AI Enlightens Wireless Communication: A Transformer Backbone for CSI Feedback

Xiao Han¹, Wang Zhiqin²*, Li Dexin¹, Tian Wenqiang¹, Liu Xiaofeng², Liu Wendong¹, Jin Shi¹, Shen Jia¹, Zhang Zhi¹, Yang Ning¹

¹ Dept. of Standards Research, OPPO, Beijing 100026, China
² China Academy of Information and Communications Technology, Beijing 100191, China
³ National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, China
* The corresponding author, email: zhiqin.wang@caict.ac.cn

Cite as: H. Xiao, Z. Wang, et al., “AI enlightens wireless communication: A transformer backbone for CSI feedback,” China Communications, 2024, vol. 21, no. 12, pp. 243-256. DOI: 10.23919/JCC.ea.2022-0186.202401

Abstract: This paper is based on the background of the 2nd Wireless Communication Artificial Intelligence (AI) Competition (WAIC) which is hosted by IMT-2020(5G) Promotion Group 5G+AI Work Group, where the framework of the eigenvector-based channel state information (CSI) feedback problem is firstly provided. Then a basic Transformer backbone for CSI feedback referred to EVCsiNet-T is proposed. Moreover, a series of potential enhancements for deep learning based (DL-based) CSI feedback including i) data augmentation, ii) loss function design, iii) training strategy, and iv) model ensemble are introduced. The experimental results involving the comparison between EVCsiNet-T and traditional codebook methods over different channels are further provided, which show the advanced performance and a promising prospect of Transformer on DL-based CSI feedback problem.

Keywords: CSI feedback; deep learning; MIMO; Transformer

I. INTRODUCTION

The physical layer is the basis for ensuring the quality of communication services for a wireless communication system. As a critical technology in physical layer for the fifth generation (5G), the massive multiple-input multiple-output (MIMO) will also be integral in the sixth generation (6G) to meet the growing needs of the mobile data traffic. Specifically, the channel state information (CSI) feedback is indispensable in massive MIMO to ensure that the base station (BS) can accurately obtain the channel quality and perform further resource allocation, data transmission, etc.

To fully explore the potential of massive MIMO, accurate CSI feedback can be conducted based on the codebook based mechanism including TypeI and enhanced TypeII (eTypeII), which have been standardized in release 16 (R16) by 3rd Generation Partnership Project (3GPP) [1–3] and widely used for CSI feedback in current 5G new radio (NR) system. However, since the performance of such a scheme depends on the codebook, beam vector quantization error by codebook is inevitable, and the complexity of codebook design and the corresponding feedback overhead will increase significantly with the growing number of antennas and subcarriers, which brings considerable challenges to codebook based methods.

The core contributions are summarized as follows.

• With 2nd Wireless Communication Artificial Intelligence (AI) Competition (WAIC) as the background, a framework of eigenvector-based CSI feedback is introduced, where the system-level...
channel model for dataset and eigenvector-based system model provide a strong and effective guidance for future system design, performance evaluation and protocol specification in task of deep learning (DL) based CSI feedback.

• Based on the provided framework, a Transformer backbone for CSI feedback referred to EVCsiNet-T is proposed, which processes all eigenvectors of the CSI as a sequence using self-attention mechanism and can potentially be a plausible and convincing baseline for research of CSI feedback in academia and recent study item of 3GPP.

• Beyond the backbone of EVCsiNet-T, a series of potential enhancements including data augmentation, loss function design, training strategy and model ensemble which can be exploited to further improve the performance of the backbone are introduced.

• To demonstrate the advanced performance and a promising prospect of Transformer on DL-based CSI feedback, we conduct various numerical experiments involving the comparison between the EVCsiNet-T and traditional codebook methods over different channels.

The structure of this paper is as follows, the background including the relevant works and the 2nd WAIC with the task of DL-based eigenvector-basedCSI feedback are introduced in Section II. The framework system model and channel model involved in the 2nd WAIC are depicted in Section III. A basic Transformer backbone for CSI feedback named EVCsiNet-T is proposed in Section IV. Moreover, a series of potential enhancements for DL-based CSI feedback is introduced in Section V. Finally, experimental results and a brief conclusion are provided in Section VI and VII, respectively.

II. BACKGROUND

2.1 Relevant Works

Since the DL has gained great successes in the fields of computer vision (CV) and natural language processing (NLP), the combination of wireless communication and DL has also attracted great attention in recent years [4–6]. The study item (SI) named “study on artificial intelligence (AI)/machine learning (ML) for NR air interface” has been established in 3GPP release 18 (R18) in which DL-based CSI feedback is regarded as an important use case [7–10]. In academia, DL-based CSI feedback has been widely studied since it can simultaneously achieve higher compression and recovery accuracy and lower feedback overhead [11–25]. CsiNet [12] is firstly proposed whose encoder and decoder are constructed by convolutional neural network (CNN) to conduct CSI compression and recovery at the user equipment (UE) and BS, respectively. Furthermore, a series of follow-up studies designed based on various kinds of CNNs [13–19] and long short-term memory (LSTM) [20] have been proposed to further improve the CSI feedback performance. It can be noticed that most studies originated from CsiNet mainly focus on the full channel state information (F-CSI) feedback, where the whole downlink channel matrix is compressed and recovered at encoder and decoder, respectively. However, the solution for CSI feedback discussed in 3GPP is based on the compression and feedback of the eigenvector of the channel matrix, which is the main application in current 5G system. Therefore, besides the DL-based F-CSI feedback, it is necessary to study the eigenvector-based CSI feedback using DL methods, which can achieve direct fair performance comparison with the existing codebook based solutions. Different from the DL-based F-CSI feedback, EVCsiNet [6] is introduced which concentrates on eigenvector-based CSI feedback. However, the current research on eigenvector-based CSI feedback is still relatively preliminary calling for more exploration, which thus is the main focus of this paper.

2.2 Wireless Communication AI Competition

In order to further explore the practical application of AI in wireless communication systems, IMT-2020(5G) Promotion Group 5G+AI Work Group held the 2nd WAIC in July 2021 with a topic of AI Enlightens Wireless Communication which is still committed to promoting the deep integration and mutual promotion of the wireless communication and AI. The 2nd WAIC focuses on the task of Performance Improvement of DL-based CSI feedback and DL-based channel estimation. As one of the tasks, DL-based CSI feedback aims to evaluate the performance of DL-based CSI recovery and feedback overhead reduction with classic data-set on 3GPP system-level channel model. Since the high degree of integration between

© China Communications Magazine Co., Ltd. · December 2024

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
### Table 1. Dataset parameter settings for 2nd WAIC.

| Parameter               | Value               |
|-------------------------|---------------------|
| Channel model           | UMa & NLoS          |
| Carrier frequency $F_c$ | 3.5GHz              |
| Subcarrier spacing $B_{sc}$ | 15KHz             |
| Number of resource blocks $N_{RB}$ | 48                  |
| Number of subbands $N_{sb}$ | 12            |
| Number of Tx antennas $N_t$ | 32               |
| Number of Rx antennas $N_r$ | 4                 |
| UE speed                | 3km/h               |
| Number of cells $N_c$   | 57                  |
| Number of UEs in training set $N_{train}$ | 3000       |
| Number of UEs in testing set $N_{test}$ | 400              |
| Number of sampling slots $N_{slot}$ | 200          |
| Number of interval slots $T$ | 50                |

the task and the industry, there are more than 300 teams involving the contestants from related companies, universities and research institutes participate in the competition. Focusing on the schedule of the competition, the two-month online results submission for participants and the one-day offline seminar for interested researchers provide a good opportunity for technical exploring and sharing.

As for the task of DL-based CSI feedback in 2nd WAIC, the participants achieve excellent results and there are six over top ten teams utilize the Transformer [26] architecture as the backbone for model design. The Transformer architecture was firstly proposed to solve problems in the field of natural language processing (NLP) and computer vision (CV) [27], and the successful utilization in 2nd WAIC shows great potential of Transformer in the DL-based CSI feedback task.

In the following, based on the background of 2nd WAIC, the details of i) dataset construction method, ii) the Transformer solution for CSI feedback and iii) the noteworthy enhancing schemes for CSI feedback are introduced. The corresponding experimental results are also provided. With the success of the competition, it becomes an unprecedented way to promote the integration of 5G and AI by involving academic and industrial research and development to solve typical wireless problems with AI tools.

### III. FRAMEWORK OF EIGENVECTOR-BASED CSI FEEDBACK

#### 3.1 Channel Model

3GPP technical report (TR) 38.901 [28] entitled “Study on channel model for frequencies from 0.5 to 100 GHz” helps to properly model and evaluate the performance of physical layer techniques using the appropriate channel model. Since the channel models in TR 38.901 are recognized and typical, the study of CSI feedback based on these channel models is appropriate and persuasive. As a benchmark channel for simulation and performance evaluation in 3GPP, the system level three-dimensional channel model is adopted in this paper to reflect the channel characteristics in realistic environments. Specifically, four types of channel models including urban-macro (UMa), urban-micro indoor-hotspot, and rural-macro are defined in TR 38.901, and we selectively consider the UMa model with pure non-line-of-sight (NLoS). The system-level time domain downlink channel matrix $H_t$ can be written as

$$H_t = \sum_{d=1}^{N_d} \tilde{H}_{d} = \sum_{d=1}^{N_d} \sum_{l=1}^{L_d} \tilde{H}_{d,l},$$

where $\tilde{H}_{d,l}$ and $L_d$ denote the channel of $l$-th sub-path and the total number of sub-paths in the $d$-th cluster, respectively. Note that the channel defined in TR 38.901 is considered, where each sub-path $\tilde{H}_{d,l}$ is generated with different angle-of-departures, angle-of-arrivals, angle-spreads in both azimuth and zenith domains, power and delay distributions, and initial phases, etc, therefore the small-scale fading of frequency selective also occurs. Moreover, due to the frequency selection characteristics of the channel, CSI of different subcarriers are different, which introduces the subband-based CSI feedback. Moreover, $N_{slot} \times N_{UE}$ channel samples are provided in related dataset, where $N_{slot}$ slots are uniformly sampled with $T$ continuous interval slots for each UE, $N_{UE} = N_{train} + N_{test}$ denotes the number of randomly distributed UEs in $N_c$ cells, $N_{train}$ and $N_{test}$ represent the numbers of UEs in training and testing set, respectively. The corresponding channel and dataset parameter settings in 2nd WAIC are shown in Table 1. Note that in order to balance the difficulty of the competition for the contestants and make the score distribution of the leaderboard reasonable, fre-
frequency selective fading of the channel is further reduced. The dataset (Data-WAIC2nd.zip) and an example code (reference model_WAIC2nd_EVCsiNetT.zip) involved in 2nd WAIC are released in the page of ‘News’ of Wireless-Intelligence (https://wireless-intelligence.com) which is a channel database website provided for AI-based wireless communication research. Since the dataset are constructed and modified based on 3GPP specification, the results on this channel dataset have strong guiding significance for further research on the future communication system design and protocol specification.

3.2 System Model

We consider a typical MIMO system with $N_t$ transmitting antennas at BS and $N_r$ receiving antennas at UE. Due to the frequency selection characteristics of the channel, CSI of different subcarriers are different, which introduces the subband-based CSI feedback by dividing the full bandwidth into multiple subbands for feedback. By conducting fast Fourier transform (FFT) on the time domain downlink channel matrix $H_t$ described in Section 3.1, the downlink channel in the frequency domain can be given as

$$H_f = [H_1, H_2, \cdots, H_{N_{sb}}],$$  \hspace{1cm} (2)

where $N_{sb}$ represents the number of subbands consisting of 4 resource blocks (RBs) as the basic feedback granularity, and $H_k \in \mathbb{C}^{N_t \times N_i}, 1 \leq k \leq N_{sb}$ indicates the downlink channel of the $k$th subband. According to the application of the current 5G system, the solution for CSI feedback is based on the compression and feedback for the eigenvector of the channel matrix. However, most existing academia researches [11–25] for DL-based CSI feedback mainly concentrate on full CSI compression and recovery, which is incomparable with the current Type I and eType II codebook based mechanisms with eigenvector CSI operation defined by 3GPP. Moreover, since the study item named ‘study on artificial intelligence (AI)/machine learning (ML) for NR air interface’ has been established in 3GPP R18 in which DL-based CSI feedback is regarded as an important use case [7–10], the study for advanced structure of neural network (NN) and comparison between DL-based methods and codebook based methods are significant and indispensable. These motivate the study of this paper, and the DL-based CSI feedback for eigenvector considered in this paper can also achieve fair comparison with the traditional codebook based mechanism including Type I and eType II in the 5G system. Assuming ideal channel estimation at UE and implementing the single layer downlink transmission, the corresponding eigenvector for the $k$th subband $w_k \in \mathbb{C}^{N_t \times 1}$ with normalization $||w_k||^2 = 1$, can be utilized as the downlink precoding vector and calculated using eigenvector decomposition, i.e.,

$$H_k^H H_k w_k = \lambda_k w_k, \hspace{1cm} (3)$$

where $\lambda_k$ represents the maximum eigenvalue of $H_k^H H_k$. As all $N_{sb}$ eigenvectors should be fed back to the BS for downlink precoding, total $N_{sb} N_t$ complex coefficients for $N_{sb}$ subbands should be compressed and recovered using NN. Thus a CSI sample $i$ across total $N_{sb}$ subbands can be written as

$$W_i = [w_{1,i}, w_{2,i}, \cdots, w_{N_{sb},i}] \in \mathbb{C}^{N_t \times N_{sb}}. \hspace{1cm} (4)$$

The gray scale of $W$ is also described in Figure 1, where x-axis indicates the subband index and y-axis indicates the corresponding Tx antenna index, and deeper color means the smaller energy of the corresponding channel element. For CSI compressing and feedback, a DL-based CSI feedback scheme is implemented as shown in Figure 2, where a CSI sample $W$ is compressed to a bitstream $s$ of length $M$ using a DL encoder at UE. Then the CSI $W$ is recovered from

![Figure 1. Illustration of gray scale of energy for one sample.](image-url)
the feedback bitstream $s$ exploiting the DL decoder at the BS. Moreover, denoting the test set $\mathcal{W}$ consisting of $|\mathcal{W}|$ CSI samples where $| \cdot |$ denotes the cardinality of a set, the average squared generalized cosine similarity (SGCS) is usually utilized to evaluate the CSI compression and recovery accuracy as follows

$$
\rho(W, W') = \frac{1}{|\mathcal{W}|} \sum_{i=1}^{N_b} \sum_{k=1}^{N_r} \frac{\|w^H_k,i w^\prime_{k,i}\|_2^2}{\|w_k,i\|_2 \|w^\prime_{k,i}\|_2^2},
$$

where $\rho(W, W') \in [0, 1]$, $\| \cdot \|_2$ denotes the $\ell_2$ norm, $W'$ represents the predicted eigenvector CSI set, $w_{k,i}$ and $w'_{k,i}$ represent the $k$-th eigenvectors of $i$-th sample in $W$ and $W'$, respectively. A higher SGCS performance indicates higher CSI compression and recovery accuracy. Based on the basic requirements of wireless communication, the 2nd WAIC focuses on the SGCS performance under different feedback overhead in a complex channel environment. The model design with $M = 48$ and $M = 128$ bits are required which correspond to low feedback overhead scenario and high feedback overhead scenario, respectively. The final score is obtained by averaging the SGCS performance of the two scenarios.

**IV. A TRANSFORMER BACKBONE FOR CSI FEEDBACK**

Recently, Transformer [26] based solution shows great potential in CSI feedback, whose natural encoder-decoder structure is quite suitable for respectively implementing at the UE and BS. By implementing Transformer for CSI feedback, a specific domain of the channel such as paths, antenna pairs, or subbands can be treated as the sequence inputs and processed by self-attention block, which is similar to the sequential processing issue in NLP and benefits the extraction of unique characteristics of the wireless channel. Here, a Transformer backbone referred to EVCsiNet-T for solving CSI feedback is introduced, where all subbands are regarded as a sequence. By splitting the real and imaginary parts of a complex CSI sample, the input of the EVCsiNet-T can be written as

$$
\hat{W} = [\text{Re}(W)^T, \text{Im}(W)^T]^T \in \mathbb{R}^{2N_t \times N_{sb}},
$$

where the subscript $i$ indicating the index of a sample is omitted to simplify the representation, and $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the real and imaginary parts of a matrix, respectively. To implement the CSI feedback, the EVCsiNet-T solves the following problem during the training phase, i.e.,

$$
\min_{\Theta_E, \Theta_D} L(\hat{W}, \hat{W}'),
$$

where $\Theta_E$ and $\Theta_D$ denote the weights of encoder and decoder of EVCsiNet-T, respectively, and $L(\cdot)$ denotes the loss function such as mean square error (MSE), minus cosine similarity, etc.
The architecture of EVCsiNet-T is depicted in Figure 3 in detail which consists of encoder and decoder. Each vector of subband in the input CSI matrix is firstly processed by an embedding layer with embedding dimension of $N_e$ and positionally encoded by adding a learnable position vector. Then $N_b$ basic blocks are sequentially introduced to conduct the feature extraction. Moreover, a flatten layer and a dense layer with $M/B$ units are implemented in order to adapt the length of feedback bits, where $B$ is the number of quantization bits and $M$ denotes the total length of feedback bits. Finally, a uniform quantization layer is employed to transform the floating number vector with length of $M/B$ to the bitstream.

As for the decoder, the feedback bitstream is converted to a vector of float numbers using the dequantization layer. Next, a dense layer with $N_eN_{sb}$ units using gaussian error linear units (GELU) activation function is employed, whose output is reshaped to a $N_e \times N_{sb}$ matrix. After that, $N_b$ basic blocks are sequentially deployed to extract the features, whose output is flattened and processed by a dense layer with $2N_bN_{sb}$ units and activation function of GELU. Finally, the reshape layer is implemented to obtain the output with the shape of the original CSI.

The basic block can be the core component in both encoder and decoder, in which the input is firstly processed by a multi-head attention layer [26] with the number of heads $N_{head}$ and then added to the shortcut from the input of the block. A layer normalization and feedforward network are sequentially conducted where the feedforward network is constructed by two dense layers with the number of units $\{k_hN_e, N_e\}$ sandwiching a GELU, where $k_h$ denotes the scaling factor of hidden layer. The shortcut from the input of feed forward network is also added to the output and a layer normalization is implemented finally.

V. POTENTIAL ENHANCEMENTS

Different from the image data and sequence data respectively in CV and NLP, the channel data obtained from wireless communication system has some unique features, which call for more suitable tactics for performance enhancing. Specifically, based on the backbone introduced in Section IV, there are also a series of enhancing schemes for DL-based CSI feedback during the 2nd WAIC, which can further improve the performance of the model and are introduced in more detail in this section.

5.1 Data Augmentation

The data augmentation approach is widely utilized to improve the diversity of the training data and alleviate the overfitting problem. There are a number of data augmentation approaches exploited in the competition including noise injection, flipping, shift, and rotation, which are introduced in detail in this subsection.

5.1.1 Noise Injection

Noise injection refers to adding noise to the training data which is shown in Figure 4 (a). The noise is typically drawn from the various statistical distributions. In general, the noise is injected with a zero-mean Gaussian distribution, under which condition the injection process can be mathematically written as

$$ W_{aug} = W + \alpha \Upsilon, \quad \eta \sim \mathcal{N}(0, \sigma^2), $$

where $W_{aug}$ denotes the augmented CSI tensor, $W$ denotes the noise-free CSI tensor, $\Upsilon$ represents the noise tensor with the same shape as the CSI tensor $W$ and with elements sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$, and $\alpha$ is the coefficient that scales the magnitude of the injected noise, respectively.

5.1.2 Flipping

The flipping method for CSI considers flip the CSI tensor according to the dimension of subband or antenna pairs. As shown in Figure 4 (b), for a CSI tensor

$$ W = \left[ w_1, w_2, \ldots, w_{N_{sb}} \right] \in \mathbb{C}^{N_t \times N_{sb}} $$

as an example, where $w_k \in \mathbb{C}^{N_t \times 1}$ denotes the corresponding eigenvector for the $k$th subband, the flipping process according to the dimension of subband can be written as

$$ W_{aug} = \left[ w_{N_{sb}}, w_{N_{sb}-1}, \ldots, w_1 \right] \in \mathbb{C}^{N_t \times N_{sb}}, $$

where $W_{aug}$ denotes the augmented CSI tensor and flipping in antenna pairs dimensions can be easily generalized.
5.1.3 Shifting

Shifting method considers to shift along a certain dimension of the tensor according to the step size, which can change the position of the content of the tensor. Figure 4 (c) shows the cyclic shifting of a CSI tensor $W$ on the dimension of subband. It shifts the last subband to the first position for $p$ times, which can be written as

$$W^{aug} = \left[ w_{N_{sb} - p + 1}, \ldots, w_{N_{sb}}, w_1, \ldots, w_{N_{sb} - p} \right]_s,$$

where $w_k \in \mathbb{C}^{N_t \times 1}$ denotes the corresponding eigenvector for the $k$th subband and $N_{sb}$ denotes the number of subbands. Different from the cyclic shifting, there can also be another method refered to random shift achieves data augmentation by randomly shuffling the subbands, which is shown in Figure 4 (d). In more detail, cyclic shift can generate at most $N_{sb} - 1$ samples based on one raw data sample while random shift does at most $(N_{sb}! - 1)$ samples. Therefore random shift can provide more diversity of data augmentation than cyclic shift. However, since the relationship of adjacent subbands does not change when cyclic shifting, thus certain frequency selective fading features are still retained in the generated data.

5.1.4 Rotation

Rotation augmentation shown in Figure 4 (e) utilizes sine and cosine functions to randomly rotate the eigenvector by a certain angle, which can be written as

$$\text{Re}\{w_{aug}\} = \cos(\theta)\text{Re}\{w\} - \sin(\theta)\text{Im}\{w\},$$
$$\text{Im}\{w_{aug}\} = \sin(\theta)\text{Re}\{w\} + \cos(\theta)\text{Im}\{w\},$$

(12)

where $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ are the real and imaginary part of the input, $w$ is one subband eigenvector of the CSI tensor $W$, $\theta$ is the rotation degree parameter which can be different for different eigenvectors, respectively.

5.2 Loss Function Design

Loss function design is a significant factor affecting the training. The minus cosine similarity and MSE which are commonly used always fail to strictly satisfy the final goal or unique features of the task, calling for novel method for designing the loss function. Here, we will give an introduction on enhanced loss function design for DL-based CSI feedback.

5.2.1 Scoring Loss

Since the final goal of the task is the pursuit of a high score according to (5), a design method can be directly adapt the scoring indicator function as the loss function, i.e.,

$$L = -\rho,$$

(13)

where $\rho$ is the scoring function in (5) to evaluate the accuracy of the CSI compression and recovery.
5.2.2 Quantization Error Compensation Loss

The quantization error can be up to half the quantization interval and are always unavoidable. Therefore the quantization error compensation loss $L_{\text{quant}}$ can be designed to reduce the error between the vector before quantization $v$ and the vector after dequantization $v'$ caused by quantization procedure, i.e.,

$$L_{\text{quant}} = L_{\text{base}}(v, v'), \quad (14)$$

where $L_{\text{base}}$ is differentiable and can be implemented using MSE, normalized MSE (NMSE) and so on, and $v$ and $v'$ are the input and output vectors of the quantization layer, respectively. Note that the quantization layer consisting of quantization and dequantization processes is generally non-differentiable in mathematics, but can still be implemented during training where same gradient is applied before and after the quantization layer. The quantization error compensation loss can be added to the original loss function forming the global loss to enhance the recovery performance, i.e.,

$$L = L_{\text{original}} + L_{\text{quant}}, \quad (15)$$

where $L_{\text{original}} = L_{\text{base}}(\hat{W}, \hat{W}')$ reduces the error between the original CSI $\hat{W}$ and the recovered CSI $\hat{W}'$, $L_{\text{quant}}$ reduces the quantization error to further improve the CSI recovery accuracy, and $L_{\text{base}}$ can also be defined as MSE, NMSE, etc.

5.3 Training Strategy

5.3.1 Learning Rate Warm-up and Decay

Since the weights of the model are randomly initialized, a large learning rate may bring instability and oscillation of the model at the beginning of training. The warmup method can make the learning rate small during the first several training epochs, where the model can gradually become stable. After that, the training can continue under a learning rate set according to certain strategies leading to a better convergence of the model. One of the prevalent learning rate warmup methods is the gradual warmup method, using which the learning rate is updated by

$$\alpha_t = \frac{t}{T_{\text{warmup}}} \alpha_{\text{max}}, \quad (16)$$

where $t \in [1, T_{\text{warmup}}]$ denotes the $t$th epoch, $T_{\text{warmup}}$ denotes the number of warm-up epochs, and $\alpha_{\text{max}}$ denotes the final warmed up learning rate, respectively.

After warm-up, learning rate decay can be further utilized which starts with the warmed up learning rate $\alpha_{\text{max}}$ and then decaying it $T_{\text{decay}}$ epochs in subsequent training. This can be beneficial for the optimization and generalization of the model [29]. As an example, the cosine decay strategy can be written as

$$\alpha_t = \alpha_{\text{min}} + \frac{1}{2} (\alpha_{\text{max}} - \alpha_{\text{min}}) \cdot \left( 1 + \cos \left( \frac{(t - T_{\text{warmup}})\pi}{T_{\text{decay}}} \right) \right), \quad (17)$$

where $t \in [T_{\text{warmup}}, T_{\text{decay}}]$ denotes the $t$th epoch and $\alpha_{\text{min}}$ represents the minimum values of the learning rate, respectively.

5.3.2 Staged Training

Another training strategy referred to staged training considers dividing the entire training procedure into several stages, where different factors, e.g., loss function, batch size, learning rate, datasets, can be adjusted in different training stages. Hence the design of the staged training is highly personalized. For example, progressive resizing is a powerful staged training method in CV for improving the training efficiency and effect, which gradually expands the size of image in consequent training stages [30]. Similar idea can...
also be applied in DL-based CSI feedback. As shown in Figure 5, one can train with a large number of feedback bits for certain epochs and then shrinks to smaller feedback bits in the next re-training stage, in which way a more effective and stable convergence can be obtained.

5.4 Model Ensemble

The model ensemble is another optional enhancing method, which can integrate the ability of multiple models and improve the accuracy performance of the ensemble model during training or testing period. For the method of model ensemble during testing period as an example which is shown in Figure 6. V encoder-decoder pairs are constructed with different conditions such as model structure, random seed and batch size for training, etc. The V encoder-decoder pairs are jointly implemented at the UE, wherein the corresponding decoders are also adapted at BS. At the UE side, the CSI $W$ is firstly processed by $V$ encoder-decoder pairs obtaining encoded bitstreams $\{s'_1, \ldots, s'_V\}$ and the decoded CSI $\{W'_1, \ldots, W'_V\}$, respectively. The decoded CSI can be utilized to select the encoder-decoder pair with the best SGCS performance, whose index is encoded by

$$b = C(j = \arg\max_{i \leq i \leq V} \rho(W, W'_i)), \quad (18)$$

where $j$ denotes the index of best encoder-decoder pair, $C(\cdot)$ denotes index encoding function and $b \in \{0, 1\}^{|\log_2 V|}$ represents the bitstream indicating the index of the best pair, respectively. The joint bitstream for index and CSI, $s = [b^T, s'_j]^T$ is fed back to the BS side, where the BS selects the decoder according to $b$ and decodes the CSI by the selected decoder, obtaining the final recovered CSI $W'$. The above model ensemble method can combine the advantages of multiple models. However, deploying multiple encoder-decoder pairs on the UE brings storage overhead at the UE side and the feedback overhead for encoding index. Moreover, considering a fixed number of total feedback bits $M = \lceil \log_2 V \rceil + M_s$, where $\lceil \log_2 V \rceil$ and $M_s$ respectively represent the length of bits overhead for encoding index and CSI tensor, the increasing number of the total models $V$ improves the diversity of ability of the ensemble model, but decreases the length of bits for encoding CSI $M_s$ resulting in lower performance of each single model. Thus the number of models $V$ is the hyperparameter that needs to be carefully tuned.

VI. EXPERIMENTS

6.1 SGCS Performance Comparison

In this section, the experimental results are provided to verify the superiority of the EVCsiNet-T compared with traditional Type1 (with mode 1) and eTypeII (with the number of orthogonal beams $L = 2$ and 4) codebook based methods [2], where the system parameters are listed in Table 1. The i) channel data in 2nd WAIC, and ii) link-level channel data of clustered delay line (CDL) A and C with delay spread of 30 and 300 ns respectively, are considered. Moreover, the embedding dimension, the number of basic blocks, quantization bits, heads in multi-head attention layer and scaling factor of hidden layer of EVCsiNet-T are set to $N_c = 512$, $N_{b=10}$, $B = 2$, $N_{\text{head}} = 16$ and...
$k_h = 2$, respectively. The cosine similarity loss function and the default adaptive momentum (Adam) optimizer with the learning rate of 0.001 are adopted to train the EVCsiNet-T for 300 epochs.

Figure 7, 8 and 9 show the SGCS performance comparison between EVCsiNet-T and traditional codebook based TypeI and eTypeII methods over different channels including CDL-A with delay spread of 30 ns (CDL-A30), CDL-C with delay spread of 300 ns (CDL-C300), and the channel in 2nd WAIC. Note that to reveal the superiority of the backbone itself the enhancing methods described in Section V are not implemented in basic EVCsiNet-T in Figure 7, 8 and 9. It can be noticed that under low feedback overhead condition the EVCsiNet-T with $M = 32$ and $M = 48$ outperform TypeI with $M = 32$ and eTypeII ($L = 2$) with $M = 49$. As for the configuration of high feedback overhead, the EVCsiNet-T with $M = 120$ also performs higher SGCS than eTypeII ($L = 4$) with $M = 128$. Moreover, taking CDL-C300 as an example, the EVCsiNet-T with $M = 32$ achieves $\rho = 0.903$ which even outperforms the $\rho = 0.868$ of eTypeII ($L = 4$) with higher feedback bits of $M = 128$. This indicates that the EVCsiNet-T can significantly reduce the feedback overhead without the cost of performance loss in comparison with traditional codebook based counterparts. Meanwhile, three different types of channels lead to different correlation features between subbands in the frequency domain. This property can be well exploited by EVCsiNet-T for CSI compression and recovery since the different subbands are treated as a sequence of inputs so that the cross-correlated features can be extracted effectively by attention mechanism, which further brings performance gain in comparison with the traditional codebook solution with relative bias in the selection of broadband beam group and subband beams.

Beyond the superior performance of EVCsiNet-T compared to the traditional codebook based methods, there are also some enhancing schemes which can empower EVCsiNet-T and further improve the performance of this backbone. For solving DL-based CSI feedback, in 2nd WAIC there are six over top ten teams use this backbone with various of enhancing schemes introduced in Section V. Note that since the effectiveness of enhancing scheme is different in the case of different datasets, backbone hyperparameters and joint implementation of enhancing schemes, the simple evaluation of single enhancing scheme is relatively unilateral. Therefore, in Table 2 we also provide the evaluation of good-performing solutions using various combinations of enhancement schemes on
Table 2. Transformer based solutions with different enhancing schemes.

| Solution ID | SGCS (48bits) | SGCS (128bits) | Scheme | Data Augmentation | Loss Function Design | Training Strategy | Model Ensemble |
|-------------|---------------|----------------|--------|-------------------|----------------------|------------------|----------------|
| 1           | 0.877         | 0.938          |        |                   |                      |                  |                |
| 2           | 0.865         | 0.935          |        |                   |                      |                  |                |
| 3           | 0.850         | 0.930          |        |                   |                      |                  |                |
| 4           | 0.849         | 0.922          |        |                   |                      |                  |                |
| 5           | 0.843         | 0.922          |        |                   |                      |                  |                |
| 6           | 0.842         | 0.921          |        |                   |                      |                  |                |

6.2 BLER Performance Comparison

As shown in Figure 10 to further reveal the effectiveness of the proposed method, the link-level block error rate (BLER) performance comparison on CDL-C300 channels for different solutions is provided with the signal-to-noise (SNR) from -1 to 6 dB over $10^4$ realizations. Moreover, without considering adaptive modulation and coding, the modulation and coding scheme (MCS) is configured as 19 [2]. The ideal channel estimation is assumed where the BS knows the perfect CSI with SGCS = 1. We can notice that the proposed EVCsiNet-T with $M = 120$ can achieve BLER $< 10^{-3}$ with SNR = 6 dB and significantly outperforms eTypeII of $L = 4$ with overhead of 128 bits. Moreover, the EVCsiNet-T method can greatly reduce the feedback overhead with similar BLER performance, since the EVCsiNet-T with $M = 48$ and 32 slightly outperform eTypeII $L = 4$ and 2 with feedback bits of 128 and 49, respectively.

6.3 Complexity Analysis

The complexity analysis of EVCsiNet-T with the number of feedback bits $M = 32, 48$ and 120 is provided in Table 3 from the perspective of floating point operations (FLOPs) and trainable parameters.

Table 3. FLOPs and number of trainable parameters.

| $M$  | Encoder FLOPs ($\times 10^7$) | Decoder FLOPs ($\times 10^7$) | Encoder Trainable Par. ($\times 10^7$) | Decoder Trainable Par. ($\times 10^7$) |
|------|------------------------------|------------------------------|---------------------------------------|---------------------------------------|
| 32   | 4.2099                       | 4.2099                       | 2.1107                                | 2.1108                                |
| 48   | 4.2111                       | 4.2111                       | 2.1113                                | 2.1114                                |
| 120  | 4.2166                       | 4.2166                       | 2.1141                                | 2.1142                                |

Obviously, the EVCsiNet-T with $M = 32$ requires $8.4198 \times 10^7$ FLOPs for inference. Since the feedback delay of the existing communication system is required to be less than 1 ms, the inference time using a common device of NVIDIA Tesla V100 SXM2 with double-precision performance of $7.8 \times 10^{12}$ floating point operations per second (FLOPS) can theoretically be 10.8 $\mu$s which meets the requirement well. It can also be noticed that the FLOPs and trainable parameters are quite close for different number of feedback bits, which indicates that the the inference time can not be affected too much by different number of feedback bits.

VII. CONCLUSION

In this paper, we first give a description of the framework of CSI feedback and its corresponding channel model involved in the 2nd WAIC. Then a Transformer backbone for CSI feedback referred to EVCsiNet-T is proposed. Moreover, a series of enhancement schemes in 2nd WAIC including i) data augmentation, ii) loss function design, iii) training strategy, and iv) model ensemble are introduced. The experimental results involving the comparison between EVCsiNet-T and
traditional codebook methods over different channels are further provided, which show the advanced performance and a promising prospect of Transformer on CSI feedback problem. Since the significance and indispensability of interpretable ML in the AI-enhanced wireless communication, we will study on this field with more theoretical interpretation and analysis in future research. Finally we sincerely hope that this article and 2nd WAIC can broaden the thinking and insights for interested researchers.

ACKNOWLEDGEMENT

We sincerely thank China Academy of Information and Communications Technology, Guangdong OPPO Mobile Telecommunications Corp., Ltd, National Mobile Communications Research Laboratory of Southeast University and vivo Mobile Communication Co., Ltd for their great help and support on the research and the holding of the 2nd WAIC. We also would like to express our gratitude to all contestants for their participation and sharing.

References

[1] 3GPP, 3GPP TS 38.212 V16.1.0, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR: Multiplexing and Channel Coding (Release 16), 2020.

[2] 3GPP, 3GPP TS 38.214 V16.1.0 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR: Physical Layer Procedures for Data (Release 16), 2020.

[3] 3GPP, 3GPP TS 38.331 V16.0.0 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Radio Resource Control (RRC) Protocol Specification (Release 16), 2020.

[4] H. Xiao, Z. Wang, et al., “Ai enlightens wireless communication: Analyses, solutions and opportunities on csi feedback,” China Communications, 2021, vol. 18, pp. 104–116.

[5] H. Xiao, W. Tian, et al., “Channelgan: Deep learning based channel modeling and generating,” IEEE Wireless Communications Letters, 2022.

[6] W. Liu, W. Tian, et al., “Evcsinet: Eigenvector-based csi feedback under 3gp link-level channels,” IEEE Wireless Communications Letters, 2021, vol. 10, no. 12, pp. 2688–2692.

[7] Qualcomm, “Rp-213599, new si: Study on artificial intelligenc (ai)/machine learning (ml) for nr air interface,” in 3GPP TSG RAN#94e, Electronic Meeting. 3GPP, 2021.

[8] OPPO, “Rp-210235, views on ai based physical layer enhancements,” in 3GPP TSG RAN Rel-18 Workshop, Electronic Meeting. 3GPP, 2021.

[9] CAICT, “Rws-210235, views on ai based physical layer enhancements,” in 3GPP TSG RAN Rel-18 Workshop, Electronic Meeting. 3GPP, 2021.

[10] CAICT and OPPO, “Rws-210236, introduction of the 1st wireless communication ai competition (waic),” in 3GPP TSG RAN Rel-18 Workshop, Electronic Meeting. 3GPP, 2021.

[11] T. Wang, C. Wen, et al., “Deep learning for wireless physical layer: Opportunities and challenges,” China Communications, 2017, vol. 14, no. 11, pp. 92–111.

[12] C. Wen, W. Shih, et al., “Deep learning for massive mimo csi feedback,” IEEE Wireless Communications Letters, 2018, vol. 7, no. 5, pp. 748–751.

[13] Y. Sun, W. Xu, et al., “Ancinet: An efficient deep learning approach for feedback compression of estimated csi in massive mimo systems,” IEEE Wireless Communications Letters, 2020, vol. 9, no. 12, pp. 2192–2196.

[14] Z. Lu, J. Wang, et al., “Multi-resolution csi feedback with deep learning in massive mimo system,” in ICC 2020-2020 IEEE International Conference on Communications (ICC), 2020, pp. 1–6.

[15] T. Chen, J. Guo, et al., “Deep learning for joint channel estimation and feedback in massive mimo systems,” arXiv preprint arXiv:2011.07242, 2020.

[16] M. B. Marshadi, Q. Yang, et al., “Distributed deep convolutional compression for massive mimo csi feedback,” IEEE Transactions on Wireless Communications, 2020.

[17] Z. Cao, W. Shih, et al., “Lightweight convolutional neural networks for csi feedback in massive mimo,” IEEE Communications Letters, 2021.

[18] J. Guo, C. Wen, et al., “Deep learning-based csi feedback for beamforming in single-and multi-cell massive mimo systems,” IEEE Journal on Selected Areas in Communications, 2020.

[19] J. Guo, C. Wen, et al., “Canet: Uplink-aided downlink channel acquisition in fdd massive mimo using deep learning,” arXiv preprint arXiv:2101.04377, 2021.

[20] C. Lu, W. Xu, et al., “Mimo channel information feedback using deep recurrent network,” IEEE Communications Letters, 2018, vol. 23, no. 1, pp. 188–191.

[21] T. Wang, C. Wen, et al., “Deep learning-based csi feedback approach for time-varying massive mimo channels,” IEEE Wireless Communications Letters, 2018, vol. 8, no. 2, pp. 416–419.

[22] J. Guo, C. Wen, et al., “Convolutional neural network-based multiple-rate compressive sensing for massive mimo csi feedback: Design, simulation, and analysis,” IEEE Transactions on Wireless Communications, 2020, vol. 19, no. 4, pp. 2827–2840.

[23] X. Li and H. Wu, “Spatio-temporal representation with deep recurrent neural network for beamforming in single-and multi-cell massive mimo systems,” in 3GPP RAN#94g, Electronic Meeting. 3GPP, 2021.

[24] T. Chen, J. Guo, et al., “A novel quantization method for deep learning-based massive mimo csi feedback,” IEEE Communications Letters, 2021.

[25] C. Lu, W. Xu, et al., “Bit-level optimized neural network for multi-antenna channel quantization,” IEEE Wireless Communications Letters, 2019, vol. 9, no. 1, pp. 87–90.
[26] A. Vaswani, N. M. Shazeer, et al., “Attention is all you need.” ArXiv, 2017, vol. abs/1706.03762.
[27] A. Dosovitskiy, L. Beyer, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” ArXiv, 2021, vol. abs/2010.11929.
[28] 3GPP, 3GPP TR 38.901 V16.1.0 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies from 0.5 to 100 GHz; (Release 16), 2020.
[29] K. You, M. Long, et al., “How does learning rate decay help modern neural networks,” arXiv: Learning, 2019.
[30] L. Liang, I. Zharkov, et al., “Guidance network with staged learning for image enhancement,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 836–845.

Biographies

Xiao Han received the B.Eng. degree and M.Eng. degree from Beijing Jiaotong University, Beijing, China, in 2017 and 2020, respectively. He is currently a researcher for standardization of communication in OPPO Research Institute, Shenzhen, China. His current research interests include the compressive sensing and deep learning for wireless physical layer.

Wang Zhiqin is the Vice President of the China Academy of Information and Communications Technology (CAICT), the Professor-level senior engineer. She is currently the Chair of the Wireless Technical Committee of China Communications Standards Association (CCSA) and the Vice Director of the Wireless and Mobile Technical Committee of China Institute of Communications (CIC). She is also serving as the Chair of the IMT-2020 (5G) Promotion Group and Chair of the IMT-2030 (6G) Promotion Group in China. She has contributed to the design, standardization, and development of 3G (TD-SCDMA), 4G (TD-LTE), and 5G mobile communication systems. She has authored or co-authored more than 60 research papers/articles, and over 20 patents in this area. Her achievements have received multiple top awards and honors by China central government, including the Grand Prize of the National Award for Scientific and Technological Progress in 2016 (the highest Prize in China), the First Prize of the National Award for Scientific and Technological Progress once, the Second Prize of the National Award for Scientific and Technological Progress twice, the National Innovation Award (once), and the Ministerial Science and Technology Awards many times. Her current research interests include the theoretical research, standardization, and industry development for 5G-Advanced and 6G.

Li Dexin received the B.S. degree from North University of China, Taiyuan, China in 2015 and M.S. degree from Northeastern University, Shenyang, China, in 2018. He is currently a researcher in OPPO research institute. His interests include AI-based physical layer wireless communication.

Tian Wenqiang received his B.S. degree from Fudan University in 2010 and his Ph.D degree from University of Chinese Academy of Sciences in 2015. Now, he is a senior Standardization Engineer of Guangdong OPPO Mobile Telecommunications Corp. Ltd. He has participated in the 5G standardization work and focused on physical layer design in 3GPP R15 and R16. Since 2018, he has started the research of 5G/6G and led the study on AI empowered PHY. He is the co-rapporteur of AI-MIMO study in IMT-2020 5G+AI Work Group, and has organized the 1st W AIC in 2020.

Liu Xiaofeng received his B.S. degree in Communication Engineering from Beijing University of Posts & Telecommunications in 2003, his M.S and Ph.D in Communication and Information Systems from Beijing University of Posts & Telecommunications, in 2006 and 2009. He is the deputy chief engineer of Mobile Communication Innovation Center of China Academy of Information and Communications Technology (CAICT), the Professor-level senior engineer. He is also serving as the Chair of IMT-2020 (5G) Promotion Group 5G+AI Work Group. He has over 20 paper published and holds over 20 patents. He is the author of 5G wireless system design and international standard and 5G wireless enhanced design and international standard. His current research focus on integration of wireless communication and AI.

Liu Wendong received the B.S. degree in 2015 and Ph.D degree with the Department of Electronic Engineering in 2020 from Tsinghua University, Beijing, China. He is currently a researcher in OPPO research institute. His interest is AI-based physical layer wireless communication.
**Jin Shi** (S’06-M’07-SM’17) received the B.S. degree in communications engineering from Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from the Southeast University, Nanjing, in 2007. From June 2007 to October 2009, he was a Research Fellow with the Adastral Park Research Campus, University College London, London, U.K. He is currently with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include space time wireless communications, random matrix theory, and information theory. He served as an Associate Editor for the *IEEE Transactions on Wireless Communications*, and *IEEE Communications Letters*, and *IET Communications*. Dr. Jin and his co-authors have been awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory and a 2010 Young Author Best Paper Award by the IEEE Signal Processing Society.

**Shen Jia** received his Bachelor degree from Department of Electronic Engineering of Tsinghua University, China in 2000, and Ph.D in Electronics degree from University of York, UK in 2005. Dr. Shen ever worked with CAICT as a director engineer. He has been a principal communications standard researcher of OPPO since 2016. Dr. Shen has joined in 3GPP 4G LTE standardization since 2005 and 5G standardization since 2016. He also contributed to large-scale 4G trials in China from 2008 to 2013, and ever served as the Leader of Testing and Specification of TD-LTE Working Group of China. Now he serves as the Vice Chair of IMT-2020 (5G) Promotion Group 5G+AI Work Group. Dr. Shen is also the rapporteur of 3GPP SA1 work item “AI/ML model transfer in 5GS” and several feature leads on 5G NR PUCCH and URLLC in 3GPP RAN1 WG. He is the inventor of over 100 patents, and the first author of books “3GPP Long-Term Evolution (LTE): Technical Principle and System Design” and “5G NR and Enhancements: From R15 to R16”.

**Zhang Zhi** received his B.S. degree from Shandong University (SDU), Jinan China in 1999, ME degree from University of Science and Technology of China (USTC), Hefei China in 2002 and Ph.D from University of Hong Kong, Hong Kong, China in 2007, all in Electrical Engineering. He has been in 3GPP RAN1 physical layer standardization & research for more than 10 years. He is currently leading physical layer standardization & research team in OPPO research institute.

**Yang Ning** is professorate senior engineer, head of Standard and Research Department in OPPO, and the research areas include 4G/5G/6G system. He participates in 3GPP and CCSA standardization work on behalf of OPPO and has made meaningful contributions to the setting of 5G standards. He has published more than ten EI/SCI indexed articles and has led or contributed to more than 200 invention patent applications. He received a doctorate degree in Electronic Circuits and Systems from Beijing University of Posts and Telecommunications in 2008.