the gNB employs a retrained model for CSI reconstruction in different scenarios, is a potential alternative to our CSI-TransNet, yet it faces specific limitations. Notably, a shared encoder in this model can lead to reduced CSI recovery accuracy at lower compression ratios (CR), as it might not capture sufficient environment-specific features [25]. Contrarily, CSI-TransNet's translation module adeptly handles diverse channel conditions, even those significantly different from the channel sampling of the backbone scenario, ensuring enhanced CSI feedback accuracy.

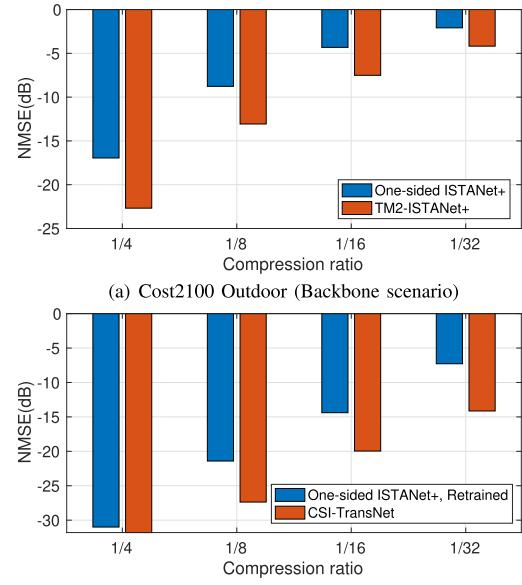


Fig. 10. NMSE comparison between the one-sided CSI feedback solution and our proposed solution in different scenarios.

To quantitatively assess the performance disparity between our CSI-TransNet and the one-sided CSI feedback model, we utilize the one-sided encoder detailed in [25], paired with an identical backbone decoder for a fair comparison. Fig. 10(a) illustrates that the trainable encoder-assisted TM2-ISTANet+ exhibits marked improvement over the one-sided ISTANet+ model, primarily due to its enhanced capability in channel characteristic extraction. Furthermore, when comparing CSI-TransNet, whose backbone model is pretrained in the Cost2100 outdoor scenario, with the retrained one-sided ISTANet+ in a new Cost2100 indoor scenario, CSI-TransNet outperforms the retrained one-sided model, particularly at smaller CRs. This superiority is demonstrated in Fig. 10(b), where CSI-TransNet exhibits a more effective feature extraction in compact formats. Recognizing the potential of the one-sided model as a simpler but less efficient alternative, our future work will investigate its capabilities and potential enhancements.

#### F. Robustness Evaluation

In this subsection, the robustness of CSI-TransNet is evaluated with a focus on the impact of codewords quantization and channel noise.

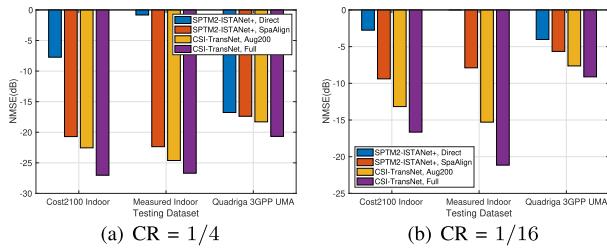


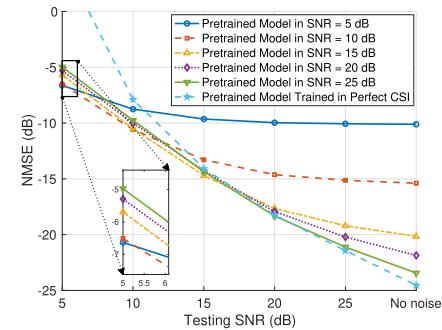
Fig. 11. NMSE at 4-bit quantization for the model pretrained in the Cost2100 Outdoor dataset to the new scenarios.

Considering the necessity of transmitting CSI feedback as a bitstream, the role of quantization in this process is crucial. While several DL-based CSI feedback studies have incorporated quantization [45], [46], we specifically assess its impact within our CSI-TransNet. We employ the  $\mu$ -law quantization technique, as detailed in [46], applying it to the dimension-compressed codewords of our backbone network, SPTM2-ISTANet+. The network undergoes fine-tuning to adapt to the quantization process. After applying 4-bit quantization, SPTM2-ISTANet+ achieves NMSE values of  $-20.2$  dB for CR = 1/4 and  $-7.3$  dB for CR = 1/16 in the Cost2100 Outdoor scenario. Subsequently, we examine the performance of 4-bit quantization on the backbone model pretrained for the Cost2100 Outdoor scenario when applied to new scenarios.

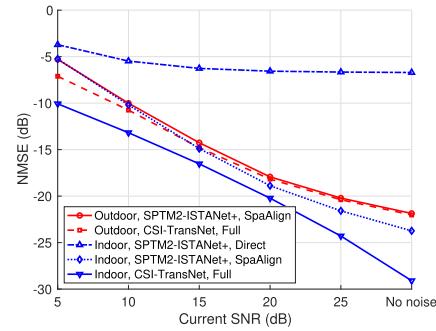
Utilizing the model pretrained with 4-bit quantization for the Cost2100 Outdoor scenario, Fig. 11(a) and Fig. 11(b) illustrate the NMSE performance of CSI-TransNet across various scenarios. Remarkably, “CSI-TransNet, Full” consistently shows significant improvements in CSI recovery accuracy for different CRs compared to directly reusing the pretrained network. This highlights the effectiveness of our plug-in CSI-to-CSI translation design in efficiently leveraging pretrained models and weights. Furthermore, “CSI-TransNet, Aug200” registers NMSE values below  $-18.0$  dB and  $-7.6$  dB at CR = 1/4 and CR = 1/16, respectively, across different scenarios. These results underscore the resilience and robustness of CSI-TransNet, even when subjected to quantization effects.

To evaluate the robustness of SPTM2-ISTANet+ and the CSI-TransNet framework, we investigate their reconstruction performance in scenarios affected by channel estimation errors caused by channel noise. Specifically, the model was trained using noisy CSI data as input while targeting clean CSI as the output, effectively enabling denoising. Fig. 12(a) explores the adaptability of the pretrained SPTM2-ISTANet+ to variations in signal-to-noise ratios (SNRs) between the training and testing phases. Concentrating on the Cost2100 outdoor scenario, we assess the massive MIMO CSI recovery performance across different SNR levels, with a CR set to 1/4. The findings in Fig. 12(a) reveal a significant performance drop in the backbone network when the testing SNR is lower than the SNR used during training, especially for models trained under ideal channel conditions. In contrast, for models pretrained at lower SNRs, the performance tends to level off when the testing SNR surpasses the training SNR.

In Fig. 12(a), we can also observe the superior performance of the model trained at SNR = 25 dB over that at SNR = 20 dB, when both are tested at 20 dB. This



(a) Generalizability of SPTM2-ISTANet+ pretrained in noisy Cost2100 outdoor



(b) Performance evaluation of CSI-TransNet using the backbone model pretrained in Cost2100 outdoor SNR = 20 dB

Fig. 12. Evaluation of SPTM2-ISTANet+ and CSI-TransNet over channel SNR variations ranges.

observation can be attributed to two factors. Firstly, training at the higher SNR of 25 dB exposes the model to cleaner data, enhancing its ability to learn channel characteristics, thereby performing better in high-quality channel scenarios. Secondly, the training process at SNR = 20 dB is more likely to be negatively impacted by noise, potentially impeding the model’s convergence to an optimal training point, thus affecting performance. Additionally, as indicated in Fig. 12(a), the model trained at SNR = 20 dB demonstrates superior denoising capabilities in lower SNR environments, such as at 5 dB, compared to its counterpart trained at SNR = 25 dB. Notably, models pretrained at SNR values of 20 dB and 25 dB in the Cost2100 outdoor scenario exhibit robust anti-noise characteristics and maintain high recovery accuracy at high SNR levels. This indicates their potential as backbone models capable of operating effectively across a wide range of SNRs and in various environments. Given that obtaining channel samples at SNR = 20 dB is relatively easier compared to SNR = 25 dB, the model pretrained in the Cost2100 outdoor scenario at SNR = 20 dB is preferred as the backbone model.

Fig. 12(b) evaluates the robustness of the CSI-TransNet framework, employing the model pretrained in the Cost2100 outdoor scenario at SNR = 20 dB and CR = 1/4 as the backbone model in a new scenario and SNR. The translation model in CSI-TransNet serves not only for cross-environment CSI-to-CSI translation but also for denoising. As depicted in Fig. 12(b), using the model pretrained at SNR = 20 dB in the Cost2100 outdoor scenario as the backbone, both the sparsity aligning and CSI-to-CSI translation capabilities of CSI-TransNet significantly enhance performance in the

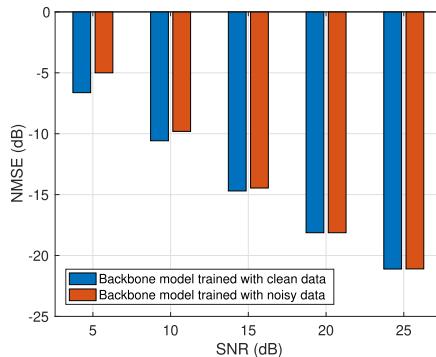


Fig. 13. Performance comparison of the model trained with clean and noisy data at different SNR levels in the Cost2100 outdoor scenario when CR = 1/4.

Cost2100 indoor noisy scenario. This observation confirms that CSI-TransNet is robust and adaptable to varying SNR conditions. Notably, the recovery accuracy of CSI-TransNet reaches nearly  $-10$  dB even when the SNR is as low as 5 dB in the Cost2100 indoor scenario, underscoring its effectiveness in challenging noise conditions.

In practical wireless systems, obtaining perfect CSI is challenging. To tackle this, we evaluate the performance of the backbone model using both pristine and corrupted CSI data. Specifically, the model was trained using noisy CSI as input against clean CSI as output for denoising (termed “clean data”), and using noisy CSI for both input and output (termed “noisy data”). Fig. 13 depicts the model’s performance across different SNR levels in the Cost2100 outdoor scenario with a CR of 1/4. As shown in Fig. 13, the model trained on noisy data performs comparably to the one trained on clean data, especially at SNR levels above 10 dB. This can be attributed to the encoder’s ability to compress the input into a lower-dimensional, noise-reduced representation, which the decoder then uses to reconstruct a denoised image. However, feedback performance deteriorates at lower SNR levels.

To improve channel acquisition in scenarios with lower SNR, denoising neural networks are beneficial, as indicated in [47] and [48]. For example, CBDNet, a convolutional denoising approach introduced in [47], enhances channel estimation accuracy by integrating a noise estimation network and reducing the discrepancy between the noisy channel matrix and its estimated noise counterpart. CBDNet significantly bolsters channel estimation in massive MIMO systems, delivering NMSE values of  $-23$  dB and  $-27$  dB at SNR levels of 10 dB and 15 dB, respectively. Although denoising optimization is not our main focus, preliminary performance evaluations of our CSI-TransNet with denoised CSI matrices at SNR = 10 dB and 15 dB are conducted.

Fig. 14 contrasts the CSI-TransNet framework’s performance using initial noisy CSI estimates versus denoised CSI estimates in training the CSI feedback networks. Employing the model pretrained in the Cost2100 outdoor scenario at a CR of 1/4 as a foundation, we observed notable enhancements when utilizing denoised CSI estimates at SNR levels of 10 dB and 15 dB. Incorporating denoising preprocessing can thus substantially elevate CSI feedback precision in low-SNR environments. Future endeavors will concentrate on integrating

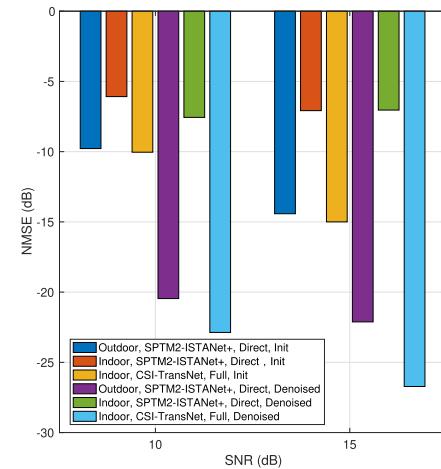


Fig. 14. Comparative performance evaluation of CSI-TransNet using initial and denoised CSI in the Cost2100 outdoor scenario when CR = 1/4.

channel denoising with low-overhead feedback optimization, aiming to enhance the resilience of DL-based CSI feedback mechanisms in practical massive MIMO systems.

#### G. Complexity Comparison

We delve into a discussion on algorithm complexity, focusing on the cost-sensitive UE encoder and translation module. Tables III and IV provide a comparative analysis of parameters and Floating Point Operations (FLOPs) for these components. We have chosen not to compare decoder networks, as the computational power and energy resources of gNBs (which are typically resource-rich) are less critical in this context.

The comparisons reveal that SPTM2-ISTANet+ substantially reduces computation requirements at the UE, achieving over 88%, 28%, and 19% reduction in comparison with TransNet, STNet, and DCRNet respectively at a CR of 1/4. The computational savings increase as the CR decreases. The number of parameters for these encoder networks is at a similar level, with the exception of TransNet, which uses at least an additional 30% parameters. Notably, our plug-in translation module significantly reduces the number of UE parameters that need updating in new scenarios, dropping from around one million parameters to merely 1.8 thousand. This is a stark contrast to conventional encoder networks. Additionally, TransNet’s computational demand is exponentially higher due to its transformer-based architecture, while SPTM2-ISTANet+ is more economical. Despite the FLOPs of the translation module being double that of SPTM2-ISTANet+ at CR = 1/4, the complexity of convolutional layers can be mitigated by applying lightweight optimization techniques [49], such as depth-wise convolution [50] and knowledge distillation [51].

For practical understanding, Table V offers runtime comparisons among different encoder networks. Tests conducted on the PyTorch platform using the Nvidia GeForce RTX 3080 GPU show that SPTM2-ISTANet+ has the shortest runtime, while TransNet has the longest. Furthermore, the CSI-TransNet framework, which includes both the TransModule and SPTM2-ISTANet+, demonstrates improved runtime efficiency compared to TransNet and STNet. This highlights the practical effectiveness of our proposed translation design in real-world scenarios.

TABLE III  
MODEL SIZE (PARAMETERS) OF ENCODER NETWORKS IN UE. M: MILLION, K: THOUSAND

	DCRNet	STNet	TransNet	SPTM2-ISTANet+	TransModule
<b>CR=</b> $\frac{1}{4}$	1.0 M	1.1 M	1.3 M	1.0 M	1.8 K
<b>CR=</b> $\frac{1}{8}$	0.5 M	0.5 M	0.8 M	0.5 M	1.8 K
<b>CR=</b> $\frac{1}{16}$	0.3 M	0.3 M	0.5 M	0.3 M	1.8 K
<b>CR=</b> $\frac{1}{32}$	0.1 M	0.1 M	0.4 M	0.1 M	1.8 K

TABLE IV  
COMPUTATIONAL COMPLEXITY (FLOPS) OF ENCODER NETWORKS IN UE. M: MILLION

	DCRNet	STNet	TransNet	SPTM2-ISTANet+	TransModule
<b>CR=</b> $\frac{1}{4}$	2.6 M	2.9 M	17.9 M	2.1 M	3.7 M
<b>CR=</b> $\frac{1}{8}$	1.6 M	2.4 M	17.3 M	1.0 M	3.7 M
<b>CR=</b> $\frac{1}{16}$	1.0 M	2.2 M	17.1 M	0.5 M	3.7 M
<b>CR=</b> $\frac{1}{32}$	0.8 M	2.0 M	16.9 M	0.3 M	3.7 M

TABLE V  
RUNTIME OF ENCODER NETWORKS (IN  $10^{-5}$  SECOND)

	DCRNet	STNet	TransNet	SPTM2-ISTANet+	CSI-TransNet	TransModule
<b>CR=</b> $\frac{1}{4}$	0.5	1.2	1.3	0.1	0.8	0.8
<b>CR=</b> $\frac{1}{8}$	0.5	1.2	1.3	0.1	0.8	0.8
<b>CR=</b> $\frac{1}{16}$	0.5	1.2	1.3	0.1	0.8	0.8
<b>CR=</b> $\frac{1}{32}$	0.5	1.2	1.3	0.1	0.8	0.8

TABLE VI  
PARAMETERS AND COMPUTATIONAL COMPLEXITY OF AUGMENTATION STRATEGIES. B: BILLION, M: MILLION, K: THOUSAND

	ChannelGAN	ADS	PRS	ADS+PRS
<b>Parameters</b>	11.7 M	0.2 K	32	0.2 K
<b>FLOPs</b>	5.4 B	-	4.1 K	4.1 K

Table VI compares parameters and FLOPs of different augmentation strategies. Unlike ChannelGAN which requires millions of parameters and billions of FLOPs, our proposed ADS+PRS only needs several thousand parameters and FLOPs to achieve a higher CSI recovery accuracy.

In our performance evaluation, we recognize that the translation module in our AI model, while modest in parameter count, requires careful consideration of its computational demands. Industry technical report, according to the information in CSI\_Table I, CSI compression 1-on-1 joint training of 3GPP TR 38.843 [52], reveals that encoder networks used by major players like Huawei, Qualcomm, Apple, and Nokia typically exceed 10 million FLOPs. Our solution's combined computational load, encompassing both the translation module and encoder network, is less than 6 million FLOPs, well within this industry-accepted threshold. Moreover, the floating point operations per second (FLOPS) capacity of contemporary smartphone GPUs, as indicated by recent data [53], can exceed 2.1 TeraFLOPS ( $\times 10^{12}$ ), affirming that the computational demands of our module are manageable for current mobile devices. Hence, the computational complexity of our translation module is in line with industry standards and suitable for modern smartphones. We are committed to further optimizing our model for an ideal balance between performance

and computational efficiency, ensuring its viability in mobile communication systems.

## VIII. CONCLUSION

This work introduces a novel solution for enhancing the training and deployment of DL models in massive MIMO CSI feedback. We consider two major obstacles to DL-based feedback frameworks: adapting to new unseen channel environments and coping with limited training datasets derived from field measurements. For new channel environments, we present an efficient scenario-adaptive CSI feedback architecture “CSI-TransNet”. CSI-TransNet exploits the plug-in CSI-to-CSI translation module to reuse the pretrained anchor CSI model with high recovery accuracy and enables a lightweight encoder update in new scenarios. Additionally, we have developed SPTM2-ISTANet+, an efficient deep unfolding-based CSI feedback network, which serves as the backbone of CSI-TransNet. Against small measurement datasets, we propose a simple yet effective data augmentation strategy grounded in domain knowledge to replace block-box GAN-based augmentation. With our proposed augmentation strategy in conjunction with CSI-TransNet, we significantly enhance CSI recovery performance. Remarkably, with only about a thousand encoder parameters requiring updates, CSI-TransNet can achieve an NMSE below  $-20$  dB using only 200 measurement channel samples in three new scenarios at a CR of 1/4.

The use of prior knowledge about channel variations has proven to be key in improving both data augmentation and the CSI feedback network. While our current focus is on leveraging insights from wireless channel behavior to refine CSI data augmentation and feedback network optimization without

auxiliary data, there is still unexplored potential. The dynamic nature of wireless channels, characterized by their temporal and spatial correlations due to shared local scattering clusters, presents opportunities to reduce redundant information in time-varying channels. Going forward, we aim to investigate precise channel variation modeling and the incorporation of side-information into model-driven DL networks. This future direction holds the promise of further optimizing CSI feedback efficiency in diverse environments.

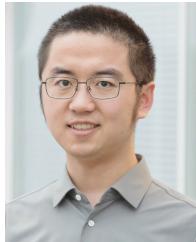
### ACKNOWLEDGMENT

The authors would like to acknowledge Yu-Chien Lin for his useful discussions in the process of preparing this manuscript.

### REFERENCES

- [1] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.
- [2] H. Son and Y.-H. Cho, "Analysis of compressed CSI feedback in MISO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1671–1674, Dec. 2019.
- [3] C. Wen, W. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.
- [4] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [5] J. Guo, C. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2827–2840, Apr. 2020.
- [6] S. Tang, J. Xia, L. Fan, X. Lei, W. Xu, and A. Nallanathan, "Dilated convolution based CSI feedback compression for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11216–11221, Oct. 2022.
- [7] X. Chen, C. Deng, B. Zhou, H. Zhang, G. Yang, and S. Ma, "High-accuracy CSI feedback with super-resolution network for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 141–145, Jan. 2022.
- [8] Y. Wang, X. Chen, H. Yin, and W. Wang, "Learnable sparse transformation-based massive MIMO CSI recovery network," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1468–1471, Jul. 2020.
- [9] Y. Cui, A. Guo, and C. Song, "TransNet: Full attention network for CSI feedback in FDD massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 903–907, May 2022.
- [10] S. Mourya, S. Amuru, and K. K. Kuchi, "A spatially separable attention mechanism for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 40–44, Jan. 2023.
- [11] Y. Lin, Z. Liu, T. Lee, and Z. Ding, "Deep learning phase compression for MIMO CSI feedback by exploiting FDD channel reciprocity," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2200–2204, Oct. 2021.
- [12] J. Guo, C.-K. Wen, and S. Jin, "CANet: Uplink-aided downlink channel acquisition in FDD massive MIMO using deep learning," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 199–214, Jan. 2022.
- [13] M. K. Shehzad, L. Rose, S. Wesemann, and M. Assaad, "ML-based massive MIMO channel prediction: Does it work on real-world data?" *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 811–815, Apr. 2022.
- [14] M. K. Shehzad, L. Rose, and M. Assaad, "Dealing with CSI compression to reduce losses and overhead: An artificial intelligence approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.
- [15] Z. Liu, M. del Rosario, and Z. Ding, "A Markovian model-driven deep learning framework for massive MIMO CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1214–1228, Feb. 2022.
- [16] T. Wang, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, Apr. 2019.
- [17] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Jul. 2019.
- [18] J. Guo, Y. Zuo, C.-K. Wen, and S. Jin, "User-centric online gossip training for autoencoder-based CSI feedback," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 559–572, Apr. 2022.
- [19] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, and C.-X. Wang, "Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 22–27, Mar. 2019.
- [20] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3133–3143, May 2020.
- [21] H. Xiao, W. Tian, W. Liu, and J. Shen, "ChannelGAN: Deep learning-based channel modeling and generating," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 650–654, Mar. 2022.
- [22] Y. Yang, F. Gao, Z. Zhong, B. Ai, and A. Alkhateeb, "Deep transfer learning-based downlink channel prediction for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7485–7497, Dec. 2020.
- [23] J. Zeng et al., "Downlink CSI feedback algorithm with deep transfer learning for FDD massive MIMO systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 4, pp. 1253–1265, Dec. 2021.
- [24] X. Li, J. Guo, C.-K. Wen, S. Jin, S. Han, and X. Wang, "Multi-task learning-based CSI feedback design in multiple scenarios," 2022, *arXiv:2204.12698*.
- [25] W. Chen, W. Wan, S. Wang, P. Sun, G. Y. Li, and B. Ai, "CSI-PPPNet: A one-sided one-for-all deep learning framework for massive MIMO CSI feedback," *IEEE Trans. Wireless Commun.*, early access, 2023.
- [26] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1828–1837.
- [27] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
- [28] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4066–4079, Aug. 2018.
- [29] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 319–345.
- [30] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [31] C. Wang, "Kernel learning for visual perception, chapter 2.2," Ph.D. dissertation, School Elect. Electron. Eng., Nanyang Technological Univ., Singapore, 2019.
- [32] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*.
- [33] F. Hejazi, K. Vuckovic, and N. Rahnavard, "DyLoc: Dynamic localization for massive MIMO using predictive recurrent neural networks," in *Proc. IEEE Conf. Comput. Commun.*, May 2021, pp. 1–9.
- [34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [35] H. Zhang, Z. Lu, X. Zhang, and J. Wang, "Data augmentation of bridging the delay gap for DL-based massive MIMO CSI feedback," 2023, *arXiv:2308.00478*.
- [36] L. Liu et al., "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, Dec. 2012.
- [37] S. De Bast and S. Pollin, (2021), "Ultra dense indoor MaMIMO CSI dataset," *IEEE Dataport*, doi: [10.21227/nr6k-8r78](https://doi.org/10.21227/nr6k-8r78).
- [38] S. Jaeckel, L. Raschkowski, K. Borner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.
- [39] M. del Rosario and Z. Ding, "Learning-based MIMO channel estimation under practical pilot sparsity and feedback compression," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1161–1174, Feb. 2023.
- [40] J. Deng, W. Shi, J. Zhang, X. Zhang, and C. Zhang, "Supervised contrastive CSI representation learning for massive MIMO positioning," *IEEE Wireless Commun. Lett.*, vol. 26, no. 8, pp. 1799–1803, Aug. 2022.
- [41] Remcom. (2023). *Wireless InSite*. Accessed: Dec. 31, 2023. [Online]. Available: <http://www.remcom.com/wireless-insite>
- [42] OpenStreetMap. Accessed: Dec. 31, 2023. [Online]. Available: <https://www.openstreetmap.org>

- [43] W. Utschick, V. Rizzello, M. Joham, Z. Ma, and L. Piazz, "Learning the CSI recovery in FDD systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6495–6507, Aug. 2022.
- [44] N. Song and T. Yang, "Machine learning enhanced CSI acquisition and training strategy for FDD massive MIMO," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2021, pp. 1–6.
- [45] C. Lu, W. Xu, S. Jin, and K. Wang, "Bit-level optimized neural network for multi-antenna channel quantization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 87–90, Jan. 2020.
- [46] Z. Liu, L. Zhang, and Z. Ding, "An efficient deep learning framework for low rate massive MIMO CSI reporting," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4761–4772, Aug. 2020.
- [47] Y. Jin, J. Zhang, B. Ai, and X. Zhang, "Channel estimation for mmWave massive MIMO with convolutional blind denoising network," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 95–98, Jan. 2020.
- [48] J. Zhao, Y. Wu, Q. Zhang, and J. Liao, "Two-stage channel estimation for mmWave massive MIMO systems based on ResNet-UNet," *IEEE Syst. J.*, vol. 17, no. 3, pp. 4291–4300, Sep. 2023.
- [49] J. Guo, J. Wang, C. Wen, S. Jin, and G. Y. Li, "Compression and acceleration of neural networks for communications," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 110–117, Aug. 2020.
- [50] Z. Cao, W. Shih, J. Guo, C. Wen, and S. Jin, "Lightweight convolutional neural networks for CSI feedback in massive MIMO," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2624–2628, Aug. 2021.
- [51] Z. Lu, X. Zhang, R. Zeng, and J. Wang, "Better lightweight network for free: Codeword mimic learning for massive MIMO CSI feedback," 2022, *arXiv:2210.16544*.
- [52] *Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface (Release 18)*, document TR 38.843, Version 1.0.0, 3GPP, Sep. 2023. [Online]. Available: <http://www.3gpp.org>
- [53] NanoReview. *Smartphone Processors Ranking*. Accessed: Dec. 28, 2023. [Online]. Available: <https://nanoreview.net/en/soc-list/rating>



**Zhenyu Liu** (Member, IEEE) received the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT) in 2020. From 2017 to 2019, he was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of California, Davis. From 2020 to 2023, he was a Post-Doctoral Fellow with the School of Computer Science (National Pilot Software Engineering School), BUPT. He is currently a 6G Research Fellow with the 5GIC and 6GIC, Institute for Communication Systems, University of Surrey. His research interests include deep learning for the physical layer optimization in 6G and wireless sensing.



**Li Wang** (Senior Member, IEEE) received the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2009. She held a visiting positions with the School of Electrical and Computer Engineering, Georgia Tech, Atlanta, GA, USA, from December 2013 to January 2015, and the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, from August to November 2015 and from July to August 2018. She is currently a Full Professor with the School of Computer Science (National Pilot Software Engineering School), BUPT, where she is also the Associate Dean and the Head of the High Performance Computing and Networking Laboratory. She is a member of the Key Laboratory of Universal Wireless Communications, Ministry of Education, China. She has authored/coauthored almost 70 journal articles and four books. Her current research interests include wireless communications, distributed networking and storage, vehicular communications, social networks, and edge AI.

She was a recipient of the 2013 Beijing Young Elite Faculty for Higher Education Award, and the Best Paper Awards from several IEEE conferences, such as IEEE ICCC 2017, IEEE GLOBECOM 2018, and IEEE WCSP 2019. She was also a recipient of Beijing Technology Rising Star Award in 2018. She was an Associate Editor of IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, the Symposium Chair of the IEEE ICC 2019 on Cognitive Radio and Networks Symposium, and the Tutorial Chair of the IEEE VTC 2019-fall. She also served as the Vice Chair of the Meetings and Conference Committee for the IEEE Communication Society (ComSoc) Asia Pacific Board for the term of 2020–2021 and chairs the special interest group (SIG) on Sensing, Communications, Caching, and Computing (C3) in Cognitive Networks for IEEE Technical Committee on Cognitive Networks. She currently serves on the editorial boards for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, *Computer Networks*, and *China Communications*. She has served on the TPC of multiple IEEE conferences, including IEEE Infocom, Globecom, International Conference on Communications, IEEE Wireless Communications and Networking Conference, and IEEE Vehicular Technology Conference in recent years.



**Lianming Xu** (Member, IEEE) received the B.Eng. degree from Hefei University of Technology, Hefei, China, in 2003, and the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2009. He is currently an Assistant Professor with the School of Electronic Engineering, BUPT. His research interests include edge intelligence, the Internet of Things, caching, and collaborative computing.



**Zhi Ding** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Cornell University in 1990.

From 1990 to 2000, he was a Faculty Member with Auburn University and later with The University of Iowa. He joined the College of Engineering, University of California at Davis (UC Davis), Davis, CA, USA, in 2000. He is currently with the Department of Electrical and Computer Engineering, UC Davis, where he holds the position of a Distinguished Professor. His research team of enthusiastic researchers

works very closely with industry to solve practical problems and contributes to technological advances. His team has collaborated with researchers around the world and welcomes self-motivated young talents as new members. He is the coauthor of the textbook *Modern Digital and Analog Communication Systems* (Oxford University Press, 5th edition, 2019). He supervised more than 30 Ph.D. dissertations since joining UC Davis. His major research interests include the areas of wireless networking, communications, signal processing, multimedia, and learning. He was a member of the Technical Committee on Statistical Signal and Array Processing and a member of the Technical Committee on Signal Processing for Communications from 1994 to 2003. He served as a Steering Committee Member for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2009 and the Chair from 2009 to 2010. He received the IEEE Communication Society's WTC Award in 2012 and the IEEE Communication Society's Education Award in 2020. He was the General Chair of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing, and the Technical Program Chair of the IEEE GLOBECOM 2006. He was the Chair of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2009 to 2010. He has served as the Chief Information Officer and the Chief Marketing Officer for the IEEE Communications Society. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1997 and from 2001 to 2004. He was an Associate Editor of IEEE SIGNAL PROCESSING LETTERS from 2002 to 2005. He was also a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2004 to 2006 and the IEEE Communications Society from 2008 to 2009.