Efficient Spatio-Temporal Predictive Learning for Massive MIMO CSI Prediction

CHENG Jiaming[1], CHEN Wei[1], LI Lun[2,3], AI Bo[1]

(1. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;
3. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202501002

https://kns.cnki.net/kcms/detail/34.1294.TN.20250303.1371.004.html, published online March 4, 2025

Manuscript received: 2025-02-23

Abstract: Accurate channel state information (CSI) is crucial for 6G wireless communication systems to accommodate the growing demands of mobile broadband services. In massive multiple-input multiple-output (MIMO) systems, traditional CSI feedback approaches face challenges such as performance degradation due to feedback delay and channel aging caused by user mobility. To address these issues, we propose a novel spatio-temporal predictive network (STPNet) that jointly integrates CSI feedback and prediction modules. STPNet employs stacked Inception modules to learn the spatial correlation and temporal evolution of CSI, which captures both the local and the global spatio-temporal features. In addition, the signal-to-noise ratio (SNR) adaptive module is designed to adapt flexibly to diverse feedback channel conditions. Simulation results demonstrate that STPNet outperforms existing channel prediction methods under various channel conditions.

Keywords: massive MIMO; deep learning; CSI prediction; CSI feedback

Citation (Format 1): CHENG J M, CHEN W, LI L, et al. Efficient spatio-temporal predictive learning for massive MIMO CSI prediction [J]. ZTE Communications, 2025, 23(1): 3 – 10. DOI: 10.12142/ZTECOM.202501002
Citation (Format 2): J. M. Cheng, W. Chen, L. Li, et al., "Efficient spatio-temporal predictive learning for massive MIMO CSI prediction," *ZTE Communications*, vol. 23, no. 1, pp. 3 – 10, Mar. 2025. doi: 10.12142/ZTECOM.202501002.

1 Introduction

Future 6G communication systems are expected to support significantly higher demands from mobile broadband services[1]. As a representative 6G scenario, ultra-massive multiple-input multiple-output (MIMO) systems critically depend on real-time, accurate, and reliable channel state information (CSI)[2]. In frequency division duplex (FDD) systems, user equipment (UE) estimates downlink CSI and feeds it back to the base station (BS) via uplink transmission. However, the increasing number of antennas has dramatically expanded the feedback overhead, thereby placing a substantial burden on limited bandwidth resources. Recently, deep learning (DL) techniques have been introduced to compress CSI and reduce feedback overhead[3 – 4]. Specifically, DL-based CSI feedback utilizes an encoder to compress the CSI into codewords at the UE and a decoder at the BS to reconstruct the CSI from these codewords[5]. This approach has been demonstrated to outperform traditional codebook-based feedback methods in terms of effectiveness[6]. In Ref. [7], SwinCFNet is proposed for a CSI feedback task, which utilizes the Swin Transformer to extract long-range dependency information from CSI.

However, due to changes in the scattering environment and user mobility, the channel varies rapidly over time. In mobile scenarios, processing delay in the CSI feedback process makes the CSI received by the BS outdated, leading to a significant degradation in system performance. The authors in Ref. [8] theoretically analyze the impact of CSI delay on the channel. To mitigate the performance degradation caused by channel aging, accurate and timely CSI prediction becomes increasingly essential, which leverages the temporal correlation between historical CSI and future channel states. Besides, in recent years, digital twins have emerged as a revolutionary technology for visualizing, predicting, and analyzing the interactions between digital models and the physical world[9]. The design of digital twins relies on the virtual mapping of physical products, using real-time data and information from the field. High-precision time series prediction of wireless channel information in physical entities is crucial to building a digital twin environment[10].

Traditional methods for CSI prediction, such as the linear extrapolation model[11] and the autoregressive (AR) model[12], rely on statistical and mathematical formulations that struggle

This work was supported in part by the Natural Science Foundation of China under Grant Nos. U2468201 and 62221001 and ZTE Industry-University-Institute Cooperation Funds under Grant No. IA20240420002.

to capture the dynamic complexity of realistic wireless channels. In contrast, DL-based models, with their capacity for capturing nonlinear relationships and their flexibility in handling large datasets, offer a promising alternative. Inspired by the great potential of the recurrent neural networks (RNNs) and their variants in time series modeling, an RNN-based predictor[13] and a long short-term memory (LSTM)-based predictor[14] have been proposed. In Ref. [15], a transformer-based parallel channel prediction model is introduced to accurately predict time-varying channels, which avoids the error propagation problem in classical sequential prediction methods. Additionally, the authors in Ref. [16] propose a joint framework for channel feedback and prediction, leveraging the convolutional LSTM (ConvLSTM) to exploit temporal correlations. However, these existing methods primarily focus on the temporal correlation, while overlooking the array and frequency correlations crucial for further improvement.

In this paper, we propose a novel spatio-temporal predictive network (STPNet) for CSI prediction in massive MIMO systems. STPNet employs a joint CSI feedback and prediction framework, where the feedback network compresses and reconstructs CSI while capturing inter-antenna and inter-subcarrier correlations. The core prediction network consists of several cascaded Inception modules to learn the spatio-temporal features from the codewords by group convolutions. Using joint training, STPNet eliminates the error propagation issues found in separate module designs. Furthermore, we introduce a signal-to-noise ratio (SNR) adaptive module to dynamically adjust input tokens according to real-time SNR variations, enabling more robust adaptation to changing communication conditions. Numerical results show that STPNet outperforms other predictive methods across diverse channel scenarios.

2 System Model

We consider the downlink of an FDD massive MIMO system with $N_t \gg 1$ transmitting antennas at the BS and a single receiving antenna at the UE. The number of sub-carriers is N_c. The received signal at the n-th subcarrier can be expressed as:

$$y_n = \boldsymbol{h}_n^H \boldsymbol{v}_n x_n + z_n \quad (1),$$

where $\boldsymbol{h}_n \in \mathbb{C}^{N_t}$, $\boldsymbol{v}_n \in \mathbb{C}^{N_t}$, $x_n \in \mathbb{C}$, and $z_n \in \mathbb{C}$ denote the channel vector, the precoding vector, the transmitted data symbol and the additive noise at the n-th subcarrier, respectively. The downlink CSI can be denoted by:

$$\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_n]^{\mathrm{T}} \in N_c \times N_t \quad (2).$$

Since the elements of the channel matrix are complex numbers, the total number of CSI parameters is $2N_cN_t$. However, as the number of antennas in future massive MIMO systems grows, the size of the CSI matrix might exceed the uplink's feedback capacity.

To tackle the challenge of payload size reduction, we implement a framework that compresses the channel matrix $\boldsymbol{H}$ into a low-dimensional codeword $\boldsymbol{s}$ of size $M\times1$ at the UE, which can be formulated as:

$$\boldsymbol{s} = f_{\mathrm{en}}(\boldsymbol{H}; \theta_{\mathrm{en}}) \quad (3),$$

where $f_{\mathrm{en}}(\cdot)$ represents the function of the encoder and θ_{en} denotes its parameter. The compression ratio (CR) is defined as $\gamma = \frac{M}{2N_cN_t}$. Then, the encoded vector $\boldsymbol{s}$ is transmitted via a noisy channel. In our work, we consider the widely used additive white Gaussian noise (AWGN) channel. The channel output vector $\hat{\boldsymbol{s}}$ received by the BS is expressed as:

$$\hat{\boldsymbol{s}} = \eta(\boldsymbol{s}, \sigma) = \boldsymbol{s} + \boldsymbol{n} \quad (4),$$

where each component of the noise vector $\boldsymbol{n}$ is independently sampled from a Gaussian distribution, i.e., $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$, and σ^2 is the noise power.

The structure of AI-based CSI feedback is illustrated in Fig. 1a. However, in high-speed mobile scenarios, the channel matrix varies rapidly over time. Due to the feedback delay and channel aging problems, directly feeding back the channel at the current slot leads to a mismatch between the feedback channel and the actual channel. To address this issue, a CSI prediction module is introduced at the BS. Our proposed AI-based joint CSI feedback and prediction framework, shown in Fig. 1b, performs prediction at the codeword level. Let $\hat{\boldsymbol{s}}^{(t)}$ and $\bar{\boldsymbol{s}}^{(t+1)}$ denote the codeword of the t-th slot and the predicted codeword of the $(t+1)$-th slot, respectively. We adopt the received historical codewords from the past P slots to simultane-

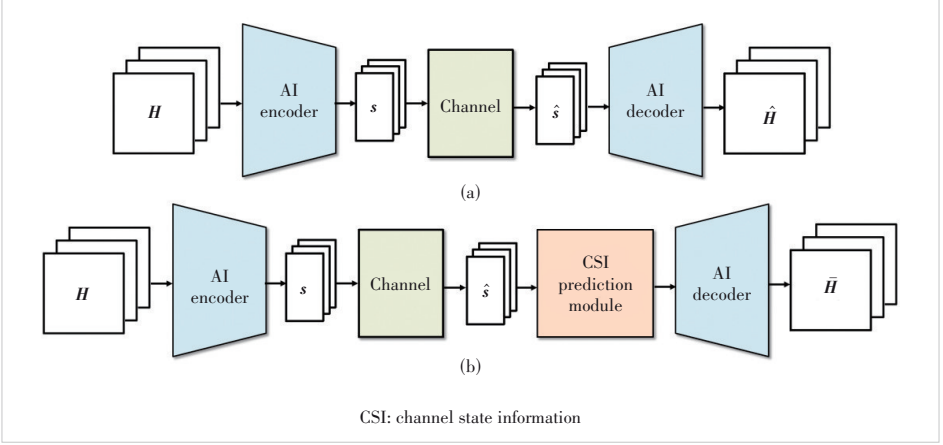


Figure 1. (a) Structure of AI-based CSI feedback; (b) Our proposed AI-based joint CSI feedback and prediction framework

ously predict the future codewords for the next L consecutive slots simultaneously, which can be expressed as:

$$(\bar{\boldsymbol{s}}^{(t+1)},\cdots,\bar{\boldsymbol{s}}^{(t+L)}) = f_{\text{pre}}(\hat{\boldsymbol{s}}^{(t-P+1)},\cdots,\hat{\boldsymbol{s}}^{(t)};\theta_{\text{pre}}) \quad (5),$$

where $f_{\text{pre}}(\cdot)$ represents the function of the prediction module and θ_{pre} denotes the corresponding parameter set. Subsequently, the BS reconstructs the channel matrix from the predicted future codewords as follows.

$$\bar{\boldsymbol{H}} = f_{\text{de}}(\bar{\boldsymbol{s}};\theta_{\text{de}}) \quad (6),$$

where $f_{\text{de}}(\cdot)$ represents the function of the decoder and θ_{de} denotes the parameter set of the decoder. $\bar{\boldsymbol{H}}$ is the recovered channel matrix.

3 Design of STPNet

3.1 Network Architecture

Compared with simple CSI feedback, joint CSI feedback and prediction can more effectively mitigate the distortion caused by feedback delays and channel aging. In a separate feedback and prediction architecture, each module is optimized and designed independently, so the local optimum of each component may not yield a globally optimal outcome. In contrast, the joint architecture employs end-to-end training to reduce error propagation between modules, resulting in more accurate predicted CSI.

Building on the advantages of the joint feedback and prediction architecture, we present an overview of our STPNet model in Fig. 2a. STPNet consists of a CSI encoder, SNR adaptive modules, a CSI prediction module and a CSI decoder. The encoder is used to compress the CSI into codewords and extract spatial features of the channel matrix at UE. The CSI prediction module, serving as the network's core, operates at the codeword level. The prediction module leverages the spatial and temporal correlation of historical channel characteristics to forecast future codewords. The SNR adaptive modules, integrated into both the encoder and decoder, dynamically modulate intermediate tokens based on instantaneous channel quality. Finally, the decoder aggregates and processes the predicted codewords to produce the final CSI output at the BS.

We employ the SwinCFNet architecture to implement the CSI encoder and decoder within STPNet. Built upon the Swin Transformer, SwinCFNet delivers superior performance in CSI feedback tasks. First, it effectively reduces feedback data while aggregating spatial-frequency domain CSI features to support the prediction module. Second, this design captures long-range dependencies, exploiting both inter-frequency and inter-antenna correlations within the channel matrix, ultimately enhancing the accuracy of the predicted output. Ref. [7] presents a detailed description of the SwinCFNet architecture.

In the core prediction module, an Inception architecture is introduced to learn the temporal evolution by capturing and updating spatio-temporal features, as shown in Fig. 2b. Motivated by Refs. [17] and [18], cascaded Inception blocks are

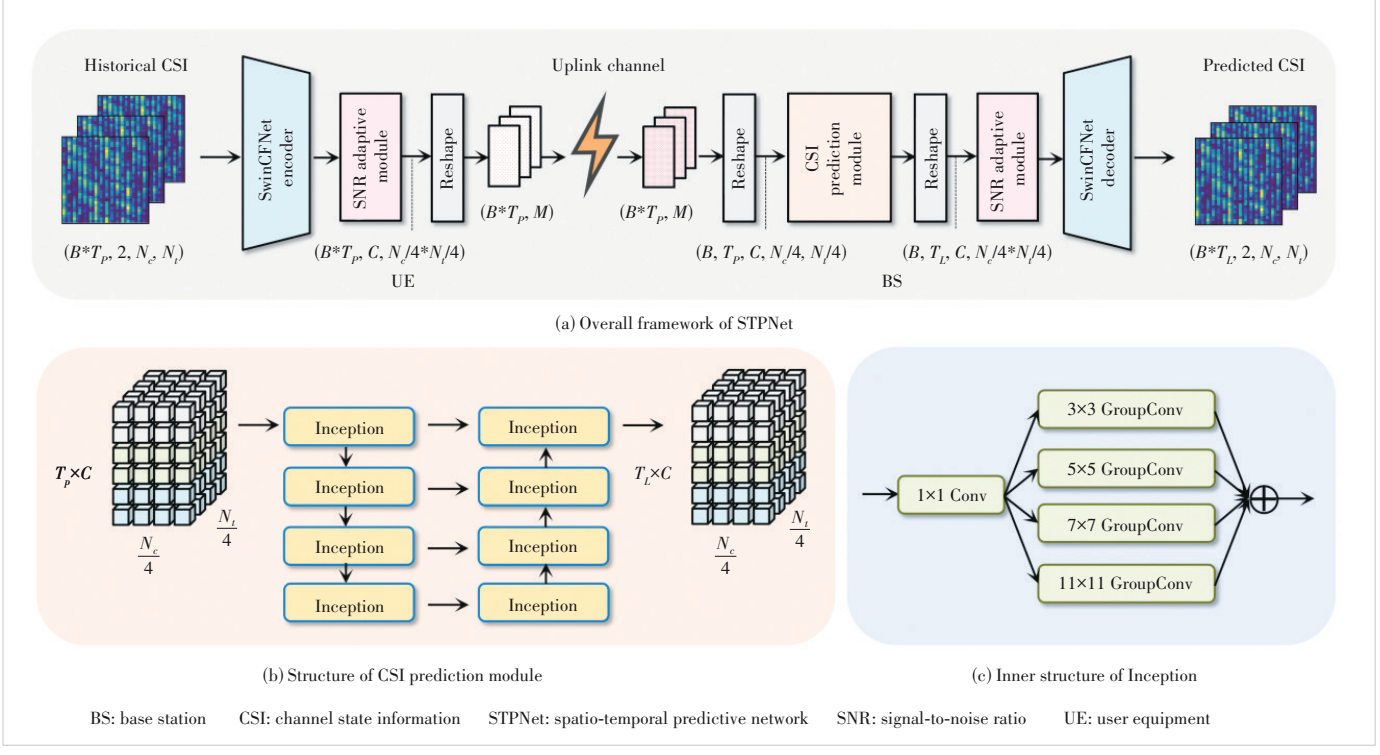


Figure 2. Network architecture of STPNet

employed. These blocks primarily consist of convolution layers with 1×1 kernels, followed by parallel GroupConv2D operations. The inner structure of Inception is illustrated in Fig. 2c. Here, the 1×1 Conv2D layer is used to increase the depth of the network and enhance representational capacity. To extract diverse local patterns, GroupConv2D layers with kernel sizes of 3×3, 5×5, 7×7, and 11×11 split the feature channels into multiple groups, each capturing different localized features. Due to the complexity of channel conditions, predicting future channels is challenging because the locations of useful features vary significantly over time. By utilizing a multi-branch Inception architecture, the cascaded modules extract both local and global features from the codewords. In the final block, outputs from convolution layers with varying kernel sizes are fused through addition, integrating multiple spatio-temporal CSI features at different scales.

Note that the joint CSI feedback and prediction model is trained in an end-to-end manner. Its parameters are updated using an adaptive moment estimation (ADAM) optimizer. The networks are trained to minimize the difference between the predicted and the ground truth CSIs. Consequently, the training loss function is defined as the mean square error (MSE) expressed as follows.

$$L(\theta_{\text{en}},\theta_{\text{pre}},\theta_{\text{de}}) = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{L}\left\| \boldsymbol{H}_i^{(t+j)} - \bar{\boldsymbol{H}}_i^{(t+j)} \right\|^2 \quad (7),$$

where T is the number of samples in the training set, and the subscript of $\boldsymbol{H}$ denotes the i-th sample in the training set. $\boldsymbol{H}_i^{(t+j)}$ and $\bar{\boldsymbol{H}}_i^{(t+j)}$ denote actual and predicted CSI at the $(t+j)$-th slot, respectively.

3.2 SNR Adaptive Module

In high-speed mobile scenarios, the uplink feedback channel undergoes rapid variations, requiring the end-to-end feedback system to adapt automatically to changing channel conditions. To address this, we introduce an SNR adaptive module (SAM), depicted in Fig. 3. The SAM is designed based on the mechanism of channel-wise soft attention[19], which identifies channel relationships and generates distinct scaling parameters for different channel states, thereby enhancing or attenuating their influence on subsequent layers[20]. By dynamically adjusting resource allocation strategies based on these varying channel states, the system implicitly modulates the source coding rates in both the encoder and decoder, ultimately achieving higher-quality transmission and CSI reconstruction.

As illustrated in Fig. 3, the SAM consists of three components: 1) SNR semantic extraction, 2) semantic embedding, and 3) feature calibration. The channel feature s is first processed by the fully connected (FC) layer and then fed into the SAM for modulation.

1) SNR semantic extraction. The uplink channel information SNR is first input into the three FC layers to generate the semantic information of the channel state. The first and second FC layers are followed by the Rectified Linear Unit (ReLU) and the last FC layer is followed by a sigmoid to restrict the output range to the interval (0, 1)[21]. It transforms SNR into an M-dimensional vector $\boldsymbol{v}_{\text{SNR}}$.

2) Semantic embedding. The input channel features and the extracted SNR semantic information $\boldsymbol{v}_{\text{SNR}}$ are fused and embedded by the element-wise product. The output will pass through the next FC layer and continue to be multiplied by $\boldsymbol{v}_{\text{SNR}}$. Following three rounds of semantic embedding, it will be restored to the same channel dimension as s via the last FC layer, and then pass through a sigmoid function to obtain the modulation scale factor.

3) Feature calibration. The resulting modulation scale factor is subsequently multiplied by the original channel characteristics to obtain the CSI feature map s'.

The SNR adaptive module integrates the SNR directly into the token processing pipeline to compute channel-wise attention, enhancing the adaptability of the network in scenarios with varying signal conditions[22]. Algorithm 1 summarizes the operation process of the proposed SAM.

Algorithm 1. Operation process of SAM

Input: The channel feature s and the uplink channel SNR
Output: The calibrated channel feature s'
1. Upgrade the channel features to M dimensions and get s_M
2. Extract the SNR semantic vector: $\boldsymbol{v}_{\text{SNR}} = \text{Sigmoid}(W_3\text{ReLU}(W_2\text{ReLU}(W_1\text{SNR} + b_1) + b_2) + b_3)$
3. Combine features and the SNR semantic vector: $\text{output}_0 = s_M \cdot \boldsymbol{v}_{\text{SNR}}$
4. **For** $i = 1 : 1 : 3$ **do**
5. Embed SNR semantic information in features: $\text{output}_i = (W_{Mi}\text{output}_{i-1} + b_{Mi}) \cdot \boldsymbol{v}_{\text{SNR}}$
6. **end for**
7. Calculate the modulation scale factor: $\mu = \text{Sigmoid}(W_C\text{output}_3 + b_C)$

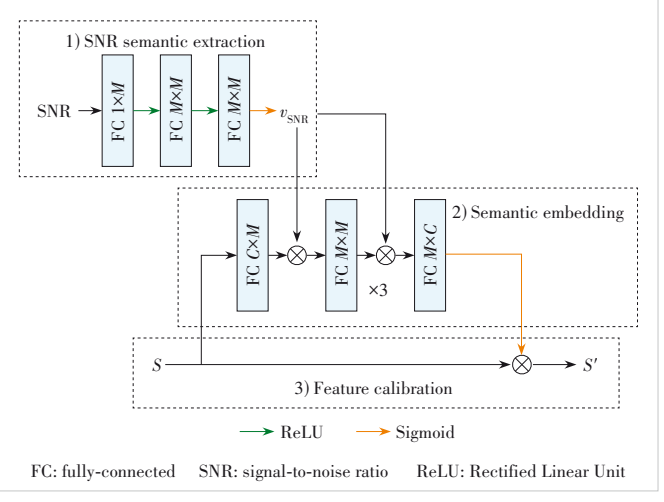


Figure 3. Architecture of SAM

8. Obtain the calibrated channel feature: $s' = s \cdot \mu$
9. **return** s'

4 Experimental Results

In this section, we present the numerical results to investigate the performance of the proposed STPNet design for joint CSI feedback and prediction.

4.1 Experiment Settings

The simulation results are based on the clustered delay line (CDL)-C channel model and the 3GPP urban macro (UMa) channel model[23], respectively. The BS employs a uniform rectangular panel array of dual-polarized antennas arranged in an 8×2 configuration. The user speed is set to 30 km/h. There are N_c = 32 subcarriers with 10 MHz bandwidth. The communication frequency f is set as 2 GHz. The lengths of historical and predicted CSIs are both set to 5. Table 1 summarizes the simulation parameters. The training and testing datasets contain 10 000 and 2 000 samples, respectively. To enhance model generalization, the prediction model is trained using uplink channels with SNR values ranging from 1 dB to 20 dB. We update the parameters with a constant learning rate of 1×10^{-3}. The batch size and the training epoch are set as 16 and 100, respectively. To evaluate model effectiveness, we quantify the accuracy of CSI prediction by using normalized mean square error (NMSE) as a quantitative metric. The NMSE is defined as:

$$\mathrm{NMSE} = \mathbb{E}\left(\frac{\left\| \boldsymbol{H} - \bar{\boldsymbol{H}} \right\|^2}{\left\| \boldsymbol{H} \right\|^2}\right) \tag{8},$$

where $\boldsymbol{H} \in \mathbb{C}^{L \times N_c \times N_t}$ denotes the ideal channel for the next L slots, and $\bar{\boldsymbol{H}} \in \mathbb{C}^{L \times N_c \times N_t}$ denotes the predicted channel.

Fig. 4a shows a sample from a single BS antenna selected for simulation from the CDL-C scenario CSI dataset. The duration of this particular sample is 10 ms. The time-varying nature of the wireless channel is captured by its autocorrelation function (ACF), as illustrated in Fig. 4b. This second-order statistic is typically influenced by factors such as the propagation geometry, the mobile's velocity, and the characteristics of the antennas[24–25]. In this paper, the DL-based approach is adopted to learn and capture the spatio-temporal correlation of CSI.

Table 1. Simulation parameters

Parameter	Value
Channel type	3GPP CDL-C and UMa[23]
Carrier frequency	2 GHz
Bandwidth	10 MHz
N_t	32
N_r	1
Number of subcarriers	32
Feedback interval	1 ms
UE speed	30 km/h

CDL: clustered delay line UE: user equipment UMa: urban macro

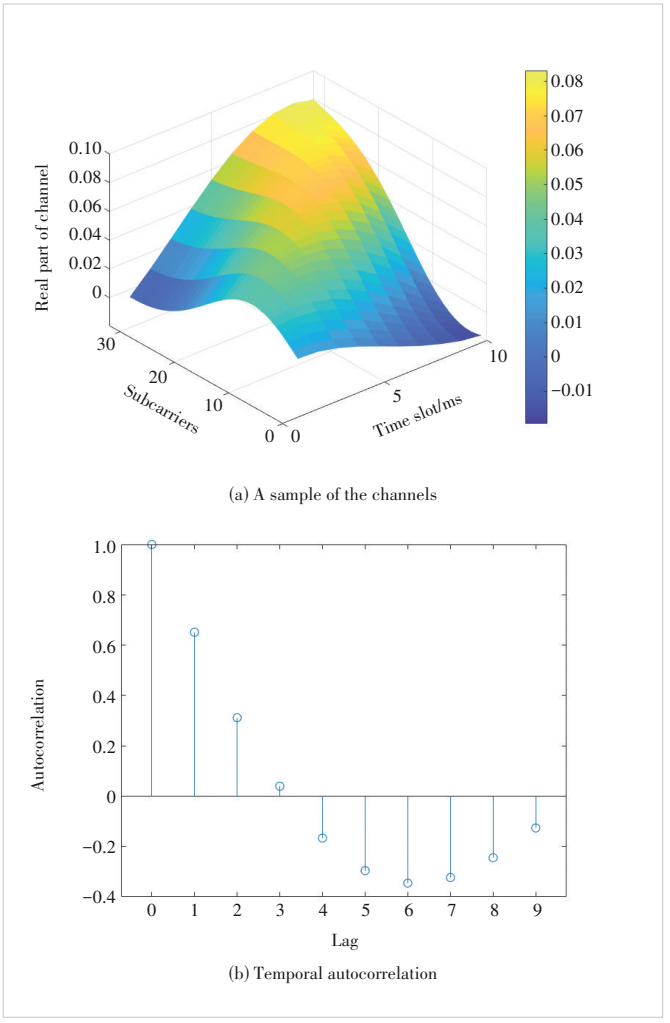


Figure 4. A sample from the CDL-C channel model CSI dataset and the temporal autocorrelation

4.2 Performance Comparison

We primarily compare our CSI prediction module with some existing ones, such as the RNN-based method[13] and the LSTM-based method[14]. To ensure a fair comparison, all baseline prediction methods are jointly trained with the CSI feedback network. The CSI feedback process is implemented using the SwinCFNet architecture with an SNR-adaptive module. Fig. 5 demonstrates the NMSE performance of the proposed and baseline methods at CR=1/4, 1/8 in the CDL-C channel model. The test SNR is set to 20 dB. The performance of non-prediction schemes represents the gaps between the reconstructed nearest historical CSI and the future CSI, which further underscores the importance of channel prediction in the feedback process.

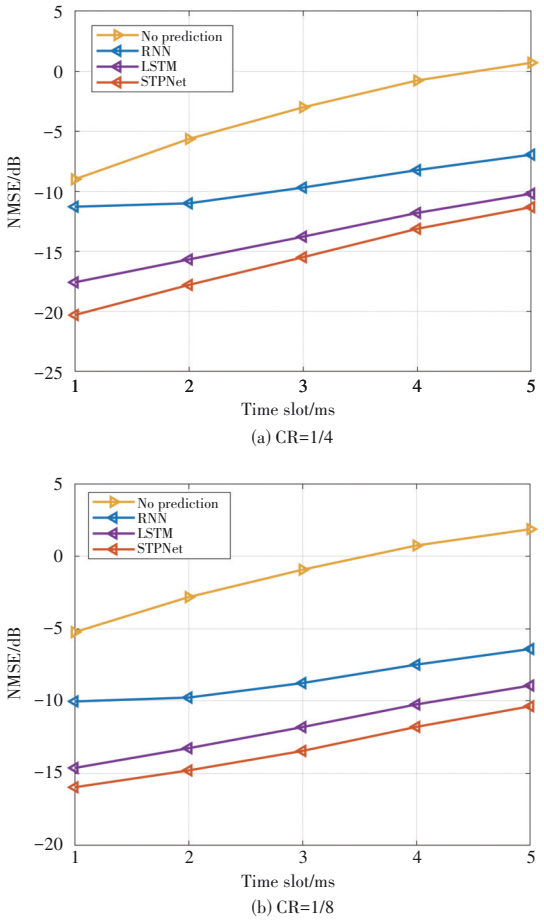


Figure 5. NMSE performance in the CDL-C channel model with CR=1/4 and 1/8

From both Figs. 5a and 5b, it is seen that the NMSE performance of all evaluated algorithms decreases over time. As illustrated in Fig. 5, the proposed Inception-based STPNet achieves the highest performance in the CDL-C channel. For example, when CR is equal to 1/4, STPNet attains NMSE gains of 6.79 dB and 2.12 dB over the RNN-based and LSTM-based methods, respectively, when predicting the channel at the second future slot. Furthermore, compared with the non-prediction scenario, STPNet improves the accuracy of the fifth time slot by more than 12 dB at CR = 1/8. Under these settings, STPNet also achieves an additional 1.43 dB NMSE improvement over the best results of other competing methods.

The improvements of the proposed channel prediction scheme in Fig. 5 come from two aspects. First, the traditional RNN-based prediction methods operate recursively, using the current time slot as input to predict the next. While effective for short-term forecasting, this approach often leads to substantial performance degradation when extrapolating over extended future intervals. In contrast, our proposed scheme predicts all future channels simultaneously, thereby breaking the recursive loop and preventing error accumulation. Second, rather than treating CSI as a time series, our method represents it as a spatial map, capturing the spatio-temporal correlations embedded in the data. By leveraging a multi-branch architecture, the Inception-based CSI prediction module effectively extracts both local and global features from stacks of temporal dynamics.

In Fig. 6, we compare the NMSE performance of STPNet and other prediction networks with CR=1/4 in the UMa channel model generated on QuaDRiGa[26]. The test SNR is set to 20 dB. Since the 3GPP UMa model randomly samples channel parameters, the resulting channels exhibit greater randomness and reduced predictability compared with the CDL-C model. Nevertheless, as shown in Fig. 6, STPNet maintains the state-of-the-art NMSE performance. Notably, for the prediction of the channel at the first future slot, the RNN-based method proves less accurate than the non-prediction approach due to the gradient vanishing problem. Compared with the non-prediction scheme and the LSTM-based method, STPNet achieves improvements in NMSE of 80.99% and 32.56%, respectively.

Furthermore, we investigate the performance of joint CSI feedback and prediction compared with separate CSI feedback and prediction. In the STPNet-separate configuration, CSI feedback and channel prediction networks are trained independently and then evaluated in series. As illustrated in Fig. 7a, the joint architecture, STPNet-joint, achieves at least a 2 dB improvement in NMSE over the STPNet-separate configuration, demonstrating the effectiveness of joint training. Fig. 7b shows the achievable sum-rate performance of different methods. The upper bound is attained by the scheme with perfect channel information available. We can also observe that STPNet-joint could approximate the near-optimal sum-rate performance attained with perfect channel information. For instance, when pre-

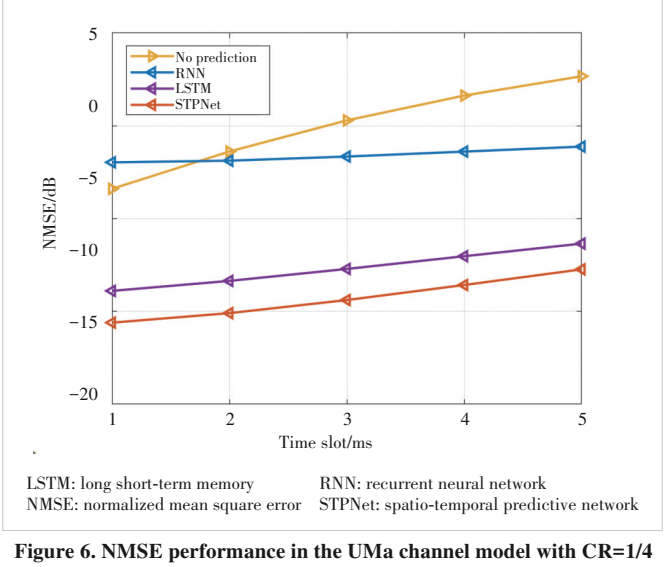
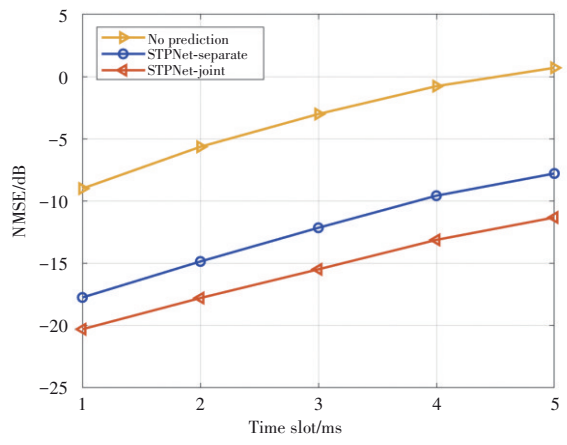
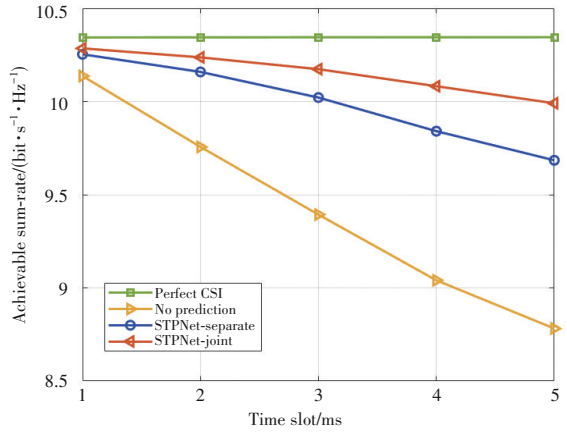


Figure 6. NMSE performance in the UMa channel model with CR=1/4



(a) NMSE performance



(b) Achievable sum-rate performance

CSI: channel state information
LSTM: long short-term memory
NMSE: normalized mean square error
RNN: recurrent neural network
STPNet: spatio-temporal predictive network

Figure 7. NMSE and achievable sum-rate performance of different architectures in the CDL-C channel model with CR=1/4

dicting the channel for the fifth future slot, STPNet-joint achieves approximately 96.57% of the sum-rate performance of the upper bound. By integrating CSI feedback and prediction, the system avoids error propagation between these two cascaded subsystems, thereby enhancing overall accuracy.

5 Conclusions

This paper presents STPNet, an efficient spatio-temporal predictive network based on a joint feedback and prediction framework. STPNet is designed to address the challenges of excessive feedback overhead and dynamic channel conditions in massive MIMO systems. The CSI prediction module is stacked with a series of Inception modules used for capturing both the local and global spatio-temporal features. By leveraging spatio-temporal features and SNR-aware modulation, STPNet achieves outstanding performance in CSI prediction accuracy and robustness, significantly outperforming traditional methods. Simulation results validate the effectiveness of the proposed framework across diverse channel scenarios, demonstrating its potential to enhance future wireless communication systems. Future work will explore extending the model to more complex and dynamic environments, further improving its adaptability and efficiency.

References

[1] CHEN W S, LIN X Q, LEE J, et al. 5G-advanced toward 6G: past, present, and future [J]. IEEE journal on selected areas in communications, 2023, 41 (6): 1592 – 1619. DOI: 10.1109/JSAC.2023.3274037
[2] CHEN W, LIU Y W, JAFARKHANI H, et al. Signal processing and learning for next generation multiple access in 6G [J]. IEEE journal of selected topics in signal processing, 2024, 18(7): 1146 – 1177. DOI: 10.1109/JSTSP.2024.3511403
[3] WEN C K, SHIH W T, JIN S. Deep learning for massive MIMO CSI feedback [J]. IEEE wireless communications letters, 2018, 7(5): 748 – 751. DOI: 10.1109/LWC.2018.2818160
[4] GAO Y CHEN J J, LI D P. Intelligence driven wireless networks in B5G and 6G era: a survey [J]. ZTE communications, 2024, 22(3): 99 – 105. DOI: 10.12142/ZTECOM.202403012
[5] YANG B LIANG X, LIU S N, et al. Intelligent 6G wireless network with multi-dimensional information perception [J]. ZTE communications, 2023, 21(2): 3 – 10. DOI: 10.12142/ZTECOM.202302002
[6] GUO Y R, CHEN W, SUN F F, et al. Deep learning for CSI feedback: one-sided model and joint multi-module learning perspectives [EB/OL]. (2024-05-09)[2024-12-12]. http://export.arxiv.org/abs/2405.05522
[7] CHENG J M, CHEN W, XU J L, et al. Swin Transformer-based CSI feedback for massive MIMO [C]//The 23rd International Conference on Communication Technology (ICCT). IEEE, 2023: 809 – 814. DOI: 10.1109/ICCT59356.2023.10419637
[8] YI X P, YANG S, GESBERT D, et al. The degrees of freedom region of temporally correlated MIMO networks with delayed CSIT [J]. IEEE transactions on information theory, 2014, 60(1): 494 – 514. DOI: 10.1109/TIT.2013.2284500
[9] MIHAI S, YAQOOB M, HUNG D V, et al. Digital twins: a survey on enabling technologies, challenges, trends and future prospects [J]. IEEE communications surveys & tutorials, 2022, 24(4): 2255 – 2291. DOI: 10.1109/COMST.2022.3208773
[10] TAN J SHA X B, DAI B, et al. Analysis of industrial Internet of Things and digital twins [J]. ZTE communications, 2021, 19(2): 53 – 60. DOI: 10.12142/ZTECOM.202102007
[11] YIN H F, WANG H Q, LIU Y Z, et al. Addressing the curse of mobility in massive MIMO with prony-based angular-delay domain channel predictions [J]. IEEE journal on selected areas in communications, 2020, 38 (12): 2903 – 2917. DOI: 10.1109/JSAC.2020.3005473
[12] BADDOUR K E, BEAULIEU N C. Autoregressive modeling for fading channel simulation [J]. IEEE transactions on wireless communications, 2005, 4(4): 1650 – 1662. DOI: 10.1109/TWC.2005.850327
[13] JIANG W, SCHOTTEN H D. Neural network-based fading channel prediction: A comprehensive overview [J]. IEEE access, 2019, 7: 118112 – 118124. DOI: 10.1109/ACCESS.2019.2937588
[14] JIANG W, SCHOTTEN H D. Deep learning for fading channel prediction [J]. IEEE open journal of the communications society, 2020, 1: 320 – 332. DOI: 10.1109/OJCOMS.2020.2982513
[15] JIANG H, CUI M Y, NG D W K, et al. Accurate channel prediction based on transformer: making mobility negligible [J]. IEEE journal on selected areas in communications, 2022, 40(9): 2717 – 2732. DOI: 10.1109/JSAC.2022.3191334
[16] REN Z Z, ZHANG X D, WANG J T. Joint CSI feedback and prediction with deep learning in high-speed scenarios [C]//Proceedings of IEEE/CIC

International Conference on Communications in China (ICCC). IEEE, 2024: 1910 – 1915. DOI: 10.1109/ICCC62479.2024.10681972
[17] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]// Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015: 1 – 9. DOI: 10.1109/CVPR.2015.7298594
[18] GAO Z Y, TAN C, WU L R, et al. SimVP: Simpler yet better video prediction [C]//Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 3160 – 3170. DOI: 10.1109/CVPR52688.2022.00317
[19] XU J L, AI B, CHEN W, et al. Wireless image transmission using deep source channel coding with attention modules [J]. IEEE transactions on circuits and systems for video technology, 2022, 32(4): 2315 – 2328. DOI: 10.1109/TCSVT.2021.3082521
[20] XU J L, AI B, WANG N, et al. Deep joint source-channel coding for CSI feedback: an end-to-end approach [J]. IEEE journal on selected areas in communications, 2023, 41(1): 260 – 273. DOI: 10.1109/JSAC.2022.3221963
[21] YANG K, WANG S X, DAI J C, et al. WITT: a wireless image transmission transformer for semantic communications [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1 – 5. DOI: 10.1109/ICASSP49357.2023.10094735
[22] DENG L T ZHAO Y R. Deep learning-based semantic feature extraction: a literature review and future directions [J]. ZTE communications, 2023, 21(2): 11 – 17. DOI: 10.12142/ZTECOM.202302003
[23] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz: TR 38.901 V18.0.0 [S]. 2024
[24] WU C, YI X P, ZHU Y M, et al. Channel prediction in high-mobility massive MIMO: from spatio-temporal autoregression to deep learning [J]. IEEE journal on selected areas in communications, 2021, 39(7): 1915 – 1930. DOI: 10.1109/JSAC.2021.3078503
[25] YUAN J D, NGO H Q, MATTHAIOU M. Machine learning-based channel prediction in massive MIMO with channel aging [J]. IEEE transactions on wireless communications, 2020, 19(5): 2960 – 2973. DOI: 10.1109/TWC.2020.2969627
[26] JAECKEL S, RASCHKOWSKI L, BÖRNER K, et al. Quasi deterministic radio channel generator, user manual and documentation [R]. Berlin, Germany: QuaDRiGa, 2021

Biographies

CHENG Jiaming received his BE degree from Beijing Jiaotong University, China in 2024, where he is currently pursuing his PhD degree. His current research interests include massive MIMO and intelligent communications.

CHEN Wei (weich@bjtu.edu.cn) received his BE and ME degrees from the Beijing University of Posts and Telecommunications, China in 2006 and 2009, respectively, and PhD degree in computer science from the University of Cambridge, UK in 2013. Later, he was a research associate with the Computer Laboratory, University of Cambridge, from 2013 to 2016. He is currently a professor with Beijing Jiaotong University, China. He has published over 130 articles and won several international awards. His current research interests include intelligent wireless communication systems and multimedia processing.

LI Lun received his MS degree in electronics and communication engineering from Harbin Institute of Technology, China in 2018. He joined ZTE Corporation in 2018, where he is currently a technical pre-research engineer. His research interests include artificial intelligence/machine learning for wireless communications.

AI Bo received his MS and PhD degrees from Xidian University, China in 2002 and 2004, respectively. He is currently a full professor with Beijing Jiaotong University, China. He has authored/coauthored eight books and published over 300 academic research articles. His research interests include the research and applications of channel measurement and channel modeling, and dedicated mobile communications for rail traffic systems. He has received many awards, such as the Distinguished Youth Foundation and the Excellent Youth Foundation from the National Natural Science Foundation of China, the Qiushi Outstanding Youth Award by Hong Kong Qiushi Foundation, the New Century Talents by the Chinese Ministry of Education, the Zhan Tianyou Railway Science and Technology Award by the Chinese Ministry of Railways, and the Science and Technology New Star Award by the Beijing Municipal Science and Technology Commission.