Accurate Channel Prediction Based on Spatial-Temporal Electromagnetic Kernel Learning

Jinke Li*, Jieao Zhu*, and Linglong Dai*, *Fellow, IEEE*

*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
*Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China
E-mails: lijk23@mails.tsinghua.edu.cn; zja21@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn

ICC 2025 - IEEE International Conference on Communications | 979-8-3315-0521-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICC52391.2025.11161116

***Abstract*—In wireless communication, accurate and efficient channel prediction is essential for addressing channel aging caused by user mobility. However, the actual channel variations over time are complex in high-mobility scenarios. This complexity makes it difficult for existing channel predictors to obtain future channels accurately. To overcome channel aging, we propose a channel prediction scheme based on spatial-temporal electromagnetic kernel learning (STEM-KL). Specifically, the STEM correlation function can capture the fundamental propagation characteristics of the wireless channel, making it suitable to use as a kernel function that incorporates prior information. For the channel prediction problem especially, we redesign the hyperparameters of the STEM kernel, including user velocity and concentration, which characterize the direction of the EM wave. The hyperparameters are obtained through kernel learning. Then, we use Bayesian inference to predict the future channels, employing the STEM kernel as the required covariance. To further improve the stability and model expressiveness, we propose a grid-based electromagnetic mixed kernel learning (GEM-KL) scheme. We design the mixed kernel to be a convex combination of multiple sub-kernels, where each of the sub-kernels corresponds to a grid point in the parameter space. This approach transforms the learning of concentration and speed hyperparameters into the learning of weights for different sub-kernels, helping the kernel learning process avoid local optima. Finally, simulation results verify that the proposed STEM-KL schemes outperform the baseline schemes.**

***Index Terms*—Channel prediction, spatial-temporal electromagnetic (STEM) correlation, grid-based electromagnetic mixed kernel learning (GEM-KL), Gaussian process regression (GPR).**

I. Introduction

The effective communication of massive MIMO system highly relies on accurate and timely channel state information (CSI). However, dynamic environments, characterized by user mobility, complicate the acquisition of CSI. According to the current 5G standard, explicit CSI acquisition, i.e., channel estimation, is performed periodically. When user mobility is high, significant channel changes may occur within a single channel estimation period, leading to outdated CSI. This phenomenon is termed as *channel aging* [1]. To address the challenges posed by channel aging, various channel prediction techniques have emerged.

Sparsity-based methods typically exploit the Doppler domain sparse structure of channel responses to predict future channels. For example, the sum-of-sinusoids model-based method [2] represents the channel response as a combination of sinusoidal waves. This method first identifies the dominant sinusoidal components and then uses the harmonic retrieval method to obtain these components for channel prediction. The authors of [3] proposed the Prony vector (PVEC) method which fits a linear prediction model to the observed channel response. Specifically, PVEC models the future channel as a linear combination of the past channels, where the combination weights are computed from the received pilot signals.

AR-based methods use autoregressive principle to process channel time series [4]. The original AR prediction method models the future channel as a weighted sum of its past values, where the weights, i.e., the AR parameters, are obtained from the autocorrelation function of channels at different times [5]. The Wiener channel predictor is a generalization of the AR predictor [6]. This approach predicts an autoregressive multivariate random process using a Wiener linear filter.

The existing two categories of channel prediction schemes were designed for prediction in the discrete-time domain. However, wireless communication systems utilize continuously varying electromagnetic fields in both space and time for signal transmission [7]. The existing methods are incompatible with the continuous characteristics of electromagnetic signals, which means they cannot accurately capture CSI within a single estimation period, leading to severe performance degradation. Therefore, it is essential to restore to the continuous electromagnetic fields to analyze channel prediction and design channel predictors based on this foundation.

In order to solve this problem, we propose a channel prediction scheme based on *electromagnetic kernel learning*, which simultaneously utilizes the spatial-temporal electromagnetic (EM) correlation characteristic of the channel. Specifically, we derive the spatial-temporal electromagnetic (STEM) correlation function of the EM channel. This STEM kernel originates from EM physics, thus it is more suitable for modeling practical wireless propagation environments than other kernel functions. The spatial-temporal electromagnetic kernel learning (STEM-KL) based channel predictor is proposed to achieve parallel prediction of future channels through the Bayesian framework. In this framework, channel prediction is divided into two sub-problems: First, the hyperparameters are obtained through STEM kernel learning. Then, through Bayesian inference, several future channels are simultaneously predicted. In order to further improve the stability and model expressiveness, we propose a grid-based electromagnetic mixed kernel learning (GEM-KL) scheme. We design a mixed kernel composed of sub-kernels, where each of the sub-kernel corresponds to a grid point in the parameter space. In this way, hyperparameters

learning is transformed into sub-kernel weights learning. Finally, through performance analysis and numerical experiments, it can be verified that the proposed GEM-KL channel predictor outperforms the PVEC and AR baselines.

II. System Model and Problem Formulation

In this section, we first review the Gaussian random field (GRF)-based channel model, based on which we introduce a probabilistic inference method for channel prediction. The autocorrelation function of the channel is selected via the maximum likelihood (ML) criterion from a GRF ensemble, and the inference is done by the minimum mean square error (MMSE) estimator.

A. GRF-Based Channel Model

Traditional channel models express channel matrices as a weighted Gaussian mixture of steering vectors, which is a discrete special case of a Gaussian random field. To capture the continuously varying properties of the wireless channel, in this section, we model the channel with a complex symmetric Gaussian random field (CSGRF). Let function $h(\boldsymbol{\rho}) : \mathbb{R}^4 \to \mathbb{C}$ represent a circularly symmetric Gaussian random field (CSGRF). The variable is $\boldsymbol{\rho} = (\mathbf{x}, t)$, where $\mathbf{x} = (x, y, z)$ represents the spatial location, t represents time indicator, and $(\mathbf{x}, t) \in \mathbb{R}^4$. For any Q points, the joint distribution of their function values $(h(\boldsymbol{\rho}_1), h(\boldsymbol{\rho}_2), \ldots, h(\boldsymbol{\rho}_Q))$ follows a multivariate Gaussian distribution, then the random field is a Gaussian random field, denoted as $h(\boldsymbol{\rho}) \sim \mathcal{GRF}(0, k(\boldsymbol{\rho}, \boldsymbol{\rho}'))$, and its probability measure is determined by their autocorrelation function

$$k(\boldsymbol{\rho}, \boldsymbol{\rho}') = \mathbb{E}\left[h(\boldsymbol{\rho}) h^*(\boldsymbol{\rho}')\right]. \tag{1}$$

The autocorrelation function is usually called the kernel, note that the kernel function of the GRF must be semi-positive definite. To enable CSGRF to represent the wireless channel, some restrictions should be imposed on $k(\boldsymbol{\rho}, \boldsymbol{\rho}')$ so that the $h(\boldsymbol{\rho})$ generated by it satisfies the EM propagation constraints. We use $h(\boldsymbol{\rho})$ to model the electric field distribution $\mathbf{E}(\boldsymbol{\rho}) : \mathbb{R}^4 \to \mathbb{C}^3$. Then, the autocorrelation function can be defined as $\mathbf{K}_{\mathbf{E}}(\boldsymbol{\rho}, \boldsymbol{\rho}') = \mathbb{E}\left[\mathbf{E}(\boldsymbol{\rho})\mathbf{E}(\boldsymbol{\rho}')^{\mathsf{H}}\right] \in \mathbb{C}^{3\times 3}$ [8]. Similarly, for a channel vector with N_{BS} components, it can also be modeled using CSGRF by constructing the autocorrelation function of $\boldsymbol{\rho}_n$ for $n = 1, 2, \ldots, N_{\mathrm{BS}}$.

B. Signal Model

We consider a massive MIMO system, in which a single base station (BS) with N_{BS} antennas serves a single user with 1 antenna. We will try to solve the problem of uplink channel prediction in a narrowband system. Consider the simplest communication scenario, assuming we use an N_{BS}-antenna base station with fully digital precoding, where each antenna is connected to a dedicated radio frequency (RF) chain. The uplink signal model is

$$\mathbf{y}_t = \mathbf{h}_t + \mathbf{n}_t, \tag{2}$$

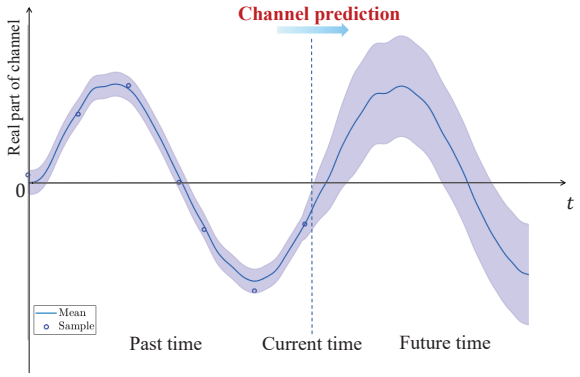

Fig. 1. An illustration of channel time-varying: Taking a component of a channel vector as an example

where $\mathbf{y}_t \in \mathbb{C}^{N_{\mathrm{BS}}\times 1}$is the BS received pilots at time t, $\mathbf{h}_t \in \mathbb{C}^{N_{\mathrm{BS}}\times 1}$is the normalized channel vector satisfying $\mathbb{E}[\|\mathbf{h_t}\|^2] = N_{\mathrm{BS}}$, and $\mathbf{n}_t$ is the complex-valued additive white Gaussian noise (AWGN) with zero mean and covariance $\gamma^{-1}\mathbf{I}_{N_{\mathrm{BS}}}$. The symbol γ represents the received signal-to-noise ratio (SNR) of the BS. We use the minimum mean square error (MMSE) criterion [9] to estimate the channel.

$$\hat{\mathbf{h}}_t^{\mathrm{MMSE}} = \mathbb{E}\left[\mathbf{h}_t|\mathbf{y}_t\right] = \boldsymbol{\Sigma}_{\mathbf{h}_t}\left(\boldsymbol{\Sigma}_{\mathbf{h}_t} + \frac{1}{\gamma}\mathbf{I}\right)^{-1}\mathbf{y}_t, \tag{3}$$

where $\boldsymbol{\Sigma}_{\mathbf{h}_t} = \mathbb{E}\left\{\mathbf{h}_t\mathbf{h}_t^{\mathsf{H}}\right\}$ is the prior covariance matrix of channel.

C. Problem Formulation

We refer to the period of channel estimation as a frame, which contains N_s time slots. Channel estimation is only performed in the first slot. In mobile scenarios, because of the Doppler effect, except for the channel at the first slot, the actual channels of the follow-up slots may have significant differences from the channel estimation result, leading to a decrease in the accuracy of the obtained CSI and thus affecting communication quality. The variations of the channel and its uncertainty over time are shown in Fig. 1. The solid curve represents the real part of the channel vector component, and the shadow area represents its uncertainty region. It can be observed that the channel uncertainty significantly increases at future time moments.

The channel prediction problem is to obtain future channels through past channels. Considering the characteristics of the GRF channel, achieving accurate channel prediction requires an appropriate autocorrelation function, i.e., the kernel. We can then predict the future channel through inference based on this kernel. The appropriate kernel form will be discussed in the next section. Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ denote model parameters of the kernel, and $\boldsymbol{\Omega}$ is the set of model parameters. $\mathbf{y} = (\mathbf{y}_1^{\mathsf{T}}, \mathbf{y}_2^{\mathsf{T}}, \ldots, \mathbf{y}_L^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{C}^{L_N\times 1}$ denotes the column vector composed of the received pilot sequences in the past L time frames, where $L_N = N_{\mathrm{BS}}L$ and $F_N = N_{\mathrm{BS}}F$. $\mathbf{h}_{\mathcal{L}} = (\mathbf{h}_1^{\mathsf{T}}, \mathbf{h}_2^{\mathsf{T}}, \ldots, \mathbf{h}_L^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{C}^{L_N\times 1}$ denotes the column vector composed of the previous L channels. $\mathbf{h}_{\mathcal{F}} = (\mathbf{h}_{(L+1)}^{\mathsf{T}}, \mathbf{h}_{(L+2)}^{\mathsf{T}}, \ldots, \mathbf{h}_{(L+F)}^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{C}^{F_N\times 1}$ denotes the column vector composed of F future channels that need to be predicted.

After accurately predicting the channel of future frames, we can reduce the frequency of channel estimation, thereby reducing the pilot overhead. By using the ML criterion to obtain kernel parameters, and then using the MMSE criterion to predict future channels, the channel prediction problem can be formulated as

$$\begin{aligned}\hat{\boldsymbol{\omega}}(\mathbf{y}) &= \arg\max_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} \ln \int p(\mathbf{y}|\mathbf{h}_{\mathcal{L}}) p(\mathbf{h}_{\mathcal{L}}|\boldsymbol{\omega}) \mathrm{d}\mathbf{h}_{\mathcal{L}}, \\ \hat{\mathbf{h}}_{\mathcal{F}}(\mathbf{y}) &= \arg\max_{\mathbf{h}_{\mathcal{F}}\in\mathbb{C}^{N_{\mathrm{BS}}F\times 1}} \ln p(\mathbf{y}|\mathbf{h}_{\mathcal{F}}) + \ln p(\mathbf{h}_{\mathcal{F}}|\hat{\boldsymbol{\omega}}(\mathbf{y})).\end{aligned} \tag{4}$$

In the following Section III, we will solve the problem in (4).

III. Proposed Spatial-Temporal Electromagnetic Kernel Learning Based Channel Prediction

In this section, the appropriate autocorrelation function, i.e., the STEM, has been applied. Then we propose a channel prediction scheme that simultaneously utilizes the spatial-temporal EM correlation between channels.

A. STEM Correlation Function

EM information can be combined with the autocorrelation function of the channel, This function is also named as electromagnetic correlation function (EMCF) [10], which is expressed as

$$\mathbf{K}_{\mathrm{EMCF}}(\boldsymbol{\rho}, \boldsymbol{\rho}') = \frac{\zeta^2}{S(\|\boldsymbol{\delta}\|)} \boldsymbol{\Sigma}(\boldsymbol{\xi}), \tag{5}$$

where $\mathbf{K}_{\mathrm{EMCF}}$ is a 3×3 complex matrix, $\mathrm{tr}(\mathbf{K}_{\mathrm{EMCF}}(\boldsymbol{\rho}, \boldsymbol{\rho}')) = \zeta^2$, $\boldsymbol{\xi} = k_0\mathbf{w} = k_0(\mathbf{x} - \mathbf{x}' + \mathbf{v}(t-t')) - \mathrm{i}\boldsymbol{\delta} \in \mathbb{C}^3$, in which $k_0 = 2\pi/\lambda_0$ is the wavenumber. The vector $\boldsymbol{\delta} \in \mathbb{C}^3$ is the concentration parameter, and its direction represents the direction in which the EM wave is concentrated. $S(\delta) = \sinh(\delta)/\delta$ is an additional normalisation factor, where $\delta = \|\boldsymbol{\delta}\| \in \mathbb{R}_+$. For time-varying channels, we incorporate the Doppler frequency shift into the EM correlation function by introducing the velocity vector $\mathbf{v}$, hence this EMCF can be referred to as the spatial-temporal kernel function (STEM-CF). We utilize the commonly used spherical Bessel functions $j_n(\xi)$ in 3D scenes to represent the correlation function $\boldsymbol{\Sigma}(\boldsymbol{\xi})$

$$\boldsymbol{\Sigma}(\boldsymbol{\xi}) = \frac{1}{6}(4j_0(\xi) - j_2(\xi))\mathbf{I_3} + \frac{1}{2}(j_2(\xi) - 2j_0(\xi))\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^{\mathsf{T}}, \tag{6}$$

where $\xi = |\boldsymbol{\xi}| = \sqrt{\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}/\xi$ denotes the normalized $\boldsymbol{\xi}$. The spherical Bessel function $j_n(\xi)$ is expressed as

$$j_n(\xi) = (-\xi)^n \left(\frac{1}{\xi}\frac{\mathrm{d}}{\mathrm{d}\xi}\right)^n \frac{\sin\xi}{\xi}. \tag{7}$$

It is important to note that $\mathbf{w} = \mathbf{x} - \mathbf{x}' + \mathbf{v}(t-t') - i\boldsymbol{\delta}/k_0$ contains the spatially varying covariates and time-varying covariates, which means that the correlation function we use is able to account for the spatial and temporal correlation in a way consistent with EM principles.

B. Gaussian Process Regression

Gaussian Process Regression (GPR) [11], also called Bayesian linear regression can obtain predictions through prior information and observation data of GRF. Specifically, for the GRF $f(x) \sim \mathcal{GRF}(\mu(x), k(x,x'))$, GPR uses observation data $y_i = f(x_i) + n_i$, $n_i \sim \mathcal{CN}(0, \sigma_n^2), i = 1, 2, \ldots, L_N$ to get a set of F_N-point prediction $\mathcal{F} = \{f(x_{L_N+1}), f(x_{L_N+2}), \ldots, f(x_{L_N+F_N})\}$.

The joint probability distribution of the observed and predicted joint vector $\mathbf{g} = [y_1, y_2, \ldots, y_{L_N}, f(x_{L_N+1}), f(x_{L_N+2}), \ldots, f(x_{L_N+F_N})]^{\mathsf{T}}$ satisfies

$$\mathbf{g} \sim \mathcal{CN}\left(\begin{bmatrix}\boldsymbol{\mu}_{\mathcal{L}} \\ \boldsymbol{\mu}_{\mathcal{F}}\end{bmatrix}, \begin{bmatrix}\mathbf{K}_{\mathcal{LL}} + \sigma_n^2\mathbf{I}_L & \mathbf{K}_{\mathcal{LF}} \\ \mathbf{K}_{\mathcal{FL}} & \mathbf{K}_{\mathcal{FF}}\end{bmatrix}\right), \tag{8}$$

where $\boldsymbol{\mu}_{\mathcal{L}} = [\mu(x_1), \mu(x_2), \ldots, \mu(x_{L_N})]^{\mathsf{T}}$ and $\boldsymbol{\mu}_{\mathcal{F}} = [\mu(x_{L_N+1}), \mu(x_{L_N+2}), \ldots, \mu(x_{L_N+F_N})]^{\mathsf{T}}$. The (m,n)-th entry of $\mathbf{K}_{\mathcal{LL}} \in \mathbb{C}^{L_N\times L_N}$ is $k(x_m, x_n)$, for all $m, n \in \{1, \ldots, L_N\}$. The (m,n)-th entry of $\mathbf{K}_{\mathcal{LF}} \in \mathbb{C}^{L_N\times F_N}$ is $k(x_m, x_n)$, for all $m \in \{1, \ldots, L_N\}$ and $n \in \{L_N+1, \ldots, L_N+F_N\}$. And $\mathbf{K}_{\mathcal{FL}} = \mathbf{K}_{\mathcal{LF}}^{\mathsf{H}} \in \mathbb{C}^{F_N\times L_N}$. The (m,n)-th entry of $\mathbf{K}_{\mathcal{FF}} \in \mathbb{C}^{F_N\times F_N}$ is $k(x_m, x_n)$, for all $m, n \in \{L_N+1, \ldots, L_N+F_N\}$. We use $\mathbf{K}_{\mathbf{y}}$ to represent $\mathbf{K}_{\mathcal{LL}} + \sigma_n^2\mathbf{I}_{L_N}$. From the Gaussian posterior formula, we can obtain

$$\begin{aligned}\boldsymbol{\mu}_{\mathcal{F}|\mathcal{L}} &= \boldsymbol{\mu}_{\mathcal{F}} + \mathbf{K}_{\mathcal{LF}}^{\mathsf{H}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}, \\ \mathbf{K}_{\mathcal{F}|\mathcal{L}} &= \mathbf{K}_{\mathcal{F}} - \mathbf{K}_{\mathcal{LF}}^{\mathsf{H}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{K}_{\mathcal{FL}}.\end{aligned} \tag{9}$$

The results of Bayesian regression are given by $\boldsymbol{\mu}_{\mathcal{F}|\mathcal{L}}$.

C. Kernel Learning

Choosing appropriate kernel parameters is an important step in constructing an effective regression model, which affects the accuracy of the kernel function in reconstructing Gaussian processes. The parameters that need to be adjusted in this process are usually referred to as hyperparameters. Assuming that the hyperparameters $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^{N_{\boldsymbol{\omega}}}$ of the adjustable kernel $k(x; x'|\boldsymbol{\omega})$ is also tunable. The process of finding the optimal hyperparameters for the STEM kernel is called kernel learning.

It is necessary to specify a criterion for evaluating whether hyperparameters are appropriate. The maximum likelihood (ML) criterion is a commonly used method, which can be expressed as

$$\hat{\boldsymbol{\omega}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\omega}\in\Omega} \ln p(\mathbf{y}|\boldsymbol{\omega}), \tag{10}$$

where the probability density function (PDF) of observing $\mathbf{y}$ given the parameter $\boldsymbol{\omega}$ is expressed as

$$p(\mathbf{y}|\boldsymbol{\omega}) = \frac{1}{\pi^{L_N+F_N}\det\mathbf{K}_{\mathbf{y}}} \exp(-\mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}). \tag{11}$$

The kernel $\mathbf{K}_{\mathbf{y}} = \mathbf{K}_{\mathbf{y}}(\boldsymbol{\omega})$ is a matrix-valued function of hyperparameter $\boldsymbol{\omega}$. Function $l(\boldsymbol{\omega}|\mathbf{y}) = \ln p(\mathbf{y}|\boldsymbol{\omega}) = -\ln\det\mathbf{K}_{\mathbf{y}} - (L_N+F_N)\ln\pi - \mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}$ is the log-likelihood function. In order to obtain the maximum likelihood estimator of the hyperparameter $\boldsymbol{\omega}$, methods such as gradient descent, conjugate gradient descent, and Newton iteration can be used. All of these methods require the derivative of the log-likelihood function with respect to $\boldsymbol{\omega}$, which is given by

$$\begin{aligned}\frac{\partial l(\boldsymbol{\omega}|\mathbf{y})}{\partial\omega_i} &= \frac{\partial}{\partial\omega_i}(-\ln\det\mathbf{K}_{\mathbf{y}} - \mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}) \\ &= \mathrm{tr}((\boldsymbol{\varpi}\boldsymbol{\varpi}^{\mathsf{H}} - \mathbf{K}_{\mathbf{y}}^{-1})\frac{\partial\mathbf{K}_{\mathbf{y}}}{\partial\omega_i}),\end{aligned} \tag{12}$$

where ω_i for $i = 1, 2, \ldots, N_{\boldsymbol{\omega}}$ represents each component of hyperparameter $\boldsymbol{\omega}$. For simplicity, let $\boldsymbol{\varpi} = \mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}$. When the

hyperparameter components are complex numbers, we need to consider the Wirtinger derivatives $(\partial/\partial\omega_{i,\mathrm{Re}} - \mathrm{i}\partial/\partial\omega_{i,\mathrm{Im}})/2$. Since $l(\boldsymbol{\omega}|\mathbf{y})$ is an analytic function of $\mathbf{K}_\mathbf{y}$, the derivative formula (12) remains unchanged.

Through gradient-based methods such as gradient ascent, these results can be used to obtain better $\boldsymbol{\omega}$ according to ML criterion.

D. Proposed Grid Electromagnetic Mixed Kernel

The gradient based hyperparameter optimization method may get stuck in local optima. Fortunately, the grid-based electromagnetic mixed kernel learning (GEM-KL) proposed in this subsection can achieve more global learning results.

Firstly, we analyze the objective function $l(\boldsymbol{\omega}|\mathbf{y})$, which can be intuitively represented as a function of kernel $\mathbf{K}_\mathbf{y}$. However, $l(\boldsymbol{\omega}|\mathbf{y})$ is not a convex/concave function of $\mathbf{K}_\mathbf{y}$. Therefore, gradient-based optimization methods are difficult to find the maximum value of $l(\boldsymbol{\omega}|\mathbf{y})$. Moreover, the kernel $\mathbf{K}_\mathbf{y}$ can be expressed as a function of the hyperparameter $\boldsymbol{\omega}$. Unfortunately, the components $\boldsymbol{\delta}, \mathbf{v}$ of $\boldsymbol{\omega}$ are not linearly related to $\mathbf{K}_\mathbf{y}$, making it difficult to directly characterize the relationship between $\boldsymbol{\omega}$ and $l(\boldsymbol{\omega}|\mathbf{y})$. In order to avoid the inconvenience caused by the non-convexity/concavity of functions, the grid-based method can be used in the parameter learning of the STEM kernel. We design a mixed kernel composed of sub-kernels, and each of the sub-kernels corresponds to a grid point in the parameter space. specifically, several fixed values of $\boldsymbol{\delta}$ and $\mathbf{v}$ are taken to be designed as the selection values for the grid. By introducing the idea of mixed kernel, k_{GEM} is a combination of multiple sub-STEM kernels. We assume that there are N_k sub-correlation kernels and each of them has a weight of $c_n \in \mathbb{R}$, $n = 1, 2, \ldots, N_k$. Specifically, the GEM kernel function is designed as

$$\begin{aligned} &k_{\mathrm{GEM}}(x_p, t_p; x_q, t_q|\boldsymbol{\omega}) \\ &\quad = \mathbf{u}_p^{\mathsf{T}}\big(\sum_{n=1}^{N_k} c_n \mathbf{K}_{\mathrm{STEM}}(x_p, t_p; x_q, t_q|\boldsymbol{\omega}_n)\big)\mathbf{u}_q, \end{aligned} \tag{13}$$

where the value of each $k_{\mathrm{GEM}}(x_p, t_p; x_q, t_q|\boldsymbol{\omega}_n)$ is on the grid $(\boldsymbol{\delta}_n, \mathbf{v}_n)$, where $\boldsymbol{\delta}_n \in \boldsymbol{\Delta}$ and $\mathbf{v}_n \in \mathbf{V}$. The grid values are uniformly sampled from the two-dimensional space defined by $\boldsymbol{\Delta} \times \mathbf{V}$. $\boldsymbol{\omega}_n \in \{\boldsymbol{\delta}_n, \mathbf{v}_n, c_n\}_{n=1}^{N_k} \subset \boldsymbol{\Omega}$ is the collection of all the hyperparameters $\boldsymbol{\omega}_n \in \boldsymbol{\Omega}$. Correspondingly, the components of the mixed correlation kernel matrix can be represented as

$$(\mathbf{K}_{\mathcal{LL},\mathrm{Mix}})_{p,q} = k_{\mathrm{GEM}}(x_p, t_p; x_q, t_q|\boldsymbol{\omega}), \tag{14}$$

The weight c_n is linearly related to the kernel $k_{\mathrm{STEM}}(x_p, t_p; x_q, t_q|\boldsymbol{\omega}_n)$ in the objective function $l(\boldsymbol{\omega}|\mathbf{y})$, so optimizing the weights $\{c_n\}_{n=1}^{N_k}$ corresponding to different $\boldsymbol{\delta}_n$ and $\mathbf{v}_n$ is sufficient to obtain the optimal hyperparameters on the grid. Let $\mathbf{c} \in \mathbf{S}_{N_k}$ denotes $\{c_n\}_{n=1}^{N_k}$, where $\mathbf{S}_{N_k}$ is the collection of the non negative vector that sum to 1.

The mixed and grid-based kernel is able to improve the fitting ability of Gaussian random fields defined by the STEM function to channel observation data. The ML problem is expressed as

$$\hat{\mathbf{c}}_{\mathrm{ML}} = \underset{\mathbf{c}\in\mathbf{S}_{N_k}}{\arg\max} \ln p(\mathbf{y}|\mathbf{c}), \tag{15}$$

The log likelihood function is

$$\begin{aligned} l(\{c_n\}_{n=1}^{N_k}, \zeta^2|\mathbf{y}) &= \ln p(\mathbf{y}|\{c_n\}_{n=1}^{N_k}) \\ &= -\ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}} - \mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y} \\ &\quad + \mathrm{const}, \end{aligned} \tag{16}$$

where $\mathbf{K}_{\mathbf{y},\mathrm{Mix}} = \mathbf{K}_{\mathcal{LL},\mathrm{Mix}} + \sigma_\mathbf{h}^2\mathbf{I}_{L_N} = \sum_{n=1}^{N_k} c_n\mathbf{K}_{\mathcal{LL},n} + \sigma_\mathbf{h}^2\mathbf{I}_{L_N}$. Let $l_r(\{c_n\}_{n=1}^{N_k}, \zeta^2|\mathbf{y}) = \ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}} + \mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y}$, we transform ML problems into finding the minimum value of l_r to eliminate the negative sign

$$\hat{\mathbf{c}}_{\mathrm{ML}} = \underset{\mathbf{c}\in\mathbf{S}_{N_k}}{\arg\min}(\ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}} + \mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y}), \tag{17}$$

where $\mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y}$ is a convex function about $\mathbf{K}_{\mathbf{y},\mathrm{Mix}}$ and $\ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}}$ is a concave function about $\mathbf{K}_{\mathbf{y},\mathrm{Mix}}$. The majorization-minimization (MM) algorithm can be used to solve the optimal hyperparameters with non-convex and non-concave objective functions through an iterative scheme.

In the majorization step, we use the first-order Taylor expansion to design the surrogate function, which approximates the upper bound of the concave part of the new objective function. To linearize $\ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}}$, the concave part, at $\mathbf{K}_{\mathbf{y},\mathrm{Mix}} = \mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)}$, i.e., $\mathbf{c} = \mathbf{c}^{(m)}$, the inequality is constructed as follow

$$\begin{aligned} l_r(\mathbf{K}_{\mathbf{y},\mathrm{Mix}}) \le &\mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y} + l_{\mathrm{CCV}}(\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)}) \\ &+\mathrm{tr}\left(\nabla l_{\mathrm{CCV}}(\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)})^{\mathsf{T}}(\mathbf{K}_{\mathbf{y},\mathrm{Mix}} - \mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)})\right) \end{aligned} \tag{18}$$

where l_r is the new objective function, $l_{\mathrm{CCV}}(\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)}) = \ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)}$ and $\big(\nabla l_{\mathrm{CCV}}(\mathbf{K})\big)_{ij} = \partial l/\partial\mathbf{K}_{ij}$. The Wirtinger derivative of l w.r.t. $\mathbf{K}_{\mathcal{LL},n}$ is given by the following formula

$$\frac{\partial l}{\partial\mathbf{K}_{\mathcal{LL},n}} = (\mathbf{g}\mathbf{g}^{\mathsf{H}} - \mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1})^*, \tag{19}$$

where $\mathbf{g} = \mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y}$. The real-variable derivative of the objective function l with respect to c_n is expressed as

$$\frac{\partial l}{\partial c_n} = 2\Re\left[\mathrm{tr}(\mathbf{K}_{\mathcal{LL},n}(\boldsymbol{\omega}_n)(\mathbf{g}\mathbf{g}^{\mathsf{H}} - \mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}))\right]. \tag{20}$$

Using formulas (18), (19) and (20), the surrogate function of the MM algorithm can be expressed as

$$\begin{aligned} l_s(c_n|c_n^{(m)}) = &\mathbf{y}^{\mathsf{H}}\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{-1}\mathbf{y} + \ln\det\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)} \\ &+2\Re\left\{\mathrm{tr}\left[((\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)})^{-1})^{\mathsf{T}}(\mathbf{K}_{\mathbf{y},\mathrm{Mix}} - \mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)})\right]\right\}. \end{aligned} \tag{21}$$

Then, in the minimization step, the weights $\left\{c_n^{(m)}\right\}_{n=1}^{N_k}$ are updated through

$$\hat{c}_n^{(m+1)} = \underset{\mathbf{c}\in\mathbf{S}_{N_k}}{\arg\min}(l_s(c_n|c_n^{(m)})), \tag{22}$$

This step requires the real-valued derivative of the surrogate function with respect to c_n, which is expressed as

$$\frac{\partial l_s}{\partial c_n} = 2\Re\left[\mathrm{tr}(\mathbf{K}_{\mathcal{LL},n}(\boldsymbol{\omega}_n)((\mathbf{K}_{\mathbf{y},\mathrm{Mix}}^{(m)})^{-1} - \mathbf{g}\mathbf{g}^{\mathsf{H}}))\right]. \tag{23}$$

These results can be used for iteratively solving the optimal weights $\{c_n\}_{n=1}^{N_k}$ in the MM algorithm. The sequence

Algorithm 1 Proposed GEM Kernel Hyperparameters Learning Algorithm.

Input: Number of sub-kernels N_k; grid hyperparameters $\{\boldsymbol{\delta}_n, \mathbf{v}_n\}_{n=1}^{N_k}$; Received pilots $\{\mathbf{y}_\ell\}_{\ell=1}^{L_N}$; Noise variance $\sigma_{\mathbf{h}}^2$; Maximum iteration number M_{iter}.
Output: Hyperparameters $\{\boldsymbol{\delta}_n, \mathbf{v}_n, \hat{c}_n\}_{n=1}^{N_k}$; $\hat{\zeta}^2$.

Initialization: $c_n^{(0)} = 1/N_k$, for $n = 1, 2, \ldots, N_k$. Learning rates of Armijo-Goldstein's optimizer. $m \leftarrow 0$.
Let $\mathbf{y} \in \mathbb{C}^{L_N \times 1}$ contain received pilots from $\{\mathbf{y}_\ell\}_{\ell=1}^{L_N}$.
for $m = 1, 2, \ldots, M_{\text{iter}}$ **do**
  Construct the GEM kernel $\mathbf{K}_{\mathbf{y},\text{Mix}}$ from hyperparameters $\{\boldsymbol{\delta}_n^{(m-1)}, \mathbf{v}_n^{(m-1)}, {c_n}^{(m-1)}\}_{n=1}^{N_k}$ by (13) and (14).
  $\mathbf{g} \leftarrow \mathbf{K}_{\mathbf{y},\text{Mix}}^{-1}\mathbf{y}$
  for $n = 1, 2, \ldots, N_k$ **do**
    Construct surrogate function $l_s(c_n|c_n^{(m-1)})$ by (21).
    Compute $\frac{\partial l_s}{\partial c_n}$ from (23).
    Update $c_n^{(m)}$ from (22) by Armijo-Goldstein's optimizer.
    Update $\mathbf{K}_{\mathbf{y},\text{Mix}}$ from $\{c_n^{(m)}\}_{n=1}^{N_k}$.
  end for
end for
$\hat{\zeta}^2 \leftarrow 2\left\|\mathbf{y}\right\|^2 / (L_N \cdot (1 + \sigma_{\mathbf{h}}^2))$
return Hyperparameters learning results $\{\boldsymbol{\delta}_n, \mathbf{v}_n, \hat{c}_n\}_{n=1}^{N_k}$, and $\hat{\zeta}^2$.

$(l_r(\mathbf{c}^{(m)}))_{m\in\mathbb{N}}$ is non-increasing since

$$l_r(\mathbf{c}^{(m+1)}) \le l_s(\mathbf{c}^{(m+1)}|\mathbf{c}^{(m)}) \le l_s(\mathbf{c}^{(m)}|\mathbf{c}^{(m)}) = l_r(\mathbf{c}^{(m)}). \tag{24}$$

The first term in the objective function (16) represents model complexity, while the second term represents data fitness. The process of maximizing the objective function l is capable of automatically balancing model complexity and data fitness. The GEM kernel parameter learning algorithm is summarized in **Algorithm 1**, and in the next subsection, we will summarize the overall GEM channel prediction algorithm.

E. Proposed GEM-KL Channel Prediction Algorithm

We set the number of base station antennas to N_{BS}, assuming that these antennas are located at $\{\mathbf{x}_n\}_{n=1}^{N_{\text{BS}}} \in \mathbb{R}^3$. We consider the spatial-temporal correlation tensor between the m-th polarization of antenna a at time t_i and the n-th polarization of antenna b at time t_j. Let $p = (a, m, i)$ and $q = (b, n, j)$, the correlation tensor can be expressed as

$$\mathbf{K}_{p,q} = \mathbf{u}_p^{\mathsf{T}} \left[\mathbf{K}_{\text{STEM}}(\mathbf{x}_p, t_p; \mathbf{x}_q, t_q)\right] \mathbf{u}_q, \tag{25}$$

where $\mathbf{u}_p$ and $\mathbf{u}_q$ represent the unit vector of antenna polarization direction. Based on formula (25), the correlation matrix between several channels in different time and space can be calculated, and the specific scheme is given by **Algorithm 2**. The proposed STEM-based channel prediction method is summarized in **Algorithm 3**. Specifically, the BS receives noisy observations at any spatial-temporal coordinate at past times and predicts the channel at future times. In this algorithm, the channels in the future or past times are modeled as a Gaussian random field. We need to first use STEM-CF to calculate the autocorrelation matrix $\mathbf{K}_{\mathbf{y}} = \mathbf{K}_{\mathcal{L}\mathcal{L}} + \sigma_n^2 \mathbf{I}_{L_N}$ of the channels at past times. And then calculate the correlation matrix between the past and future channels. Finally, Bayesian inference is used to obtain the future channels. The performance of the proposed channel prediction algorithm will be evaluated in the next section.

Algorithm 2 Channels Correlation Matrix Design.

Input: Hyperparameters $\{\boldsymbol{\delta}_n, \mathbf{v}_n, \hat{c}_n\}_{n=1}^{N_k}$, and $\hat{\zeta}^2$, channel indices $p \in \mathcal{P}$; $q \in \mathcal{Q}$, $p_{\min}$, $p_{\max}$, $q_{\min}$, $q_{\max}$.
Output: The correlation matrix between the channels in set $\mathcal{P}$ and the channels in set $\mathcal{Q}$: $\mathbf{K}_{\mathcal{P}\mathcal{Q}}$.

Let $\mathbf{K}_{\mathcal{P}\mathcal{Q}} \in \mathbb{C}^{|\mathcal{P}|\times|\mathcal{Q}|}$, $p = p_{\min}$, $q = q_{\min}$.
for $p = p_{\min}, p_{\min} + 1, \ldots, p_{\max}$ **do**
  for $q = q_{\min}, q_{\min} + 1, \ldots, q_{\max}$ **do**
    Calculate the STEM function: $\mathbf{K}_{pq} \leftarrow \mathbf{u}_p^{\mathsf{T}} \mathbf{K}_{\text{STEM}}(\mathbf{x}_p, t_p; \mathbf{x}_q, t_q|\boldsymbol{\omega})\mathbf{u}_q$ according to (5).
  end for
end for
return The correlation matrix $\mathbf{K}_{\mathcal{P}\mathcal{Q}}$.

Algorithm 3 Proposed STEM Channel Predictor.

Input: Past channel indices $l \in \mathcal{L}$; Future channel indices $f \in \mathcal{F}$; Received pilots $y_l, l \in \mathcal{L}$; Noise variance σ_n^2.
Output: Predicted future channels $\hat{\mathbf{H}}_{\mathcal{F}}$.

Obtain GEM hyperparameters $\{\boldsymbol{\delta}_n, \mathbf{v}_n, \hat{c}_n\}_{n=1}^{N_k}$ and $\hat{\zeta}^2$ according to **Algorithm 1**;
Compute the correlation matrix of past channels $\mathbf{K}_{\mathcal{L}\mathcal{L}}$ and the correlation matrix between the past channels and the future channels $\mathbf{K}_{\mathcal{F}\mathcal{L}}$ according to **Algorithm 2**.
$\mathbf{K}_{\mathbf{y}} = \mathbf{K}_{\mathcal{L}\mathcal{L}} + \sigma_n^2 \mathbf{I}_{L_N}$.
$\mathbf{g} \leftarrow \mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}$.
Reconstruct the predicted futrue channels $\hat{\mathbf{H}}_{\mathcal{F}} \leftarrow \mathbf{K}_{\mathcal{F}\mathcal{L}}\mathbf{g}$ according to (9).
return Channel prediction result $\hat{\mathbf{H}}_{\mathcal{F}}$.

IV. Simulation Results

In order to show the performance of our proposed STEM-KL and GEM-KL channel prediction schemes, the simulation results of some channel predictors are provided in this section.

Simulation setup. In the following channel prediction simulation, we evaluate the performance of various predictors using a standard 3GPP TR 38.901 CDL channel. The 128-element ULA is considered. The center of the antenna array is located at $(0, 0, 0)$, ULA is located on the x-axis, and the user moves in the xoz plane. The carrier frequency is set to $f_c = 3.5$ GHz. We set the period of transmitting pilot signals to $0.625\,\text{ms}$. The unit vector of antenna polarization direction is $\mathbf{u} = (0, 1, 0)^{\mathsf{T}}$.

Baseline algorithms. The no-prediction scheme refers to comparing the current estimated channel with the future channel. The AR predictor is given by the autoregressive modeling [4]. The PVEC predictor is given by the prony vector prediction method [3].

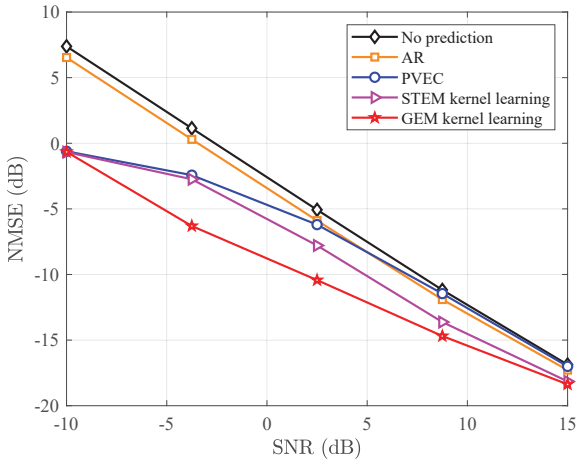

Fig. 2. The NMSE performance versus SNR in CDL channel model at maximum Doppler velocity of $72\,\mathrm{km/h}$.

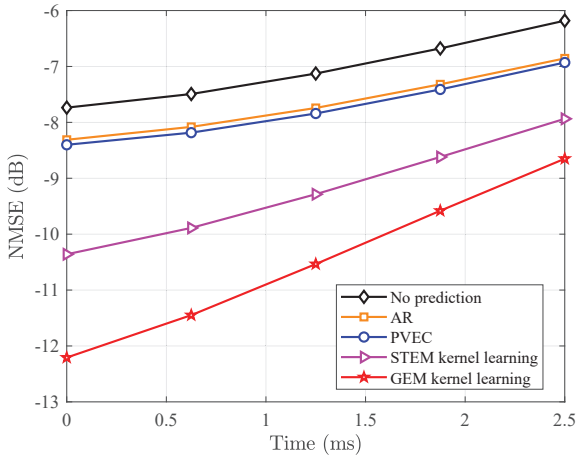

Fig. 3. The NMSE performance versus time in CDL channel model at the maximum Doppler velocity of $72\,\mathrm{km/h}$.

All schemes are evaluated using normalized mean square error (NMSE) performance, which is defined as

$$\mathrm{NMSE} = \mathbb{E}\left[\frac{\|\hat{\mathbf{h}} - \mathbf{h}\|^2}{\|\mathbf{h}\|^2}\right], \tag{26}$$

The NMSE performance versus SNR for different channel prediction schemes under the CDL-A channel model is plotted in Fig. 2. The channels from the past two frames are used to predict the channels for the next frame. It can be observed that the proposed STEM-KL scheme performs better than baseline schemes with a maximum Doppler velocity of $72\,\mathrm{km/h}$. Among them, the GEM-KL scheme can achieve the best performance. For example, when SNR is $2.5\,\mathrm{dB}$, compared to the PVEC scheme, the GEM kernel learning scheme can achieve NMSE performance gains of approximately $4.4\,\mathrm{dB}$ at $v = 72\,\mathrm{km/h}$, respectively.

In addition, we plot Fig. 3 to show the temporal variation of different channel prediction schemes. When SNR is $5\,\mathrm{dB}$, the channels from the past two frames are used to predict the channels for the next five frames. It can be observed that the proposed GEM-KL predictor also has the best NMSE performance in predicting the channels of subsequent frames. Taking the prediction of the channel for the second future frame as an example, compared with the PVEC channel prediction scheme, the proposed GEM-KL method achieves $3.8\,\mathrm{dB}$ improvement in NMSE performance in the scenarios of $v = 72\,\mathrm{km/h}$.

V. Conclusions

In this paper, we design a high-accuracy channel predictor through STEM kernel learning. We use the STEM correlation function as a kernel function and redesign the hyperparameters of the STEM kernel, including user velocity and concentration to fit time-varying channels. The hyperparameters are obtained through kernel learning. Then, we use GPR to predict future channels, using the STEM kernel as the required covariance. In order to improve the stability, we design the GEM kernel to be a convex combination of multiple sub-kernels, where each of the sub-kernels corresponds to a grid point in the parameter space. Finally, we test the proposed STEM-KL and GEM-KL channel prediction scheme, achieving improved performance over other baseline methods.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2023YFB3811503), in part by the National Natural Science Foundation of China (Grant No. 62325106), and in part by the National Natural Science Foundation of China (Grant No. 62031019).

References

[1] 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, Sep. 2013.

[2] F. Pena-Campos, R. Carrasco-Alvarez, O. Longoria-Gandara, and R. Parra-Michel, "Estimation of fast time-varying channels in OFDM systems using two-dimensional prolate," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 898–907, Jan. 2013.

[3] 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, Jun. 2020.

[4] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Sep. 2005.

[5] L. Chen, M. Loschonsky, and L. M. Reindl, "Autoregressive modeling of mobile radio propagation channel in building ruins," *IEEE Trans. Microwave Theory Tech.*, vol. 60, no. 5, pp. 1478–1489, Mar. 2012.

[6] Z. Qin, H. Yin, and W. Li, "Eigenvector prediction-based precoding for massive MIMO with mobility," *arXiv preprint arXiv:2308.12619*, Aug. 2023.

[7] Z. Wan, J. Zhu, Z. Zhang, L. Dai, and C.-B. Chae, "Mutual information for electromagnetic information theory based on random fields," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 1982–1996, Feb. 2023.

[8] J. Zhu, Z. Wan, L. Dai, M. Debbah, and H. V. Poor, "Electromagnetic information theory: Fundamentals, modeling, applications, and open problems," *IEEE Wireless Commun.*, Jan. 2024.

[9] 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, Jul. 2022.

[10] J. Zhu, Z. Wan, L. Dai, and T. J. Cui, "Can electromagnetic information theory improve wireless systems? A channel estimation example," *arXiv preprint arXiv:2310.12446*, Oct. 2023.

[11] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, Aug. 2018.