# A Tensor-Structured Approach to Dynamic Channel Prediction for Massive MIMO Systems with Temporal Non-Stationarity

Hongwei Hou, *Graduate Student Member, IEEE*, Yafei Wang, *Graduate Student Member, IEEE*, Yiming Zhu, *Graduate Student Member, IEEE*, Xinping Yi, *Member, IEEE*, Wenjin Wang, *Member, IEEE*, Dirk T. M. Slock, *Life Fellow, IEEE*, Shi Jin, *Fellow, IEEE*

***Abstract*—In moderate- to high-mobility scenarios, channel state information (CSI) varies rapidly and becomes temporally non-stationary, leading to severe performance degradation in the massive multiple-input multiple-output (MIMO) transmissions. To address this issue, we propose a tensor-structured approach to dynamic channel prediction (TS-DCP) for massive MIMO systems with temporal non-stationarity, exploiting both dual-timescale and cross-domain correlations. Specifically, due to inherent spatial consistency, non-stationary channels over long-timescales can be approximated as stationary on short-timescales, decoupling complicated temporal correlations into more tractable dual-timescale ones. To exploit such property, we propose the sliding frame structure composed of multiple pilot orthogonal frequency-division multiplexing (OFDM) symbols, which capture short-timescale correlations within frames by Doppler domain modeling and long-timescale correlations across frames by Markov/autoregressive processes. Building on this, we develop the Tucker-based spatial-frequency-temporal domain channel model, incorporating angle-delay-Doppler (ADD) domain channels and factor matrices parameterized by ADD domain grids. Furthermore, we model cross-domain correlations of ADD domain channels within each frame, induced by clustered scattering, through the Markov random field and tensor-coupled Gaussian distribution that incorporates high-order neighborhood structures. Following these probabilistic models, we formulate the TS-DCP problem as variational free energy (VFE) minimization, and unify different inference rules through the structure design of trial beliefs. This formulation results in the dual-layer VFE optimization process and yields the online TS-DCP algorithm, where the computational complexity is reduced by exploiting tensor-structured operations. Numerical simulations demonstrate the significant superiority of the proposed algorithm over benchmarks in terms of channel prediction performance.**

***Index Terms*—Massive MIMO, channel prediction, variational inference, tensor representation**

## I. Introduction

DRIVEN by the demand for higher data rates, massive multiple-input multiple-output (MIMO) and orthogonal frequency-division multiplexing (OFDM) techniques are expected to keep playing vital roles in future communication systems thanks to their significant capacity gain and enhanced spectral efficiency [1]–[5]. In massive MIMO-OFDM systems, precoding and detection rely on the sufficiently accurate channel state information (CSI), making CSI acquisition critical in achieving high performance [6]–[8].

In time division duplex (TDD) systems, the channel reciprocity enables the acquisition of CSI in the uplink for use in the downlink, thus reducing CSI acquisition overhead. However, large Doppler frequencies and temporal non-stationarity in moderate- to high-mobility scenarios lead to severe channel aging, causing significant performance degradation. This issue becomes more challenging as future communication systems evolve to the upper mid-band [9], due to the linearity of the Doppler frequency to the carrier frequency. To address these challenges, channel prediction has emerged as a promising solution, which captures temporal correlations in historical CSI to predict future CSI, thereby combating channel aging [10].

### A. Previous Works

In the stationary propagation environments, Doppler domain modeling demonstrates significant superiority among temporal correlation models for channel prediction. Specifically, [11] and [12] investigate Doppler frequency estimation by spectrum estimation algorithms for narrowband and wideband MIMO systems, respectively. Leveraging the high angle and delay resolution of massive MIMO-OFDM systems, the angle-delay domain channel estimation is performed, followed by Doppler frequency extraction from effective angle-delay domain channel taps [13]–[15]. To mitigate the effect of channel estimation error on channel prediction, [16] and [17] investigate the joint channel estimation and prediction in the angle-delay-Doppler (ADD) domain. When it comes to the non-stationary channels, the time-varying Doppler frequencies are approximated by polynomials [18], enabling channel prediction via polynomial Fourier transformation and the orthogonal matching pursuit algorithm. By treating Doppler frequencies as imperfection parameters, the Doppler variations over time are also captured by autoregressive (AR) processes in [19], [20], which facilitate the exploitation of channel sparsity.

Hongwei Hou, Yafei Wang, Yiming Zhu, and Wenjin Wang are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with Purple Mountain Laboratories, Nanjing 211100, China (e-mail: hongweihou@seu.edu.cn; wangyf@seu.edu.cn; ymzhu@seu.edu.cn; wangwj@seu.edu.cn).

Xinping Yi and Shi Jin are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: xyi@seu.edu.cn; jinshi@seu.edu.cn).

Dirk T. M. Slock is with the Department of Communication Systems, EURECOM, 06410 Biot, France (e-mail: Dirk.Slock@eurecom.fr).

Beyond the Doppler domain modeling, the AR model has also been extensively investigated, especially in non-stationary environments. Specifically, channel models based on angle and delay domain sparse representation are presented in [21]–[23], where AR processes capture the temporal correlation of sparse domain channels. As such, sparse Bayesian learning (SBL) and expectation maximization algorithms are employed for channel tracking and AR parameter learning. By incorporating the AR process of angle domain channels, [24] presents the online group SBL algorithm, which exploits the joint sparsity of the angle domain across subcarriers and the structured sparsity within one symbol. To further explore the inter-subcarrier correlations, angle-delay domain sparse representations are developed in [25], where spatial-temporal AR processes capture the residual temporal correlations of neighboring elements. Without the perfect channel knowledge, [26], [27] adopt spatial domain high-order AR processes to enhance the robustness of channel prediction algorithms in practical systems, while [28] employs the deep learning framework. Due to the continuous evolution of propagation environments, the angle variations over time are also captured by AR processes and tracked by the Kalman filter (KF) [29]. In [30], [31], the Markov processes are introduced for the angle and angle-delay domain channel support, capturing the dynamic sparsity. Inspired by AR processes, [32] employs the Taylor series for temporal correlations of angle-delay domain channels and develops a linear regression-based algorithm.

### B. Motivations and Contributions

Channel prediction in moderate- to high-mobility scenarios is challenging due to intractable temporal correlations arising from dynamic propagation environments. However, the inherent spatial consistency of channels offers a practical simplification: non-stationary channels over long-timescales can be approximated as stationary on short-timescales [33]–[35], thus decoupling the complicated temporal correlations into more tractable dual-timescale correlations. Moreover, the clustered distribution of scatterers induces cross-domain correlations with significant structured patterns in both channel sparsity and power [36], [37]. By capturing these correlations, channel prediction performance in dynamic propagation environments can be significantly enhanced, which motivates our work.

In this paper, we investigate the tensor-structured approach to dynamic channel prediction (TS-DCP) for massive MIMO systems with temporal non-stationarity. The main contributions of this paper are summarized as follows:

- To exploit dual-timescale correlations, we propose the sliding frame structure composed of multiple pilot OFDM symbols for channel prediction, characterizing the short- and long-timescale correlations by intra- and inter-frame ones. Specifically, we develop the tensor-structured received signal model in the spatial-frequency-temporal (SFT) domain, incorporating ADD domain channels and factor matrices parameterized by ADD domain grids. In this representation, the intra-frame correlations are captured by Doppler domain modeling, while the inter-frame correlations of support and power of ADD domain channels are modeled using Markov/AR process.
- To capture cross-domain correlations, we introduce the concept of high-order neighbors under the tensor representation, defining the interaction region of cross-domain correlations for each ADD domain grid. Building on this, we employ the Markov random fields (MRF) and tensor-coupled Gaussian distributions (TCGD) to model the ADD domain sparse structures and power variations, which are caused by the clustering distribution of scatterers. Specifically, the MRF characterizes the support of ADD domain channels within each frame, while the TCGD models the power variations in the AR process of inter-frame ADD domain channels, providing more accurate representations of practical channels.
- Building on the probabilistic models, we formulate the TS-DCP problem as variational free energy (VFE) minimization and present the online approximation to enable real-time prediction. By designing the structure of trial beliefs, this problem is interpreted as a dual-layer optimization process, ultimately yielding the online TS-DCP algorithm. In this algorithm, different inference rules are unified under the VFE minimization framework, and tensor-structured operations are leveraged to significantly reduce the computational complexity. Numerical simulations demonstrate the superior channel prediction performance of the proposed algorithms compared to the state-of-the-art benchmarks.

### C. Organization and Notations

*1) Organization:* The remainder of this paper is organized as follows. In Section II, we develop the SFT domain channel model and its Tucker-based representations with factor matrices parameterized by ADD domain grids. Based on probabilistic models, the TS-DCP problem is formulated under the VFE minimization framework in Section III. Following this, the online TS-DCP algorithm is developed in Section IV, where the inner- and outer-layer are detailed respectively. Numerical simulations in Section V demonstrate the superiority of proposed algorithms over benchmarks in terms of the computational complexity and channel prediction performance. Finally, Section VI concludes the paper.

*2) Standard 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. $[\cdot]_{i_1,\ldots,i_D}$ is the $(i_1,\ldots,i_D)$-th element of $D$-order tensor. $\mathsf{D}\{\cdot\}$ and $\mathsf{H}\{\cdot\}$ denote the Kullback-Leibler divergence (KLD) and differential entropy, respectively. The outer and Hadamard products are represented by the operator $\circ$ and $\odot$, respectively. $\mathsf{E}\{\cdot\}$ and $\mathsf{V}\{\cdot\}$ denote the expectation and variance operators, respectively. $\mathsf{diag}\{\cdot\}$ and $\mathsf{Re}\{\cdot\}$ denote the diagonal and real part operators, respectively. The key abbreviations are summarized in Table I.

*3) Tensor Notations:* The tensor operations and definitions in this paper align with the counterparts in [38]. For a $D$-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{N_1 \times N_2 \times \ldots \times N_D}$, the mode-$d$ matrixization $\mathbf{X}_d \in \mathbb{C}^{N_d \times N_1 \ldots N_{d-1} N_{d+1} \ldots N_D}$ arranges the mode-$d$ fibers of

TABLE I
SUMMARY OF KEY ABBREVIATIONS

| Abbreviation | Full Description |
|---|---|
| MIMO | Multiple-Input Multiple-Output |
| OFDM | Orthogonal Frequency-Division Multiplexing |
| TDD | Time Division Duplex |
| CSI | Channel State Information |
| SFT | Spatial-Frequency-Temporal |
| ADD | Angle-Delay-Doppler |
| MRF | Markov Random Field |
| TCGD | Tensor-Structured Coupled Gaussian Distribution |
| MMSE | Minimum Mean Square Error |
| KLD | Kullback-Leibler Divergence |
| VFE | Variational Free Energy |
| BFE | Bethe Free Energy |
| MCC | Marginal Consistency Constraints |
| SSCC | Sufficient Statistical Consistency Constraints |

this tensor, obtained by fixing the index along the $d$-th dimension and varying the others, into its column vectors. The high-order cyclic shift operations is defined by $[\boldsymbol{\mathcal{X}}^{[\mathbf{r}]}]_{n_1,\ldots,n_D} = [\boldsymbol{\mathcal{X}}]_{\bar{n}_1,\ldots,\bar{n}_D}$ with $\bar{n}_d = (n_d + r_d) \mod N_d$. Given the tensors $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}$ with the same size, the inner product is defined as

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{n_1} \sum_{n_2} \cdots \sum_{n_D} [\boldsymbol{\mathcal{X}}]_{n_1,n_2,\ldots,n_D} [\boldsymbol{\mathcal{Y}}]^*_{n_1,n_2,\ldots,n_D}, \quad (1)$$

which is also the high-order extension of matrix inner product, and we define $\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle}$ as the high-order extension of Frobenius norm. We also define $\ell_1$ norm of $\boldsymbol{\mathcal{X}}$, given by $\|\boldsymbol{\mathcal{X}}\|_1 = \sum_{n_1} \sum_{n_2} \cdots \sum_{n_D} |[\boldsymbol{\mathcal{X}}]_{n_1,n_2,\ldots,n_D}|$. The mode-$d$ tensor-matrix multiplication of tensor $\boldsymbol{\mathcal{X}}$ and matrix $\mathbf{U}_d \in \mathbb{C}^{K_d \times N_d}$ is denoted as $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_d \mathbf{U}_d$, given by

$$[\boldsymbol{\mathcal{Y}}]_{n_1,\ldots,n_{d-1},k_d,n_{d+1},\ldots,n_D} = \sum\nolimits_{n_d} [\mathbf{U}]_{k_d,n_d} [\boldsymbol{\mathcal{X}}]_{n_1,\ldots,n_D}. \quad (2)$$

The tensor-matrix multiplication $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_d \mathbf{U}_d$ is equivalent to matrix-matrix multiplication $\mathbf{Y}_d = \mathbf{U}_d \mathbf{X}_d$, and satisfies the commutativity across dimensions. $\mathsf{CN}(\boldsymbol{\mathcal{X}}; \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{E}})$ denotes the joint probability density function (PDF) of $\boldsymbol{\mathcal{X}}$, whose entries are independently drawn from circularly symmetric complex Gaussian distributions with element-wise mean $\boldsymbol{\mathcal{U}}$ and variance $\boldsymbol{\mathcal{E}}$. $\boldsymbol{\mathcal{C}}(x)$ denotes the tensor with all elements being $x$. The element-wise function applications to tensors follow the convention that, for any tensor $\boldsymbol{\mathcal{X}}$ and function $f(\cdot)$, we have $[f(\boldsymbol{\mathcal{X}})]_{n_1,\ldots,n_D} = f([\boldsymbol{\mathcal{X}}]_{n_1,\ldots,n_D})$.

## II. SYSTEM MODEL

In this paper, we consider the massive MIMO-OFDM system with a base station (BS) equipped with a uniform planar array (UPA) of $N_{\text{an}} = N_{\text{h}} \times N_{\text{v}}$ antennas and single-antenna mobile terminals (MTs), where $N_{\text{h}}$ and $N_{\text{v}}$ denote the number of horizontal and vertical antennas, respectively. This system operates in the TDD mode under the assumption of perfect calibration, and the frame structure conforms to the the fifth generation (5G) New Radio (NR) specifications [39]. The OFDM symbol duration, cyclic prefix (CP) duration, and subcarrier spacing are denoted by $\Delta T_{\text{sym}}$, $\Delta T_{\text{cp}}$, and $\Delta f$ respectively, leading to the total OFDM symbol duration with CP given by $\Delta T = \Delta T_{\text{sym}} + \Delta T_{\text{cp}}$.

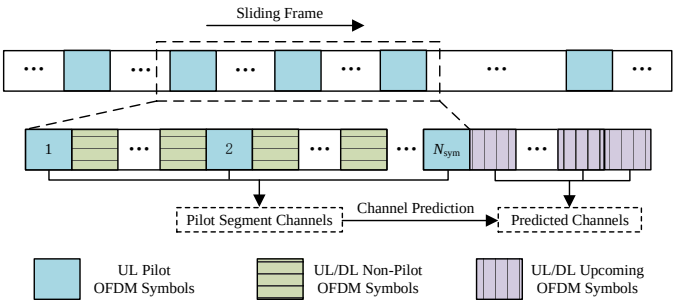



Fig. 1. Sliding frame structure for dynamic channel prediction.

The MTs employ sounding reference signal as pilots for uplink channel sounding, which is characterized by comb-type pilot patterns with uniform spacing in both the temporal and frequency domains. Specifically, pilot symbols are inserted every $\Delta \bar{T} = N_{\text{IS}} \Delta T$ in the temporal domain, and every $\Delta \bar{f} = N_{\text{TC}} \Delta f$ in the frequency domain, where $N_{\text{IS}}$ and $N_{\text{TC}}$ denote the number of pilot interval symbols and transmission combs, respectively. Since channels exhibit inherent spatial consistency in practical systems, the physical parameters of each path remain unchanged on short-timescale and vary smoothly on long-timescale [33]–[35]. To capture this dual-timescale correlation, we introduce the sliding frame structure shown in Fig. 1, where each frame contains multiple pilot OFDM symbols. Within each frame, intra-frame correlations characterize short-timescale channel variations, while inter-frame correlations capture long-timescale dynamic correlations across frames. In each frame, the BS collects the received symbols in the pilot segment and predicts channels of all symbols from the last observed pilot symbol to the first upcoming pilot symbol, thus combating channel aging.

### A. Channel and Signal Models

The channel impulse response (CIR) at the $(n_{\text{h}}, n_{\text{v}})$-th antenna and the $n_{\text{F}}$-th frame is expressed as[1]

$$h_{n_{\text{F}},n_{\text{h}},n_{\text{v}}}(t,\tau) = \sum\nolimits_{l=1}^{L_{n_{\text{F}}}} g_{n_{\text{F}},l} e^{j2\pi t \nu_{n_{\text{F}},l}} \delta(\tau - \tau_{n_{\text{F}},n_{\text{h}},n_{\text{v}},l}), \quad (3)$$

where $L_{n_{\text{F}}}$ denotes the number of paths, $\tau_{n_{\text{F}},n_{\text{h}},n_{\text{v}},l} \triangleq \tau_{n_{\text{F}},l} + [(n_{\text{h}}-1)d_{\text{h},n_{\text{F}},l} + (n_{\text{v}}-1)d_{\text{v},n_{\text{F}},l}]/c$ denote the propagation delay of the $l$-th path from the MT to the $(n_{\text{h}}, n_{\text{v}})$-th under the far-field assumption, $d_{\text{h},n_{\text{F}},l} \triangleq d\mathsf{sin}(\psi_{\text{el},n_{\text{F}},l})\mathsf{sin}(\psi_{\text{az},n_{\text{F}},l})$ and $d_{\text{v},n_{\text{F}},l} \triangleq d\mathsf{cos}(\psi_{\text{el},n_{\text{F}},l})$ denote the propagation distance difference in horizontal and vertical dimensions of the $l$-th path, respectively, $g_{n_{\text{F}},l}$, $\tau_{n_{\text{F}},l}$, $\nu_{n_{\text{F}},l}$, $\psi_{\text{el},n_{\text{F}},l}$, and $\psi_{\text{az},n_{\text{F}},l}$ denote the complex gain, delay, Doppler frequency, elevation and azimuth angle of the $l$-th path, respectively, $c$ and $d$ denote the speed of light and inter-antenna spacing, respectively.

By taking the Fourier transform of CIRs and stacking the result in all antennas and pilot resource elements in the $n_{\text{F}}$-th frame, the SFT domain channel at pilot segments is given as

$$\boldsymbol{\mathcal{H}}_{n_{\text{F}}} = \sum\nolimits_{l=1}^{L_{n_{\text{F}}}} g_{n_{\text{F}},l} \mathbf{a}_{\text{h}}(\theta_{n_{\text{F}},l}) \circ \mathbf{a}_{\text{v}}(\phi_{n_{\text{F}},l}) \circ \mathbf{b}(\tau_{n_{\text{F}},l}) \circ \mathbf{c}(\nu_{n_{\text{F}},l}), \quad (4)$$

[1]Since there is no pilot contamination between MTs, we focus on the typical MT in the cell and ignore the MT index for simplicity of expressions.

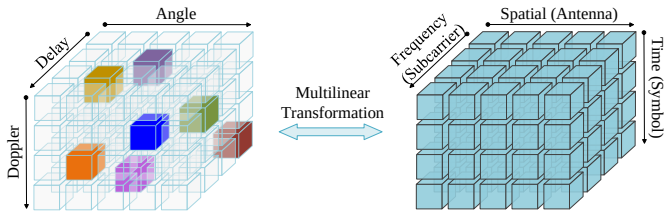



Fig. 2. The diagrams of the SFT domain and ADD domain channel tensors. In this case, only the horizontal angle and antenna domains are shown.

where $\mathbf{a}_{\mathrm{h}}(\theta)$, $\mathbf{a}_{\mathrm{v}}(\phi)$, $\mathbf{b}(\tau)$ and $\mathbf{c}(\nu)$ denote the horizontal angle, vertical angle, delay and Doppler domain steering vectors, defined by $[\mathbf{a}_{\mathrm{h}}(\theta)]_{n_{\mathrm{h}}} \triangleq \exp(-j2\pi(n_{\mathrm{h}}-1)\theta)$, $[\mathbf{a}_{\mathrm{v}}(\phi)]_{n_{\mathrm{v}}} \triangleq \exp(-j2\pi(n_{\mathrm{v}}-1)\phi)$, $[\mathbf{b}(\tau)]_{n_{\mathrm{sc}}} \triangleq \exp(-j2\pi(n_{\mathrm{sc}}-1)\Delta\bar{f}\tau)$ and $[\mathbf{c}(\nu)]_{n_{\mathrm{sym}}} \triangleq \exp(j2\pi(n_{\mathrm{sym}}-1)\Delta\bar{T}\nu)$, respectively, $\theta_{n_{\mathrm{F}},l} \triangleq d_{\mathrm{h},n_{\mathrm{F}},l}/\lambda$ and $\phi_{n_{\mathrm{F}},l} \triangleq d_{\mathrm{v},n_{\mathrm{F}},l}/\lambda$ denote the horizontal and vertical directional cosines of the $l$-th path, respectively, and $\lambda$ denotes the wavelength.

After the cyclic prefix removal and OFDM demodulation, the received signal tensor at pilot segments is expressed as

$$\boldsymbol{\mathcal{Y}}_{n_{\mathrm{F}}} = \boldsymbol{\mathcal{X}} \odot \boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}} + \boldsymbol{\mathcal{Z}}_{n_{\mathrm{F}}}, \tag{5}$$

where $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{N_{\mathrm{h}}\times N_{\mathrm{v}}\times N_{\mathrm{sc}}\times N_{\mathrm{sym}}}$, $\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}} \in \mathbb{C}^{N_{\mathrm{h}}\times N_{\mathrm{v}}\times N_{\mathrm{sc}}\times N_{\mathrm{sym}}}$ and $\boldsymbol{\mathcal{Z}}_{n_{\mathrm{F}}} \in \mathbb{C}^{N_{\mathrm{h}}\times N_{\mathrm{v}}\times N_{\mathrm{sc}}\times N_{\mathrm{sym}}}$ denote the pilot tensor, the SFT domain channel and the additive white Gaussian noise at pilot segments, respectively. Since pilot symbols in the uplink training phase are known at BS, we assume that $\boldsymbol{\mathcal{X}}$ is all-one tensor without loss of generality.

### *B. Tucker-Based Representations*

Due to the limited scattering propagation environments, the SFT domain channels exhibit low-rank structures, which can be exploited by the unstructured tensor decomposition (TD) framework. However, the uniqueness condition of such framework is typically not satisfied in practical propagation environments. This limitation can be mitigated by leveraging the inherent structure of factor matrices [40]–[42].

Based on (4), the factor matrices of SFT domain channels are parameterized by ADD domain grids, defined as

$$\mathbf{A}_{\mathrm{h}}(\tilde{\boldsymbol{\theta}}_{n_{\mathrm{F}}}) = [\mathbf{a}_{\mathrm{h}}(\tilde{\theta}_{1,n_{\mathrm{F}}}),\ldots,\mathbf{a}_{\mathrm{h}}(\tilde{\theta}_{K_{\mathrm{h}},n_{\mathrm{F}}})] \in \mathbb{C}^{N_{\mathrm{h}}\times K_{\mathrm{h}}}, \tag{6a}$$

$$\mathbf{A}_{\mathrm{v}}(\tilde{\boldsymbol{\phi}}_{n_{\mathrm{F}}}) = [\mathbf{a}_{\mathrm{v}}(\tilde{\phi}_{1,n_{\mathrm{F}}}),\ldots,\mathbf{a}_{\mathrm{v}}(\tilde{\phi}_{K_{\mathrm{v}},n_{\mathrm{F}}})] \in \mathbb{C}^{N_{\mathrm{v}}\times K_{\mathrm{v}}}, \tag{6b}$$

$$\mathbf{B}(\tilde{\boldsymbol{\tau}}_{n_{\mathrm{F}}}) = [\mathbf{b}(\tilde{\tau}_{1,n_{\mathrm{F}}}),\ldots,\mathbf{b}(\tilde{\tau}_{K_{\mathrm{de}},n_{\mathrm{F}}})] \in \mathbb{C}^{N_{\mathrm{sc}}\times K_{\mathrm{de}}}, \tag{6c}$$

$$\mathbf{C}(\tilde{\boldsymbol{\nu}}_{n_{\mathrm{F}}}) = [\mathbf{c}(\tilde{\nu}_{1,n_{\mathrm{F}}}),\ldots,\mathbf{c}(\tilde{\nu}_{K_{\mathrm{do}},n_{\mathrm{F}}})] \in \mathbb{C}^{N_{\mathrm{sym}}\times K_{\mathrm{do}}}, \tag{6d}$$

where $K_{\mathrm{x}}$, $\mathrm{x} \in \{\mathrm{h},\mathrm{v},\mathrm{de},\mathrm{do}\}$ denote the number of grids, $\boldsymbol{\theta}_{n_{\mathrm{F}}}$, $\boldsymbol{\phi}_{n_{\mathrm{F}}}$, $\boldsymbol{\tau}_{n_{\mathrm{F}}}$, and $\boldsymbol{\nu}_{n_{\mathrm{F}}}$ denote the ADD domain grids. Following most prior work on dynamic grids, the numbers of angle, delay, and Doppler domain grids are typically set proportional to the numbers of antennas, subcarriers, and OFDM symbols, respectively, thereby offering trade-offs between computational complexity and model flexibility.

Therefore, the SFT domain channels are represented by the Tucker-based model, given by

$$\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}} = \boldsymbol{\mathcal{G}}_{n_{\mathrm{F}}} \times_1 \mathbf{A}_{\mathrm{h}}(\tilde{\boldsymbol{\theta}}_{n_{\mathrm{F}}}) \times_2 \mathbf{A}_{\mathrm{v}}(\tilde{\boldsymbol{\phi}}_{n_{\mathrm{F}}}) \times_3 \mathbf{B}(\tilde{\boldsymbol{\tau}}_{n_{\mathrm{F}}}) \times_4 \mathbf{C}(\tilde{\boldsymbol{\nu}}_{n_{\mathrm{F}}}), \tag{7}$$

where $\boldsymbol{\mathcal{G}}_{n_{\mathrm{F}}} \in \mathbb{C}^{K_{\mathrm{h}}\times K_{\mathrm{v}}\times K_{\mathrm{de}}\times K_{\mathrm{do}}}$ denotes ADD domain channels, with elements being the path gain of the specific ADD domain grid and corresponding to the specific rank-one multi-path component. Based on this representation, channel prediction can be achieved by ADD domain channel acquisition, which leverages the inter-antenna, inter-subcarrier, and inter-symbol correlations. The specific SFT and ADD domain channels are illustrated in Fig. 2, where only the horizontal angle and antenna domains are shown for visualization. It can be observed that the ADD domain channels contain non-zero values contributed by only a few scatterers, revealing the inherent low-rank structure of SFT domain channels.

**Remark 1.** The Tucker-based model captures the multi-linear structure of channels, offering the following advantages:

- **Concise representation**: The high-order structure of the ADD domain channels allows for the concise representation of high-order neighbors in structured prior models.
- **Efficient computation**: The low-dimensional factor matrix in each domain enables significantly reduced computational complexity of the multi-linear model.

## III. Problem Formulation

Although the multi-linear rank is crucial for ADD domain channel acquisition, its optimal selection remains challenging due to the presence of noise. This challenge motivates the introduction of Bayesian perspectives, which inherently enable automatic rank determination and eliminate the need for the explicit estimation [43]–[46]. Toward this end, we develop probabilistic models for channel prediction with temporal non-stationarity, formulating the dynamic channel prediction problem based on the VFE minimization. Following this, the online approximation and dual-layer optimization of the VFE is achieved through the structure design of trail beliefs.

### *A. Probabilistic Models*

*1) Observation and Multi-linear Transformation:* Following (5), the PDF of the observation model is expressed as

$$\mathsf{P}(\boldsymbol{\mathcal{Y}}_{(N_{\mathrm{F}})} \,|\, \boldsymbol{\mathcal{H}}_{(N_{\mathrm{F}})}) = \prod\nolimits_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \underbrace{\mathsf{CN}(\boldsymbol{\mathcal{Y}}_{n_{\mathrm{F}}}; \boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{C}}(\sigma_z^2))}_{\mathsf{P}_{Y,n_{\mathrm{F}}}}, \tag{8}$$

where $\boldsymbol{\mathcal{Y}}_{(N_{\mathrm{F}})} \triangleq \{\boldsymbol{\mathcal{Y}}_{n_{\mathrm{F}}}\}_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}}$ and $\boldsymbol{\mathcal{H}}_{(N_{\mathrm{F}})} \triangleq \{\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}\}_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}}$ denote the sets of observations and SFT domain channels for $N_{\mathrm{F}}$ frames, respectively.

For multi-linear transformations, it is challenging to explicitly capture the dependencies between SFT domain channels and ADD domain grids. To simplify these dependencies, we uniformly sample the ADD domain to replace the dynamic grids in (7) with fixed sampling grids. This sampling strategy provides accurate approximations of the continuous physical parameters when the numbers of antennas, subcarriers, and OFDM symbols are sufficiently large. However, the deviations between sampling grids and ground-truth parameters are inevitable in the practical systems with limited dimensions, which can be captured through ADD domain perturbation parameters on top of uniformly sampling grids. Specifically, we define the ADD domain perturbed grids as $\tilde{\boldsymbol{\chi}}_{n_{\mathrm{F}}} \triangleq \bar{\boldsymbol{\chi}} + \boldsymbol{\Delta\chi}_{n_{\mathrm{F}}}$

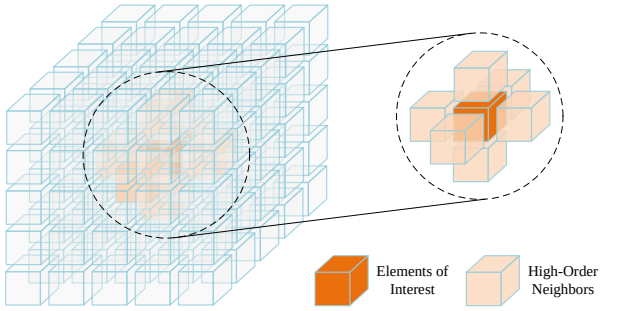



Fig. 3. The three-dimensional example of high-order neighbors. In this case, we have $\mathcal{N} \triangleq \{(0,0,\pm1),(0,\pm1,0),(\pm1,0,0)\}$.

for $\boldsymbol{\chi} \in \{\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\tau}, \boldsymbol{\nu}\}$, where $\bar{\boldsymbol{\chi}}$ and $\boldsymbol{\Delta\chi}_{n_\mathrm{F}}$ denote the sampling grids and perturbation parameters, respectively. Although the resolution of uniform sampling is typically determined by the discrete Fourier transform, slight oversampling can further improve model accuracy with only a marginal increase in the computational complexity. By treating perturbation parameters as imperfection parameters due to the off-grid effect, the PDF of multi-linear transformation is given as

$$\mathsf{P}(\boldsymbol{\mathcal{H}}_{(N_\mathrm{F})} \,|\, \boldsymbol{\mathcal{G}}_{(N_\mathrm{F})}; \mathcal{P}_{\mathrm{OG},(N_\mathrm{F})}) = \prod\nolimits_{n_\mathrm{F}=1}^{N_\mathrm{F}} \underbrace{\mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}} \,|\, \boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \mathcal{P}_{\mathrm{OG},n_\mathrm{F}})}_{\mathsf{P}_{HG,n_\mathrm{F}}}, \tag{9a}$$

$$\begin{aligned}\mathsf{P}_{HG,n_\mathrm{F}} \propto \delta(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}} - \boldsymbol{\mathcal{G}}_{n_\mathrm{F}} \times_1 \mathbf{A}_\mathrm{h}(\bar{\boldsymbol{\theta}} + \boldsymbol{\Delta\theta}_{n_\mathrm{F}}) \times_2 \mathbf{A}_\mathrm{v}(\bar{\boldsymbol{\phi}} + \boldsymbol{\Delta\theta}_{n_\mathrm{F}}) \\ \times_3 \mathbf{B}(\bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau}_{n_\mathrm{F}}) \times_4 \mathbf{C}(\bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu}_{n_\mathrm{F}})),\end{aligned} \tag{9b}$$

where $\boldsymbol{\mathcal{G}}_{(N_\mathrm{F})} \triangleq \{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\}_{n_\mathrm{F}=1}^{N_\mathrm{F}}$ denotes the set of ADD domain channels for $N_\mathrm{F}$ frames, $\mathcal{P}_{\mathrm{OG},(N_\mathrm{F})} \triangleq \cup_{n_\mathrm{F}=1}^{N_\mathrm{F}} \mathcal{P}_{\mathrm{OG},n_\mathrm{F}}$ denotes the sets of perturbation parameters in all frames, and $\mathcal{P}_{\mathrm{OG},n_\mathrm{F}} \triangleq \{\boldsymbol{\Delta\theta}_{n_\mathrm{F}}, \boldsymbol{\Delta\phi}_{n_\mathrm{F}}, \boldsymbol{\Delta\tau}_{n_\mathrm{F}}, \boldsymbol{\Delta\nu}_{n_\mathrm{F}}\}$.

*2) ADD Domain Channels:* Beyond sparsity, the ADD domain channels exhibit long-timescale correlations due to spatial consistency and cross-domain correlations due to the cluster scattering, which can be captured to further enhance channel prediction. Specifically, the ADD domain channels are modeled by two hidden random processes, expressed as

$$\mathsf{P}(\boldsymbol{\mathcal{G}}_{(N_\mathrm{F})} \,|\, \boldsymbol{\mathcal{S}}_{(N_\mathrm{F})}, \boldsymbol{\mathcal{Q}}_{(N_\mathrm{F})}) = \prod\nolimits_{n_\mathrm{F}=1}^{N_\mathrm{F}} \underbrace{\delta(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}} - \boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \odot \boldsymbol{\mathcal{Q}}_{n_\mathrm{F}})}_{\mathsf{P}_{GSQ,n_\mathrm{F}}}, \tag{10}$$

where $\boldsymbol{\mathcal{S}}_{(N_\mathrm{F})} \triangleq \{\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}\}_{n_\mathrm{F}=1}^{N_\mathrm{F}}$, and $\boldsymbol{\mathcal{Q}}_{(N_\mathrm{F})} \triangleq \{\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}\}_{n_\mathrm{F}=1}^{N_\mathrm{F}}$, the binary-valued tensor $\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}$ and complex-valued tensor $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$ describe the sparsity and power of the ADD domain channel, respectively. Since ADD domain channel tensor elements that are not visible in one frame may become visible in subsequent frames under environments with tenporal non-stationarity, it is crucial to employ hidden value tensors for the modeling of all potential path gains [30].

Due to the smooth variation of the physical parameters of each path across frames, the hidden support tensor is modeled as a first-order Markov process, given by

$$\mathsf{P}(\boldsymbol{\mathcal{S}}_{(N_\mathrm{F})}; \boldsymbol{\mathcal{M}}) = \prod\nolimits_{n_\mathrm{F}=1}^{N_\mathrm{F}} \mathsf{P}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \,|\, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}; \boldsymbol{\mathcal{M}}), \tag{11}$$

where $\boldsymbol{\mathcal{S}}_0 \triangleq \boldsymbol{\mathcal{C}}(0)$, $\boldsymbol{\mathcal{M}}$ denotes the transition factor and controls the sparsity transition probability across frames. In this model, $\mathsf{P}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \,|\, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}; \boldsymbol{\mathcal{M}})$ denotes the transition probabilistic models of the hidden support tensor, given by

$$\mathsf{P}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \,|\, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}; \boldsymbol{\mathcal{M}}) = \mathsf{P}_{\mathrm{MRF}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}) \mathsf{P}_{\mathrm{TC}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \,|\, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}; \boldsymbol{\mathcal{M}}), \tag{12}$$

where $\mathsf{P}_{\mathrm{TC}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \mid \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}; \boldsymbol{\mathcal{M}})$ denotes the long-timescale probabilistic model across frames, given by

$$\begin{aligned}\mathsf{P}_{\mathrm{TC}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \,|\, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}; \boldsymbol{\mathcal{M}}) \propto \\ \exp(\langle \boldsymbol{\mathcal{M}}, (2\boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1} - \boldsymbol{\mathcal{C}}(1)) \odot (2\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} - \boldsymbol{\mathcal{C}}(1)) \rangle),\end{aligned} \tag{13}$$

where the conversion from binary to bipolar support tensor is inspired by the Ising model [47]–[49] and $\mathsf{P}_{\mathrm{MRF}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}})$ denotes the intra-frame cross-domain probabilistic model, given by

$$\begin{aligned}\mathsf{P}_{\mathrm{MRF}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}) \propto \\ \prod\nolimits_{\mathbf{r} \in \mathcal{N}} \exp(\gamma \langle (2\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} - \boldsymbol{\mathcal{C}}(1)), (2\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}^{[\mathbf{r}]} - \boldsymbol{\mathcal{C}}(1)) \rangle),\end{aligned} \tag{14}$$

where $\gamma$ controls the strength of cross-domain correlations, and $\mathcal{N}$ denotes the relative index set of high-order neighbors, whose three-dimensional example is shown in Fig. 3.

Similarly, the hidden value tensor is modeled based on the hidden value tensor of the previous frame, given by

$$\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}} = (\boldsymbol{\mathcal{C}}(1) - \boldsymbol{\mathcal{L}}) \odot \boldsymbol{\mathcal{Q}}_{n_\mathrm{F}-1} + \boldsymbol{\mathcal{L}} \odot \boldsymbol{\mathcal{W}}_{n_\mathrm{F}}, \tag{15}$$

where $\boldsymbol{\mathcal{L}}$ denotes the transition factor and controls the long-timescale temporal correlations, $\boldsymbol{\mathcal{Q}}_0 \triangleq \boldsymbol{\mathcal{C}}(0)$, $\boldsymbol{\mathcal{W}}_{n_\mathrm{F}}$ following TCGD controls the variations of the hidden value tensor across frames, given by

$$\mathsf{P}(\boldsymbol{\mathcal{W}}) \propto \mathsf{CN}(\boldsymbol{\mathcal{W}}; \boldsymbol{\mathcal{C}}(0), \boldsymbol{\mathcal{V}}) \prod\nolimits_{\mathbf{r} \in \mathcal{N}} \mathsf{CN}(\boldsymbol{\mathcal{W}}; \boldsymbol{\mathcal{C}}(0), \gamma^{-1} \boldsymbol{\mathcal{V}}^{[\mathbf{r}]}). \tag{16}$$

It implies that the power of each element in $\boldsymbol{\mathcal{W}}_{n_\mathrm{F}}$ is controlled not only by its own hyperparameter but also by the hyperparameters within $\mathcal{N}$.

**Remark 2.** The temporal variations of hyperparameters are assumed to be negligible over the duration of interest, enabling joint processing of multiple frames to enhance hyperparameter learning. This assumption is typically satisfied by selecting the proper number of frames for hyperparameter learning.

### *B. Dual-Layer Online VFE Minimization Formulation*

Building on the above probabilistic models, the hierarchical Bayesian model is depicted in Fig. 4, where only the interaction between two frames is shown. Therefore, the acquisition of ADD domain channels based on the minimum mean square error (MMSE) criterion is expressed as

$$\hat{\boldsymbol{\mathcal{G}}}_{(N_\mathrm{F})} = \int \boldsymbol{\mathcal{G}}_{(N_\mathrm{F})} \mathsf{P}(\boldsymbol{\mathcal{G}}_{(N_\mathrm{F})} \,|\, \boldsymbol{\mathcal{Y}}_{(N_\mathrm{F})}; \mathcal{P}_{\mathrm{OG},(N_\mathrm{F})}, \mathcal{P}_{\mathrm{HP}}), \tag{17}$$

where $\mathsf{P}(\boldsymbol{\mathcal{G}}_{(N_\mathrm{F})} \mid \boldsymbol{\mathcal{Y}}_{(N_\mathrm{F})}; \mathcal{P}_{\mathrm{OG},(N_\mathrm{F})}, \mathcal{P}_{\mathrm{HP}})$ denotes the posterior PDF, $\mathcal{P}_{\mathrm{HP}} \triangleq \{\boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{V}}\}$ denotes the set of hyperparameters.

Since the marginal posterior PDFs involve multiple integrals of the joint posterior PDF, the exact MMSE estimator suffers from prohibitive computational complexity in massive MIMO systems. To address this issue, we adopt variational inference techniques to approximate the complicated PDF by

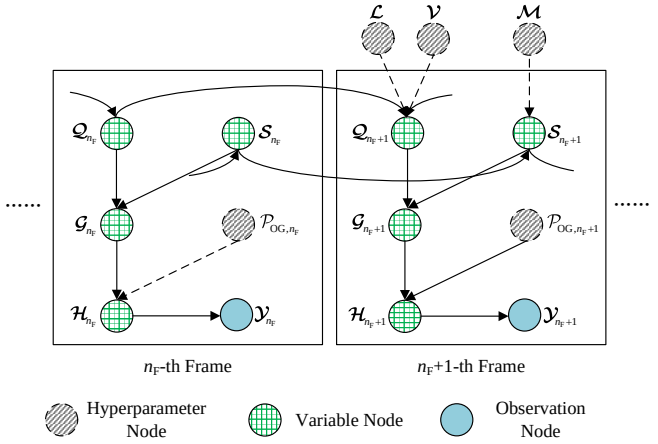



Fig. 4. Hierarchical Bayesian model for TS-DCP problem.

trial beliefs with simplified structures [50]. Based on the KLD minimization criterion, such approximation is formulated as

$$\hat{\mathsf{b}} = \arg\min_{\mathsf{b}} F_{\mathrm{V}}, \tag{18}$$

where $\mathsf{b} \triangleq \mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{G}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{S}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{Q}}_{(N_{\mathrm{F}})}; \mathcal{P}_{\mathrm{OG},(N_{\mathrm{F}})}, \mathcal{P}_{\mathrm{HP}})$ denotes the trial belief, $F_{\mathrm{V}}$ denotes the joint VFE and is defined as the KLD between the trial belief and joint PDF.

Because the variables and hyperparameters are shared across frames, a straightforward VFE minimization involves storing all received signals and jointly optimizing the trial beliefs for all frame. However, this prevents the real-time inference and is meaningless for the channel prediction task. To circumvent this issue, we simplify the trial belief to the factorized form inspired by variational messaging algorithms [51], given as

$$\begin{aligned}&\mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{G}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{S}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{Q}}_{(N_{\mathrm{F}})}; \mathcal{P}_{\mathrm{OG},(N_{\mathrm{F}})}, \mathcal{P}_{\mathrm{HP}}) \approx \\ &\quad \mathsf{b}(\mathcal{P}_{\mathrm{HP}})\mathsf{b}(\mathcal{P}_{\mathrm{OG},(N_{\mathrm{F}})})\mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{G}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{S}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{Q}}_{(N_{\mathrm{F}})}),\end{aligned} \tag{19}$$

where the hyperparameter and perturbation parameter beliefs are constrained to uninformative Dirac-Delta functions parameterized by unknown ground-truth, given as

$$\mathsf{b}(\mathcal{P}_{\mathrm{HP}}) = \mathsf{b}(\boldsymbol{\mathcal{M}})\mathsf{b}(\boldsymbol{\mathcal{L}})\mathsf{b}(\boldsymbol{\mathcal{V}}), \tag{20a}$$

$$\mathsf{b}(\boldsymbol{\mathcal{P}}) = \delta(\boldsymbol{\mathcal{P}} - \hat{\boldsymbol{\mathcal{P}}}), \boldsymbol{\mathcal{P}} \in \{\boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{V}}\}, \tag{20b}$$

and

$$\mathsf{b}(\mathcal{P}_{\mathrm{OG},(N_{\mathrm{F}})}) = \prod\nolimits_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \mathsf{b}(\mathcal{P}_{\mathrm{OG},n_{\mathrm{F}}}), \tag{21a}$$

$$\mathsf{b}(\mathcal{P}_{\mathrm{OG},n_{\mathrm{F}}}) = \mathsf{b}(\boldsymbol{\Delta\theta}_{n_{\mathrm{F}}})\mathsf{b}(\boldsymbol{\Delta\phi}_{n_{\mathrm{F}}})\mathsf{b}(\boldsymbol{\Delta\tau}_{n_{\mathrm{F}}})\mathsf{b}(\boldsymbol{\Delta\nu}_{n_{\mathrm{F}}}), \tag{21b}$$

$$\mathsf{b}(\boldsymbol{\chi}) = \delta(\boldsymbol{\chi} - \hat{\boldsymbol{\chi}}), \boldsymbol{\chi} \in \{\boldsymbol{\Delta\theta}_{n_{\mathrm{F}}}, \boldsymbol{\Delta\phi}_{n_{\mathrm{F}}}, \boldsymbol{\Delta\tau}_{n_{\mathrm{F}}}, \boldsymbol{\Delta\nu}_{n_{\mathrm{F}}}\}, \tag{21c}$$

where $\hat{\boldsymbol{\mathcal{P}}}$ and $\hat{\boldsymbol{\chi}}$ denote the ground-truth hyperparameters and perturbation parameters, respectively. This simplification is justified by the observation that the model hyperparameters $\mathcal{P}_{\mathrm{HP}}$, which capture long-timescale and cross-domain correlations in ADD domain channels, vary much more slowly than the instantaneous channels. Therefore, the model hyperparameters are assumed to remain stable within each frame, as is the case for perturbation parameters. For the shared variables across successive frames, we approximate them by the posterior mean. This further decouples the trial belief into the factorized form of per-frame trial belief, given by

$$\begin{aligned}&\mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{G}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{S}}_{(N_{\mathrm{F}})}, \boldsymbol{\mathcal{Q}}_{(N_{\mathrm{F}})}) \approx \\ &\qquad \prod\nolimits_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \mathsf{b}(\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{G}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{S}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{Q}}_{n_{\mathrm{F}}}),\end{aligned} \tag{22}$$

where $\mathsf{b}(\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{G}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{S}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{Q}}_{n_{\mathrm{F}}})$ denotes the trial belief of the $n_{\mathrm{F}}$-th frame.

Building on the structure design of trial beliefs, the joint VFE minimization is reformulated as the dual-layer optimization. The outer-layer focuses on perturbation parameter and hyperparameter learning, while the inner-layer aims to minimize the VFE with learned parameters from the outer-layer. Specifically, the inner-layer operates in the online manner, restricting the optimization of trial beliefs to a single frame, as described in the following proposition.

**Proposition 1.** Given belief structures in (19), (20), (21), annd (22), the VFE optimization in the inner-layer is given by

$$\hat{\mathsf{b}}_{n_{\mathrm{F}}} = \arg\min_{\mathsf{b}_{n_{\mathrm{F}}}} \bar{F}_{\mathrm{V},n_{\mathrm{F}}}, \tag{23}$$

where $\mathsf{b}_{n_{\mathrm{F}}} \triangleq \mathsf{b}(\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{G}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{S}}_{n_{\mathrm{F}}}, \boldsymbol{\mathcal{Q}}_{n_{\mathrm{F}}})$ denotes the per-frame trial belief, $\bar{F}_{\mathrm{V},n_{\mathrm{F}}}$ denotes the per-frame VFE given by (24) at the top of the next page, $\hat{\mathcal{P}}_{\mathrm{HP}} \triangleq \{\hat{\boldsymbol{\mathcal{M}}}, \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}}\}$ and $\hat{\mathcal{P}}_{\mathrm{OG},n_{\mathrm{F}}} \triangleq \{\hat{\boldsymbol{\Delta\theta}}_{n_{\mathrm{F}}}, \hat{\boldsymbol{\Delta\phi}}_{n_{\mathrm{F}}}, \hat{\boldsymbol{\Delta\tau}}_{n_{\mathrm{F}}}, \hat{\boldsymbol{\Delta\nu}}_{n_{\mathrm{F}}}\}$ denote the set of hyperparameters and perturbation parameters learned from the outer-layer, respectively.

*Proof.* Please refer to Appendix A. □

## IV. Online Tensor-Structured Dynamic Channel Prediciton Algorithm

Building upon the dual-layer VFE optimization process, the inference of the inner- and outer-layers is discussed in Section IV-A and Section IV-B, respectively, and ultimately yields the online TS-DCP algorithm.

### A. Multi-Linear Inference with Structured Priors

The VFE minimization in (23) is formulated as the alternating minimization between two modules, as shown in Fig. 5. The first module, the multi-linear observation (MO) module, focuses on observation and multi-linear transformation models, while the second, the structured prior (SP) module, incorporates ADD domain channel models. Therefore, the trial belief of the $n_{\mathrm{F}}$-th frame is constrained by

$$\mathsf{b}_{n_{\mathrm{F}}} = \mathsf{b}_{\mathrm{MO},n_{\mathrm{F}}}\mathsf{b}_{\mathrm{HS},n_{\mathrm{F}}}\mathsf{b}_{\mathrm{HV},n_{\mathrm{F}}}, \tag{25}$$

where $\mathsf{b}_{\mathrm{MO},n_{\mathrm{F}}}$ is optimized in the MO module, while $\mathsf{b}_{\mathrm{HS},n_{\mathrm{F}}}$ and $\mathsf{b}_{\mathrm{HV},n_{\mathrm{F}}}$ are optimized in the SP module.

*1) MO Module:* Since the MO model is a multi-linear generalization of the linear observation model, we employ the Bethe method [52], [53] based on joint PDF factorization and region partitioning and constrain the trial belief as

$$\mathsf{b}_{\mathrm{MO},n_{\mathrm{F}}} = (\mathsf{b}_{Y,n_{\mathrm{F}}}\mathsf{b}_{HG,n_{\mathrm{F}}}\mathsf{b}_{G,n_{\mathrm{F}}})(\mathsf{f}_{H,n_{\mathrm{F}}}\mathsf{f}_{G,n_{\mathrm{F}}}^{N})^{-1} \tag{26}$$

where $\mathsf{b}_{Y,n_{\mathrm{F}}}$, $\mathsf{b}_{HG,n_{\mathrm{F}}}$, and $\mathsf{b}_{G,n_{\mathrm{F}}}$ denote the factor beliefs for the observation, multi-linear transformation, and prior models,

$$\bar{F}_{\mathrm{V},n_\mathrm{F}} \triangleq \mathsf{D}[\mathsf{b}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{G}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}) \parallel \mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{G}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{Y}}_{n_\mathrm{F}} \mid \boldsymbol{\mathcal{S}}_{n_\mathrm{F}-1}, \boldsymbol{\mathcal{Q}}_{n_\mathrm{F}-1}; \hat{\mathcal{P}}_{\mathrm{OG},n_\mathrm{F}}, \hat{\mathcal{P}}_{\mathrm{HP}})], \tag{24}$$

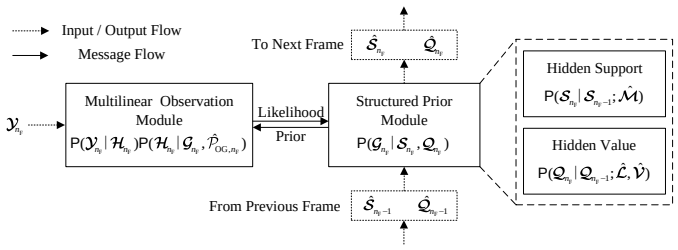



Fig. 5. Module diagram of the multi-linear inference with structured priors.

respectively, $\mathsf{f}_{H,n_\mathrm{F}}$ and $\mathsf{f}_{G,n_\mathrm{F}}$ denote the variable beliefs for $\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}$ and $\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}$, respectively, and $N \triangleq N_\mathrm{h}N_\mathrm{v}N_\mathrm{sc}N_\mathrm{sym}$ denotes the repetitive counts to be excluded in the Bethe method.

Therefore, the VFE in the MO module is converted to the Bethe free energy (BFE), expressed as

$$\begin{aligned} F_{\mathrm{MO},n_\mathrm{F}} =& \mathsf{D}[\mathsf{b}_{Y,n_\mathrm{F}} \parallel \mathsf{P}_{Y,n_\mathrm{F}}] + \mathsf{D}[\mathsf{b}_{HG,n_\mathrm{F}} \parallel \mathsf{P}_{HG,n_\mathrm{F}}] + \\ & \mathsf{D}[\mathsf{b}_{G,n_\mathrm{F}} \parallel \mathsf{P}^{\mathrm{pri}}_{G,n_\mathrm{F}}] + \mathsf{H}[\mathsf{f}_{H,n_\mathrm{F}}] + N\mathsf{H}[\mathsf{f}_{G,n_\mathrm{F}}], \end{aligned} \tag{27}$$

where $\mathsf{P}^{\mathrm{pri}}_{G,n_\mathrm{F}}$ denotes the prior PDF approximation of ADD domain channels from the SP module. To guarantee global dependencies between beliefs, we impose marginal consistency constraints (MCCs) for factor and variable beliefs sharing the same variables. However, sinde the MCCs for continuous random variables remain intractable, we relax them to first-order (FO) and second-order (SO) sufficient statistical consistency constraints (SSCCs), denoted as

$$\mathsf{E}_{\mathsf{b}_{Y,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{b}_{HG,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{f}_{H,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\}, \tag{28a}$$

$$\mathsf{V}_{\mathsf{b}_{Y,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\} = \mathsf{V}_{\mathsf{b}_{HG,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\} = \mathsf{V}_{\mathsf{f}_{H,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\}, \tag{28b}$$

$$\mathsf{E}_{\mathsf{b}_{HG,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{b}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{f}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\}, \tag{28c}$$

$$\mathsf{V}_{\mathsf{b}_{HG,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\} = \mathsf{V}_{\mathsf{b}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\} = \mathsf{V}_{\mathsf{f}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\}. \tag{28d}$$

Building upon the above BFE construction and MCC relaxation, the inference in the MO module is formulated as the constrained BFE minimization problem, given by

$$\arg\min_{\mathsf{b}_{\mathrm{MO},n_\mathrm{F}}} F_{\mathrm{MO},n_\mathrm{F}}, \quad \text{s.t.(28)}, \tag{29}$$

whose fixed-point equations are summarized in Proposition 2.

**Proposition 2.** Based on the Lagrange multiplier method [54], the fixed-point equations of (29) are given by

$$\mathsf{b}_{Y,n_\mathrm{F}} \propto \mathsf{P}_{Y,n_\mathrm{F}}\mathsf{CN}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{H,b_H}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}})^{\odot -1}), \tag{30a}$$

$$\begin{aligned} \mathsf{b}_{HG,n_\mathrm{F}} \propto & \mathsf{P}_{HG,n_\mathrm{F}}\mathsf{CN}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{H,b_{HG}}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{H,b_{HG}}_{n_\mathrm{F}})^{\odot -1}) \\ & \mathsf{CN}(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{G,b_{HG}}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{G,b_{HG}}_{n_\mathrm{F}})^{\odot -1}), \end{aligned} \tag{30b}$$

$$\mathsf{b}_{G,n_\mathrm{F}} \propto \mathsf{P}^{\mathrm{pri}}_{G,n_\mathrm{F}}\mathsf{CN}(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{G,b_G}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{G,b_G}_{n_\mathrm{F}})^{\odot -1}), \tag{30c}$$

$$\begin{aligned} \mathsf{f}_{H,n_\mathrm{F}} \propto & \mathsf{CN}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{H,b_H}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}})^{\odot -1}) \\ & \mathsf{CN}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{H,b_{HG}}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{H,b_{HG}}_{n_\mathrm{F}})^{\odot -1}), \end{aligned} \tag{30d}$$

$$\begin{aligned} \mathsf{f}_{G,n_\mathrm{F}} \propto & \mathsf{CN}(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{G,b_{HG}}_{n_\mathrm{F}}, -N(\boldsymbol{\mathcal{E}}^{G,b_{HG}}_{n_\mathrm{F}})^{\odot -1}) \\ & \mathsf{CN}(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{G,b_G}_{n_\mathrm{F}}, -N(\boldsymbol{\mathcal{E}}^{G,b_G}_{n_\mathrm{F}})^{\odot -1}), \end{aligned} \tag{30e}$$

**Algorithm 1** Single Iteration of MO Module

**Input:**

$\boldsymbol{\mathcal{Y}}_{n_\mathrm{F}}$, $\sigma_z^2$: Observation and noise variance.

$\mathsf{P}^{\mathrm{pri}}_{G,n_\mathrm{F}}$: Message of $\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}$ from SP module.

$\hat{\mathcal{P}}_{\mathrm{OG},n_\mathrm{F}}$: Estimation result of $\mathcal{P}_{\mathrm{OG},n_\mathrm{F}}$.

**Output:**

$\mathsf{P}^{\mathrm{lik}}_{G,n_\mathrm{F}}$: Message of $\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}$ to SP module.

$\boldsymbol{\mathcal{E}}^{G,b_{HG}}_{n_\mathrm{F}} = -(\mathsf{V}_{\mathsf{b}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\})^{\odot -2} - \boldsymbol{\mathcal{E}}^{G,b_G}_{n_\mathrm{F}}/N$

$\boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}} = ((\boldsymbol{\mathcal{E}}^{G,b_{HG}}_{n_\mathrm{F}})^{\odot -1} \times_1 |\mathbf{A}_\mathrm{h}(\hat{\boldsymbol{\theta}}_{n_\mathrm{F}})|^{\odot 2}$
$\times_2 |\mathbf{A}_\mathrm{v}(\hat{\boldsymbol{\phi}}_{n_\mathrm{F}})|^{\odot 2} \times_3 |\mathbf{B}(\hat{\boldsymbol{\tau}}_{n_\mathrm{F}})|^{\odot 2} \times_4 |\mathbf{C}(\hat{\boldsymbol{\nu}}_{n_\mathrm{F}})|^{\odot 2})^{\odot -1}$

$\boldsymbol{\mathcal{N}}^{H,b_H}_{n_\mathrm{F}} = -\boldsymbol{\mathcal{U}}^{H,b_H}_{n_\mathrm{F}} \oslash \boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}} + \hat{\boldsymbol{\mathcal{G}}}_{n_\mathrm{F}} \times_1 \mathbf{A}_\mathrm{h}(\hat{\boldsymbol{\theta}}_{n_\mathrm{F}})$
$\times_2 \mathbf{A}_\mathrm{v}(\hat{\boldsymbol{\phi}}_{n_\mathrm{F}}) \times_3 \mathbf{B}(\hat{\boldsymbol{\tau}}_{n_\mathrm{F}}) \times_4 \mathbf{C}(\hat{\boldsymbol{\nu}}_{n_\mathrm{F}})$

$\mathsf{P}^{\mathrm{pri}}_{H,n_\mathrm{F}} = \mathsf{CN}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{H,b_H}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}})^{\odot -1})$

$\mathsf{b}_{H,n_\mathrm{F}} \propto \mathsf{P}_{Y,n_\mathrm{F}}\mathsf{P}^{\mathrm{pri}}_{H,n_\mathrm{F}}$, $\hat{\boldsymbol{\mathcal{H}}}_{n_\mathrm{F}} = \mathsf{E}_{\mathsf{b}_{H,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\}$

$\boldsymbol{\mathcal{E}}^{H,b_{HG}}_{n_\mathrm{F}} = -(\mathsf{V}_{\mathsf{b}_{H,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\})^{\odot -2} - \boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}}$

$\boldsymbol{\Delta\mathcal{N}}_{n_\mathrm{F}} = \boldsymbol{\mathcal{U}}^{H,b_H}_{n_\mathrm{F}} \oslash \boldsymbol{\mathcal{E}}^{H,b_{HG}}_{n_\mathrm{F}} + \hat{\boldsymbol{\mathcal{H}}}_{n_\mathrm{F}} - \boldsymbol{\mathcal{N}}^{H,b_H}_{n_\mathrm{F}}$

$\boldsymbol{\mathcal{U}}^{H,b_H}_{n_\mathrm{F}} = \boldsymbol{\Delta\mathcal{N}}_{n_\mathrm{F}} \oslash ((\boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}})^{\odot -1} + (\boldsymbol{\mathcal{E}}^{H,b_{HG}}_{n_\mathrm{F}})^{\odot -1})$

$\boldsymbol{\mathcal{J}}^{H,b_H}_{n_\mathrm{F}} = \boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}} \odot (1 + \boldsymbol{\mathcal{E}}^{H,b_H}_{n_\mathrm{F}} \odot (\mathsf{V}_{\mathsf{b}_{H,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\})^{\odot 2})$

$\boldsymbol{\mathcal{J}}^{G,b_G}_{n_\mathrm{F}} = \boldsymbol{\mathcal{J}}^{H,b_H}_{n_\mathrm{F}} \times_1 |\mathbf{A}^H_\mathrm{h}(\hat{\boldsymbol{\theta}}_{n_\mathrm{F}})|^{\odot 2} \times_2 |\mathbf{A}^H_\mathrm{v}(\hat{\boldsymbol{\phi}}_{n_\mathrm{F}})|^{\odot 2}$
$\times_3 |\mathbf{B}^H(\hat{\boldsymbol{\tau}}_{n_\mathrm{F}})|^{\odot 2} \times_4 |\mathbf{C}^H(\hat{\boldsymbol{\nu}}_{n_\mathrm{F}})|^{\odot 2}$

$\boldsymbol{\mathcal{E}}^{G,b_G}_{n_\mathrm{F}} = ((\boldsymbol{\mathcal{J}}^{G,b_G}_{n_\mathrm{F}})^{\odot -1} - (N\boldsymbol{\mathcal{E}}^{H,b_{HG}}_{n_\mathrm{F}})^{\odot -1})^{\odot -1}$

$\boldsymbol{\mathcal{N}}^{G,b_G}_{n_\mathrm{F}} = \hat{\boldsymbol{\mathcal{G}}}_{n_\mathrm{F}} + (\boldsymbol{\mathcal{E}}^{G,b_G}_{n_\mathrm{F}})^{\odot -1} \odot (\boldsymbol{\mathcal{U}}^{R,b_R}_{n_\mathrm{F}} \times_1 \hat{\mathbf{A}}^H_\mathrm{h}(\hat{\boldsymbol{\theta}}_{n_\mathrm{F}})$
$\times_2 \hat{\mathbf{A}}^H_\mathrm{v}(\hat{\boldsymbol{\phi}}_{n_\mathrm{F}}) \times_3 \hat{\mathbf{B}}^H(\hat{\boldsymbol{\tau}}_{n_\mathrm{F}}) \times_4 \hat{\mathbf{C}}^H(\hat{\boldsymbol{\nu}}_{n_\mathrm{F}}))$

$\mathsf{P}^{\mathrm{lik}}_{G,n_\mathrm{F}} = \mathsf{CN}(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \boldsymbol{\mathcal{N}}^{G,b_G}_{n_\mathrm{F}}, -(\boldsymbol{\mathcal{E}}^{G,b_G}_{n_\mathrm{F}})^{\odot -1})$

$\mathsf{b}_{G,n_\mathrm{F}} \propto \mathsf{P}^{\mathrm{pri}}_{G,n_\mathrm{F}}\mathsf{P}^{\mathrm{lik}}_{G,n_\mathrm{F}}$, $\hat{\boldsymbol{\mathcal{G}}}_{n_\mathrm{F}} = \mathsf{E}_{\mathsf{b}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\}$

where $\boldsymbol{\mathcal{N}}^{\mathrm{x},b_\mathrm{x}}_{n_\mathrm{F}}$ is defined in (31) at the top of the next page, $\boldsymbol{\mathcal{U}}^{\mathrm{x},b_\mathrm{x}}_{n_\mathrm{F}}$ and $\boldsymbol{\mathcal{E}}^{\mathrm{x},b_\mathrm{x}}_{n_\mathrm{F}}$ are the Lagrange multipliers for SSCCs.

*Proof.* Please refer to Appendix B. □

By substituting fixed-point equations in Proposition 2 into SSCCs, the iteration equations for the Lagrange multipliers can be derived. These equations are organized based on the message flow backward and forward from observations to ADD domain channels, summarized in Algorithm 1.

*2) SP Module:* In this module, due to correlations between elements in hidden support tensors, the structured trial beliefs are still constructed by the Bethe method, given by

$$\mathsf{b}_{\mathrm{HS},n_\mathrm{F}} = (\mathsf{b}^{\mathrm{lik}}_{S,n_\mathrm{F}}\mathsf{b}^{\mathrm{MRF}}_{S,n_\mathrm{F}}\mathsf{b}^{\mathrm{TC}}_{S,n_\mathrm{F}})(\mathsf{f}^{F+1}_{S,n_\mathrm{F}})^{-1}, \tag{32}$$

where $\mathsf{b}^{\mathrm{lik}}_{S,n_\mathrm{F}}$, $\mathsf{b}^{\mathrm{MRF}}_{S,n_\mathrm{F}}$, and $\mathsf{b}^{\mathrm{TC}}_{S,n_\mathrm{F}}$ denote the factor beliefs introduced for likelihood PDF, MRF model, and long-timescale correlation model, respectively, $\mathsf{f}_{S,n_\mathrm{F}}$ denotes the variable belief introduced for $\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}$, $F$ denotes the number of high-order neighbors. Therefore, the VFE of hidden support tensors in the SP module is expressed as

$$\begin{aligned} F_{\mathrm{HS},n_\mathrm{F}} =& \mathsf{D}[\mathsf{b}^{\mathrm{lik}}_{S,n_\mathrm{F}} \parallel \mathsf{P}^{\mathrm{lik}}_{S,n_\mathrm{F}}] + \mathsf{D}[\mathsf{b}^{\mathrm{MRF}}_{S,n_\mathrm{F}} \parallel \mathsf{P}^{\mathrm{MRF}}_{S,n_\mathrm{F}}] + \\ & \mathsf{D}[\mathsf{b}^{\mathrm{TC}}_{S,n_\mathrm{F}} \parallel \mathsf{P}^{\mathrm{TC,OL}}_{S,n_\mathrm{F}}] + (F+1)\mathsf{H}[\mathsf{f}_{S,n_\mathrm{F}}], \end{aligned} \tag{33}$$

$$\boldsymbol{\mathcal{N}}_{n_\mathrm{F}}^{H,b_H} = -\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{H,b_H} \oslash \boldsymbol{\mathcal{E}}_{n_\mathrm{F}}^{H,b_H} + \mathsf{E}_{\mathsf{b}_{Y,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\}, \boldsymbol{\mathcal{N}}_{n_\mathrm{F}}^{H,b_{HG}} = -\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{H,b_{HG}} \oslash \boldsymbol{\mathcal{E}}_{n_\mathrm{F}}^{H,b_{HG}} + \mathsf{E}_{\mathsf{b}_{Y,n_\mathrm{F}}}\{\boldsymbol{\mathcal{H}}_{n_\mathrm{F}}\}, \tag{31a}$$

$$\boldsymbol{\mathcal{N}}_{n_\mathrm{F}}^{G,b_{HG}} = -\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{G,b_{HG}} \oslash \boldsymbol{\mathcal{E}}_{n_\mathrm{F}}^{G,b_{HG}} + \mathsf{E}_{\mathsf{b}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\}, \boldsymbol{\mathcal{N}}_{n_\mathrm{F}}^{G,b_G} = -\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{G,b_G} \oslash \boldsymbol{\mathcal{E}}_{n_\mathrm{F}}^{G,b_G} + \mathsf{E}_{\mathsf{b}_{G,n_\mathrm{F}}}\{\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}\}. \tag{31b}$$

**Algorithm 2** Single Iteration of SP Module

**Input:**

$\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{lik}}$, $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}}$: Message of $\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}$ and $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$ from MO module.

$\hat{\mathcal{P}}_{\mathrm{HP}}$: Estimation result of $\mathcal{P}_{\mathrm{HP}}$.

**Output:**

$\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{pri}}$, $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{pri}}$: Message of $\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}$ and $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$ to MO module.

$\bar{\boldsymbol{\mathcal{U}}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF}} = \sum_{\bar{\mathbf{r}}\in\mathcal{N}\setminus\mathbf{r}} \boldsymbol{\mathcal{U}}_{n_\mathrm{F},\bar{\mathbf{r}}}^{\mathrm{MRF}}, \forall \mathbf{r} \in \mathcal{N}$

$\boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{S,\mathrm{MRF}} = \boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{lik}} + \boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{TC}} + \bar{\boldsymbol{\mathcal{U}}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF}}, \forall \mathbf{r} \in \mathcal{N}$

$\boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF},1} = 1 + \exp(2\mathcal{C}(\gamma) + \boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{S,\mathrm{MRF}}), \forall \mathbf{r} \in \mathcal{N}$

$\boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF},2} = \exp(2\mathcal{C}(\gamma)) + \exp(\boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{S,\mathrm{MRF}}), \forall \mathbf{r} \in \mathcal{N}$

$\boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF}} = \boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF},1} - \boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF},2}, \forall \mathbf{r} \in \mathcal{N}$

$\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{MRF}} = \sum_{\mathbf{r}\in\mathcal{N}} \boldsymbol{\mathcal{U}}_{n_\mathrm{F},\mathbf{r}}^{\mathrm{MRF}}, \forall \mathbf{r} \in \mathcal{N}$

$\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{pri}} \propto \exp(-\langle \boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{TC}} + \boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{MRF}}, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}} \rangle)$

$\mathsf{b}_{S,n_\mathrm{F}} \propto \mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{lik}} \mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{pri}}, \mathsf{b}_{Q,n_\mathrm{F}} \propto \mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}} \mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{pri}}$

$\hat{\boldsymbol{\mathcal{S}}}_{n_\mathrm{F}} = \mathsf{E}_{\mathsf{b}_{S,n_\mathrm{F}}}\{\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}\}$, $\hat{\boldsymbol{\mathcal{Q}}}_{n_\mathrm{F}} = \mathsf{E}_{\mathsf{b}_{Q,n_\mathrm{F}}}\{\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}\}$

where $\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{lik}}$ denotes the likelihood PDF of hidden support tensors, and $\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{TC,OL}}$ denotes the temporal correlation model with an online approximation, defined by $\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{TC,OL}} \triangleq \mathsf{P}_{\mathrm{TC}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}} | \hat{\boldsymbol{\mathcal{S}}}_{n_\mathrm{F}-1}; \hat{\boldsymbol{\mathcal{M}}})$. Furthermore, the MCCs of binary-valued hidden support tensor are equivalent to FO-SSCCs, given by

$$\mathsf{E}_{\mathsf{b}_{S,n_\mathrm{F}}^{\mathrm{MRF}}}\{\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{b}_{S,n_\mathrm{F}}^{\mathrm{TC}}}\{\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{b}_{S,n_\mathrm{F}}^{\mathrm{lik}}}\{\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}\} = \mathsf{E}_{\mathsf{f}_{S,n_\mathrm{F}}}\{\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}\}, \tag{34}$$

where $\mathsf{b}_{S,n_\mathrm{F}}^{\mathrm{MRF}}$ denotes the factor belief of MRFs, given by

$$\mathsf{b}_{S,n_\mathrm{F}}^{\mathrm{MRF}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}) = \prod\nolimits_{\mathbf{r}\in\mathcal{N}} \mathsf{b}_{S,n_\mathrm{F},\mathbf{r}}(\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}}^{[\mathbf{r}]}). \tag{35}$$

Therefore, the inference of hidden support tensors in the SP module is formulated as

$$\arg\min_{\mathsf{b}_{\mathrm{HS},n_\mathrm{F}}} F_{\mathrm{HS},n_\mathrm{F}}, \quad \text{s.t.(34)}, \tag{36}$$

and can be organized through belief propagation rules.

Due to the conditional independency of the elements in $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$ given $\hat{\boldsymbol{\mathcal{Q}}}_{n_\mathrm{F}-1}$, no additional constraints are required on the trial beliefs of $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$, thus simplifying the VFE minimization. By directly minimizing the KLD, the fixed-point equations of hidden value tensors in the SP module are expressed as

$$\mathsf{b}_{\mathrm{HV},n_\mathrm{F}} \propto \mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}} \mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{TC,OL}}, \tag{37}$$

respectively, where $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}}$ and $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{TC,OL}}$ denote the likelihood PDF and the temporal correlation model with online approximation of hidden value tensors, respectively, and we define $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{TC,OL}} \triangleq \mathsf{P}(\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}} \mid \hat{\boldsymbol{\mathcal{Q}}}_{n_\mathrm{F}-1}; \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}})$.

Similar to the MO module, the iterative equations of Lagrange multipliers in the SP module are summarized in Algorithm 2. In this context, $\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{lik}}$ and $\boldsymbol{\mathcal{U}}_{n_\mathrm{F}}^{\mathrm{TC}}$ denote log-likelihood ratios of $\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{lik}}$ and $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}}$, respectively.

*3) Closing the Loops:* Based on the iterative equations within the MO and SP modules, a critical step in closing the loop of online TS-DCP algorithm is the message exchange between these modules, including $\mathsf{P}_{G,n_\mathrm{F}}^{\mathrm{pri}}$, $\mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{lik}}$, and $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}}$. Following the joint PDF factorization, the joint posterior PDF of $\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}$, $\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}$, and $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$ is

$$\mathsf{P}(\boldsymbol{\mathcal{G}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{S}}_{n_\mathrm{F}}, \boldsymbol{\mathcal{Q}}_{n_\mathrm{F}} \mid \boldsymbol{\mathcal{Y}}_{n_\mathrm{F}}) \propto \mathsf{P}_{GSQ,n_\mathrm{F}} \mathsf{P}_{S,n_\mathrm{F}}^{\mathrm{pri}} \mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{pri}} \mathsf{P}_{G,n_\mathrm{F}}^{\mathrm{lik}}. \tag{38}$$

Therefore, the prior and likelihood PDFs can be obtained by marginalizing the joint posterior PDF and excluding contributions of other variables. However, zero elements in $\boldsymbol{\mathcal{S}}_{n_\mathrm{F}}$ lead to the unobservability of the corresponding elements in $\boldsymbol{\mathcal{Q}}_{n_\mathrm{F}}$, thus yielding uninformative likelihood PDFs and preventing the VFE minimization. To mitigate this issue, we adopt the threshold-based Gaussian sum approximation [55] in $\mathsf{P}_{Q,n_\mathrm{F}}^{\mathrm{lik}}$.

### *B. Perturbation Parameter and Hyperparameter Learning*

Beyond the MO and SP modules, the perturbation parameter and hyperparameter learning refinies the grids and hyperparameters for MO and SP modules, respectively.

*1) Perturbation Parameter:* To reduce computational complexity, we employ the sequential optimization approach for the perturbation parameter. In the example of the horizontal angle domain, the perturbation parameter learning rule based on the VFE minimization is formulated as

$$\begin{aligned}\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}}^{\star} &\overset{(a)}{=} \arg\min_{\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}}} \mathsf{D}[\mathsf{b}_{HG,n_\mathrm{F}} \parallel \mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}} | \boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \hat{\mathcal{P}}_{\mathrm{OG},n_\mathrm{F}})] \\ &= \arg\max_{\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}}} \underbrace{\int \mathsf{b}_{HG,n_\mathrm{F}} \ln \mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\mathrm{F}} | \boldsymbol{\mathcal{G}}_{n_\mathrm{F}}; \hat{\mathcal{P}}_{\setminus\mathrm{h},n_\mathrm{F}})}_{f(\hat{\boldsymbol{\Delta}}\theta_{n_\mathrm{F}})},\end{aligned} \tag{39}$$

where $(a)$ is due to only this KLD term depending on $\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}}$, $\hat{\mathcal{P}}_{\setminus\mathrm{h},n_\mathrm{F}} \triangleq \{\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}}, \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}_{n_\mathrm{F}}^{\star}, \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}_{n_\mathrm{F}}^{\star}, \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}_{n_\mathrm{F}}^{\star}\}$ includes the perturbation parameters currently being optimized and previously optimized perturbation parameters of other domains. Through the first-order Taylor approximation of factor matrices, the objective function is approximated as in Proposition 3.

**Proposition 3.** The objective function in (39) is approximated as the quadratic function with respect to $\hat{\boldsymbol{\Delta}}\theta_{n_\mathrm{F}}$, given by

$$f(\hat{\boldsymbol{\Delta}}\theta_{n_\mathrm{F}}) = \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}}^T \boldsymbol{\Pi}_{\mathrm{h}} \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}} - 2\boldsymbol{\mu}_{\mathrm{h}}^T \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\mathrm{F}} + C, \tag{40}$$

where $C$ denotes the constant term independent of $\hat{\boldsymbol{\Delta}}\theta_{n_\mathrm{F}}$, $\boldsymbol{\Pi}_{\mathrm{h}}$ and $\boldsymbol{\mu}_{\mathrm{h}}$ are given by

$$\begin{aligned}\boldsymbol{\Pi}_{\mathrm{h}} =& (\dot{\mathbf{A}}_{\mathrm{h}}^H(\bar{\boldsymbol{\theta}}) \dot{\mathbf{A}}_{\mathrm{h}}(\bar{\boldsymbol{\theta}}))^* \odot \\ & \sum\nolimits_n (\hat{\mathbf{g}}_{n_\mathrm{F},n}^{\mathrm{h}} (\hat{\mathbf{g}}_{n_\mathrm{F},n}^{\mathrm{h}})^H + \mathsf{diag}\{\boldsymbol{\varepsilon}_{G,n_\mathrm{F},n}^{\mathrm{h}}\}),\end{aligned} \tag{41a}$$

$$\begin{aligned}\boldsymbol{\mu}_{\mathrm{h}} =& \sum\nolimits_n \mathsf{Re}\{\mathsf{diag}^H\{\hat{\mathbf{g}}_{n_\mathrm{F},n}^{\mathrm{h}}\} \dot{\mathbf{A}}_{\mathrm{h}}^H(\bar{\boldsymbol{\theta}}) \Delta\hat{\mathbf{h}}_{n_\mathrm{F},n}^{(1)}\} - \\ & \sum\nolimits_n \mathsf{Re}\{\mathsf{diag}\{\dot{\mathbf{A}}_{\mathrm{h}}^H(\bar{\boldsymbol{\theta}}) \mathbf{A}_{\mathrm{h}}(\bar{\boldsymbol{\theta}})\} \odot \boldsymbol{\varepsilon}_{G,n_\mathrm{F},n}^{\mathrm{h}}\},\end{aligned} \tag{41b}$$

where $\dot{\mathbf{A}}_{\mathrm{h}}(\boldsymbol{\theta})$ denotes the first-order derivative of $\mathbf{A}_{\mathrm{h}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, $\mathbf{g}^{\mathrm{h}}_{n_{\mathrm{F}},n}$, $\boldsymbol{\varepsilon}^{\mathrm{h}}_{G,n_{\mathrm{F}},n}$, and $\Delta\hat{\mathbf{h}}^{(1)}_{n_{\mathrm{F}},n}$ denote the first-mode fibers of $\hat{\boldsymbol{\mathcal{G}}}^{\mathrm{h}}_{n_{\mathrm{F}}}$, $\boldsymbol{\mathcal{E}}^{\mathrm{h}}_{G,n_{\mathrm{F}}}$, and $\Delta\hat{\boldsymbol{\mathcal{H}}}_{n_{\mathrm{F}}}$, respectively, $\hat{\boldsymbol{\mathcal{G}}}^{\mathrm{h}}_{n_{\mathrm{F}}}$, $\boldsymbol{\mathcal{E}}^{\mathrm{h}}_{G,n_{\mathrm{F}}}$, and $\Delta\hat{\boldsymbol{\mathcal{H}}}_{n_{\mathrm{F}}}$ are defined by (42) at the top of the next page.

*Proof.* Please refer to Appendix C. □

Therefore, the perturbation parameter learning is formulated as the quadratic optimization problem and solved by well-known approaches with low-complexity, as are the learning rules in other domains. By leveraging efficient tensor operations, the auxiliary matrices and vectors can also be constructed in the low-complexity manner.

*2) Hyperparameter Learning:* Different from the perturbation parameters, the hyperparameters in $\hat{\mathcal{P}}_{\mathrm{HP}}$ are shared across all frames and learned adaptively based on the current observations. For the $N_{\mathrm{F}}$-th frame, the hyperparameter learning rules based on the VFE minimization are expressed as

$$\hat{\boldsymbol{\mathcal{M}}}^{\star} = \arg\max_{\hat{\boldsymbol{\mathcal{M}}}} \sum_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \int \mathsf{b}_{S,n_{\mathrm{F}}} \ln \mathsf{P}_{\mathrm{TC}}(\boldsymbol{\mathcal{S}}_{n_{\mathrm{F}}} \,|\, \boldsymbol{\mathcal{S}}_{n_{\mathrm{F}}-1}; \hat{\boldsymbol{\mathcal{M}}}), \quad (43a)$$

$$\{\hat{\boldsymbol{\mathcal{L}}}^{\star}, \hat{\boldsymbol{\mathcal{V}}}^{\star}\} = \arg\max_{\{\hat{\boldsymbol{\mathcal{L}}},\hat{\boldsymbol{\mathcal{V}}}\}} \sum_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \int \mathsf{b}_{Q,n_{\mathrm{F}}} \ln \mathsf{P}(\boldsymbol{\mathcal{Q}}_{n_{\mathrm{F}}} \,|\, \boldsymbol{\mathcal{Q}}_{n_{\mathrm{F}}-1}; \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}}). \quad (43b)$$

where other terms in (23) are independent of $\hat{\boldsymbol{\mathcal{M}}}$, $\hat{\boldsymbol{\mathcal{L}}}$, and $\hat{\boldsymbol{\mathcal{V}}}$. Due to the hyperparameter coupling in $\boldsymbol{\mathcal{Q}}_{n_{\mathrm{F}}}$, it is intractable to find an analytic solution for $\hat{\boldsymbol{\mathcal{V}}}^{\star}$. To address this, we first learn $\bar{\boldsymbol{\mathcal{V}}}$ instead and then approximate $\hat{\boldsymbol{\mathcal{V}}}^{\star}$ as [56], [57]

$$\hat{\boldsymbol{\mathcal{V}}}^{\star} \approx \bar{\boldsymbol{\mathcal{V}}} + \gamma \sum\nolimits_{\mathbf{r}\in\mathcal{N}} \bar{\boldsymbol{\mathcal{V}}}^{[\mathbf{r}]}, \quad (44)$$

where $\bar{\boldsymbol{\mathcal{V}}}$ is given in (45) at the top of the next page. Besides, the learning rules of transition and temporal correlation factors are summarized in Proposition 4, henceforth all hyperparameters can be learned in the element-wise manner.

**Proposition 4.** The learning rules of transition and temporal correlation factors are given by

$$\hat{\boldsymbol{\mathcal{M}}}^{\star} = \mathsf{ln}(1+\boldsymbol{\mathcal{K}}_M) - \mathsf{ln}(1-\boldsymbol{\mathcal{K}}_M), \quad (46a)$$

$$(N_{\mathrm{F}}-1)\bar{\boldsymbol{\mathcal{V}}} \odot |\hat{\boldsymbol{\mathcal{L}}}^{\star}|^{\odot 2} + \boldsymbol{\mathcal{K}}_{L,1} \odot \hat{\boldsymbol{\mathcal{L}}}^{\star} + \boldsymbol{\mathcal{K}}_{L,0} = 0, \quad (46b)$$

where $\boldsymbol{\mathcal{K}}_M$, $\boldsymbol{\mathcal{K}}_{L,1}$ and $\boldsymbol{\mathcal{K}}_{L,2}$ are defined by

$$\boldsymbol{\mathcal{K}}_M = \frac{1}{N_{\mathrm{F}}} \sum\nolimits_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} (2\hat{\boldsymbol{\mathcal{S}}}_{n_{\mathrm{F}}} - 1) \odot (2\hat{\boldsymbol{\mathcal{S}}}_{n_{\mathrm{F}}-1} - 1), \quad (47a)$$

$$\boldsymbol{\mathcal{K}}_{L,1} \triangleq \sum\nolimits_{n_{\mathrm{F}}=2}^{N_{\mathrm{F}}} (|\hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}-1}|^{\odot 2} - \mathsf{Re}\{\hat{\boldsymbol{\mathcal{Q}}}^{*}_{n_{\mathrm{F}}-1} \odot \hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}}\}), \quad (47b)$$

$$\boldsymbol{\mathcal{K}}_{L,0} \triangleq -\sum\nolimits_{n_{\mathrm{F}}=2}^{N_{\mathrm{F}}} (|\hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}} - \hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}-1}|^{\odot 2} + \boldsymbol{\mathcal{E}}_{Q,n_{\mathrm{F}}}). \quad (47c)$$

*Proof.* Please refer to Appendix D. □

### C. Online TS-DCP Algorithm

Following the dual-layer optimization framework in Section III-B, the proposed online TS-DCP algorithm is summarized as Algorithm 3. Given the estimation of ADD domain channels and perturbation parameters, channel prediction is achieved by the transformation from Doppler to temporal domain, given as

$$\boldsymbol{\mathcal{H}}^{\mathrm{CP}}_{n_{\mathrm{F}}} = \hat{\boldsymbol{\mathcal{G}}}_{n_{\mathrm{F}}} \times_1 \mathbf{A}_{\mathrm{h}}(\bar{\boldsymbol{\theta}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_{\mathrm{F}}}) \times_2 \mathbf{A}_{\mathrm{v}}(\bar{\boldsymbol{\phi}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}_{n_{\mathrm{F}}}) \times_3 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}_{n_{\mathrm{F}}}) \times_4 \tilde{\mathbf{C}}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}_{n_{\mathrm{F}}}), \quad (48)$$

where $\tilde{\mathbf{C}}(\boldsymbol{\nu}) = [\tilde{\mathbf{c}}(\nu_1), \ldots, \tilde{\mathbf{c}}(\nu_{K_{\mathrm{do}}})]$ denotes the factor matrix in the temporal domain for channel prediction, $\tilde{\mathbf{c}}(\nu)$ denotes the Doppler domain steering vector for channel prediction, defined as $[\tilde{\mathbf{c}}(\nu)]_{n_{\mathrm{cp}}} = \exp(j2\pi(T_0 + n_{\mathrm{cp}}\Delta T)\nu)$, $T_0 = (N_{\mathrm{sym}} - 1)\Delta\bar{T}$ and $n_{\mathrm{cp}}$ denote the prediction origin and prediction length, respectively. Since Algorithm 3 operates in the sliding frame manner, it enables dynamic frame reorganization as new pilot OFDM symbols become available, facilitating real-time channel prediction.

**Algorithm 3** Online TS-DCP Algorithm

**Input:**
$\boldsymbol{\mathcal{Y}}_{(N_{\mathrm{F}})}$, $\sigma_z^2$: Observation and noise variance.
**Output:**
$\hat{\boldsymbol{\mathcal{H}}}_{(N_{\mathrm{F}})}$: Predicted SFT domain channels.

```
Initialize M̂, L̂, V̂, Ŝ_0, and Q̂_0.
for n_F = 1, ..., N_F do
    for t = 1 : T do
        Obtain the prior PDF of G_{n_F}.
        Execute Algorithm 1.
        Obtain the likelihood PDFs of S_{n_F} and Q_{n_F}.
        Execute Algorithm 2.
        Perturbation parameter / hyperparameter learning.
    end for
    Predict the channel based on (48).
end for
```

**Remark 3.** By designing various belief structures and MCC relaxations, we integrate the message passing rules-including variational message passing [51], belief propagation [52], expectation propagation variant [53], and expectation maximization [58] into Algorithm 3 under the VFE minimization framework. This integration enables different message-passing rule scheduling strategies from the alternating optimization perspective, further reducing computational complexity [59]. Moreover, the probabilistic models are compatible with quantized signals and hybrid analog-digital architectures, offering a practical approach to reducing memory overhead.

## V. Simulation Results

### A. Simulation Configuration

*1) Scenario Setting:* To validate the exactness of proposed channel and probabilistic models, we employ the QuaDRiGa channel simulator, which generates massive MIMO-OFDM channels consistent with the third Generation Partnership Program (3GPP) NR specifications [60] and has been validated in various field trials [61]. In the QuaDRiGa channel simulator, we consider the 3GPP urban macro (UMa) non-line-of-sight (NLOS) scenarios, where each channel contains 20 clusters consisting of 20 subpaths with similar physical parameters. As MTs traverse the trajectory, the contribution

$$\hat{\boldsymbol{\mathcal{G}}}^{\mathrm{h}}_{n_{\mathrm{F}}} = \hat{\boldsymbol{\mathcal{G}}}_{n_{\mathrm{F}}} \times_2 \mathbf{A}_{\mathrm{v}}(\bar{\boldsymbol{\phi}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}^{\star}_{n_{\mathrm{F}}}) \times_3 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}^{\star}_{n_{\mathrm{F}}}) \times_4 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}^{\star}_{n_{\mathrm{F}}}), \tag{42a}$$

$$\boldsymbol{\mathcal{E}}^{\mathrm{h}}_{G,n_{\mathrm{F}}} \triangleq \boldsymbol{\mathcal{E}}_{G,n_{\mathrm{F}}} \times_2 |\mathbf{A}_{\mathrm{v}}(\bar{\boldsymbol{\phi}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}^{\star}_{n_{\mathrm{F}}})|^{\odot 2} \times_3 |\mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}^{\star}_{n_{\mathrm{F}}})|^{\odot 2} \times_4 |\mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}^{\star}_{n_{\mathrm{F}}})|^{\odot 2}, \tag{42b}$$

$$\Delta\hat{\boldsymbol{\mathcal{H}}}_{n_{\mathrm{F}}} = \hat{\boldsymbol{\mathcal{H}}}_{n_{\mathrm{F}}} - \hat{\boldsymbol{\mathcal{G}}}_{n_{\mathrm{F}}} \times_1 \mathbf{A}_{\mathrm{h}}(\bar{\boldsymbol{\theta}}) \times_2 \mathbf{A}_{\mathrm{v}}(\bar{\boldsymbol{\phi}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}^{\star}_{n_{\mathrm{F}}}) \times_3 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}^{\star}_{n_{\mathrm{F}}}) \times_4 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}^{\star}_{n_{\mathrm{F}}}). \tag{42c}$$

$$\bar{\boldsymbol{\mathcal{V}}} = \frac{1}{N_{\mathrm{F}}}(|\hat{\boldsymbol{\mathcal{Q}}}_1|^{\odot 2} + \boldsymbol{\mathcal{E}}_{Q,1} + \sum_{n_{\mathrm{F}}=2}^{N_{\mathrm{F}}} (|\hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}}|^{\odot 2} + \boldsymbol{\mathcal{E}}_{Q,n_{\mathrm{F}}} - 2(1-\hat{\boldsymbol{\mathcal{L}}}^{\star}) \odot \mathsf{Re}\{\hat{\boldsymbol{\mathcal{Q}}}^{*}_{n_{\mathrm{F}}-1} \odot \hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}}\} + (1-\hat{\boldsymbol{\mathcal{L}}}^{\star}) \odot |\hat{\boldsymbol{\mathcal{Q}}}_{n_{\mathrm{F}}-1}|^{\odot 2}) \oslash (\hat{\boldsymbol{\mathcal{L}}}^{\star})^{\odot 2}). \tag{45}$$

TABLE II
SCENARIO PARAMETERS

| Paramter | Value |
|---|---|
| Carrier Frequency | $f_{\mathrm{c}} = 6.7$ GHz |
| OFDM Symbol Duration | $\Delta T_{\mathrm{sym}} = 33.33$ $\mu$s |
| Cyclic Prefix Duration | $\Delta T_{\mathrm{cp}} = 2.34$ $\mu$s |
| Subcarrier Spacing | $\Delta f = 30$ kHz |
| Number of Pilot Interval Symbols | $N_{\mathrm{IS}} = 14$ |
| Number of Transmission Combs | $N_{\mathrm{TC}} = 4$ |
| Number of BS Antennas | $(N_{\mathrm{h}}, N_{\mathrm{v}}) = (32, 16)$ |
| Number of Pilot Subcarriers | $N_{\mathrm{sc}} = 64$ |
| Number of Pilot Symbols | $N_{\mathrm{sym}} = 8$ |
| BS Height | $h_{\mathrm{BS}} = 25$ m |
| MT Height | $h_{\mathrm{BS}} = 1.5$ m |
| MT Distribution Radius | $r_{\mathrm{MT}} = 200$ m |

of scatterers evolves with time, and the spatial consistency of the channels is maintained. Unless specified otherwise, the simulation configurations follow the system model in Section II, with parameters summarized in Table II.

*2) Benchmarks and Performance Metric:* To demonstrate the superiority of the proposed online TS-DCP algorithm, we select the following state-of-the-art algorithms as benchmarks:

- **VKF** [26]: Estimates spatial domain channels by the least square (LS) algorithm, with temporal correlations captured by the AR-based vector KF.
- **FIT** [32]: Estimates angle-delay domain channels by the alternating LS (ALS) algorithm, with temporal correlations captured by the first-order Taylor series.
- **PAD** [13], **MPAD** [14]: Estimates spatial domain channels by linear MMSE (LMMSE) algorithm, with temporal correlations captured by Prony and matrix pencil methods for effective angle-delay domain taps.

Besides, we include **Online TS-DCP** algorithm with pre-sampled grids and unstructured independent prior as benchmarks, referred to as **Online TS-DCP (PG)** and **Online TS-DCP (UIP)**, respectively. For benchmarks that only predict channels for future pilot symbols, 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 SFT domain channels is adopted as the performance metric, defined by

$$\mathrm{TNMSE} = \frac{1}{N_{\mathrm{F}}} \sum_{n_{\mathrm{F}}=1}^{N_{\mathrm{F}}} \frac{\|\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}}_{n_{\mathrm{F}}} - \boldsymbol{\mathcal{H}}^{\mathrm{CP}}_{n_{\mathrm{F}}}\|_F^2}{\|\boldsymbol{\mathcal{H}}^{\mathrm{CP}}_{n_{\mathrm{F}}}\|_F^2}, \tag{49}$$

where $\boldsymbol{\mathcal{H}}^{\mathrm{CP}}_{n_{\mathrm{F}}}$ and $\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}}_{n_{\mathrm{F}}}$ denote the groud-truth and predicted SFT domain channels of the upcoming OFDM symbols in the $n_{\mathrm{F}}$-th frame, respectively. To ensure fairness, we evaluate all algorithms on the same QuaDRiGa-generated channel data.

### *B. Computational Complexity Analysis*

*1) Preliminaries:* Before delving into the computational complexity analysis, we first introduce some preliminaries regarding the computational cost of tensor-matrix multiplication. Specifically, for a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{N_1 \times \ldots \times N_D}$ and a matrix $\mathbf{U}_d \in \mathbb{C}^{K_d \times N_d}$, the mode-$d$ tensor-matrix multiplication is implemented based on tensor matricization, following the equivalence $\mathbf{Y}_d = \mathbf{U}_d \mathbf{X}_d$. This suggests that the computational complexity of a tensor-matrix multiplication along the $d$-th mode is equivalent to the matrix multiplication of $\mathbf{U}_d$ and $\mathbf{X}_d$, with the order of $\mathcal{O}(K_d N_1 \ldots N_D)$. It is worth noting that the computational complexity of multi-mode tensor-matrix multiplications depends on the execution order, owing to the commutativity across different modes. In general, the execution order that minimizes the size of intermediate tensors will result in lower computational complexity.

*2) Asymptotic Order:* Building upon the above preliminaries, we analyze the computational complexity of the proposed algorithms, which arises primarily from tensor-matrix multiplications and scalar arithmetic operations. The former is the main computational burden in the MO module (Algorithm 1) and perturbation parameter learning, while the latter dominates the SP module (Algorithm 2) and hyperparameter learning. To illustrate the computational complexity reduction achieved by tensor-matrix multiplication compared with matrix-vector multiplication, we consider a typical case in which the size of the ADD and SFT domain channels is of the same order—an assumption that typically holds in practice. Therefore, the computational complexity order of tensor-matrix multiplications are given by $\mathcal{O}(N_{\mathrm{h}} N_{\mathrm{v}} N_{\mathrm{sc}} N_{\mathrm{sym}} \bar{N})$, where $\bar{N} \triangleq N_{\mathrm{h}} + N_{\mathrm{v}} + N_{\mathrm{sc}} + N_{\mathrm{sym}}$. As for the scalar arithmetic operations, the computational complexity scales linearly with the size of ADD and SFT domain channels, i.e., $\mathcal{O}(N_{\mathrm{h}} N_{\mathrm{v}} N_{\mathrm{sc}} N_{\mathrm{sym}})$, and is significantly smaller than that of tensor-matrix multiplications. Therefore, the computational complexity per iteration of Algorithm 3 is on the order of $\mathcal{O}(N_{\mathrm{h}} N_{\mathrm{v}} N_{\mathrm{sc}} N_{\mathrm{sym}} \bar{N})$, and can be further reduced to $\mathcal{O}(N_{\mathrm{h}} N_{\mathrm{v}} N_{\mathrm{sc}} N_{\mathrm{sym}} \bar{N}_{\log})$ in **Online TS-DCP (PG)** due to its fast Fourier transform-based implementation, with $\bar{N}_{\log} \triangleq \log_2(N_{\mathrm{h}}) + \log_2(N_{\mathrm{v}}) + \log_2(N_{\mathrm{sc}}) + \log_2(N_{\mathrm{sym}})$. In contrast, when matrix-vector multiplications are employed

TABLE III
COMPUTATIONAL COMPLEXITY: ASYMPTOTIC ORDER AND FLOPS

| Algorithm | Asymptotic Order | FLOPs | | |
|---|---|---|---|---|
| | | $(N_{\mathrm{h}}, N_{\mathrm{v}}) = (32, 16)$ | $(N_{\mathrm{h}}, N_{\mathrm{v}}) = (32, 32)$ | $(N_{\mathrm{h}}, N_{\mathrm{v}}) = (64, 32)$ |
| This work (per iteration) | | | | |
| Online TS-DCP | $\mathcal{O}(N_{\mathrm{h}}N_{\mathrm{v}}N_{\mathrm{sc}}N_{\mathrm{sym}}\bar{N})$ | $2.15 \times 10^8$ | $4.88 \times 10^8$ | $1.20 \times 10^9$ |
| Online TS-DCP (UIP) | $\mathcal{O}(N_{\mathrm{h}}N_{\mathrm{v}}N_{\mathrm{sc}}N_{\mathrm{sym}}\bar{N})$ | $1.92 \times 10^8$ | $4.40 \times 10^8$ | $1.11 \times 10^9$ |
| Online TS-DCP (PG) | $\mathcal{O}(N_{\mathrm{h}}N_{\mathrm{v}}N_{\mathrm{sc}}N_{\mathrm{sym}}\bar{N}_{\log})$ | $3.46 \times 10^7$ | $6.98 \times 10^7$ | $1.41 \times 10^8$ |
| VKF | $\mathcal{O}(P^3N_{\mathrm{h}}^3N_{\mathrm{v}}^3)$ | $2.29 \times 10^{10}$ | $1.83 \times 10^{11}$ | $1.47 \times 10^{12}$ |
| FIT | $\mathcal{O}(TRN_{\mathrm{h}}N_{\mathrm{v}}N_{\mathrm{sc}}N_{\mathrm{sym}})$ | $1.05 \times 10^{10}$ | $2.10 \times 10^{10}$ | $4.19 \times 10^{10}$ |
| PAD, MPAD | $\mathcal{O}(N_{\mathrm{h}}^3N_{\mathrm{v}}^3 + N_{\mathrm{h}}^2N_{\mathrm{v}}^2N_{\mathrm{sc}}N_{\mathrm{sym}})$ | $2.01 \times 10^9$ | $1.29 \times 10^{10}$ | $9.02 \times 10^{10}$ |

without the separable factor matrices, the computational complexity per iteration will increase to $\mathcal{O}(N_{\mathrm{h}}^2N_{\mathrm{v}}^2N_{\mathrm{sc}}^2N_{\mathrm{sym}}^2)$.

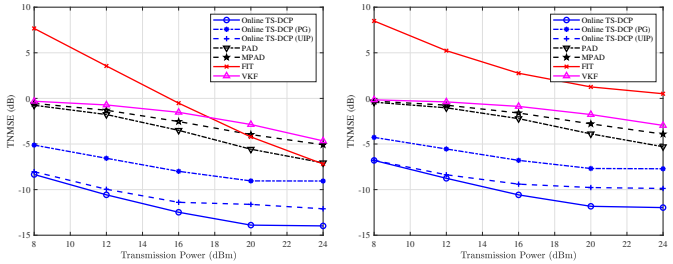
For the benchmarks, the computational complexity of **PAD** and **MPAD** is primarily dominated by the LMMSE estimation for all pilot symbols and subcarriers with shared spatial correlations, resulting in the computational complexity of $\mathcal{O}(N_{\mathrm{h}}^3N_{\mathrm{v}}^3 + N_{\mathrm{h}}^2N_{\mathrm{v}}^2N_{\mathrm{sc}}N_{\mathrm{sym}})$. In the case of **VKF**, the critical step is the AR parameter learning shared among all pilot subcarriers, with the computational complexity of $\mathcal{O}(P^3N_{\mathrm{h}}^3N_{\mathrm{v}}^3)$, where $P$ denotes the AR order. Alternatively, the spatial correlations in **PAD**, **MPAD**, and **VKF** can also be estimated for each resource element to enhance the channel prediction performance in practical environments—albeit at the cost of significantly higher computational complexity. For **FIT**, its computational complexity is governed by the ALS algorithm, i.e., $\mathcal{O}(TRN_{\mathrm{h}}N_{\mathrm{v}}N_{\mathrm{sc}}N_{\mathrm{sym}})$, where $T$ denotes the number of iterations, and $R$ denotes the predefined tensor rank, typically set as the maximum number of identifiable paths in the environment. The asymptotic orders for the proposed algorithms and benchmarks are summarized in Table III.

*3) FLOPs:* Beyond the asymptotic order, the number of floating-point operations (FLOPs) offers a more direct and practical measure of computational cost, and serves as another critical metric for algorithm efficiency. In this analysis, we set $K_{\mathrm{h}} = N_{\mathrm{h}}$, $K_{\mathrm{v}} = N_{\mathrm{v}}$, $K_{\mathrm{de}} = N_{\mathrm{sc}}(N_{\mathrm{TC}}\Delta T_{\mathrm{cp}}/\Delta T_{\mathrm{sym}})$, and $K_{\mathrm{do}} = 2N_{\mathrm{sym}}$, where delay domain grids beyond the CP-limited range are discarded, and slight oversampling in the Doppler domain improves channel prediction accuracy. Given these configurations and the scenario parameters detailed in Table II, the per-iteration FLOP counts for **Online TS-DCP**, **Online TS-DCP (UIP)**, and **Online TS-DCP (PG)** are approximately $2.15 \times 10^8$, $1.92 \times 10^8$, and $3.46 \times 10^7$, respectively. In the benchmarks, the FLOP counts for **VKF**, **FIT**, and **PAD** (**MPAD**) are approximately $2.29 \times 10^{10}$, $1.05 \times 10^{10}$, and $2.01 \times 10^9$, respectively, where the AR order $P$ in **VKF** is set to 4, the tensor rank $R$ and the number of iterations $T$ for **FIT** are adapted to propagation environments, with average values of approximately 100 and 50, respectively. To assess the scalability of computational cost with respect to antenna array size, we report the FLOP counts for enlarged antenna arrays in Table III. These results, consistent with the asymptotic order, confirm that the proposed TS-DCP algorithms maintain favorable scalability, highlighting the computational advantage in large-scale array settings.

### C. Performance Evaluation

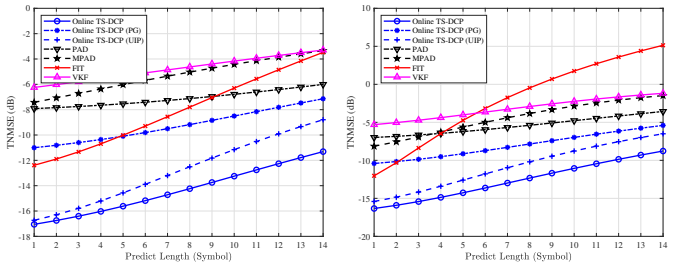
*1) TNMSE for Different Transmission Power:* The TNMSE of the proposed algorithm and benchmarks versus transmission power for $v = 60$ km/h and $v = 120$ km/h are shown in Fig. 6. It is clear that the proposed algorithms significantly outperform the benchmarks across the full transmission power range. Since **FIT** depends on the ALS algorithm, it struggles to incorporate statistical CSI and noise variance, resulting in a noticeable performance drop in the low transmission power regime. Furthermore, **FIT** fails to achieve a TNMSE below 0 dB at $v = 120$ km/h due to the inaccuracy of the first-order Taylor series in modeling temporal correlations at high mobility. While **VKF** benefits from high-order AR processes and outperforms **FIT** in the low transmission power regime, it still struggles due to the neglect of inter-subcarrier and inter-antenna correlations. As a result, **VKF** fails to achieve a TNMSE below $-5$ dB at $P_{\mathrm{T}} = 24$ dBm, even at $v = 60$ km/h. By exploiting statistical CSI by LMMSE estimators and modeling temporal correlations with Prony and matrix pencil methods, **PAD** and **MPAD** outperform both **FIT** and **VKF** across the full transmission power range, with only marginal inferiority to **FIT** at $P_{\mathrm{T}} \geq 20$ dBm and $v = 60$ km/h. Among the proposed algorithms, **Online TS-DCP** is superior to both **Online TS-DCP (PG)** and **Online TS-DCP (UIP)**, owing to the perturbation parameter learning and structured prior. Notably, the proposed algorithms still significantly outperform all benchmarks, even with pre-sampled grids and the unstructured independent prior, highlighting the improvement of channel prediction performance by Doppler domain modeling.

*2) TNMSE for Different Prediction Length:* To illustrate the variation of channel prediction performance with prediction length, we provide the performance for symbols between the current pilot symbols and the first future pilot symbols in Fig. 7. As expected, the performance of the proposed algorithms and the benchmarks deteriorates with increasing prediction length, which is consistent with prior channel prediction studies. Since **FIT** captures the temporal correlations by first-order Taylor series, the modeling error rises with the increasing prediction length, leading to an 8 dB drop in TNMSE at $v = 60$ km/h and a 15 dB drop at $v = 120$ km/h. The TNMSE of **PAD** and **MPAD** is more stable with increasing prediction length, which suggests that Prony and matrix pencil methods provides a more accurate temporal correlation modeling than the first-order Taylor series. The limited exploitation of inter-antenna and inter-subcarrier correlations hinders **VKF** from achieving



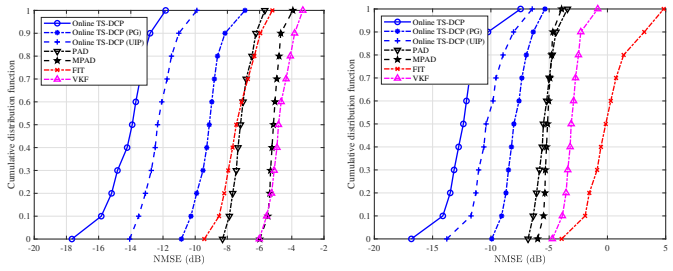

(a) $v = 60$ km/h (b) $v = 120$ km/h

Fig. 6. Channel prediction performance versus transmission power.




(a) $v = 60$ km/h (b) $v = 120$ km/h

Fig. 7. Channel prediction performance versus predict length for $P_{\mathrm{T}} = 24$ dBm.

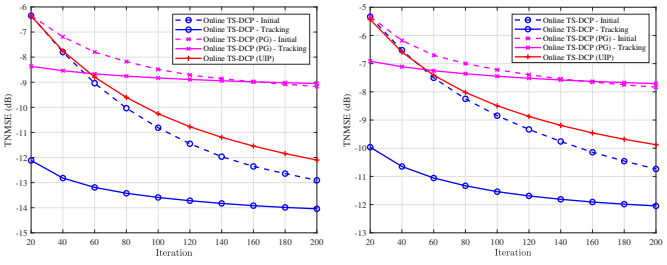



(a) $v = 60$ km/h (b) $v = 120$ km/h

Fig. 8. CDF of channel prediction error for $P_{\mathrm{T}} = 24$ dBm.

satisfactory performance across all prediction lengths. Due to the gridless algorithm in **FIT**, **Online TS-DCP (PG)** is slightly inferior to it for short prediction length, demonstrating the necessity of perturbation parameter learning. Notably, **Online TS-DCP** and **Online TS-DCP (UIP)** outperform all benchmarks across the prediction lengths. Specifically, for the first future non-pilot symbols and pilot symbols, **Online TS-DCP** achieves TNMSEs below $-16$ dB and $-11$ dB at $v = 60$ km/h and below $-15$ dB and $-9$ dB at $v = 120$ km/h, which are unattainable by all other algorithms. The TNMSE gap between **Online TS-DCP (UIP)** and **Online TS-DCP** widens to more than 2 dB with increasing prediction length, indicating that the structured prior contributes to the acquisition of more accurate ADD domain channels.

*3) CDF of Prediction Errors:* In Fig. 8, we present the cumulative distribution function (CDF) of the prediction errors, measured in terms of normalized mean square error (NMSE), for the proposed algorithms and benchmarks. **FIT** and **PAD** exhibit similar prediction error distributions at $v = 60$ km/h,




(a) $v = 60$ km/h (b) $v = 120$ km/h

Fig. 9. Convergence of the proposed algorithms for $P_{\mathrm{T}} = 24$ dBm.

while **FIT** is inferior to **PAD** at $v = 120$ km/h, but both outperform **VKF**. This highlights the inaccuracy of the first-order Taylor series at high mobility and emphasizes the necessity of capturing correlations across antennas and subcarriers for accurate channel prediction. Owing to the inherent similarity between Prony and matrix pencil methods, **PAD** and **MPAD** exhibit nearly identical error distributions, with a consistent performance gap observed under moderate mobility scenarios. Furthermore, **Online TS-DCP (PG)**, **Online TS-DCP (UIP)**, and **Online TS-DCP** achieve the lowest prediction error for both mobility scenarios. Specifically, the 90-th percentile of the prediction error for these algorithms is below $-12.5$ dB, $-11$ dB, and $-8$ dB at $v = 60$ km/h, and below $-10$ dB, $-7$ dB, and $-6$ dB at $v = 120$ km/h, respectively.

*4) TNMSE for Different Scenarios:* To demonstrate the robustness of the proposed algorithms, we consider three typical outdoor scenarios defined in 3GPP specficiations [60]: UMa, urban micro (UMi), and rural macro (RMa). Specifically, UMa features moderate multipath with frequent NLOS conditions, UMi exhibits rich multipath with large delay and angle spreads, and RMa is characterized by dominant line-of-sight paths with weak scattering and small delay spreads. The TNMSE of the proposed algorithms and benchmarks across different propagation scenarios at $P_{\mathrm{T}} = 24$ dBm is summarized in Table IV for both $v = 60$ km/h and $v = 120$ km/h. In the benchmarks, **FIT** suffers from substantial degradation with TNMSEs exceeding 0 dB at $v = 120$ km/h, due to inaccurate temporal correlation modeling based on the first-order Taylor series. Despite employing high-order AR models that outperform **FIT** at $v = 120$ km/h, the channel prediction performance of **VKF** remains limited, with TNMSE exceeding $-5$ dB in most scenarios. Due to the incorporation of angle-delay domain structure and improved temporal correlation modeling, **PAD** and **MPAD** deliver better channel prediction performance than **FIT** and **VKF**, but still remain inferior to the proposed algorithms. Among the proposed algorithms, the proposed **Online TS-DCP** achieves the best performance across all scenarios, achieving TNMSE performance below $-12$ dB at $v = 60$ km/h and below $-10$ dB at $v = 120$ km/h. The simplified variants of the proposed algorithm, **Online TS-DCP (PG)** and **Online TS-DCP (UIP)** still surpass all benchmarks. These trends, previously detailed for UMa scenarios, are consistently observed in both UMi and RMa settings. This reveals that, despite minor performance degradation due to model simplifications, the core design of the proposed TS-

TABLE IV
TNMSE (DB) OF CHANNEL PREDICTION VERSUS SCENARIOS FOR $P_{\mathrm{T}} = 24$ DBM

| Velocity | Scenario | Algorithm | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Online TS-DCP | Online TS-DCP (UIP) | Online TS-DCP (PG) | PAD | MPAD | FIT | VKF |
| 60 km/h | UMa | $\mathbf{-13.98}$ | $-12.10$ | $-9.06$ | $-7.06$ | $-5.08$ | $-7.19$ | $-4.66$ |
| | UMi | $\mathbf{-12.37}$ | $-11.43$ | $-8.60$ | $-7.33$ | $-4.70$ | $-6.15$ | $-4.52$ |
| | RMa | $\mathbf{-16.56}$ | $-13.51$ | $-9.91$ | $-9.73$ | $-5.72$ | $-4.88$ | $-5.01$ |
| 120 km/h | UMa | $\mathbf{-11.97}$ | $-9.87$ | $-7.72$ | $-5.30$ | $-3.91$ | $0.51$ | $-2.96$ |
| | UMi | $\mathbf{-10.40}$ | $-8.54$ | $-6.96$ | $-4.34$ | $-3.49$ | $2.05$ | $-2.79$ |
| | RMa | $\mathbf{-13.74}$ | $-11.54$ | $-8.31$ | $-5.80$ | $-4.42$ | $3.43$ | $-3.64$ |

DCP approaches remains effective across diverse scenarios.

*5) Convergence Behaviors:* To observe the trend of channel prediction performance with respect to the number of iterations, we present the convergence performance of **Online TS-DCP**, **Online TS-DCP (UIP)**, and **Online TS-DCP (PG)** at $P_{\mathrm{T}} = 24$ dBm in Fig. 9, separating the initial and tracking phases. It is evident that **Online TS-DCP** and **Online TS-DCP (UIP)** exhibit fast convergence rates and achieve favorable convergence results in both phases, owing to the model accuracy improvement enabled by perturbation parameter learning. Compared to **Online TS-DCP (PG)**, **Online TS-DCP** and **Online TS-DCP (UIP)** reduce the TNMSE from approximately $-9$ dB to nearly $-12$ dB and $-13$ dB at $v = 60$ km/h, respectively, while from about $-8$ dB to below $-9.5$ dB and $-11$ dB at $v = 120$ km/h, respectively. Besides, the structured prior design in **Online TS-DCP** and **Online TS-DCP (PG)** enables a rapid reduction in TNMSE during the initial iterations of the tracking phase, allowing for early termination to reduce computational complexity based on practical system performance requirements. Specifically, in the early iterations of the tracking phase, **Online TS-DCP** and **Online TS-DCP (PG)** achieve the TNMSE gains approximately $1.5\sim2$ dB and $4.5\sim5.5$ dB compared to the initial phase, respectively, with the gap narrowing as the number of iterations increases.

## VI. CONCLUSION

This paper investigated TS-DCP for massive MIMO with temporal non-stationarity in moderate- to high-mobility scenarios. Specifically, by framing the pilot symbols, we captured intra-frame and inter-frame correlations through Doppler domain modeling and Markov/AR processes, respectively. Besides, we employed MRF and TCGD to model significant cross-domain correlations of the sparsity structure and power patterns in the ADD domain channel. Building on the probabilistic models, the TS-DCP was formulated as the VFE minimization problem, unifying different variational inference rules through the structure design of the trial beliefs. By leveraging the multi-linear structure of channels, the proposed online TS-DCP algorithm achieved substantial reductions in computational complexity. Numerical results demonstrated significant superiority of proposed algorithms over benchmarks in channel prediction performance. In the future, we will investigate deep learning approaches to further enhance the channel prediction performance in massive MIMO-OFDM systems under complicated propagation environments.

## APPENDIX A
## PROOF OF PROPOSITION 1

The joint VFE in (18) is given by

$$F_{\mathrm{V}} = F_{\mathrm{NE}} - F_{\mathrm{EL}}, \tag{50}$$

where $F_{\mathrm{NE}}$ and $F_{\mathrm{EL}}$ denote the negative entropy and expected log-likelihood subfunctions, respectively, given in (51) at the top of the next page. By substituting (19), (20), and (21) into (50), they are simplified as (52) at the top of the next page.

After that, we further factorize the joint PDF as (53) at the top of the next page based on Bayes rules. By substituting (22) and (53) into (52), and incorporating normalization constraints of trial beliefs, we complete the proof.

## APPENDIX B
## PROOF OF PROPOSITION 2

The problem (29) can be solved by the Lagrange multiplier method, and the Lagrange function is expressed as

$$L_{\mathrm{MO},n_{\mathrm{F}}} = F_{\mathrm{MO},n_{\mathrm{F}}} + L_{\text{MO-FO},n_{\mathrm{F}}} + L_{\text{MO-SO},n_{\mathrm{F}}}, \tag{54}$$

where $L_{\text{MO-FO},n_{\mathrm{F}}}$ and $L_{\text{MO-SO},n_{\mathrm{F}}}$ are the subfunctions of FO-SSCCs and SO-SSCCs defined in (57a) and (57b) at the top of the next page, respectively.

We begin with the fixed-point equations for factor beliefs $\mathsf{b}_{Y,n_{\mathrm{F}}}$, $\mathsf{b}_{HG,n_{\mathrm{F}}}$, and $\mathsf{b}_{G,n_{\mathrm{F}}}$, given in (30a), (30b), and (30c), respectively. By differentiating $L_{\mathrm{MO},n_{\mathrm{F}}}$ with respect to $\mathsf{b}_{Y,n_{\mathrm{F}}}$ and setting the result to zero, we can obtain the fixed-point equation of $\mathsf{b}_{Y,n_{\mathrm{F}}}$, which satisfies

$$\begin{aligned}\ln(\mathsf{b}_{Y,n_{\mathrm{F}}}) =& \ln(\mathsf{P}_{Y,n_{\mathrm{F}}}) - 2\mathrm{Re}\{\langle \boldsymbol{\mathcal{U}}_{n_{\mathrm{F}}}^{H,b_H}, \boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}\rangle\} + \\ & \langle \boldsymbol{\mathcal{E}}_{n_{\mathrm{F}}}^{H,b_H}, \|\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}} - \mathsf{E}_{\mathsf{b}_{Y,n_{\mathrm{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}\}\|^{\odot 2}\rangle + C_{b_Y,n_{\mathrm{F}}},\end{aligned} \tag{55}$$

where $C_{b_Y,n_{\mathrm{F}}}$ denotes the term independent of $\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}$. The fixed-point equation of $\mathsf{b}_{Y,n_{\mathrm{F}}}$ is reformulated as in (30a), where the Gaussian distribution factor arises from the relaxation from MCCs to FO- and SO-SSCCs. Similarly, the fixed-point equations for $\mathsf{b}_{HG,n_{\mathrm{F}}}$ and $\mathsf{b}_{G,n_{\mathrm{F}}}$ given in (30b) and (30c) can also be derived by differentiating $L_{\mathrm{MO},n_{\mathrm{F}}}$ with respect to $\mathsf{b}_{HG,n_{\mathrm{F}}}$ and $\mathsf{b}_{G,n_{\mathrm{F}}}$, respectively, and setting the results to zero. In the fixed-point eqaution of $\mathsf{b}_{HG,n_{\mathrm{F}}}$, we substitute the FO-SSCCs of $\mathsf{b}_{Y,n_{\mathrm{F}}}$ and $\mathsf{b}_{HG,n_{\mathrm{F}}}$ with respect to $\boldsymbol{\mathcal{H}}_{n_{\mathrm{F}}}$, as well as $\mathsf{b}_{G,n_{\mathrm{F}}}$ and $\mathsf{b}_{HG,n_{\mathrm{F}}}$ with respect to $\boldsymbol{\mathcal{G}}_{n_{\mathrm{F}}}$ to the derivatives.

Next, we move to the fixed-point equation for the beliefs of the variable beliefs $\mathsf{f}_{H,n_{\mathrm{F}}}$ and $\mathsf{f}_{G,n_{\mathrm{F}}}$, given by (30d) and (30e), respectively. By differentiating $L_{\mathrm{MO},n_{\mathrm{F}}}$ with respect to

$$F_{\text{NE}} = \int \mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}; \mathcal{P}_{\text{OG},(N_\text{F})}, \mathcal{P}_{\text{HP}}) \ln[\mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}; \mathcal{P}_{\text{OG},(N_\text{F})}, \mathcal{P}_{\text{HP}})], \tag{51a}$$

$$F_{\text{EL}} = \int \mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}; \mathcal{P}_{\text{OG},(N_\text{F})}, \mathcal{P}_{\text{HP}}) \ln[\mathsf{P}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}, \boldsymbol{\mathcal{Y}}_{(N_\text{F})}; \mathcal{P}_{\text{OG},(N_\text{F})}, \mathcal{P}_{\text{HP}})]. \tag{51b}$$

$$F_{\text{NE}} = \int \mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}) \ln[\mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})})], \tag{52a}$$

$$F_{\text{EL}} = \int \mathsf{b}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}) \ln[\mathsf{P}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}, \boldsymbol{\mathcal{Y}}_{(N_\text{F})}; \hat{\mathcal{P}}_{\text{OG},(N_\text{F})}, \hat{\mathcal{P}}_{\text{HP}})]. \tag{52b}$$

$$\begin{aligned}
&\mathsf{P}(\boldsymbol{\mathcal{H}}_{(N_\text{F})}, \boldsymbol{\mathcal{G}}_{(N_\text{F})}, \boldsymbol{\mathcal{S}}_{(N_\text{F})}, \boldsymbol{\mathcal{Q}}_{(N_\text{F})}, \boldsymbol{\mathcal{Y}}_{(N_\text{F})}; \hat{\mathcal{P}}_{\text{OG},(N_\text{F})}, \hat{\mathcal{P}}_{\text{HP}}) \\
&\quad \propto \prod\nolimits_{n_\text{F}=1}^{N_\text{F}} [\mathsf{P}(\boldsymbol{\mathcal{Y}}_{n_\text{F}} \mid \boldsymbol{\mathcal{H}}_{n_\text{F}}) \mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\text{F}} \mid \boldsymbol{\mathcal{G}}_{n_\text{F}}; \hat{\mathcal{P}}_{\text{OG},n_\text{F}}) \mathsf{P}(\boldsymbol{\mathcal{G}}_{n_\text{F}} \mid \boldsymbol{\mathcal{S}}_{n_\text{F}}, \boldsymbol{\mathcal{Q}}_{n_\text{F}}) \mathsf{P}(\boldsymbol{\mathcal{S}}_{n_\text{F}} \mid \boldsymbol{\mathcal{S}}_{n_\text{F}-1}; \hat{\boldsymbol{\mathcal{M}}}) \mathsf{P}(\boldsymbol{\mathcal{Q}}_{n_\text{F}} \mid \boldsymbol{\mathcal{Q}}_{n_\text{F}-1}; \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}})] \\
&\quad \propto \prod\nolimits_{n_\text{F}=1}^{N_\text{F}} \mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\text{F}}, \boldsymbol{\mathcal{G}}_{n_\text{F}}, \boldsymbol{\mathcal{S}}_{n_\text{F}}, \boldsymbol{\mathcal{Q}}_{n_\text{F}}, \boldsymbol{\mathcal{Y}}_{n_\text{F}} \mid \boldsymbol{\mathcal{S}}_{n_\text{F}-1}, \boldsymbol{\mathcal{Q}}_{n_\text{F}-1}; \hat{\mathcal{P}}_{\text{OG},n_\text{F}}, \hat{\mathcal{P}}_{\text{HP}})
\end{aligned} \tag{53}$$

$\mathsf{f}_{H,n_\text{F}}$ and setting the result to zero, we obtain the fixed-point equation of $\mathsf{f}_{H,n_\text{F}}$, which satisfies

$$\begin{aligned}
\ln(\mathsf{f}_{H,n_\text{F}}) =& \langle \boldsymbol{\mathcal{E}}_{n_\text{F}}^{H,b_H} + \boldsymbol{\mathcal{E}}_{n_\text{F}}^{H,b_{HG}}, \|\boldsymbol{\mathcal{H}}_{n_\text{F}} - \mathsf{E}_{\mathsf{f}_{H,n_\text{F}}}\{\boldsymbol{\mathcal{H}}_{n_\text{F}}\}\|^{\odot 2} \rangle \\
& - 2\text{Re}\{\langle \boldsymbol{\mathcal{U}}_{n_\text{F}}^{H,b_H} + \boldsymbol{\mathcal{U}}_{n_\text{F}}^{H,b_{HG}}, \boldsymbol{\mathcal{H}}_{n_\text{F}} \rangle\} + C_{f_{H,n_\text{F}}},
\end{aligned} \tag{56}$$

where $C_{f_{H,n_\text{F}}}$ denotes the term independent of $\boldsymbol{\mathcal{H}}_{n_\text{F}}$. Due to the FO-SSCCs of $\mathsf{b}_{Y,n_\text{F}}$ and $\mathsf{f}_{H,n_\text{F}}$ with respect to $\boldsymbol{\mathcal{H}}_{n_\text{F}}$, the fixed-point equation of $\mathsf{f}_{H,n_\text{F}}$ is reformulated as in (30d), where the product of Gaussian distribution factors arises from the relaxation from MCCs to FO- and SO-SSCCs of both $\mathsf{b}_{Y,n_\text{F}}$ and $\mathsf{b}_{HG,n_\text{F}}$. Similarly, the fixed-point equations for $\mathsf{f}_{G,n_\text{F}}$ given in (30e) can also be derived by differentiating $L_{\text{MO},n_\text{F}}$ with respect to $\mathsf{f}_{G,n_\text{F}}$, substituting the FO-SSCCs of $\mathsf{b}_{G,n_\text{F}}$ and $\mathsf{f}_{G,n_\text{F}}$ with respect to $\boldsymbol{\mathcal{G}}_{n_\text{F}}$, and setting the result to zero.

## Appendix C
## Proof of Proposition 3

Due to the singularity in Dirac Delta functions, the conditional PDF $\mathsf{P}(\boldsymbol{\mathcal{H}}_{n_\text{F}} | \boldsymbol{\mathcal{G}}_{n_\text{F}}; \hat{\mathcal{P}}_{\setminus \text{h},n_\text{F}})$ is approximated by the Gaussian distribution with variance approaching zero. On this basis, the objective function of perturbation parameter learning in the horizontal angle domain is given as $f(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}}) = f_1(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}}) + f_2(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}})$, where the objective subfunctions $f_1(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}})$ and $f_2(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}})$ are defined in (63) at the top of the next page.

The parametrized factor matrix $\mathbf{A}_\text{h}(\bar{\boldsymbol{\theta}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}})$ is approximated by the first-order Taylor series, given by

$$\mathbf{A}_\text{h}(\bar{\boldsymbol{\theta}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}}) \approx \mathbf{A}_\text{h}(\bar{\boldsymbol{\theta}}) + \dot{\mathbf{A}}_\text{h}(\bar{\boldsymbol{\theta}}) \text{diag}\{\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}}\}, \tag{58}$$

where $\dot{\mathbf{A}}_\text{h}(\cdot)$ denotes the derivative matrix of the horizontal angle domain parametrized factor matrix with respect to the horizontal direction cosine. Based on the Taylor approximation and the commutativity of tensor-matrix multiplication across dimensions, the first subfunction $f_1(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}})$ is expressed as

$$\begin{aligned}
f_1(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}}) &= \|\Delta\hat{\boldsymbol{\mathcal{H}}}_{n_\text{F}} - \hat{\boldsymbol{\mathcal{G}}}_{n_\text{F}}^\text{h} \times_1 \dot{\mathbf{A}}_\text{h}(\bar{\boldsymbol{\theta}}) \text{diag}\{\hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}}\}\|_F^2 \\
&= \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}}^T \boldsymbol{\Pi}_{\text{h},1} \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}} - 2\boldsymbol{\mu}_{\text{h},1}^T \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}} + C_1,
\end{aligned} \tag{59}$$

where $C_1$ denotes the term independent of $\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}}$, $\boldsymbol{\Pi}_{\text{h},1}$ and $\boldsymbol{\mu}_{\text{h},1}$ denote the auxiliary matrix and vector, given by

$$\boldsymbol{\Pi}_{\text{h},1} = \sum\nolimits_n (\dot{\mathbf{A}}_\text{h}^H(\bar{\boldsymbol{\theta}}) \dot{\mathbf{A}}_\text{h}(\bar{\boldsymbol{\theta}}))^* \odot (\hat{\mathbf{g}}_{n_\text{F},n} \hat{\mathbf{g}}_{n_\text{F},n}^H), \tag{60a}$$

$$\boldsymbol{\mu}_{\text{h},1} = \sum\nolimits_n \text{Re}\{\text{diag}^H\{\mathbf{g}_{n_\text{F},n}^\text{h}\} \dot{\mathbf{A}}_\text{h}^H(\bar{\boldsymbol{\theta}}) \Delta\hat{\mathbf{h}}_{n_\text{F},n}^{(1)}\}. \tag{60b}$$

Similarly, the second subfunction $f_2(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}})$ is expressed as

$$\begin{aligned}
f_2(\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}}) &= \|\boldsymbol{\mathcal{E}}_{G,n_\text{F}}^\text{h} \times_1 |\tilde{\mathbf{A}}_\text{h}(\bar{\boldsymbol{\theta}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}})|^{\odot 2}\|_1 \\
&= \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}}^T \boldsymbol{\Pi}_{\text{h},2} \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}} - 2\boldsymbol{\mu}_{\text{h},2}^T \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_\text{F}} + C_2,
\end{aligned} \tag{61}$$

where $C_2$ denotes the term independent of $\hat{\boldsymbol{\Delta}}\theta_{n_\text{F}}$, $\boldsymbol{\Pi}_{\text{h},2}$ and $\boldsymbol{\mu}_{\text{h},2}$ denote the auxiliary matrix and vector, given by

$$\boldsymbol{\Pi}_{\text{h},2} = \sum\nolimits_n (\dot{\mathbf{A}}_\text{h}^H(\bar{\boldsymbol{\theta}}) \dot{\mathbf{A}}_\text{h}(\bar{\boldsymbol{\theta}}))^* \odot \text{diag}\{\boldsymbol{\varepsilon}_{G,n_\text{F},n}^\text{h}\}, \tag{62a}$$

$$\boldsymbol{\mu}_{\text{h},2} = -\sum\nolimits_n \text{Re}\{\text{diag}\{\dot{\mathbf{A}}_\text{h}^H(\bar{\boldsymbol{\theta}}) \tilde{\mathbf{A}}_\text{h}(\bar{\boldsymbol{\theta}})\} \odot \boldsymbol{\varepsilon}_{G,n_\text{F},n}^\text{h}\}. \tag{62b}$$

This concludes the proof.

## Appendix D
## Proof of Proposition 4

We begin with the learning rule for the transition factor $\hat{\boldsymbol{\mathcal{M}}}$ of hidden state tensors. The objective function (43a) is decomposed into the summation of $N_\text{F}$ subfunctions, given by

$$f_{\text{HS,TC}}(\hat{\boldsymbol{\mathcal{M}}}) = \sum\nolimits_{n_\text{F}=1}^{N_\text{F}} f_{\text{HS,TC},n_\text{F}}(\hat{\boldsymbol{\mathcal{M}}}), \tag{64}$$

where $f_{\text{HS,TC}}(\hat{\boldsymbol{\mathcal{M}}})$ denotes the objective function in (43a) and $f_{\text{HS,TC},n_\text{F}}(\hat{\boldsymbol{\mathcal{M}}})$ denotes the $n_\text{F}$-th subfunction, given by

$$f_{\text{HS,TC},n_\text{F}}(\hat{\boldsymbol{\mathcal{M}}}) = \mathsf{E}_{\mathsf{b}_{S,n_\text{F}}}\{\ln \mathsf{P}_\text{TC}(\boldsymbol{\mathcal{S}}_{n_\text{F}} \mid \boldsymbol{\mathcal{S}}_{n_\text{F}-1}; \hat{\boldsymbol{\mathcal{M}}})\}. \tag{65}$$

To incorporate the normalization factor involving $\hat{\boldsymbol{\mathcal{M}}}$ in temporal correlation models, we specify its expression as

$$\begin{aligned}
C_{\text{TC},n_\text{F}} =& 2^{K_\text{h} K_\text{v} K_\text{de} K_\text{do}} \prod_{k_\text{h},k_\text{v},k_\text{de},k_\text{do}} \exp(-[\hat{\boldsymbol{\mathcal{M}}}]_{k_\text{h},k_\text{v},k_\text{de},k_\text{do}}) \\
& \prod_{k_\text{h},k_\text{v},k_\text{de},k_\text{do}} [\exp(2[\hat{\boldsymbol{\mathcal{M}}}]_{k_\text{h},k_\text{v},k_\text{de},k_\text{do}}) + 1].
\end{aligned} \tag{66}$$

$$L_{\text{MO-FO},n_{\text{F}}} = 2\text{Re}\{\langle \boldsymbol{\mathcal{U}}_{n_{\text{F}}}^{H,b_H}, [\mathsf{E}_{\mathsf{b}_{Y,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\} - \mathsf{E}_{\mathsf{f}_{H,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\}]\rangle + \langle \boldsymbol{\mathcal{U}}_{n_{\text{F}}}^{H,b_{HG}}, [\mathsf{E}_{\mathsf{b}_{HG,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\} - \mathsf{E}_{\mathsf{f}_{H,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\}]\rangle +$$
$$\langle \boldsymbol{\mathcal{U}}_{n_{\text{F}}}^{G,b_G}, [\mathsf{E}_{\mathsf{b}_{G,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\} - \mathsf{E}_{\mathsf{f}_{G,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\}]\rangle + \langle \boldsymbol{\mathcal{U}}_{n_{\text{F}}}^{G,b_{HG}}, [\mathsf{E}_{\mathsf{b}_{HG,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\} - \mathsf{E}_{\mathsf{f}_{G,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\}]\rangle\}, \tag{57a}$$

$$L_{\text{MO-SO},n_{\text{F}}} = \langle \boldsymbol{\mathcal{E}}_{n_{\text{F}}}^{H,b_H}, [\mathsf{V}_{\mathsf{b}_{Y,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\} - \mathsf{V}_{\mathsf{f}_{H,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\}]\rangle + \langle \boldsymbol{\mathcal{E}}_{n_{\text{F}}}^{H,b_{HG}}, [\mathsf{V}_{\mathsf{b}_{HG,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\} - \mathsf{V}_{\mathsf{f}_{H,n_{\text{F}}}}\{\boldsymbol{\mathcal{H}}_{n_{\text{F}}}\}]\rangle$$
$$\langle \boldsymbol{\mathcal{E}}_{n_{\text{F}}}^{G,b_G}, [\mathsf{V}_{\mathsf{b}_{G,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\} - \mathsf{V}_{\mathsf{f}_{G,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\}]\rangle + \langle \boldsymbol{\mathcal{E}}_{n_{\text{F}}}^{G,b_{HG}}, [\mathsf{V}_{\mathsf{b}_{HG,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\} - \mathsf{V}_{\mathsf{f}_{G,n_{\text{F}}}}\{\boldsymbol{\mathcal{G}}_{n_{\text{F}}}\}]\rangle. \tag{57b}$$

$$f_1(\hat{\boldsymbol{\Delta}}\theta_{n_{\text{F}}}) = \|\hat{\boldsymbol{\mathcal{H}}}_{n_{\text{F}}} - \hat{\boldsymbol{\mathcal{G}}}_{n_{\text{F}}} \times_1 \mathbf{A}_{\text{h}}(\bar{\theta} + \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_{\text{F}}}) \times_2 \mathbf{A}_{\text{v}}(\bar{\phi} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}_{n_{\text{F}}}^{\star}) \times_3 \mathbf{B}(\bar{\tau} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}_{n_{\text{F}}}^{\star}) \times_4 \mathbf{C}(\bar{\nu} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}_{n_{\text{F}}}^{\star})\|_F^2, \tag{63a}$$

$$f_2(\hat{\boldsymbol{\Delta}}\theta_{n_{\text{F}}}) = \|\boldsymbol{\mathcal{E}}_{G,n_{\text{F}}} \times_1 |\mathbf{A}_{\text{h}}(\bar{\theta} + \hat{\boldsymbol{\Delta}}\boldsymbol{\theta}_{n_{\text{F}}})|^{\odot 2} \times_2 |\mathbf{A}_{\text{v}}(\bar{\phi} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}_{n_{\text{F}}}^{\star})|^{\odot 2} \times_3 |\mathbf{B}(\bar{\tau} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}_{n_{\text{F}}}^{\star})|^{\odot 2} \times_4 |\mathbf{C}(\bar{\nu} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}_{n_{\text{F}}}^{\star})|^{\odot 2}\|_1, \tag{63b}$$

Therefore, the derivative of the $n_{\text{F}}$-th subfunction with respect to $[\hat{\boldsymbol{\mathcal{M}}}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}}$ is given by

$$\frac{\partial f_{\text{HS,TC},n_{\text{F}}}(\hat{\boldsymbol{\mathcal{M}}})}{\partial [\hat{\boldsymbol{\mathcal{M}}}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}}} \approx \frac{1 - \exp(2[\hat{\boldsymbol{\mathcal{M}}}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}})}{1 + \exp(2[\hat{\boldsymbol{\mathcal{M}}}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}})} +$$
$$(2[\hat{\boldsymbol{\mathcal{S}}}_{n_{\text{F}}}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}} - 1)(2[\hat{\boldsymbol{\mathcal{S}}}_{n_{\text{F}}-1}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}} - 1), \tag{67}$$

where the approximation follows from the online approximation. By summing the derivative of all objective subfunctions and setting the result to zero, we have

$$[\hat{\boldsymbol{\mathcal{M}}}^{\star}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}} = \ln \frac{1 + [\hat{\boldsymbol{\mathcal{K}}}_M]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}}}{1 - [\hat{\boldsymbol{\mathcal{K}}}_M]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}}}, \tag{68}$$

where $[\hat{\boldsymbol{\mathcal{K}}}_M]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}}$ is defined by

$$[\hat{\boldsymbol{\mathcal{K}}}_M]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}} = \tag{69}$$
$$\frac{1}{N_{\text{F}}} \sum\nolimits_{n_{\text{F}}=1}^{N_{\text{F}}} (2[\hat{\boldsymbol{\mathcal{S}}}_{n_{\text{F}}}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}} - 1)(2[\hat{\boldsymbol{\mathcal{S}}}_{n_{\text{F}}-1}]_{k_{\text{h}},k_{\text{v}},k_{\text{de}},k_{\text{do}}} - 1).$$

The derivation of the learning rule for transition factors can be completed by organizing the above results into tensor form.

Similar to the case of transition factors $\hat{\boldsymbol{\mathcal{M}}}$, the objective function (43b) is also decomposed as

$$f_{\text{HV,TC}}(\hat{\boldsymbol{\mathcal{L}}}) = \sum\nolimits_{n_{\text{F}}=1}^{N_{\text{F}}} f_{\text{HV,TC},n_{\text{F}}}(\hat{\boldsymbol{\mathcal{L}}}), \tag{70}$$

where $f_{\text{HV,TC}}(\hat{\boldsymbol{\mathcal{L}}})$ denotes the objective function in (43b) and $f_{\text{HV,TC},n_{\text{F}}}(\hat{\boldsymbol{\mathcal{L}}})$ denotes the $n_{\text{F}}$-th subfunction, given by

$$f_{\text{HV,TC},n_{\text{F}}}(\hat{\boldsymbol{\mathcal{L}}}) = \mathsf{E}_{\mathsf{b}_{Q,n_{\text{F}}}}\{\ln \mathsf{P}(\boldsymbol{\mathcal{Q}}_{n_{\text{F}}} \mid \boldsymbol{\mathcal{Q}}_{n_{\text{F}}-1}; \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}})\}, \tag{71}$$

where $\mathsf{P}(\boldsymbol{\mathcal{Q}}_{n_{\text{F}}} \mid \boldsymbol{\mathcal{Q}}_{n_{\text{F}}-1}; \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}})$ is expressed as

$$\mathsf{P}(\boldsymbol{\mathcal{Q}}_{n_{\text{F}}} \mid \boldsymbol{\mathcal{Q}}_{n_{\text{F}}-1}; \hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{V}}}) \propto$$
$$\begin{cases} \mathsf{CN}(\boldsymbol{\mathcal{Q}}_{n_{\text{F}}}; (1 - \hat{\boldsymbol{\mathcal{L}}})\boldsymbol{\mathcal{Q}}_{n_{\text{F}}-1}, \hat{\boldsymbol{\mathcal{L}}}^{\odot 2} \odot \bar{\boldsymbol{\mathcal{V}}}), & n_{\text{F}} > 1 \\ \mathsf{CN}(\boldsymbol{\mathcal{Q}}_{n_{\text{F}}}; \boldsymbol{\mathcal{C}}(0), \bar{\boldsymbol{\mathcal{V}}}), & n_{\text{F}} = 1 \end{cases}, \tag{72}$$

and $\bar{\boldsymbol{\mathcal{V}}}$ is defined in (44). It can be observed that $f_{\text{HV,TC},1}(\hat{\boldsymbol{\mathcal{L}}})$ is independent of $\hat{\boldsymbol{\mathcal{L}}}$, implying that only the subfunctions with $n_{\text{F}} > 1$ need to be considered. By differentiating the subfunctions with $n_{\text{F}} > 1$ with respect to $\hat{\boldsymbol{\mathcal{L}}}$, we yield

$$-\frac{\partial f_{\text{HV,TC},n_{\text{F}}}(\hat{\boldsymbol{\mathcal{L}}})}{\partial \hat{\boldsymbol{\mathcal{L}}}} \approx$$
$$2[\bar{\boldsymbol{\mathcal{V}}} \odot \hat{\boldsymbol{\mathcal{L}}}^{\odot 2} + \boldsymbol{\mathcal{K}}_{L,1,n_{\text{F}}} \odot \hat{\boldsymbol{\mathcal{L}}} + \boldsymbol{\mathcal{K}}_{L,0,n_{\text{F}}}] \oslash (\hat{\boldsymbol{\mathcal{L}}}^{\odot 3} \bar{\boldsymbol{\mathcal{V}}}), \tag{73}$$

where the approximation also follows from the online approximation, $\boldsymbol{\mathcal{K}}_{L,1,n_{\text{F}}}$ and $\boldsymbol{\mathcal{K}}_{L,0,n_{\text{F}}}$ are defined as

$$\boldsymbol{\mathcal{K}}_{L,1,n_{\text{F}}} \triangleq |\hat{\boldsymbol{\mathcal{Q}}}_{n_{\text{F}}-1}|^{\odot 2} - \text{Re}\{\hat{\boldsymbol{\mathcal{Q}}}_{n_{\text{F}}-1}^{*} \odot \hat{\boldsymbol{\mathcal{Q}}}_{n_{\text{F}}}\}, \tag{74a}$$

$$\boldsymbol{\mathcal{K}}_{L,0,n_{\text{F}}} \triangleq -|\hat{\boldsymbol{\mathcal{Q}}}_{n_{\text{F}}} - \hat{\boldsymbol{\mathcal{Q}}}_{n_{\text{F}}-1}|^{\odot 2} - \boldsymbol{\mathcal{E}}_{Q,n_{\text{F}}}. \tag{74b}$$

The derivation of the learning rule for temporal correlation factor is completed by summing the derivatives of the subfunctions with $n_{\text{F}} > 1$ and enforcing the non-zero constraints on both $\hat{\boldsymbol{\mathcal{L}}}$ and $\bar{\boldsymbol{\mathcal{V}}}$, which concludes the proof.

## 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] L. Lu, G. Y. Li *et al.*, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[3] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[4] 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.

[5] H. Jin, K. Liu *et al.*, "Massive MIMO evolution towards 3GPP Release 18," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1635–1654, Jun. 2023.

[6] A. Liu, L. Lian *et al.*, "Downlink channel estimation in multiuser massive MIMO with hidden Markovian sparsity," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4796–4810, Sep. 2018.

[7] H. Lee, H. Choi *et al.*, "Downlink channel reconstruction for spatial multiplexing in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6154–6166, Sep. 2021.

[8] Y. Zhu, G. Sun *et al.*, "OFDM-based massive grant-free transmission over frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4543–4558, Jul. 2022.

[9] H. Miao, J. Zhang *et al.*, "Sub-6 GHz to mmWave for 5G-advanced and beyond: Channel measurements, characteristics and impact on system performance," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1945–1960, Jun. 2024.

[10] 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.

[11] 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.

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

[13] H. Yin, H. Wang *et al.*, "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] Z. Qin, H. Yin *et al.*, "A partial reciprocity-based channel prediction framework for FDD massive MIMO with high mobility," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9638–9652, Nov. 2022.

[15] W. Li, H. Yin *et al.*, "A multi-dimensional matrix pencil-based channel prediction method for massive MIMO with mobility," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2215–2230, Apr. 2023.

[16] Y. Zhu, J. Zhuang *et al.*, "Joint channel estimation and prediction for massive MIMO with frequency hopping sounding," *IEEE Trans. Wireless Commun.*, 2024.

[17] D. Shi, L. Song *et al.*, "Channel acquisition for HF skywave massive MIMO-OFDM communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 4074–4089, Jun. 2023.

[18] X. Wang, Y. Shi *et al.*, "Channel prediction with time-varying Doppler spectrum in high-mobility scenarios: A polynomial Fourier transform based approach and field measurements," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7116–7129, Nov. 2023.

[19] Y. Wan, G. Liu *et al.*, "Robust multi-user channel tracking scheme for 5G new radio," *IEEE Trans. Wireless Commun.*, 2023.

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

[21] J. Ma, S. Zhang *et al.*, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.

[22] M. Li, S. Zhang *et al.*, "Time-varying massive MIMO channel estimation: Capturing, reconstruction, and restoration," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7558–7572, Nov. 2019.

[23] S. Srivastava, M. S. Kumar *et al.*, "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.

[24] S. Srivastava, C. S. K. Patro *et al.*, "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.

[25] C. Wu, X. Yi *et al.*, "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.

[26] H. Kim, S. Kim *et al.*, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.

[27] P. Wang, J. Li *et al.*, "Two-stage based partial online optimization channel prediction for massive MIMO in high-mobility," *IEEE Commun. Lett.*, vol. 28, no. 2, pp. 417–421, Feb. 2024.

[28] J. Xu, L. Li *et al.*, "Learning to estimate: A real-time online learning framework for MIMO-OFDM channel estimation," *IEEE Trans. Wireless Commun.*, 2024.

[29] J. Zhao, H. Xie *et al.*, "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.

[30] L. Lian, A. Liu, and V. K. Lau, "Exploiting dynamic sparsity for downlink FDD-massive MIMO channel tracking," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2007–2021, Apr. 2019.

[31] X. Liu, W. Wang *et al.*, "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.

[32] W. Peng, W. Li *et al.*, "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.

[33] X. Cheng, Z. Huang, and L. Bai, "Channel nonstationarity and consistency for beyond 5G and 6G: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1634–1669, 3rd Quart. 2022.

[34] Z. Huang, X. Cheng, and X. Yin, "A general 3D non-stationary 6G channel model with time-space consistency," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3436–3450, May 2022.

[35] H. Hou, Y. Wang *et al.*, "Joint beam alignment and Doppler estimation for fast time-varying wideband mmWave channels," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10 895–10 910, Sep. 2024.

[36] J. Bian, C.-X. Wang *et al.*, "A general 3D non-stationary wireless channel model for 5G and beyond," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3211–3224, May 2021.

[37] ——, "A novel 3D beam domain channel model for massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1618–1632, Mar. 2023.

[38] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[39] 3rd Generation Partnership Project (3GPP), "NR; physical channels and modulation," 3GPP, Technical Specification (TS) 38.211, 2025. [Online]. Available: https://www.3gpp.org

[40] M. Sørensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, Nov. 2013.

[41] J. H. d. M. Goulart, M. Boizard *et al.*, "Tensor CP decomposition with structured factor matrices: Algorithms and performance," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 757–769, Jun. 2015.

[42] J. E. Cohen and N. Gillis, "Dictionary-based tensor canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1876–1889, Apr. 2017.

[43] C. K. Thomas and D. Slock, "BP-VB-EP based static and dynamic sparse Bayesian learning with Kronecker structured dictionaries," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 9095–9099.

[44] W.-C. Chang and Y. T. Su, "Sparse Bayesian learning based tensor dictionary learning and signal recovery with application to MIMO channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 847–859, Apr. 2021.

[45] L. Cheng, Z. Chen *et al.*, "Towards flexible sparsity-aware modeling: Automatic tensor rank learning using the generalized hyperbolic prior," *IEEE Trans. Signal Process.*, vol. 70, pp. 1834–1849, Apr. 2022.

[46] X. Xu, S. Zhang *et al.*, "Sparse Bayesian learning based channel extrapolation for RIS assisted MIMO-OFDM," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5498–5513, Aug. 2022.

[47] M. Zhang, X. Yuan, and Z.-Q. He, "Variance state propagation for structured sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 68, pp. 2386–2400, 2020.

[48] W. Xu, Y. Xiao *et al.*, "Joint scattering environment sensing and channel estimation based on non-stationary Markov random field," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 3903–3917, May 2023.

[49] X. Liu, W. Wang *et al.*, "Structured hybrid message passing based channel estimation for massive MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7491–7507, Jun. 2023.

[50] C. Zhang, J. Bütepage *et al.*, "Advances in variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 2008–2026, Aug. 2019.

[51] J. Winn, C. M. Bishop, and T. Jaakkola, "Variational message passing," *J. Mach. Learn. Res.*, vol. 6, no. 4, p. 661–694, 2005.

[52] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.

[53] D. Zhang, X. Song *et al.*, "Unifying message passing algorithms under the framework of constrained Bethe free energy minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4144–4158, Jul. 2021.

[54] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[55] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013.

[56] J. Fang, Y. Shen *et al.*, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.

[57] 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.

[58] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[59] H. Hou, X. He *et al.*, "Beam-delay domain channel estimation for mmWave XL-MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 4, pp. 646–661, May 2024.

[60] 3rd Generation Partnership Project (3GPP), "Study on channel model for frequencies from 0.5 to 100 GHz (Release 18)," 3GPP, Technical Report (TR) 38.901, 2024. [Online]. Available: https://www.3gpp.org

[61] S. Jaeckel, L. Raschkowski *et al.*, "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.