Spectral Temporal Graph Neural Network for Massive MIMO CSI Prediction

Sharan Mourya, Pavan Reddy, SaiDhiraj Amuru, and Kiran Kumar Kuchi

***Abstract*—In the realm of 5G communication systems, the accuracy of Channel State Information (CSI) prediction is vital for optimizing performance. This letter introduces a pioneering approach: the Spectral-Temporal Graph Neural Network (STEM GNN), which fuses spatial relationships and temporal dynamics of the wireless channel using the Graph Fourier Transform (GFT). We compare the STEM GNN approach with conventional Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and Transformer models for CSI prediction. Our findings reveal a significant enhancement in overall communication system performance through STEM GNNs. For instance, in one scenario, STEM GNN achieved a spectral efficiency of 4.683 bps/Hz which is 16.5% higher than that of a transformer, 63% higher than an LSTM and 198% higher than that of an RNN. The spectral-temporal analysis capabilities of STEM GNNs capture intricate patterns often overlooked by traditional models, offering improvements in beamforming and interference mitigation.**

***Index Terms*—Graph neural networks, STEM GNN, CSI feedback, CSI prediction, STNET, massive MIMO.**

I. Introduction

IN TIME division duplex (TDD) cellular networks, channel reciprocity enables downlink channel state information (DL-CSI) inference from uplink channel state information (UL-CSI), obviating the need for extra estimation or feedback. However, this reciprocity principle doesn't hold for frequency division duplex (FDD) systems, introducing challenges in DL-CSI acquisition. Prior methods [1], [2] required receivers to predict and feedback DL-CSI, incurring computational load and overheads. Addressing this, researchers have delved into channel prediction techniques which leverage historical CSI correlations to forecast future channel conditions. Existing methods mainly fall into model-based or neural network-based categories. Model-based strategies employ techniques like linear extrapolation, sum-of-sinusoids, and autoregression (AR) [1] to capture dynamic channel behavior. However, complexities arising from factors like multipath and Doppler effects pose challenges to precise modeling. Consequently, data-driven approaches have gained traction [1] which train neural networks to mirror real-world channel behaviors.

Manuscript received 14 February 2024; accepted 27 February 2024. Date of publication 1 March 2024; date of current version 10 May 2024. The associate editor coordinating the review of this article and approving it for publication was Q. Guo. *(Corresponding author: Sharan Mourya.)*

Sharan Mourya is with the Electrical and Computer Engineering Department, University of Illinois at Urbana–Champaign, Urbana, IL, USA (e-mail: sharanmourya7@gmail.com).

Pavan Reddy is with Wisig Networks Pvt. Ltd., Hyderabad 500089, India.

SaiDhiraj Amuru and Kiran Kumar Kuchi are with the Electrical Engineering Department, Indian Institute of Technology Hyderabad, Sangareddy 502285, India.

Digital Object Identifier 10.1109/LWC.2024.3372148

In [2], authors employed feed-forward neural networks for CSI prediction, directly linking uplink and downlink CSI. However, their assumption of a bijective user-channel matrix relationship isn't always applicable. To address this, a more advanced model using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) was proposed in [3], notably outperforming methods like Maximum Likelihood (ML) and Minimum Mean Squared Error (MMSE). A similar data-driven model based on CNN and LSTM was also presented by Wang et al. [4] for CSI prediction. Recurrent Neural Networks (RNNs) were employed in some strategies [1], [5], often relying on CNNs for feature extraction, despite being primarily designed for real-valued image tasks. To alleviate this limitation, Zhang et al. introduced a complex-valued 3D CNN for CSI prediction in [6], improving accuracy over real-valued networks. While LSTMs excel in capturing long-range correlations compared to RNNs, they face the vanishing gradient problem. To circumvent this, a transformer-based CSI prediction model was introduced in [7]. Transformers employ self-attention mechanisms, overcoming vanishing gradients and preserving long-range correlations more effectively than LSTMs. Additionally, transformers resolve the sequential prediction problem, where predicting each future CSI matrix relies on historical CSI and the latest predicted CSI, leading to error propagation. Transformers process data in parallel, sidestepping the sequential prediction challenge and efficiently forecasting future CSI matrices.

Another category of deep learning architectures with high retention of long-range correlation is the Graph Neural Network (GNN) [8]. When dealing with long-range correlations, traditional neural networks like feed-forward or recurrent neural networks may struggle due to their local receptive fields and sequential nature. However, GNNs are well-suited to address this challenge as they incorporate message passing between nodes in a graph iteratively. Each node aggregates information from its neighboring nodes and updates its representation based on the gathered information. This process allows information to propagate across the graph, enabling nodes to capture correlations with more distant nodes. On the other hand, one of the fundamental components of GNNs is the graph convolutional layer [8] which incorporates both local and non-local information, enabling nodes to capture correlations with their neighbors as well as more distant nodes in the graph making them highly suitable for wireless communications [9].

In this letter, we propose to use a spectral temporal (STEM) graph neural network [10] for CSI prediction. Originally STEM GNN was proposed to solve the problem of time series forecasting [11] on data with complex relational structures represented as graphs. STEM GNNs are particularly useful

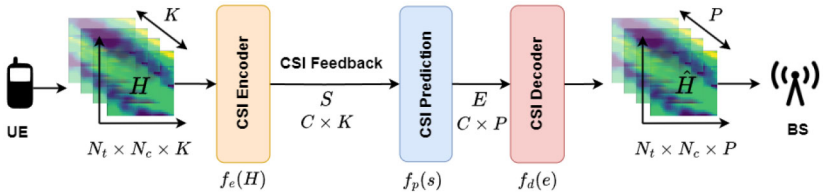


Fig. 1. System Model.

when dealing with multivariate time series data, where each time step contains multiple variables (e.g., temperature, humidity, pressure, etc.) and the interactions between variables can be represented as a graph. We model CSI prediction as a multivariate time-series forecasting problem and to the best of our knowledge, we are the first to do it.

II. System Model

We consider a MIMO orthogonal frequency division multiplexing (OFDM) system. In Fig. 1, the base station (BS) is outfitted with an array of antennas organized in an $M \times N$ rectangular pattern, with each antenna featuring dual polarization resulting in a total of $N_t = M \times N \times 2$ transmit antennas and the user equipment (UE) has a single antenna. The number of sub-carriers is N_c. Given that the CSI prediction operates independently for various users, it is sufficient to illustrate the principle using a single user. In this context, we consider the communication channel between the user and the BS to be represented by a three-dimensional (3D) channel model, conforming to the specifications outlined in the 3rd Generation Partnership Project (3GPP) for 5G communications [12]. This channel is denoted as H. The received signal y_n on the n^{th} sub-carrier can be denoted by

$$y_n = {h_n}^H x_n + w_n, \tag{1}$$

where, $h_n \in \mathbb{C}^{N_t \times 1}$, $x_n \in \mathbb{C}$, and $w_n \in \mathbb{C}$ are the channel, symbol and additive noise respectively. The channel matrix H of size $N_c \times N_t$ is given by

$$H = \left[h_1, h_2, \ldots, h_{N_c}\right]^H. \tag{2}$$

The matrix representing this channel consists of a total of $2N_tN_C$ elements, considering that the entries within the matrix are of a complex nature. However, in practical scenarios, transmitting such a sizable matrix proves to be unfeasible for a massive MIMO system. To address this challenge, we adopt a strategy to reduce the payload size. This entails compressing the matrix H, which is of a 3D nature, into a more compact one-dimensional vector with dimensions $C \times 1$, as depicted in Fig. 1. To quantify CSI compression, we introduce the compression ratio denoted by $\gamma = \frac{C}{2N_CN_t}$. The compressed one-dimensional vector is then transmitted to the BS. Upon reception, a CSI prediction model is applied to this vector to forecast future CSI in its compressed form. Subsequently, a decoder is employed to reconstruct the estimated channel matrix, denoted as $\hat{H}$, from the predicted vector. This entire sequence of operations can be described through the following equations:

$$s = f_e(H), \quad e = f_p(s), \quad \hat{H} = f_d(e),$$

where f_e is the encoder, f_p is the CSI prediction model, and f_d is the decoder. s, e, and $\hat{H}$ are the compressed, predicted, and estimated channel matrices respectively. Using the estimated channel matrix $\hat{H}$, BS utilizes the CSI prediction model to estimate the channel matrices for the future frames. We address the number of such frames as the horizon (denoted by P) and it would be specified in the CSI prediction model while training.

III. Spectral Temporal Graph Neural Network

Having established the system model, let's delve into the CSI prediction model depicted in Fig. 1. Conventional CSI prediction relies on auto-regression [1], [13] or hidden Markov chains [14], which capture temporal correlations between historical and future CSI but neglect spatial correlations. Thus, deep learning emerged, employing CNNs to capture the spatial relationships in CSI which is viewed as a spatial map, with each element representing a channel aspect. CNN-based models transform raw CSI into images highlighting patterns like signal fluctuations and interference.

Typically, CNN-based CSI models include an RNN or LSTM to capture temporal trends [1], [3], [4], [5] by taking historical CSI sequences as input to learn patterns and trends. RNNs and LSTMs retain memory, accurately inferring future states. However, this separate spatial-temporal approach misses the concurrent spatio-temporal correlations present in CSI evolution. We propose a unified model using GNNs, effectively extracting these concurrent correlations for CSI prediction. STEM GNNs [10] offer a profound approach to comprehending and predicting multivariate time-series patterns intertwined in space and time. Predicting such a series is intricate [11], requiring simultaneous consideration of intra-series and inter-series correlations. In CSI prediction, these map to spatial and temporal correlations respectively. STEM GNNs adeptly handle both simultaneously within the spectral domain by merging Graph Fourier Transform (GFT) [8], for spatial correlations and Discrete Fourier Transform (DFT), for temporal dependencies, into a single framework.

A. Problem Formulation

Let us establish the graph that captures the evolution of the channel as $G = (S, W)$, where $\{S = s_{t-K}, s_{t-K+1}, \ldots, s_{t-1}\} \in \mathbb{R}^{C \times K}$ signifies the historical compressed CSI input. In this context, $s(i, j)$ represents the latent representation vector with dimension (nodes) C, and K represents the number of timestamps. S is obtained from the UE after CSI compression and it is the codeword transmitted to the BS. The compressed CSI values observed at timestamp

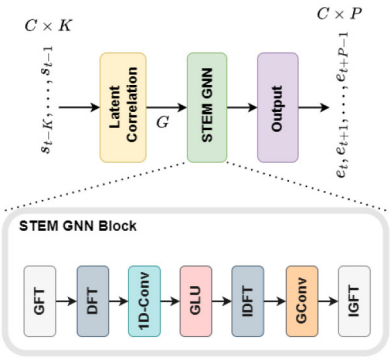


Fig. 2. STEM GNN Architecture.

t are symbolized as $s_t \in \mathbb{R}^C$. The adjacency matrix is denoted as $W \in \mathbb{R}^{C \times C}$, with $w_{ij} > 0$ indicating the presence of a connection between nodes i and j, where the magnitude of w_{ij} quantifies the strength of this connection. The procedure to find W is discussed in the following section.

Given the observed values from the K preceding timestamps $S = \{s_{t-K}, \ldots, s_{t-1}\}$, the objective of CSI prediction revolves around anticipating the node values within a multivariate temporal graph $G = (S, W)$ for the subsequent P timestamps, noted as $E = \{e_t, e_{t+1}, \ldots, e_{t+P-1}\}$. These values are deducible through the predictive model f_p governed by parameters θ.

$$E = \{e_t, e_{t+1}, \ldots, e_{t+P-1}\} = f_p(s_{t-K}, \ldots, s_{t-1}; G; \theta). \tag{3}$$

Now, using the decoder f_d we decompress the predicted compressed CSI to obtain the future CSI as shown in Fig. 1. Horizon (P) and window size (K) depend on various factors such as multi-path fading, user mobility, and coherence time and optimizing them with respect to the spectral efficiency is outside the scope of this letter.

B. STEM GNN Architecture

STEM GNN consists of three main layers each having its own importance: latent correlation layer, STEM GNN layer, and output layer as shown in Fig. 2. The latent correlation layer is tasked with finding the adjacency matrix (W) of the graph G. As it is difficult to construct an adjacency matrix of historical compressed CSI (S) by hand, a self-attention layer [15] is used to learn the latent correlations between multiple compressed CSI vectors automatically. In this way, the model exploits correlations to form an adjacency matrix in a data-driven fashion. The extracted adjacency matrix W along with the graph structure G are fed to the next layer. STEM GNN layer has a couple of STEM GNN blocks (as shown in Fig. 2) interconnected with skip connections and addition blocks [10]. This forms the crust of the CSI prediction model as this takes G as an input and extracts the spatio-temporal correlations from the historical compressed CSI which are then used by the output layer to create accurate predictions for the following P timestamps.

The STEM GNN block encompasses several operations, including GFT and its inverse IGFT, DFT and its inverse IDFT, a 1D convolution layer (1D-Conv), followed by a gated linear unit (GLU) activation, and the pivotal graph convolution (GConv). In the initial steps, GFT transforms the graph G into a spectral matrix representation, rendering compressed CSI independent in the spatial domain. Subsequently, DFT shifts univariate time-series components into the frequency domain, and 1D convolution with GLU captures feature patterns before reverting through inverse DFT. The cornerstone GConv operation, crucial in any GNN, propagates information across the graph. The sequence concludes with inverse GFT. This entire process is visually depicted in Fig. 2.

TABLE I
CHANNEL MODEL

Scenario	Urban Macro (UMa)
Antenna Config.	$[8, 2, 2] \Rightarrow N_t = 8 \times 2 \times 2 = 32$
No. of Subcarriers (N_c)	32
No. of consecutive timestamps (T)	10,000
Bandwidth	20 MHz
User Mobility	3Kmph, 30Kmph, 120Kmph

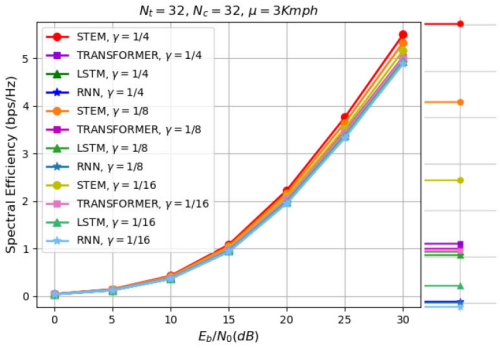
IV. RESULTS

Utilizing the STEM GNN architecture and the problem formulation outlined earlier, we conducted a sequence of simulations using the 3GPP channel model, characterized by the following specifications mentioned in Table I. The BS antenna is modelled by a uniform rectangular panel array, comprising 8×2 panels. The antenna panel is dual polarized hence the antenna configuration is [8, 2, 2]. We choose window size (K) in Eq. (3), i.e., the number of previous timestamps to input to the CSI prediction model to be 12. Dimension of the compressed CSI, C is chosen from {128, 256, 512} which translates to compression ratio of $\gamma = \{1/16, 1/8, 1/4\}$ respectively. With these settings, we compare our approach with RNN, LSTM and transformer-based CSI prediction models as they outperform many of the traditional CSI prediction techniques [1]. In addition, for CSI compression and feedback we use STNet [16] as it has the best performance per resource utilization among all the other CSI feedback networks [17].

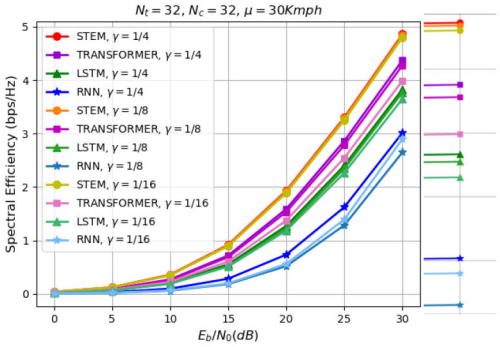
A. Spectral Efficiency Performance

Using transmit precoding, a massive MIMO system can attain high capacities. In this regard, we employ the prevalent linear Zero-Forcing (ZF) transmit precoding technique [16] to assess the overall enhancement in communication system performance brought about by distinct CSI feedback approaches. The evaluation process involves plotting the spectral efficiency of each method against the signal-to-noise ratio (SNR) for varying compression ratios. The resulting depiction can be observed in Fig. 2.

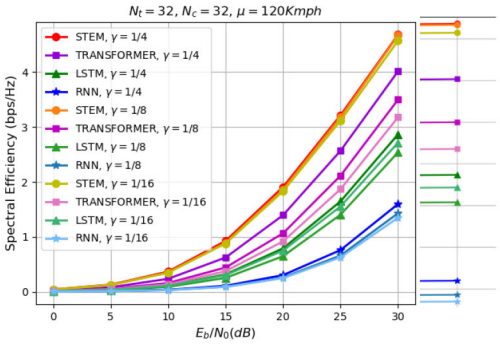
Firstly, we note that for every approach, the spectral efficiency increases with γ which is expected as the loss of information in channel matrices is low at high compression ratios. Secondly, the difference in the spectral efficiency performances of STEM GNN, RNN, LSTM and transformer becomes wider and more evident as the mobility (μ) of the UE increases from 3*Kmph* to 120*Kmph* which is expected



(a) Mobility, $\mu = 3Kmph$



(b) Mobility, $\mu = 30Kmph$



(c) Mobility, $\mu = 120Kmph$

Fig. 3. Spectral Efficiency vs E_b/N_0 plots comparing STEM GNN with RNN, LSTM, and transformer for CSI prediction with various mobility values, $\mu = \{3, 30, 120\}$ at $K = 12$.

behaviour as unlike other approaches, STEM GNN is equipped with enough representation power to generalize the compressed CSI for higher mobilities. For instance, at $\mu = 120Kmph$ and $E_b/N_0 = 30dB$, STEM GNN achieves a spectral efficiency of 4.683 bps/Hz which is 16.5% higher than that of transformer, 63% higher than LSTM and 198% higher than that of RNN. For a 20 MHz bandwidth channel, this is equivalent to a total data rate of 93.6 MHz, 80.36 MHz, 57.38 MHz and 31.96 for STEM GNN, transformer, LSTM, and RNN respectively. Additionally, stability of STEM GNN with varying mobility is substantially higher than other approaches. For instance, at $E_b/N_0 = 30dB$, spectral efficiency of STEM GNN only changes by 1.4% when mobility changes from 30*Kmph* to 120*Kmph* whereas the spectral efficiency of the transformer, LSTM, and RNN

TABLE II
COMPLEXITY

S.No	Model	C	Runtime (sec)	Parameters
1	RNN	512	0.0010	1299136
		256	0.0007	884896
		128	0.0005	626777
2	LSTM	512	0.0016	1502560
		256	0.0013	942592
		128	0.0010	841268
3	Transformer	512	0.0073	1716000
		256	0.0067	1018400
		128	0.0055	969600
4	STEM	512	0.0095	1867329
		256	0.0089	1266241
		128	0.0079	1113153

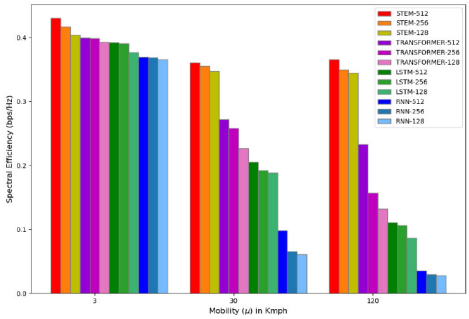


Fig. 4. Spectral efficiency performance of RNN, LSTM, transformer, and STEM GNN-based CSI prediction models for various compression ratios (γ) and mobilities (μ) at $E_b/N_0 = 30dB$.

drops by 14.1%, 46.1%, and 64.5% respectively as shown in Fig. 4.

B. Complexity

In Table II, we tabulated the number of parameters used by each approach along with the runtime of the model for a single channel usage. We notice that the complexity reduces with the compression ratio which is expected as the model size decreases with the latent space. We also notice that STEM GNN is computationally more complex than the other approaches. For instance, at $C = 512$ STEM GNN is 2.2*ms* slower than the transformer. However, the tremendous increase in spectral efficiency and stability over large mobilities still makes STEM GNN preferrable over other approaches.

C. CSI Prediction Performance

The accuracy of CSI prediction by various networks can be measured using root mean squared error (RMSE) [10] between the original future CSI and the predicted future CSI. We plotted the RMSE values for RNN, LSTM, transformer, and STEM GNN networks for various compression ratios and horizons as shown in Fig. 5. It should be noted that the RMSE of STEM GNN is much lower when compared to RNN, LSTM, and transformer networks. For instance, at $P = 3$ and $\gamma = 1/4$, STEM GNN achieved an RMSE of 1.5832 whereas RNN, LSTM, and transformer achieved an RMSE of 24.4404, 22.2726, and 20.166 which are more than 12-fold higher than that of STEM GNN.

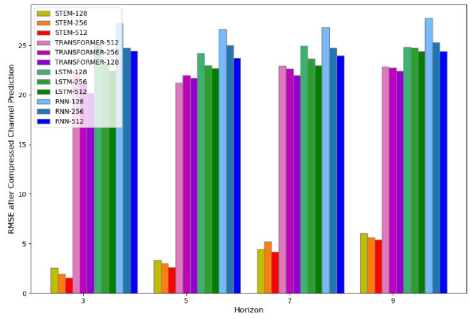


Fig. 5. RMSE performance of RNN, LSTM, transformer, and STEM GNN-based CSI prediction models for various compression ratios (γ) and horizons (P) at $\mu = 3Kmph$.

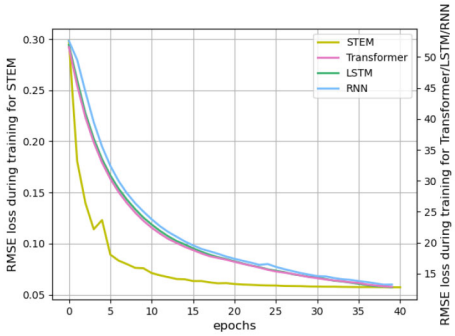


Fig. 6. Training loss vs the number of epochs for RNN, LSTM, transformer, and STEM GNN-based CSI prediction models at the horizon, $P = 5$, $\mu = 3Kmph$ and $\gamma = 1/4$.

D. Training Performance

Training of all the networks was achieved on an Nvidia RTX 3060 graphic card.[1] Number of epochs was set to 100, batch size to 50, and learning rate to 0.0001. RMSProp optimizer is used along with RMSE as the loss function. Training loss is an important metric in understanding the stability and convergence of a network with respect to the dataset. Therefore, we plotted a training loss plot varying with the number of epochs in Fig. 6. From the figure, it can be seen that the training loss of STEM GNN is at least two orders of magnitude less than that of RNN, LSTM, and transformer. In addition, the STEM GNN model also converges significantly faster than others.

V. Conclusion

In this letter, we study the Spectral-Temporal Graph Neural Network (STEM GNN) approach for CSI prediction in 5G communication systems and demonstrate its superiority over traditional RNN, LSTM, and transformer models. Leveraging both spatial relationships and temporal dynamics through the Graph Fourier Transform, STEM GNNs exhibit a remarkable improvement in overall communication system performance.

[1]Source code: https://github.com/sharanmourya/CSI-Prediction/.

References

[1] W. Jiang and H. D. Schotten, "Neural network-based fading channel prediction: A comprehensive overview," *IEEE Access*, vol. 7, pp. 118112–118124, 2019.

[2] Y. Yang, F. Gao, G. Y. Li, and M. Jian, "Deep learning-based downlink channel prediction for FDD massive MIMO system," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1994–1998, Nov. 2019.

[3] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 227–236, Jan.–Mar. 2020.

[4] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access*, vol. 7, pp. 96105–96112, 2019.

[5] W. Jiang and H. D. Schotten, "Deep learning for fading channel prediction," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 320–332, 2020.

[6] Y. Zhang et al., "CV-3DCNN: Complex-valued deep learning for CSI prediction in FDD massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 266–270, Feb. 2021.

[7] H. Jiang, M. Cui, D. W. K. Ng, and L. Dai, "Accurate channel prediction based on transformer: Making mobility negligible," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2717–2732, Sep. 2022.

[8] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021. [Online]. Available: https://doi.org/10.1109

[9] S. Mourya, P. Reddy, S. Amuru, and K. K. Kuchi, "Graph neural networks-based user pairing in wireless communication systems," 2023, *arXiv:2306.00717*.

[10] D. Cao et al., "Spectral temporal graph neural network for multivariate time-series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 17766–17778.

[11] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philosoph. Trans. Roy. Soc. A*, vol. 379, no. 2194, 2021, Art. no. 20200209.

[12] Q. Zhu et al., "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Sophia Antipolis, France, Rep. TR 38.901, 2021.

[13] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive MIMO: From spatio-temporal autoregression to deep learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1915–1930, May 2021.

[14] Z. Chen and R. C. Qiu, "Prediction of channel state for cognitive radio using higher-order hidden Markov model," in *Proc. IEEE SoutheastCon (SoutheastCon)*, 2010, pp. 276–282.

[15] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[16] S. Mourya, S. D. 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.

[17] J. Guo, C. K. Wen, S. Jin, and G. Y. Li, "Overview of deep learning-based CSI feedback in massive MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8017–8045, Oct. 2022.