Dynamic Channel Prediction for Massive MIMO Systems with Temporal Non-Stationarity

Hongwei Hou*†, Yafei Wang*†, Yiming Zhu*†, Wenjin Wang*†, Dirk T. M. Slock‡
* National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
† Purple Mountain Laboratories, Nanjing 211100, China
‡ Department of Communication Systems, EURECOM, 06410 Biot, France
Email: {hongweihou, wangyf, ymzhu, wangwj}@seu.edu.cn, Dirk.Slock@eurecom.fr

***Abstract*—This paper investigates dynamic channel prediction for massive MIMO under temporal non-stationarity by leveraging dual-timescale and cluster correlations. We propose the sliding frame structure including pilot OFDM symbols to exploit dual-timescale correlations, i.e., intra-frame and inter-frame correlations, through Doppler-domain modeling and AR processes, respectively. This formulation leads to the spatial-frequency-temporal domain channel model and the correlated angle-delay-Doppler (ADD) domain representation, facilitating the joint exploitation of inter-antenna, inter-subcarrier, and inter-symbol correlations. In AR processes for long-timescale correlation modeling, the pattern-coupled variance is introduced to capture the energy leakage effects resulting from finite ADD domain resolutions. Building upon this structured prior model, we develop the dynamic channel prediction algorithm based on the simplified dynamic approximate message passing framework, with hyperparameters optimized via the expectation-maximization algorithm. Numerical simulations demonstrate the superiority of the proposed algorithm in terms of channel prediction performance.**

***Index Terms*—Massive MIMO, channel prediction, temporal non-stationarity, approximate message passing**

I. Introduction

The massive multiple-input multiple-output (MIMO) and orthogonal frequency-division multiplexing (OFDM) are key technologies for future communication systems to offer substantial capacity gains. However, the performance depends on accurate channel state information (CSI) [1]. In time division duplex (TDD) systems, CSI accuracy based on channel reciprocity is significantly degraded due to mobility-induced Doppler effects and temporal non-stationarity, particularly at higher carrier frequencies. To mitigate this issue, channel prediction has emerged as a promising solution, leveraging temporal correlations in historical CSI to estimate future CSI and counteract channel aging [2].

In the prior efforts, Doppler-domain modeling has demonstrated superior performance among temporal correlation models for channel prediction in stationary environments. Specifically, spectrum estimation techniques are employed for Doppler frequency estimation in both narrowband and wideband MIMO [3], [4]. In massive MIMO-OFDM systems, high-resolution angle-delay representations enable the extraction of Doppler frequencies for each angle-delay tap [5], while joint channel estimation and prediction algorithms help mitigate estimation errors [6]. For non-stationary channels, time-varying Doppler frequencies are approximated using polynomial fitting [7], facilitating channel prediction via polynomial Fourier transformation and the orthogonal matching pursuit algorithm. By treating Doppler frequencies as uncertain parameters,autoregressive (AR) processes are employed to capture Doppler variations over time [8], thereby enhancing the exploitation of channel sparsity.

Beyond Doppler domain modeling, AR models have been extensively explored, particularly in non-stationary environments. Specifically, sparse representations in the angle-delay domain facilitate AR-based channel tracking [9], where sparse Bayesian learning (SBL) and expectation-maximization aiding parameter estimation. Bulding on this, [10] presents the online group SBL algorithm that exploits both inter-subcarrier joint sparsity in the angle domain and structured sparsity within a single symbol. To further explore inter-subcarrier correlations, angle-delay domain sparse representations are developed in [11], where spatial-temporal AR processes capture the residual temporal correlations among neighboring elements. Besides, [12] adopts high-order AR processes to model temporal correlations in the spatial domain, improving channel prediction robustness in practical systems. Due to the continuous evolution of propagation environments, angle variations over time are also captured by AR processes and tracked using the Kalman filter (KF) [13]. In [14], the Markov processes are introduced for angle and angle-delay domain channel support, modeling dynamic sparsity. Inspired by AR processes, [15] employs the Taylor series for temporal correlations of angle-delay domain channels, developing a linear regression-based prediction algorithm.

This paper investigates dynamic channel prediction (DCP) for massive MIMO systems with temporal non-stationarity. Specifically, we propose the sliding frame structure including pilot OFDM symbols to exploit dual-timescale correlations, i.e., intra-frame and inter-frame correlations, through Doppler-domain modeling and AR processes, respectively. Based on this frame structure, we formulate a spatial-frequency-temporal (SFT) domain channel model and its corresponding angle-delay-Doppler (ADD) domain representation. To capture energy leakage effects in the ADD domain, we model correlated power patterns within the action regions using a pattern-coupled variance in the AR models. On this basis, we propose the channel prediction algorithm through the simplified dynamic approximate message passing (SD-AMP) algorithm, where unknown hyperparameters are

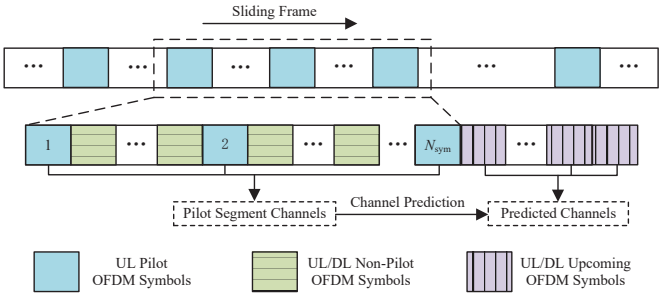

Fig. 1. Sliding frame structure for dynamic channel prediction.

learned by expectation-maximization (EM) algorithms.

Notations: The imaginary unit is represented by $j = \sqrt{-1}$. x, $\mathbf{x}$, and $\mathbf{X}$ denote scalars, column vectors, and matrices, respectively. The transpose, conjugate, and conjugate-transpose operations are represented by the superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$, respectively. The symbols $\mathbb{C}$ denote the complex number fields. The operators $\odot$ and $()^{\odot 2}$ denote the element-wise product and square operations, respectively. $\mathsf{diag}\{\cdot\}$ and $\mathsf{Re}\{\cdot\}$ denote the diagonal and real part operators, respectively. $\mathsf{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C})$ denotes the circular complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$.

II. Signal and Channel Model

We consider the massive MIMO-OFDM system with a base station (BS) equipped with a uniform linear array (ULA) of N_{an} antennas and single-antenna mobile terminals (MTs). This system operates in the TDD mode with the frame structure shown in Fig. 1, and adopts sounding reference signal in the third Generation Partnership Program (3GPP) New Radio (NR) specifications as pilots for uplink channel sounding. The MTs employ comb-type pilot patterns to mitigate pilot contamination, which are uniformly spaced at intervals of $\Delta\bar{T}$ and $\Delta\bar{f}$ in the temporal and frequency domains, respectively. In this system, the BS collects the received symbols in the pilot segment and predicts channels of all symbols from the last current pilot symbol to the first future pilot symbol, thus combating channel aging.

Assuming that physical parameters of each path remain unchanged within frames and vary smoothly across frames, the spatial domain channel at the n_{F}-th frame is expressed as the superposition of $L_{n_{\text{F}}}$ paths, given by[1]

$$\mathbf{h}^{\text{S}}_{n_{\text{F}}}(t,f) = \sum_{l=1}^{L_{n_{\text{F}}}} g_{n_{\text{F}},l} e^{j2\pi(t\nu_{n_{\text{F}},l} - f\tau_{n_{\text{F}},l})} \mathbf{a}(\theta_{n_{\text{F}},l}) \tag{1}$$

where $g_{n_{\text{F}},l}$, $\tau_{n_{\text{F}},l}$, $\nu_{n_{\text{F}},l}$, and $\theta_{n_{\text{F}},l}$ denote the path gain, delay, Doppler frequency, and directional cosine of the l-th path, respectively, and $\mathbf{a}(\theta)$ denotes the angle domain steering vector, defined by $[\mathbf{a}(\theta)]_{n_{\text{an}}} \triangleq \exp(-j2\pi(n_{\text{an}}-1)\theta)$. By sampling the spatial domain channels at all pilot OFDM symbols and subcarriers, the SFT domain channels are given by

$$\mathbf{h}_{n_{\text{F}}} = \sum_{l=1}^{L_{n_{\text{F}}}} g_{n_{\text{F}},l} \mathbf{a}(\theta_{n_{\text{F}},l}) \otimes \mathbf{b}(\tau_{n_{\text{F}},l}) \otimes \mathbf{c}(\nu_{n_{\text{F}},l}), \tag{2}$$

where $\mathbf{b}(\tau)$ and $\mathbf{c}(\nu)$ denote the delay and Doppler domain steering vectors, defined as $[\mathbf{b}(\tau)]_{n_{\text{sc}}} \triangleq \exp(j2\pi(n_{\text{sc}} - 1)\Delta\bar{f}\tau)$ and $[\mathbf{c}(\nu)]_{n_{\text{sym}}} \triangleq \exp(j2\pi(n_{\text{sym}} - 1)\Delta\bar{T}\nu)$, respectively.

From (2), the SFT domain channels are superposed by several paths, implying the channel sparsity in the ADD domain. To this end, we sample directional cosines, delays, and Doppler frequencies uniformly at intervals $\theta_\Delta = 1/K_{\text{an}}$, $\tau_\Delta = 1/(K_{\text{de}}\Delta\bar{f})$, and $\nu_\Delta = 1/(K_{\text{do}}\Delta\bar{T})$, where K_{an}, K_{de} and K_{do} denote the number of sampling grids in the angle, delay, and Doppler domains, respectively. The sampling sets are defined by

$$\Upsilon_{\text{an}} = \left\{\bar{\theta}_{k_{\text{an}}} : -\frac{1}{2} + (k_{\text{an}} + \frac{1}{2})\theta_\Delta, k_{\text{an}} \in \mathcal{K}_{\text{an}}\right\}, \tag{3a}$$

$$\Upsilon_{\text{de}} = \left\{\bar{\tau}_{k_{\text{de}}} : -\frac{1}{2\Delta\bar{f}} + (k_{\text{de}} + \frac{1}{2})\tau_\Delta, k_{\text{de}} \in \mathcal{K}_{\text{de}}\right\}, \tag{3b}$$

$$\Upsilon_{\text{do}} = \left\{\bar{\nu}_{k_{\text{do}}} : -\frac{1}{2\Delta\bar{T}} + (k_{\text{do}} + \frac{1}{2})\nu_\Delta, k_{\text{do}} \in \mathcal{K}_{\text{do}}\right\}, \tag{3c}$$

where $K_{\text{x}} \triangleq \{0, 1, \ldots, K_{\text{x}} - 1\}, \text{x} \in \{\text{an}, \text{de}, \text{do}\}$ denotes the sampling point index set. By selecting sufficiently large K_{an}, K_{de}, and K_{do}, the SFT domain channels are well approximated as the linear combination of basis functions determined by ADD domain parameters.

Therefore, the SFT domain channels at pilot segments can be approximated as

$$\mathbf{h}_{n_{\text{F}}} \approx \boldsymbol{\Psi}\mathbf{g}_{n_{\text{F}}}, \tag{4}$$

where $\boldsymbol{\Psi} \triangleq \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$ and $\mathbf{g}_{n_{\text{F}}}$ denote the transformation matrix and ADD domain channel, respectively, and each column of $\boldsymbol{\Psi}$ and each element of $\mathbf{g}_{n_{\text{F}}}$ are specified by ADD domain parameter $(\bar{\theta}_{k_{\text{an}}}, \bar{\tau}_{k_{\text{de}}}, \bar{\nu}_{k_{\text{do}}})$, and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ denote the spatial, frequency, and temporal domain transformation matrix, whose columns are given by $\mathbf{a}(\bar{\theta}_{k_{\text{an}}})$, $\mathbf{b}(\bar{\tau}_{k_{\text{de}}})$, and $\mathbf{c}(\bar{\nu}_{k_{\text{do}}})$, respectively. Similarly, the SFT domain channels to be predicted can be approximated as

$$\mathbf{h}^{\text{CP}}_{n_{\text{F}}} \approx \tilde{\boldsymbol{\Psi}}\mathbf{g}_{n_{\text{F}}}, \tag{5}$$

where $\tilde{\boldsymbol{\Psi}} \triangleq \mathbf{A} \otimes \mathbf{B} \otimes \tilde{\mathbf{C}}$ and $\tilde{\mathbf{C}}$ denotes the transformation matrix for channel prediction.

Since the pilot symbols in the uplink training phase are known at BS, the pilot symbols are assumed to be all-one without loss of generality. Therefore, in the uplink training phase, the pilot segment received signal at the n_{sc}-th subcarrier is expressed as

$$\mathbf{y}_{n_{\text{F}}} = \boldsymbol{\Psi}\mathbf{g}_{n_{\text{F}}} + \mathbf{z}_{n_{\text{F}}}, \tag{6}$$

where $\mathbf{z}_{n_{\text{F}}}$ denotes the additive white Gaussian noise (AWGN) vector at pilot segments with zero mean and variance σ_z^2. Facilitated by Doppler domain modeling, channel prediction can be achieved by the transformation from

[1]We focus on the typical MT in the cell and ignore the MT index for simplicity of expressions.

Doppler domain to temporal domain based on the estimation of ADD domain channels.

III. Probabilistic Model and Problem Formulation

With the proposed sparse representation, the dynamic channel prediction problem is to estimate the ADD domain channels based on the received signal and ultimately extrapolate to the SFT domain channels in the non-pilot segments. The ADD domain channel can be expressed as the posterior mean based on the minimum mean square error (MMSE) criterion, given by

$$\mathbf{g}^{\star}_{(N_{\mathrm{F}})} = \int \mathbf{g}_{(N_{\mathrm{F}})} \mathsf{P}(\mathbf{g}_{(N_{\mathrm{F}})} \,|\, \mathbf{y}_{(N_{\mathrm{F}})}) \mathrm{d}\mathbf{g}_{(N_{\mathrm{F}})}, \tag{7}$$

where $\mathsf{P}(\mathbf{g}_{(N_{\mathrm{F}})} \mid \mathbf{y}_{(N_{\mathrm{F}})})$ denotes the posterior probability density function, $\mathbf{y}_{(N_{\mathrm{F}})} \triangleq \{\mathbf{y}_{n_{\mathrm{F}}}\}_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}}$ and $\mathbf{g}_{(N_{\mathrm{F}})} \triangleq \{\mathbf{g}_{n_{\mathrm{F}}}\}_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}}$ denote the sets of received signals and ADD domain channels for N_{F} frames, respectively. From the received signal model and the ADD domain channel representation, $\mathsf{P}(\mathbf{g}_{(N_{\mathrm{F}})} \,|\, \mathbf{y}_{(N_{\mathrm{F}})})$ can be expressed as

$$\mathsf{P}(\mathbf{g}_{(N_{\mathrm{F}})} \,|\, \mathbf{y}_{(N_{\mathrm{F}})}) \propto \mathsf{P}(\mathbf{y}_{(N_{\mathrm{F}})} \,|\, \mathbf{g}_{(N_{\mathrm{F}})}) \mathsf{P}(\mathbf{g}_{(N_{\mathrm{F}})}). \tag{8}$$

Based on the signal model, the channel transfer function can be modeled as

$$\mathsf{P}(\mathbf{y}_{(N_{\mathrm{F}})} \,|\, \mathbf{g}_{(N_{\mathrm{F}})}) = \prod_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \mathsf{P}(\mathbf{y}_{n_{\mathrm{F}}} \,|\, \mathbf{g}_{n_{\mathrm{F}}}), \tag{9a}$$

$$\mathsf{P}(\mathbf{y}_{n_{\mathrm{F}}} \,|\, \mathbf{g}_{n_{\mathrm{F}}}) \propto \mathsf{CN}(\mathbf{y}_{n_{\mathrm{F}}}; \mathbf{\Psi} \mathbf{g}_{n_{\mathrm{F}}}, \sigma_z^2 \mathbf{I}). \tag{9b}$$

The ADD domain channel exhibits three key characteristics:

- Sparsity: Due to the limited scattering propagation environments, the channel is inherently sparse.
- Inter-frame Correlation: The physical parameters vary smoothly across frames, resulting in strong correlations between adjacent frames.
- Intra-frame Correlation: Finite resolution causes energy leakage, inducing significant correlations among neighboring elements within the ADD domain.

Therefore, we model ADD domain channels by

$$\mathbf{g}_{n_{\mathrm{F}}} = (1 - \boldsymbol{\alpha}) \odot \mathbf{g}_{n_{\mathrm{F}}-1} + \boldsymbol{\alpha} \odot \mathbf{w}_{n_{\mathrm{F}}}, \tag{10}$$

where $\mathbf{g}_0 \triangleq \mathbf{0}$, $\boldsymbol{\alpha}$ denotes the transition factor and controls the long-timescale temporal correlations, $\mathbf{w}_{n_{\mathrm{F}}}$ controls the power variations across frames. To capture cluster correlations from the energy leakage effect, $\mathbf{w}_{n_{\mathrm{F}}}$ are modeled as the coupled Gaussian distribution, whose PDF is given by

$$\mathsf{P}(\mathbf{w}_{n_{\mathrm{F}}} \,|\, \boldsymbol{\chi}) \propto \prod_{k} \mathsf{CN}([\mathbf{w}_{n_{\mathrm{F}}}]_k; 0, \underbrace{([\boldsymbol{\chi}]_k^{-1} + \gamma \sum_{\bar{k} \in \mathcal{N}_k} [\boldsymbol{\chi}]_{\bar{k}}^{-1})^{-1}}_{[\bar{\boldsymbol{\chi}}]_k}), \tag{11}$$

where $\boldsymbol{\chi}$ controls the power of $\mathbf{q}_{n_{\mathrm{F}}}$, $\mathcal{N}_k$ denotes the index set of neighboring elements of k-th element in the ADD domain, and γ reflects the strength of the energy leakage effect.

Building upon such probabilistic models, the marginal posterior PDF can be given as

$$\mathsf{P}(\mathbf{g}_{(N_{\mathrm{F}})} \,|\, \mathbf{y}_{(N_{\mathrm{F}})}) \propto \prod_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \mathsf{P}(\mathbf{y}_{n_{\mathrm{F}}} \,|\, \mathbf{g}_{n_{\mathrm{F}}}) \mathsf{P}(\mathbf{g}_{n_{\mathrm{F}}} \,|\, \mathbf{g}_{n_{\mathrm{F}}-1}; \boldsymbol{\alpha}, \boldsymbol{\chi}). \tag{12}$$

Due to the presence of unknown hyperparameters, it is infeasible to obtain the exact marginal posterior PDF as well as the MMSE estimator. Even if the hyperparameters are known, the exact MMSE estimator still involves matrix inversion, which suffers from extremely high computational complexity in large-scale systems. To address this issue, we employ the approximate message passing (AMP) algorithms to approximate the marginal posterior PDF based on the factor graph principle, and learn the unknown hyperparameters through the EM algorithm, which is detailed in the subsequent section.

IV. SD-AMP Based Dynamic Channel Prediction

Based on the probabilistic models above, the factor graph can be partitioned into factor subgraphs of multiple frames, and each subgraph incorporates the linear models.

A. Linear Model

For the linear model, we begin with the equivalent prior model of ADD domain channels, given by

$$\begin{aligned} &\mathsf{P}_{\text{pri}}(\mathbf{g}_{n_{\mathrm{F}}}; \boldsymbol{\rho}, \boldsymbol{\alpha}, \boldsymbol{\chi}) \propto \\ &\quad \mathsf{CN}(\mathbf{g}_{n_{\mathrm{F}}}; (1 - \boldsymbol{\alpha}) \odot \mathbf{g}_{n_{\mathrm{F}}-1}, \mathsf{diag}\{\boldsymbol{\alpha}^{\odot 2} \odot \bar{\boldsymbol{\chi}}\})], \end{aligned} \tag{13}$$

Given this equivalent Gaussian prior model for ADD domain channels in each frame, various statistical inference methods, such as sparse Bayesian learning, can be employed. However, these methods still face computational challenges due to the matrix inversion complexity. With the help of the Taylor series expansion and central limit theorem, the message between $\mathbf{g}_{n_{\mathrm{F}}}$ and $\mathbf{y}_{n_{\mathrm{F}}}$ can be well approximated as the Gaussian distribution [16], and depends only on the nodes instead of edges in the factors subgraph, thus guaranteeing the tractability of the linear model part.

B. Message Across Frames

The straightforward approach for the message exchange across frames involves storing all received signals and performing message passing between all frames. As the number of frames increases, this approach results in significant storage overhead, computational complexity, and processing latency. Inspired by the dynamic AMP algorithm, we further simplify its filtering mode by exploiting only the posterior mean of the ADD domain channels from the previous frame, while ignoring the posterior estimation uncertainty. Specifically, such simplification implies that the message from the factor node $\mathsf{P}(\mathbf{g}_{n_{\mathrm{F}}} \,|\, \mathbf{g}_{n_{\mathrm{F}}-1})$ to the ADD domain channel $\mathbf{g}_{n_{\mathrm{F}}}$ is approximated as

$$\begin{aligned} &\mathsf{M}_{\mathsf{P}_{g,n_{\mathrm{F}}} \to \mathbf{g}_{n_{\mathrm{F}}}} \propto \\ &\quad \mathsf{CN}(\mathbf{g}_{n_{\mathrm{F}}}; (1 - \boldsymbol{\alpha}) \odot \hat{\mathbf{g}}_{n_{\mathrm{F}}}, \mathsf{diag}\{\boldsymbol{\alpha}^{\odot 2} \odot \bar{\boldsymbol{\chi}}\}), \end{aligned} \tag{14}$$

Algorithm 1 SD-AMP Based Dynamic Channel Prediction

Input: Observation signal $\mathbf{y}_{(N_\mathrm{F})}$, Noise variance σ_z^2.
Output: Channel prediction results $\hat{\mathbf{h}}^{\mathrm{CP}}_{(N_\mathrm{F})}$.

1: Initialize the hyperparameters $\boldsymbol{\alpha}^\star$ and $\boldsymbol{\chi}^\star$.
2: Set the initial ADD channel state $\hat{\mathbf{g}}_0 = 0$.
3: **for** $n_\mathrm{F} = 1, \ldots, N_\mathrm{F}$ **do**
4: **for** $t_{\text{out}} = 1 : T_{\text{out}}$ **do**
5: **for** $t_{\text{in}} = 1 : T_{\text{in}}$ **do**
6: $\boldsymbol{\varepsilon}^{y,\mathrm{lik}}_{n_\mathrm{F}} = |\boldsymbol{\Psi}|^{\odot 2}\boldsymbol{\varepsilon}^{g,\mathrm{post}}_{n_\mathrm{F}}$
7: $\boldsymbol{\mu}^{y,\mathrm{lik}}_{n_\mathrm{F}} = \boldsymbol{\Psi}\hat{\mathbf{g}}_{n_\mathrm{F}} - \boldsymbol{\varepsilon}^{y,\mathrm{lik}}_{n_\mathrm{F}} \odot \boldsymbol{\mu}^{y,\mathrm{res}}_{n_\mathrm{F}}$
8: $\boldsymbol{\mu}^{y,\mathrm{res}}_{n_\mathrm{F}} = (\mathbf{y}_{n_\mathrm{F}} - \boldsymbol{\mu}^{y,\mathrm{lik}}_{n_\mathrm{F}}) \oslash (\boldsymbol{\varepsilon}^{y,\mathrm{lik}}_{n_\mathrm{F}} + \sigma_z^2)$
9: $\boldsymbol{\varepsilon}^{y,\mathrm{res}}_{n_\mathrm{F}} = (\boldsymbol{\varepsilon}^{y,\mathrm{lik}}_{n_\mathrm{F}} + \sigma_z^2)^{\odot -1}$
10: $\boldsymbol{\varepsilon}^{g,\mathrm{lik}}_{n_\mathrm{F}} = (|\boldsymbol{\Psi}^H|^{\odot 2}\boldsymbol{\varepsilon}^{y,\mathrm{res}}_{n_\mathrm{F}})^{\odot -1}$
11: $\boldsymbol{\mu}^{g,\mathrm{lik}}_{n_\mathrm{F}} = \hat{\mathbf{g}}_{n_\mathrm{F}} + \boldsymbol{\varepsilon}^{g,\mathrm{lik}}_{n_\mathrm{F}} \odot (\boldsymbol{\Psi}^H \boldsymbol{\mu}^{y,\mathrm{res}}_{n_\mathrm{F}})$
12: $\boldsymbol{\mu}^{g,\mathrm{pri}}_{n_\mathrm{F}} = (1-\boldsymbol{\alpha}) \odot \hat{\mathbf{g}}_{n_\mathrm{F}}, \boldsymbol{\varepsilon}^{g,\mathrm{pri}}_{n_\mathrm{F}} = \boldsymbol{\alpha}^{\odot 2} \odot \bar{\boldsymbol{\chi}}$
13: $\boldsymbol{\varepsilon}^{g,\mathrm{post}}_{n_\mathrm{F}} = (\boldsymbol{\varepsilon}^{g,\mathrm{lik}}_{n_\mathrm{F}} \odot \boldsymbol{\varepsilon}^{g,\mathrm{pri}}_{n_\mathrm{F}}) \oslash (\boldsymbol{\varepsilon}^{g,\mathrm{pri}}_{n_\mathrm{F}} + \boldsymbol{\varepsilon}^{g,\mathrm{lik}}_{n_\mathrm{F}})$
14: $\hat{\mathbf{g}}_{n_\mathrm{F}} = \boldsymbol{\varepsilon}^{g,\mathrm{post}}_{n_\mathrm{F}} \odot (\boldsymbol{\mu}^{g,\mathrm{lik}}_{n_\mathrm{F}} \odot \boldsymbol{\varepsilon}^{g,\mathrm{pri}}_{n_\mathrm{F}} + \boldsymbol{\mu}^{g,\mathrm{pri}}_{n_\mathrm{F}} \odot \boldsymbol{\varepsilon}^{g,\mathrm{lik}}_{n_\mathrm{F}})$
15: **end for**
16: Refine the hyperparameters $\boldsymbol{\alpha}^\star$ and $\boldsymbol{\chi}^\star$.
17: **end for**
18: Predict the channel $\hat{\mathbf{h}}^{\mathrm{CP}}_{n_\mathrm{F}}$ based on (19).
19: **end for**

where $\hat{\mathbf{g}}_{n_\mathrm{F}}$ denotes the posterior mean of the previous frame. Given the AMP algorithm in the subgraph of linear models, the posterior PDF of $\mathbf{g}_{n_\mathrm{F}}$ can be approximated by

$$\begin{aligned} \mathsf{b}^{\mathrm{post}}(\mathbf{g}_{n_\mathrm{F}}) &\propto \mathsf{M}_{\mathsf{P}_{g,n_\mathrm{F}} \to \mathbf{g}_{n_\mathrm{F}}} \mathsf{M}_{\mathsf{P}_{y,n_\mathrm{F}} \to \mathbf{g}_{n_\mathrm{F}}} \\ &\propto \mathsf{CN}(\mathbf{g}_{n_\mathrm{F}}; \boldsymbol{\mu}^{g}_{n_\mathrm{F}}, \mathsf{diag}\{\boldsymbol{\varepsilon}^{g,\mathrm{post}}_{n_\mathrm{F}}\}) \end{aligned} \tag{15}$$

where $\mathsf{M}_{\mathsf{P}_{y,n_\mathrm{F}} \to \mathbf{g}_{n_\mathrm{F}}}$ is the message from the factor node $\mathsf{P}(\mathbf{y}_{n_\mathrm{F}} \mid \mathbf{g}_{n_\mathrm{F}})$ to the ADD domain channel $\mathbf{g}_{n_\mathrm{F}}$ and exhibits the Gaussian form. Therefore, the posterior estimation of $\mathbf{g}_{n_\mathrm{F}}$ can be given by the Gaussian reproduction lemma.

C. Hyperparameter Learning

The hyperparameters of the AR model are assumed to be known in the inter- and intra-frame message passing process, which should be learned adaptively based on the environment in practical systems. Following the EM algorithm, the learning rules of transition factor $\boldsymbol{\alpha}$ can be expressed as element-wise quadratic equations, given by

$$(N_\mathrm{F} - 1)[\bar{\boldsymbol{\chi}}]_k [\boldsymbol{\alpha}]_k^2 + [\boldsymbol{\lambda}_1]_k [\boldsymbol{\alpha}]_k + [\boldsymbol{\lambda}_0]_k = 0, \forall k, \tag{16}$$

where the coefficients are defined as

$$\boldsymbol{\lambda}_1 = \sum_{n_\mathrm{F}=2}^{N_\mathrm{F}} (|\hat{\mathbf{g}}_{n_\mathrm{F}-1}|^{\odot 2} - \mathsf{Re}\{\hat{\mathbf{g}}^*_{n_\mathrm{F}-1} \odot \hat{\mathbf{g}}_{n_\mathrm{F}}\}), \tag{17a}$$

$$\boldsymbol{\lambda}_0 = -\sum_{n_\mathrm{F}=2}^{N_\mathrm{F}} (|\hat{\mathbf{g}}_{n_\mathrm{F}} - \hat{\mathbf{g}}_{n_\mathrm{F}-1}|^{\odot 2} + \boldsymbol{\varepsilon}^{g,\mathrm{post}}_{n_\mathrm{F}}), \tag{17b}$$

where $\bar{\boldsymbol{\chi}}$ denotes the uncoupled variance hyperparameters and can be learned by EM algorithms. Therefore, the optimal transfer factor learning can be achieved by the rooting formula, and it should be noted that only results in the range of $(0, 1)$ are selected. Due to the coupling in power variation hyperparameters, it is intractable to find an analytic solution for $\boldsymbol{\chi}^\star$. To address this, we first learn hyperparameter $\bar{\boldsymbol{\chi}}$ instead and then approximate $\boldsymbol{\chi}^\star$ as [17], [18]

$$[\boldsymbol{\chi}^\star]_k \approx [\bar{\boldsymbol{\chi}}]_k + \gamma \sum_{\bar{k} \in \mathcal{N}_k} [\bar{\boldsymbol{\chi}}]_{\bar{k}}. \tag{18}$$

The detailed proof of the hyperparameter learning rules can be found in the full version of this paper [19].

D. Channel Prediction

Building upon the ADD domain representations of SFT domain channels, the channel prediction can be achieved by

$$\hat{\mathbf{h}}^{\mathrm{CP}}_{n_\mathrm{F}} = \tilde{\boldsymbol{\Psi}} \hat{\mathbf{g}}_{n_\mathrm{F}}. \tag{19}$$

Furthermore, to avoid large delays caused by the collection of pilot OFDM symbols, we employ channel prediction based on sliding windows instead of frames. This approach enables dynamic frame reorganization as new pilot OFDM symbols become available, facilitating real-time channel prediction. The overall dynamic channel prediction algorithm is summarized in Algorithm 1, where T_{in} and T_{out} denote the number of iterations for SD-AMP and EM algorithms, respectively, and Line 6 and 10 can be further simplified due to the constant modulus property of transformation matrix $\boldsymbol{\Psi}$.

An alternative interpretation of the EM algorithm arises from the variational free energy perspective, which can be viewed as one of the message passing rules under an uninformative Dirac delta prior model. This perspective offers useful intuition for message scheduling design, suggesting that the E-step need not be fully converged before proceeding to the M-step, which can alleviate the computational burden when hyperparameters are still unreliable in the early iterations.

TABLE I
SCENARIO PARAMTERS

Paramter	Value
Carrier Frequency	$f_\mathrm{c} = 6.7$ GHz
Pilot Symbol Interval	$\Delta\bar{T} = 14 \times 35.68$ μs
Pilot Subcarrier Spacing	$\Delta\bar{f} = 4 \times 30$ kHz
Number of BS Antennas	$N_{\text{an}} = 32$
Number of Pilot Subcarriers	$N_{\text{sc}} = 64$
Number of Pilot Symbols	$N_{\text{sym}} = 8$
Velocity of MTs	$v = 60$ km/h

V. SIMULATION RESULTS

A. Configuration and Benchmarks

In this section, we present simulation results to evaluate the channel prediction performance of the proposed algorithm. We employ the QuaDRiGa channel simulator [20], which generates massive MIMO-OFDM channels consistent with 3GPP NR specifications [21]. We consider the 3GPP urban macro (UMa) non-line-of-sight (NLOS) scenarios, and the spatial consistency of the channels is enabled. Unless specified otherwise, the simulation configurations follow Section II, with parameters summarized in Table I.

To demonstrate the superiority of the proposed dynamic channel prediction algorithm, we select the following state-of-the-art algorithms as benchmarks:

- **VKF** [12]: Estimates spatial domain channels by the least square (LS) algorithm, with temporal correlations captured by the AR-based vector KF.
- **FIT** [15]: Estimates angle-delay domain channels by the alternating LS (ALS) algorithm, with temporal correlations captured by the first-order Taylor series.
- **PAD** [5]: Estimates angle-delay domain channels by the linear MMSE (LMMSE) algorithm, with temporal correlations captured by the Prony method.

Since **VKF** predicts channels only for future pilot symbols (i.e., fixed-length channel prediction), we predict channels on non-pilot symbols through MMSE interpolation, with prior knowledge of the maximum Doppler frequency and transmission power. The time-averaged normalized mean square error (TNMSE) of the spatial-frequency domain channels is adopted as the metric, defined by

$$\text{TNMSE} = \frac{1}{N_{\text{F}}} \sum_{n_{\text{F}}=1}^{N_{\text{F}}} \frac{\|\hat{\mathbf{h}}_{n_{\text{F}}}^{\text{CP}} - \mathbf{h}_{n_{\text{F}}}^{\text{CP}}\|_2^2}{\|\mathbf{h}_{n_{\text{F}}}^{\text{CP}}\|_2^2}. \tag{20}$$

B. Computational Complexity Analysis

The computational complexity of the proposed algorithm is dominated by matrix–vector multiplications, which can be significantly reduced by exploiting the Kronecker structure of the transformation matrices in the ADD domain. Specifically, the matrix–vector multiplication can be performed domain by domain; while focusing on the current domain, the remaining domains can be processed in batches. The computational complexity order of low-complexity matrix-vector multiplications in the SFT domains are defined by

$$C_{\text{an}} \triangleq N_{\text{an}} K_{\text{an}} (K_{\text{de}} K_{\text{do}} + N_{\text{sc}} N_{\text{sym}}), \tag{21a}$$

$$C_{\text{sc}} \triangleq N_{\text{sc}} K_{\text{de}} (N_{\text{an}} K_{\text{do}} + K_{\text{an}} N_{\text{sym}}), \tag{21b}$$

$$C_{\text{sym}} \triangleq N_{\text{sym}} K_{\text{do}} (N_{\text{an}} N_{\text{sc}} + K_{\text{an}} K_{\text{de}}), \tag{21c}$$

respectively. Therefore, the computational complexity of Algorithm 1 is expressed as $\mathcal{O}((C_{\text{an}} + C_{\text{sc}} + C_{\text{sym}}))$, where $T = T_{\text{out}} T_{\text{in}}$ denotes the number of iterations. To gain further insight into the complexity reduction achieved through domain-by-domain processing, we assume that the size of each dimension in the ADD-domain channel is on the same order as in the SFT-domain channel, which is typically the case in practical systems. Under this assumption, the complexity can be further simplified to $\mathcal{O}(T N_{\text{an}} N_{\text{sc}} N_{\text{sym}} (N_{\text{an}} + N_{\text{sc}} + N_{\text{sym}}))$, whereas ignoring the Kronecker structure leads to a computational complexity of $\mathcal{O}(T N_{\text{an}}^2 N_{\text{sc}}^2 N_{\text{sym}}^2)$. For inference algorithms such as expectation propagation or SBL, the Kronecker structure cannot be directly exploited. As a result, the cubic complexity renders these methods computationally prohibitive under the current simulation configuration.

In the bencmarks, the computational complexity of **PAD** is dominated by the LMMSE algorithm for each pilot symbol and subcarrier, which is $\mathcal{O}(N_{\text{an}}^3 N_{\text{sc}} N_{\text{sym}})$. For **VKF**, the main contributor to computational complexity is the AR parameter learning, expressed as $\mathcal{O}(P^3 N_{\text{an}}^3 N_{\text{sc}})$, where P denotes the AR order, typically on the same order as the pilot OFDM symbols. In the case of **FIT**, its computational complexity is controlled by the ALS algorithm with $\mathcal{O}(T R N_{\text{an}} N_{\text{sc}} N_{\text{sym}})$, where T denotes the number of iterations and R denotes the predefined tensor rank, typically set as the maximum number of possible paths in the environment.

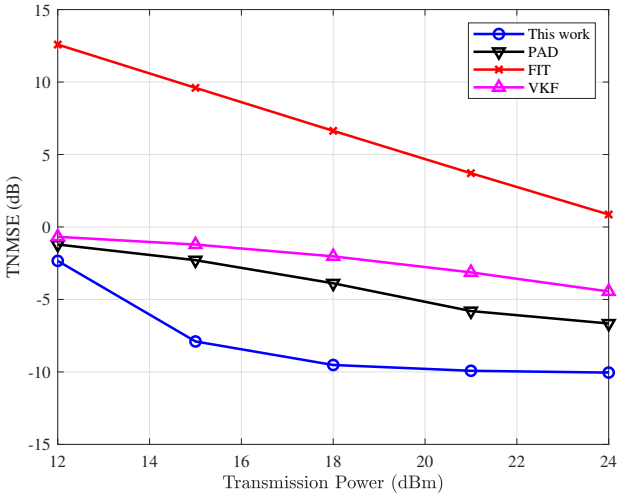

Fig. 2. TNMSE versus transmission power.

C. Performance Evaluation and Discussion

1) Transmission Power: The channel prediction performance of the proposed algorithm and benchmark schemes as a function of transmission power is illustrated in Fig. 2. Since **FIT** is unable to incorporate statistical CSI and noise variance, it fails to achieve a TNMSE below 0 dB throughout the entire transmission power range. Although **VKF** leverages high-order AR processes and achieves better performance than **FIT**, it still suffers from the omission of inter-subcarrier and inter-antenna correlations. Consequently, **VKF** is unable to attain a TNMSE below -5 dB, even at $P_{\text{T}} = 24$ dBm. In contrast, **PAD** effectively exploits statistical CSI through LMMSE-based channel denoising and captures temporal correlations using the Prony method. As a result, it outperforms both **FIT** and **VKF** across the entire transmission power range. Overall, the proposed SD-AMP based dynamic channel prediction algorithm in Algorithm 1 achieves superior performance compared to all benchmarks, demonstrating the benefits of Doppler domain and structured prior modeling in enhancing channel prediction accuracy.

2) Prediction Length: To analyze the impact of prediction length on channel prediction performance, we evaluate the performance at $P_{\text{T}} = 18$ dBm for symbols between the current pilot symbols and the first future pilot symbols, as shown in Fig. 3. As expected, the performance of both the proposed algorithm and the benchmark schemes degrades with increasing prediction length, consistent with previous studies on channel prediction. Since **FIT** captures temporal correlations using a first-order Taylor series approximation, its modeling error increases as the prediction length grows,

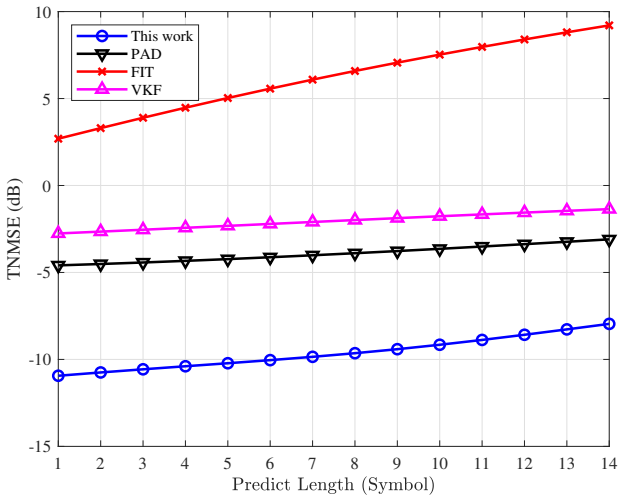

Fig. 3. TNMSE versus predict length ($P_{\mathrm{T}} = 18$ dBm).

resulting in a 7 dB drop in TNMSE. In contrast, the TNMSE of **PAD** and **VKF** remains more stable with increasing prediction length, indicating that the Prony method and AR models provide a more robust representation of temporal correlations in moderate- to high-mobility scenarios compared to the first-order Taylor series. However, due to the under-utilization of inter-antenna and inter-subcarrier correlations, **VKF** struggles to maintain satisfactory performance across all prediction lengths, while **PAD** demonstrates significantly better performance than **VKF**. The proposed algorithm consistently outperforms all benchmark schemes across different prediction lengths. Specifically, for the first future non-pilot symbols and pilot symbols, the proposed algorithm achieves TNMSEs below -10 dB and -8 dB, which are beyond the reach of all benchmarks.

VI. Conclusion

In this paper, we presented the dynamic channel prediction framework for massive MIMO systems under temporal non-stationarity. By modeling intra-frame correlations in the Doppler domain and inter-frame correlations via AR processes, the SFT domain channel and its ADD domain representation are established. The incorporation of AR model and pattern-coupled variances effectively capture the inter-frame and cluster correlations due to temporal non-stationarity and energy leakage effects. To achieve efficient channel prediction, we developed the SD-AMP based channel prediction algorithm, where hyperparameters are learned via the EM approach. Numerical simulations validated the superiority of the proposed method over existing benchmarks, demonstrating enhanced prediction accuracy and improved robustness to non-stationary variations.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62371122 and the China Scholarship Council (CSC). EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG and Norton LifeLock.

References

[1] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

[2] M. Guillaud and D. T. Slock, "A specular approach to MIMO frequency-selective channel tracking and prediction," in *Proc. IEEE Signal Process. Advances in Wireless Commun.*, 2004, pp. 59–63.

[3] R. O. Adeogun, P. D. Teal, and P. A. Dmochowski, "Extrapolation of MIMO mobile-to-mobile wireless channels using parametric-model-based prediction," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4487–4498, Oct. 2015.

[4] R. Adeogun, P. Teal, and P. Dmochowski, "Parametric schemes for prediction of wideband MIMO wireless channels," *arXiv preprint arXiv:1408.0581*, 2014.

[5] 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.

[6] Y. Zhu, J. Zhuang, G. Sun, H. Hou, L. You, and W. Wang, "Joint channel estimation and prediction for massive MIMO with frequency hopping sounding," *IEEE Trans. Wireless Commun.*, 2024.

[7] 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.

[8] Y. Wan and A. Liu, "A two-stage 2D channel extrapolation scheme for TDD 5G NR systems," *IEEE Trans. Wireless Commun.*, 2024.

[9] S. Srivastava, M. S. Kumar, A. Mishra, S. Chopra, A. K. Jagannatham, and L. Hanzo, "Sparse doubly-selective channel estimation techniques for OSTBC MIMO-OFDM systems: A hierarchical Bayesian Kalman filter based approach," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4844–4858, Aug. 2020.

[10] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and L. Hanzo, "Sparse, group-sparse, and online Bayesian learning aided channel estimation for doubly-selective mmWave hybrid MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5843–5858, Sep. 2021.

[11] 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.

[12] 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.

[13] J. Zhao, H. Xie, F. Gao, W. Jia, S. Jin, and H. Lin, "Time varying channel tracking with spatial and temporal BEM for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5653–5666, Aug. 2018.

[14] X. Liu, W. Wang, X. Song, X. Gao, and G. Fettweis, "Sparse channel estimation via hierarchical hybrid message passing for massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7118–7134, Nov. 2021.

[15] W. Peng, W. Li, W. Wang, X. Wei, and T. Jiang, "Downlink channel prediction for time-varying FDD massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1090–1102, Sep. 2019.

[16] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. 45th Asilomar Conf. Signals, Syst. Comput., Pacific Grov*. IEEE, 2011, pp. 799–803.

[17] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.

[18] J. Fang, L. Zhang, and H. Li, "Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2920–2930, Jun. 2016.

[19] H. Hou, Y. Wang, Y. Zhu, X. Yi, W. Wang, D. Slock, and S. Jin, "A tensor-structured approach to dynamic channel prediction for massive MIMO systems with temporal non-stationarity," *arXiv preprint arXiv:2412.06713*, Dec. 2024.

[20] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.

[21] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 18)," *Technical Report TR 38.901*, 2020.