

Dual-Band Super-Resolution Channel Prediction in High-Mobility MIMO Systems

Yiliang Sang, Ke Ma[✉], Zhaocheng Wang[✉], *Fellow, IEEE*, and Sheng Chen[✉], *Life Fellow, IEEE*

Abstract—For multiple-input multiple-output systems, channel prediction is crucial for mitigating channel aging in mobile scenarios. The existing channel prediction schemes typically require strictly equal sampling intervals of historical and predicted channel sequences, which imposes enormous pilot overhead in high-mobility scenarios with frequent channel estimation. To tackle this problem, we investigate the super-resolution channel prediction, where the future channel sequence is predicted at a finer temporal resolution without additional channel estimation. Specifically, we theoretically analyze the physics process underlying super-resolution channel prediction to show that the measurement of Doppler phase rotation faces the challenging issue of phase ambiguity in high-mobility and high-frequency scenarios. To address this issue, a deep learning-based dual-band fusion approach is proposed to adaptively integrate the low-frequency information for accurate Doppler phase measurement. To realize accurate channel prediction at a finer temporal resolution, we propose the physics feature-inspired neural ordinary differential equation with modulated-periodic-based multi-layer perceptron for effectively learning the dynamics of fast time-varying channels. Simulation results verify that our proposed scheme outperforms existing channel prediction schemes and it maintains robust performance in high-mobility scenarios.

Index Terms—Multiple-input multiple-output, channel prediction, high-mobility scenarios, Doppler shift, deep learning.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) is one of the fundamental technologies in wireless communication systems [1], [2]. To fully exploit the potential of MIMO, channels need to be acquired at the base station (BS). As the numbers of users and antennas rapidly increase, however, the pilot overhead of channel estimation becomes excessive. Although the existing channel estimation works have proposed to utilize compressed sensing [3], [4], codebook-based

schemes [5], [6], and deep learning (DL) [7], [8] to reduce the overhead, channel aging in mobile user equipment (UE) scenarios exacerbates the problem of pilot overhead [9] due to the need for frequent channel estimation. To resolve these issues, channel prediction has been widely adopted to obtain the future channel sequence without additional channel estimation [10].

The conventional channel prediction schemes mainly include the autoregressive (AR) model [11], [12] and the linear extrapolation model [13]. Specifically, the work [11] proposed a vector Kalman filter-based predictor with AR parameter estimation. The study [12] designed a spatial-temporal AR to leverage the channel sparsity in the angle-delay domain, while a Prony-based angle-delay domain prediction scheme was derived in [13]. Nonetheless, these schemes heavily rely on some simplified theoretical assumptions on channel models, which may not be consistent with the complex time-varying characteristics of MIMO channels encountered in practical scenarios due to the multi-path and Doppler effects.

Since DL has strong capabilities to adaptively capture the complex correlations in channel sequences, it offers a powerful solution to tackle the aforementioned problem. Recently, DL-based channel prediction schemes have been widely studied to flexibly adapt to practical channels [14], [15], [16], [17], [18], [19], [20]. Specifically, recurrent neural network (RNN) was utilized in [14], [15], and [16] to extract the temporal correlations of channel sequences. The works [17] and [18] used convolutional neural network (CNN) to capture the spatial features of channel sequences, followed by a long short-term memory (LSTM) network for channel prediction. The study [19] proposed a spatial-temporal neural network with the attention mechanism, while a transformer-based predictor was designed in [20].

However, the specific sampling interval of the channel sequence is not modeled in the above schemes, and therefore the sampling intervals of historical and predicted channel sequences are required to be strictly equal. In other words, only the channel sequence at the same temporal resolution as the historical channel sequence can be accurately predicted. Unfortunately, in high-mobility scenarios, the sampling interval of predicted channel sequence needs to be quite short to mitigate the severe channel aging, which inevitably leads to enormous pilot overhead due to frequent channel estimation.

Owing to the fact that channel sequences at different temporal resolutions fundamentally reflect the continuous-time variations in the same channel, they exhibit the temporal correlations. Thus, leveraging the historical channel sequence

Received 13 May 2024; revised 10 October 2024; accepted 30 November 2024. Date of publication 5 December 2024; date of current version 18 June 2025. This work was supported by the National Natural Science Foundation of China under Grant U22B2057. The associate editor coordinating the review of this article and approving it for publication was M. Matthaiou. (*Corresponding authors: Zhaocheng Wang; Ke Ma.*)

Yiliang Sang and Ke Ma are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: sangyl23@mails.tsinghua.edu.cn; ma-k19@mails.tsinghua.edu.cn).

Zhaocheng Wang is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China (e-mail: zcwang@tsinghua.edu.cn).

Sheng Chen is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: sqc@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/TCOMM.2024.3511954

can potentially predict the future channel sequence at a finer temporal resolution, which is termed the super-resolution channel prediction. As a promising solution, the neural ordinary differential equation (ODE) [21], [22], [23], [24], with its powerful continuous-time learning architecture, can be used to model channel prediction at a finer temporal resolution. In particular, the ODE-RNN was proposed in [25] to learn the temporal dynamics for continuous-time channel prediction, while a tensor neural ODE was employed in [26] to implement channel prediction at a finer temporal resolution. In the study [27], the ODE-LSTM architecture that combines the neural ODE with LSTM was designed.

Nevertheless, these existing neural ODE-based channel prediction schemes are purely data-driven, and they neglect the physics process underlying super-resolution channel prediction. Specifically, the temporal channel variations arise from the Doppler phase rotation of each path, which is proportional to the UE velocity and the carrier frequency [12], [13], [20]. In high-mobility and high-frequency scenarios, the Doppler phase rotation between adjacent channel estimations may span multiple integer cycles (multiple of 2π), where the specific multiple integer cycles cannot be determined. However, to predict the channel sequence at a finer temporal resolution, the actual Doppler phase rotation of each path needs to be accurately measured. Thus, the uncertainty in determining the exact multiple integer cycles between adjacent channel estimations can impede the measurement of actual Doppler phase rotation of each path. This issue is referred to as phase ambiguity, which causes serious performance degradation of super-resolution channel prediction in high-mobility and high-frequency scenarios.

Fortunately, in the non-stand-alone (NSA) architecture widely deployed in beyond 5G and 6G systems [28], [29], [30], high- and low-frequency antennas are generally co-located at the same BS [31], [32], [33]. Therefore, dual-band channels share the same surrounding multi-path environment, and their corresponding paths generated by the same scatterer are demonstrated to exhibit the spatial congruence including the analogous angles and delays in field experiments [34], [35], [36], [37]. More importantly, since the Doppler phase rotation is proportional to the carrier frequency, the Doppler phase rotation within the same sampling interval for the low-frequency band is smaller than that for the high-frequency band. Therefore, utilizing the low-frequency channel sequence can more accurately measure its actual Doppler phase rotation in each path. Furthermore, based on the spatial congruence of paths in dual-band channels, the accurately measured actual Doppler phase rotation in each path from the low-frequency band can be used to assist in the measurement of actual Doppler phase rotation in the corresponding path for the high-frequency band, overcoming the issue of phase ambiguity.

In addition, the existing neural ODE-based channel prediction schemes encounter difficulties in effectively fitting temporal channel dynamics in high-mobility scenarios, which hinders accurate channel prediction at a finer temporal resolution. In particular, in high-mobility scenarios, the magnitude part of channel dynamics is a slowly time-varying function [26], [38], while the phase part is a periodic function

with rapid temporal variations due to the drastic changing Doppler phase rotation. However, the existing neural ODE-based channel prediction schemes employ piecewise linear and monotonic activation functions, such as rectified linear unit (ReLU), tanh and sigmoid, and hence they cannot efficiently fit the rapidly time-varying and periodic phase part.

In this paper, we propose a dual-band fusion approach to integrate the low-frequency information for assisting in the super-resolution channel prediction in the high-frequency band, so that the issue of phase ambiguity in high-mobility scenarios can be tackled by the accurate measurement of actual Doppler phase rotation in the low-frequency band. To effectively extract the Doppler phase rotation features in each path from the multi-path channels, DL is adopted to adaptively implement the dual-band fusion. Furthermore, in order to efficiently fit channel dynamics in high-mobility scenarios, we propose the physics feature-inspired neural ODE with modulated-periodic-based multi-layer perceptron (MLP), wherein the periodic MLP exploits rapidly time-varying and periodic sine function as nonlinear activations to facilitate learning of the phase part in channel dynamics, and the modulated MLP utilizes piecewise linear ReLU as nonlinear activations to fit the remaining magnitude part. Simulation results demonstrate that our proposed scheme outperforms existing channel prediction schemes and it maintains robust performance in high-mobility scenarios. The main contributions of this paper can be summarized as follows.

- We theoretically analyze the physics process underlying super-resolution channel prediction to demonstrate that the accurate measurement of actual Doppler phase rotation faces the issue of phase ambiguity in high-mobility and high-frequency scenarios.
- We propose a DL-based dual-band fusion approach to adaptively integrate the low-frequency information for addressing the issue of phase ambiguity in the high-frequency band.
- We investigate the physics features of channel dynamics in high-mobility scenarios, and show that the phase part of channel dynamics exhibits rapid temporal variations and periodicity, while the magnitude part changes slowly over time.
- To specifically exploit these physics features of channel dynamics, we propose the neural ODE with modulated-periodic-based MLP to efficiently fit channel dynamics in high-mobility scenarios for accurate channel prediction at a finer temporal resolution.

Table I offers a brief comparison of our proposed scheme with existing channel prediction schemes. In contrast to existing schemes, the proposed scheme can handle the issue of phase ambiguity by integrating the low-frequency information and sufficiently leverage the physics features of channel dynamics for accurate super-resolution channel prediction.

This paper is organized as follows. Section II presents the system model. In Section III, the physics process underlying super-resolution channel prediction is investigated, where the phase ambiguity is elaborated. The proposed dual-band fusion approach is presented in Section IV, and our physics feature-inspired neural ODE with modulated-periodic-based MLP is

TABLE I
COMPARISON BETWEEN PROPOSED AND EXISTING CHANNEL PREDICTION SCHEMES

Schemes	Finer temporal resolution prediction	Handling phase ambiguity	Exploiting physics features of channel dynamics	Key points
Conventional DL-based [14]–[20]	×	×	×	Require strictly equal sampling intervals of historical and predicted channel sequences
Neural ODE-based [25]–[27]	✓	×	×	Leverage purely data-driven neural ODE architecture for continuous-time channel prediction
Proposed	✓	✓	✓	Integrate low-frequency information for handling phase ambiguity and utilize physics features of channel dynamics

elaborated in Section V. Section VI presents the simulation results. Our conclusions are drawn in Section VII.

Notations: $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ denote the $m \times n$ real and complex spaces, respectively. Matrices and vectors are denoted by boldface capital and lower-case letters, respectively. $[\cdot]_i$ represents the i -th element of a vector or the i -th column of a matrix, and $[\cdot]_{i,j}$ denotes the (i, j) -th element of a matrix. $j = \sqrt{-1}$ represents the imaginary unit. $(\cdot)^T$ and $(\cdot)^H$ express the transpose and conjugate transpose, respectively. I_N denotes the $N \times N$ identity matrix. For a complex number x , $|x|$ denotes its magnitude, and $\angle x$ represents its phase. $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the 2-norm and infinite norm of a vector, respectively. $\lfloor \cdot \rfloor$ is the floor function, and $\text{mod}(a, b)$ denotes a modulo b . \odot represents the element-wise product, while \circ is the function composition operator. \rightarrow denotes the mapping, implying that given an input, there exists only one unique output. \mathcal{U} and \mathcal{N} denote the uniform distribution and the Gaussian distribution, respectively. $\log(\cdot)$ denotes the logarithm function with base e .

II. SYSTEM MODEL

We consider the NSA architecture, where high- and low-frequency antennas are deployed at the same BS. The C-band as 3.5 GHz and sub-1G band as 0.9 GHz are utilized to represent respectively the high- and low-frequency bands. Our objective is to exploit the sub-1G information to assist in the C-band channel prediction. Moreover, we consider the time division duplexing (TDD) based multi-user MIMO system, where a BS serves K UEs. For simplicity, each UE is equipped with one C-band antenna and one sub-1G antenna but our proposed scheme can be directly extended to the scenario with multiple-antenna UEs. The BS is equipped with dual-polarized antennas in the uniform planar array for dual bands, and the numbers of C-band/sub-1G antennas in the horizontal and vertical directions are M_h/\overline{M}_h and M_v/\overline{M}_v , respectively. Hence, the total numbers of C-band and sub-1G antennas at the BS are given by $M = 2M_hM_v$ and $\overline{M} = 2\overline{M}_h\overline{M}_v$, respectively. To clearly distinguish two frequency bands, the variables of sub-1G band are marked with overline throughout.

A. Channel Model

The 3D time-varying multi-path channel model [39] that has been widely used in existing channel prediction works [40], [41] is adopted for the C-band channel. Denote the C-band channel vector at time t between the BS and the k -th UE as

$\mathbf{h}^{(k)}(t) \in \mathbb{C}^{M \times 1}$, whose m -th element $[\mathbf{h}^{(k)}(t)]_m$ is given by

$$[\mathbf{h}^{(k)}(t)]_m = \sum_{l_p=1}^{L_p^{(k)}} \alpha_{l_p}^{(k)} e^{j2\pi \frac{(\mathbf{r}_{\text{tx}, l_p}^{(k)})^T \mathbf{d}_{\text{tx}, m}}{\lambda}} e^{j2\pi f_{D, l_p}^{(k)} t} e^{-j2\pi f_c \tau_{l_p}^{(k)}}, \quad (1)$$

where $L_p^{(k)}$ denotes the number of paths, $\alpha_{l_p}^{(k)}$, $f_{D, l_p}^{(k)}$ and $\tau_{l_p}^{(k)}$ represent the complex path gain, Doppler shift and delay of the l_p -th path, respectively, while λ denotes the wavelength, f_c is the carrier frequency, and $\mathbf{d}_{\text{tx}, m}$ is the location vector of the m -th BS antenna. Furthermore, let us define the spherical unit vector \mathbf{r} at azimuth angle ϕ and elevation angle θ as

$$\mathbf{r}(\phi, \theta) = [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta]^T. \quad (2)$$

Then, $\mathbf{r}_{\text{tx}, l_p}^{(k)}$ in (1), which denotes the spherical unit vector at azimuth departure angle $\phi_{l_p, \text{AOD}}^{(k)}$ and elevation departure angle $\theta_{l_p, \text{ZOD}}^{(k)}$, can be written as

$$\mathbf{r}_{\text{tx}, l_p}^{(k)} = \mathbf{r}(\phi_{l_p, \text{AOD}}^{(k)}, \theta_{l_p, \text{ZOD}}^{(k)}). \quad (3)$$

In (1), the temporal channel variations are caused by the Doppler phase rotation $2\pi f_{D, l_p}^{(k)} t$ of each path, while the other channel parameters are usually assumed to be slow-changing over time [12], [20], [26]. However, when the UE velocity v is very high, the time-varying features of these other channel parameters also need to be modeled [19], [42]. Achieving accurate channel prediction under the scenario that these other channel parameters change relatively fast over time is left for future work. Furthermore, the Doppler shift $f_{D, l_p}^{(k)}$ in (1), which is proportional to the UE velocity $v^{(k)}$ and the carrier frequency f_c [12], [13], [20], can be written as

$$f_{D, l_p}^{(k)} = \frac{f_c (\mathbf{r}_{\text{rx}, l_p}^{(k)})^T \mathbf{v}^{(k)}}{c}, \quad (4)$$

where c is the light speed, and $\mathbf{r}_{\text{rx}, l_p}^{(k)}$ represents the spherical unit vector at azimuth arrival angle $\phi_{l_p, \text{AOA}}^{(k)}$ and elevation arrival angle $\theta_{l_p, \text{ZOA}}^{(k)}$, which can be expressed similarly to (3), while $\mathbf{v}^{(k)}$ denotes the k -th UE velocity vector at travel azimuth angle $\phi_v^{(k)}$ and travel elevation angle $\theta_v^{(k)}$, which can be expressed as

$$\mathbf{v}^{(k)} = v^{(k)} \mathbf{r}(\phi_v^{(k)}, \theta_v^{(k)}). \quad (5)$$

For the sub-1G band, the channel vector at time t between the BS and the k -th UE $\overline{\mathbf{h}}^{(k)}(t) \in \mathbb{C}^{\overline{M} \times 1}$ can be modeled in the way similar to that for its C-band counterpart.

The wireless channel results from the interaction between the transmitted signals and the surrounding multi-path environment around BS and UE [33]. In the NSA architecture, since C-band and sub-1G antennas are deployed at the same BS, the dual-band channels share a common multi-path environment for a given UE, so that their paths generated by the same scatterer exhibit spatial congruence, including similar azimuth and elevation arrival angles [34], [35], [36], [37]. For the azimuth and elevation arrival angles of dual bands, $\phi_{l_p, \text{AOA}}^{(k)}$ and $\bar{\phi}_{l_p, \text{AOA}}^{(k)}$ as well as $\theta_{l_p, \text{ZOA}}^{(k)}$ and $\bar{\theta}_{l_p, \text{ZOA}}^{(k)}$, the analogy can be formulated as

$$\bar{\phi}_{l_p, \text{AOA}}^{(k)} = \phi_{l_p, \text{AOA}}^{(k)} + \Delta\phi_{\text{AOA}}, \quad (6)$$

$$\bar{\theta}_{l_p, \text{ZOA}}^{(k)} = \theta_{l_p, \text{ZOA}}^{(k)} + \Delta\theta_{\text{ZOA}}, \quad (7)$$

where $\Delta\phi_{\text{AOA}}$ and $\Delta\theta_{\text{ZOA}}$ denote dual-band angular deviations of azimuth and elevation arrival angles, respectively.

B. Transmission Model

To estimate the channels at the BS, all the UEs transmit the orthogonal pilots periodically. Let us define the period of channel estimation as T_e . The uplink received signal at time jT_e , $\mathbf{Y}_j \in \mathbb{C}^{M \times K}$, can be expressed as

$$\mathbf{Y}_j = \sqrt{\rho_{\text{UE}}} \mathbf{H}_j \mathbf{\Psi} + \mathbf{N}_j, \quad (8)$$

where ρ_{UE} is the UE transmit power, $\mathbf{H}_j \in \mathbb{C}^{M \times K}$ is the channel matrix $\mathbf{H}(t) = [\mathbf{h}^{(1)}(t), \mathbf{h}^{(2)}(t), \dots, \mathbf{h}^{(K)}(t)] \in \mathbb{C}^{M \times K}$ at time jT_e , i.e., $\mathbf{H}_j = \mathbf{H}(jT_e)$, while $\mathbf{\Psi} \in \mathbb{C}^{K \times K}$ is the pilot matrix satisfying $\mathbf{\Psi} \mathbf{\Psi}^H = \mathbf{I}_K$, and $\mathbf{N}_j \in \mathbb{C}^{M \times K}$ represents the corresponding additional white Gaussian noise (AWGN) matrix whose elements have variance σ^2 . According to the received signal \mathbf{Y}_j , the estimated channel matrix $\tilde{\mathbf{H}}_j = [\tilde{\mathbf{h}}_j^{(1)}, \tilde{\mathbf{h}}_j^{(2)}, \dots, \tilde{\mathbf{h}}_j^{(K)}] \in \mathbb{C}^{M \times K}$ can be obtained by the classical least square (LS) or minimum mean square error (MMSE) channel estimation [43]. Then, based on the well-known reciprocity between uplink and downlink channels in the TDD system [44], the estimated uplink channel matrix $\tilde{\mathbf{H}}_j$ can be used for downlink channel prediction.

Furthermore, the spectral efficiency in downlink data transmission R_j , which has been widely used in existing channel prediction works [11], [20], [38], [45], is utilized to evaluate the theoretical system performance, which can be written as

$$R_j = \sum_{k=1}^K \log_2 \left(1 + \frac{\rho_{\text{BS}} |(\mathbf{h}_j^{(k)})^H \mathbf{q}_j^{(k)}|^2}{\sum_{m \neq k} \rho_{\text{BS}} |(\mathbf{h}_j^{(k)})^H \mathbf{q}_j^{(m)}|^2 + \sigma^2} \right), \quad (9)$$

where ρ_{BS} is the BS transmit power, $\mathbf{h}_j^{(k)} \in \mathbb{C}^{M \times 1}$ is the channel $\mathbf{h}^{(k)}(t)$ at time jT_e , i.e., $\mathbf{h}_j^{(k)} = \mathbf{h}^{(k)}(jT_e)$, and $\mathbf{q}_j^{(k)} \in \mathbb{C}^{M \times 1}$ represents the precoding vector satisfying $\|\mathbf{q}_j^{(k)}\|_2 = 1$, which is designed based on $\tilde{\mathbf{H}}_j$.

III. SUPER-RESOLUTION CHANNEL PREDICTION AND PHASE AMBIGUITY IN HIGH-MOBILITY SCENARIOS

In this section, we firstly introduce the frameworks of conventional channel prediction and super-resolution channel

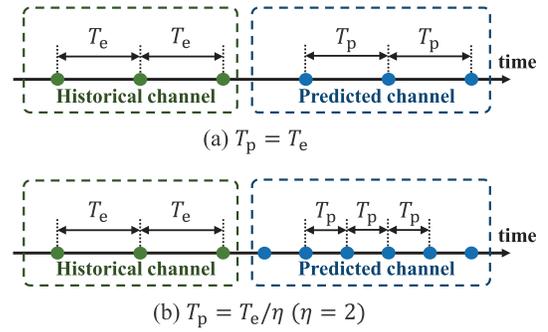


Fig. 1. (a) Conventional channel prediction framework, and (b) Super-resolution channel prediction framework.

prediction. Next, we theoretically analyze the physics process underlying super-resolution channel prediction, wherein the phase ambiguity is elaborated.

A. Channel Prediction Problem Formulation

In wireless communication systems, the number of UEs is growing exponentially [16], which imposes substantial pilot overhead of channel estimation. Moreover, to mitigate channel aging in mobile UE scenarios, frequent channel estimation aggravates the problem of pilot overhead.

To this end, channel prediction has been extensively studied to obtain the future channel sequence based on the temporal correlations [14], [15], [16], [17], [18], [19], [20]. Specifically, by using orthogonal pilots among different UEs, channel prediction can be applied to each UE separately [11], [20], [38]. Thus the design of channel prediction schemes can focus on a single UE, and the UE superscript k is omitted. Besides, as the number of UEs becomes large, nearby UEs usually exhibit spatial correlations [46], [47], which can be utilized for further improving the performance of channel prediction. Exploiting spatial correlations among nearby UEs is left for future channel prediction works.

As depicted in Fig. 1(a), the existing works typically formulate channel prediction as time series prediction with the same temporal resolution. Consequently, the prediction period of future channel sequence T_p must be strictly equal to the estimation period of historical channel sequence T_e in the conventional channel prediction framework, which can be formulated as

$$\{\tilde{\mathbf{h}}_j\}_{j=1}^J \rightarrow \{\hat{\mathbf{h}}_{J+p}\}_{p=1}^P, \quad (10)$$

where $\hat{\mathbf{h}}_{J+p} \in \mathbb{C}^{M \times 1}$ is the predicted channel vector at time $(J+p)T_e$, and J and P denote the lengths of historical and predicted channel sequences, respectively, while the mapping \rightarrow implies that it is mathematically feasible to uniquely determine the future channel sequence $\{\hat{\mathbf{h}}_{J+p}\}_{p=1}^P$ solely based on the historical channel sequence $\{\tilde{\mathbf{h}}_j\}_{j=1}^J$. In high-mobility scenarios, channel aging is extremely severe. To mitigate this problem, the prediction period T_p needs to be very short. Unfortunately, since the estimation period T_e is strictly equal to T_p in the conventional channel prediction

framework, it imposes excessive pilot overhead of channel estimation.

Owing to the fact that channel sequences at different temporal resolutions inherently reflect the continuous-time variations in the same channel $\mathbf{h}(t)$, they exhibit the temporal correlations. Thus, the historical channel sequence has great potential for predicting the future channel sequence at a finer temporal resolution, which is termed the super-resolution channel prediction. As shown in Fig. 1(b), we assume that the estimation period T_e is η times the prediction period T_p , i.e., $T_e/T_p = \eta > 1$. In this way, the future channel sequence is predicted at a finer temporal resolution to mitigate severe channel aging, while maintaining the pilot overhead of channel estimation unchanged. Let us define $\hat{\mathbf{h}}_{(J,p)} \in \mathbb{C}^{M \times 1}$ as the predicted channel vector at time $JT_e + pT_p$. Without additional information, however, it is mathematically infeasible to uniquely determine the future channel sequence at a finer temporal resolution $\{\hat{\mathbf{h}}_{(J,p)}\}_{p=1}^P$ from the historical channel sequence $\{\tilde{\mathbf{h}}_j\}_{j=1}^J$ in high-mobility and high-frequency scenarios. The reason is elaborated in the next subsection.

B. Phase Ambiguity of Super-Resolution Channel Prediction

As the temporal channel variations are determined by the Doppler phase rotation at each path, the objective of channel prediction essentially is predicting the corresponding Doppler phase rotation. We present the phase ambiguity theorem to explain the infeasibility of predicting $\{\hat{\mathbf{h}}_{(J,p)}\}_{p=1}^P$ solely based on $\{\tilde{\mathbf{h}}_j\}_{j=1}^J$ in high-mobility and high-frequency scenarios.

Theorem 1 (Phase Ambiguity in Single-Band Systems): If the number of integer cycles of C-band Doppler phase rotation between adjacent channel estimations, $n_{D,l_p} = \lfloor \frac{2\pi f_{D,l_p} T_e}{2\pi} \rfloor$, satisfies $n_{D,l_p} \geq 1$, predicting $\{\hat{\mathbf{h}}_{(J,p)}\}_{p=1}^P$ solely based on $\{\tilde{\mathbf{h}}_j\}_{j=1}^J$ is infeasible.

Proof: Due to the periodic feature of Doppler phase rotation $2\pi f_{D,l_p} T_e$ in $\{\tilde{\mathbf{h}}_j\}_{j=1}^J$, only the Doppler phase rotation within one integer cycle, $\text{mod}(2\pi f_{D,l_p} T_e, 2\pi)$, can be measured, while n_{D,l_p} cannot be determined. In order to calculate $\text{mod}(2\pi f_{D,l_p} T_p, 2\pi)$ for predicting $\{\hat{\mathbf{h}}_{(J,p)}\}_{p=1}^P$, $2\pi f_{D,l_p} T_e$ needs to be estimated because $T_p < T_e$. However, when $n_{D,l_p} \geq 1$, the uncertainty in determining n_{D,l_p} means that it is infeasible to distinguish the desired $2\pi f_{D,l_p} T_e$ from various elements in $\{\text{mod}(2\pi f_{D,l_p} T_e, 2\pi) + 2\pi n_{D,l_p} \mid n_{D,l_p} \in \mathbb{N}_+\}$. Therefore, predicting $\{\hat{\mathbf{h}}_{(J,p)}\}_{p=1}^P$ solely based on $\{\tilde{\mathbf{h}}_j\}_{j=1}^J$ is infeasible if $n_{D,l_p} \geq 1$. \square

The above phase ambiguity is similar to the well-known spatial aliasing effect in phased array antennas [48], [49], [50], which states that it is infeasible to reconstruct a spatial signal if the spacing of the antenna elements is greater than half of the wavelength of this signal.

IV. DUAL-BAND FUSION APPROACH FOR SUPER-RESOLUTION CHANNEL PREDICTION

A. Motivation

In the NSA architecture, since the C-band and sub-1G channels share the common surrounding multi-path environment, their corresponding paths generated by the same scatterer exhibit the spatial congruence, especially for the analogous $\phi_{l_p, \text{AOA}}$ and $\bar{\phi}_{l_p, \text{AOA}}$, and the analogous $\theta_{l_p, \text{ZOA}}$ and $\bar{\theta}_{l_p, \text{ZOA}}$ of each path [34], [35], [36], [37]. Based on this spatial congruence, we have the following corollary.

Corollary 1: For a dual-band system, assuming that $\phi_{l_p, \text{AOA}} = \bar{\phi}_{l_p, \text{AOA}}$ and $\theta_{l_p, \text{ZOA}} = \bar{\theta}_{l_p, \text{ZOA}}$, then the C-band and sub-1G Doppler phase rotations within one integer cycle, $\text{mod}(2\pi f_{D,l_p} T_e, 2\pi)$ and $\text{mod}(2\pi \bar{f}_{D,l_p} \bar{T}_e, 2\pi)$, satisfy

$$\begin{aligned} & \text{mod}(2\pi f_{D,l_p} T_e, 2\pi) - \frac{f_c T_e}{f_c \bar{T}_e} \text{mod}(2\pi \bar{f}_{D,l_p} \bar{T}_e, 2\pi) \\ &= 2\pi \left(\frac{f_c T_e}{f_c \bar{T}_e} \bar{n}_{D,l_p} - n_{D,l_p} \right), \end{aligned} \quad (11)$$

where $\bar{n}_{D,l_p} = \lfloor \frac{2\pi \bar{f}_{D,l_p} \bar{T}_e}{2\pi} \rfloor$ is the number of integer cycles of sub-1G Doppler phase rotation within \bar{T}_e .

Proof: When $\phi_{l_p, \text{AOA}} = \bar{\phi}_{l_p, \text{AOA}}$ and $\theta_{l_p, \text{ZOA}} = \bar{\theta}_{l_p, \text{ZOA}}$, the spherical unit vectors of the dual bands are the same, i.e., $\mathbf{r}_{\text{rx}, l_p} = \bar{\mathbf{r}}_{\text{rx}, l_p}$, according to the definition of spherical unit vector (3) for $\mathbf{r}_{\text{rx}, l_p}$ and the similar form for $\bar{\mathbf{r}}_{\text{rx}, l_p}$. Thus, for a given UE, the Doppler phase rotations of the dual bands satisfy $\frac{2\pi f_{D,l_p} T_e}{2\pi \bar{f}_{D,l_p} \bar{T}_e} = \frac{f_c T_e}{f_c \bar{T}_e}$, and (11) can be obtained by substituting $\text{mod}(2\pi f_{D,l_p} T_e, 2\pi) + 2\pi n_{D,l_p}$ for $2\pi f_{D,l_p} T_e$ and $\text{mod}(2\pi \bar{f}_{D,l_p} \bar{T}_e, 2\pi) + 2\pi \bar{n}_{D,l_p}$ for $2\pi \bar{f}_{D,l_p} \bar{T}_e$. \square

In Corollary 1, it can be observed that the mappings of dual-band Doppler phase rotations within one integer cycle can be utilized to infer n_{D,l_p} and \bar{n}_{D,l_p} , which can further be exploited to reconstruct the actual C-band Doppler phase rotation by $2\pi f_{D,l_p} T_e = \text{mod}(2\pi f_{D,l_p} T_e, 2\pi) + 2\pi n_{D,l_p}$. Note that when $\phi_{l_p, \text{AOA}}$ and $\bar{\phi}_{l_p, \text{AOA}}$ as well as $\theta_{l_p, \text{ZOA}}$ and $\bar{\theta}_{l_p, \text{ZOA}}$ exist deviations, the information from the mappings of dual-band Doppler phase rotations within one integer cycle can still be utilized to infer $2\pi f_{D,l_p} T_e$, because the dual-band angular deviations are very limited in practical scenarios [34], [37].

However, different pairs of n_{D,l_p} and \bar{n}_{D,l_p} can map to the same $2\pi \left(\frac{f_c T_e}{f_c \bar{T}_e} \bar{n}_{D,l_p} - n_{D,l_p} \right)$ in (11), indicating that the phase ambiguity still exists in dual-band systems. In the subsequent theorem, we investigate this phase ambiguity and find that the solution spaces of n_{D,l_p} and \bar{n}_{D,l_p} are usually smaller than \mathbb{N}_+ when the phase ambiguity exists in dual-band systems. Since \mathbb{N}_+ is the solution space of n_{D,l_p} when the phase ambiguity exists in single-band systems according to Theorem 1, integrating the sub-1G information can assist in accurate super-resolution channel prediction in the C-band.

Theorem 2 (Phase Ambiguity in Dual-Band Systems): For a dual-band system, the solution space of $(n_{D,l_p}, \bar{n}_{D,l_p})$, $\mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p})$, which contains all elements mapping to the

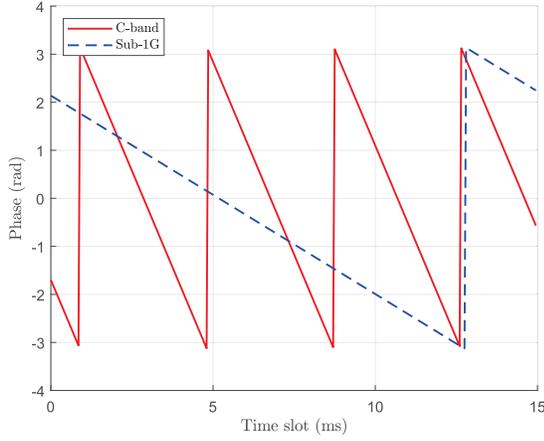


Fig. 2. Doppler phase rotations of C-band and sub-1G band versus time.

same $2\pi\left(\frac{f_c T_e}{f_c \bar{T}_e} \bar{n}_{D,l_p} - n_{D,l_p}\right)$, is given by

$$\begin{aligned} \mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p}) = & \\ \left\{ \left(\text{mod}(n_{D,l_p}, p_D) + n_d p_D, \text{mod}(\bar{n}_{D,l_p}, q_D) + n_d q_D \right) \mid n_d \in \mathbb{N}, \right. & \\ \left. p_D, q_D \in \mathbb{N}_+, \frac{p_D}{q_D} = \frac{f_c T_e}{f_c \bar{T}_e} \text{ in irreducible form} \right\}, & \quad (12) \end{aligned}$$

where “in irreducible form” means that $\frac{p_D}{q_D}$ is the irreducible fraction of $\frac{f_c T_e}{f_c \bar{T}_e}$.

Proof: Let us define the different pairs of n_{D,l_p} and \bar{n}_{D,l_p} that map to the same $2\pi\left(\frac{f_c T_e}{f_c \bar{T}_e} \bar{n}_{D,l_p} - n_{D,l_p}\right)$ as $n_{D,l_p} + \Delta n_{D,l_p}$ and $\bar{n}_{D,l_p} + \Delta \bar{n}_{D,l_p}$, respectively, where $\Delta n_{D,l_p}, \Delta \bar{n}_{D,l_p} \in \mathbb{Z}_{\neq 0}$ denote the corresponding integer differences. The phase ambiguity in dual-band systems can be expressed as

$$\frac{f_c T_e}{f_c \bar{T}_e} \bar{n}_{D,l_p} - n_{D,l_p} = \frac{f_c T_e}{f_c \bar{T}_e} (\bar{n}_{D,l_p} + \Delta \bar{n}_{D,l_p}) - (n_{D,l_p} + \Delta n_{D,l_p}). \quad (13)$$

(13) can be further simplified into

$$\frac{\Delta n_{D,l_p}}{\Delta \bar{n}_{D,l_p}} = \frac{f_c T_e}{f_c \bar{T}_e} = \frac{p_D}{q_D}, \quad (14)$$

where $p_D, q_D \in \mathbb{N}_+$, and $\frac{p_D}{q_D}$ is the irreducible fraction of $\frac{f_c T_e}{f_c \bar{T}_e}$.

Since $\Delta n_{D,l_p}, \Delta \bar{n}_{D,l_p} \in \mathbb{Z}_{\neq 0}$, (14) indicates that when $\Delta n_{D,l_p} = n_d p_D$ and $\Delta \bar{n}_{D,l_p} = n_d q_D$ with $n_d \in \mathbb{Z}$, the corresponding mapping of $2\pi\left(\frac{f_c T_e}{f_c \bar{T}_e} \bar{n}_{D,l_p} - n_{D,l_p}\right)$ remains the same. Therefore, $(\text{mod}(n_{D,l_p}, p_D), \text{mod}(\bar{n}_{D,l_p}, q_D))$ is a feasible solution in $\mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p})$. Meanwhile, since the integer cycle of Doppler phase rotation is non-negative, $(\text{mod}(n_{D,l_p}, p_D), \text{mod}(\bar{n}_{D,l_p}, q_D))$ is also the smallest solution in $\mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p})$. Consequently, all feasible solutions of $(n_{D,l_p}, \bar{n}_{D,l_p})$ in $\mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p})$ follow the form of $(\text{mod}(n_{D,l_p}, p_D) + n_d p_D, \text{mod}(\bar{n}_{D,l_p}, q_D) + n_d q_D)$ with $n_d \in \mathbb{N}$, which completes the proof. \square

Remark 1: Since the Doppler phase rotation is proportional to the carrier frequency, the velocity of Doppler phase rotation in the sub-1G band is slower than that in the C-band, as shown in Fig. 2. When the sub-1G Doppler phase rotation does not span one integer cycle, (11) is equivalent to

$2\pi f_{D,l_p} T_e = \frac{f_c T_e}{f_c \bar{T}_e} \text{mod}(2\pi \bar{f}_{D,l_p} \bar{T}_e, 2\pi)$, which means that the estimated sub-1G Doppler phase rotation can be used to accurately infer the C-band Doppler phase rotation.

Remark 2: Since p_D and q_D are typically larger than 1, the solution spaces of n_{D,l_p} and \bar{n}_{D,l_p} in $\mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p})$ are usually smaller than \mathbb{N}_+ , implying that the system can still benefit from the dual-band information when the phase ambiguity exists in both C-band and sub-1G band. Meanwhile, as the numbers of integer cycles of dual-band Doppler phase rotations are limited owing to finite UE velocity in practical scenarios, the feasible solutions of n_{D,l_p} and \bar{n}_{D,l_p} in $\mathcal{S}_d(n_{D,l_p}, \bar{n}_{D,l_p})$ are typically very limited. For instance, when $f_c = 3.5$ GHz, $\bar{f}_c = 0.9$ GHz, $T_e = 40$ ms, and $\bar{T}_e = 5$ ms, the fraction $\frac{f_c T_e}{f_c \bar{T}_e} = \frac{3.5 \times 40}{0.9 \times 5} = \frac{280}{9}$, thus p_D and q_D are 280 and 9, respectively. This implies that the dual-band system encounters the phase ambiguity only if $n_{D,l_p} \geq \text{mod}(n_{D,l_p}, p_D) + p_D \geq 280$ according to (12). However, $n_{D,l_p} \geq 280$ means the UE velocity $v \geq 2160$ km/h according to (4), which is impossible in practical scenarios.

B. Problem Formulation

We assume that the sub-1G estimation period \bar{T}_e is $1/\Gamma$ of the C-band estimation period T_e , where $\Gamma \in \mathbb{N}_+$ for simplicity. This means that each C-band channel vector $\tilde{\mathbf{h}}_j$ corresponds to a sub-1G channel sequence $\left\{ \tilde{\mathbf{h}}_{(j-1,\gamma)} \right\}_{\gamma=1}^{\Gamma}$ with length Γ , wherein $\tilde{\mathbf{h}}_{(j-1,\gamma)} \in \mathbb{C}^{\bar{M} \times 1}$ represents the estimated sub-1G channel vector at time $(j-1)T_e + \gamma\bar{T}_e$. The objective of this paper is to utilize the channel sequence of dual bands $\left\{ \tilde{\mathbf{h}}_j, \left\{ \tilde{\mathbf{h}}_{(j-1,\gamma)} \right\}_{\gamma=1}^{\Gamma} \right\}_{j=1}^J$ for accurate C-band super-resolution prediction of $\left\{ \hat{\mathbf{h}}_{(j,p)} \right\}_{p=1}^P$.

Due to the complex multi-path effect and noise in practical wireless communication systems, accurately extracting the Doppler phase rotation features of each path from estimated channels for super-resolution prediction is challenging for conventional channel prediction schemes [12], [14], [18], [20]. Considering that DL enjoys strong fitting capabilities to capture intricate correlations, we propose to utilize DL to effectively model the dual-band fusion-based super-resolution channel prediction.

C. CNN-LSTM-Based Dual-Band Fusion Model

The proposed DL-based dual-band fusion model can be expressed as

$$\left\{ \mathbf{u}_j \right\}_{j=1}^J = \left\{ f_d \left(\tilde{\mathbf{h}}_j, \left\{ \tilde{\mathbf{h}}_{(j-1,\gamma)} \right\}_{\gamma=1}^{\Gamma}; \Omega_d \right) \right\}_{j=1}^J, \quad (15)$$

where \mathbf{u}_j is the fused feature vector at time jT_e , and $f_d(\cdot)$ is the dual-band fusion model with parameters Ω_d . We use CNN and LSTM to fit $f_d(\cdot)$, as depicted in Fig. 3, which consists of two components: C-band feature extraction model and sub-1G feature extraction model. We now detail them one by one.

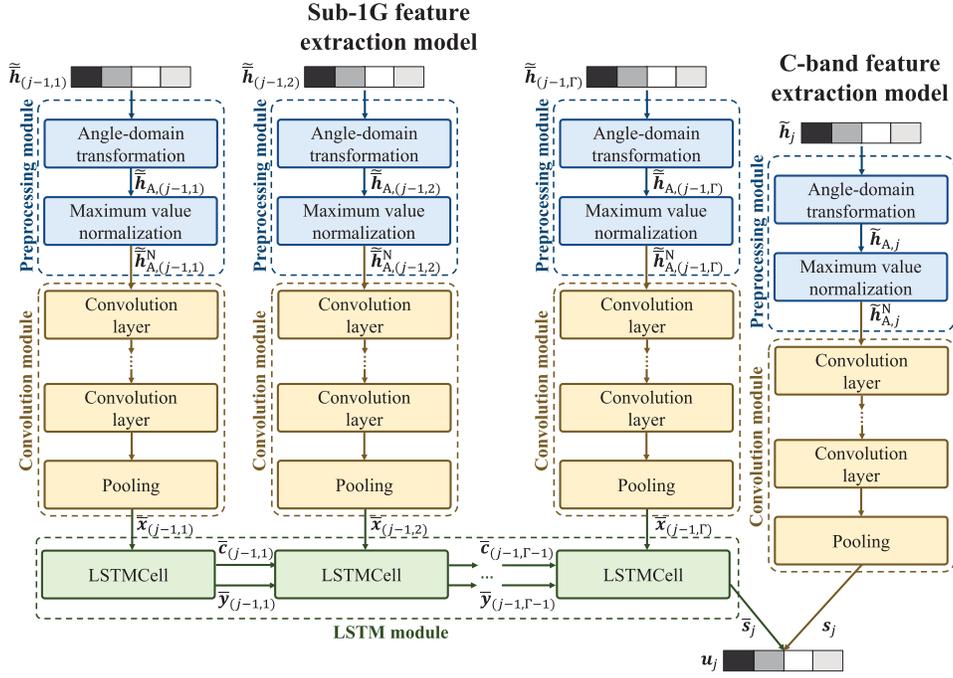


Fig. 3. Structure of proposed CNN-LSTM-based dual-band fusion model.

1) *C-Band Feature Extraction Model:* The C-band feature extraction can be partitioned into the preprocessing module and the convolution module. Specifically, in the preprocessing module, since the angle-domain channel usually presents the sparsity that is beneficial to take full advantage of DL [12], [51], [52], the C-band channel vector \tilde{h}_j is transformed to the angle domain from the spatial domain via discrete Fourier transform (DFT), which can be written as

$$\tilde{h}_{A,j} = \mathbf{F}_A^H \tilde{h}_j, \quad (16)$$

where $\tilde{h}_{A,j} \in \mathbb{C}^{M \times 1}$ denotes the corresponding angle-domain channel vector, $\mathbf{F}_A = \text{diag}(\mathbf{B}, \mathbf{B}) \in \mathbb{C}^{M \times M}$ is the angle-domain unitary matrix for dual-polarized directions, and $\mathbf{B} \in \mathbb{C}^{M/2 \times M/2}$ is the DFT matrix. Furthermore, due to the large dynamic ranges of $\tilde{h}_{A,j}$, the maximum value normalization is adopted to obtain $\tilde{h}_{A,j}^N = \frac{\tilde{h}_{A,j}}{\|\tilde{h}_{A,j}\|_\infty}$. In addition, the complex-value elements of $\tilde{h}_{A,j}^N$ are split into their real and imaginary parts, which are then stacked on two feature channels for subsequent feature extraction.

In the convolution module, multiple convolution layers are sequentially employed to capture the C-band features, where each convolution layer is followed by a batch normalization layer and a nonlinear activation layer. After the last convolution layer, a pooling layer is used to reduce the feature space, where the output is the ultimate C-band feature vector s_j corresponding to time jT_e .

2) *Sub-1G Feature Extraction Model:* Similar to the C-band feature extraction, the spatial-domain sub-1G channel vector $\tilde{h}_{(j-1,\gamma)}$ is transformed to the angle domain and normalized. Then, the convolution module is employed to capture the sub-1G features, where the output is denoted as $\bar{x}_{(j-1,\gamma)}$. Next, $\bar{x}_{(j-1,\gamma)}$ is used by LSTM to extract the Doppler

phase rotation features of sub-1G channel sequence, where the process in the basic LSTM cell can be expressed as

$$\left\{ \bar{y}_{(j-1,\gamma)}, \bar{c}_{(j-1,\gamma)} \right\} = \text{LSTMCell} \left(\bar{x}_{(j-1,\gamma)}, \left\{ \bar{y}_{(j-1,\gamma-1)}, \bar{c}_{(j-1,\gamma-1)} \right\} \right). \quad (17)$$

The pair of state vectors $\left\{ \bar{y}_{(j-1,\gamma)}, \bar{c}_{(j-1,\gamma)} \right\}$ represents the hidden state and the memory cell corresponding to time $(j-1)T_e + \gamma T_e$, respectively.

Finally, the hidden state of the last LSTM cell $\bar{y}_{(j-1,\Gamma)}$ corresponding to time $(j-1)T_e + \Gamma T_e = jT_e$ is used as the sub-1G feature vector \bar{s}_j , which is further concatenated with the C-band feature vector s_j to obtain the fused feature vector u_j at time jT_e for sequential super-resolution channel prediction.

V. NEURAL ODE WITH MODULATED-PERIODIC-BASED MLP FOR SUPER-RESOLUTION CHANNEL PREDICTION

Based on the above fused feature series $\{u_j\}_{j=1}^J$, the super-resolution channel prediction can be expressed as

$$\left\{ \hat{h}_{(J,p)} \right\}_{p=1}^P = f_s \left(\{u_j\}_{j=1}^J; \Omega_s \right), \quad (18)$$

where $f_s(\cdot)$ is the super-resolution channel prediction model with parameters Ω_s . Note that the sampling interval of input sequence $\{u_j\}_{j=1}^J$ corresponds to the estimation period T_e , while the sampling interval of output sequence $\left\{ \hat{h}_{(J,p)} \right\}_{p=1}^P$ is the prediction period T_p . We introduce the neural ODE architecture for channel prediction at a finer temporal resolution. Then, by analyzing the physics features of channel dynamics in high-mobility scenarios, we propose the modulated-periodic-based MLP to enhance prediction performance.

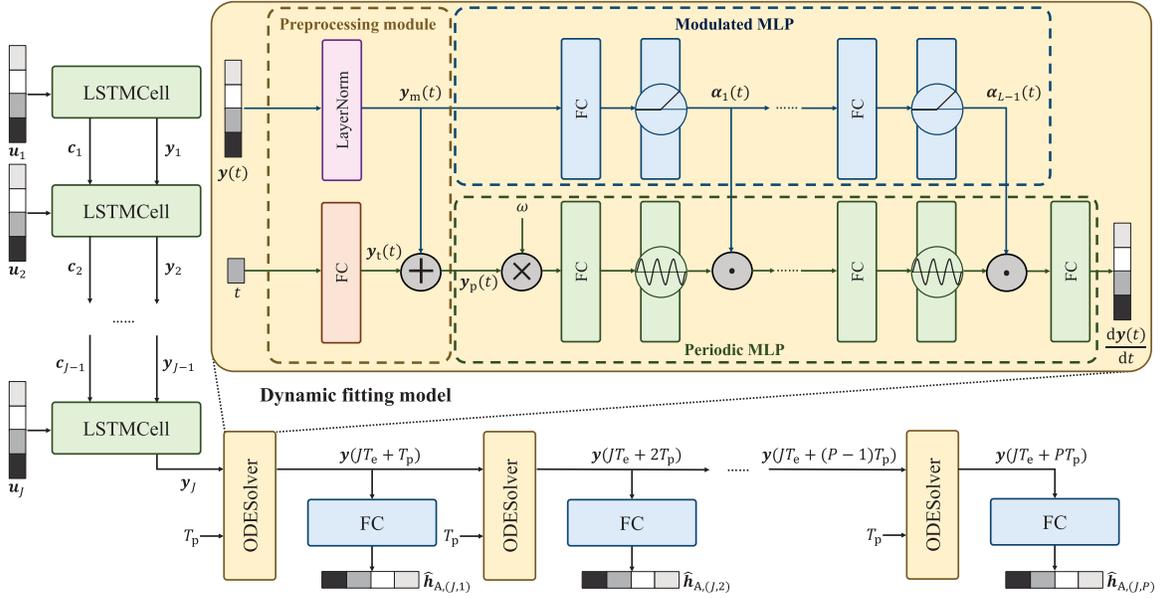


Fig. 4. Proposed neural ODE architecture with modulated-periodic-based MLP for super-resolution channel prediction.

A. Neural ODE Architecture

To fit $f_s(\cdot)$, the existing works [14], [15], [16], [17], [18] usually adopted the classic RNN architecture to extract the temporal correlations of $\{\mathbf{u}_j\}_{j=1}^J$. Taking LSTM as an example, define $\{\mathbf{y}_j, \mathbf{c}_j\}$ as the hidden state and the memory cell, respectively, with the corresponding state transfer similar to (17). However, since $\{\mathbf{y}_j, \mathbf{c}_j\}_{j=1}^J$ is updated based on $\{\mathbf{u}_j\}_{j=1}^J$, the state transfer in LSTM only operates at a single temporal resolution T_e , so that the conventional LSTM cannot accurately model channel prediction at a finer temporal resolution T_p .

To address this problem, the neural ODE, as a continuous-time learning architecture, can be used to effectively fit $f_s(\cdot)$ [21], [22], [23], [24]. The neural ODE learns the dynamics of continuous-time state $\mathbf{y}(t)$ using neural networks, which naturally incorporate the state transfer in arbitrary temporal resolutions. Mathematically, the neural ODE can be expressed as

$$\frac{d\mathbf{y}(t)}{dt} = g(\mathbf{y}(t), t; \Omega_0), \quad (19)$$

where $g(\cdot)$ denotes the dynamic fitting model with parameters Ω_0 . If the initial state $\mathbf{y}(t_0)$ is obtained, the state $\mathbf{y}(t)$ at time t can be inferred by the ODE solver as follows

$$\mathbf{y}(t) = \int_{t_0}^t g(\mathbf{y}(\tau), \tau; \Omega_0) d\tau + \mathbf{y}(t_0), \quad (20)$$

where the integrator can be implemented by various numerical methods, such as Euler method and Runge-Kutta method.

To obtain a robust initial state $\mathbf{y}(t_0)$, the neural ODE can be integrated with LSTM, as shown in Fig. 4. Specifically, LSTM is first employed to adaptively capture the temporal correlations of $\{\mathbf{u}_j\}_{j=1}^J$, where its hidden state \mathbf{y}_J at time JT_e is utilized as the initial state $\mathbf{y}(t_0)$. Based on $\mathbf{y}(t_0)$, the state transfer at a finer temporal resolution T_p is realized by

the ODE solver (20). Moreover, the cascaded ODE solvers are used to jointly predict $\{\hat{\mathbf{h}}_{A,(J,p)}\}_{p=1}^P$ with length P , where the output state of the p -th ODE solver $\mathbf{y}(t_0 + pT_p)$ serves as the initial state of the $(p+1)$ -th ODE solver. This initial state $\mathbf{y}(t_0 + pT_p)$, together with T_p , which is the duration required to characterize the channel dynamics and the interval of numerical integration, is utilized to predict $\mathbf{y}(t_0 + (p+1)T_p)$. These P ODE solvers share the same parameters due to their identical temporal resolution, effectively reducing model storage overhead. Then, P FC layers with the same parameters are exploited to transform the state sequence $\{\mathbf{y}(t_0 + pT_p)\}_{p=1}^P$ to the corresponding predicted angle-domain channel sequence $\{\hat{\mathbf{h}}_{A,(J,p)}\}_{p=1}^P$, where $\hat{\mathbf{h}}_{A,(J,p)} \in \mathbb{C}^{M \times 1}$ is the predicted angle-domain channel vector at time $JT_e + pT_p$. Finally, $\{\hat{\mathbf{h}}_{A,(J,p)}\}_{p=1}^P$ can be acquired by applying the inverse transform \mathbf{F}_A on (16).

B. Physics Feature-Inspired MLP With Modulated-Periodic Activation Function

For the neural ODE, the design of dynamic fitting model $g(\cdot)$ plays a vital role in learning $\frac{d\mathbf{y}(t)}{dt}$. Considering that $\frac{d\mathbf{y}(t)}{dt}$ corresponds to the temporal channel dynamics $\frac{d\mathbf{h}(t)}{dt}$, the physics features of $\frac{d\mathbf{h}(t)}{dt}$ are investigated for designing a dedicated dynamic fitting model.

1) *Physics Features of Channel Dynamics in High-Mobility Scenarios:* According to (1), the m -th element of $\frac{d\mathbf{h}(t)}{dt}$ can be expressed as

$$\begin{aligned} & \frac{d[\mathbf{h}(t)]_m}{dt} \\ &= \frac{d \sum_{l_p=1}^{L_p} \alpha_{l_p} e^{j2\pi \frac{(\tau_{lx,l_p})^T \mathbf{d}_{lx,m}}{\lambda}} e^{j2\pi f_{d,l_p} t} e^{-j2\pi f_c \tau_{l_p}}}{dt} \end{aligned}$$

$$\begin{aligned}
&= \sum_{l_p=1}^{L_p} \alpha_{l_p} e^{j2\pi \frac{(\tau_{ix,l_p})^T \mathbf{d}_{ix,m}}{\lambda}} \frac{d e^{j2\pi f_{D,l_p} t}}{dt} e^{-j2\pi f_c \tau_{l_p}} \\
&= \sum_{l_p=1}^{L_p} j2\pi f_{D,l_p} \alpha_{l_p} e^{j2\pi \frac{(\tau_{ix,l_p})^T \mathbf{d}_{ix,m}}{\lambda}} e^{j2\pi f_{D,l_p} t} e^{-j2\pi f_c \tau_{l_p}} \\
&= \sum_{l_p=1}^{L_p} |2\pi f_{D,l_p} \alpha_{l_p}| e^{j \left(\angle(j\alpha_{l_p}) + 2\pi \left(\frac{(\tau_{ix,l_p})^T \mathbf{d}_{ix,m}}{\lambda} + f_{D,l_p} t - f_c \tau_{l_p} \right) \right)}. \tag{21}
\end{aligned}$$

The magnitude part $F_{l_p}(\cdot)$ and the phase part $G_{l_p}(\cdot)$ in $\frac{d\mathbf{h}(t)}{dt}$ can be written respectively as

$$F_{l_p}(\cdot) = |2\pi f_{D,l_p} \alpha_{l_p}|, \tag{22}$$

$$G_{l_p}(\cdot) = e^{j \left(\angle(j\alpha_{l_p}) + 2\pi \left(\frac{(\tau_{ix,l_p})^T \mathbf{d}_{ix,m}}{\lambda} + f_{D,l_p} t - f_c \tau_{l_p} \right) \right)}. \tag{23}$$

Since $F_{l_p}(\cdot)$ is determined by slowly time-varying channel parameters f_{D,l_p} and α_{l_p} , $F_{l_p}(\cdot)$ changes slowly over time. These slowly time-varying parameters can be extracted from the historical channel sequence, making $F_{l_p}(\cdot)$ a function of the state $\mathbf{y}(t)$ that contains the information of historical channel sequence. By contrast, due to the drastic changing Doppler phase rotation $2\pi f_{D,l_p} t$, $G_{l_p}(\cdot)$ is a rapidly time-varying and periodic function of the state $\mathbf{y}(t)$ and time t .

In the existing neural ODE-based channel prediction schemes [21], [22], [23], [24], [25], [26], [27], tanh-, sigmoid- or ReLU-based MLP is used to construct the dynamic fitting model for learning $\frac{d\mathbf{h}(t)}{dt}$. However, the rapidly time-varying and periodic features of phase part $G_{l_p}(\cdot)$ pose significant challenges to the learning of $\frac{d\mathbf{h}(t)}{dt}$ for all tanh-, sigmoid- and ReLU-based MLPs for two reasons. First, since tanh, sigmoid and ReLU can all be regarded as piecewise linear, the fitting mechanism of neural networks that use them as nonlinear activations is essentially a form of piecewise linear approximation [53], [54], [55], [56], [57]. Furthermore, because the magnitude of the Fourier transform of this piecewise linear approximation-based fitted function decreases as the change rate of fitted function with its input increases (see Appendix A for a detailed analysis), neural networks with piecewise linear activations can effectively learn the slowly time-varying magnitude part $F_{l_p}(\cdot)$, but they are insufficient to approximate the rapidly time-varying phase part $G_{l_p}(\cdot)$. Second, due to their monotonic nature, it is difficult for MLPs based on tanh, sigmoid and ReLU to accurately fit the periodic phase part $G_{l_p}(\cdot)$ in channel dynamics.

2) *Modulate-Periodic-Based MLP*: To address the above problems, we propose a modulated-periodic-based MLP to effectively learn channel dynamics in high-mobility scenarios, as depicted in Fig. 4, which consists of a preprocessing module, a periodic-based MLP and a modulated-based MLP. In the preprocessing module, the layer normalization is applied on $\mathbf{y}(t) \in \mathbb{R}^{n_o \times 1}$ to speed up training. The normalized vector, denoted as $\mathbf{y}_m(t) \in \mathbb{R}^{n_o \times 1}$, is fed to the modulated MLP for learning $F_{l_p}(\cdot)$. A FC layer is adopted for time t to scale up its dimension, yielding the output vector $\mathbf{y}_l(t) \in \mathbb{R}^{n_o \times 1}$. The sum of $\mathbf{y}_l(t)$ and $\mathbf{y}_m(t)$ is utilized as the input vector $\mathbf{y}_p(t) \in \mathbb{R}^{n_o \times 1}$ for the periodic MLP to fit $G_{l_p}(\cdot)$. Different

from the traditional neural ODE, therefore, two parallel MLPs are employed to effectively learn channel dynamics $\frac{d\mathbf{h}(t)}{dt}$.

2.1) The periodic MLP is dedicated to the fitting of rapidly time-varying and periodic $G_{l_p}(\cdot)$, which uses sine function as the nonlinear activation for two reasons. First, when the values of the weight matrix in the MLP that define the frequency of sine function are large, sine function is a rapidly time-varying function, thereby enhancing its fitting capability of rapidly time-varying feature in $G_{l_p}(\cdot)$. Second, as sine function is inherently periodic, it can naturally model the periodic feature in $G_{l_p}(\cdot)$. Specifically, the periodic MLP is defined by

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{W}_L (\psi_{L-1} \circ \psi_{L-2} \circ \dots \circ \psi_1) (\mathbf{y}_p(t)) + \mathbf{b}_L, \tag{24}$$

$$\psi_l(\mathbf{y}_{p,l}(t)) = \alpha_l(t) \odot \sin(\mathbf{W}_l \mathbf{y}_{p,l}(t) + \mathbf{b}_l), \tag{25}$$

where L is the number of FC layers, $\psi_l(\cdot)$ represents the l -th layer of the periodic MLP, which consists of a FC layer with the weight matrix $\mathbf{W}_l \in \mathbb{R}^{n_l \times n_{l-1}}$, the bias vector $\mathbf{b}_l \in \mathbb{R}^{n_l \times 1}$ and the following sine function $\sin(\cdot)$, and $\mathbf{y}_{p,l}(t) \in \mathbb{R}^{n_{l-1} \times 1}$ is the corresponding input vector with $\mathbf{y}_{p,1}(t) = \mathbf{y}_p(t)$. Inspired by the modulated magnitude sine function form of $\frac{d\mathbf{h}(t)}{dt}$ in (21), sine function is modulated by the magnitude parameter $\alpha_l(t) \in \mathbb{R}^{n_l \times 1}$, which models the magnitude part $F_{l_p}(\cdot)$.

However, due to the periodic feature of sine function, the training of MLP with sine activations often converges to undesired local minima. To address this problem, we adopt a principled initialization scheme in [58] to ensure that the input distribution before sine activation, $\mathbf{W}_l \mathbf{y}_{p,l}(t) + \mathbf{b}_l$, mainly falls within $[-\pi/2, \pi/2]$ for $l \in \{2, 3, \dots, L-1\}$, which can be written as

$$[\mathbf{W}_l]_{i,j} \sim \mathcal{U}(-\sqrt{6/n_{l-1}}, \sqrt{6/n_{l-1}}), l \in \{1, 2, \dots, L-1\}. \tag{26}$$

Additionally, to facilitate the fitting of rapidly time-varying and periodic feature in $G_{l_p}(\cdot)$, the input vector $\mathbf{y}_p(t)$ needs to be multiplied by a large hyperparameter ω to ensure that $\sin(\mathbf{W}_1 \omega \mathbf{y}_p(t) + \mathbf{b}_1)$ spans multiple integer cycles, becoming a rapidly varying and periodic function of $\mathbf{y}_p(t)$.

Remark 3: The study [58] demonstrated that the output vector of the first layer, $\mathbf{y}_{p,2}(t)$, follows the arcsine distribution $\text{Arcsin}(-1, 1)$. By using the initialization scheme (26) for the weight matrix of the second layer \mathbf{W}_2 , the dot product of \mathbf{W}_2 and $\mathbf{y}_{p,2}(t)$, $\sum_{j=1}^{n_1} [\mathbf{W}_2]_{i,j} [\mathbf{y}_{p,2}(t)]_j$, follows the normal distribution according to the central limit theorem. Then, through numerical approximations, the normal distribution through the sine activations yields again the arcsine distribution $\text{Arcsin}(-1, 1)$. By analogy, the dot products of \mathbf{W}_l and $\mathbf{y}_{p,l}(t)$ for $l \in \{2, 3, \dots, L-1\}$ all follow the normal distribution by using the initialization scheme (26) in each layer. Thus $\mathbf{W}_l \mathbf{y}_{p,l}(t) + \mathbf{b}_l$ mainly falls within $[-\pi/2, \pi/2]$, solving the training difficulty of sine-based MLP.

2.2) The magnitude parameter $\alpha_l(t)$ is learned by the modulated MLP that fits the magnitude part $F_{l_p}(\cdot)$. Since $F_{l_p}(\cdot)$ is a slowly time-varying function, the modulated MLP uses piecewise linear ReLU as the nonlinear activation. Consequently, the modulated MLP can be written as

$$\alpha_l(t) = \psi'_l \circ \psi'_{l-1} \circ \dots \circ \psi'_1(\mathbf{y}_m(t)), \tag{27}$$

$$\psi'_l(\mathbf{y}_{m,l}(t)) = \text{ReLU}(\mathbf{W}'_l \mathbf{y}_{m,l}(t) + \mathbf{b}'_l), \quad (28)$$

where $\psi'_l(\cdot)$ is the l -th layer of modulated MLP, which consists of a FC layer with the weight matrix $\mathbf{W}'_l \in \mathbb{R}^{n_l \times n_{l-1}}$, the bias vector $\mathbf{b}'_l \in \mathbb{R}^{n_l \times 1}$ and the ReLU activation function, and $\mathbf{y}_{m,l}(t) \in \mathbb{R}^{n_{l-1} \times 1}$ is the input vector with $\mathbf{y}_{m,1}(t) = \mathbf{y}_m(t)$.

C. Model Training and Prediction

Let $\widehat{\mathbf{H}}_{(J,p)} = [\widehat{\mathbf{h}}_{(J,p)}^{(1)}, \widehat{\mathbf{h}}_{(J,p)}^{(2)}, \dots, \widehat{\mathbf{h}}_{(J,p)}^{(K)}] \in \mathbb{C}^{M \times K}$ denote the predicted C-band channel sequence of K UEs at time $JT_e + pT_p$. Based on $\widehat{\mathbf{H}}_{(J,p)}$ and the corresponding perfect channel matrix $\mathbf{H}_{(J,p)} = [\mathbf{h}_{(J,p)}^{(1)}, \mathbf{h}_{(J,p)}^{(2)}, \dots, \mathbf{h}_{(J,p)}^{(K)}] \in \mathbb{C}^{M \times K}$, the spectral efficiency $R_{(J,p)}$ can be calculated according to (9). To optimize the proposed DL models, we adopt the negative average spectral efficiency widely used in multi-user MIMO systems [59], [60] as the loss function, which can be expressed as

$$\text{loss} = -R_{\text{avg}}, \quad (29)$$

$$R_{\text{avg}} = \frac{1}{P} \sum_{p=1}^P R_{(J,p)}, \quad (30)$$

where R_{avg} denotes the average spectral efficiency. Note that $\mathbf{H}_{(J,p)}$ can be obtained by channel estimation with the finer temporal resolution T_p and long pilot sequence to ensure high signal-to-noise ratio (SNR).

The proposed dual-band super-resolution channel prediction scheme consists of two stages: training and prediction. In the training stage, conventional channel estimation schemes are employed to periodically collect the dual-band channels, where the estimation periods of C-band and sub-1G are T_p and \bar{T}_e , respectively. To construct a training data sample, the historical C-band channel sequence with temporal resolution T_e can be obtained by downsampling the estimated C-band channel sequence with temporal resolution T_p by a factor of T_e/T_p . This, together with the estimated sub-1G channel sequence, jointly serves as the input. The corresponding future C-band channel sequence with temporal resolution T_p is used as the label. Based on the constructed training data set, the loss function (29) is calculated to optimize the model parameters. Once the model has been sufficiently trained, it switches to the prediction stage. At this stage, conventional channel estimation schemes measure the C-band and sub-1G channel sequences with estimation periods T_e and \bar{T}_e , respectively. The estimated C-band and sub-1G channel sequences are jointly utilized to predict the future C-band channel sequence with a finer temporal resolution T_p . Since T_e is generally much larger than T_p , the pilot overhead of C-band channel estimation can be significantly reduced.

VI. SIMULATION STUDY

A. Simulation Setup

We consider a multi-user MIMO system, where a BS serves $K = 5$ single-antenna UEs. In our simulations, the channel model of 3GPP TR 38.901 for urban macro-cell scenarios is used to generate the data set [39]. In this channel model, line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios are

TABLE II
DUAL-BAND CHANNEL PARAMETERS

Parameters	C-band	Sub-1G
Center frequency f_c, \bar{f}_c	3.5 GHz	0.9 GHz
Number of BS horizontal antennas M_h, \bar{M}_h	4	4
Number of BS vertical antennas M_v, \bar{M}_v	4	4
Number of BS antennas M, \bar{M}	32	32
Number of paths L_p, \bar{L}_p	12	12
Delay spread $\tau_s, \bar{\tau}_s$	98.3 ns	112.0 ns
Estimation period T_e, \bar{T}_e	40 ms	5 ms
Prediction period T_p, \bar{T}_p	5 ms	/
Estimation SNR, SNR	5 dB	5 dB

both included, where the LOS probability is calculated based on the 2D distance d_{2D} between BS and UE together with the UE height h_{UE} . The complex channel gain α_{l_p} is primarily determined by the path loss and initial phase. More details can be found in Subsection 7.5 of [39]. In the LOS scenario, the path loss is calculated as

$$\text{PL}_{\text{LOS}}[\text{dB}] = 28 + 22\log_{10}(d_{3D}) + 20\log_{10}(f_c), \quad (31)$$

where d_{3D} is the 3D distance between BS and UE and recall that f_c is the carrier frequency. In the NLOS scenario, the path loss is expressed as

$$\begin{aligned} \text{PL}_{\text{NLOS}} &= \max(\text{PL}_{\text{LOS}}, \text{PL}'_{\text{NLOS}}), \\ \text{PL}'_{\text{NLOS}}[\text{dB}] &= 13.54 + 39.08\log_{10}(d_{3D}) \\ &\quad + 20\log_{10}(f_c) - 0.6(h_{\text{UE}} - 1.5). \end{aligned} \quad (32)$$

The initial phase is uniformly distributed within $(-\pi, \pi)$ [40]. Moreover, the channel delay τ_{l_p} is calculated as

$$\tau_{l_p} = -1.5\tau_s \log(X_{l_p}), \quad (34)$$

where τ_s is the delay spread and $X_{l_p} \sim \mathcal{U}(0, 1)$. In our simulations, the cell radius is 300 m, and the BS height is set to 25 m [41]. The specific dual-band parameters are listed in Table II. Unless otherwise stated, Gaussian distribution is adopted to model dual-band angular deviations with $\Delta\phi_{\text{AOA}} \sim \mathcal{N}(0, \sigma_{\text{AOA}}^2)$ and $\Delta\theta_{\text{ZOA}} \sim \mathcal{N}(0, \sigma_{\text{ZOA}}^2)$ [31], where σ_{AOA} and σ_{ZOA} both are set to 5° . The default UE velocity is set to $v = 80$ km/h. The default lengths of historical and predicted channel sequences are set to $J = 9$ and $P = 7$. To calculate the C-band spectral efficiency of downlink transmission, the BS transmit power ρ_{BS} is set to 35 dBm, the noise factor is 5 dB, and the zero-forcing precoding is employed for simultaneously serving all UEs [61].

The parameters of the proposed DL models are specified in Table III. Here, L_c and L_l are the numbers of convolution layers and LSTM layers, respectively, while f_i and f_o denote the numbers of input and output feature channels. We adopt a (3, 3, 1) CNN structure, i.e., the kernel size, stride, and padding size of the convolution layer are set to 3, 3 and 1. LReLU denotes the leaky ReLU. The layer normalization is used before LSTM to accelerate convergence, while the dropout strategy is exploited in the FC layer to avoid overfitting. For the ODE solver, the Euler method [25], [27] is selected as the integrator. ‘‘Preprocessing FC’’ denotes the FC layer for scaling up the dimension of time t . To ensure that

TABLE III
PROPOSED DL MODELS

Models	Layers	Structures
C-band feature extraction model	Convolution	$L_c = 3, f_i = 2, f_o = 128, (3, 3, 1), \text{BatchNorm, LReLU}$
	Pooling	$f_i = 128, f_o = 128, \text{average-pooling}$
Sub-1G feature extraction model	Convolution	$L_c = 3, f_i = 2, f_o = 128, (3, 3, 1), \text{BatchNorm, LReLU}$
	Pooling	$f_i = 128, f_o = 128, \text{average-pooling}$
	LSTM	$L_1 = 1, f_i = 128, f_o = 128, \text{LayerNorm}$
ODE-LSTM model	LSTM	$L_1 = 1, f_i = 256, f_o = 256, \text{LayerNorm}$
	FC	$f_i = 256, f_o = 2N_t, \text{dropout} = 0.02$
Dynamic fitting model	Preprocessing FC	$f_i = 1, f_o = 256$
	Periodic MLP	$L = 4, f_i = 256, f_o = 1024, \text{Sine}$
	Modulated MLP	$L = 3, f_i = 256, f_o = 1024, \text{ReLU}$

$\sin(\mathbf{W}_1 \omega \mathbf{y}_p(t) + \mathbf{b}_1)$ spans multiple integer cycles to fit the rapidly time-varying function, the hyperparameter ω is set to 30 [58]. The training data set and the test data set consist of 48,640 and 2,560 samples, respectively. The batch size is set to 64. The models are trained for 200 epochs using the Adam optimizer with the initial learning rate $\text{lr} = 1 \times 10^{-5}$.

In our simulation, we compare the following schemes:

- **Proposed dual-band fusion:** Use CNN to extract dual-band spatial features, and then combine LSTM with the neural ODE using modulated-periodic-based MLP to learn temporal channel dynamics.
- **Proposed without sub-1G:** Use CNN to extract C-band spatial features, and then combine LSTM with the neural ODE using modulated-periodic-based MLP to learn temporal channel dynamics.
- **ODE-LSTM:** Use CNN to extract C-band spatial features, and then combine LSTM with the neural ODE to learn temporal channel dynamics [27].
- **CNN-LSTM:** Use CNN to extract C-band spatial features and LSTM to capture temporal correlations [17].

For fair comparisons, CNN and LSTM in all the schemes adopt the same structure.

B. Investigation of Proposed Schemes

1) *Investigation of CNN Structure:* First, to ensure the satisfactory performance with a relatively small model complexity, the average spectral efficiency convergence performance of our proposed dual-band fusion scheme with respect to the numbers of CNN layers L_c and output feature channels f_o is investigated in Fig. 5. Observe that the proposed dual-band fusion models with different CNN structures all converge within 100 epochs. It can also be seen that the average spectral efficiency improves when L_c increases from 2 to 3 and f_o increases from 64 to 128, as deeper and wider neural networks have stronger fitting capability. Further increasing L_c and f_o does not improve the performance but imposes higher model complexity. Thus, the CNN structure

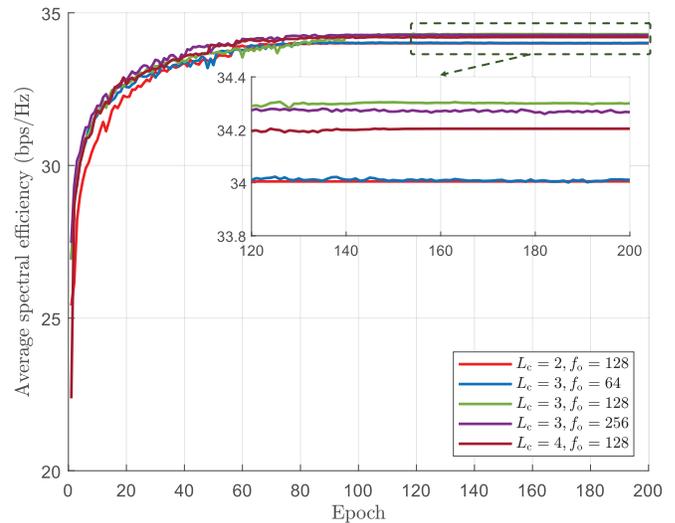


Fig. 5. Convergence performance comparison of proposed dual-band fusion scheme with different CNN structures.

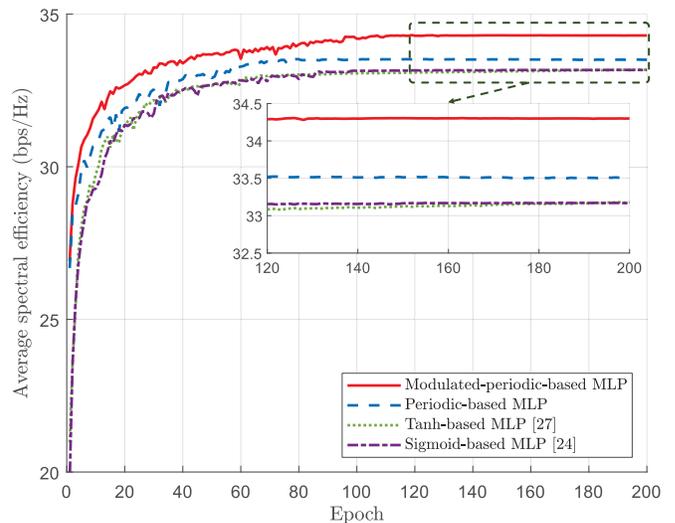


Fig. 6. Convergence performance comparison of our proposed scheme with different dynamic fitting models.

with $L_c = 3$ and $f_o = 128$ is appropriate for our dual-band fusion scheme.

2) *Investigation of Dynamic Fitting Model:* We further investigate the convergence performance of our proposed scheme with those of different dynamic fitting models in Fig. 6. Periodic-based MLP represents that the modulated MLP is removed in the proposed modulated-periodic-based MLP. For a fair comparison, the numbers of FC layers in the periodic-based, tanh-based [27] and sigmoid-based [24] MLPs are equal to the sum of the numbers of FC layers in the modulated and periodic MLPs of our proposed modulated-periodic-based MLP. It can be seen that the performance of the periodic-based MLP is superior to those of the tanh-based and sigmoid-based MLPs, which indicates that using sine function is beneficial to effectively learning the rapidly time-varying and periodic phase part in channel dynamics. It can also be observed that our proposed modulated-periodic-based MLP outperforms the periodic-based MLP, which indicates

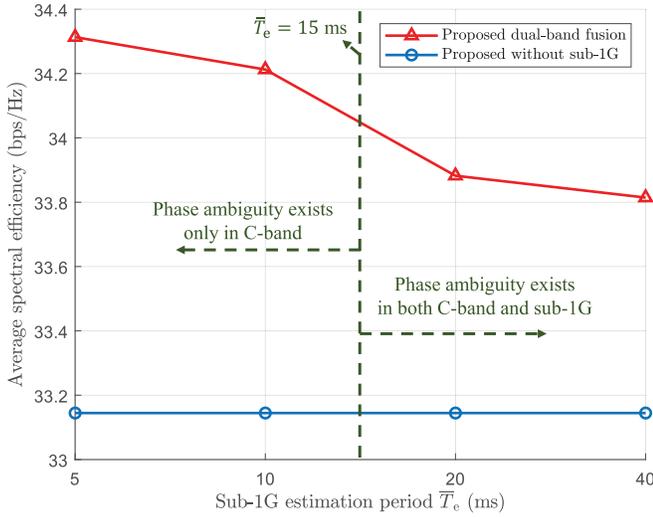


Fig. 7. Average spectral efficiency with respect to sub-1G estimation period \bar{T}_e for proposed dual-band fusion scheme.

that the modulated MLP with ReLU can efficiently fit the remaining slowly time-varying magnitude part. Therefore, the proposed DL model adopts the modulated-periodic-based MLP to construct the dynamic fitting model in the subsequent simulations.

3) *Validation of Phase Ambiguity Theorem in Dual-Band Systems*: To validate the theorem of phase ambiguity in dual-band systems given in Subsection IV-A, the impact of sub-1G estimation period \bar{T}_e on our dual-band fusion scheme is evaluated in Fig. 7, given UE velocity $v = 80$ km/h and C-band estimation period $T_e = 40$ ms. According to (4), the maximum C-band Doppler phase rotation within T_e is $20.74\pi > 2\pi$, thus the phase ambiguity exists in the C-band. Meanwhile, the estimation period for the maximum sub-1G Doppler phase rotation to span one integer cycle is 15 ms. Therefore, in Fig. 7, when \bar{T}_e is 5 ms or 10 ms, the phase ambiguity does not exist in the sub-1G band, whereas when \bar{T}_e is 20 ms or 40 ms, both the C-band and sub-1G band face the phase ambiguity issue. It can be observed that the proposed dual-band fusion scheme significantly outperforms “Proposed without sub-1G” when \bar{T}_e is 5 ms and 10 ms, since the estimated sub-1G Doppler phase rotation can be used to accurately infer the Doppler phase rotation in the C-band, as elucidated in Remark 1. As \bar{T}_e increases from 10 ms to 20 ms and 40 ms, although the proposed dual-band fusion scheme suffers from slight performance degradation, it still surpasses “Proposed without sub-1G”. This is because the feasible solutions for dual-band Doppler phase rotations are much fewer than those in single-band systems, as discussed in Remark 2.

4) *Impact of Dual-Band Angular Deviations*: Table IV investigates the impact of dual-band angular deviations $\Delta\phi_{AOA}$ and $\Delta\theta_{ZOA}$ on the average spectral efficiency performance of our proposed dual-band fusion scheme. To model $\Delta\phi_{AOA}$ and $\Delta\theta_{ZOA}$, we employ the Gaussian distribution with $\Delta\phi_{AOA} \sim \mathcal{N}(0, \sigma_{AOA}^2)$ and $\Delta\theta_{ZOA} \sim \mathcal{N}(0, \sigma_{ZOA}^2)$ as in [31], and the uniform distribution with $\Delta\phi_{AOA} \sim \mathcal{U}(-a_{AOA}, a_{AOA})$ and $\Delta\theta_{ZOA} \sim \mathcal{U}(-a_{ZOA}, a_{ZOA})$ as in [32], respectively. It can be seen that the average spectral efficiency performance decreases

TABLE IV
IMPACT OF DUAL-BAND ANGULAR DEVIATIONS

Angular deviations	Performance (bps/Hz)
$\sigma_{AOA} = \sigma_{ZOA} = 1^\circ$	34.40
$\sigma_{AOA} = \sigma_{ZOA} = 5^\circ$	34.31
$\sigma_{AOA} = \sigma_{ZOA} = 10^\circ$	34.08
$\sigma_{AOA} = \sigma_{ZOA} = 30^\circ$	33.75
$a_{AOA} = a_{ZOA} = 1^\circ$	34.42
$a_{AOA} = a_{ZOA} = 5^\circ$	34.37
$a_{AOA} = a_{ZOA} = 10^\circ$	34.29
$a_{AOA} = a_{ZOA} = 30^\circ$	33.88

as dual-band angular deviations increase, because the spatial congruence of dual bands becomes weaker. Moreover, even in the cases with $\sigma_{AOA} = \sigma_{ZOA} = 30^\circ$ or $a_{AOA} = a_{ZOA} = 30^\circ$, our proposed dual-band fusion scheme still significantly outperforms “Proposed without sub-1G”, whose performance is 33.15 bps/Hz as shown in Fig. 7.

5) *Explicit Illustration of Dual-Band Fusion to Address Phase Ambiguity*: Given the C-band estimation period $T_e = 10$ ms, sub-1G estimation period $\bar{T}_e = 5$ ms, and C-band prediction period $T_p = 5$ ms, Fig. 8 presents two data samples of true and predicted channels of the strongest angle index in the C-band, showing their real and imaginary parts. Note that 30 ms, 40 ms, and 50 ms are synchronized with the input time slots of 0 ms, 10 ms, and 20 ms, whereas 25 ms, 35 ms, and 45 ms are not. It can be observed that the predicted channel without using sub-1G information achieves smaller errors at the time slots of 30 ms, 40 ms, and 50 ms, but exhibits larger errors at the time slots of 25 ms, 35 ms, and 45 ms. This verifies that only using the C-band channel sequence for super-resolution prediction encounters the phase ambiguity issue. By contrast, the proposed dual-band fusion scheme achieves very accurate predictions at all time slots, and moreover it significantly outperforms the scheme without using sub-1G information, which demonstrates that the phase ambiguity issue is effectively addressed by dual-band fusion.

C. Performance Comparison

1) *Comparison of Complexity and Running Time*: The numbers of model parameters, numbers of floating-point operations (FLOPs), and running times for different schemes are summarized in Table V, with the running time measured on NVIDIA GeForce RTX 4090. It can be seen that the schemes using the neural ODE require more model parameters, FLOPs, and running time, because these schemes need to fit the channel dynamics and conduct predictions through numerical integration. In particular, the running time of our proposed dual-band fusion scheme, 4.283 ms, is sufficient to predict the channel sequence with the prediction period $T_p = 5$ ms. The results of Table V show that our proposed dual-band fusion scheme imposes higher complexity and running time than the existing DL-based schemes but as will be shown in the subsequent results, it achieves much more accurate channel prediction.

2) *Performance Comparison Under Different UE Velocities*: Fig. 9 shows the average spectral efficiency as the function

TABLE V
COMPLEXITY AND RUNNING TIME COMPARISON OF DIFFERENT SCHEMES

Schemes	Number of model parameters	Number of FLOPs	Running time
Proposed dual-band fusion	5.073×10^6	1.934×10^8	4.283 ms
Proposed without sub-1G	4.711×10^6	1.788×10^8	3.623 ms
ODE-LSTM [27]	3.135×10^6	1.127×10^8	3.047 ms
CNN-LSTM [17]	1.561×10^6	0.073×10^8	0.667 ms

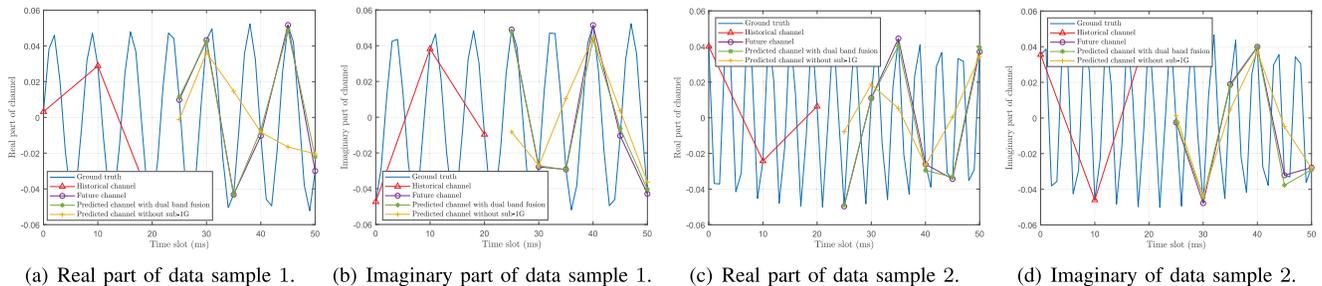


Fig. 8. Comparison of true and predicted channels in the C-band at different time slots without sub-1G information and with dual-band fusion.

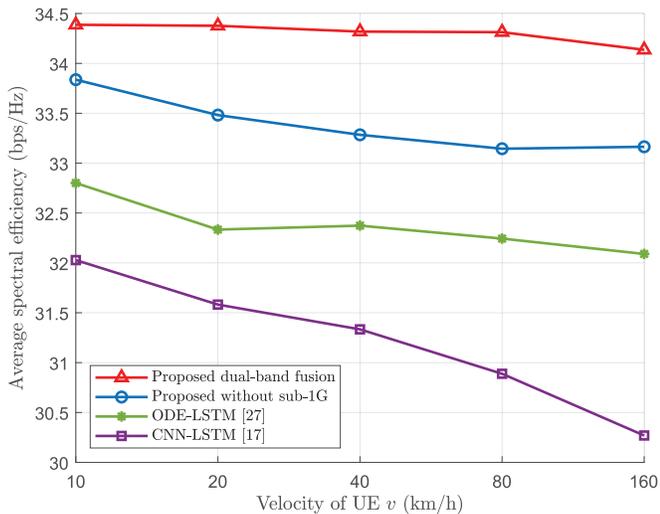


Fig. 9. Average spectral efficiency with respect to UE velocity v for different channel prediction schemes.

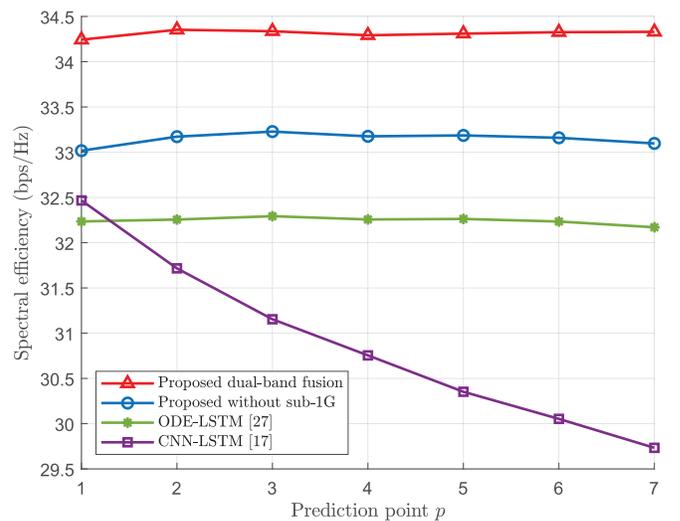


Fig. 10. Spectral efficiency at different predicted points p for different channel prediction schemes.

of UE velocity v for different channel prediction schemes. It can be seen that the CNN-LSTM baseline has the poorest performance for the super-resolution channel prediction, and its performance degrades dramatically in high-mobility scenarios. This is because the conventional LSTM is incapable of modeling channel prediction at a finer temporal resolution. It can also be seen that our proposed scheme without sub-1G information outperforms the ODE-LSTM baseline considerably, which verifies that the proposed modulated-periodic-based MLP can effectively fit channel dynamics by fully exploiting the physics features. Moreover, the proposed dual-band fusion scheme achieves much better performance than its counterpart without sub-1G feature extraction, especially in high-mobility scenarios, which validates the necessity of exploiting the low-frequency sub-1G information to address the phase ambiguity issue in the high-frequency C-band prediction. In addition, the average spectral efficiency in

$v = 160$ km/h is only 0.73% lower than that in $v = 10$ km/h for our proposed dual-band fusion scheme. This indicates that our proposed dual-band fusion scheme can maintain robust performance in high-mobility scenarios.

3) *Performance Comparison at Different Prediction Points:* Fig. 10 compares the spectral efficiency at different prediction points p for various channel prediction schemes. It can be seen that the CNN-LSTM baseline outperforms the ODE-LSTM baseline when $p=1$. This is because the prediction of CNN-LSTM is based on sequential processing techniques [16], [20], which prioritize accurately fitting the channel $\hat{\mathbf{h}}_{(J,1)}$ with $p=1$ as the input for the subsequent predictions of $\{\hat{\mathbf{h}}_{(J,p)}\}_{p=2}^P$. However, for the prediction of the entire channel sequence, the performance of CNN-LSTM is significantly inferior to that of ODE-LSTM, which verifies that LSTM cannot model channel prediction at a finer temporal resolution. Moreover, our proposed dual-band fusion scheme attains the

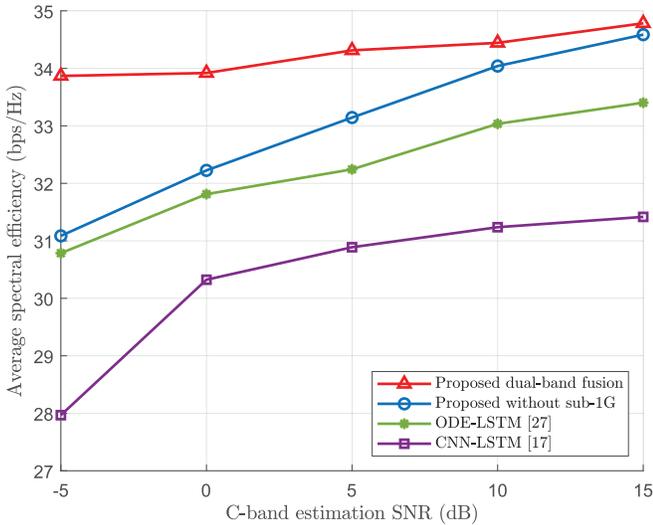


Fig. 11. Average spectral efficiency as function of C-band estimation SNR for various channel prediction schemes.

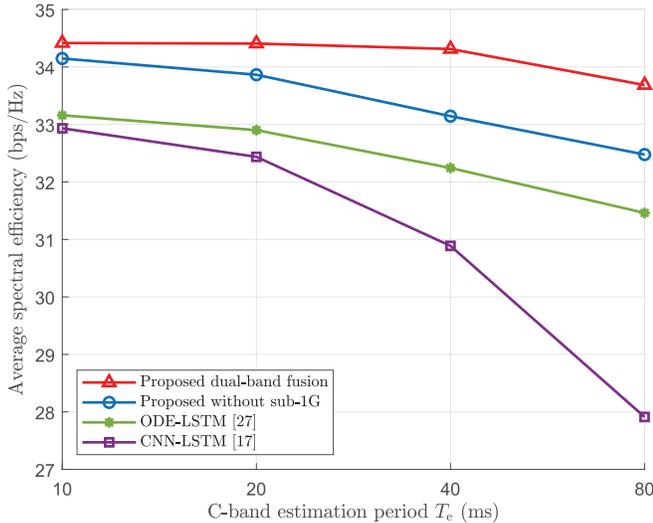


Fig. 12. Average spectral efficiency as function of C-band estimation period T_e for various channel prediction schemes.

best performance and it outperforms the ODE-LSTM by around 6%, in terms of the average spectral efficiency.

4) *Performance Comparison with Respect to C-band Estimation SNR*: Fig. 11 investigates the impact of C-band estimation SNR on the average spectral efficiency for different channel prediction schemes. It can be observed that the performance of the proposed scheme without sub-1G information approaches our proposed dual-band fusion scheme when the C-band estimation SNR is 15 dB. This is because the sub-1G estimation SNR is fixed at 5 dB, and the information of C-band channel itself plays the dominant role in channel prediction, especially at high C-band estimation SNR. Moreover, it can be seen that our dual-band fusion scheme consistently outperforms the two baselines across all C-band estimation SNRs, particularly in low-SNR scenarios. Observe that our proposed dual-band fusion scheme experiences only a 2.7% average spectral efficiency degradation when the

SNR reduces from 15 dB to -5 dB. This verifies its robust performance with respect to C-band estimation SNR.

5) *Performance Comparison with Respect to C-band Estimation Period*: Fig. 12 shows the average spectral efficiency as the function of the C-band estimation period T_e for different channel prediction schemes. As anticipated, the average spectral efficiency of all the schemes decreases as the C-band estimated period T_e increases, since the temporal correlations of channel sequence become weaker. However, it can be observed that the proposed dual-band fusion scheme markedly surpasses the two baselines, and it achieves a robust performance with respect to the C-band estimation period T_e . Specifically, the degradation of average spectral efficiency from $T_e = 10$ ms to $T_e = 80$ ms for our dual-band fusion scheme is only 2.12%, which is much smaller than 5.12% for the ODE-LSTM and 15.25% for the CNN-LSTM.

VII. CONCLUSION

In this paper, we have investigated the super-resolution channel prediction at a finer temporal resolution without additional channel estimation. Our contribution has been twofold. First, we have theoretically analyzed that phase ambiguity in high-mobility and high-frequency scenarios impedes the accurate measurement of actual Doppler phase rotation, thereby degrading the performance of super-resolution channel prediction. To address this problem, we have proposed a DL-based dual-band fusion approach to integrate the low-frequency sub-1G information, where CNN and LSTM are used to adaptively fuse the features of C-band and sub-1G channel sequences. Second, inspired by the physics features of channel dynamics in high-mobility scenarios, we have proposed the neural ODE with modulated-periodic-based MLP to enhance prediction performance. Simulation results have demonstrated that our proposed schemes outperform existing channel prediction schemes and maintain robust performance in high-mobility scenarios.

APPENDIX

A. Function Fitting Property of MLP With Piecewise Linear Activations

The following derivation is inspired by [54] and [53]. For simplicity, let us consider the scenario of a 2-layer MLP with ReLU activations but the derivation can be extended to other piecewise linear activations. The fitted function by this MLP, $f_{\text{ReLU}}(\cdot)$, can be expressed as

$$f_{\text{ReLU}}(\mathbf{x}) = \mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2, \quad (35)$$

where $\mathbf{x} \in \mathbb{R}^{n_0 \times 1}$ is the input vector, $\mathbf{W}_l \in \mathbb{R}^{n_l \times n_{l-1}}$ and $\mathbf{b}_l \in \mathbb{R}^{n_l \times 1}$ are the weight matrix and bias vector of the l -th layer, $l \in \{1, 2\}$. (35) can be further reformulated as

$$\begin{aligned} f_{\text{ReLU}}(\mathbf{x}) &= \sum_{j=1}^{n_1} [\mathbf{W}_2]_j \text{ReLU}([\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1]_j) + \mathbf{b}_2 \\ &= \sum_{j=1}^{n_1} [\mathbf{W}_{\text{equ},2}]_j [\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1]_j + \mathbf{b}_2, \end{aligned} \quad (36)$$

where $\mathbf{W}_{\text{equ},2} \in \mathbb{R}^{n_2 \times n_1}$ is the equivalent weight matrix, whose j -th column satisfies

$$[\mathbf{W}_{\text{equ},2}]_j = \begin{cases} [\mathbf{W}_2]_j, & \text{if } [\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1]_j \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (37)$$

It can be seen that when the j -th neuron $[\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1]_j$ is greater than 0, it is multiplied by a weight vector $[\mathbf{W}_{\text{equ},2}]_j$, contributing $[\mathbf{W}_{\text{equ},2}]_j [\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1]_j$ for the output $f_{\text{ReLU}}(\mathbf{x})$. Thus, the fitting mechanism of ReLU-based MLP is essentially a form of piecewise linear approximation. By analogy, the L -layer MLP with ReLU activations can be explicitly written as [53]

$$f_{\text{ReLU}}(\mathbf{x}) = \sum_{\epsilon} 1_{P_{\epsilon}}(\mathbf{x})(\mathbf{W}_{\epsilon} \mathbf{x} + \mathbf{b}_{\epsilon}), \quad (38)$$

where ϵ is the index of linear regions P_{ϵ} , and $1_{P_{\epsilon}}$ is the indicator function on P_{ϵ} , while $\mathbf{W}_{\epsilon} \in \mathbb{R}^{n_L \times n_0}$ and $\mathbf{b}_{\epsilon} \in \mathbb{R}^{n_L \times 1}$ are the corresponding weight matrix and bias vector, respectively.

Furthermore, the Fourier transform of $f_{\text{ReLU}}(\mathbf{x})$ can be formulated as [54]

$$f_{\text{ReLU},F}(\mathbf{k}) = \int f_{\text{ReLU}}(\mathbf{x}) e^{-j\mathbf{k}\mathbf{x}} d\mathbf{x} = \int \sum_{\epsilon} 1_{P_{\epsilon},F}(\mathbf{k}) \frac{\mathbf{W}_{\epsilon} \mathbf{k}}{\|\mathbf{k}\|_2^2}, \quad (39)$$

where $\mathbf{k} \in \mathbb{R}^{n_0 \times 1}$ is the frequency corresponding to \mathbf{x} and represents the change rate of $f_{\text{ReLU}}(\mathbf{x})$ with \mathbf{x} , while $f_{\text{ReLU},F}(\mathbf{k})$ and $1_{P_{\epsilon},F}(\mathbf{k})$ are the Fourier transforms of $f_{\text{ReLU}}(\mathbf{x})$ and $1_{P_{\epsilon}}(\mathbf{x})$, respectively. Note that the exact form of $1_{P_{\epsilon},F}(\mathbf{k})$ is omitted due to its complexity, but its magnitude decreases as $\|\mathbf{k}\|_2$ increases [54]. Consequently, as the frequency \mathbf{k} increases, the magnitude of $f_{\text{ReLU},F}(\mathbf{k})$ in (39) rapidly decreases.

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.
- [4] Z. Gao et al., "Compressive-sensing-based grant-free massive access for 6G massive communication," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7411–7435, Mar. 2024.
- [5] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.
- [6] Z. Xiao, H. Dong, L. Bai, P. Xia, and X.-G. Xia, "Enhanced channel estimation and codebook design for millimeter-wave communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9393–9405, Oct. 2018.
- [7] Z. Gao, S. Liu, Y. Su, Z. Li, and D. Zheng, "Hybrid knowledge-data driven channel semantic acquisition and beamforming for cell-free massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 964–979, Sep. 2023.
- [8] Y. Wang, Z. Gao, D. Zheng, S. Chen, D. Gunduz, and H. V. Poor, "Transformer-empowered 6G intelligent networks: From massive MIMO processing to semantic communication," *IEEE Wireless Commun.*, vol. 30, no. 6, pp. 127–135, Dec. 2023.
- [9] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, 2013.
- [10] C. Kong, C. Zhong, A. K. Papazafiroopoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4879–4893, Dec. 2015.
- [11] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.
- [12] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive MIMO: From spatio-temporal autoregression to deep learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1915–1930, Jul. 2021.
- [13] H. Yin, H. Wang, Y. Liu, and D. Gesbert, "Addressing the curse of mobility in massive MIMO with Prony-based angular-delay domain channel predictions," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2903–2917, Dec. 2020.
- [14] W. Jiang and H. D. Schotten, "Neural network-based fading channel prediction: A comprehensive overview," *IEEE Access*, vol. 7, pp. 118112–118124, 2019.
- [15] M. K. Shehzad, L. Rose, S. Wesemann, and M. Assaad, "ML-based massive MIMO channel prediction: Does it work on real-world data?" *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 811–815, Apr. 2022.
- [16] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2960–2973, May 2020.
- [17] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 227–236, Jan. 2020.
- [18] C. Eom and C. Lee, "Hybrid neural network-based fading channel prediction for link adaptation," *IEEE Access*, vol. 9, pp. 117257–117266, 2021.
- [19] G. Liu, Z. Hu, L. Wang, J. Xue, H. Yin, and D. Gesbert, "Spatio-temporal neural network for channel prediction in massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8003–8016, Dec. 2022.
- [20] H. Jiang, M. Cui, D. W. K. Ng, and L. Dai, "Accurate channel prediction based on transformer: Making mobility negligible," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2717–2732, Sep. 2022.
- [21] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 6572–6583.
- [22] Y. Rubanova, R. T. Q. Chen, and D. Duvenaud, "Latent ODEs for irregularly-sampled time series," in *Proc. NeurIPS*, Jan. 2019, pp. 5320–5330.
- [23] M. Lechner and R. Hasani, "Learning long-term dependencies in irregularly-sampled time series," 2020, *arXiv:2006.04418*.
- [24] Y. Chen et al., "ContiFormer: Continuous-time transformer for irregular time series modeling," in *Proc. NeurIPS*, New Orleans, LA, USA, 2023, pp. 1–33.
- [25] Z. Xiao, Z. Zhang, Z. Chen, Z. Yang, and R. Jin, "Mobile MIMO channel prediction with ODE-RNN: A physics-inspired adaptive approach," in *Proc. IEEE 33rd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2022, pp. 1301–1307.
- [26] M. Cui, H. Jiang, Y. Chen, Y. Du, and L. Dai, "Continuous-time channel prediction based on tensor neural ordinary differential equation," *China Commun.*, vol. 21, no. 1, pp. 163–174, Jan. 2024.
- [27] K. Ma, F. Zhang, W. Tian, and Z. Wang, "Continuous-time mmWave beam prediction with ODE-LSTM learning architecture," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 187–191, Jan. 2023.
- [28] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii, "Future steps of LTE-A: Evolution toward integration of local area and wide area systems," *IEEE Wireless Commun.*, vol. 20, no. 1, pp. 12–18, Feb. 2013.
- [29] Z. Ren, S. Wu, and A. Zhao, "Coexist design of sub-6 GHz and millimeter-wave antennas for 5G mobile terminals," in *Proc. Int. Symp. Antennas Propag. (ISAP)*, Oct. 2018, pp. 1–2.
- [30] F. Gao, B. Lin, C. Bian, T. Zhou, J. Qian, and H. Wang, "FusionNet: Enhanced beam prediction for mmWave communications using sub-6 GHz channel and a few pilots," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8488–8500, Dec. 2021.
- [31] K. Ma, D. He, H. Sun, and Z. Wang, "Deep learning assisted mmWave beam prediction with prior low-frequency information," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.
- [32] A. Ali, N. González-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1052, Feb. 2018.

- [33] K. Ma, S. Du, H. Zou, W. Tian, Z. Wang, and S. Chen, "Deep learning assisted mmWave beam prediction for heterogeneous networks: A dual-band fusion approach," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 115–130, Jan. 2023.
- [34] M. Hashemi, C. E. Koksal, and N. B. Shroff, "Out-of-band millimeter wave beamforming and communications to achieve low latency and high energy efficiency in 5G systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 875–888, Feb. 2018.
- [35] M. S. Sim, Y.-G. Lim, S. H. Park, L. Dai, and C.-B. Chae, "Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation," *IEEE Access*, vol. 8, pp. 51634–51646, 2020.
- [36] X. Zhang, L. Zhong, and A. Sabharwal, "Directional training for FDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5183–5197, Aug. 2018.
- [37] P. Kyösti, P. Zhang, A. Pärssinen, K. Haneda, P. Koivumäki, and W. Fan, "On the feasibility of out-of-band spatial channel information for millimeter-wave beam search," *IEEE Trans. Antennas Propag.*, vol. 71, no. 5, pp. 4433–4443, May 2023.
- [38] Y. Liu and O. Simeone, "Learning how to transfer from uplink to downlink via hyper-recurrent neural network for FDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 7975–7989, Oct. 2022.
- [39] *Study Channel Model for Frequencies From 0.5 To 100 GHz (Release 17)*, Standard 3GPP, TR 38.901, Version 17.0.0, 2022.
- [40] L. Wang, G. Liu, J. Xue, and K.-K. Wong, "Channel prediction using ordinary differential equations for MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2111–2119, Feb. 2023.
- [41] G. Liu, Z. Hu, L. Wang, H. Zhang, J. Xue, and M. Matthaiou, "A hypernetwork based framework for non-stationary channel prediction," *IEEE Trans. Veh. Technol.*, vol. 73, no. 6, pp. 8338–8351, Jun. 2024.
- [42] X. Wang, Y. Shi, W. Xin, T. Wang, G. Yang, and Z. Jiang, "Channel prediction with time-varying Doppler spectrum in high-mobility scenarios: A polynomial Fourier transform based approach and field measurements," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7116–7129, Nov. 2023.
- [43] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.
- [44] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.
- [45] Z. Feng and B. Clerckx, "Deep reinforcement learning for multi-user massive MIMO with channel aging," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 1, pp. 360–375, Oct. 2023.
- [46] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [47] X. Du and A. Sabharwal, "Massive MIMO channels with inter-user angle correlation: Open-access dataset, analysis and measurement-based validation," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1602–1616, Feb. 2022.
- [48] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1003–1010, Mar. 2007.
- [49] J. Dmochowski, J. Benesty, and S. Affes, "On spatial aliasing in microphone arrays," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1383–1395, Apr. 2009.
- [50] I. Ben Hagai, F. M. Fazi, and B. Rafaely, "Generalized sampling expansion for functions on the sphere," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5870–5879, Nov. 2012.
- [51] D. Burghal et al., "Enhanced AI-based CSI prediction solutions for massive MIMO in 5G and 6G systems," *IEEE Access*, vol. 11, pp. 117810–117825, 2023.
- [52] K. Ma, Y. Sang, Y. Ming, J. Lian, C. Tian, and Z. Wang, "Improving the performance of R17 type-II codebook with deep learning," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2023, pp. 4823–4828.
- [53] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. S. Dickstein, "On the expressive power of deep neural networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 2847–2854.
- [54] N. Rahaman et al., "On the spectral bias of neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5301–5310.
- [55] M. Tancik et al., "Fourier features let networks learn high frequency functions in low dimensional domains," in *Proc. NeurIPS*, Vancouver, BC, Canada, Jan. 2020, pp. 7537–7547.
- [56] Z. Xiao, Z. Zhang, C. Huang, X. Chen, C. Zhong, and M. Debbah, "C-GRBFNet: A physics-inspired generative deep neural network for channel representation and prediction," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2282–2299, Aug. 2022.
- [57] I. Mehta, M. Gharbi, C. Barnes, E. Shechtman, R. Ramamoorthi, and M. Chandraker, "Modulated periodic activations for generalizable local functional representations," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14194–14203.
- [58] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. NIPS*, 2020, pp. 7462–7473.
- [59] F. Sohrabi, K. M. Attiah, and W. Yu, "Deep learning for distributed channel feedback and multiuser precoding in FDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4044–4057, Jul. 2021.
- [60] J. Guo, C. Wen, and S. Jin, "Deep learning-based CSI feedback for beamforming in single- and multi-cell massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1872–1884, Jul. 2021.
- [61] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.