Tensor-Structured Bayesian Channel Prediction for XL-MIMO Systems

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

***Abstract*—This paper investigates channel prediction for extremely large-scale multiple-input multiple-output (XL-MIMO) systems under near-field (NF) propagation, where spherical wavefronts and high-dimensional channel representations make both modeling and inference computationally challenging. To fully exploit the correlations in the spatial, frequency, and temporal domains, we develop a tensor-structured channel model and its beam-delay-Doppler (BDD) domain representation, where Doppler-domain features enable physically interpretable extrapolation over arbitrary prediction horizons within the stationarity interval. To handle NF propagation without incurring a combinatorial grid explosion, we propose a hybrid beam-domain strategy that discretizes only the angle dimension while treating the NF slope parameters as continuous, learnable hyperparameters, together with perturbation-based grid refinement to support continuous-valued physical parameters without dense sampling. Building on the resulting Bayesian formulation with a sparse prior, we design an approximate inference algorithm within the generalized approximate message passing (GAMP) framework, where both the mean-type and variance-type matrix-vector multiplications are implemented via mode-wise tensor contractions that exploit the separable Kronecker structure of the transform matrix, thereby avoiding explicit Kronecker matrices and substantially reducing complexity. Numerical results using the practical channel simulator QuaDRiGa demonstrate that the proposed approach achieves significantly improved channel prediction performance compared with existing methods.**

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

I. INTRODUCTION

As next-generation communication systems evolve toward increasingly data-intensive applications, extremely large-scale multiple-input multiple-output (XL-MIMO) are being widely adopted to meet growing capacity demands [2], [3]. However, the system performance is vulnerable to channel aging, i.e., the mismatch between the estimated and actual channels caused by temporal variations, rendering timely and accurate channel state information (CSI) essential. In response to this, channel prediction has emerged as a promising approach to mitigate channel aging by exploiting temporal correlations inherent in historical CSI.

In massive MIMO systems, early works exploit spectral estimation for channel prediction, with theoretical performance bounds derived to characterize the fundamental limits [4], [5]. For orthogonal frequency-division multiplexing (OFDM)-based systems, Doppler frequencies are extracted from dominant angle-delay domain taps to suppress the noise [6], [7]. Moreover, joint channel estimation and prediction frameworks based on beam-delay-Doppler (BDD) representations are proposed to support the arbitrary-length channel prediction [8], [9]. Despite these advances, channel prediction for XL-MIMO systems equipped with the extremely large aperture arrays (ELAAs) remains largely unexplored, as existing studies predominantly focus on channel estimation or tracking. Specifically, to address near-field (NF) propagation in XL-MIMO, polar-domain representations have been proposed to exploit channel sparsity in the beam domain [10], [11]. Most recently, [12] presents an initial attempt at channel prediction in XL-MIMO by approximating spherical wavefronts and extracting Doppler information via wavefront transformation. However, the high-dimensional channel representations and spherical wavefront effects in XL-MIMO systems lead to prohibitive computational complexity and performance degradation when using existing methods.

To this end, we propose a tensor-structured Bayesian channel prediction (TSBCP) algorithm for XL-MIMO systems operating under NF propagation. Beyond the far-field BDD framework in [9], we extend the channel model and algorithm design to account for NF spherical wavefronts. The main contributions are as follows.

- *Channel model:* We develop a tensor-structured SFT channel model incorporating NF spherical wavefront effects, which exclusively impact the spatial domain due to the separability of the different domains.
- *Probabilistic model:* We propose a hybrid beam-domain strategy that discretizes only angles while treating NF slope parameters as hyperparameters learned within the EM framework, avoiding combinatorial grid explosion.
- *Tensor-structured algorithm:* Due to the separability of different domains, we implement mean-type and variance-type multiplications within the GAMP framework using low-complexity mode-wise tensor contraction, with the M-step auxiliary matrix treated similarly.

Numerical results using QuaDRiGa demonstrate superior channel prediction performance over existing methods.

Notation: Boldface lowercase and uppercase letters denote vectors and matrices, respectively. Calligraphic letters denote

An extended version of this work, which further addresses spatial non-stationarity, has been published in [1].

tensors. $(\cdot)^T$, $(\cdot)^H$, $\otimes$, $\circ$, $\odot$, $\mathrm{vec}\{\cdot\}$, and $|\cdot|^{\circ 2}$ denote transpose, Hermitian transpose, Kronecker product, outer product, Hadamard product, vectorization, and element-wise magnitude squared, respectively. $\mathbb{E}\{\cdot\}$, $\delta(\cdot)$, and $\mathcal{CN}(\cdot)$ denote expectation, Dirac delta function, and circularly symmetric complex Gaussian distribution, respectively. $[\mathbf{a}]_i$ denotes the i-th element of vector $\mathbf{a}$. $\times_n$ denotes the n-mode product of a tensor with a matrix. $\mathsf{CN}(\boldsymbol{\mathcal{X}};\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{E}})$, $\mathsf{BG}(\boldsymbol{\mathcal{X}};\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{E}})$, 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 $\boldsymbol{\mathcal{M}}$.

II. System Model

We consider an XL-MIMO-OFDM system under time-division duplexing (TDD). The base station (BS) is equipped with a ULA of $N_{\text{an}} \gg 1$ antennas and serves a typical single-antenna mobile terminal (MT) moving with constant velocity. In this system, the OFDM symbol duration and subcarrier spacing are denoted by ΔT and Δf, respectively, with $\Delta T \Delta f = 1$. The BS employs sounding reference signals (SRS) [13] for channel sounding, where pilots appear every N_{IS} OFDM symbols and N_{TC} subcarriers, corresponding to spacings $\Delta\bar{T} = N_{\text{IS}}\Delta T$ and $\Delta\bar{f} = N_{\text{TC}}\Delta f$. To support real-time prediction, we adopt a sliding-frame protocol, where the BS uses received pilots in the current frame to predict channels for subsequent OFDM symbols until the next pilot arrives, and the procedure repeats as the frame slides.

A. Signal and Channel Model

Based on a ray-based model, the channel impulse response (CIR) at the n_{an}-th antenna is [14]

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

where β_l, $\tilde{\nu}_l$, and $\tilde{\tau}_l$ denote the complex gain, Doppler frequency, and the reference delay of the l-th path, respectively, and $\Delta\tilde{\tau}_{l,n_{\text{an}}}$ captures the antenna-dependent delay deviation induced by spherical wavefronts. With the constant-projection constraint $\tilde{r}_{l,n_{\text{an}}} \cos\tilde{\theta}_{l,n_{\text{an}}} = \tilde{r}_{l,1}\cos\tilde{\theta}_{l,1}$, the distance is approximated (up to the second order) as [3], [15]

$$\tilde{r}_{l,n_{\text{an}}} \approx \tilde{r}_l - (n_{\text{an}}-1)d\tilde{\phi}_l + (n_{\text{an}}-1)^2 d^2 \tilde{\eta}_l, \quad (2)$$

where $\tilde{\phi}_l \triangleq \sin\tilde{\theta}_{l,1}$, $\tilde{r}_l \triangleq \tilde{r}_{l,1}$, and $\tilde{\eta}_l \triangleq (1-\tilde{\phi}_l^2)/(2\tilde{r}_l)$ is the NF "slope" parameter that quantifies the deviation from planar wavefronts. Accordingly, we have

$$\Delta\tilde{\tau}_{l,n_{\text{an}}} = [-(n_{\text{an}}-1)d\tilde{\phi}_l + (n_{\text{an}}-1)^2 d^2\tilde{\eta}_l]/c. \quad (3)$$

Assuming the CIR is constant within one OFDM symbol, after Fourier transformation and stacking across antennas, subcarriers, and pilot symbols, the spatial-frequency-temporal (SFT) channel tensor can be expressed as

$$\boldsymbol{\mathcal{H}} = \sum_{l=1}^{L} \beta_l\, \mathbf{a}_{\text{NF}}(\tilde{\phi}_l, \tilde{\eta}_l) \circ \mathbf{b}(\tilde{\tau}_l) \circ \mathbf{c}(\tilde{\nu}_l), \quad (4)$$

where $\mathbf{b}(\tau)$ and $\mathbf{c}(\nu)$ are frequency and temporal domain steering vectors with $[\mathbf{b}(\tau)]_{n_{\text{sc}}} = \exp\big(-j2\pi(n_{\text{sc}}-1)\Delta\bar{f}\tau\big)$ and $[\mathbf{c}(\nu)]_{n_{\text{sym}}} = \exp\big(j2\pi(n_{\text{sym}}-1)\Delta\bar{T}\nu\big)$, respectively, $\mathbf{a}_{\text{NF}}(\phi,\eta)$ is the NF spatial domain steering vector, defined as $[\mathbf{a}_{\text{NF}}(\phi,\eta)]_{n_{\text{an}}} = \exp(j2\pi(n_{\text{an}}-1)d[\phi-(n_{\text{an}}-1)d\eta]/\lambda)$ with $\lambda = c/f_{\text{c}}$. Here, the linear phase term $(n_{\text{an}}-1)d\phi/\lambda$ captures the far-field planar wavefront, while the quadratic phase term $-(n_{\text{an}}-1)^2 d^2\eta/\lambda$ accounts for the NF spherical wavefront deviation. When $\eta = 0$, the NF steering vector reduces to the standard far-field array response. After CP removal and OFDM demodulation, the received signal is

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \odot \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{Z}}, \quad (5)$$

where $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Z}}$ denote the pilot tensor and AWGN, respectively. Since pilots are known at both BS and MT, we set $\boldsymbol{\mathcal{X}}$ as all-one tensor without loss of generality.

B. Tucker-Based BDD Domain Representation

To incorporate physical priors and enable scalable inference, we construct BDD domain factor matrices, given by

$$\mathbf{A}(\boldsymbol{\phi},\boldsymbol{\eta}) = [\mathbf{a}_{\text{NF}}(\phi_1,\eta_1),\ldots,\mathbf{a}_{\text{NF}}(\phi_{K_{\text{be}}},\eta_{K_{\text{be}}})], \quad (6a)$$

$$\mathbf{B}(\boldsymbol{\tau}) = [\mathbf{b}(\tau_1),\ldots,\mathbf{b}(\tau_{K_{\text{de}}})], \quad (6b)$$

$$\mathbf{C}(\boldsymbol{\nu}) = [\mathbf{c}(\nu_1),\ldots,\mathbf{c}(\nu_{K_{\text{do}}})], \quad (6c)$$

where K_{be}, K_{de}, and K_{do} denote the number of beam, delay, and Doppler domain grids, respectively. In practice, $K_{\text{be}} = N_{\text{an}}$ and $K_{\text{de}} = N_{\text{sc}}$ match the respective observation dimensions, while $K_{\text{do}} \geq N_{\text{sym}}$ can moderately oversample the Doppler domain to improve prediction accuracy. The vectors $\boldsymbol{\phi}$, $\boldsymbol{\eta}$, $\boldsymbol{\tau}$, and $\boldsymbol{\nu}$ denote the BDD domain grids. Then, the SFT channel admits the Tucker form

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\boldsymbol{\phi},\boldsymbol{\eta}) \times_2 \mathbf{B}(\boldsymbol{\tau}) \times_3 \mathbf{C}(\boldsymbol{\nu}), \quad (7)$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{C}^{K_{\text{be}}\times K_{\text{de}}\times K_{\text{do}}}$ is the BDD domain channel tensor. Building upon this representation, the channel prediction is achieved by recovering $\boldsymbol{\mathcal{G}}$ and the associated BDD grids from $\boldsymbol{\mathcal{Y}}$, and Doppler domain features enable physically interpretable extrapolation over arbitrary prediction horizons.

III. Tensor-Structured Channel Prediction

To overcome the reliance on prior sparsity knowledge in Tucker-based channel recovery, we adopt a Bayesian formulation that enables data-driven sparsity inference.

A. Probabilistic Model and Problem Formulation

1) Observation and Multi-linear Transformation: From (5), the likelihood is

$$\mathsf{P}(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}) \propto \mathsf{CN}(\boldsymbol{\mathcal{Y}};\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{C}}(\sigma_z^2)). \quad (8)$$

For tractable Bayesian inference, we start from fixed coarse BDD domain grids obtained via uniform sampling. However, under NF propagation, the array response depends jointly on angle and slope parameters, and a straightforward discretization over both dimensions leads to a combinatorial explosion of grids, rendering inference computationally infeasible.

To address this challenge, we adopt a hybrid beam domain strategy: only angle parameters are discretized, while the slope parameters are modeled as continuous hyperparameters to be learned from data. This design decouples discretization from NF modeling, dramatically reducing the number of grids while preserving the ability to accurately capture spherical wavefront. Furthermore, to bridge the gap between discrete grids and continuous physical parameters, we refine the coarse BDD grids via perturbation parameters: $\boldsymbol{\phi} = \bar{\boldsymbol{\phi}} + \Delta\boldsymbol{\phi}, \boldsymbol{\tau} = \bar{\boldsymbol{\tau}} + \Delta\boldsymbol{\tau}, \boldsymbol{\nu} = \bar{\boldsymbol{\nu}} + \Delta\boldsymbol{\nu}$. This sampling-and-refinement mechanism enables continuous-valued parameter recovery using lightweight updates, avoiding dense grid sampling while maintaining high modeling fidelity. Accordingly, the multi-linear transformation in (7) is modeled as

$$\begin{aligned}\mathsf{P}(\boldsymbol{\mathcal{H}}\,|\boldsymbol{\mathcal{G}}; \Delta\boldsymbol{\phi}, \Delta\boldsymbol{\tau}, \Delta\boldsymbol{\nu}, \boldsymbol{\eta}) &\propto \delta(\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}} + \Delta\boldsymbol{\phi}, \boldsymbol{\eta}) \\ &\times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \Delta\boldsymbol{\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \Delta\boldsymbol{\nu})),\end{aligned} \tag{9}$$

where $\boldsymbol{\eta}$ becomes the NF slope hyperparameters.

2) BDD Domain Channel: The low-rank SFT channel corresponds to a sparse BDD domain representation. We adopt a Bernoulli-Gaussian (BG) prior given by

$$\mathsf{P}(\boldsymbol{\mathcal{G}}; \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{V}}) \propto \mathsf{BG}(\boldsymbol{\mathcal{G}}; \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{C}}(0), \boldsymbol{\mathcal{V}}), \tag{10}$$

where the hyperparameters $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{V}}$ denote the element-wise sparsity rate (probability of being nonzero) and the variance of nonzero entries, respectively, which control the sparsity pattern and power profile of $\boldsymbol{\mathcal{G}}$.

3) MMSE-Based Problem Formulation: Given the models above, the MMSE estimate of the BDD domain channel is formulated as the posterior mean with the knowledge of hyperparameters $\mathcal{P}_{\text{HP}} \triangleq \{\Delta\boldsymbol{\phi}, \Delta\boldsymbol{\tau}, \Delta\boldsymbol{\nu}, \boldsymbol{\eta}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{V}}\}$. However, since $\mathcal{P}_{\text{HP}}$ is environment-dependent and unknown, we employ the expectation-maximization (EM) framework to learn it from observations [16], which alternates between: (i) an E-step that computes the posterior given current hyperparameters, and (ii) an M-step that updates $\mathcal{P}_{\text{HP}}$ by maximizing the expected complete-data log-likelihood.

B. TS-EM-AMP Algorithm for Channel Prediction

1) Tensor-Structured Approximate Inference (E-Step): Following probabilistic models and EM framework, the observation tensor is given by

$$\begin{aligned}\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}} + \hat{\Delta}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}) \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\Delta}\boldsymbol{\tau}) \\ \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \hat{\Delta}\boldsymbol{\nu}) + \boldsymbol{\mathcal{Z}},\end{aligned} \tag{11}$$

where $\hat{\Delta}\boldsymbol{\phi}$, $\hat{\Delta}\boldsymbol{\tau}$, $\hat{\Delta}\boldsymbol{\nu}$, and $\hat{\boldsymbol{\eta}}$ are the current hyperparameter estimates from the M-step. By vectorizing (11), we obtain the linear model expressed as

$$\text{vec}(\boldsymbol{\mathcal{Y}}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\text{vec}(\boldsymbol{\mathcal{G}}) + \text{vec}(\boldsymbol{\mathcal{Z}}), \tag{12}$$

where we define $\boldsymbol{\Phi} \triangleq \mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A}$. However, the size of vectorized observations and BDD domain channels scales as $K_{\text{be}}K_{\text{de}}K_{\text{do}}$, rendering exact Bayesian inference computationally intractable.

To obtain a scalable approximation, we adopt the GAMP framework [17], which converts the posterior inference into an iterative procedure involving matrix-vector multiplications with $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^H$ (for mean-type updates), element-wise magnitude-squared matrix products with $\mathbf{S} \triangleq |\boldsymbol{\Phi}|^{\circ 2}$ and $\mathbf{S}^T$ (for variance-type updates), together with component-wise denoising. The computational bottleneck thus lies in repeatedly computing these matrix-vector products.

The key observation is that the Kronecker structure $\boldsymbol{\Phi} = \mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A}$ in (12) carries over to the variance matrix $\mathbf{S} = |\mathbf{C}|^{\circ 2} \otimes |\mathbf{B}|^{\circ 2} \otimes |\mathbf{A}|^{\circ 2}$, since element-wise operations preserve the Kronecker structure. By exploiting this structure, the products $\boldsymbol{\Phi}\mathbf{u}$, $\boldsymbol{\Phi}^H\mathbf{v}$, $\mathbf{S}\mathbf{u}$, and $\mathbf{S}^T\mathbf{v}$ can all be computed via mode-wise tensor-matrix multiplications with $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ (or their element-wise counterparts). This is an implementation technique that avoids explicit construction of the Kronecker matrices and reduces the per-iteration complexity to linear scaling with tensor dimensions, while providing posterior means and variances of $\boldsymbol{\mathcal{G}}$ (and $\boldsymbol{\mathcal{H}}$) as sufficient statistics for the subsequent M-step. To improve convergence, adaptive damping [18] is applied to the GAMP iterates.

2) Hyperparameter Learning (M-Step): In the M-step, we update the BDD domain grid perturbations and the NF slope parameters by maximizing the expected complete-data log-likelihood under the current E-step posteriors.

Proposition 1. Given the posterior means $\hat{\boldsymbol{\mathcal{H}}}$ and $\hat{\boldsymbol{\mathcal{G}}}$ from the E-step, the M-step update of $\{\Delta\boldsymbol{\phi}, \boldsymbol{\eta}, \Delta\boldsymbol{\tau}, \Delta\boldsymbol{\nu}\}$ reduces to the following residual minimization problem, given by

$$\{\hat{\Delta}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}, \hat{\Delta}\boldsymbol{\tau}, \hat{\Delta}\boldsymbol{\nu}\} = \underset{\Delta\boldsymbol{\phi}, \boldsymbol{\eta}, \Delta\boldsymbol{\tau}, \Delta\boldsymbol{\nu}}{\arg\min}\ f(\Delta\boldsymbol{\phi}, \boldsymbol{\eta}, \Delta\boldsymbol{\tau}, \Delta\boldsymbol{\nu}), \tag{13}$$

where constant terms independent of the hyperparameters are omitted and the objective function is the expected complete-data log-likelihood, simplified as

$$\begin{aligned}f(\Delta\boldsymbol{\phi}, \boldsymbol{\eta}, \Delta\boldsymbol{\tau}, \Delta\boldsymbol{\nu}) = \|\hat{\boldsymbol{\mathcal{H}}} - \hat{\boldsymbol{\mathcal{G}}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}} + \Delta\boldsymbol{\phi}, \boldsymbol{\eta}) \\ \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \Delta\boldsymbol{\tau}) \times_3 \mathbf{C}(\bar{\boldsymbol{\nu}} + \Delta\boldsymbol{\nu})\|_{\text{F}}^2.\end{aligned} \tag{14}$$

Proof. This result comes from sharply peaked Gaussian approximation of the Dirac delta function in the probabilistic model and constant magnitudes of factor matrices. □

Proposition 2. With $\Delta\boldsymbol{\chi} \triangleq [\Delta\boldsymbol{\phi}^T, \boldsymbol{\eta}^T]^T$, the objective in (13) admits the following quadratic approximations via first-order Taylor expansions of the factor matrices:

$$J_{\text{be}}(\Delta\boldsymbol{\chi}) \approx \Delta\boldsymbol{\chi}^T\boldsymbol{\Pi}_{\text{be}}\Delta\boldsymbol{\chi} - 2\mathsf{Re}\{\boldsymbol{\mu}_{\text{be}}^T\}\Delta\boldsymbol{\chi} + C_{\text{be}}, \tag{15a}$$

$$J_{\text{de}}(\Delta\boldsymbol{\tau}) \approx \Delta\boldsymbol{\tau}^T\boldsymbol{\Pi}_{\text{de}}\Delta\boldsymbol{\tau} - 2\mathsf{Re}\{\boldsymbol{\mu}_{\text{de}}^T\}\Delta\boldsymbol{\tau} + C_{\text{de}}, \tag{15b}$$

$$J_{\text{do}}(\Delta\boldsymbol{\nu}) \approx \Delta\boldsymbol{\nu}^T\boldsymbol{\Pi}_{\text{do}}\Delta\boldsymbol{\nu} - 2\mathsf{Re}\{\boldsymbol{\mu}_{\text{do}}^T\}\Delta\boldsymbol{\nu} + C_{\text{do}}, \tag{15c}$$

where $\boldsymbol{\Pi}_{\cdot}$, $\boldsymbol{\mu}_{\cdot}$ denote the quadratic and linear coefficients computed from the E-step posteriors $\hat{\boldsymbol{\mathcal{G}}}$, $\hat{\boldsymbol{\mathcal{H}}}$ and derivatives of the factor matrices, and $C_{\cdot}$ are constants. Closed-form solutions follow directly. Detailed expressions and derivations are given in [9, Appendix C] and [1]. The construction of auxiliary matrices benefits from the separable structure across domains, as intermediate quantities are computed via mode-wise tensor contractions.

The BG hyperparameters are updated using standard EM rules [19]. Since K_{be}, K_{de}, and K_{do} are of the same order as N_{an}, N_{sc}, and N_{sym}, respectively, the overall per-iteration complexity of the algorithm is $\mathcal{O}(T_{\text{G}}N_{\text{an}}N_{\text{sc}}N_{\text{sym}}(N_{\text{an}}+N_{\text{sc}}+N_{\text{sym}}))$, where T_{G} is the number of GAMP iterations per EM

Algorithm 1 TS-EM-AMP Algorithm for Channel Prediction

Input: Observation $\boldsymbol{\mathcal{Y}}$ and noise variance σ_z^2.

Output: Predicted channel $\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}}$.

1: Initialize: $\hat{\boldsymbol{\mathcal{M}}} = \boldsymbol{\mathcal{C}}(0.5)$, $\hat{\boldsymbol{\mathcal{V}}} = \boldsymbol{\mathcal{C}}(\|\boldsymbol{\mathcal{Y}}\|_{\mathrm{F}}^2/(N_{\mathrm{an}}N_{\mathrm{sc}}N_{\mathrm{sym}}))$, $\hat{\boldsymbol{\mathcal{G}}} = \boldsymbol{\mathcal{C}}(0)$, $\hat{\Delta}\boldsymbol{\phi} = \hat{\Delta}\boldsymbol{\tau} = \hat{\Delta}\boldsymbol{\nu} = \mathbf{0}$, $\hat{\boldsymbol{\eta}} = \mathbf{0}$.

2: **for** $t = 1, \ldots, T_{\mathrm{M}}$ **do**

3: **E-step:** run GAMP with tensor-structured matrix-vector products and adaptive damping [18] to obtain posterior means/variances of $\boldsymbol{\mathcal{G}}$ (and $\boldsymbol{\mathcal{H}}$).

4: **M-step:** update $\hat{\Delta}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}, \hat{\Delta}\boldsymbol{\tau}, \hat{\Delta}\boldsymbol{\nu}$ via Proposition 2, and update $\hat{\boldsymbol{\mathcal{M}}}, \hat{\boldsymbol{\mathcal{V}}}$ via BG-EM in [19].

5: **end for**

6: Predict $\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}}$ via (16).

step, compared with $\mathcal{O}(T_{\mathrm{G}}N_{\mathrm{an}}^2N_{\mathrm{sc}}^2N_{\mathrm{sym}}^2)$ for a direct GAMP implementation without tensor-structured computation [1].

3) Channel Prediction and Algorithm Summary: After EM converges, the future channels are predicted by extrapolating the Doppler steering vectors, given by

$$\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}} = \hat{\boldsymbol{\mathcal{G}}} \times_1 \mathbf{A}(\bar{\boldsymbol{\phi}} + \hat{\Delta}\boldsymbol{\phi}, \hat{\boldsymbol{\eta}}) \times_2 \mathbf{B}(\bar{\boldsymbol{\tau}} + \hat{\Delta}\boldsymbol{\tau}) \times_3 \tilde{\mathbf{C}}(\bar{\boldsymbol{\nu}} + \hat{\Delta}\boldsymbol{\nu}), \quad (16)$$

where $[\tilde{\mathbf{c}}(\nu)]_{n_{\mathrm{cp}}} = \exp(j2\pi(T_0 + n_{\mathrm{cp}}\Delta T)\nu)$, $T_0 = (N_{\mathrm{sym}} - 1)\Delta\bar{T}$, and $\tilde{\mathbf{C}}(\boldsymbol{\nu}) = [\tilde{\mathbf{c}}(\nu_1), \ldots, \tilde{\mathbf{c}}(\nu_{K_{\mathrm{do}}})]$.

IV. Simulation Results

A. Simulation Configuration

1) Scenario Setting: To validate the exactness of proposed channel and probabilistic models, we employ the QuaDRiGa that generates XL-MIMO-OFDM channels consistent with the third Generation Partnership Program (3GPP) specifications [20] and has been validated in various field trials [21]. Unless specified otherwise, we consider the 3GPP urban macro (UMa) non-line-of-sight (NLOS) scenarios, with configurations in Section II and parameters in Table I.

2) Benchmarks and Performance Metric: We select the following algorithms as benchmarks

- **VKF** [22]: Estimates spatial domain channels by the least square (LS) algorithm, with temporal correlations captured by the AR-based vector Kalman filter.
- **FIT** [23]: Estimates beam-delay domain channels by the alternating LS (ALS) algorithm, with temporal correlations captured by the first-order Taylor series.
- **PAD** [6], **WTMP** [12]: 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.

For these benchmarks, 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

$$\mathrm{NMSE} = \|\hat{\boldsymbol{\mathcal{H}}}^{\mathrm{CP}} - \boldsymbol{\mathcal{H}}^{\mathrm{CP}}\|_{\mathrm{F}}^2/\|\boldsymbol{\mathcal{H}}^{\mathrm{CP}}\|_{\mathrm{F}}^2, \quad (17)$$

TABLE I
Scenario Parameters

Parameter	Value
Carrier Frequency	$f_{\mathrm{c}} = 15$ GHz
OFDM Symbol Duration	$\Delta T_{\mathrm{sym}} = 16.67\ \mu$s
Cyclic Prefix Duration	$\Delta T_{\mathrm{cp}} = 1.17\ \mu$s
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

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

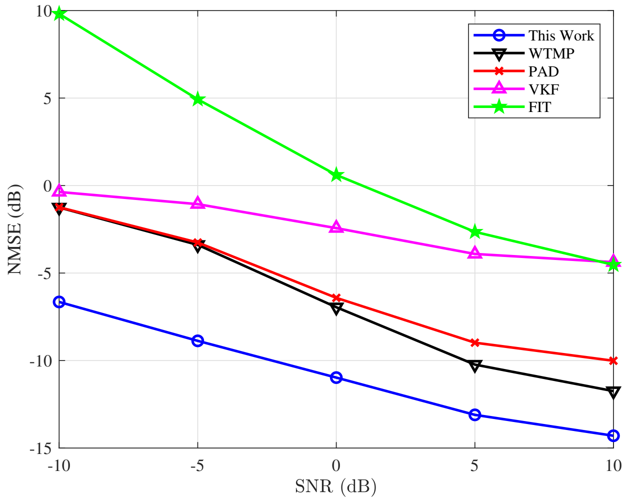
B. Performance Evaluation

Fig. 1a shows the NMSE versus SNR. **FIT** and **VKF**, which lack beam domain modeling, yield the poorest performance, with **FIT** further degraded at low SNR. **PAD** and **WTMP** benefit from beam and Doppler domain modeling, where **WTMP** gains from spherical wave modeling. The proposed algorithm significantly outperforms all benchmarks, achieving at least 2 dB gain at SNR = 10 dB, with more substantial improvements in the low-SNR regime.

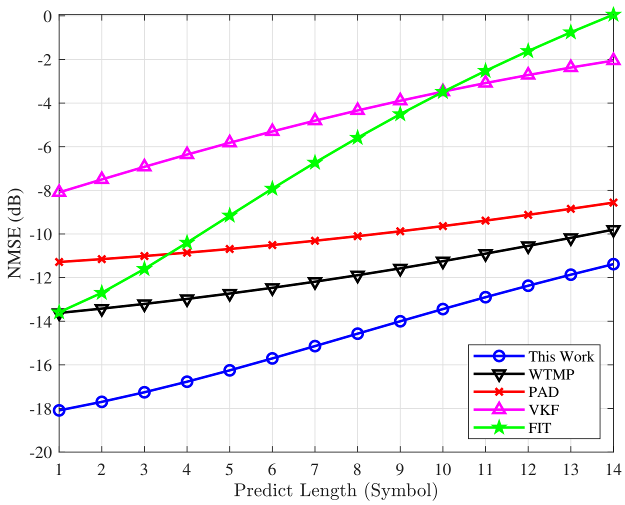
Fig. 1b shows the NMSE versus prediction length at SNR = 10 dB. **FIT** degrades sharply (about 14 dB) due to first-order Taylor extrapolation errors, while **VKF** exhibits about 6 dB degradation. **PAD** and **WTMP** show improved stability owing to Prony and matrix pencil methods. The proposed algorithm consistently outperforms all benchmarks, achieving NMSEs of approximately -18 dB and -11.5 dB for the first future non-pilot and pilot symbols, respectively.

V. Conclusion

This paper studied channel prediction for XL-MIMO systems under NF propagation and addressed the associated computational challenges through a tensor-structured Bayesian framework. By exploiting the inherent multi-linear structure of SFT domain channels and the BDD domain representations, channel prediction was formulated as the sparse BDD domain channel recovery from noisy observations. An efficient tensor-structured algorithm based on GAMP was developed to enable scalable inference. Numerical results based on the QuaDRiGa demonstrate that the proposed approach achieves significantly improved channel prediction accuracy compared with existing methods, validating its effectiveness for practical XL-MIMO systems. The proposed framework can be further extended to scenarios where the observation involves bilinear mixing of multiple unknown factors, by introducing bridging variables that decompose the problem into conditionally independent linear and bilinear sub-modules with dedicated message passing for each, as demonstrated in [1].



(a) NMSE versus SNR.



(b) NMSE versus prediction length at SNR = 10 dB.

Fig. 1. NMSE of channel prediction.

Acknowledgment

This work was supported in part by the NSFC under Grants U25A20392, 62371122, and 62471129. EURECOM's research was partially supported by its industrial members, the French projects CellFree6G (ANR) and PERSEUS (PEPR-FN), and the H2030 EU RI project CONVERGE. The work of H. Hou was supported by China Scholarship Council (CSC).

References

[1] H. Hou, Y. Wang, X. Yi, W. Wang, D. T. M. Slock, and S. Jin, "Tensor-structured Bayesian channel prediction for upper mid-band XL-MIMO systems," *IEEE Trans. Commun.*, vol. 74, pp. 3968–3983, 2026.

[2] Z. Wang, J. Zhang, H. Du, W. E. I. Sha, B. Ai, D. Niyato, and M. Debbah, "Extremely large-scale MIMO: Fundamentals, challenges, solutions, and future directions," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 117–124, 2024.

[3] H. Hou, X. He, T. Fang, X. Yi, W. Wang, and S. Jin, "Beam-delay domain channel estimation for mmWave XL-MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 4, pp. 646–661, May 2024.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 3rd Generation Partnership Project (3GPP), "NR; physical channels and modulation," 3GPP, Technical Specification (TS) 38.211, 2025. [Online]. Available: https://www.3gpp.org

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

[15] 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.

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

[17] 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.

[18] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE, 2015, pp. 2021–2025.

[19] 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.

[20] 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

[21] 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.

[22] 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.

[23] 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.