

Research Article

# Large Models Enabled Ubiquitous Wireless Sensing

Shun Hu<sup>1</sup>

1. School of Science and Engineering, The Chinese University of Hong Kong (Shen Zhen), China

In the era of 5G communication, the knowledge of channel state information (CSI) is crucial for enhancing network performance. This paper explores the utilization of language models for spatial CSI prediction within MIMO-OFDM systems. We begin by outlining the significance of accurate CSI in enabling advanced functionalities such as adaptive modulation. We review existing methodologies for CSI estimation, emphasizing the shift from traditional to data-driven approaches. Then a novel framework for spatial CSI prediction using realistic environment information is proposed, and experimental results demonstrate the effectiveness. This research paves way for innovative strategies in managing wireless networks.

Corresponding author: Shun Hu, [223010027@link.cuhk.edu.cn](mailto:223010027@link.cuhk.edu.cn)

## I. Introduction

In the background of widespread 5G, the users around the globe benefit from high-speed characteristics of such a new communication technology. Meanwhile, the demand for more reliable and low-latency transmission is also increasing. Efficient management and adjustment of large-scale antenna arrays is an essential key to optimizing communication. Traditional management techniques are typically reactive, relying on feedback from users to adjust to the constantly changing wireless environment. But recent research emphasizes a forward-looking approach using CSI prediction. This proactive prediction-based management allows for more dynamic adjustment of network resources.<sup>[1]</sup>

To explain the importance of CSI, it is needed to point out that CSI is essential for advanced functionalities of communication system such as precoding, bit-loading, adaptive modulation, channel-aware scheduling, and beamforming. In a MIMO setup, having precise knowledge of the CSI at the transmitter enables significantly higher channel capacity compared to transmission without CSI, improving the reliability of traditional communication systems.<sup>[2]</sup>

In this context, CSI prediction is becoming an integral part. People used to utilize different approaches to predict channels, and various parametric models have been explored, such as the autoregressive (AR) model, the sum-of-sinusoids model, and the polynomial extrapolation model. As for concrete computation, a prediction algorithm called Prony-based angular-delay domain (PAD) was introduced as well. Despite their usefulness, these methods may struggle to accurately capture the complex features of real-world channels.

In recent years, researchers have sought innovative techniques to tackle the complexity of CSI prediction, ranging from traditional algorithms to modern algorithms, so the technique has evolved significantly. Meanwhile, AI tools are growing at a high pace, which helps in increasing the precision of prediction.

The content of this report includes:

- Foundational Concepts in Wireless Channels: An overview of wireless channel theory, channel modeling, Massive MIMO, and OFDM fundamentals, establishing a basis for advanced prediction techniques.
- Time-Series Channel Prediction with *LLM4CP*: Replication of time-series prediction experiments using the “*LLM4CP*” framework.
- Spatial CSI Prediction Studies: Exploration of spatial CSI prediction with large models.

The contributions of this paper include:

- Application of Large models for Spatial CSI Prediction.
- Development of a preprocessing framework which merges geometric-physical knowledge for robust prediction.

The notations which could be used are presented in the table below.

$\mathbf{A}_{p \times q}$	$p \times q$ matrix
$\ \mathbf{A}\ _2$	Frobenius norm of matrix
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{A}^H$	Hermitian of $\mathbf{A}$

## II. Review on Related Works

The foundation of modern CSI prediction can be traced back to early works like<sup>[3]</sup>, which addressed the challenge of adaptive channel estimation in MIMO systems, focusing on the use of feedback channels to

optimize transmission. The work by<sup>[4]</sup> highlights the challenges in traditional methods of estimating channel coefficients.

Building on this,<sup>[5]</sup> address the limitations imposed by the dimensionality of CSI in massive MIMO systems. They propose leveraging compressive sensing and deep learning to enhance the efficiency of CSI estimation, as a groundbreaking work.

In 2020, <sup>[6]</sup> further advanced the discussion by comparing traditional Kalman filtering techniques with emerging machine learning methods. Then the research shifts towards data-driven methodologies is echoed in the work of<sup>[7]</sup>, who thoroughly explore the potential of deep learning to enhance CSI prediction accuracy.

<sup>[8]</sup> contribute to this discourse by proposing a novel 3D CNN framework for predicting future CSI in highly mobile environments. This trend towards utilizing advanced neural network architectures is further supported by <sup>[9]</sup>, who introduce a domain transformation technique to enhance training efficiency for channel prediction, highlighting the ongoing refinement of machine learning applications in CSI estimation.

The exploration of meta-learning and deep denoising techniques by <sup>[10]</sup> illustrates how such new strategies can alleviate the challenges posed by limited pilot data, thereby enhancing the performance of channel predictors in real-world scenarios. Apart from that, <sup>[11]</sup> expand the range of channel prediction to spatial-temporal domain, where two different tools CNN and convolutional variation of LSTM is combined to be namely Conv-LSTM to capture the inherent complexity of massive MIMO systems, which can be found in <sup>[10]</sup>.

In a more recent study, <sup>[12]</sup> delve into the applicability of meta-learning for adaptive predictors to new environments, emphasizing the necessity for models that can quickly adjust to varying conditions without extensive retraining.

In the context of advancing 6G technologies, <sup>[1]</sup> introduces a novel framework, the digital radio twin, which leverages U-Net for accurate spatial-CSI prediction.

Finally, <sup>[13]</sup> provides a comprehensive survey on LLMs in telecommunications, highlighting their capabilities in managing complex network tasks and optimizing CSI prediction processes.

### **III. Basic Concepts of Channels**

#### *A. General Theory of Channel Models*

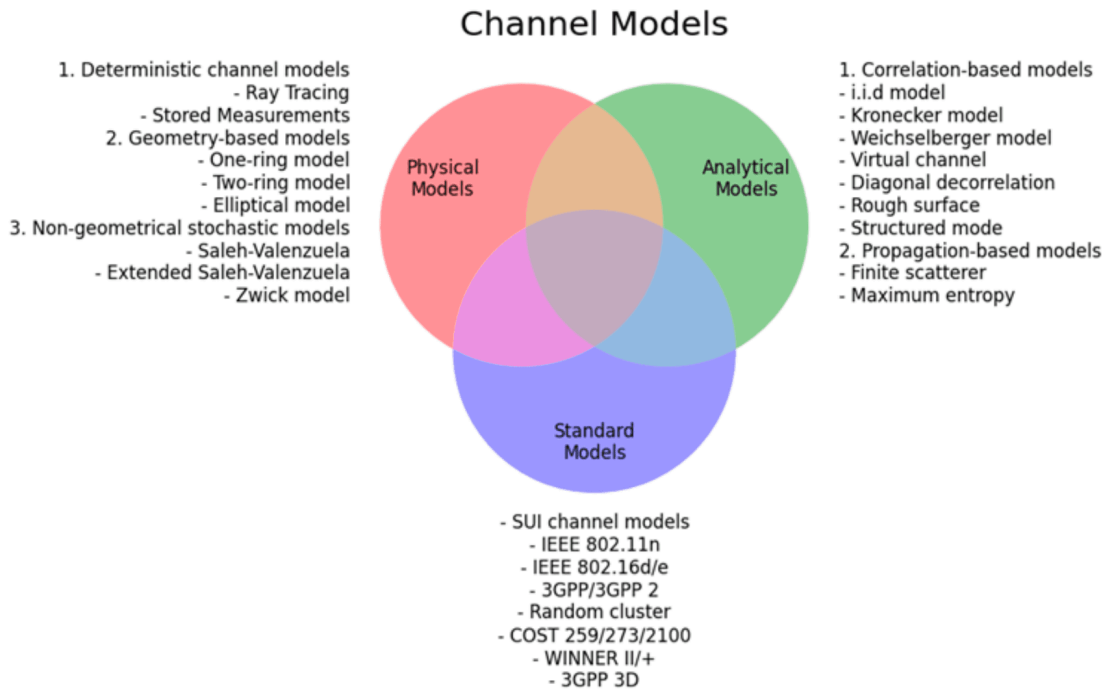
In wireless communication systems, the channel is a critical factor in determining the overall performance.

In a MIMO system with  $m$  transmitting antennas and  $n$  receiving antennas, the channel is typically

represented by a matrix, named CSI, as shown below:

$$H(t, \tau) = \begin{bmatrix} h_{11}(t, \tau) & h_{12}(t, \tau) & \dots & h_{n1}(t, \tau) \\ h_{21}(t, \tau) & h_{22}(t, \tau) & \dots & h_{n2}(t, \tau) \\ \vdots & \vdots & \ddots & \vdots \\ h_{1m}(t, \tau) & h_{2m}(t, \tau) & \dots & h_{nm}(t, \tau) \end{bmatrix} \quad (1)$$

The models of channels can be categorized in several ways, and the diagram is shown in Fig.1, which shows three ways of classification: physical models, analytical models, and standard models.



**Figure 1.** Modern classification of wireless channels

Physical models use basic theories in physics to describe the multipath propagation environment between the transmitter and receiver, considering factors like the direction of arrival (DoA), direction of departure (DoD), signal amplitude, and the delay of multipath components (MPCs).

Analytical models focus on the channel impulse response or transfer function between specific pairs of transmitter and receiver antennas. These models do not directly account for wave propagation but instead synthesize a channel gain matrix that describes the overall system behavior.

Standard models are designed mostly by industrial organizations and companies to assist in new wireless communication technologies.

## *B. OFDM Systems*

In modern digital wireless communication systems, orthogonal frequency division multiplexing (OFDM) technology is widely used to combat frequency-selective fading in multipath propagation environments, such as in standard IEEE 802.11a/g/n.

## **IV. WINNER Models and Ray Tracing Models**

Some of the channel dataset used in this work is generated by QuaDRiGa, which is based on the earlier WINNER model. Other data are generated using ray-tracing models, for realistic modelling and more stable prediction.

### *A. WINNER*

WINNER supports a wide frequency range, making it suitable for mmWave and for scenarios with diverse transceiver techniques. WINNER models are adopted by QuaDRiGa. <sup>[14]</sup>

### *B. Ray Tracing*

Another important channel modeling used in this study is based on ray tracing, which is a deterministic method used to model wave propagation by simulating the paths of electromagnetic waves through an environment. Unlike statistical models that provide average estimates, ray tracing computes the interactions between waves and obstacles using the principles of geometric optics. This results in a more precise description of the channel characteristics.

### *C. Softwares Using Ray Tracing*

WinProp (Wireless In-building Propagation) integrates a range of deterministic, empirical, and hybrid models, making it well-suited for wireless network planning and coverage analysis. Winprop's interests go further than just principle-level verification since it can also be used to generate realistic datasets. WinProp supports the import of CAD files, offering flexibility in modeling different environments.

In practice, One can adopt different approaches based on the electrical construction size and complexity (Fig.2). It is reasonable that in large urban scenarios, more approximations will be introduced in simulation, instead of solving complex wave equations.

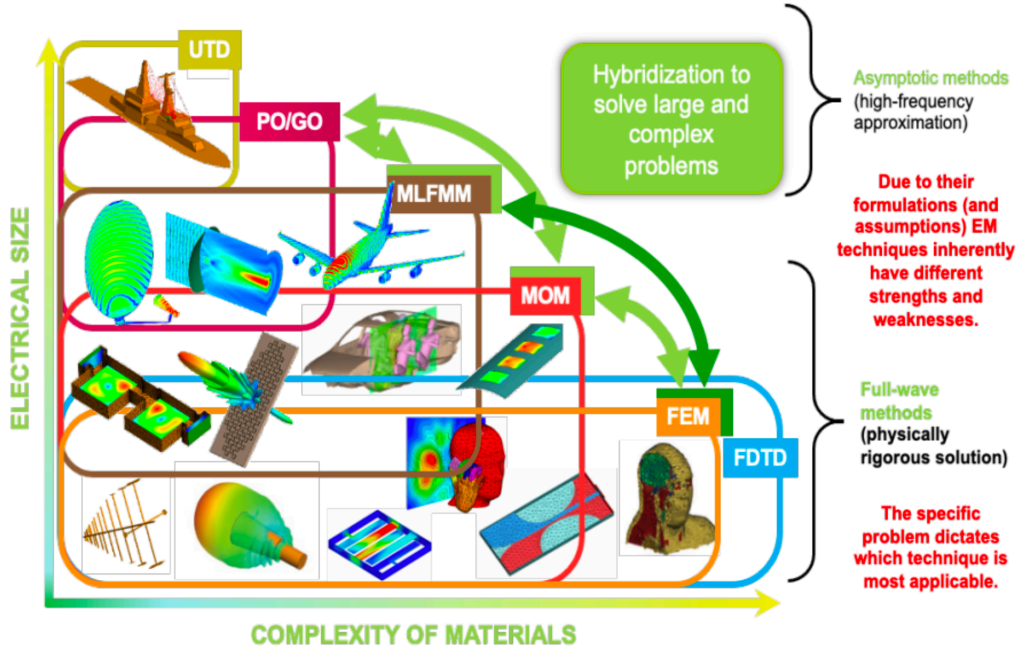


Figure 2. Numerical Methods of Winprop, with more approximation introduced when the problem's scale becomes larger.

## V. Reproducing Time-Series Channel Prediction Experiment

As part of research, we will review and validate one related work in the field of intelligent channel prediction. The experiment in the paper *LLM<sub>4</sub>CP: Adapting Large Language Models for Channel Prediction* (in short “LLM<sub>4</sub>CP”) will be reproduced, where a pre-trained LLM is applied to predict future downlink CSI based on historical uplink CSI.

### A. Modelling of Channel in LLM<sub>4</sub>CP

The base station is modeled with a dual-polarized uniform planar array (UPA), consisting of  $N_t = N_h \times N_v$  antennas. The users are set to have motion, which induces Doppler shifts and contributes to the time variation of the channel.

The received signal at UE is given by:

$$y_k = h_k^H w_k x_k + n_k, \quad (2)$$

where  $h_k$  is the downlink CSI for the  $k$ -th subcarrier,  $w_k$  is the transmit precoder,  $x_k$  is the transmitted symbol, and  $n_k$  is additive white Gaussian noise (AWGN). The precoding vector  $w_k$  is designed based on  $h_k$ , and any inaccuracy in the predicted CSI can result in suboptimal transmission rates.

## B. Formulation of Channel Prediction Problem

In LLM<sub>4</sub>CP, the goal is to predict future downlink CSI over  $L$  RBs using historical uplink CSI from  $P$  RBs. In mathematical formulation, we hope to minimize NMSE:

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\sum_{i=1}^L \|\hat{\mathbf{H}}_d^{S+i} - \mathbf{H}_d^{S+i}\|_F^2}{\sum_{i=1}^L \|\mathbf{H}_d^{S+i}\|_F^2} \right\} \quad (3)$$

Where:

- $\mathbb{E}$  represents the expectation over all the data samples.
- $\hat{\mathbf{H}}_d^{S+i}$  represents the predicted downlink CSI at time  $S + i$ .
- $\mathbf{H}_d^{S+i}$  represents the actual downlink CSI at time  $S + i$ .

The uplink and downlink CSI are assumed to share some statistical correlation due to channel reciprocity, which can be leveraged to infer future downlink conditions.

## C. Experiment Methodology

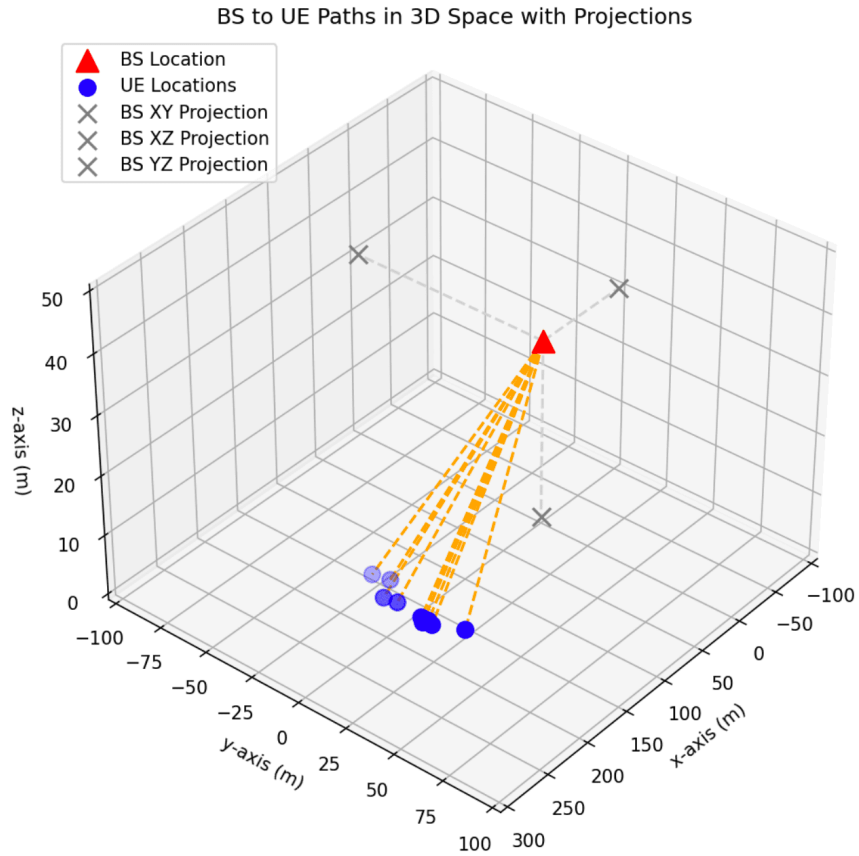
### 1. CSI Generation

The first step of the experiment is to generate channel data. To replicate real-world communication environments, we configure standardized 3GPP 38.901 Urban Macro(Uma) scenario using QuaDRiGa. Users are set to move linearly at speeds ranging from 10 km/h to 100 km/h.

Parameter	Value
Center Frequency	2.4 GHz
BS Antenna	4x4 dual-polarized array
Antenna Tilting Angle	7 degrees
User Device	Single omni-directional antenna
User Speed	10.1100 km/h in 0.1 km/h increments
Velocity Sampling Period	0.5 ms
Sample Duration	19 cycle lengths

**Table I.** Parameters Configuration

In order to have a direct comprehension of the experimental setup, the locations of UE and BS are depicted in Fig.3. The height of BS is 30m and the height of UEs is 0m.



**Figure 3.** Locations of UEs and BS in LLM<sub>4</sub>CP’s experiment setup. For consideration of simplicity, only LOS paths are shown. But remember that NLOS paths also exist and they are caused by randomly generated scatter points in the space, under QuaDriGa built-in algorithms.

## 2. CSI Data Preprocessing

The complex channel matrix  $\mathbf{H}_f$  is constructed for each transmit antenna  $j$  using the temporal length of historical CSI data.



### 3. Model Training

Both frequency domain and delay domain information are feeded into the model. Note that most layers of the pre-trained GPT-2 are frozen during training, while other parameters of the network are trainable. The number of trainable parameters is relatively small.

Model/Speed(km/h)	10	20	30	40	50	60	70	80	90	100
GPT(LLM4CP)	0.02083	0.02296	0.02485	0.02608	0.02983	0.03292	0.03849	0.04231	0.04986	0.05747
Transformer	0.03138	0.03286	0.03429	0.03232	0.04195	0.04371	0.05647	0.06176	0.07890	0.08816
CNN	0.02882	0.03075	0.03101	0.03113	0.03823	0.04109	0.05071	0.05773	0.07105	0.08244
GRU	0.05893	0.06010	0.05556	0.04230	0.06078	0.05665	0.08096	0.08246	0.10270	0.09824
LSTM	0.07976	0.07787	0.06596	0.04831	0.07003	0.06375	0.08819	0.09591	0.12286	0.11560
RNN	0.16996	0.13275	0.10381	0.07724	0.10493	0.08652	0.11641	0.12414	0.15519	0.20657
No Prediction	0.05873	0.13305	0.25106	0.41127	0.59231	0.79652	1.02281	1.25426	1.47749	1.68876
PAD	0.08252	0.12101	0.14623	0.11190	0.17296	0.18224	0.23769	0.15349	0.18944	0.18723

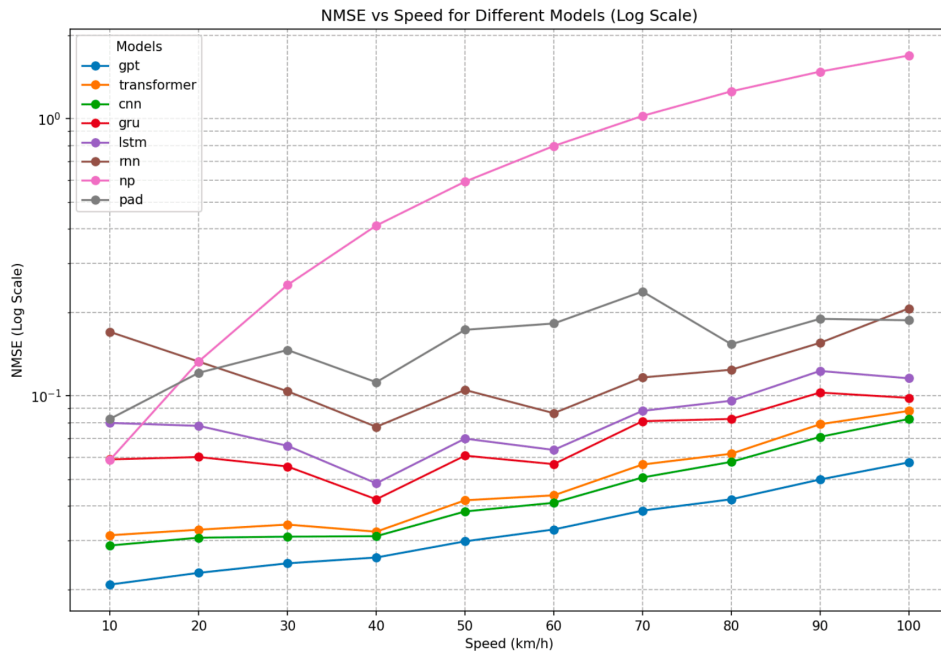
Table II. NMSE of prediction for different models across speeds, under TDD scenario

Model/Speed(km/h)	10	20	30	40	50	60	70	80	90	100
GPT(LLM4CP)	0.40161	0.34878	0.35639	0.40715	0.34129	0.38660	0.40247	0.33781	0.36739	0.40094
Transformer	0.56755	0.49624	0.47638	0.54351	0.48867	0.52255	0.51155	0.48147	0.49504	0.51693
CNN	0.97241	0.96062	0.96784	0.95742	0.95547	0.94933	0.94312	0.97714	0.96508	0.97834
GRU	0.62848	0.60466	0.60371	0.61128	0.59004	0.59892	0.59454	0.60476	0.60998	0.61508
LSTM	0.69523	0.65747	0.64816	0.65555	0.62290	0.64631	0.63110	0.64763	0.64663	0.66668
RNN	0.60296	0.55399	0.53408	0.52743	0.51272	0.50512	0.50732	0.52599	0.53564	0.58028
No Prediction	2.01695	1.92730	1.95513	1.93453	1.99573	1.88920	1.92420	1.85007	1.89675	1.82435

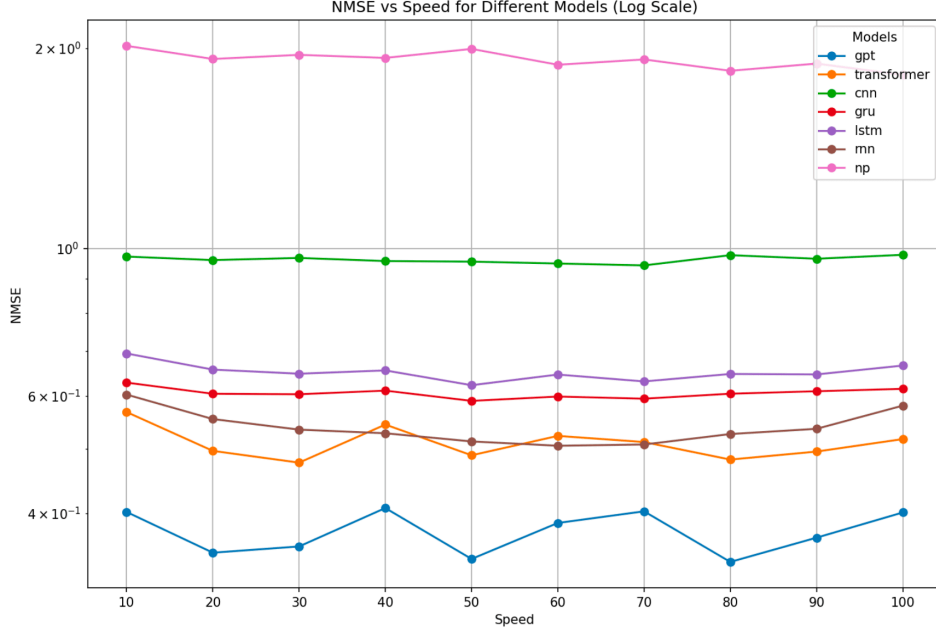
Table III. NMSE of prediction for different models across speeds, under FDD scenario

#### 4. Model Evaluation

To evaluate the performance of the proposed LLM<sub>4</sub>CP model, we compare its NMSE with several baselines across different user velocities for both TDD and FDD systems, and these include PAD, RNN, LSTM, GRU (Gate recurrent unit), CNN and traditional Transformers (Fig.4 and Fig.5).



**Figure 4.** The LLM<sub>4</sub>CP and other baselines' NMSE performance in relation to various user velocities for TDD systems



**Figure 5.** The LLM4CP and other baselines' NMSE performance in relation to various user velocities for FDD systems

## VI. QuaDRiGa Dataset for Spatial CSI Prediction

Our work will be focused on CSI spatial prediction from now on. This section serves as a heuristic startup and tools will be refined in later sections. Spatial prediction could be beneficial when the BS acquires knowledge about a user's location but communication rate between BS and this user is relatively slow.

We still adopt QuaDRiGa for data generation, with the main adjustment being that all pedestrians are stationary to focus on spatial rather than temporal prediction. Users are positioned within a 3D area (horizontal range: -100 m to 100 m, height: 0 to 3 m).

### A. Training Procedure

Model training employs a customized GPT-2-based structure, with AdamW optimizer. During training, the input to the model is each position  $\mathbf{p} \in \mathbb{R}^3$ , and the output is the predicted CSI value  $\mathbf{H}^{de}$  for the corresponding position. The predicted CSI  $\hat{\mathbf{H}}$  is hoped to be as close as possible to the ground-truth CSI  $\mathbf{H}^{gt}$ .

The training process uses a smooth L1 loss function,  $\mathcal{L}_{\text{SmoothL1}}$ . This loss function is a combination of absolute error and squared error:

$$\mathcal{L}_{\text{SmoothL1}} = \begin{cases} 0.5 \|\mathbf{H}^{de} - \mathbf{H}^{gt}\|^2, & \text{if } \|\mathbf{H}^{de} - \mathbf{H}^{gt}\|_F^2 < 1 \\ \|\mathbf{H}^{de} - \mathbf{H}^{gt}\|_F^2 - 0.5, & \text{otherwise} \end{cases} \quad (4)$$

### *B. Results of Spatial CSI Prediction Based on QuaDRiGa*

The model incorporates dual convolutional layers before GPT-2. Data from single time stamp leads to the lowest observed validation NMSE of 0.17. Then, model is trained using multiple time slices. This leads to a noticeable improvement in performance and the least NMSE is 0.035.

### *C. Limitations of Current Approach and Improvements*

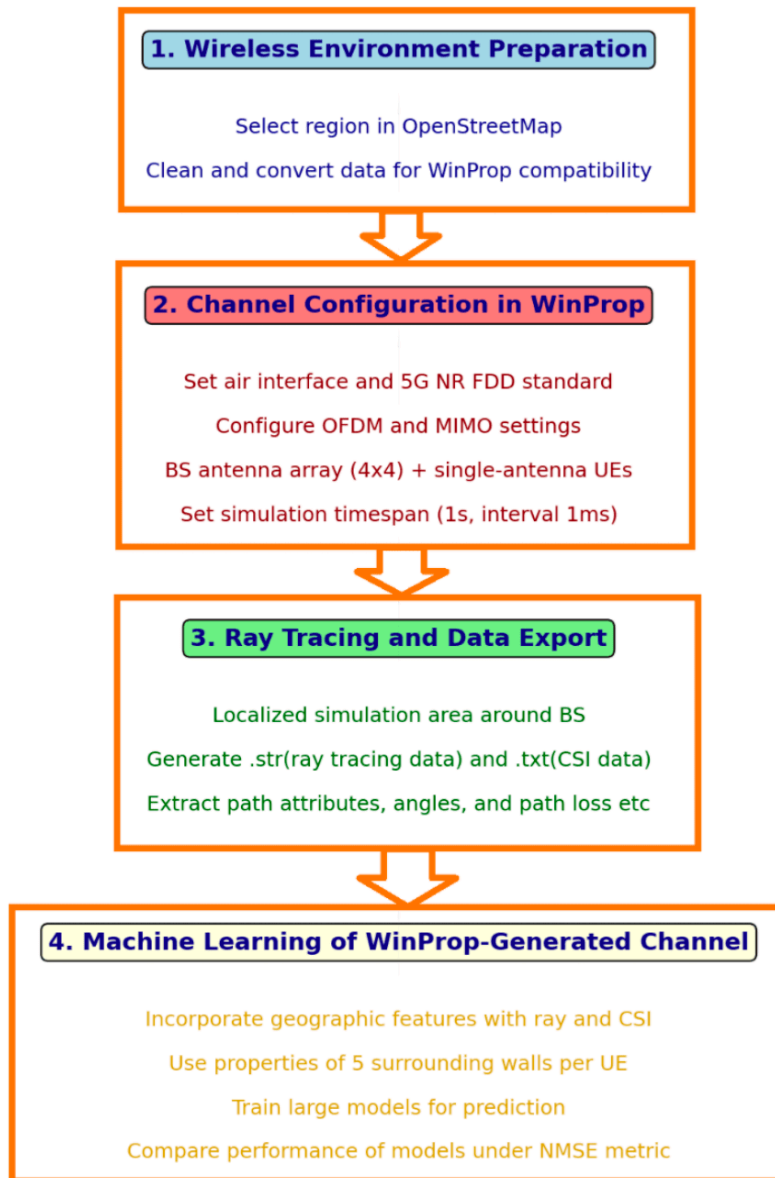
While the current work demonstrates seemly ideal results, several limitations still need to be addressed. The greatest drawback is that, current implementation uses a statistical channel. The scatter points in space are random, resulting in fluctuating channel conditions that do not correspond to physical existence. This randomness can adversely affect the stability and accuracy of predictions.

## **VII. MIMO-OFDM Channel Modeling in WinProp for Realistic Spatial CSI Prediction**

### *A. Methodology of Channel Modelling*

To achieve a realistic channel that accounts for surroundings and data frames in real-world settings, and realize high-level customization, we utilize WinProp.

The whole procedure from channel simulation to machine learning is depicted in Fig.6.

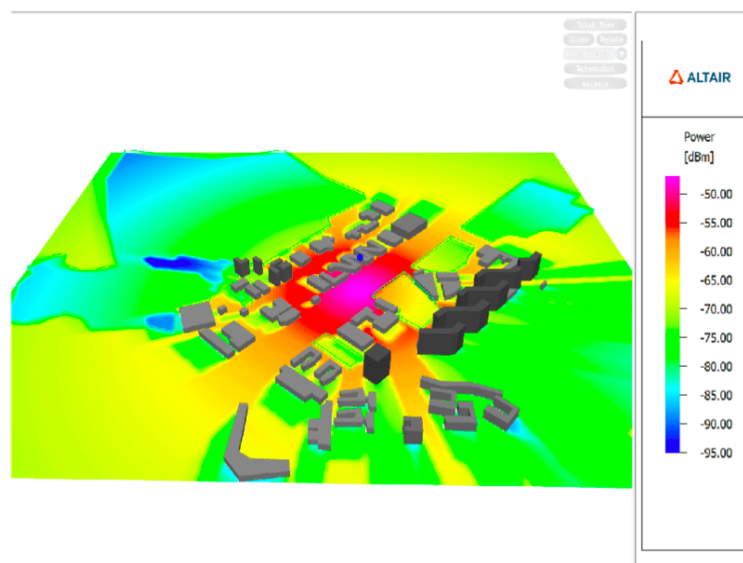


**Figure 6.** Procedure from physical environment build-up, channel generation, to model training

The first step involves preparing an environmental layout. We select the area of CUHK(SZ) in OpenStreetMap to obtain an OSM file, then clean and convert the data via QGIS for compatibility with WinProp, shown in Fig.7. The OSM file provides a detailed model of buildings, streets, and vegetation zones. Once imported, the simulation can be conducted in the area defined by the OSM file.



(a)



(b)

**Figure 7.** (a) The selected region of CUHK(SZ) as the place for ray-tracing simulation. (b) The wide range simulation result shown in power distribution, with a tiny blue dot at the center representing the BS. Notice that the simulation in the picture is coarse with distant UEs. For model training, more dense UEs are simulated in a smaller region.

Then the OFDM protocol is selected in air interface settings, and the 5G NR FDD standard description file is imported to WinProp for consistency with current wireless communication protocols. It is possible to custom-tailor parameters such as subcarrier spacing, modulation schemes, and bandwidth of the system. For industrial-grade simulators like WinProp, without configuration the air interface, the BS will only send simple modulated waves, and in this case discussing about data frame and RBs etc. is impossible.

The OFDM symbols per slot are set to 14, ensuring the simulation closely reflects high-capacity, high-frequency communication. For MIMO configurations, we implement a 4x4 BS antenna array combined with single-antenna UE.

In WinProp simulation fashion, each UE's position is uniformly distributed in predefined zones, instead of randomly generated according to a probabilistic function. The distance between UEs in the grid is 0.1m. Meanwhile the BS antenna is placed near the center.

### *B. Ray Tracing and Data Export*

Ray tracing is conducted in a rectangle inside the map. Although it is allowed to simulate the channel for only one time-slot in WinProp, it will bring a static result, which makes CSI output impossible. To obtain CSI of the channel, the dynamic simulation is carried out, with the time span of simulation at 1s and the time interval being 1ms.

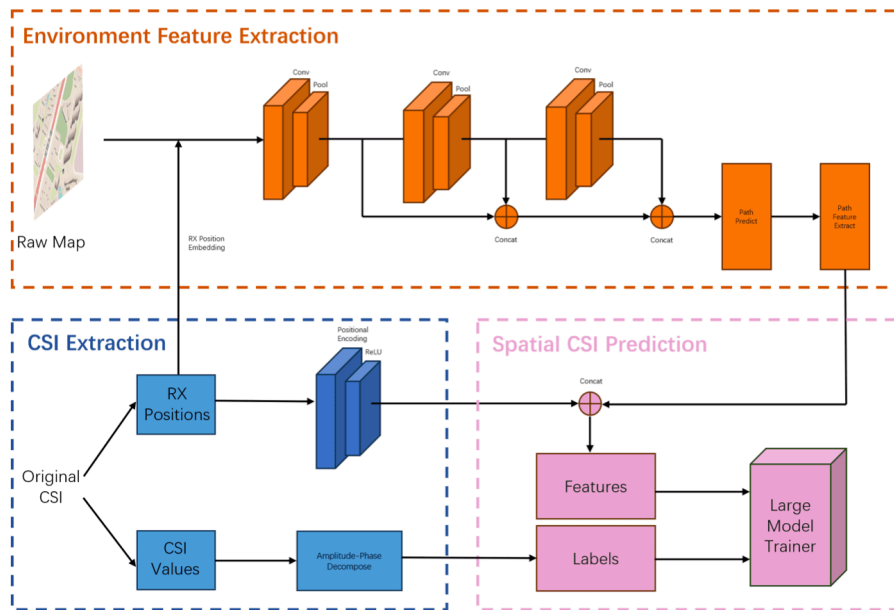
Each simulation run generates essential parameters for machine learning integration. The .str file encapsulates the attributes of each path, such as loss of path, along with the UEs interacted, but it is not measurable in practice, thus deserted. The .txt file contains CSI values for all locations and time-stamps considered in the simulation.

### *C. Machine Learning of WinProp-Generated Channel for CSI Prediction*

Once the geographic features have been extracted, they need to be merged with CSI data to train models, corresponding to the step 4 of Fig.6. To achieve this, features require suitable representation. If a model predicts well across different points or even across regions, it means that the model can infer the CSI accurately from the geographic information which can be practically measured.

In our initial approach, the walls nearby each UE is treated as a segment characterized by distance and orientation relative to the UE. With such two geometric parameters, it is believed that the surrounding of UE can be grasped by the large model, and other material-related parameters will be learned indirectly throughout the epochs. For every UE, the properties of top 5 nearest walls are extracted, which has computational efficiency but lacks of details.

To overcome insufficient representation of features, a new approach is adopted. As shown in Fig.8, the first stage involves embedding TX and RX positions into a raw geographical map and extracting environmental features. The map is then processed through a CNN module. The features are processed through multiple convolution and pooling layers to capture hierarchical representations of the environment. Intermediate features are concatenated to preserve multiscale information.



**Figure 8.** Novel architecture of machine learning targeted at informing CSI accurately from environmental knowledge, with prediction capability enhanced by multiple-feature fusing technique

Simultaneously, raw CSI data is represented in a machine-learning-friendly format. The RX positions are encoded using a positional encoding module, while the CSI values undergo an amplitude-phase decomposition to separate the magnitude and phase information.

The final stage integrates the environmental features and CSI data to predict spatial CSI. Features extracted from the environment and encoded RX locations are again concatenated as tensors before being fed into the learning model. The tensors are saved for reuse by multiple models. Notice that feeding positions and the raw map separately into the model will cause in a degenerate effect of feature learning, due to that the features are not aligned.

Multiple machine learning models are employed to predict spatial CSI by leveraging knowledge of physical environments. The models utilized include GPT-2, VAE, Transformer, Diffusion Model which incorporates a

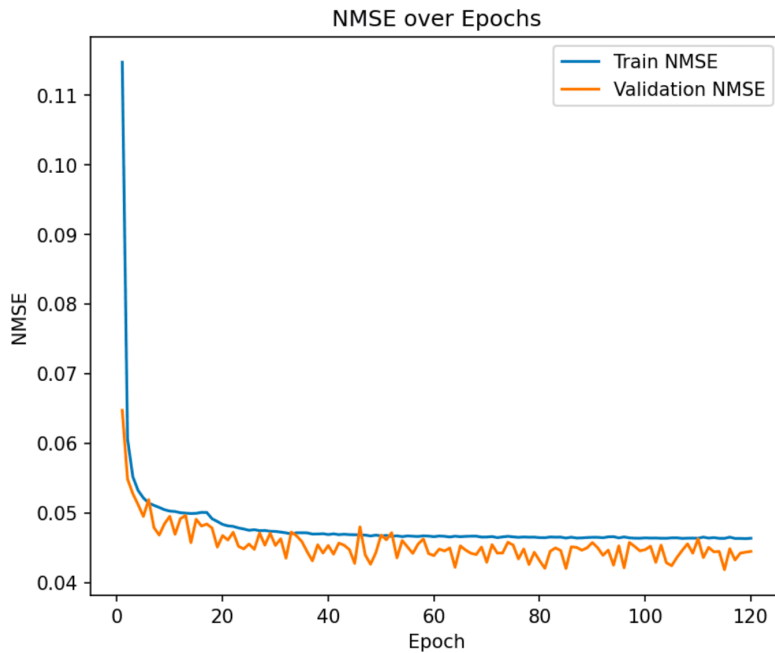


U-Net inside, and MLP as baseline.

#### D. Results and Discussion

The models are all evaluated using NMSE metric.

In this task, the VAE achieves the best performance, represented by a smooth and gradually decreasing NMSE curve in both training and validation sets, as shown in Fig.9. Meanwhile the time cost by VAE training is less than typical large models.



**Figure 9.** The learning curve of VAE, exhibiting high accuracy and low NMSE in the task of predicting spatial CSI distributed in an area of CUHK(SZ)

The Diffusion Model also shows robust performance in prediction. During early epochs, the NMSE might fluctuate. However, as training progresses, the curve converges and stabilizes. Considering no obvious progress in reducing NMSE after 30 epochs, early-stop is adopted and the lowest NMSE of validation set is 0.125 (Fig.10).

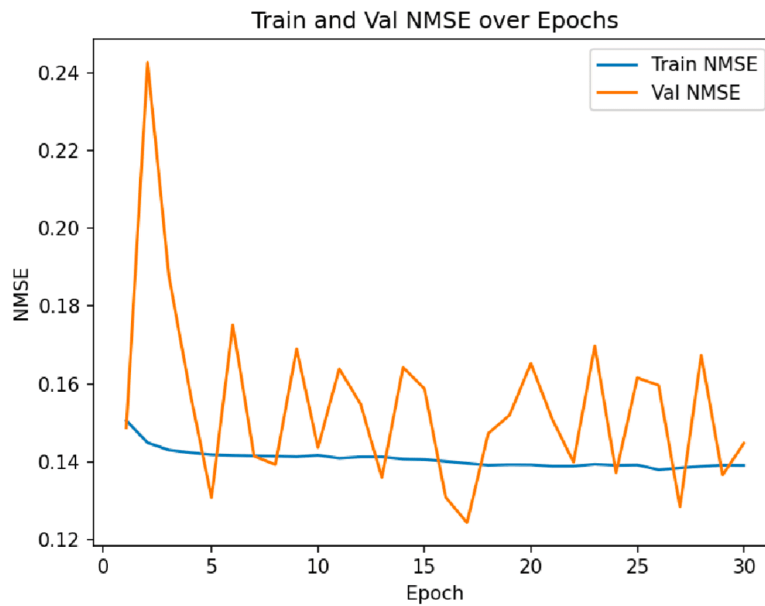


Figure 10. The learning curve of Diffusion model

GPT-2 and Transformer models, which are primarily tailored for sequence modeling, exhibit moderate initial improvement in NMSE, with the curve plateauing early. At the same time, The MLP model, being the simplest architecture, yields the highest NMSE across all epochs. The learning curves of all these three models decrease slowly and sporadically, reflecting their limited ability to capture CSI function. The validation NMSE of them are significantly higher than the training NMSE, indicating lack of fitting (Fig.11).

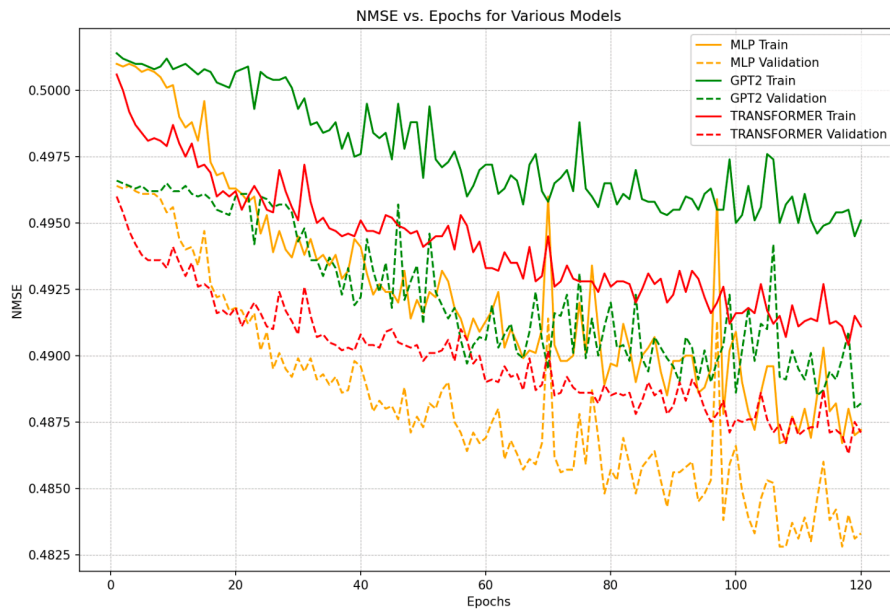


Figure 11. Learning curves of GPT-2, Transformer and MLP

Epoch	Transformer		GPT-2		MLP		VAE		Diffusion	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation	Training	Validation
1	0.5006	0.4960	0.5014	0.4966	0.5010	0.4964	0.1148	0.0647	0.1514	0.1496
10	0.4987	0.4941	0.5008	0.4962	0.5002	0.4956	0.0502	0.0495	0.1424	0.1445
20	0.4962	0.4918	0.5007	0.4960	0.4963	0.4918	0.0483	0.0467	0.1400	0.1661
30	0.4951	0.4908	0.4993	0.4943	0.4945	0.4899	0.0473	0.0453	0.1398	0.1456
60	0.4933	0.4890	0.4972	0.4906	0.4913	0.4869	0.0466	0.0439		
90	0.4919	0.4871	0.4916	0.4876	0.4902	0.4866	0.0464	0.0457		
120	0.4931	0.4863	0.4911	0.4871	0.4872	0.4833	0.0463	0.0445		

Table IV. Training and Validation NMSE Across Models and Epochs

## VIII. Conclusion

In this paper, the study starts with probabilistic modeling like QuaDRiGa, and switches to a more robust, physics-driven approach. By integrating raytracing-based channel generation technique, we successfully create a highly realistic dataset which reflects the sophisticated nature of real-world urban scenario. Combined with the dataset, the use of large models like GPT-2 and VAE provide an effective solution for predicting spatial CSI. Nevertheless, the incorporation of physical characteristics of the paths of the signal reduces the uncertainty of the function to be learned by the model, hence reliably raises the accuracy of prediction. The novel methodology is believed to be adaptive to totally different and complex urban maps.

## IX. Future Developments and Possible Applications

Future work on CSI prediction with machine learning could advance along multiple fronts. Adopting architectures such as Mistral AI or Llama models, or models capable of handling complex graphs is a promising direction. For the latter idea, applying GNN, VoxNet or PointNet could be helpful, but the challenge of the feature engineering also exists.

The application potential for CSI prediction is considerable. One of the most immediate applications lies in optimizing beamforming and antenna selection strategies in 5G and emerging 6G networks. Another promising application is in vehicular and mobile robotics communications, where vehicles or drones require robust and dynamic connection capabilities. In such cases, real-time CSI prediction enables adaptive signal transmission, leading to improved reliability in motion-intensive scenarios. In addition, the integration of these predictive models with other intelligent systems in smart cities or IoT environments offers potential for advancements in resource allocation, where predicted CSI data could inform better network resource deployment.

## References

1. <sup>a</sup>Zhang L, Sun H, Zeng Y, Hu RQ. "Spatial Channel State Information Prediction with Generative AI: Towards Holographic Communication and Digital Radio Twin." *arXiv*. 2024. [arXiv:2401.08023](https://arxiv.org/abs/2401.08023).
2. <sup>^</sup>Al-Asadi A, Al-Saedi IRK, Alwane SK, Li H, Alzubaidi L. "Enhanced MIMO CSI Estimation Using ACCPM with Limited Feedback." *Sensors*. 23 (18): 7965, Sep. 2023. doi:[10.3390/s23187965](https://doi.org/10.3390/s23187965).
3. <sup>^</sup>Gheryani M (2009). "A new approach to the design of adaptive MIMO wireless communication systems." <http://core.ac.uk/download/211515877.pdf>.
4. <sup>^</sup>Abd El-Moaty AM, Zerguine A (2018). "Sparse Channel Estimation with Gradient-Based Algorithms: A comparative Study." <https://arxiv.org/pdf/1812.04196>.

5. <sup>△</sup>Liu Z, Zhang L, Ding Z (2019). "Overcoming the Channel Estimation Barrier in Massive MIMO Communication Systems." <https://arxiv.org/pdf/1912.10573>.
6. <sup>△</sup>Kim H, Kim S, Lee H, Jang C, et al. "Massive MIMO Channel Prediction: Kalman Filtering vs. Machine Learning," 2020. <https://arxiv.org/pdf/2009.09967>.
7. <sup>△</sup>Pecorella T, Fantacci R, Picano B (2020). "Improving CSI Prediction Accuracy with Deep Echo State Networks in 5G Networks." <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7697607/>.
8. <sup>△</sup>Zhang Y, Alkhateeb A, Madadi P, Jeon J, et al. "Predicting Future CSI Feedback For Highly-Mobile Massive MIMO Systems," 2022. <https://arxiv.org/pdf/2202.02492>.
9. <sup>△</sup>Ko B, Kim H, Choi J (2022). "Massive MIMO Channel Prediction Using Machine Learning: Power of Domain Transformation." <https://arxiv.org/pdf/2208.04545>.
10. <sup>△</sup><sub>b</sub>Kim H, Choi J, Love DJ (2022). "Massive MIMO Channel Prediction Via Meta-Learning and Deep Denoising: Is a Small Dataset Enough?" <https://arxiv.org/pdf/2210.08770>.
11. <sup>△</sup>Zhou T, Zhang H, Ai B, Xue C, Liu L. "Deep-Learning-Based Spatial-Temporal Channel Prediction for Smart High-Speed Railway Communication Networks." *IEEE Transactions on Wireless Communications*. 21 (7): 5333 – 5345, July 2022. doi:[10.1109/TWC.2021.3139384](https://doi.org/10.1109/TWC.2021.3139384). <https://ieeexplore.ieee.org/document/9676455>.
12. <sup>△</sup>Du Y, Chang Liew S, Chen K, Shao Y. "The Power of Large Language Models for Wireless Communication System Development: A Case Study on FPGA Platforms," 2023. <https://arxiv.org/pdf/2307.07319>.
13. <sup>△</sup>Zhou H, Hu C, Yuan Y, Cui Y, et al. "Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities," 2024. <https://arxiv.org/pdf/2405.10825>.
14. <sup>△</sup>QF. Channel Simulators for mmWave and 5G Applications. 2017. S2CID [56454171](https://doi.org/10.1109/S2CID.2017.56454171).

## Declarations

**Funding:** No specific funding was received for this work.

**Potential competing interests:** No potential competing interests to declare.