Tensor-Structured Bayesian Channel Prediction for Upper Mid-Band XL-MIMO Systems

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

***Abstract*—The upper mid-band balances coverage and capacity for the future cellular systems and also embraces extremely large-scale multiple-input multiple-output (XL-MIMO) systems, offering enhanced spectral and energy efficiency. However, these benefits are significantly degraded under mobility due to channel aging, and further exacerbated by the unique near-field (NF) and spatial non-stationarity (SnS) propagation in such systems. To address this challenge, we propose a novel channel prediction approach that incorporates dedicated channel modeling, probabilistic representations, and Bayesian inference algorithms for this emerging scenario. Specifically, we develop tensor-structured channel models in both the spatial-frequency-temporal (SFT) and beam-delay-Doppler (BDD) domains, which capture the NF and SnS propagation effects and leverage temporal correlations among multiple snapshots for channel prediction. In this model, the factor matrices of multi-linear transformations are parameterized by BDD domain grids and SnS factors, where beam domain grids are jointly determined by angles and slopes under spatial-chirp based NF representations. To enable tractable inference, we replace these environment-dependent BDD domain grids with uniformly sampled ones, and introduce perturbation parameters in each domain to mitigate grid mismatch. We further propose a hybrid beam domain strategy that integrates angle-only sampling with slope hyperparameterization to avoid the computational burden of explicit slope sampling. On this basis, we develop tensor-structured bi-layer inference (TS-BLI) algorithm under the expectation-maximization (EM) framework, which reduces the computational complexity by leveraging the inherent separation across different domains. In the E-step, we develop the bi-layer factor graph representation to isolate the bilinear mixing in the spatial domain induced by SnS propagation, thus facilitating bi-layer iterations using approximate inference techniques. In the M-step, we leverage an alternating strategy for hyperparameter learning, with closed-form rules derived by the quadratic approximation of objective functions. Numerical simulations based on a near-practical channel simulator developed upon QuaDRiGa with SnS extensions demonstrate the superior channel prediction performance of the proposed algorithm.**

***Index Terms*—XL-MIMO, channel prediction, tensor representation, near-field, spatial non-stationarity**

Hongwei Hou, Yafei Wang, 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; 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).

I. Introduction

As next-generation cellular communication systems strive to support data-intensive applications, the demand for spectrum that enables both high capacity and broad coverage has become increasingly critical [1], [2]. Motivated by this demand, growing attention has turned to the newly defined frequency range (FR), with which FR3 (upper mid-band, 7.125-24.25 GHz) has emerged as the most promising candidate that offers a favorable trade-off between the spectral congestion of FR1 (sub-7 GHz, 0.41-7.125 GHz) and the severe propagation losses of FR2 (millimeter-wave, 24.25-71.0 GHz) [3], [4].

The relatively short wavelengths of FR3 make it feasible to deploy extremely large aperture arrays (ELAA) at the base station (BS) [5], which have been recognized as a key enabling technology to sustainably boost both spectral and energy efficiency. To unleash this potential, the acquisition of channel state information (CSI) is essential, yet increasingly challenging under mobility due to channel aging, i.e., the mismatch between the estimated and actual channels caused by temporal variations. This challenge is further exacerbated in FR3 systems, where the higher carrier frequencies induce more significant channel aging than in FR1, and the richer multipath propagation compared to FR2 renders straightforward Doppler compensation ineffective [6]. Therefore, channel prediction has emerged as a promising solution by exploiting temporal correlations in historical CSI to predict future CSI. However, the large physical aperture of ELAA brings scatterers and mobile terminals (MTs) within the Rayleigh distance, and also results in non-uniform visibility across the array [7]. These effects give rise to unique channel characteristics in FR3, incorporating both near-field (NF) and spatial non-stationarity (SnS) propagation, which reshape the beam domain channel representations and further complicate channel prediction.

A. Previous Works

Before delving into the challenges specific to FR3 systems, it is crucial to review representative channel prediction

techniques developed for massive multiple-input multiple-output (MIMO) systems. Among the earliest contributions, [8], [9] explore channel prediction using spectral estimation techniques and parametric channel models, where [9] also derives performance bounds that quantify the theoretical limits of channel prediction. In the context of massive MIMO with orthogonal frequency-division multiplexing (OFDM), [10], [11] extract Doppler frequencies from dominant angle-delay domain channel taps, while [12] adopts autoregressive (AR) modeling as an alternative approach, with both achieving significant gains. For more practical channel dynamics, [13], [14] incorporate time-varying Doppler frequencies through polynomial expansions and AR processes, as exemplified by [13] through validation on measured channel data. To alleviate the effect of channel estimation error, [15]–[17] unify the channel estimation and prediction based on beam-delay-Doppler (BDD) domain representations, and achieve the minimum mean square error (MMSE)-based channel prediction.

Despite extensive research on channel prediction in massive MIMO systems, this topic remains largely unexplored in the context of extremely large-scale MIMO (XL-MIMO) equipped with ELAA. Most existing works focus on channel estimation or tracking, which aim to reconstruct the currently observed channels rather than to predict future, unobserved ones. Specifically, [18]–[20] propose polar domain representations based on the angle-distance sampling to exploit the inherent sparsity under NF propagation, where [18] introduces grid refinement to mitigate quantization errors. Besides NF propagation, recent studies have focused on capturing SnS, which has emerged as another fundamental channel characteristic in XL-MIMO systems. Assuming local spatial stationarity within small subarrays, [21] and [22] partition the ELAA into uniform sub-arrays, while [23] proposes an adaptive segmentation strategy to enhance modeling accuracy. In contrast, [24] directly models the antenna domain sparsity induced by SnS, focusing exclusively on the structural support of active elements. Building upon this, [25]–[27] explicitly incorporate inter-antenna correlations caused by multipath propagation using beam domain representations. For modeling simplification, [25] and [27] assume shared visibility regions across all paths and directional beam-based sub-channel decomposition, respectively, while [26] designs multi-module architectures that alternate between beam domain channel and SnS factor inference. Most recently, [28] presents the first attempt to address channel prediction in XL-MIMO systems, where the spherical wavefronts are approximated as near-planar ones based on wavefront transformations (WT), and Doppler frequencies are extracted by the WT matrix pencil (WTMP) approach to facilitate channel prediction.

B. Motivations and Contributions

In upper mid-band XL-MIMO systems, the higher carrier frequency than FR1 and the richer multipath than FR2 bring the increased Doppler frequencies and spread, which emphasizes the necessity of channel prediction in combating more severe channel aging. However, the deployment of ELAA introduces the coexistence of NF and SnS propagation, rendering the exploitation of inter-antenna correlation challenging and further degrading the channel prediction performance. Despite these challenges, existing efforts remain limited—either ignoring SnS propagation or restricting attention to the channel estimation task—thereby leaving a critical gap in channel prediction for upper mid-band XL-MIMO systems.

To bridge this gap, we investigate tensor-structured Bayesian channel prediction for upper mid-band XL-MIMO systems in this paper, pioneering the integration of NF and SnS propagation effects into channel prediction. The main contributions of this work are summarized as follows:

- Within the fixed-length observation window, we develop the tensor-structured spatial-frequency-temporal (SFT) domain and BDD domain channel models with parameterized factor matrices, incorporating both NF and SnS propagation effects. Specifically, the spatial domain factor matrices are parameterized by instantaneous angles and variation rates (i.e., slopes) under the spatial-chirp representation of NF propagation, while antenna-dependent visibility factors capture the SnS propagation. This modeling framework enables arbitrary-length channel prediction through the transformation from the Doppler domain to the temporal domain, while the beam and delay domains further enhance the channel prediction by capturing the inter-antenna and inter-subcarrier correlations.
- To enable tractable inference, we propose dedicated probabilistic models for channel prediction, which capture multi-linear transformations, BDD domain channel and SnS factor sparsity. Specifically, we sample the BDD domain uniformly and introduce perturbation parameters in each domain to alleviate grid mismatch, thereby replacing dynamic grids in multi-linear transformations by fixed ones. We further adopt a hybrid beam domain sampling strategy to bypass the increased computational complexity due to extra slope sampling under NF propagation, which integrates angle-only sampling and slope hyper-parameterization to strikes the balance between model complexity and expressive capacity.
- Following these probabilistic models, we propose the tensor-structured bi-layer inference (TS-BLI) algorithm within the expectation-maximization (EM) framework for channel prediction. Specifically, we present a bi-layer factor graph in the E-step of the EM framework to isolate the bilinear mixing in the spatial domain induced by SnS propagation, thereby deriving tractable iterations via approximate inference techniques. In the M-step, the alternating optimization is employed to shrink the search space of hyperparameters, while approximating the objective functions yields near-optimal closed-form update rules. By leveraging the inherent separation across different domains, the proposed tensor-structured inference not only decouples linear and bilinear mixing, but also avoids multi-domain joint processing to significantly reduce computational complexity.

C. Organization and Notations

1) Organization: The SFT domain channel model and its Tucker-based BDD domain representations are developed in

Section II. Based on the probabilistic models, we formulate the MMSE-based channel prediction problem and propose the corresponding TS-MLI algorithm in Section III and Section IV, respectively. Section V presents the numerical simulations, and Section VI concludes the paper.

2) 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 (KL) divergence and differential entropy, respectively. The outer product, element-wise multiplication, and division are denoted by $\circ$, $\odot$, and $\oslash$ 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.

3) Tensor Notations: The tensor operations and definitions in this paper align with the counterparts in [29]. 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 this tensor into its column vectors, obtained by fixing the index along the d-th mode and varying the others. 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 ℓ_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 $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_d \mathbf{U}_d$, expressed as

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

and equivalent to $\mathbf{Y}_d = \mathbf{U}_d\mathbf{X}_d$. In analogy with the Einstein product [30], we define the special case of tensor-tensor multiplications for $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M_1\times\ldots\times M_{d-1}\times P\times M_{d+1}\times\ldots\times M_D}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_1\times\ldots\times M_{d-1}\times Q\times M_{d+1}\times\ldots\times M_D}$ as $\mathbf{Z} = \boldsymbol{\mathcal{X}} \times_{-d} \boldsymbol{\mathcal{Y}}$. This operation applies to all modes except for the d-th one and is equivalent to the multiplication of their mode-d matricizations, i.e., $\mathbf{Z} = \mathbf{X}_d\mathbf{Y}_d^T$. $\mathsf{CN}(\boldsymbol{\mathcal{X}};\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{E}})$, $\mathsf{BG}(\boldsymbol{\mathcal{X}};\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{E}})$, and $\boldsymbol{\delta}(\boldsymbol{\mathcal{X}})$ denote the tensor-structured joint probability density functions (PDFs) of $\boldsymbol{\mathcal{X}}$, whose entries are independently drawn from the corresponding scalar complex Gaussian (CN), Bernoulli–Gaussian (BG), and Dirac delta distributions, respectively, with element-wise parameters of mean $\boldsymbol{\mathcal{U}}$, variance $\boldsymbol{\mathcal{E}}$, and sparsity $\boldsymbol{\mathcal{M}}$. Specifically, the scalar CN distribution with mean μ and variance σ^2 is defined as

$$\mathsf{p}_{\mathrm{CN}}(x;\mu,\sigma^2) = (\pi\sigma^2)^{-1}\exp\left(-|x-\mu|^2/\sigma^2\right), \quad (3)$$

and the scalar BG distribution with sparsity ρ, mean μ, and variance σ^2 is defined as

$$\mathsf{p}_{\mathrm{BG}}(x;\rho,\mu,\sigma^2) = (1-\rho)\delta(x) + \rho\mathsf{p}_{\mathrm{CN}}(x;\mu,\sigma^2), \quad (4)$$

where $\delta(\cdot)$ denotes the Dirac delta function that equals zero everywhere except at zero and integrates to one. $\boldsymbol{\mathcal{C}}(x)$ denotes the constant tensor with all elements being x.

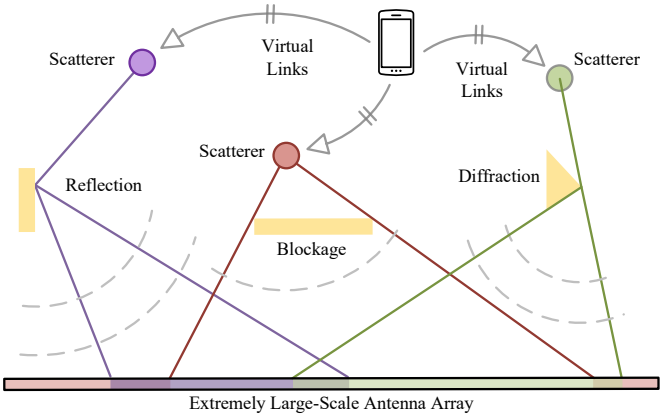


Fig. 1. NF and SnS propagation in XL-MIMO systems, with reflection, blockage, and diffraction illustrated as examples.

II. System Model

In this paper, we consider the XL-MIMO-OFDM system in the upper mid-band under the time-division duplexing (TDD) mode. The BS, equipped with a ULA of $N_{\mathrm{an}} \gg 1$ antennas, serves one typical single-antenna mobile terminal (MT) that moves along a 2D trajectory at a constant velocity[1]. The deployment of ELAAs induces both NF and SnS propagation effects in XL-MIMO systems, as illustrated in Fig. 1 and mathematically modeled in the subsequent subsection. In this context, NF arises when the last-hop scatterer (LHS) is located within the Rayleigh distance, while SnS results from phenomena such as partial reflection, blockage, and diffraction.

In this system, the OFDM symbol duration, cyclic prefix (CP) duration, and subcarrier spacing are denoted by ΔT_{sym}, ΔT_{cp}, and Δf respectively, leading to the total OFDM symbol duration with CP given by $\Delta T = \Delta T_{\mathrm{sym}} + \Delta T_{\mathrm{cp}}$. The MTs employ sounding reference signal (SRS) as pilots for uplink channel sounding [31], which is characterized by comb-type pilot patterns with uniform spacing in both temporal and frequency domains. Specifically, pilot symbols are inserted every N_{IS} OFDM symbols and N_{TC} subcarriers in the temporal and frequency domains, respectively, corresponding to pilot spacings of $\Delta\bar{T} = N_{\mathrm{IS}}\Delta T$ and $\Delta\bar{f} = N_{\mathrm{TC}}\Delta f$. With this configuration, we introduce the sliding frame structure shown in Fig. 2, where each frame contains multiple pilot OFDM symbols. In each sliding frame, the BS collects received SRSs within the current frame as inputs to predict the channels of all subsequent OFDM symbols until the next SRS appears. As the frame slides forward over time, this process repeats continuously to enable uninterrupted channel prediction, thereby mitigating channel aging.

A. Signal and Channel Models

Based on ray-tracing models, the channel impulse response (CIR) at the n_{an}-th antenna consists of L paths, expressed as

$$h_{n_{\mathrm{an}}}(t,\tau) = \sum_{l=1}^{L} \beta_l \tilde{s}_{l,n_{\mathrm{an}}} \exp(j2\pi t\tilde{\nu}_l)\delta(\tau - \tilde{\tau}_l - \Delta\tilde{\tau}_{l,n_{\mathrm{an}}}), \quad (5)$$

[1]Although multiple MTs may exist, each is processed independently owing to orthogonal pilot allocation.

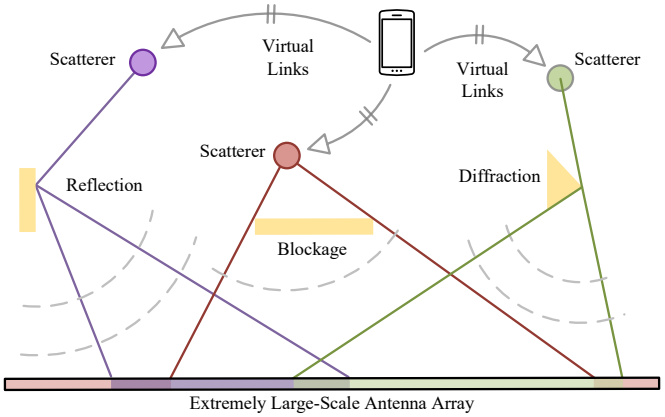


Fig. 2. Sliding frame structure for channel prediction in upper mid-band XL-MIMO-OFDM systems under TDD mode.

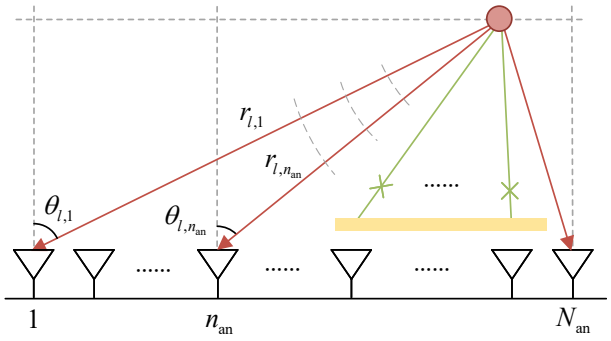


Fig. 3. NF and SnS propagations in the upper mid-band XL-MIMO systems.

where β_l and $\tilde{\nu}_l$ denote complex path gain and Doppler frequency of the l-th path, respectively, $\tilde{\tau}_l$ denotes the reference propagation delay from the LHS to the first antenna, $\tilde{s}_{l,n_{\text{an}}}$ denotes the SnS factor of the l-th path at the n_{an}-th antenna, and $\Delta\tilde{\tau}_{l,n_{\text{an}}}$ denotes the propagation delay difference between the n_{an}-th and the first antenna for the l-th path[2], respectively.

To effectively capture the NF and SnS characteristics of XL-MIMO propagation, we develop the geometric relationships among the BS array and scatterers, with the example of blockage illustrated in Fig. 3. This geometry provides the foundation for defining the SnS factor $\tilde{s}_{l,n_{\text{an}}}$ and the delay deviation $\Delta\tilde{\tau}_{l,n_{\text{an}}}$ in the CIR expression. Specifically, the limited spatial extent of blockage leads to non-uniform visibility across the array, implying that these paths are visible only to a subset of antennas, which can be captured by $\tilde{s}_{l,n_{\text{an}}}$. The channel measurements from both academia and industry have demonstrated that binary-valued SnS factors are sufficient to capture this essential features [7], [33], thereby setting $\tilde{s}_{l,n_{\text{an}}} \in \{0,1\}$. In addition, due to the non-negligible spherical wavefronts under NF propagation, the propagation delay varies non-linearly with respect to the antenna index. Geometrically, this non-linear variation follows the constant projection constraint; that is, the propagation distances from different antennas to the LHS maintain identical projections onto the array-normal axis, given by

$$\tilde{r}_{l,n_{\text{an}}} \cos\tilde{\theta}_{l,n_{\text{an}}} = \tilde{r}_{l,1} \cos\tilde{\theta}_{l,1}, \tag{6}$$

where $\tilde{r}_{l,n_{\text{an}}}$ and $\tilde{\theta}_{l,n_{\text{an}}}$ denote the propagation distance and the angle of arrival between the n_{an}-th antenna and LHS of the l-th path, respectively, with the former further expressed as[3]

$$\begin{aligned}\tilde{r}_{l,n_{\text{an}}} &\overset{(a)}{=} \sqrt{\left[\tilde{r}_{l,1}\cos\tilde{\theta}_{l,1}\right]^2 + \left[\tilde{r}_{l,1}\sin\tilde{\theta}_{l,1} - (n_{\text{an}}-1)d\right]^2} \\ &\overset{(b)}{\approx} \tilde{r}_{l,1} - (n_{\text{an}}-1)d\tilde{\phi}_{l,1} + (n_{\text{an}}-1)^2 d^2 \frac{1-\tilde{\phi}_{l,1}^2}{2\tilde{r}_{l,1}},\end{aligned} \tag{7}$$

[2]We consider only phase variations induced by spherical wavefronts. This approximation holds when the propagation distance exceeds $\zeta D/(2\sqrt{1-\zeta^2})$ [32], where D is the array aperture and ζ is amplitude variation thresholds.

[3]We retain up to the second-order term and ignore the high-order ones, which are valid for propagation distances exceeding $D^{4/3}/(2\lambda^{1/3})$ [34].

where (a) and (b) follow from the constant projection constraint above and the second-order Taylor series of $\sqrt{1+x}$ at $x=0$, respectively, $\tilde{\phi}_{l,1} \triangleq \sin\tilde{\theta}_{l,1}$ and $\tilde{r}_{l,1}$ denote the reference directional cosine and propagation distance of the l-th path, referred to as $\tilde{\phi}_l$ and $\tilde{r}_l$, respectively, and d denotes the inter-antenna spacing. With such definition, we have

$$\Delta\tilde{\tau}_{l,n_{\text{an}}} = [-(n_{\text{an}}-1)d\tilde{\phi}_l + (n_{\text{an}}-1)^2 d^2 \tilde{\eta}_l]/c, \tag{8}$$

where $\tilde{\eta}_l \triangleq (1-\tilde{\phi}_l^2)/(2\tilde{r}_l)$ denotes slope parameters of the l-th path, characterizing the deviation of the spherical wavefront from the planar one under NF propagation. In this context, the term "slope" stems from the spatial-chirp representation of the array response, which captures the variation of instantaneous direction cosine with respect to the antenna index.

Due to the sufficiently short duration of OFDM symbols, CIRs are assumed to be constant within one OFDM symbol. By Fourier transforming and stacking the CIRs of all antennas and pilot resource elements, the spatial-frequency-temporal (SFT) domain channel $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{N_{\text{an}}\times N_{\text{sc}}\times N_{\text{sym}}}$ is given by

$$\boldsymbol{\mathcal{H}} = \sum_{l=1}^{L} \beta_l (\tilde{\mathbf{s}}_l \odot \mathbf{a}_{\text{SS}}(\tilde{\phi}_l, \tilde{\eta}_l)) \circ \mathbf{b}(\tilde{\tau}_l) \circ \mathbf{c}(\tilde{\nu}_l), \tag{9}$$

where $\tilde{\mathbf{s}}_l = [\tilde{s}_{l,1}, \ldots, \tilde{s}_{l,N_{\text{an}}}]^T \in \mathbb{C}^{N_{\text{an}}\times 1}$ denotes the SnS factor of the l-th path, $\mathbf{a}_{\text{SS}}(\phi,\eta) \in \mathbb{C}^{N_{\text{an}}\times 1}$ denotes beam domain steering vector without SnS propagation effect, defined as $[\mathbf{a}_{\text{SS}}(\phi,\eta)]_{n_{\text{an}}} = \exp(j2\pi(n_{\text{an}}-1)d[\phi_l - (n_{\text{an}}-1)d\eta_l]/\lambda)$ with $\lambda = c/f_{\text{c}}$, f_{c}, and c being the wavelength, carrier frequency, and speed of light, respectively, $\mathbf{b}(\tau) \in \mathbb{C}^{N_{\text{sc}}\times 1}$ and $\mathbf{c}(\nu) \in \mathbb{C}^{N_{\text{sym}}\times 1}$ denote delay and Doppler domain steering vectors, defined as $[\mathbf{b}(\tau)]_{n_{\text{sc}}} = \exp(-j2\pi(n_{\text{sc}}-1)\Delta\bar{f}\tau)$ and $[\mathbf{c}(\nu)]_{n_{\text{sym}}} = \exp(j2\pi(n_{\text{sym}}-1)\Delta\bar{T}\nu)$, respectively, and we also define the beam domain steering vector with SnS propagation effect as $\mathbf{a}(\phi,\eta,\mathbf{s}) \triangleq \mathbf{a}_{\text{SS}}(\phi,\eta) \odot \mathbf{s}$.

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

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \odot \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{Z}}, \tag{10}$$

where $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Z}}$ denote the pilot tensor and the additive white Gaussian noise at pilot segments, respectively. Since the pilot symbols are known at both BS and MTs, we assume that $\boldsymbol{\mathcal{X}}$ is an all-one tensor without loss of generality.

Remark 1. *Due to the inherent spatial consistency of wireless channels, we assume that the physical parameters of each*

path, including BDD domain grids and SnS factors, remain unchanged over the duration of interest.

B. Tucker-Based Representation

While the low-rank characterization of SFT domain channels enables the efficient representation within the tensor decomposition framework [35], [36], the practical lack of the uniqueness guarantee for this decomposition necessitates the incorporation of explicit physical knowledge as structural priors. To this end, we construct the factor matrices $\mathbf{A}(\boldsymbol{\phi},\boldsymbol{\eta},\mathbf{S}) \in \mathbb{C}^{N_{\text{an}}\times K_{\text{be}}}$, $\mathbf{B}(\boldsymbol{\tau}) \in \mathbb{C}^{N_{\text{sc}}\times K_{\text{de}}}$, and $\mathbf{C}(\boldsymbol{\nu}) \in \mathbb{C}^{N_{\text{sym}}\times K_{\text{do}}}$ of SFT domain channels based on the beam-delay-Doppler (BDD) domain grids and the corresponding SnS factors, defined as

$$\mathbf{A}(\boldsymbol{\phi},\boldsymbol{\eta},\mathbf{S}) = [\mathbf{a}(\phi_1,\eta_1,\mathbf{s}_1),\ldots,\mathbf{a}(\phi_{K_{\text{be}}},\eta_{K_{\text{be}}},\mathbf{s}_{K_{\text{be}}})], \quad (11a)$$

$$\mathbf{B}(\boldsymbol{\tau}) = [\mathbf{b}(\tau_1),\ldots,\mathbf{b}(\tau_{K_{\text{de}}})], \quad (11b)$$

$$\mathbf{C}(\boldsymbol{\nu}) = [\mathbf{c}(\nu_1),\ldots,\mathbf{c}(\nu_{K_{\text{do}}})], \quad (11c)$$

where K_{be}, K_{de}, and K_{do} denote the number of BDD domain grids, respectively, which are set to balance model expressiveness and complexity, $\boldsymbol{\phi} = [\phi_1,\ldots,\phi_{K_{\text{be}}}]^T$, $\boldsymbol{\eta} = [\eta_1,\ldots,\eta_{K_{\text{be}}}]^T$, $\boldsymbol{\tau} = [\tau_1,\ldots,\tau_{K_{\text{de}}}]^T$, and $\boldsymbol{\nu} = [\nu_1,\ldots,\nu_{K_{\text{do}}}]^T$ denote the environment-dependent dynamic grids in the BDD domain, $\mathbf{S} = [\mathbf{s}_1,\ldots,\mathbf{s}_{K_{\text{be}}}] \in \mathbb{C}^{N_{\text{an}}\times K_{\text{be}}}$, and $\mathbf{s}_{k_{\text{be}}}$ denotes the SnS factors corresponding to different beam domain grids.

With the aid of these factor matrices, the SFT domain channel is represented by the Tucker model, as given by

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\boldsymbol{\phi},\boldsymbol{\eta},\mathbf{S}) \times_2 \mathbf{B}(\boldsymbol{\tau}) \times_3 \mathbf{C}(\boldsymbol{\nu}), \quad (12)$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{C}^{K_{\text{be}}\times K_{\text{de}}\times K_{\text{do}}}$ denotes the BDD domain channel under these factor matrices. The factor matrices $\mathbf{A}(\boldsymbol{\phi},\boldsymbol{\eta},\mathbf{S})$, $\mathbf{B}(\boldsymbol{\tau})$, and $\mathbf{C}(\boldsymbol{\nu})$ represent the linear transformations from beam, delay, and Doppler domains to spatial, frequency, and temporal domains, respectively, which collectively constitute the multi-linear transformation from the BDD domain channel to the SFT domain channel.

Following this model, channel prediction is achieved by the acquisition of the BDD domain channel, along with its corresponding grids and SnS factors, from the received signal. In this framework, the Doppler domain features serve as the key predictors, since they allow an explicit and physically interpretable transformation into the temporal domain, thereby enabling the extrapolation of future channels over arbitrary prediction lengths. In parallel, the beam and delay domains encode the spatial and frequency domain correlations, respectively, where the former captures both SnS and NF propagation characteristics. By incorporating all these correlations into the multi-linear transformation, the proposed mechanism enhances the robustness of channel prediction even under noisy conditions or complex propagation environments.

Remark 2. *The prior work can be viewed as special cases of the proposed model under specific conditions. When we have $\mathbf{S} = \mathbf{1}_{N_{an}\times K_{be}}$, the representation degenerates to the spatial stationary model in [28] without delay variations. Moreover, the representation further degenerates to the model in the conventional massive MIMO systems for $\boldsymbol{\eta} = \mathbf{0}$, which is consistent with the temporal stationary case in [17]. By adjusting the parametrization of spatial domain factor matrices, the proposed model is also compatible with different NF propagation representations, in which the beam domain parameters are instantiated as angle-distance tuples [18], [20], coordinates [21], or spatial-chirp tuples [37].*

III. Problem Formulation

In the Tucker-based representations, the acquisition of BDD domain channels is critically dependent on prior knowledge of sparsity, yet it is typically unknown or costly to obtain. This limitation motivates a paradigm shift towards a Bayesian approach, which offers an elegant solution through automatic sparsity determination: sparsity is inferred directly from the data, instead of being required as a predefined input [38], [39]. To this end, we develop comprehensive probabilistic models for the observations, multi-linear transformations, BDD domain channels, and SnS factors, thus formulating the channel prediction task in a Bayesian context.

A. Probabilistic Model

1) Observation and Multi-linear Transformation: Based on the signal model in (10), the observation model is given as

$$\mathsf{P}(\boldsymbol{\mathcal{Y}} \,|\, \boldsymbol{\mathcal{H}}) \propto \mathsf{CN}(\boldsymbol{\mathcal{Y}}; \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{C}}(\sigma_z^2)). \quad (13)$$

To facilitate the inference of BDD domain channels, BDD domain grids, and SnS factors with given observations, it is essential to develop probabilistic models for multi-linear transformations defined in (12). However, the environment-dependent dynamic grids in the multi-linear transformation complicates this model, which motivates the alternative of fixed grids to dynamic ones for tractable Bayesian inference. To guarantee sufficient coverage of the BDD domain, we build fixed grids via uniform sampling, yet challenges persist due to the NF propagation effect in the beam domain. Specifically, the spatial-chirp characterization of the array response necessitates joint sampling over both angle and slope parameters, which significantly increases computational burden due to the enlarged spatial domain factor matrix. To alleviate this burden, we employ a sampling-hyperparameterization strategy that decouples angle and slope discretization: only angles are sampled in the beam domain, while slope parameters are treated as learnable hyperparameters. This approach avoids the combinatorial explosion of joint sampling, thereby boosting inference efficiency without compromising modeling flexibility, as the learnable slopes enable the representation to adaptively capture the underlying propagation geometry.

Despite fixed grids enabling tractable Bayesian inference, such discretization in the BDD domain inherently limits the ability to capture continuous physical parameters. Instead of employing densely sampled grids directly, we incorporate perturbation parameters into the coarsely sampled ones to capture deviations from the ground-truth physical parameters. Specifically, the perturbed parameters for angles, delays, and Doppler frequencies are expressed as $\boldsymbol{\phi} \triangleq \bar{\boldsymbol{\phi}} + \boldsymbol{\Delta\phi}$, $\boldsymbol{\tau} \triangleq \bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau}$, and $\boldsymbol{\nu} \triangleq \bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu}$, respectively, where $\bar{\boldsymbol{\phi}}$, $\bar{\boldsymbol{\tau}}$, and $\bar{\boldsymbol{\nu}}$ denote coarsely sampled grids, while $\boldsymbol{\Delta\phi}$, $\boldsymbol{\Delta\tau}$, and

$\boldsymbol{\Delta\nu}$ denote the corresponding perturbation parameters that refine the discretizations. With such sampling-then-refinement strategy, the PDF of multi-linear transformation is given as

$$\begin{aligned}\mathsf{P}(\boldsymbol{\mathcal{H}} \mid \boldsymbol{\mathcal{G}}, \mathbf{S}; \boldsymbol{\Delta\phi}, \boldsymbol{\Delta\tau}, \boldsymbol{\Delta\nu}, \boldsymbol{\eta}) \propto \\ \delta(\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}} + \boldsymbol{\Delta\phi}, \boldsymbol{\eta}, \mathbf{S}) \\ \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu})). \end{aligned} \tag{14}$$

An interesting perspective is that $\boldsymbol{\eta}$ can also be interpreted as perturbation parameters, since they quantify the deviation relative to the zero-slope reference grid (i.e., the far-field case). This further highlights the flexibility of the proposed strategy: both NF propagation and continuous-valued physical parameters can be captured through lightweight refinements instead of costly dense sampling.

2) BDD Domain Channels and SnS Factors: The low-rank property of the SFT domain channel corresponds directly to the sparsity of its BDD domain representation, since each dominant rank-1 component in the SFT domain maps to one significant propagation path characterized by the unique BDD domain grid. Therefore, the BDD domain channels are modeled by BG distribution, expressed as

$$\mathsf{P}(\boldsymbol{\mathcal{G}}; \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{V}}) \propto \mathsf{BG}(\boldsymbol{\mathcal{G}}; \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{C}}(0), \boldsymbol{\mathcal{V}}), \tag{15}$$

where $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{V}}$ denote the model hyperparameters, capturing the sparsity and power of BDD domain channels, respectively.

The SnS factor is captured by Bernoulli distributions as

$$\mathsf{P}(\mathbf{S}; \boldsymbol{\Gamma}) \propto \exp(\langle \boldsymbol{\Gamma}, \mathbf{S} \rangle), \tag{16}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{N_{\text{an}} \times K_{\text{be}}}$ denotes the hyperparameters, characterizing the strength of the SnS propagation. This model can be extended to structured probabilistic models to capture potential dependencies among SnS factors. It is also feasible to employ continuous-valued probabilistic models for non-binary SnS factors. In this work, we focus on the independent binary-valued case; nevertheless, the aforementioned extensions are fully compatible with our framework.

B. Channel Prediction Problem Formulation

Following these probabilistic models, we formulate the acquisition of BDD domain channels and SnS factors under the MMSE criterion, which can be expressed as the posterior expectations and given by

$$\hat{\boldsymbol{\mathcal{G}}} = \mathsf{E}\{\boldsymbol{\mathcal{G}}\}, \hat{\mathbf{S}} = \mathsf{E}\{\mathbf{S}\}, \tag{17}$$

where the expectations are taken with respect to the joint posterior PDF $\mathsf{P}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}}, \mathbf{S} \mid \boldsymbol{\mathcal{Y}}; \mathcal{P}_{\text{HP}})$, and the hyperparameters are collected as $\mathcal{P}_{\text{HP}} \triangleq \{\boldsymbol{\Delta\phi}, \boldsymbol{\Delta\tau}, \boldsymbol{\Delta\nu}, \boldsymbol{\eta}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\Gamma}\}$. Note that the estimation of BDD domain grids is converted into the corresponding hyperparameter learning through the simplification strategy of multilinear transformation models.

However, these posterior expectations require evaluating posterior PDFs that involve unknown hyperparameters, which are environment-dependent and not directly observable in practice. To address this, we adopt the EM framework to learn these hyperparameters and thereby adapt to diverse propagation environments [40]. Specifically, the inference iteratively alternates between the E-step and M-step, detailed as follows:

- **E-Step**: Given the current estimate of model hyperparameters, $\hat{\mathcal{P}}_{\text{HP}}$, from M-step, the posterior PDF $\mathsf{P}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}}, \mathbf{S} \mid \boldsymbol{\mathcal{Y}}; \hat{\mathcal{P}}_{\text{HP}})$ is computed by combining the observation and multi-linear transformation models, along with the BDD domain channel and SnS factor prior models.
- **M-Step**: Given the posterior PDF, $\mathsf{P}(\boldsymbol{\mathcal{G}} \mid \boldsymbol{\mathcal{Y}}; \hat{\mathcal{P}}_{\text{HP}})$, derived from the E-step, the model hyperparameters are refined through the maximization of the expected log-likelihood function, expressed as

$$\hat{\mathcal{P}}_{\text{HP}}^{(t_{\text{M}})} = \arg\max_{\mathcal{P}_{\text{HP}}} \mathsf{Q}(\mathcal{P}_{\text{HP}}, \hat{\mathcal{P}}_{\text{HP}}^{(t_{\text{M}}-1)}), \tag{18}$$

  where t_{M} denotes the M-step iteration index, and the Q-function is defined as

$$\mathsf{Q}(\mathcal{P}_{\text{HP}}, \hat{\mathcal{P}}_{\text{HP}}^{(t_{\text{M}}-1)}) \triangleq \mathsf{E}\{\ln(\mathsf{P}(\boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}}, \mathbf{S}; \mathcal{P}_{\text{HP}}))\}, \tag{19}$$

  where the above expectation is evaluated with respect to $\mathsf{P}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}}, \mathbf{S} \mid \boldsymbol{\mathcal{Y}}; \hat{\mathcal{P}}_{\text{HP}}^{(t_{\text{M}}-1)})$ given previous hyperparameter estimations. The M-step iteration index is omitted hereafter to simplify notations. Unless stated otherwise, hyperparameters on the right-hand side of any equation represent values from the previous iteration, while those on the left correspond to the updated ones.

In this framework, the environment-dependent hyperparameters are automatically inferred from observations, including the sparsity and power of BDD domain channels, the strengths of SnS and NF propagation effects, and the perturbation parameters in the BDD domain. The framework remains compatible when a subset of hyperparameters is available from external sources, such as historical statistics [41] or channel knowledge maps [42], as these hyperparameters will be treated as known and excluded from the M-step updates.

IV. Tensor-Structured Bi-Layer Inference for Bayesian Channel Prediction

While the EM framework facilitates MMSE-based channel prediction with unknown hyperparameters, it remains computationally demanding due to the posterior PDF evaluation and hyperparameter learning. To alleviate this, we streamline the former in the E-step based on the factor graph principle, and incorporate approximate inference techniques to enhance computational efficiency. In the M-step, hyperparameters are learned via an alternating optimization strategy to avoid joint search in large parameter spaces, which collectively constitutes the TS-BLI algorithm with the E-step for channel prediction.

A. E-Step: Bi-Layer Factor Graph Representation

In the E-step, the development of approximate inference techniques is challenged by its underlying affine matrix factorization form: the unknown SnS factors introduce bilinear mixing in the spatial domain, while the frequency and temporal domains preserve the linear mixing of conventional massive MIMO systems. The unique bilinear structure in the spatial domain leads to extra randomness in the factor matrices beyond the BDD domain channel, which motivates the design of tractable and interpretable Bayesian inference algorithms.

To address this challenge, we introduce the spatial-delay-Doppler (SDD) domain channel to bridge between the SFT and

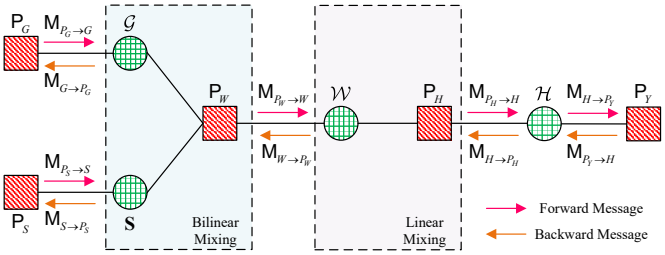


Fig. 4. Factor graph representation for E-step, where red squares and green circles represent factor and variable nodes, respectively.

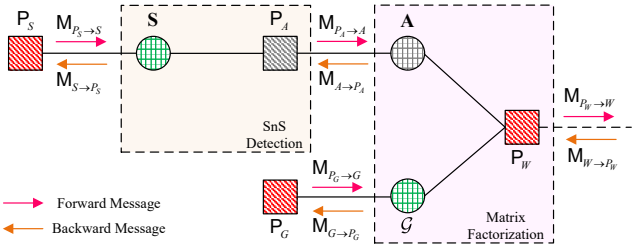


Fig. 5. Equivalent factor graph in the bilinear mixing module, where the gray square and circle represent auxiliary factor and variable nodes, respectively.

BDD domain channels, allowing only linear mixing between the SDD and SFT domain channels, while bilinear mixing is incorporated between the SDD and BDD domain channels. This bridge enables a progressive inference trajectory, starting from the SFT domain, advancing through the SDD domain, and ultimately reaching the BDD domain, where each stage involves only one type of mixing: either linear or bilinear.

As such, we comes to the evaluation of the augmented posterior PDF with the SDD domain channel $\boldsymbol{\mathcal{W}} \in \mathbb{C}^{N_{\text{an}} \times K_{\text{de}} \times K_{\text{do}}}$, which can be factorized by

$$\begin{aligned}\mathsf{P}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{G}}, \mathbf{S} \,|\, \boldsymbol{\mathcal{Y}}; \hat{\mathcal{P}}_{\text{HP}}) \propto \mathsf{P}(\boldsymbol{\mathcal{Y}} \,|\, \boldsymbol{\mathcal{H}}) \mathsf{P}(\boldsymbol{\mathcal{H}} \,|\, \boldsymbol{\mathcal{W}}; \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}, \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}) \\ \mathsf{P}(\boldsymbol{\mathcal{W}} \,|\, \boldsymbol{\mathcal{G}}, \mathbf{S}; \hat{\boldsymbol{\Delta}}\phi, \hat{\eta}) \mathsf{P}(\boldsymbol{\mathcal{G}}; \hat{\boldsymbol{\mathcal{M}}}, \hat{\boldsymbol{\mathcal{V}}}) \mathsf{P}(\mathbf{S}; \hat{\boldsymbol{\Gamma}}),\end{aligned} \tag{20}$$

where the factorized probabilistic models above are denoted, in order, as P_Y, P_H, P_W, P_G, and P_S, respectively, and the newly introduced probabilistic models are specified by

$$\begin{aligned}&\mathsf{P}(\boldsymbol{\mathcal{H}} \,|\, \boldsymbol{\mathcal{W}}; \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}, \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}) \propto \\ &\qquad \delta(\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{W}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu})), \qquad \text{(21a)} \\ &\mathsf{P}(\boldsymbol{\mathcal{W}} \,|\, \boldsymbol{\mathcal{G}}, \mathbf{S}; \hat{\boldsymbol{\Delta}}\phi, \hat{\eta}) \propto \delta(\boldsymbol{\mathcal{W}} - \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\bar{\phi} + \hat{\boldsymbol{\Delta}}\phi, \hat{\eta}, \mathbf{S})). \qquad \text{(21b)}\end{aligned}$$

With these probabilistic models, the equivalence of the MMSE estimators stems from the fact that the posterior PDF required in the E-step can be obtained by marginalizing its augmented counterpart, expressed by

$$\mathsf{P}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}}, \mathbf{S} \mid \boldsymbol{\mathcal{Y}}; \hat{\mathcal{P}}_{\text{HP}}) \propto \int_{\boldsymbol{\mathcal{W}}} \mathsf{P}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{G}}, \mathbf{S} \mid \boldsymbol{\mathcal{Y}}; \hat{\mathcal{P}}_{\text{HP}}). \tag{22}$$

Following the factorization of the augmented posterior PDF, the bi-layer factor graph representation employed in the E-step is illustrated in Fig. 4, where scalar nodes are merged to improve visual clarity. The graph is partitioned into two subgraphs corresponding to the linear and bilinear mixing parts, respectively, each leveraging externally defined equivalent prior and likelihood models to facilitate inference [43]. Within this factor graph, we follow the sum-product rules, with messages passed iteratively forward and backward between the observations and BDD domain channels until convergence.

However, exact inference remains computationally intensive due to the high-dimensional integrations involved in posterior expectation computations and the dependencies between messages along the edges. To mitigate this burden, we employ approximate inference techniques within each layer, based on the central limit theorem and Taylor series, thereby enabling message representations using Gaussian parameterizations and eliminates edge dependencies. It is worth noting that practical systems are finite in size, which may introduce approximation bias and consequently degrade inference performance. While rigorous analytical results for finite-sized systems with general factor matrices remain an open problem, empirical simulations confirm that the asymptotic Gaussian approximation adopted in this work remains effective.

1) Linear Mixing Module: In this module, the message $\mathsf{M}_{P_W \to W}$ and $\mathsf{M}_{H \to P_H}$ serve as equivalent prior and likelihood models of the linear mixing model, respectively, to develop an equivalent signal model, which is expressed as

$$\hat{\boldsymbol{\mathcal{H}}}_{\text{lik}} = \boldsymbol{\mathcal{W}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}) + \boldsymbol{\mathcal{Z}}_{H,\text{lik}}, \tag{23}$$

where $\hat{\boldsymbol{\mathcal{H}}}_{\text{lik}} = \boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Z}}_{H,\text{lik}} = \boldsymbol{\mathcal{Z}}$ denote the equivalent observation and noise of this module, respectively, $\boldsymbol{\mathcal{Z}}_{H,\text{lik}}$ follows an independent Gaussian distribution with zero mean and identical variance σ_z^2. By matricizing this model along the frequency and temporal domains, we obtain the linear model with multiple measurement vectors, given by

$$\hat{\mathbf{H}}_{\text{lik}} = (\mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}) \otimes \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}))\mathbf{W} + \mathbf{Z}_{H,\text{lik}}, \tag{24}$$

where we define $\hat{\mathbf{H}}_{\text{lik}} \in \mathbb{C}^{N_{\text{sym}} N_{\text{sc}} \times N_{\text{an}}}$, $\mathbf{W} \in \mathbb{C}^{K_{\text{do}} K_{\text{de}} \times N_{\text{an}}}$, and $\mathbf{Z}_{H,\text{lik}} \in \mathbb{C}^{N_{\text{sym}} N_{\text{sc}} \times N_{\text{an}}}$ as the matrixization of $\hat{\boldsymbol{\mathcal{H}}}_{\text{lik}}$, $\boldsymbol{\mathcal{W}}$, and $\boldsymbol{\mathcal{Z}}_{H,\text{lik}}$ with respect to the frequency and temporal domains, respectively. This implies that the spatial domain observations are treated as independent, allowing this module to focus on the linear mixing part of this multi-linear transformation.

When the number of pilot subcarriers and OFDM symbols tends to infinity, the key input and output messages can be approximated as Gaussian distributions, given by

$$\mathsf{M}_{P_H \to H} \propto \mathsf{CN}(\boldsymbol{\mathcal{H}}; \hat{\boldsymbol{\mathcal{H}}}_{\text{pri}}, \boldsymbol{\mathcal{E}}_{H,\text{pri}}), \tag{25a}$$

$$\mathsf{M}_{W \to P_W} \propto \mathsf{CN}(\boldsymbol{\mathcal{W}}; \hat{\boldsymbol{\mathcal{W}}}_{\text{lik}}, \boldsymbol{\mathcal{E}}_{W,\text{lik}}), \tag{25b}$$

where the parameters in $\mathsf{M}_{P_H \to H}$ and $\mathsf{M}_{W \to P_W}$ are computed from Line 2 to Line 8 of Algorithm 1, following the general ideas of the generalized approximate message passing (GAMP) procedure [44].

2) Bilinear Mixing Module: Similar to the linear mixing module, the message $\mathsf{M}_{W \to P_W}$, interpreted as the equivalent likelihood, is combined with the prior models for the BDD domain channels and SnS factors, denoted by $\mathsf{M}_{P_G \to G}$ and $\mathsf{M}_{P_S \to S}$, respectively, to develop the equivalent signal model in this module, given by

$$\hat{\boldsymbol{\mathcal{W}}}_{\text{lik}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\bar{\phi} + \hat{\boldsymbol{\Delta}}\phi, \hat{\eta}, \mathbf{S}) + \boldsymbol{\mathcal{Z}}_{W,\text{lik}}, \tag{26}$$

where $\boldsymbol{\mathcal{Z}}_{W,\text{lik}}$ denotes the equivalent noise of this module, assumed to follow an independent Gaussian distribution with zero mean and variance $\boldsymbol{\mathcal{E}}_{W,\text{lik}}$. The matrixization along the spatial domain of this model can be expressed as

$$\hat{\mathbf{W}}_{\text{lik}} = \mathbf{A}(\bar{\boldsymbol{\phi}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}, \mathbf{S})\mathbf{G} + \mathbf{Z}_{W,\text{lik}}, \tag{27}$$

where we define $\hat{\mathbf{W}}_{\text{lik}} \in \mathbb{C}^{N_{\text{an}} \times K_{\text{de}} K_{\text{do}}}$, $\mathbf{G} \in \mathbb{C}^{K_{\text{be}} \times K_{\text{de}} K_{\text{do}}}$, and $\mathbf{Z}_{W,\text{lik}} \in \mathbb{C}^{N_{\text{an}} \times K_{\text{de}} K_{\text{do}}}$ as the matrixization of $\hat{\boldsymbol{\mathcal{W}}}_{\text{lik}}$, $\boldsymbol{\mathcal{G}}$, and $\boldsymbol{\mathcal{Z}}_{W,\text{lik}}$ with respect to the spatial domain, respectively.

This module builds upon the structured bilinear formulation, which incorporates the fact that the unknown spatial domain factor matrices is parametrized by SnS factors. To address such structure, we split it into two sub-modules: matrix factorization (MF) and SnS detection (SnSD), where latter provides the prior model of spatial domain factor matrices and the former returns the likelihood model for the SnS factor detection.

Specifically, the parameterized structure of spatial domain factor matrices is temporarily disregarded in the MF sub-module, resulting in the simplified bilinear model consistent with the standard formulation. From the factor graph perspective, this can be interpreted as augmenting the model with a new set of auxiliary variable and factor nodes corresponding to the spatial domain factor matrices, as shown in Fig. 5. When the number of antennas tends to infinity, the Gaussian approximation of input and output messages are given by

$$\mathsf{M}_{P_W \to W} \propto \mathsf{CN}(\boldsymbol{\mathcal{W}}; \hat{\boldsymbol{\mathcal{W}}}_{\text{pri}}, \boldsymbol{\mathcal{E}}_{W,\text{pri}}), \tag{28a}$$

$$\mathsf{M}_{G \to P_G} \propto \mathsf{CN}(\boldsymbol{\mathcal{G}}; \hat{\boldsymbol{\mathcal{G}}}_{\text{lik}}, \boldsymbol{\mathcal{E}}_{G,\text{lik}}), \tag{28b}$$

$$\mathsf{M}_{A \to P_A} \propto \mathsf{CN}(\mathbf{A}; \hat{\mathbf{A}}_{\text{lik}}, \boldsymbol{\Sigma}_{A,\text{lik}}), \tag{28c}$$

where the paramters in $\mathsf{M}_{P_W \to W}$, $\mathsf{M}_{G \to P_G}$, and $\mathsf{M}_{A \to P_A}$ are computed from Line 9 to Line 19 of Algorithm 1, which follows the bilinear GAMP (BiGAMP) procedure [45].

In the SnSD sub-module, the inference begins with the deterministic parameterized and component-wise structure of spatial domain factor matrices, where the message $\mathsf{M}_{P_A \to A}$ is characterized by the Dirac Delta function, given by

$$\mathsf{M}_{P_A \to A} \propto \delta(\mathbf{A} - \mathbf{A}_{\text{SS}} \odot \mathbf{S}), \tag{29}$$

where $\mathbf{A}_{\text{SS}} \in \mathbb{C}^{N_{\text{an}} \times K_{\text{be}}}$ denotes the spatial stationary spatial domain factor matrix with the hyperparameters $\hat{\boldsymbol{\Delta}}\boldsymbol{\phi}$ and $\hat{\boldsymbol{\eta}}$ from M-step, defnied by

$$\mathbf{A}_{\text{SS}} = [\mathbf{a}_{\text{SS}}(\bar{\phi}_1 + \hat{\Delta}\phi_1, \hat{\eta}_1), \ldots, \mathbf{a}_{\text{SS}}(\bar{\phi}_{K_{\text{be}}} + \hat{\Delta}\phi_{K_{\text{be}}}, \hat{\eta}_{K_{\text{be}}})]. \tag{30}$$

3) Closing the Loop: To close the loop, we start with the missing messages in this bi-layer factor graph. Specifically, the messages $\mathsf{M}_{P_G \to G}$, $\mathsf{M}_{P_S \to S}$, and $\mathsf{M}_{P_Y \to H}$ are determined by the BDD domain channel prior model, the SnS factor prior model, and the observation model, respectively.

The posterior estimation of hidden variables involved in messages also plays a critical role. Following the sum-product rules, the posterior PDFs of SFT, SDD, and BDD domain channels can be directly obtained by

$$\mathsf{b}_H \propto \mathsf{M}_{P_H \to H} \mathsf{M}_{H \to P_H}, \tag{31a}$$

$$\mathsf{b}_W \propto \mathsf{M}_{P_W \to W} \mathsf{M}_{W \to P_W}, \tag{31b}$$

$$\mathsf{b}_G \propto \mathsf{M}_{P_G \to G} \mathsf{M}_{G \to P_G}, \tag{31c}$$

Algorithm 1 Single Iteration of E-Step

Input: External Information $(\boldsymbol{\mathcal{Y}}, \sigma_z^2)$, hyperparameters estimated from M-step ($\hat{\mathcal{P}}_{\text{HP}}$), and intermediate results from the previous E-step.

Output: Intermediate results to the next E- and M-step.

1: **for** $t_{\text{E}} = 1, \ldots, T_{\text{E}}$ **do**

// Linear Mixing Module

2: $\quad \boldsymbol{\mathcal{E}}_{H,\text{pri}} = \boldsymbol{\mathcal{E}}_{W,\text{post}} \times_2 |\mathbf{B}|^{\odot 2} \times_3 |\mathbf{C}|^{\odot 2}$

3: $\quad \hat{\boldsymbol{\mathcal{H}}}_{\text{pri}} = \hat{\boldsymbol{\mathcal{W}}} \times_2 \mathbf{B} \times_3 \mathbf{C} - \hat{\boldsymbol{\mathcal{H}}}_{\text{res}} \odot \boldsymbol{\mathcal{E}}_{H,\text{pri}}$

4: $\quad$ Obtain the posterior mean $\hat{\boldsymbol{\mathcal{H}}}$ and variance $\boldsymbol{\mathcal{E}}_{H,\text{post}}$ of $\boldsymbol{\mathcal{H}}$ based on (31a).

5: $\quad \boldsymbol{\mathcal{E}}_{H,\text{res}} = (\boldsymbol{\mathcal{E}}_{H,\text{post}} - \boldsymbol{\mathcal{E}}_{H,\text{pri}}) \oslash (\boldsymbol{\mathcal{E}}_{H,\text{pri}})^{\odot 2}$

6: $\quad \hat{\boldsymbol{\mathcal{H}}}_{\text{res}} = (\hat{\boldsymbol{\mathcal{H}}} - \hat{\boldsymbol{\mathcal{H}}}_{\text{pri}}) \oslash \boldsymbol{\mathcal{E}}_{H,\text{pri}}$

7: $\quad \boldsymbol{\mathcal{E}}_{W,\text{lik}} = (\boldsymbol{\mathcal{E}}_{H,\text{res}} \times_2 |\mathbf{B}^H|^{\odot 2} \times_3 |\mathbf{C}^H|^{\odot 2})^{\odot -1}$

8: $\quad \hat{\boldsymbol{\mathcal{W}}}_{\text{lik}} = \hat{\boldsymbol{\mathcal{W}}} + \boldsymbol{\mathcal{E}}_{W,\text{lik}} \odot (\boldsymbol{\mathcal{H}}_{\text{res}} \times_2 \mathbf{B}^H \times_3 \mathbf{C}^H)$

// Bilinear Mixing Module

9: $\quad \boldsymbol{\mathcal{E}}_{W,\text{plug-in}} = \boldsymbol{\mathcal{E}}_{G,\text{post}} \times_1 |\hat{\mathbf{A}}|^{\odot 2} + |\hat{\boldsymbol{\mathcal{G}}}|^{\odot 2} \times_1 \boldsymbol{\Sigma}_{A,\text{post}}$

10: $\quad \hat{\boldsymbol{\mathcal{W}}}_{\text{pri}} = \hat{\boldsymbol{\mathcal{G}}} \times_1 \hat{\mathbf{A}} - \hat{\boldsymbol{\mathcal{W}}}_{\text{res}} \odot \boldsymbol{\mathcal{E}}_{W,\text{plug-in}}$

11: $\quad \boldsymbol{\mathcal{E}}_{W,\text{pri}} = \boldsymbol{\mathcal{E}}_{W,\text{plug-in}} + \boldsymbol{\mathcal{E}}_{G,\text{pri}} \times_1 \boldsymbol{\Sigma}_{A,\text{post}}$

12: $\quad$ Obtain the posterior mean $\hat{\boldsymbol{\mathcal{W}}}$ and variance $\boldsymbol{\mathcal{E}}_{W,\text{post}}$ of $\boldsymbol{\mathcal{W}}$ based on (31b).

13: $\quad \boldsymbol{\mathcal{E}}_{W,\text{res}} = (\boldsymbol{\mathcal{E}}_{W,\text{post}} - \boldsymbol{\mathcal{E}}_{W,\text{pri}}) \oslash (\boldsymbol{\mathcal{E}}_{W,\text{pri}})^{\odot 2}$

14: $\quad \hat{\boldsymbol{\mathcal{W}}}_{\text{res}} = (\hat{\boldsymbol{\mathcal{W}}} - \hat{\boldsymbol{\mathcal{W}}}_{\text{pri}}) \oslash \boldsymbol{\mathcal{E}}_{W,\text{pri}}$

15: $\quad \boldsymbol{\mathcal{E}}_{G,\text{lik}} = (\boldsymbol{\mathcal{E}}_{W,\text{res}} \times_1 |\hat{\mathbf{A}}^H|^{\odot 2})^{\odot -1}$

16: $\quad \hat{\boldsymbol{\mathcal{G}}}_{\text{lik}} = \hat{\boldsymbol{\mathcal{G}}} - \hat{\boldsymbol{\mathcal{G}}} \odot \boldsymbol{\mathcal{E}}_{G,\text{lik}} \odot (\boldsymbol{\mathcal{E}}_{W,\text{res}} \times_1 \boldsymbol{\Sigma}_{A,\text{post}}) + \boldsymbol{\mathcal{E}}_{G,\text{lik}} \odot (\hat{\boldsymbol{\mathcal{W}}}_{\text{res}} \times_1 \hat{\mathbf{A}}^H)$

17: $\quad$ Obtain the posterior mean $\hat{\boldsymbol{\mathcal{G}}}$ and variance $\boldsymbol{\mathcal{E}}_{G,\text{post}}$ of $\boldsymbol{\mathcal{G}}$ based on (31c).

18: $\quad \boldsymbol{\Sigma}_{A,\text{lik}} = (\boldsymbol{\mathcal{E}}_{W,\text{res}} \times_{-1} |\hat{\boldsymbol{\mathcal{G}}}|^{\odot 2})^{\odot -1}$

19: $\quad \hat{\mathbf{A}}_{\text{lik}} = \hat{\mathbf{A}} - \hat{\mathbf{A}} \odot \boldsymbol{\Sigma}_{A,\text{lik}} \odot (\boldsymbol{\mathcal{E}}_{W,\text{res}} \times_{-1} \boldsymbol{\mathcal{E}}_{G,\text{post}}) + \boldsymbol{\Sigma}_{A,\text{lik}} \odot (\hat{\boldsymbol{\mathcal{W}}}_{\text{res}} \times_{-1} \hat{\boldsymbol{\mathcal{G}}}^*)$

20: $\quad$ Obtain the posterior mean $\hat{\mathbf{A}}$ and variance $\boldsymbol{\Sigma}_{A,\text{post}}$ of $\mathbf{A}$ based on (33) and (34).

21: **end for**

where $\mathsf{M}_{H \to P_H} = \mathsf{M}_{P_Y \to H}$, and the posterior mean and variance are computed over the corresponding PDFs.

For spatial domain factor matrices, the equivalent likelihood model of SnS factors is specified by the message $\mathsf{M}_{A \to P_A}$ from the MF sub-module and channel model, expressed as

$$\hat{\mathbf{A}}_{\text{lik}} = \mathbf{A}_{\text{SS}} \odot \mathbf{S} + \mathbf{Z}_{A,\text{lik}}, \tag{32}$$

where $\mathbf{Z}_{A,\text{lik}}$ denotes the equivalent noise with variance $\boldsymbol{\Sigma}_{A,\text{lik}}$. Therefore, the posterior PDF is given by

$$\mathsf{P}(\mathbf{S} \,|\, \hat{\mathbf{A}}_{\text{lik}}; \hat{\boldsymbol{\Gamma}}) \propto \exp(\langle \hat{\boldsymbol{\Gamma}}, \mathbf{S} \rangle) \mathsf{CN}(\hat{\mathbf{A}}_{\text{lik}}; \mathbf{A}_{\text{SS}} \odot \mathbf{S}, \boldsymbol{\Sigma}_{A,\text{lik}}), \tag{33}$$

Using the above posterior PDF, the posterior mean and variance of spatial domain factor matrices can be given by

$$\hat{\mathbf{A}} = \mathbf{A}_{\text{SS}} \odot \hat{\mathbf{S}}, \tag{34a}$$

$$\boldsymbol{\Sigma}_{A,\text{post}} = |\mathbf{A}_{\text{SS}}|^{\odot 2} \odot \hat{\mathbf{S}} \odot (1 - \hat{\mathbf{S}}), \tag{34b}$$

where $\hat{\mathbf{S}}$ denotes the posterior expectation of $\mathbf{S}$.

In summary, the message passing process is summarized in Algorithm 1 as one E-step, where $\hat{\mathbf{A}} \triangleq \mathbf{A}(\bar{\boldsymbol{\phi}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{S}})$, $\mathbf{B} \triangleq \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau})$, and $\mathbf{C} \triangleq \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu})$ denote shorthand

representations of factor matrices under estimated SnS factors and hyperparameters. To guarantee convergence under finite-sized systems and ill-conditioned factor matrices, damping is adopted by updating the posterior estimates as weighted combinations of the current and previous ones.

B. M-Step: Model Hyperparameter Learning

In the M-step, we employ an alternating optimization strategy to reduce the hyperparameter search space. Specifically, we start with perturbation and slope parameters in the factor matrices, with updating rules decoupled as

$$\{\hat{\boldsymbol{\Delta\tau}}, \hat{\boldsymbol{\Delta\nu}}\} = \underset{\boldsymbol{\Delta\tau},\boldsymbol{\Delta\nu}}{\arg\max} \underbrace{\mathsf{E}\{\ln(\mathsf{P}(\boldsymbol{\mathcal{H}} \,|\, \boldsymbol{\mathcal{W}}; \boldsymbol{\Delta\tau}, \boldsymbol{\Delta\nu}))\}}_{\triangleq\, -J_{\tau,\nu}(\boldsymbol{\Delta\tau},\boldsymbol{\Delta\nu})}, \quad (35a)$$

$$\{\hat{\boldsymbol{\Delta\phi}}, \hat{\boldsymbol{\eta}}\} = \underset{\boldsymbol{\Delta\phi},\boldsymbol{\eta}}{\arg\max} \underbrace{\mathsf{E}\{\ln(\mathsf{P}(\boldsymbol{\mathcal{W}} \,|\, \boldsymbol{\mathcal{G}}, \mathbf{S}; \boldsymbol{\Delta\phi}, \boldsymbol{\eta}))\}}_{\triangleq\, -J_{\phi,\eta}(\boldsymbol{\Delta\phi},\boldsymbol{\eta})}, \quad (35b)$$

where the expectation is taken over the posterior PDFs from the E-step, and such decoupling inherently results from the factorized forms in (20). The objective functions can be further simplified as the negative residual energy of SFT and SDD domain channels, as summarized in Proposition 1.

Proposition 1. *The negative objective functions $J_{\tau,\nu}$ and $J_{\phi,\eta}$ are given by*

$$J_{\tau,\nu} = \|\hat{\boldsymbol{\mathcal{H}}} - \hat{\boldsymbol{\mathcal{W}}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu})\|_F^2, \quad (36a)$$

$$J_{\phi,\eta} = \|\hat{\boldsymbol{\mathcal{W}}} - \hat{\boldsymbol{\mathcal{G}}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}} + \boldsymbol{\Delta\phi}, \boldsymbol{\eta}, \hat{\mathbf{S}})\|_F^2, \quad (36b)$$

respectively, where $\hat{\boldsymbol{\mathcal{H}}}$, $\hat{\boldsymbol{\mathcal{W}}}$, $\hat{\boldsymbol{\mathcal{G}}}$, and $\hat{\mathbf{S}}$ denote the posterior estimates obtained from the previous E-step.

Proof. By approximating the Dirac Delta function in (35a) with the sharply peaked Gaussian PDF, the objective function of linear mixing is given by [17]

$$J_{\tau,\nu} = \mathsf{E}\{\|\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{W}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu})\|_F^2\}, \quad (37)$$

With the posterior independence derived from the E-step, the objective function is further decomposed into $J_{\tau,\nu} = J_{\tau,\nu,1} + J_{\tau,\nu,2}$, where sub-functions are defined as

$$J_{\tau,\nu,1} \triangleq \|\hat{\boldsymbol{\mathcal{H}}} - \hat{\boldsymbol{\mathcal{W}}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu})\|_F^2, \quad (38a)$$

$$J_{\tau,\nu,2} \triangleq \|\boldsymbol{\mathcal{E}}_{W,\text{post}} \times_2 |\mathbf{B}(\bar{\boldsymbol{\tau}} + \boldsymbol{\Delta\tau})|^{\odot 2} \times_3 |\mathbf{C}(\bar{\boldsymbol{\nu}} + \boldsymbol{\Delta\nu})|^{\odot 2}\|_1. \quad (38b)$$

Since the magnitudes of the factor matrices are, by definition, independent of $\boldsymbol{\Delta\tau}$ and $\boldsymbol{\Delta\nu}$, the sub-function $J_{\tau,\nu,2}$ is also constant with respect to these variables. Therefore, the objective function for hyperparameter learning corresponding to linear mixing reduces to (36a), and the bilinear mixing case can be derived analogously as (36b). □

By leveraging the first-order Taylor series of factor matrices, these objective functions can be further approximated as quadratic forms and summarized in Proposition 2.

Proposition 2. *The negative objective functions in* (36) *are approximated as the following quadratic forms, given by*

$$J_{\phi,\eta}(\boldsymbol{\chi}) \approx \boldsymbol{\chi}^T \boldsymbol{\Pi}_{\phi,\eta} \boldsymbol{\chi} - 2\mathsf{Re}\{\boldsymbol{\mu}_{\phi,\eta}^T\}\boldsymbol{\chi} + C_{\phi,\eta}, \quad (39a)$$

$$J_{\tau}(\boldsymbol{\Delta\tau}) \approx \boldsymbol{\Delta\tau}^T \boldsymbol{\Pi}_{\tau} \boldsymbol{\Delta\tau} - 2\mathsf{Re}\{\boldsymbol{\mu}_{\tau}^T\}\boldsymbol{\Delta\tau} + C_{\tau}, \quad (39b)$$

$$J_{\nu}(\boldsymbol{\Delta\nu}) \approx \boldsymbol{\Delta\nu}^T \boldsymbol{\Pi}_{\nu} \boldsymbol{\Delta\nu} - 2\mathsf{Re}\{\boldsymbol{\mu}_{\nu}^T\}\boldsymbol{\Delta\nu} + C_{\nu}, \quad (39c)$$

where $\boldsymbol{\chi} \triangleq [\boldsymbol{\Delta\phi}^T, \boldsymbol{\eta}^T]^T$ denotes beam domain perturbation and slope parameters, $C_{\phi,\eta}$, C_{τ}, and C_{ν} denote constant terms, the quadratic and linear coefficients are given by

$$\boldsymbol{\Pi}_{\phi,\eta} = (\dot{\mathbf{A}}^H(\bar{\boldsymbol{\phi}}, \mathbf{0}, \hat{\mathbf{S}})\dot{\mathbf{A}}(\bar{\boldsymbol{\phi}}, \mathbf{0}, \hat{\mathbf{S}}))^* \odot (\mathbf{1}_{2\times 2} \otimes (\hat{\boldsymbol{\mathcal{G}}} \times_{-1} \hat{\boldsymbol{\mathcal{G}}}^*)), \quad (40a)$$

$$\boldsymbol{\mu}_{\phi,\eta} = \sum_n \mathsf{diag}^H\{\mathbf{1}_{2\times 1} \otimes \hat{\mathbf{g}}_n\}\dot{\mathbf{A}}^H(\bar{\boldsymbol{\phi}}, \mathbf{0}, \hat{\mathbf{S}})\hat{\mathbf{r}}_{\phi,\eta,n}, \quad (40b)$$

$$\boldsymbol{\Pi}_{\tau} = (\dot{\mathbf{B}}^H(\bar{\boldsymbol{\tau}})\dot{\mathbf{B}}(\bar{\boldsymbol{\tau}}))^* \odot (\hat{\boldsymbol{\mathcal{W}}}_{\tau} \times_{-2} \hat{\boldsymbol{\mathcal{W}}}_{\tau}^*), \quad (40c)$$

$$\boldsymbol{\mu}_{\tau} = \sum_n \mathsf{diag}^H\{\hat{\mathbf{w}}_{\tau,n}\}\dot{\mathbf{B}}^H(\bar{\boldsymbol{\tau}})\hat{\mathbf{r}}_{\tau,n}, \quad (40d)$$

$$\boldsymbol{\Pi}_{\nu} = (\dot{\mathbf{C}}^H(\bar{\boldsymbol{\nu}})\dot{\mathbf{C}}(\bar{\boldsymbol{\nu}}))^* \odot (\hat{\boldsymbol{\mathcal{W}}}_{\nu} \times_{-3} \hat{\boldsymbol{\mathcal{W}}}_{\nu}^*), \quad (40e)$$

$$\boldsymbol{\mu}_{\nu} = \sum_n \mathsf{diag}^H\{\hat{\mathbf{w}}_{\nu,n}\}\dot{\mathbf{C}}^H(\bar{\boldsymbol{\nu}})\hat{\mathbf{r}}_{\nu,n}, \quad (40f)$$

where $\dot{\mathbf{A}}(\cdot,\cdot,\cdot) \triangleq [\dot{\mathbf{A}}_{\phi}(\cdot,\cdot,\cdot), \dot{\mathbf{A}}_{\eta}(\cdot,\cdot,\cdot)]$ denotes the aggregate derivative matrix, $\dot{\mathbf{A}}_{\phi}(\cdot,\cdot,\cdot)$ and $\dot{\mathbf{A}}_{\eta}(\cdot,\cdot,\cdot)$ denote the first-order derivatives of $\mathbf{A}(\cdot,\cdot,\cdot)$ with respect to the first and second arguments, respectively, $\dot{\mathbf{B}}(\cdot)$ and $\dot{\mathbf{C}}(\cdot)$ denote the first-order derivatives of $\mathbf{B}(\cdot)$ and $\mathbf{C}(\cdot)$, respectively, $\hat{\mathbf{g}}_n$, $\hat{\mathbf{w}}_{\tau,n}$ and $\hat{\mathbf{w}}_{\nu,n}$ denotes the n-th fibers of beam-delay-Doppler domain channel $\hat{\boldsymbol{\mathcal{G}}}$, spatial-delay-temporal domain channel $\hat{\boldsymbol{\mathcal{W}}}_{\tau} \triangleq \hat{\boldsymbol{\mathcal{W}}} \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta\nu}})$ and spatial-frequency-Doppler domain channel $\hat{\boldsymbol{\mathcal{W}}}_{\nu} \triangleq \hat{\boldsymbol{\mathcal{W}}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta\tau}})$, respectively, $\hat{\mathbf{r}}_{\phi,\eta,n}$, $\hat{\mathbf{r}}_{\tau,n}$, and $\hat{\mathbf{r}}_{\nu,n}$ denote the n-th fibers of residual channels $\hat{\boldsymbol{\mathcal{R}}}_{\phi,\eta}$, $\hat{\boldsymbol{\mathcal{R}}}_{\tau}$, and $\hat{\boldsymbol{\mathcal{R}}}_{\nu}$ along spatial, frequency, and temporal domains, respectively, with corresponding residual channels defined as

$$\hat{\boldsymbol{\mathcal{R}}}_{\phi,\eta} = \hat{\boldsymbol{\mathcal{W}}} - \hat{\boldsymbol{\mathcal{G}}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}}, \mathbf{0}, \hat{\mathbf{S}}), \quad (41a)$$

$$\hat{\boldsymbol{\mathcal{R}}}_{\tau} = \hat{\boldsymbol{\mathcal{H}}} - \hat{\boldsymbol{\mathcal{W}}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta\nu}}), \quad (41b)$$

$$\hat{\boldsymbol{\mathcal{R}}}_{\nu} = \hat{\boldsymbol{\mathcal{H}}} - \hat{\boldsymbol{\mathcal{W}}} \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta\tau}}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}}). \quad (41c)$$

Proof. The proof follows a similar approach to our previous work [17, Appendix C]. □

Based on the approximation of objective functions, closed-form learning rules for the hyperparameters can be derived. Alternatively, the Taylor series can be expanded around the previous learning result, enabling the reuse of intermediate tensors in E-step to avoid redundant computations.

When it comes to the SnS factor prior model, we have

$$\hat{\boldsymbol{\Gamma}} = \underset{\boldsymbol{\Gamma}}{\arg\max}\, \mathsf{E}\{\ln \mathsf{P}(\mathbf{S}; \boldsymbol{\Gamma})\}. \quad (43)$$

By taking the derivative of $\boldsymbol{\Gamma}$ and setting it to zero, the learning rules are given by

$$\hat{\boldsymbol{\Gamma}} = \hat{\mathbf{S}} \oslash (1 + \hat{\mathbf{S}}). \quad (44)$$

Since the hyperparamter learning rules in the BDD domain prior channel model are consistent with the techniques in

$$[\boldsymbol{\mathcal{R}}_G]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}} = \frac{[\hat{\boldsymbol{\mathcal{M}}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}}\mathsf{CN}([\hat{\boldsymbol{\mathcal{G}}}_{\text{lik}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}};0,[\hat{\boldsymbol{\mathcal{V}}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}}+[\boldsymbol{\mathcal{E}}_{G,\text{lik}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}})}{(1-[\hat{\boldsymbol{\mathcal{M}}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}})\mathsf{CN}([\hat{\boldsymbol{\mathcal{G}}}_{\text{lik}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}};0,[\boldsymbol{\mathcal{E}}_{G,\text{lik}}]_{k_{\text{be}},k_{\text{de}},k_{\text{do}}})}. \tag{42}$$

[46], the detailed derivations are omitted due to the space limitations, with the results summarized as

$$\hat{\boldsymbol{\mathcal{M}}} = \boldsymbol{\mathcal{R}}_G \oslash (\boldsymbol{\mathcal{C}}(1) + \boldsymbol{\mathcal{R}}_G), \tag{45a}$$

$$\hat{\boldsymbol{\mathcal{V}}} = (\boldsymbol{\mathcal{E}}_{G,\text{post}} + |\hat{\boldsymbol{\mathcal{G}}}|^{\odot 2}) \oslash \hat{\boldsymbol{\mathcal{M}}}, \tag{45b}$$

where $\boldsymbol{\mathcal{R}}_G$ is defined in (42) at the top of the next page.

C. Algorithm Description and Discussion

1) Algorithm Discussion: The proposed TS-BLI algorithm for channel prediction is summarized as Algorithm 2. Specifically, the channel prediction is achieved by the transformation from Doppler domain to temporal domain, formulated as

$$\begin{aligned}\boldsymbol{\mathcal{H}}^{\text{CP}} = \hat{\boldsymbol{\mathcal{G}}} \times_1 \mathbf{A}(\bar{\phi} + \hat{\boldsymbol{\Delta}}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{S}}) \\ \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\tau}) \times_3 \tilde{\mathbf{C}}(\bar{\boldsymbol{\nu}} + \hat{\boldsymbol{\Delta}}\boldsymbol{\nu}),\end{aligned} \tag{46}$$

where $\tilde{\mathbf{C}}(\boldsymbol{\nu}) = [\tilde{\mathbf{c}}(\nu_1), \ldots, \tilde{\mathbf{c}}(\nu_{K_{\text{do}}})]$ 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 $[\tilde{\mathbf{c}}(\nu)]_{n_{\text{cp}}} = \exp(j2\pi(T_0 + n_{\text{cp}}\Delta T)\nu)$, $T_0 = (N_{\text{sym}} - 1)\Delta\bar{T}$ and n_{cp} denote the prediction origin and prediction length, respectively. At the beginning of each sliding frame, as illustrated in Fig. 2, the BS collects the received SRSs, which constitute the observation tensor $\boldsymbol{\mathcal{Y}}$, together with the noise variance σ_z^2, as the input of Algorithm 2. The algorithm then infers the physical parameters and predicts the channels $\hat{\boldsymbol{\mathcal{H}}}^{\text{CP}}$ for all subsequent OFDM symbols until the next SRS appears.

The prefix "tensor-structured" of Algorithm 2 stems from the inherent multi-linear structure of channels, which allows for the natural decoupling of the linear and bilinear mixing, as well as the simplification of hyperparameter learning in the multi-linear transformations. This structure not only supports bi-layer inference across different types of mixing, but also improves computational efficiency via tensor operations.

Remark 3. *An alternative interpretation of EM arises from the variational free energy perspective [47], where it is regarded as a special case of message passing under an uninformative Dirac delta prior. While not explicitly employed in this work, it offers useful intuition for the message schedule design. Specifically, the E-step need not fully converge before proceeding to the M-step, thereby alleviating the computational burden when hyperparameters remain unreliable in early iterations.*

2) Discussion: While we focus on the ULA in this work for clarity, the proposed framework can be generalized to other antenna configurations. For other fixed antenna systems, it functions as a "plug-and-play" solution, requiring only the adaptation of the spatial domain factor matrix to the new array manifold under the spatial stationarity assumption. In the case of two-dimensional arrays, such as the uniform planar array (UPA), the observations from the second spatial domain can be modeled as independent measurements, which offers a favorable trade-off between computational complexity and prediction performance. The adaptability of the proposed framework also extends to flexible antenna systems [48]–[50], where the non-fixed antenna positions introduce an alternative parametric form for spatial domain factor matrices. In such systems, our algorithm can be adapted by repurposing the submodule within the bilinear mixing layer, originally designed for SnS detection, to infer the antenna positions.

Algorithm 2 Tensor-Structured Bi-Layer Inference

Input: Observation tensor $\boldsymbol{\mathcal{Y}}$ and noise variance σ_z^2.
Output: Channel Prediction result $\hat{\boldsymbol{\mathcal{H}}}^{\text{CP}}$.

1: Initialize $\hat{\boldsymbol{\mathcal{M}}}$, $\hat{\boldsymbol{\mathcal{V}}}$, $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\mathcal{G}}}$, $\hat{\mathbf{S}}$, and all hyperparameters.
2: **for** $t_{\text{M}} = 1, \ldots, T_{\text{M}}$ **do**
3: Execute Algorithm 1.
4: Learn hyperparameters based on Proposition 1, Proposition 2, (44), and (45).
5: **end for**
6: Predict the channel based on (46).

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 XL-MIMO-OFDM channels consistent with the third Generation Partnership Program (3GPP) specifications [51] and has been validated in various field trials [52]. To generate near-practical channels under SnS propagation, we augment the 3GPP-compliant QuaDRiGa channel generation procedure by incorporating the statistical characteristics and instantaneous values of visibility as large-scale parameters (LSPs) and small-scale parameters (SSPs), respectively. Subsequently, SnS factors are applied at a cluster-level granularity to synthesize the final channel coefficients. This approach is well-aligned with recent 3GPP standardization discussions and is also consistent with industrial measurements [53], [54]. Under the SnS propagation effect, we further assume that each antenna element has an identical and independent visibility probability of 0.5 on each path.

Unless specified otherwise, we consider the 3GPP urban macro (UMa) non-line-of-sight (NLOS) scenarios, with the simulation configurations in Section II and parameters in Table I. With these system configurations and the normalized maximum amplitude variation $\zeta = 0.99$, the approximations adopted in NF propagation modeling remains valid.

2) Benchmarks and Performance Metric: To demonstrate the superiority of the proposed algorithm, we select the following algorithms as benchmarks:

- **VKF** [55]: Estimates spatial domain channels by the least square (LS) algorithm, with temporal correlations captured by the AR-based vector Kalman filter.

TABLE I
SCENARIO PARAMETERS

Parameter	Value
Carrier Frequency	$f_{\mathrm{c}} = 15$ GHz
OFDM Symbol Duration	$\Delta T_{\mathrm{sym}} = 16.67$ μs
Cyclic Prefix Duration	$\Delta T_{\mathrm{cp}} = 1.17$ μs
Subcarrier Spacing	$\Delta f = 60$ kHz
Number of Pilot Interval Symbols	$N_{\mathrm{IS}} = 14$
Number of Transmission Combs	$N_{\mathrm{TC}} = 4$
Number of BS Antennas	$N_{\mathrm{an}} = 128$
Number of Pilot Subcarriers	$N_{\mathrm{sc}} = 128$
Number of Pilot Symbols	$N_{\mathrm{sym}} = 10$
Minimum MT Distribution Radius	$r_{\min} = 10$ m
Velocity of MTs	$v_{\mathrm{MT}} = 60$ km/h

- **FIT** [56]: Estimates beam-delay domain channels by the alternating LS (ALS) algorithm, with temporal correlations captured by the first-order Taylor series.
- **PAD** [10], **WTMP** [28]: Extract dominant beam-delay domain taps by orthogonal matching pursuit (OMP) algorithm, with temporal correlations captured by Prony and matrix pencil methods for these taps.

Among these benchmarks, **VKF** and **FIT** are independent of specific beam domain structures, allowing them to be compatible with XL-MIMO systems. Besides, **PAD** serves as the representative state-of-the-art channel prediction algorithms for massive MIMO systems, while **WTMP** further incorporates the NF propagation into the beam domain through the WT matrices. For benchmarks that only predict channels for the future pilot symbols, we predict channels on the non-pilot symbols through MMSE interpolation, with prior knowledge of the maximum Doppler frequency and transmission power. The normalized mean square error (NMSE) of SFT domain channels is adopted as the performance metric, defined by

$$\mathrm{NMSE} = \frac{\|\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}} - \boldsymbol{\mathcal{H}}^{\mathrm{CP}}\|_F^2}{\|\boldsymbol{\mathcal{H}}^{\mathrm{CP}}\|_F^2}, \tag{47}$$

where $\boldsymbol{\mathcal{H}}^{\mathrm{CP}}$ and $\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}}$ denote the groud-truth and predicted SFT domain channels of upcoming OFDM symbols, respectively.

B. Computational Complexity

1) Preliminaries: Before delving into the computational complexity analysis, we first introduce some preliminaries regarding the computational cost of tensor operations, including tensor-matrix multiplication and tensor-tensor multiplication. Specifically, for the mode-d tensor-matrix multiplication between 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 operation is implemented based on tensor matricization and follows 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 $\mathcal{O}(K_d N_1 \ldots N_D)$. Besides, for tensor-tensor multiplication of tensors $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{N_1 \times \ldots \times N_d \times \ldots \times N_D}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{N_1 \times \ldots \times M_d \times \ldots \times N_D}$ along the d-th mode, the operation also relies on the tensor matricization and follows the equivalence $\mathbf{Z} = \mathbf{X}_d \mathbf{Y}_d^T$. This implies that the computational complexity of tensor-tensor multiplication along the d-th modal is equivalent to the matrix multiplication of matrices $\mathbf{X}_d$ and $\mathbf{Y}_d$, yielding the computational complexity of $\mathcal{O}(M_d N_1 \ldots N_D)$. 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 now analyze the computational complexity of the proposed algorithms, which arises primarily from tensor-matrix multiplications, tensor-tensor multiplications, and scalar arithmetic operations. The first two operations constitute the main computational burden in the linear mixing module, the matrix factorization of the bilinear mixing module, and the perturbation parameter learning of the multi-linear transformation model, while the scalar operations dominate the SnS detection and the hyperparameter learning in the prior models of the SnS factors and the BDD domain channel.

To highlight the computational savings of tensor operations and to allow an execution-order–independent comparison, we assume that the size of BDD domain, SDD domain, and SFT domain channels are of the same order, which typically holds in practical systems. Under this assumption, the computational complexities of tensor operations in the spatial, frequency, and temporal domains are given by $\mathcal{O}(N_{\mathrm{an}}^2 N_{\mathrm{sc}} N_{\mathrm{sym}})$, $\mathcal{O}(N_{\mathrm{an}} N_{\mathrm{sc}}^2 N_{\mathrm{sym}})$, and $\mathcal{O}(N_{\mathrm{an}} N_{\mathrm{sc}} N_{\mathrm{sym}}^2)$. Regarding scalar arithmetic operations, the computational complexity scales linearly with the size of BDD domain, SDD domain, and SFT domain channels, as well as SnS factors, leading to the order of $\mathcal{O}(N_{\mathrm{an}} N_{\mathrm{sc}} N_{\mathrm{sym}})$ and $\mathcal{O}(N_{\mathrm{an}}^2)$, respectively. These costs are significantly smaller than those of tensor-matrix multiplications and tensor-tensor multiplications. Therefore, the computational complexity per iteration of the proposed algorithm is on the order of $\mathcal{O}(N_{\mathrm{an}} N_{\mathrm{sc}} N_{\mathrm{sym}} \bar{N})$, where $\bar{N} \triangleq N_{\mathrm{an}} + N_{\mathrm{sc}} + N_{\mathrm{sym}}$ denotes the aggregate dimension of the SFT domain channel. In contrast, when matrix-vector multiplications are employed without leveraging the separable structure of the factor matrices, the computational complexity per iteration will increase to the order of $\mathcal{O}(N_{\mathrm{an}}^2 N_{\mathrm{sc}}^2 N_{\mathrm{sym}}^2)$.

In the benchmarks, the computational complexity of both **PAD** and **WTMP** is dominated by the OMP algorithm, which can also exploit the separable structure of spatial-frequency domain factor matrices. Specifically, the computational complexity scales as $\mathcal{O}(T N_{\mathrm{an}} N_{\mathrm{sc}} N_{\mathrm{sym}} (N_{\mathrm{an}} F_{\mathrm{NF}} + N_{\mathrm{sc}} + T))$, with $F_{\mathrm{NF}} = 1$ for **PAD** and $F_{\mathrm{NF}} > 1$ for **WTMP** due to increased beam domain sampling induced by NF propagation. For **VKF**, the critical step is the learning of AR parameters shared across all pilot subcarriers, yielding the computational complexity of $\mathcal{O}(P^3 N_{\mathrm{an}}^3)$, where P denotes the AR order, typically on the same order as the pilot OFDM symbols. In the case of **FIT**, its computational complexity is controlled by the ALS algorithm with $\mathcal{O}(TRN_{\mathrm{an}} N_{\mathrm{sc}} N_{\mathrm{sym}})$, where T denotes the number of iterations and R denotes the predefined tensor rank, typically set as the maximum number of possible paths in the given environment. The computational complexity of the proposed algorithm and benchmarks are summarized in Table II.

3) FLOPs: Beyond the asymptotic computational complexity order, the number of floating-point operations (FLOPs)

TABLE II
COMPUTATIONAL COMPLEXITY

Algorithm	Asymptotic Order	FLOPs		
		$N_{\text{an}} = 128$	$N_{\text{an}} = 256$	$N_{\text{an}} = 512$
This work (per iteration)	$\mathcal{O}(N_{\text{an}} N_{\text{sc}} N_{\text{sym}} \bar{N})$	3.45×10^8	1.21×10^9	4.49×10^9
VKF	$\mathcal{O}(N_{\text{an}}^3 P^3)$	4.08×10^9	3.26×10^{10}	2.61×10^{11}
FIT	$\mathcal{O}(TRN_{\text{an}} N_{\text{sc}} N_{\text{sym}})$	6.55×10^9	1.31×10^{10}	2.62×10^{10}
PAD	$\mathcal{O}(TN_{\text{an}} N_{\text{sc}} N_{\text{sym}} (N_{\text{an}} + N_{\text{sc}} + T))$	1.39×10^{10}	3.00×10^{10}	6.93×10^{10}
WTMP	$\mathcal{O}(TN_{\text{an}} N_{\text{sc}} N_{\text{sym}} (N_{\text{an}} F_{\text{NF}} + N_{\text{sc}} + T))$	2.45×10^{10}	7.25×10^{10}	2.40×10^{11}

offers a more direct and practical measure of computational cost, and thus serves as another critical metric for algorithm efficiency. Under the scenario parameters detailed in Table I, we adopt nominal resolution with uniform sampling in the beam and delay domains, leading to $K_{\text{be}} = N_{\text{an}}$, and $K_{\text{de}} = N_{\text{sc}}(N_{\text{TC}} \Delta T_{\text{cp}} / \Delta T_{\text{sym}})$, where delay samples beyond the CP-limited range are discarded. In the Doppler domain, slight oversampling is applied to improve channel prediction accuracy, with $K_{\text{do}} = 2N_{\text{sym}}$. Notably, the multi-mode tensor-matrix multiplication required by the transformation from the SFT domain to the SDD domain channels are executed sequentially across the delay (frequency) and Doppler (temporal) domains, while the inverse transformation, i.e., from SDD to SFT domains, is executed in the reverse order. Given these configurations, the per-iteration FLOP count for the proposed algorithm is approximately 3.45×10^8. For the benchmarks, the FLOP counts of benchmark algorithms including **VKF**, **FIT**, **PAD**, and **MPAD** are approximately 4.08×10^9, 6.55×10^9, 1.39×10^{10}, and 2.45×10^{10}, respectively. Specifically, the AR order P in **VKF** is set to 9, which keeps the same order as the available pilot OFDM symbols N_{sym} within the sliding window. The tensor rank R and the number of iterations T for **FIT** are adaptively tuned according to the propagation environments, averaging around 100 and 50, respectively. The number of iteration T for PAD and **WTMP** depends on the number of identifiable paths in the environment, analogous to the tensor rank R in **FIT** and also averages around 100. To evaluate the scalability of computational cost with respect to the antenna array size, we report the FLOP counts for enlarged arrays in Table II. The results, aligned with the asymptotic orders, confirm that the proposed algorithm maintains favorable scalability, highlighting the computational advantage for large-scale array settings in the upper mid-band systems.

C. Performance Evaluation

To comprehensively evaluate the adaptability and performance of the proposed algorithm and benchmarks, we present numerical simulations under two distinct propagation conditions: one with and one without the SnS propagation effect. For these simulations, neither the proposed algorithm nor the benchmarks are provided with any prior knowledge regarding the presence of SnS propagation, thereby ensuring a fair and practical performance evaluation.

1) Convergence Behavior: To evaluate the convergence behavior of the proposed algorithm, we present the NMSE versus

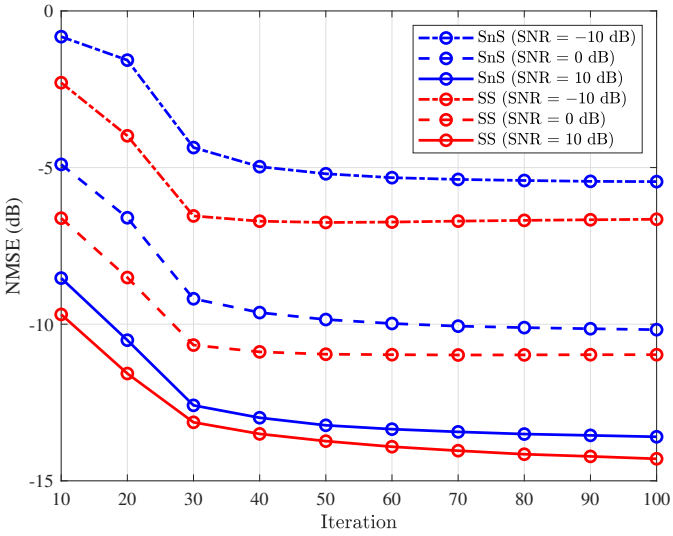


Fig. 6. The convergence performance of the proposed algorithm.

the number of iterations under different signal-to-noise ratio (SNR) levels in Fig. 6. It is noteworthy that approximately 30 iterations are sufficient to achieve nearly the full performance gains across various SNRs, while additional iterations result in only marginal refinements. Due to the rapid decrease in NMSE during the initial iterations, early termination can be adopted to further reduce computational complexity, depending on the performance requirements of practical systems. A key trend, first observed here and consistently confirmed in subsequent simulations, is that the channel prediction performance is nearly identical in scenarios with and without SnS propagation, with a slight advantage for the latter. This finding demonstrates the robustness of the proposed algorithm by incorporating SnS propagation and its statistical properties into the approximate inference and EM framework. It also suggests that the marginal performance gain arises from the simpler and more tractable beam-domain channel representation without SnS propagation effects.

2) NMSE versus SNR: The NMSE of the proposed algorithm and benchmarks versus SNR is shown in Fig. 7. Since **FIT** and **VKF** are independent of beam domain structure, they exhibit nearly identical channel prediction performance with and without SnS propagation. However, the absence of beam domain modeling results in the poorest channel prediction performance, while the **FIT** algorithm further suffers from substantial degradation in low-SNR regime. Owing to the incorporation of beam and Doppler domain modeling, both

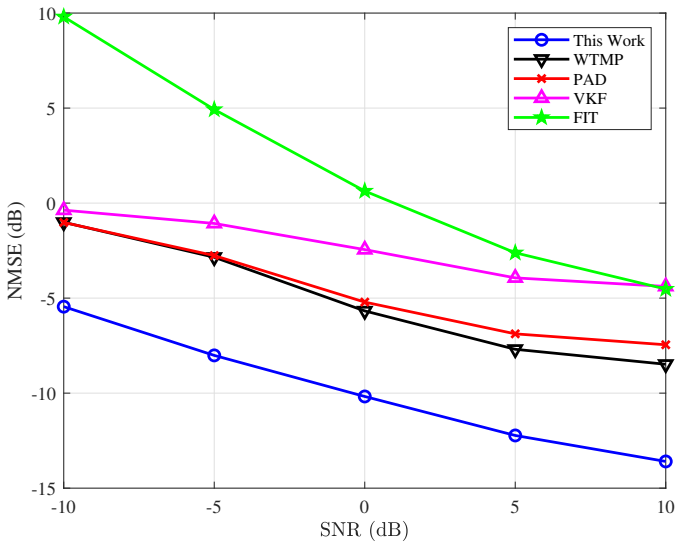


Fig. 7. NMSE of channel prediction versus SNR: (a) with SnS propagation effect, (b) without SnS propagation effect.

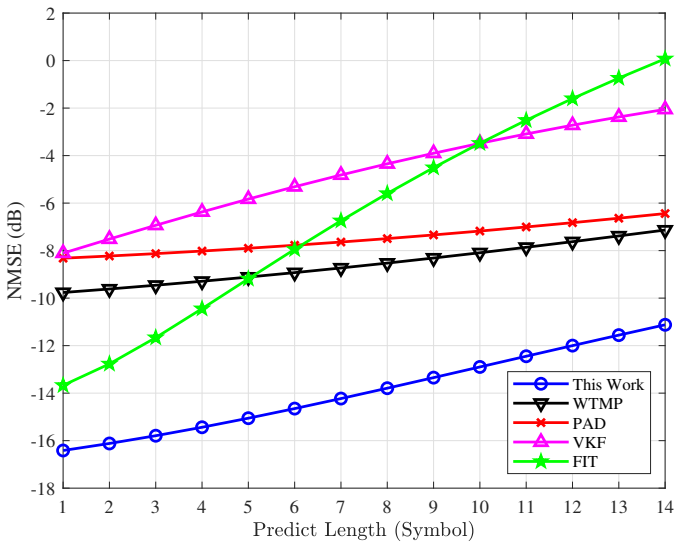


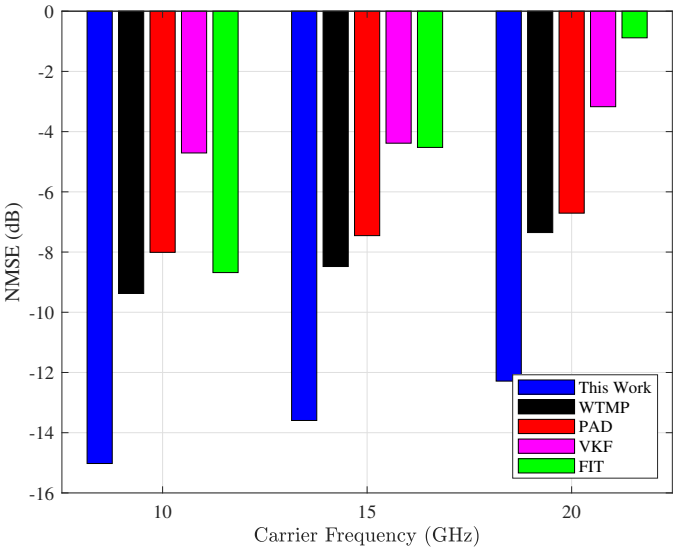
Fig. 8. NMSE of channel prediction versus prediction length at SNR $= 10$ dB: (a) with SnS propagation effect, (b) without SnS propagation effect.

PAD and **WTMP** achieve superior performance among the benchmarks, where **WTMP** benefits from spherical wave modeling and thus outperforms **PAD**. Under SnS propagation, both **PAD** and **WTMP** exhibit the performance degradation of over 2 dB compared to the case without SnS propagation at an SNR of 10 dB, indicating that SnS propagation disrupts the beam domain representation. The proposed algorithm significantly outperforms all the benchmarks, highlighting the performance gains brought by Doppler domain and SnS propagation modeling in channel prediction. For instance, at an SNR of 10 dB, it achieves at least 5 dB gain in the presence of SnS propagation and 2 dB in its absence, with even more substantial benefits observed in the low-SNR regime.

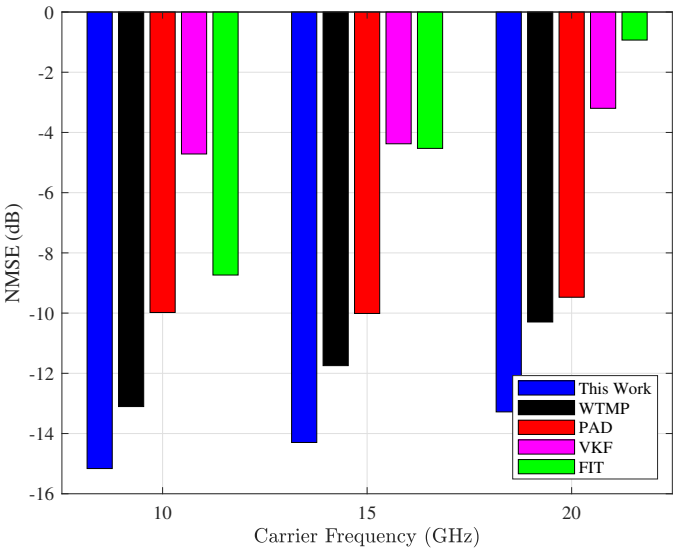
3) NMSE versus Prediction Length: To illustrate the variation of channel prediction performance with prediction length, we present the NMSEs for symbols between the current pilot symbols and the first future pilot symbols in Fig. 8. As expected, the performance of both the proposed algorithm and the benchmarks degrades as the prediction length increases, consistent with prior studies on channel prediction. Under SnS propagation, **FIT** outperforms all other benchmarks at short prediction lengths due to the adoption of ALS algorithms, which not only capture the dominant channel components but also naturally accommodate the presence of SnS. However, the performance of **FIT** degrades sharply by approximately 14 dB from the first non-pilot OFDM symbol to the next pilot OFDM symbol, owing to the temporal domain factor matrix extrapolation based on the first-order Taylor series. This approximation increasingly deviates from the non-linear temporal variations, leading to cumulative modeling errors and amplified variance over larger prediction lengths. In contrast, **VKF**, which employs AR-based temporal correlation modeling, achieves better long-term predictions than **FIT**, though it still exhibits about 6 dB degradation as prediction length increases. Both **PAD**

and **WTMP** demonstrate improved stability across prediction lengths, indicating that the Prony and matrix pencil methods offer more accurate temporal correlation modeling than both the first-order Taylor series and AR models. In comparison, the proposed algorithm consistently outperforms all benchmark methods across all prediction lengths. For the first future non-pilot and pilot symbols, it achieves NMSEs below -16 dB and -11 dB, respectively, with SnS propagation effects, and approximately -18 dB and -11.5 dB without them.

4) NMSE versus Carrier Frequency: Along with the typical carrier frequency of 15 GHz, we also evaluate the channel prediction performance of the proposed algorithm at both higher (20 GHz) and lower (10 GHz) frequencies within the upper mid-band, as shown in Fig. 9. Due to the linear dependence of Doppler frequency on carrier frequency, all benchmarks as well as the proposed algorithm suffer from performance degradation as the carrier frequency grows, while the severe performance drop of approximately 9 dB observed for **FIT** further highlights its limited capability in capturing temporal correlations. In contrast, the other benchmarks and the proposed algorithm maintain relatively stable performance over these carrier frequencies. An interesting observation is that the performance gap between **PAD** and **WTMP** without SnS propagation shrinks from over 3 dB to approximately 0.5 dB as the carrier frequency increases. This is due to the reduction in Rayleigh distance at higher carrier frequencies when the number of antennas remains fixed, which weakens NF propagation and consequently reduces the performance difference between **PAD** and **WTMP**. Across all carrier frequencies, the proposed algorithm consistently delivers the best and most robust channel prediction performance, demonstrating its effectiveness for upper mid-band systems.



(a)



(b)

Fig. 9. NMSE of channel prediction versus carrier frequencies at SNR $= 10$ dB: (a) with SnS propagation effect, (b) without SnS propagation effect.

VI. Conclusion

This paper investigated tensor-structured Bayesian channel prediction for upper mid-band XL-MIMO systems, explicitly addressing the challenges posed by NF and SnS propagation. By developing multi-linear SFT domain channel models and probabilistic models with perturbation-aware BDD sampling, the proposed method effectively captures the inherent sparsity of channels and facilitates physics-inspired channel prediction. Following the probabilistic models, we formulated the channel prediction under the MMSE criterion with unknown model hyperparameters, and developed the TS-BLI algorithm within the EM framework for channel prediction. In this algorithm, the multi-linear structure of channels allows for the bi-layer factor graph representation and facilitates tensor operations that reduce computational complexity. Simulation results demonstrate that the proposed algorithm significantly outperforms existing benchmarks in channel prediction accuracy, underscoring its potential for practical deployment in future large-scale wireless communication systems.

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] M. Na, J. Lee, G. Choi, T. Yu, J. Choi, J. Lee, and S. Bahk, "Operator's perspective on 6G: 6G services, vision, and spectrum," *IEEE Commun. Mag.*, vol. 62, no. 8, pp. 178–184, Aug. 2024.

[3] 3rd Generation Partnership Project (3GPP), "Study on the 7 to 24 GHz frequency range for NR," 3GPP, Technical Specification (TS) 38.820, 2020. [Online]. Available: https://www.3gpp.org

[4] S. Kang, M. Mezzavilla, S. Rangan, A. Madanayake, S. B. Venkatakrishnan, G. Hellbourg, M. Ghosh, H. Rahmani, and A. Dhananjay, "Cellular wireless networks in the upper mid-band," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 2058–2075, Mar. 2024.

[5] J. Tian, Y. Han, X. Li, S. Jin, and C.-K. Wen, "Mid-band extra large-scale MIMO system: Channel modeling and performance analysis," *IEEE Trans. Commun.*, vol. 73, no. 2, pp. 1025–1041, Feb. 2025.

[6] H. Miao, J. Zhang, P. Tang, L. Tian, X. Zhao, B. Guo, and G. Liu, "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. 2023.

[7] Z. Yuan, J. Zhang, Y. Ji, G. F. Pedersen, and W. Fan, "Spatial non-stationary near-field channel modeling and validation for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 71, no. 1, pp. 921–933, Jan. 2023.

[8] 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.

[9] 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.

[10] H. Yin, H. Wang, Y. Liu, and D. Gesbert, "Addressing the curse of mobility in massive MIMO with prony-based angular-delay domain channel predictions," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2903–2917, Dec. 2020.

[11] Z. Qin, H. Yin, Y. Cao, W. Li, and D. Gesbert, "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.

[12] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive MIMO: From spatio-temporal autoregression to deep learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1915–1930, Jul. 2021.

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

[14] Y. Wan, G. Liu, A. Liu, and M.-J. Zhao, "Robust multi-user channel tracking scheme for 5G new radio," *IEEE Trans. Wireless Commun.*, 2023.

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

[16] D. Shi, L. Song, W. Zhou, X. Gao, C.-X. Wang, and G. Y. Li, "Channel acquisition for HF skywave massive MIMO-OFDM communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 4074–4089, Jun. 2023.

[17] H. Hou, Y. Wang, Y. Zhu, X. Yi, W. Wang, D. T. M. Slock, and S. Jin, "A tensor-structured approach to dynamic channel prediction for massive MIMO systems with temporal non-stationarity," *IEEE Trans. Wireless Commun.*, 2025.

[18] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, Jan. 2022.

[19] Y. Lu and L. Dai, "Near-field channel estimation in mixed LoS/NLoS environments for extremely large-scale MIMO systems," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3694–3707, Jun. 2023.

[20] Y. Chen, Y. Wang, Z. Wang, and Z. Han, "Angular-distance based channel estimation for holographic MIMO," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 6, pp. 1684–1702, Jun. 2024.

[21] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 633–637, May 2020.

[22] Y. Chen and L. Dai, "Non-stationary channel estimation for extremely large-scale MIMO," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7683–7697, Jul. 2024.

[23] S. Yang, P. An, P. Yang, X. Cao, D. O. Wu, and T. Q. Quek, "Adaptive subarray segmentation: A new paradigm of spatial non-stationary near-field channel estimation for XL-MIMO systems," *arXiv preprint arXiv:2503.04211*, 2025.

[24] Y. Zhu, H. Guo, and V. K. Lau, "Bayesian channel estimation in multi-user massive MIMO with extremely large antenna array," *IEEE Trans. Signal Process.*, vol. 69, pp. 5463–5478, Sep. 2021.

[25] A. Tang, J.-B. Wang, Y. Pan, W. Zhang, X. Zhang, Y. Chen, H. Yu, and R. C. De Lamare, "Joint visibility region and channel estimation for extremely large-scale MIMO systems," *IEEE Trans. Commun.*, vol. 72, no. 10, pp. 6087–6101, Oct. 2024.

[26] W. Xu, A. Liu, M.-j. Zhao, and G. Caire, "Joint visibility region detection and channel estimation for XL-MIMO systems via alternating MAP," *IEEE Trans. Signal Process.*, vol. 72, pp. 4827–4842, Oct. 2024.

[27] A. Tang, J.-B. Wang, Y. Pan, W. Zhang, Y. Chen, H. Yu, and R. C. de Lamare, "Spatially non-stationary XL-MIMO channel estimation: A three-layer generalized approximate message passing method," *IEEE Trans. Signal Process.*, vol. 73, pp. 356–371, Dec. 2024.

[28] W. Li, H. Yin, Z. Qin, and M. Debbah, "Wavefront transformation-based near-field channel prediction for extremely large antenna array with mobility," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 15 613–15 626, Oct. 2024.

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

[30] M. Brazell, N. Li, C. Navasca, and C. Tamon, "Solving multilinear systems via tensor inversion," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 2, pp. 542–570, 2013.

[31] 3rd Generation Partnership Project (3GPP), "NR; physical channels and modulation," 3GPP, Technical Specification (TS) 38.211, 2025. [Online]. Available: https://www.3gpp.org

[32] J. Sherman, "Properties of focused apertures in the Fresnel region," *IRE Trans. Antennas Propag.*, vol. 10, no. 4, pp. 399–408, Jul. 1962.

[33] R1-2406199, "Views on channel model adaptation/extension of TR38.901 for 7-24GHz," Vivo, 3GPP TSG RAN WG1 Meeting 118, 2024.

[34] K. T. Selvan and R. Janaswamy, "Fraunhofer and Fresnel distances: Unified derivation for aperture antennas," *IEEE Antennas Propag. Mag.*, vol. 59, no. 4, pp. 12–15, Aug. 2017.

[35] J. Du, M. Han, Y. Chen, L. Jin, H. Wu, and F. Gao, "Tensor decompositions for integrated sensing and communications," *IEEE Commun. Mag.*, vol. 62, no. 9, pp. 128–134, Sep. 2024.

[36] Y. Wang, H. Hou, X. Yi, W. Wang, and S. Jin, "Towards unified AI models for MU-MIMO communications: A tensor equivariance framework," *arXiv preprint arXiv:2406.09022*, 2024.

[37] X. Shi, J. Wang, Z. Sun, and J. Song, "Spatial-chirp codebook-based hierarchical beam training for extremely large-scale massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 2824–2838, Apr. 2024.

[38] L. Cheng, Z. Chen, Q. Shi, Y.-C. Wu, and S. Theodoridis, "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.

[39] X. Shao, L. Cheng, X. Chen, C. Huang, and D. W. K. Ng, "Reconfigurable intelligent surface-aided 6G massive access: Coupled tensor modeling and sparse Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10 145–10 161, Dec. 2022.

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

[41] Y. Wang, V. N. Ha, K. Ntontin, H. Yan, W. Wang, S. Chatzinotas, and B. Ottersten, "Statistical CSI-based distributed precoding design for OFDM-cooperative multi-satellite systems," *arXiv preprint arXiv:2505.08038*, 2025.

[42] X. Wang, Y. Shi, T. Wang, Y. Huang, Z. Hu, L. Chen, and Z. Jiang, "Channel knowledge map aided channel prediction with measurements-based evaluation," *IEEE Trans. Commun.*, vol. 73, no. 5, pp. 3622–3636, 2025.

[43] Q. Zou, H. Zhang, and H. Yang, "Multi-layer bilinear generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 69, pp. 4529–4543, Jul. 2021.

[44] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*. IEEE, Jul. 2011, pp. 2168–2172.

[45] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[46] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[47] D. Zhang, X. Song, W. Wang, G. Fettweis, and X. Gao, "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.

[48] L. Zhu, W. Ma, W. Mei, Y. Zeng, Q. Wu, B. Ning, Z. Xiao, X. Shao, J. Zhang, and R. Zhang, "A tutorial on movable antennas for wireless networks," *IEEE Commun. Surveys Tuts.*, 2025.

[49] W. K. New, K.-K. Wong, H. Xu, C. Wang, F. R. Ghadi, J. Zhang, J. Rao, R. Murch, P. Ramírez-Espinosa, D. Morales-Jimenez *et al.*, "A tutorial on fluid antenna system for 6G networks: Encompassing communication theory, optimization methods and hardware designs," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 4, pp. 2325–2377, Aug. 2025.

[50] Y. Liu, Z. Wang, X. Mu, C. Ouyang, X. Xu, and Z. Ding, "Pinching-antenna systems: Architecture designs, opportunities, and outlook," *IEEE Commun. Mag.*, 2025.

[51] 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

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

[53] R1-2410492, "Channel model adaptation or extension of TR38.901 for 7-24GHz," Qualcomm, 3GPP TSG RAN WG1 Meeting 119, 2024.
[54] R1-2406385, "Views on channel model adaptation/extension of TR38.901 for 7-24GHz," CATT, 3GPP TSG RAN WG1 Meeting 118, 2024.
[55] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.
[56] W. Peng, W. Li, W. Wang, X. Wei, and T. Jiang, "Downlink channel prediction for time-varying FDD massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1090–1102, Sep. 2019.

Hongwei Hou (Graduate Student Member, IEEE) received the B.E. degree in information engineering from Southeast University, Nanjing, China, in 2020. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. From 2024 to 2025, he was a visiting Ph.D. student at the Communication Systems Department, EURECOM, France. His research interests are mainly on massive MIMO communications, statistical signal processing, and deep learning.

Yafei Wang (Graduate Student Member, IEEE) received the B.E. degree in communication engineering from Tianjin University, Tianjin, China, in 2021. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. Since December 2024, he has been a Visiting Student at the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg. His research interests are mainly on satellite communications, deep learning, and signal processing.

Xinping Yi (Member, IEEE) received his Ph.D. degree in Electronics and Communications in 2015 from Télécom ParisTech, Paris, France. He is currently a Professor at the National Mobile Communications Research Laboratory, Southeast University, China. Prior to that, he has been an Assistant Professor at University of Liverpool, United Kingdom, a Postdoctoral Research Associate at Technische Universität Berlin, Germany, and a Research Assistant at EURECOM, France. His main research interests include network information theory, statistical learning theory, graph machine learning, and their applications in wireless communications and trustworthy artificial intelligence.

Wenjin Wang (Member, IEEE) received the Ph.D. degree in communication and information systems from Southeast University, Nanjing, China, in 2011. From 2010 to 2014, he was with the School of System Engineering, University of Reading, Reading, U.K. He is currently a Professor with the National Mobile Communications Research Laboratory, Southeast University. His research interests include advanced signal processing for future wireless communications and satellite communications. Dr. Wang was awarded the Best Paper Award from IEEE WCSP'09. He was also awarded the first grade Technological Invention Award of the State Education Ministry of China in 2009.

Dirk T.M. Slock (Life Fellow, IEEE) received an electronics engineering degree from Ghent University, Belgium in 1982. In 1984 he was awarded a Fulbright scholarship for Stanford University, USA, where he received the MSEE, MS in Statistics, and PhD in EE in 1986, 1989 and 1989 resp. After two years with Philips Research, he joined EURECOM in 1991 where he is now professor. He teaches statistical signal processing (SSP) and signal processing techniques for wireless communications. His research interests include SSP for wireless communications and audio processing. He invented semi-blind channel estimation, the chip equalizer-correlator receiver used by 3G HSDPA mobile terminals, spatial multiplexing cyclic delay diversity (MIMO-CDD) now part of LTE, and his work led to Single Antenna Interference Cancellation (SAIC) integrated in the GSM standard in 2006. Recent keywords are multi-cell multi-user (Massive) MIMO, imperfect CSI, distributed resource allocation, variational and empirical Bayesian learning techniques, large random matrix analysis, audio source separation, location estimation and exploitation. He graduated 46 PhD students, leading to an edited book and 600+ papers, 10 best paper awards and 4 honorary mentions. He was the General Chair of the IEEE-SP SPAWC'06 and IWAENC'14 workshops, EUSIPCO'15 and MeditCom'25. He participated in 10+ French projects (2M€ funding), 10+ EU projects (4M€ funding), 10+ direct research contracts and 10 PhD student scholarships. In 2000 he cofounded SigTone, a start-up developing music signal processing products, and in 2014 Nestwave, a start-up on Ultra Low-Power Mobile Positioning (acquired by NextNav in 2023). He is a (Life) Fellow of IEEE and EURASIP. In 2018 he received the URSI France medal and became a Knight of the French "Ordre des Palmes Académiques" in 2023.

Shi Jin (Fellow, IEEE) received the B.S. degree in communications engineering from Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from Southeast University, Nanjing, in 2007. From June 2007 to October 2009, he was a Research Fellow at the Adastral Park Research Campus, University College London, London, U.K. He is currently affiliated with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include wireless communications, random matrix theory, and information theory. He serves as an Area Editor for the IEEE Transactions on Communications and IET Electronics Letters. He was previously an Associate Editor for the IEEE Transactions on Wireless Communications, IEEE Communications Letters, and IET Communications. Dr. Jin and his co-authors were awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory, a 2022 Best Paper Award, and a 2010 Young Author Best Paper Award by the IEEE Signal Processing Society, and 2023 Jack Neubauer Memorial Award by the IEEE Vehicular Technology Society.