

Special Topic: Integration of Large AI Model and 6G

# Large AI model for delay-Doppler domain channel prediction in 6G OTFS-based vehicular networks<sup>†</sup>

Jianzhe XUE<sup>1</sup>, Dongcheng YUAN<sup>1</sup>, Zhanxi MA<sup>1</sup>, Tiankai JIANG<sup>1</sup>, Yu SUN<sup>1</sup>,  
Haibo ZHOU<sup>1\*</sup> & Xuemin SHEN<sup>2</sup><sup>1</sup>*School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China*<sup>2</sup>*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo N2L3G1, Canada*

Received 14 November 2024/Revised 23 February 2025/Accepted 30 April 2025/Published online 20 June 2025

**Abstract** Channel prediction is crucial for high-mobility vehicular networks, as it enables the anticipation of future channel conditions and the proactive adjustment of communication strategies. However, achieving accurate vehicular channel prediction is challenging due to significant Doppler effects and rapid channel variations resulting from high-speed vehicle movement and complex propagation environments. In this paper, we propose a novel delay-Doppler (DD) domain channel prediction framework tailored for high-mobility vehicular networks. By transforming the channel representation into the DD domain, we obtain an intuitive, sparse, and stable depiction that closely aligns with the underlying physical propagation processes, effectively reducing the complex vehicular channel to a set of time-series parameters with enhanced predictability. Furthermore, we leverage the large artificial intelligence (AI) model to predict these DD-domain time-series parameters, capitalizing on its advanced ability to model temporal correlations. The zero-shot capability of the pre-trained large AI model facilitates accurate channel predictions without requiring task-specific training, while subsequent fine-tuning on specific vehicular channel data further improves prediction accuracy. Extensive simulation results demonstrate the effectiveness of our DD-domain channel prediction framework and the superior accuracy of the large AI model in predicting time-series channel parameters, thereby highlighting the potential of our approach for robust vehicular communication systems.

**Keywords** large AI model, channel prediction, delay-Doppler domain, OTFS, vehicular networks

**Citation** Xue J Z, Yuan D C, Ma Z X, et al. Large AI model for delay-Doppler domain channel prediction in 6G OTFS-based vehicular networks. *Sci China Inf Sci*, 2025, 68(7): 170306, <https://doi.org/10.1007/s11432-024-4426-0>

## 1 Introduction

The high mobility and complex environments of vehicular networks pose significant challenges for wireless communication, characterized by doubly-selective signal distortion and rapid fluctuations in channel path loss, which hinder hyper reliable and low latency communications (HRLLC) [1–4]. High vehicle speeds cause time-selective fading due to the Doppler effect, while multi-path propagation from obstacles leads to frequency-selective fading. Orthogonal time frequency space (OTFS) modulation has emerged as a promising sixth-generation (6G) technology for vehicular communications by operating in the delay-Doppler (DD) domain, effectively mitigating both types of fading and enhancing communication robustness in dynamic environments [5]. However, rapid fluctuations in channel path loss can reduce the signal-to-noise ratio (SNR) below the required thresholds, compromising reliability. Therefore, channel prediction becomes essential, allowing networks to anticipate future conditions and proactively adjust communication parameters to maintain stable and reliable connections [6–10].

Channel prediction in vehicular networks is exceptionally challenging due to the combination of high vehicle mobility and complex propagation environments [11]. The rapid movement of vehicles induces significant Doppler effects, resulting in a severe Doppler shift [12]. This Doppler shift leads to a very short channel coherence time, which can be less than 1 ms in vehicular scenarios. Additionally, multi-path propagation causes transmitted signals to arrive at the receiver via multiple paths due to reflections from buildings, vehicles, and other obstacles. This results in multiple copies of the signal arriving at slightly different times and with different phases. These copies interfere constructively or destructively,

\* Corresponding author (email: haibozhou@nju.edu.cn)

<sup>†</sup>Special Topic: The 27th Annual Meeting of the China Association for Science and Technology

leading to rapid fluctuations in signal amplitude known as fast fading and deep fading. Consequently, the time-frequency (TF) domain channels exhibit doubly-selective fading owing to the combined effects of multi-path propagation and Doppler shifts. Moreover, characterizing each time and frequency point in the TF domain results in an extensive number of channel parameters. This complex and swift variation of the TF domain vehicular channel makes accurate channel prediction exceedingly difficult [13].

To this end, we leverage the DD domain based on OTFS modulation to provide an intuitive, sparse, and stable channel representation, which is particularly beneficial for channel prediction in vehicular networks [14]. Firstly, the DD domain channel representation offers an intuitive depiction of the wireless channel by characterizing it through channel response coefficients, propagation delays, and Doppler shifts of each sub-path separately. These parameters directly reflect the underlying physical phenomena of signal propagation, facilitating a clearer understanding and more accurate modeling of signal behavior as vehicles move [15]. Secondly, the channel representation in the DD domain is inherently sparse, as only a few significant sub-paths contribute meaningfully to the received signal [16]. This sparsity reduces the number of parameters that need to be estimated and predicted, in contrast to the dense grid of coefficients typical of TF domain representations. Thirdly, the DD domain offers a stable representation over time, even under rapid vehicular movement [17]. While TF domain coefficients may fluctuate rapidly due to high mobility and dynamic environments, the DD domain parameters tend to vary more gradually. This relative stability enhances the predictability of the channel. Therefore, with the DD domain channel representation, the complex vehicular channel variations can be transformed into variations of a set of time series parameters with good predictability in the DD domain. Consequently, channel prediction in vehicular networks becomes the prediction of these time-series DD domain parameters.

To accurately predict time series parameters in the DD domain, we employ the large artificial intelligence (AI) model, specifically transformer-based neural networks with billions of parameters, for channel prediction in vehicular networks. The large AI model offers significant advantages over small models, such as long short-term memory (LSTM) networks, by effectively capturing the complex temporal correlations inherent in time series data and enabling more precise modeling of intricate channel variation patterns [18–21]. A notable benefit of the large AI model is its zero-shot prediction capability, allowing it to be cost-effectively and flexibly deployed without the need for specialized training on vehicular channel data [22]. Pre-trained on vast amounts of general time series data, these models develop a comprehensive understanding of temporal dynamics that can be applied to unseen channel conditions, thereby allowing them to generalize effectively and make accurate predictions even in new environments. Furthermore, the large AI model can be fine-tuned with actual vehicular channel data to further enhance their prediction accuracy, adapting their learned representations to the specific characteristics of channel data [23]. This makes the large AI model particularly powerful for addressing the challenges of channel prediction in high-mobility vehicular networks, outperforming smaller models in both accuracy and robustness.

In this paper, we propose a DD domain channel prediction framework for high-mobility vehicular networks, and employ the large AI model to predict the time-series parameters inherent in the DD domain channel representation. Specifically, in OTFS-based vehicular networks, channel estimation is naturally performed in the DD domain, providing a representation where channel parameters vary continuously and stably over time with excellent predictability. We transform the vehicular channel prediction problem into a time series prediction task for DD domain parameters, including the response coefficient, propagation delay, and Doppler shift of each sub-path. This significantly reduces the complexity associated with doubly-selective vehicular channel prediction. Additionally, by employing the large AI model, we achieve accurate and robust predictions of DD domain channel parameters. The zero-shot capability of the large AI model allows for low-cost and flexible channel prediction deployment without prior specialized training, while fine-tuning with specific vehicular channel data further enhances prediction accuracy. The combination of the stability of DD domain channel representations and the advanced predictive capability of the large-scale AI model enables our approach to effectively address the challenges posed by high mobility and complex propagation environments of vehicular networks. Besides, we demonstrate the effectiveness of our channel prediction approach through extensive simulations. Three main contributions of this paper are summarized as follows.

- We utilize OTFS modulation in vehicular networks to effectively counteract the signal distortions caused by doubly-selective channels. Meanwhile, OTFS modulation introduces the DD domain, offering a stable, sparse, and intuitive representation of the vehicular channel that aligns closely with the physical propagation environment.
- We propose a DD domain channel prediction framework tailored for vehicular networks character-

ized by high mobility and complex environments. Leveraging the DD domain channel representation, channel prediction is transformed into a time series prediction task of DD domain parameters, which vary continuously and predictably over time.

- We leverage the large AI model to accurately and robustly predict time-series DD domain channel parameters, capitalizing on its advanced capacity to model temporal correlations. Its zero-shot capability facilitates low-cost deployment of channel prediction, while subsequent fine-tuning further enhances prediction accuracy. Compared to traditional DL approaches, the large AI model offers a flexible, accurate, and cost-effective solution for vehicular channel prediction.

The remainder of this paper is organized as follows. Section 2 reviews related work on channel prediction. Sections 3 and 4 introduce the vehicular channel model and the DD domain channel representation, respectively, and formulate the DD domain channel prediction problem. Section 5 presents the large AI model used for channel prediction. Experimental results are given in Section 6, and conclusion is drawn in Section 7.

## 2 Related work

Channel prediction plays a crucial role in enhancing vehicular network performance, with numerous studies focusing on this area. Unlike traditional static channel prediction, vehicular networks require methods that address the rapid and complex changes in vehicular channels caused by vehicle movements. The channel prediction problem in high-mobility massive multiple input multiple output (MIMO) system is investigated in angle-delay domain, and a spatio-temporal autoregressive model-based unsupervised-learning method is proposed [24]. To cope with rapid channel variation, based on the estimated mobility parameters, a channel prediction approach incorporating the mobility of both scatterers and transceivers is proposed [25]. A model-driven high-accuracy and low-complexity channel state information (CSI) prediction scheme is presented by pre-estimating the parameter range with the short-time Fourier transform and obtaining the initial values of the parameters with the polynomial Fourier transform [26].

Deep learning (DL) models with their powerful spatial-temporal data mining capabilities have shown great promise in improving channel prediction accuracy by capturing spatial and temporal correlations [27–32]. A DL channel prediction approach using LSTM networks is proposed to enhance CSI forecasting in edge computing networks for intelligent connected vehicles, which outperforms conventional models like autoregressive integrated moving average model through extensive simulations [33]. Gruformer achieves accurate channel prediction while making training more stable by integrating the gate recurrent unit (GRU) module into the transformer architecture [34]. A transformer-based parallel channel prediction scheme is proposed to address the channel aging problem in mobile environments, enabling accurate predictions across multiple frames simultaneously and mitigating error propagation, thus achieving negligible performance loss in practical communication systems under mobile environments [35].

Inspired by the success of generative pre-trained transformer (GPT) in natural language processing, some researchers have begun exploring whether large AI models could enhance channel prediction performance, given that time series channel data can be similarly represented as sequential data [36–42]. LLM4CP is a pre-trained large language model (LLM) empowered method for predicting future downlink CSI based on historical uplink data, utilizing tailored modules for cross-modality knowledge transfer, achieving high accuracy with low training and inference costs in massive MIMO systems [43]. CSI-GPT integrates a swin transformer-based channel acquisition network with a variational auto-encoder-based channel sample generator and federated-tuning to efficiently acquire downlink CSI in massive MIMO systems [44]. CSI-LLM is an LLM-based channel prediction approach, which models the historical variable-step sequences, utilizing the next-token generation ability of LLM to predict the CSI of the next step [45]. However, these studies focus primarily on channel prediction in static environments. To the best of our knowledge, this work represents an early effort to apply the large AI model for channel prediction in high-mobility environments, specifically in scenarios like vehicular networks.

## 3 System model

In vehicular networks, multi-path propagation leads to frequency-selective fading as signals arrive via multiple sub-paths with different delays. While channels in static scenarios that only contain the multi-path effect can be approximated as the linear time-invariant system, vehicle mobility introduces Doppler

shifts, making the channel time-variant. Consequently, vehicular channels exhibit both frequency-selective and time-selective fading, resulting in rapid channel fluctuations. Accurately modeling vehicular network channels necessitates capturing both multi-path and Doppler effects arising from continuous vehicle movement and complex propagation environments. High spatial-temporal continuity is essential to represent the rapid channel variations caused by constantly changing vehicle positions and speeds. Statistical channel models can efficiently consider both multi-path and Doppler effects with low computation cost, but they lack the spatial-temporal continuity required for high-mobility scenarios. Ray-tracing channel models provide detailed channel simulations, but they are computationally demanding and challenging in achieving high spatial-temporal resolution.

To accurately model both multi-path and Doppler effects with high spatial-temporal continuity and resolution, we introduce the quasi-deterministic channel model. This channel model generates the propagation environment through a statistical distribution and then allows transmitters and receivers to interact with the environment in a deterministic way. In this section, this channel model is introduced in terms of large-scale fading and small-scale fading.

### 3.1 Large-scale fading

Large-scale fading describes the average signal attenuation over long distances and is characterized by parameters such as the Rician K-factor, shadow fading, and angular domain parameters. Fluctuations in these parameters result from environmental conditions and changes in the relative positions of the transmitter and receiver. Achieving smooth transitions in large-scale parameters and good spatial consistency is crucial for accurate large-scale fading modeling in time-varying channels.

Specifically, the Rician K-factor quantifies the ratio of direct sub-path gain to scattered sub-path gain, reflecting the relative strengths of line-of-sight (LOS) and non-line-of-sight (NLOS) components. The spatially correlated Rician K-factor  $\tilde{R}_K$  is calculated as

$$\tilde{R}_K(\mathbf{p}) = R_{K\mu} + R_{K\sigma} \cdot \tilde{X}^{R_K}(\mathbf{p}), \quad (1)$$

where  $\tilde{X}^{R_K}$  is the spatially correlated random variable,  $\mathbf{p}$  is position vector,  $R_{K\mu}$  is the expectation of reference value, and  $R_{K\sigma}$  is the standard deviation of reference value. The  $R_{K\mu}$  is calculated as

$$R_{K\mu} = R_K^0 + R_K^f \log_{10}(f_G) + R_K^d \log_{10}(d_{2D}) + R_K^h \log_{10}(h_B) + R_K^\alpha \alpha_R, \quad (2)$$

where  $R_K^f$ ,  $R_K^d$ ,  $R_K^h$ , and  $R_K^\alpha$  represent the impacts of dependent variables in scenarios, including carrier frequency  $f_G$  (Hz), 2D horizontal distance  $d_{2D}$  (m), base station height  $h_B$  (m), and elevation angle seen from the RX  $\alpha_R$  (rad). And  $R_K^0$  is the baseline value of the parameter, providing a foundational level for Rician K-factor under specific environmental conditions. The impact of each parameter varies according to statistical distributions specific to different environments. The spatially correlated random variable  $\tilde{X}^{R_K}$  is calculated as

$$\tilde{X}^{R_K}(\mathbf{p}) = \frac{\tilde{X}^{R_K}(\mathbf{p}_{tx}) + \tilde{X}^{R_K}(\mathbf{p}_{rx})}{2 \cdot \sqrt{\rho(d)+1}}, \quad (3)$$

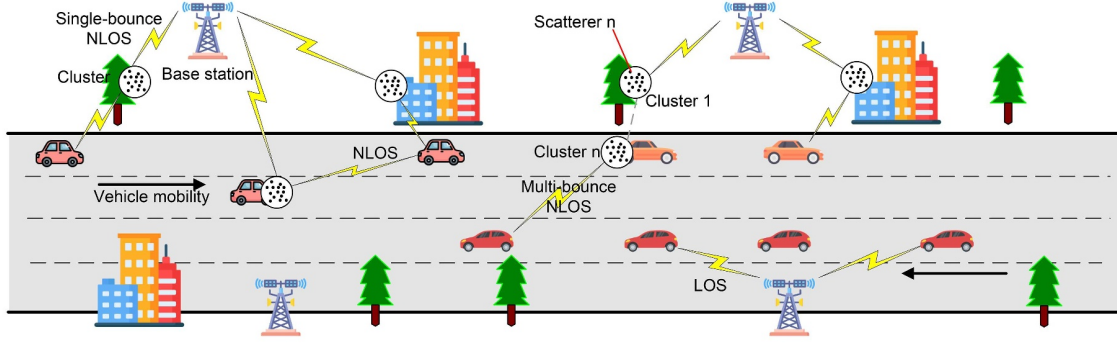
where  $\mathbf{p}_{tx}$  and  $\mathbf{p}_{rx}$  mean the position vector of transmitter and receiver.  $\rho(d)$  is the autocorrelation function defined as

$$\rho(d) = \begin{cases} \exp(-\frac{d^2}{d_\lambda^2}), & \text{for } d < d_\lambda, \\ \exp(-\frac{d}{d_\lambda}), & \text{for } d \geq d_\lambda, \end{cases} \quad (4)$$

where  $d_\lambda$  is the decorrelation distance threshold. When the real distances  $d$  is smaller than  $d_\lambda$ , the correlation decays following a Gaussian distribution, otherwise, it decays according to an exponential distribution. This procedure can be similarly applied to other large-scale parameters, such as shadow fading, generating reference values, and spatially correlated variables in the same manner.

Additionally, all large-scale parameters are cross-correlated, as environmental factors often influence multiple parameters simultaneously. The inter-parameter correlation is determined by a covariance matrix, denoted by a positive definite matrix  $\mathbf{R}$  as

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{R_K, F_S} \\ \rho_{F_S, R_K} & 1 \end{pmatrix}, \quad (5)$$



**Figure 1** (Color online) Quasi deterministic channel modeling for high-mobility vehicular networks.

where  $F_S$  is the shadow fading, and  $\rho_{F_S, R_K}$  is the cross-correlation between the shadow fading  $F_S$  and the Rician K-factor  $R_K$ . This process ensures the authenticity and spatial consistency of large-scale fading in dynamic vehicular network environments.

Besides, large-scale general path loss  $L_P$  quantifies the reduction in signal gain due to distance-based attenuation. Unlike other large-scale parameters, it lacks spatial correlation and uniformly affects all sub-paths, independent of environmental changes. It provides a baseline reference for average signal attenuation and is calculated as

$$L_P = L_P^0 + L_P^f \log_{10}(f_{\text{GHz}}) + L_P^d \log_{10}(d_{2D}) + L_P^h \log_{10}(h_B) + L_P^\alpha \alpha_R, \quad (6)$$

where  $L_P^f$ ,  $L_P^d$ ,  $L_P^h$ , and  $L_P^\alpha$  represent the impacts of frequency, horizontal distance, base station height, and elevation angle, respectively. Also,  $L_P^0$  is the baseline path loss, providing a foundational level for average signal attenuation.

### 3.2 Small-scale fading

Small-scale fading is modeled using geometric models that capture channel variations at the wavelength level, including rapid time-varying fading caused by multi-path effects and Doppler shifts. In these models, large obstacles like buildings and trees are represented as clusters, which are groups of scatterers that affect the signal path, leading to phenomena such as path diffraction and reflection. As shown in Figure 1, the propagation signal may pass through zero or multiple clusters. For multiple sub-paths, each sub-path  $i$  is characterized by variables  $[\tau_i, P_i, \psi_i]$ , where  $\tau_i$  is the propagation delay,  $P_i$  is the gain, and  $\psi_i$  is the phase. The channel model updates the signal propagation path in real time based on the movements of the transmitter, receiver, and clusters.

To accurately represent signal strength and phase variations due to antenna directionality and polarization in dynamic propagation environments, it is essential to consider the transmitter antenna's radiation pattern and polarization characteristics. The antenna composite gain  $g^A$  encapsulates the overall effect of these antenna properties. Specifically, the amplitude of  $g^A$  reflects the relative antenna gain for a given direction and polarization, while its phase component describes the antenna's phase shift. Combining  $g^A$  with propagation model, the signal composite gain  $g_i^S$  and the phase  $\psi_i$  of each sub-path are obtained as

$$g_i^S = \frac{1}{\sqrt{\sum_{n=1}^{n_s} P_{in}}} \sum_{n=1}^{n_s} g^A \cdot e^{j\psi_{in}} \cdot \sqrt{P_{in}}, \quad (7a)$$

$$\psi_i = \frac{2\pi}{\lambda} \cdot f(d_i, \theta_i, \phi_i), \quad (7b)$$

where the phase is influenced by the combined effects of the path length  $d_i$ , the elevation angle  $\theta_i$ , and the azimuth angle  $\phi_i$ , with results evaluated at the wavelength  $\lambda$  level. For the LOS path,  $d_i$  is the direct path length, while for the NLOS path,  $d_i$  is the total length after considering scattering.

### 3.3 Quasi deterministic channel parameters

To accurately represent both micro and macro channel characteristics, large-scale fading parameters and small-scale fading parameters need to be integrated, converting into channel coefficients and delay

matrices that encompass multi-path effects and Doppler shifts. The total propagation delay  $\tau_i$  of each sub-path is calculated based on the geometric model as

$$\tau_i = \frac{1}{c} \cdot \sum_{n=1}^{n_s} d_i^n, \quad (8)$$

where  $n_s$  is the number of scatterers,  $d_i^n$  is the distance between scatterers, and  $c$  is the speed of light. The channel response coefficient  $g_i$  of each sub-path is obtained as

$$g_i = L_P \cdot L_I \cdot g_i^S, \quad (9)$$

where  $L_I$  is the combined influence of large-scale parameters. As vehicles move, the path length and angle of arrival of each sub-path vary, causing changes in  $\psi_i$  and  $\tau_i$ . The rapid change of angle of arrival caused by vehicle movement manifests as the Doppler frequency shift. Therefore, the quasi deterministic channel model can consider both multi-path and Doppler effects with high spatial-temporal continuity and resolution for high-mobility vehicular networks.

## 4 OTFS and delay-Doppler domain channel

In this section, we present the mathematical formulations of OTFS modulation, along with the corresponding DD domain channel model. Then, based on the DD domain channel representation, we establish the DD domain channel prediction framework and formulate the channel prediction problem as the time series DD domain parameters prediction task.

### 4.1 OTFS modulation

OTFS modulation is an advanced scheme that maps transmitted symbols in the DD domain, allowing each symbol to experience a nearly constant channel gain even in environments with high Doppler shifts. By spreading each DD-domain symbol across the entire TF domain, OTFS fully exploits the diversity inherent in vehicular channels. Consequently, this modulation technique transforms time-varying multi-path vehicular channels into sparse and stable DD-domain representations, thereby enhancing the robustness and reliability of vehicular communications in dynamic wireless environments.

Let  $M$  denote the number of delay bins (subcarriers),  $N$  denote the number of Doppler bins (time slots),  $T$  denote the symbol duration, and  $\Delta f = 1/T$  denote the subcarrier spacing. Firstly, a set of quadrature amplitude modulation (QAM) symbols  $\mathbf{X}_{DD}[l, k] \in \mathbb{C}^{M \times N}$ , where  $l = 1, \dots, M$  and  $k = 1, \dots, N$ , is arranged on a two-dimensional DD domain grid. This DD domain symbol matrix undergoes a transformation into the TF domain,  $\mathbf{X}_{TF}[m, n] \in \mathbb{C}^{M \times N}$ , where  $m = 1, \dots, M$  and  $n = 1, \dots, N$ , by utilizing the inverse symplectic finite Fourier transform (ISFFT) as

$$\mathbf{X}_{TF}[m, n] = \frac{1}{\sqrt{MN}} \sum_{k=1}^N \sum_{l=1}^M \mathbf{X}_{DD}[l, k] e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})}. \quad (10)$$

Subsequently, the Heisenberg transform is applied to  $\mathbf{X}_{TF}$  based on the pulse waveform  $g_{tx}(t)$  to generate the time domain transmit signal  $s(t)$  as

$$s(t) = \sum_{n=1}^N \sum_{m=1}^M \mathbf{X}_{TF}[m, n] g_{tx}(t - nT) e^{j2\pi m \Delta f (t - nT)}. \quad (11)$$

After passing the vehicular channel, the time-domain received signal  $r(t)$  at the receiver is converted back to the TF domain matrix  $\mathbf{Y}_{TF}[m, n] \in \mathbb{C}^{M \times N}$  through the Wigner transform as

$$\mathbf{Y}_{TF}[m, n] = \int g_{rx}^*(t - nT) r(t) e^{-j2\pi m \Delta f (t - nT)} dt, \quad (12)$$

where  $g_{rx}(t)$  is the matched pulse waveform and superscript  $*$  denotes the conjugation. The received DD domain symbol matrix,  $\mathbf{Y}_{DD}[l, k] \in \mathbb{C}^{M \times N}$ , is then derived using the symplectic finite Fourier transform (SFFT) as

$$\mathbf{Y}_{DD}[l, k] = \frac{1}{\sqrt{MN}} \sum_{n=1}^N \sum_{m=1}^M \mathbf{Y}_{TF}[m, n] e^{-j2\pi(\frac{nk}{N} - \frac{ml}{M})}. \quad (13)$$

## 4.2 Delay-Doppler domain channel

Channel prediction in vehicular networks is particularly challenging due to the high mobility of vehicles and the complexity of the propagation environment, which results in rapid and intricate variations in the wireless channel. In the TF domain, each time and frequency point must be characterized, leading to an extensive number of channel parameters. Moreover, TF domain channels exhibit rapid fluctuations owing to the combined effects of multi-path propagation and Doppler shifts. This vast, rapidly varying parameter set significantly diminishes the predictability of the channel, thereby reducing prediction accuracy. Transforming the channel representation to the DD domain addresses these challenges effectively by providing an intuitive, sparse, and stable channel representation.

With high time resolution channel response coefficient  $g_i$  and propagation delay  $\tau_i$  provided by the quasi deterministic channel model, we can calculate each sub-paths Doppler shift from the phase changes in  $g(i, t)$  across neighboring snapshots. Specifically, the Doppler shift  $\nu(i)$  can be obtained as

$$\nu(i) = \frac{\Delta \arg(g(i, t_0))}{2\pi\Delta t}, \quad (14)$$

where  $\Delta \arg(g(i, t_0))$  is the phase difference between two neighboring snapshots at time  $t_0$ , and  $\Delta t$  is the time interval between them. Assuming that within the short duration of a data frame, the relative motion between transmitter and receiver is approximately constant, the delay and Doppler shift of each sub-path remain constant. Then, the time-continuous gain of each sub-path  $h_i(\tau_i, t)$  can be approximately expressed as

$$h_i(\tau_i, t) = g(i, t_0) \cdot e^{j2\pi\nu(i)(t-\tau_i)}, \quad (15)$$

where  $g(i, t_0)$  is the initial channel coefficient at time  $t_0$ . This expression captures the Doppler-induced phase variation over time. In general, the channel with multi-path fading and Doppler shift can be modeled as a linear time-invariant system, so the input-output relationship of this channel can be obtained as

$$r(t) = \sum_{i=1}^{N_P} h_i(\tau_i, t) s(t - \tau_i) = \sum_{i=1}^{N_P} g(i, t_0) \cdot e^{j2\pi\nu(i)(t-\tau_i)} s(t - \tau_i), \quad (16)$$

where  $r(t)$  is the received signal,  $s(t)$  is the transmitted signal, and  $N_P$  is the total number of sub-paths.

Traditional orthogonal frequency division multiplexing (OFDM) works in the TF domain and therefore uses the TF domain channel representation. By taking a Fourier transform, the TF domain channel impulse response can be obtained as

$$H(f, t) = \sum_{i=1}^{N_P} g(i, t_0) e^{-j2\pi\nu(i)\tau_i} e^{-j2\pi(f\tau_i - \nu(i)t)}, \quad (17)$$

where  $f$  is frequency. For the case of the static channel where Doppler shift  $\nu(i)$  equals 0,  $H(f, t)$  is reduced to time-independent frequency response  $H(f)$ , which is known as the frequency-fading channel. In this case, the channel prediction only needs to predict channel response at different frequencies. However, the high mobility of vehicles makes the channel time-variant, introducing the doubly-selective fading and making every element in  $H(f, t)$  becomes different. This significantly increases the number of parameters that need to be processed for channel prediction. Therefore, it will be very complicated to make predictions for the high-mobility vehicular channel in the TF domain.

In the OTFS system, the delay-Doppler channel response is expressed in terms of a group of sub-paths with different gain, delay, and Doppler as

$$h(\tau, \nu) = g(i, t_0) \cdot e^{-j2\pi\nu(i)\tau_i} \delta(\tau - \tau_i) \delta(\nu - \nu_i), \quad (18)$$

where  $\delta(\cdot)$  is the Dirac delta function. This formulation offers a sparse and quasi-time-invariant channel representation in the DD domain channel, closely reflecting the physical properties of the environment. This is a sparse wireless channel representation in the DD domain for the limited number of sub-paths. Therefore, the received signal  $r(t)$  can be expressed as

$$r(t) = \iint h(\tau, \nu) e^{j2\pi\nu t} s(t - \tau) d\nu d\tau. \quad (19)$$

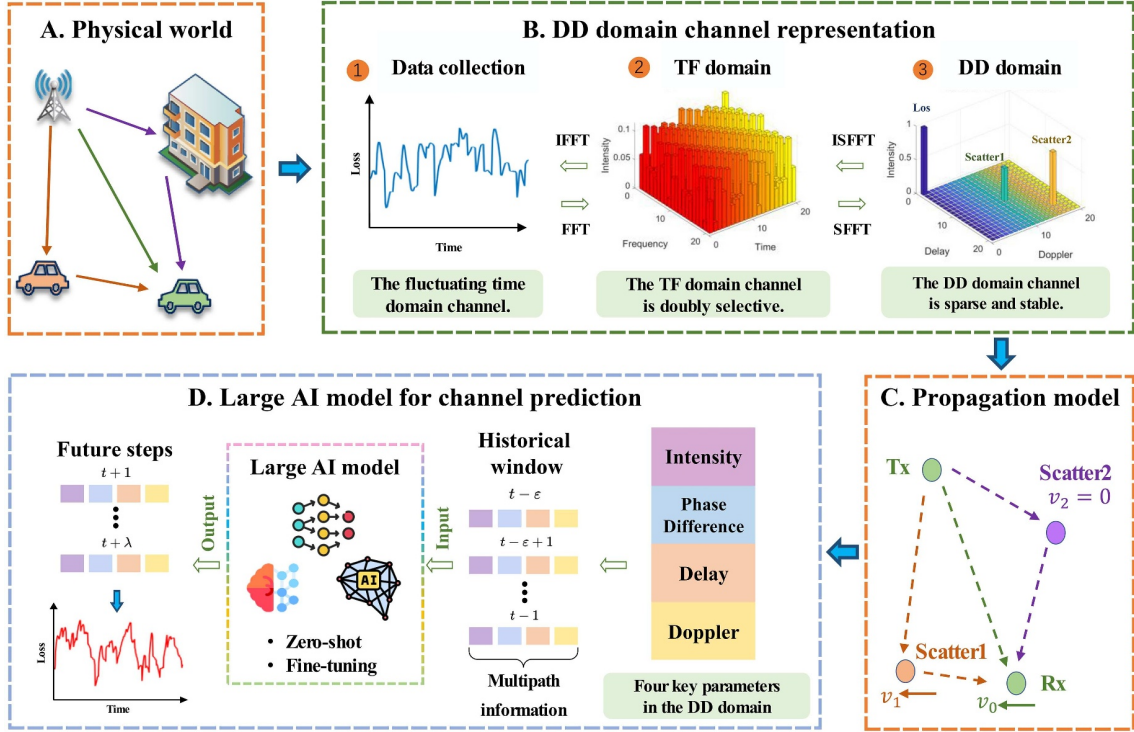


Figure 2 (Color online) Delay-Doppler domain channel prediction with the large AI model.

Finally, we obtain the DD domain channel representation for the high-mobility scenario, in which the wireless channel at a given time snapshot  $t_0$  can be completely represented by parameters including the channel response coefficient  $g(i, t_0)$ , delay spread  $\tau_i$ , and Doppler shift  $\nu_i$  of each sub-path. The DD domain channel representation with these parameters is defined as

$$H_{t_0}^{\text{DD}} = \{g(i, t_0), \tau_i, \nu_i\}, i = 1, \dots, N_{\text{P}}. \quad (20)$$

This is an intuitive, sparse, and stable channel representation, which is particularly beneficial for channel prediction in vehicular networks.

### 4.3 Channel prediction problem formulation

With the DD domain channel representation, the problem of vehicular network channel prediction becomes a task of predicting the set of the DD domain channel parameters. We leverage the neural network for channel prediction for its powerful time series prediction capability. However, the inputs to neural networks usually require real numbers, but the channel response coefficients are complex, so it is necessary to represent the channel response coefficients as real numbers. Since the intensity of the channel response coefficients varies slowly with time and its phase varies rapidly and periodically with time at a relatively constant rate due to the Doppler effect, we express the channel response coefficients based on the form of polar co-ordinates by expressing them as intensity  $|g(i, t_0)|$  and phase differences  $\Delta\varphi_{i, t_0}$ . To satisfy the real number requirement of the neural network, the DD domain channel parameters are reformulated as

$$H_{t_0}^{\text{DD}} = \{|g(i, t_0)|, \Delta\varphi_{i, t_0}, \tau_i, \nu_i\}, \text{ for } i = 1, \dots, N_{\text{P}}. \quad (21)$$

As shown in Figure 2, the task of the DD domain channel prediction framework can be formalized as follows. Given the historical and current DD domain channel parameters of  $\varepsilon$  time snapshots, represented as  $\{H_{t-\varepsilon+1}^{\text{DD}}, H_{t-\varepsilon+2}^{\text{DD}}, \dots, H_t^{\text{DD}}\}$ , the goal of channel prediction is to accurately forecast the time series DD domain channel parameters for the subsequent  $\lambda$  steps, defined as  $\{H_{t+1}^{\text{DD}}, H_{t+2}^{\text{DD}}, \dots, H_{t+\lambda}^{\text{DD}}\}$ . Therefore, the core of the DD domain vehicular channel prediction task is to build the mapping function  $\mathcal{F}(\cdot)$ ,

$$\{H_{t+1}^{\text{DD}}, H_{t+2}^{\text{DD}}, \dots, H_{t+\lambda}^{\text{DD}}\} = \mathcal{F}(\{H_{t-\varepsilon}^{\text{DD}}, H_{t-\varepsilon+1}^{\text{DD}}, \dots, H_{t-1}^{\text{DD}}\}), \quad (22)$$

which aims to leverage temporal correlations within historical time series DD domain channel parameters to anticipate future channel conditions.

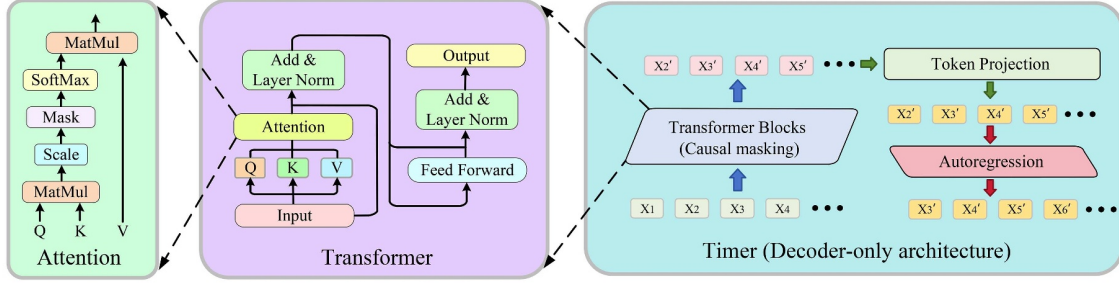


Figure 3 (Color online) Large AI model with Transformer as the backbone.

## 5 Large AI model for channel prediction

In recent years, the large AI model has become a powerful tool for processing and forecasting time series by leveraging large amounts of data and sophisticated architecture. In this section, the transformer architecture is firstly presented, serving as the backbone of the large AI model. Then we introduce the time series large AI model used for channel prediction.

### 5.1 Transformer backbone

Transformers have become the backbone of the large AI model due to their outstanding performance and efficiency. Their architecture, featuring self-attention mechanisms and parallel processing capabilities, excels at capturing dependencies and correlations in data, leading to outstanding results in various applications. These attributes are particularly advantageous for channel prediction tasks, where understanding temporal correlations and dependencies is crucial.

The core feature of the transformer architecture is the self-attention mechanism, which enables the model to consider all positions in the input sequence simultaneously. Given a sequential input  $I$ , the query  $Q$ , key  $K$ , value  $V$  are computed as

$$\begin{cases} Q = IW^Q, \\ K = IW^K, \\ V = IW^V, \end{cases} \quad (23)$$

where  $W^Q$ ,  $W^K$ ,  $W^V$  are learnable weight matrices. The self-attention mechanism is then calculated as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (24)$$

where  $d_k$  is the dimension of the key vectors used as a scaling factor,  $\text{softmax}(\cdot)$  is an activation function. Building upon the self-attention mechanism, the transformer implements multi-head attention. It allows the model to learn information from different representation subspaces, thereby enriching its ability to capture a wider range of dependencies. The multi-head attention is expressed as

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_c | c = 1, 2, \dots, C)W^O, \quad (25)$$

$$\text{head}_c = \text{Attention}(QW_c^Q, KW_c^K, VW_c^V), \quad (26)$$

where  $C$  is the number of the parallel attention heads,  $W_c^Q$ ,  $W_c^K$ ,  $W_c^V$  are the weight matrices for the  $c$ -th head, and  $W^O$  is a learnable projection matrix.

### 5.2 Time series large AI model

We employ a large AI model named Timer for channel prediction. As shown in Figure 3, Timer, short for time series transformer, is a large pre-trained model specifically designed for processing time-series data, utilizing the Transformer architecture and generative pre-training to provide a unified approach to various time-series tasks [19]. For specific down-stream applications, Timer can be fine-tuned on task-specific data, allowing it to adapt to the nuances of the target task while leveraging knowledge gained during pre-training. Pre-trained time series large AI model offers the advantage of low-cost and rapid

deployment of channel prediction through zero-shot learning, where the model performs prediction tasks on previously unseen channel data by leveraging its learned knowledge. Additionally, fine-tuning further enhances channel prediction accuracy by retraining the pre-trained model with specific channel data.

Assume a univariate variate time series sequence  $X = \{x_1, \dots, x_{UJ}\}$  of unified context length  $UJ$  is the input to the Timer model. Timer tokenizes the input data into consecutive time segments that encompass the series variations as

$$j_u = \{x_{(u-1)J+1}, \dots, x_{uJ}\} \in \mathbb{R}^J, u = 1, \dots, U, \quad (27)$$

where  $J$  is the length of each time segment and  $U$  is the number of tokens of the input data. Timer employs a decoder-only Transformer with dimension  $D$  and  $L$  layers, and applies generative pre-training on  $U$  tokens of the sequence. The computations are as follows:

$$z_u^0 = W_e j_u + \text{TE}_u, u = 1, \dots, U, \quad (28)$$

$$Z^l = \text{TransformerBlock}(Z^{l-1}), l = 1, \dots, L, \quad (29)$$

$$\{\hat{j}_{u+1}\} = Z^L W_d, u = 1, \dots, U, \quad (30)$$

where  $W_e, W_d \in \mathbb{R}^{D \times J}$  are the encoding and decoding matrices for token embeddings in  $H = \{h_u\} \in \mathbb{R}^{U \times D}$ , respectively.  $\text{TE}_u$  is the optional time stamp embedding. Using the causal attention within the decoder-only transformer, the model autoregressively generates  $\hat{j}_{u+1}$  as the prediction for the next token given  $j_u$ . The pre-training objective is formulated to minimize the mean squared error between the predicted and actual tokens,

$$\mathcal{L} = \frac{1}{UJ} \sum_{u=2}^U \|j_u - \hat{j}_u\|_2^2, u = 2, \dots, U. \quad (31)$$

This loss function provides token-wise supervisory signals, ensuring that the generated tokens at each position are independently supervised. By leveraging the Transformer architecture's ability to capture long-range dependencies through self-attention, Timer effectively models the temporal correlations in time series data, enhancing its predictive capabilities.

By pre-training on large and diverse time series datasets, Timer learns to recognize general patterns and trends across different sequences, automatically extracting important temporal features and understanding underlying temporal relationships. This pre-training endows Timer with exceptional zero-shot learning capability for processing time series data. Since DD domain channel parameters vary regularly over time, forming a time series, they can be effectively predicted using Timer. Leveraging its learned temporal features, Timer can generalize to unseen channel data, enabling it to perform channel prediction tasks without prior training on specific channel data. Furthermore, fine-tuning Timer with collected channel data can significantly improve its performance, demonstrating its effectiveness in vehicular network prediction. Therefore, the zero-shot capability of a large AI model like Timer enables low-cost channel prediction deployment, while fine-tuning further enhances prediction accuracy.

## 6 Simulation results

In this section, we present simulation results to evaluate the effectiveness of our proposed approach. The experimental setups for the vehicular network scenario and the large AI model are first given. Then, we analyze the prediction accuracy in terms of both DD domain parameters and total path loss, respectively.

### 6.1 Experimental settings

We consider vehicle-to-infrastructure (V2I) channels in a 500-meter-long highway scenario, where a base station with a single antenna is located 100 m from the road center. The road has two lanes in each direction, spaced 5 m apart. Vehicles, equipped with single antennas, travel at two different classic speeds including 60 and 120 km/h. For each pair of side-by-side vehicles traveling in the same direction, one is designated for training and the other for testing. The vehicular channels are generated using the quasi-deterministic radio channel generator (QuaDRiGa), considering both LOS and NLOS conditions [46]. The channels are sampled with a temporal resolution of 0.5 ms. The pre-processing of DD domain channel parameters is performed using z-score normalization. Each of the four parameter types is normalized independently, converting data with varying scales and distributions into a common scale with a mean of

**Table 1** Comparison of prediction errors for different velocities in LOS channel scenario. The best results are in bold.

Model	Speed (km/h)	Delay (s)	Doppler shift (Hz)	Intensity (dB)	Phase difference (rad)
LSTM	60	1.06E−09	0.1239	0.0647	0.0041
	120	2.30E−09	0.3870	0.1010	0.0124
GRU	60	3.44E−10	0.0959	0.0559	0.0031
	120	8.96E−10	0.2774	0.0877	0.0092
Time-MoE (zero-shot)	60	4.16E−10	0.0202	0.0438	0.0013
	120	7.05E−10	<b>0.0504</b>	0.2488	0.0084
Timer (zero-shot)	60	1.64E−11	0.0342	0.0202	0.0010
	120	5.67E−11	0.2041	0.0986	0.0062
Timer (fine-tuning)	60	<b>6.06E−12</b>	<b>0.0176</b>	<b>0.0081</b>	<b>0.0005</b>
	120	<b>2.30E−11</b>	0.1215	<b>0.0519</b>	<b>0.0036</b>

**Table 2** Comparison of prediction errors for different velocities in NLOS channel scenario. The best results are in bold.

Model	Speed (km/h)	Delay (s)	Doppler shift (Hz)	Intensity (dB)	Phase difference (rad)
LSTM	60	4.36E−09	0.4370	0.2253	0.0128
	120	3.68E−09	1.0428	0.2600	0.0330
GRU	60	1.79E−09	0.4117	0.2280	0.0124
	120	1.57E−09	0.9375	0.2525	0.0301
Time-MoE (zero-shot)	60	1.19E−09	0.0819	0.1820	0.0054
	120	1.38E−09	<b>0.1360</b>	0.5234	0.0178
Timer (zero-shot)	60	7.26E−11	0.1353	0.0393	0.0043
	120	1.43E−10	0.5045	0.1218	0.0160
Timer (fine-tuning)	60	<b>3.14E−11</b>	<b>0.0809</b>	<b>0.0226</b>	<b>0.0026</b>
	120	<b>6.33E−11</b>	0.3212	<b>0.0791</b>	<b>0.0104</b>

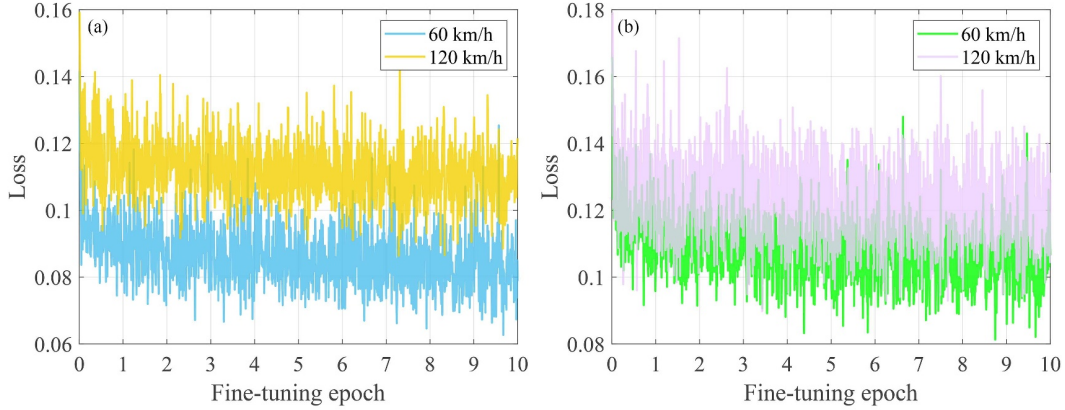
0 and a standard deviation of 1. In addition, the performance of channel prediction is evaluated using a weighted mean absolute error (MAE), where the weighting is determined by the intensity of the different sub-paths. The GPU used in the experiment is the NVIDIA GeForce RTX 4090D.

For the time series large AI model named Timer, the model dimension is configured to 1024 and the feed forward dimension is 2048, while it employs 8 heads within the multi-head attention mechanism [19]. During the fine-tuning process, the learning rate is  $1 \times 10^{-5}$  and the loss function is MAE. Another large AI model named Time-MoE (mixture of experts) is also used for zero-shot prediction [47]. Two classical models including LSTM and GRU are used to predict the channel after training. The input time series length is 20 and the output time series length is 10, indicating that the model predicts future channel conditions based on 10 ms of historical data to forecast the subsequent 5 ms.

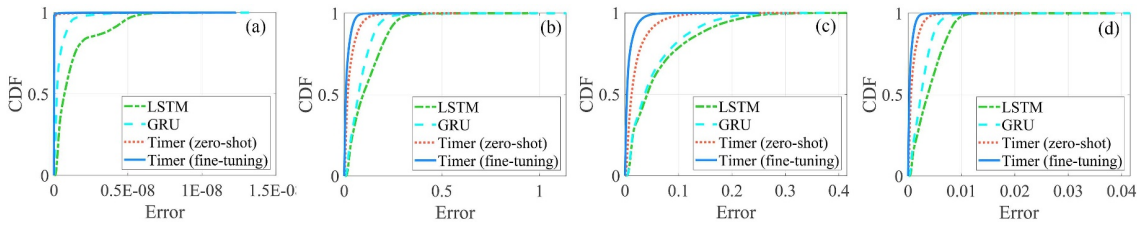
## 6.2 Prediction accuracy of delay-Doppler domain parameters

Tables 1 and 2 present a comprehensive analysis of different models' effectiveness in predicting channel conditions in vehicular networks under both LOS and NLOS scenarios. The results highlight the significant advantage of the fine-tuned Timer model, which consistently outperforms classical models and zero-shot large AI models in most metrics with vehicle speeds of 60 and 120 km/h. In LOS conditions, the fine-tuned Timer demonstrates its ability to effectively capture channel dynamics even at higher vehicle speeds, the scenario where traditional models like LSTM and GRU struggle to adapt. In NLOS conditions, where channel characteristics become even more complex due to multi-path propagation, the fine-tuned Timer maintains its advantage, highlighting its robustness in handling the severe Doppler shifts and multi-path effects characteristic of NLOS environments. The results also underscore the value of the zero-shot capability inherent in large AI models like Timer and Time-MoE, which enables reasonably accurate channel predictions without specific training, thus facilitating low-cost deployment in new scenarios. Moreover, fine-tuning further refines Timers predictive accuracy and enables it to adapt to the unique characteristics of vehicular channels, emphasizing the additional accuracy gained through fine-tuning for task-specific data. In comparison, classical models like LSTM and GRU exhibit considerably higher prediction errors across both LOS and NLOS scenarios, especially as vehicle speed increases, indicating their limitations in capturing the rapid and complex dynamics of vehicular channels.

Figure 4 shows the loss curves of Timer during the fine-tuning process under LOS and NLOS channel



**Figure 4** (Color online) Loss curves of fine-tuning for the large AI model. (a) LOS channel scenario; (b) NLOS channel scenario.



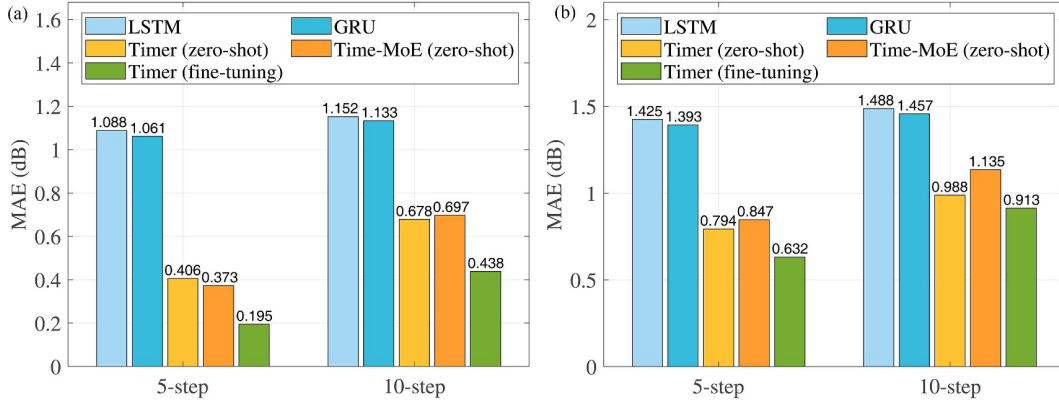
**Figure 5** (Color online) CDF of prediction errors for different delay-Doppler domain parameters. (a) Delay; (b) Doppler shift; (c) intensity; (d) phase difference.

conditions at vehicle speeds of 60 and 120 km/h. In the LOS scenario shown in Figure 4(a), the loss curve for 120 km/h consistently remains higher than that for 60 km/h. This indicates that as vehicle speed increases, the complexity of the channel characteristics in the LOS scenario also increases, resulting in higher loss values. In contrast, in the NLOS scenario shown in Figure 4(b), the loss curves for both speeds overlap more substantially, indicating that NLOS conditions are already sufficiently complex such that further increases in speed do not markedly exacerbate this complexity. Across both scenarios, all curves stabilize after approximately 1000 fine-tuning steps, illustrating Timer’s rapid convergence and efficiency in adapting to varying channel conditions. This rapid convergence highlights Timer’s efficiency in adapting to different vehicular channel prediction scenario, requiring only a relatively short fine-tuning period to perform effectively in specific tasks. This adaptation is beneficial in dynamic vehicular environments, where the large AI model can be accurately, flexibly, and cost-effectively deployed.

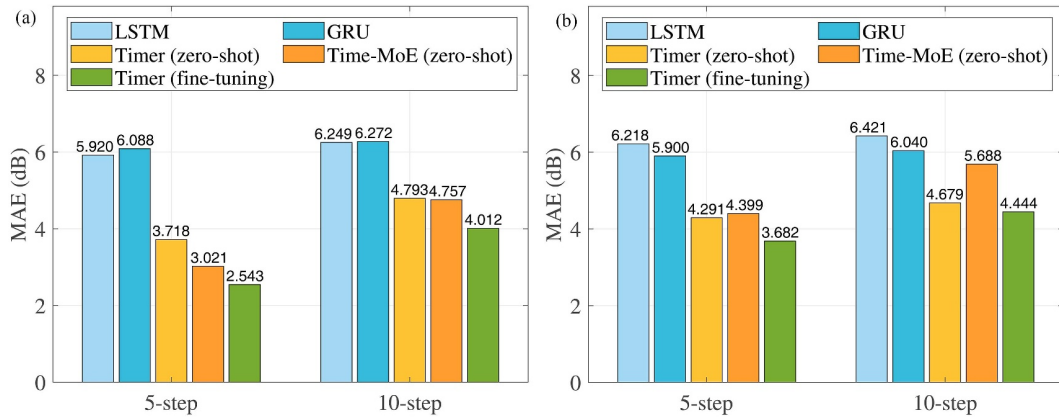
Figure 5 shows the cumulative distribution function (CDF) of prediction errors for various models across four different DD domain channel parameters in the LOS scenario at 60 km/h. The fine-tuned Timer consistently outperforms all other models, with CDF curves that reach 1 quickly for each parameter, indicating minimal prediction errors and high accuracy. Timer in zero-shot mode performs well but falls short of the fine-tuned version, highlighting the benefits of fine-tuning. In contrast, LSTM and GRU display broader CDF curves, reflecting higher prediction errors and less reliability across all DD domain channel parameters. Specifically, the fine-tuned Timer excels in capturing Doppler shift and phase difference accurately, which are critical for handling mobility-induced channel variations, while LSTM and GRU lag considerably, suggesting limitations in adapting to dynamic vehicular conditions. Overall, this analysis underscores the advantage of Timer, particularly when fine-tuned, as an accurate and robust model for channel prediction in high-mobility vehicular environments.

### 6.3 Prediction accuracy of path loss

Figures 6 and 7 compare the path loss prediction performance of five models across LOS and NLOS conditions at vehicle speeds of 60 and 120 km/h, with prediction step length of 5-step (2.5 ms) and 10-step (5 ms). The path loss is obtained by combining the DD domain parameters at each time step. The zero-shot capabilities of Timer and Time-MoE also perform well, indicating the advantage of large AI models for low-cost and generalized deployment without specific training for each environment. Moreover, fine-tuning further enhances Timers accuracy, allowing it to capture unique channel characteristics. The



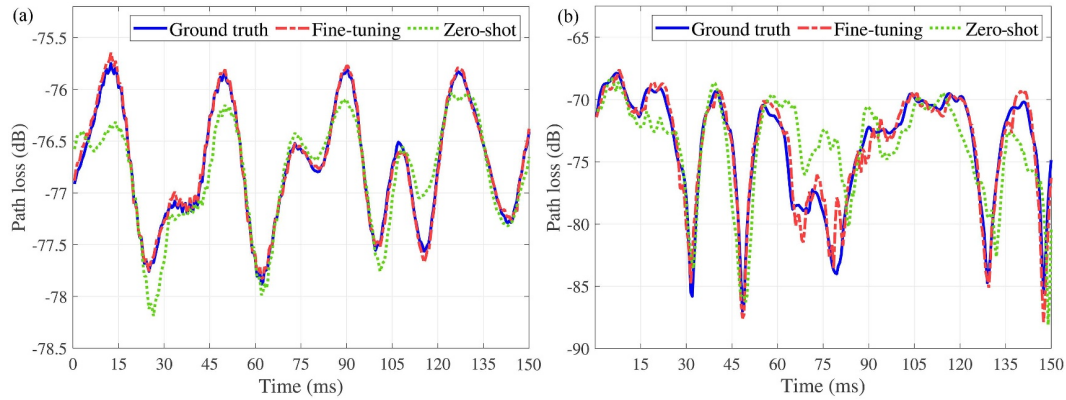
**Figure 6** (Color online) Comparison of mean absolute errors in LOS channel scenario. (a) Speed at 60 km/h; (b) speed at 120 km/h.



**Figure 7** (Color online) Comparison of mean absolute errors in NLOS channel scenario. (a) Speed at 60 km/h; (b) speed at 120 km/h.

fine-tuned Timer model consistently achieves the lowest MAE across all scenarios, demonstrating its adaptability and robustness in dynamic vehicular environments. Classical models, such as LSTM and GRU, show higher MAEs, particularly in NLOS conditions and at 120 km/h, highlighting their limitations in handling the complex, rapidly fluctuating channels typical of vehicular networks. When comparing the 5-step and 10-step prediction performance, all models exhibit higher MAE at the longer prediction steps due to the increased uncertainty. Notably, the fine-tuned Timer model shows a comparatively smaller increase in MAE, indicating its superior ability to capture and extrapolate channel dynamics over longer time spans. Overall, these results emphasize that Timer, especially when fine-tuned, offers a powerful combination of precision, adaptability, and cost-effective deployment potential for vehicular channel prediction.

Figure 8 shows a visual comparison of predicted path loss in the LOS scenario with the vehicle speed of 60 km/h, comparing the ground truth path loss values with predictions from the fine-tuned Timer model and the zero-shot Timer model over two different 15 ms time periods: one with small path loss variation and another with large path loss variation. In the small variation case shown in Figure 8(a), where path loss fluctuates within a 2 dB range, the fine-tuned Timer model closely follows the real path loss trajectory, demonstrating its ability to accurately capture subtle variations. The zero-shot Timer, while capturing the general trend, shows larger deviations from the true values, especially at peaks and troughs, indicating reduced accuracy without fine-tuning. In the large variation case shown in Figure 8(b), where path loss fluctuates by nearly 10 dB, the fine-tuned Timer again closely mirrors the real path loss, maintaining accuracy even with significant fluctuations. This consistency in performance highlights the fine-tuned model’s adaptability to both stable and highly dynamic channel conditions, a critical feature in high-mobility vehicular environments. The zero-shot Timer, however, struggles to match the rapid variations, showing more pronounced deviations from the real path loss, particularly during sharp changes. These observations underscore the advantage of fine-tuning, which enables Timer



**Figure 8** (Color online) Visualization of the predicted path loss using the large AI model. (a) Path loss with small variation; (b) path loss with large variation.

to adapt its predictions closely to real data patterns, improving accuracy and making it more reliable in dynamic conditions.

The results show the value of zero-shot capability in large AI models like Timer, allowing for cost-effective channel prediction deployment without specific training, which can be useful for vehicular networks with complex channel conditions. Meanwhile, fine-tuning significantly enhances Timers predictive accuracy, aligning it more closely with true path loss values, particularly in scenarios with rapid changes. This comparison highlights that while zero-shot models provide a baseline performance, fine-tuning is essential for achieving high accuracy in complex, high-variation environments.

## 7 Conclusion

In this paper, we have proposed a DD domain channel prediction framework for high-mobility vehicular networks, in which the large AI model is leveraged for the DD domain channel parameter prediction. By utilizing the DD domain, we have transformed complex vehicular channels into a set of time series parameters with high predictability. Extensive simulation results have demonstrated that the zero-shot capability of large AI models can provide low-cost channel prediction deployments, while fine-tuning can further achieve high prediction accuracy in complex vehicular networks. Moreover, our framework is highly applicable to a broad range of high-mobility communication systems, including unmanned aerial vehicle (UAV) and satellite-terrestrial communications. For the future work, we will investigate link adaptation in vehicular networks aided by channel prediction.

**Acknowledgements** This work was supported in part by National Natural Science Foundation of China (Grant Nos. 623B2052, 62271244) and Natural Science Fund for Distinguished Young Scholars of Jiangsu Province (Grant No. BK20220067).

## References

- Noor-A-Rahim M, Liu Z, Lee H, et al. 6G for vehicle-to-everything (V2X) communications: enabling technologies, challenges, and opportunities. *Proc IEEE*, 2022, 110: 712–734
- You X H, Huang Y M, Zhang C, et al. When AI meets sustainable 6G. *Sci China Inf Sci*, 2025, 68: 110301
- Chen J, Liang X, Xue J, et al. Evolution of RAN architectures toward 6G: motivation, development, and enabling technologies. *IEEE Commun Surv Tut*, 2024, 26: 1950–1988
- He L, Li F C, Xu H K, et al. Blockchain-based vehicular edge computing networks: the communication perspective. *Sci China Inf Sci*, 2023, 66: 172301
- Hadani R, Rakib S, Tsatsanis M, et al. Orthogonal time frequency space modulation. In: *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, 2017. 1–6
- Li Z, Wang C X, Huang C, et al. A GAN-GRU based space-time predictive channel model for 6G wireless communications. *IEEE Trans Veh Technol*, 2024, 73: 9370–9386
- Liang G, Hu J, Yang K, et al. Data augmentation for predictive digital twin channel: learning multi-domain correlations by convolutional TimeGAN. *IEEE J Sel Top Signal Process*, 2024, 18: 18–33
- Xu W, Yang Z, Ng D W K, et al. Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing. *IEEE J Sel Top Signal Process*, 2023, 17: 9–39
- Chen Z, Zhang Z, Yang Z, et al. Channel deduction: a new learning framework to acquire channel from outdated samples and coarse estimate. *IEEE J Sel Areas Commun*, 2025, 43: 944–958
- Shi X, Xuan X, Huang X, et al. Digital twin empowered wireless positioning: prospects, architecture, and challenges. *IEEE Netw*, 2025. doi: 10.1109/MNET.2025.3544694
- Joo J, Park M C, Han D S, et al. Deep learning-based channel prediction in realistic vehicular communications. *IEEE Access*, 2019, 7: 27846–27858
- Xue J, Yu K, Zhang T, et al. Cooperative deep reinforcement learning enabled power allocation for packet duplication URLLC in multi-connectivity vehicular networks. *IEEE Trans Mobile Comput*, 2024, 23: 8143–8157

- 13 Wang C X, Huang J, Wang H, et al. 6G wireless channel measurements and models: trends and challenges. *IEEE Veh Technol Mag*, 2020, 15: 22–32
- 14 Lin H, Yuan J, Yu W, et al. Multi-carrier modulation: an evolution from time-frequency domain to delay-Doppler domain. 2023. ArXiv:2308.01802
- 15 Mohammed S K, Hadani R, Chockalingam A, et al. OTFS—predictability in the delay-Doppler domain and its value to communication and radar sensing. *IEEE BITS Inform Theor Mag*, 2023, 3: 7–31
- 16 Yuan W, Zhou L, Dehkordi S K, et al. From OTFS to DD-ISAC: integrating sensing and communications in the delay Doppler domain. *IEEE Wireless Commun*, 2024, 31: 152–160
- 17 Yuan W, Li S, Wei Z, et al. New delay Doppler communication paradigm in 6G era: a survey of orthogonal time frequency space (OTFS). *China Commun*, 2023, 20: 1–25
- 18 Liang Y, Wen H, Nie Y, et al. Foundation models for time series analysis: a tutorial and survey. In: *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024. 6555–6565
- 19 Liu Y, Zhang H, Li C, et al. Timer: generative pre-trained transformers are large time series models. In: *Proceedings of the International Conference on Machine Learning (ICML)*, 2024. 32369–32399
- 20 Qu G, Chen Q, Wei W, et al. Mobile edge intelligence for large language models: a contemporary survey. *IEEE Commun Surv Tut*, 2025. doi: 10.1109/comst.2025.3527641
- 21 Long S, Tan J, Mao B, et al. A survey on intelligent network operations and performance optimization based on large language models. *IEEE Commun Surv Tut*, 2025. doi: 10.1109/COMST.2025.3526606
- 22 Gruver N, Finzi M, Qiu S, et al. Large language models are zero-shot time series forecasters. In: *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2023. 19622–19635
- 23 Han Z, Gao C, Liu J, et al. Parameter-efficient fine-tuning for large models: a comprehensive survey. 2024. ArXiv:2403.14608
- 24 Wu C, Yi X, Zhu Y, et al. Channel prediction in high-mobility massive MIMO: from spatio-temporal autoregression to deep learning. *IEEE J Sel Areas Commun*, 2021, 39: 1915–1930
- 25 Peng F, Zhang S, Jiang Z, et al. A novel mobility induced channel prediction mechanism for vehicular communications. *IEEE Trans Wireless Commun*, 2023, 22: 3488–3502
- 26 Wang X, Shi Y, Xin W, et al. Channel prediction with time-varying Doppler spectrum in high-mobility scenarios: a polynomial Fourier transform based approach and field measurements. *IEEE Trans Wireless Commun*, 2023, 22: 7116–7129
- 27 Jiang W, Schotten H D. Neural network-based fading channel prediction: a comprehensive overview. *IEEE Access*, 2019, 7: 118112
- 28 Xue J, Xu Y, Wu W, et al. Sparse mobile crowdsensing for cost-effective traffic state estimation with spatio-temporal transformer graph neural network. *IEEE Internet Things J*, 2024, 11: 16227–16242
- 29 Atluri G, Karpadne A, Kumar V. Spatio-temporal data mining. *ACM Comput Surv*, 2018, 51: 1–41
- 30 Sun Y, Wang Y, Jiang G, et al. Deep learning-based power usage effectiveness optimization for IoT-enabled data center. *Peer-to-Peer Netw Appl*, 2024, 17: 1702–1719
- 31 Guo B, Chen T, He S, et al. Generalized global ranking-aware neural architecture ranker for efficient image classifier search. In: *Proceedings of the ACM International Conference on Multimedia (MM)*, 2022. 3730–3741
- 32 Zhang C, Hu C, Wu T, et al. Achieving efficient and privacy-preserving neural network training and prediction in cloud environments. *IEEE Trans Dependable Secure Comput*, 2023, 20: 4245–4257
- 33 Liu G, Xu Y, He Z, et al. Deep learning-based channel prediction for edge computing networks toward intelligent connected vehicles. *IEEE Access*, 2019, 7: 114487
- 34 Zhang S, Zhang S, Mao Y, et al. Transformer-based channel prediction for rate-splitting multiple access-enabled vehicle-to-everything communication. *IEEE Trans Wireless Commun*, 2024, 23: 12717–12730
- 35 Jiang H, Cui M, Ng D W K, et al. Accurate channel prediction based on transformer: making mobility negligible. *IEEE J Sel Areas Commun*, 2022, 40: 2717–2732
- 36 Zhou H, Hu C, Yuan Y, et al. Large language model (LLM) for telecommunications: a comprehensive survey on principles, key techniques, and opportunities. *IEEE Commun Surv Tut*, 2024. doi: 10.1109/COMST.2024.3465447
- 37 Yu L, Shi L, Zhang J, et al. ChannelGPT: a large model to generate digital twin channel for 6G environment intelligence. 2024. ArXiv:2410.13379
- 38 Jin M, Wang S, Ma L, et al. Time-LLM: time series forecasting by reprogramming large language models. In: *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024
- 39 Chen Z, Zhang Z, Yang Z. Big AI models for 6G wireless networks: opportunities, challenges, and research directions. *IEEE Wireless Commun*, 2024, 31: 164–172
- 40 Sheng Y, Huang K, Liang L, et al. Beam prediction based on large language models. *IEEE Wireless Commun Lett*, 2025, 14: 1406–1410
- 41 Xu M, Niyato D, Chen J, et al. Generative AI-empowered simulation for autonomous driving in vehicular mixed reality metaverses. *IEEE J Sel Top Signal Process*, 2023, 17: 1064–1079
- 42 Sun Y, Zhou H, Cheng B, et al. LOSEC: local semantic capture empowered large time series model for IoT-enabled data centers. *IEEE Internet Things J*, 2025, 12: 13144–13156
- 43 Liu B, Liu X, Gao S, et al. LLM4CP: adapting large language models for channel prediction. *J Commun Inf Netw*, 2024, 9: 113–125
- 44 Zeng Y, Qiao L, Gao Z, et al. CSI-GPT: integrating generative pre-trained transformer with federated-tuning to acquire downlink massive MIMO channels. *IEEE Trans Veh Technol*, 2025, 74: 5187–5192
- 45 Fan S, Liu Z, Gu X, et al. CSI-LLM: a novel downlink channel prediction method aligned with LLM pre-training. 2024. ArXiv:2409.00005
- 46 Jaeckel S, Raschkowski L, Borner K, et al. QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials. *IEEE Trans Antenn Propagat*, 2014, 62: 3242–3256
- 47 Shi X, Wang S, Nie Y, et al. Time-MoE: billion-scale time series foundation models with mixture of experts. In: *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025