Joint Channel and Data Estimation for Multiuser Extremely Large-Scale MIMO Systems

Kabuto Arai, *Graduate Student Member, IEEE*, Koji Ishibashi, *Senior Member, IEEE*, Hiroki Iimori, *Member, IEEE*, Paulo Valente Klaine, and Szabolcs Malomsoky

***Abstract*—This paper proposes a joint channel and data estimation (JCDE) algorithm for uplink multiuser extremely large-scale multiple-input-multiple-output (XL-MIMO) systems. The initial channel estimation is formulated as a sparse reconstruction problem based on the angle and distance sparsity under the near-field propagation condition. This problem is solved using non-orthogonal pilots through an efficient low complexity two-stage compressed sensing algorithm. Furthermore, the initial channel estimates are refined by employing a JCDE framework driven by both non-orthogonal pilots and estimated data. The JCDE problem is solved by sequential expectation propagation (EP) algorithms, where the channel and data are alternately updated in an iterative manner. In the channel estimation phase, integrating Bayesian inference with a model-based deterministic approach provides precise estimations to effectively exploit the near-field characteristics in the beam-domain. In the data estimation phase, a linear minimum mean square error (LMMSE)-based filter is designed at each sub-array to address the correlation due to energy leakage in the beam-domain arising from the near-field effects. Numerical simulations reveal that the proposed initial channel estimation and JCDE algorithm outperforms the state-of-the-art approaches in terms of channel estimation, data detection, and computational complexity.**

***Index Terms*—Extremely large-scale-MIMO (XL-MIMO), near-field, joint channel and data estimation, compressed sensing**

I. Introduction

Extremely large-scale multiple-input-multiple-output (XL-MIMO) has emerged as a promising technology, enabling sharp directive beamforming and extensive spatial multiplexing [2], [3]. However, the significant increase in antenna aperture leads to an expansion of the Rayleigh distance [4], [5], defined as the border between the near-field and far-field regions. Thus, the near-field effects in XL-MIMO systems may not be negligible in some practically-relevant circumstances, such as in small area coverage with high carrier frequency [6].

Unlike the conventional far-field, the near-field channel depends not only on angles but also on distances from signal sources such as user equipments (UEs) and scatterers. Hence, conventional channel estimation methods such as [7], [8], which exploit the beam-domain sparsity under the assumption of planar wavefront, experience performance degradation in the near-field due to energy leakage effects in the beam-domain. To tackle this issue, the authors in [6] have proposed a polar-domain simultaneous orthogonal matching pursuit (P-SOMP) algorithm, which leverages the angle and distance sparsity known as polar-sparsity arising from the near-field peculiar characteristics. In P-SOMP, polar (angle-distance) grids are generated by spatially quantizing the polar-domain to utilize compressed sensing techniques. However, in multiuser systems, P-SOMP requires orthogonal pilots to separate multiple UEs. As such, this approach results in non-negligible overhead as the number of UEs grows, especially in XL-MIMO systems capable of spatially multiplexing many UEs.

Considering these challenges, a near-field channel estimation algorithm, which works even with non-orthogonal pilots, has been proposed in [9]. However, due to the non-orthogonality among pilots, inter-user interference still remains, so it is necessary to jointly estimate all UE channel components. As a result, this joint estimation significantly increases computational complexity because it requires UE-wise polar grids, which leads to a large grid size. Therefore, the authors in [9] proposed a 2D-compressive sampling matching pursuit (CoSaMP) algorithm, which is based on the CoSaMP algorithm in the polar-UE 2D domain constructed by UE-wise polar grids. While the 2D-CoSaMP algorithm can mitigate computational complexity, its estimation performance is hindered by overfitting to noisy measurements, which results from the inverse operation on over-sampled estimates. Consequently, subsequent data detection suffers from severe deterioration, particularly with high-order modulation.

One of the prospective solutions to obtain accurate channel estimate with non-orthogonal pilots is joint channel and data estimation (JCDE) [10]–[14], where not only pilot sequences but also estimated data symbols are utilized as pilot replicas. The JCDE problem can be formulated as a bilinear inference problem (BIP). One of the prominent algorithms based on a Bayesian framework for BIP is bilinear generalized approximate message passing (BiGAMP) [15]. BiGAMP is an extension of GAMP, originally designed for a high-dimensional generalized-linear problem by utilizing loopy belief propagation (BP) with central limit theorem (CLT) and Taylor-series approximations based on large system limit to simplify the BP update[1]. Due to the heavy dependency on the large system assumption of BiGAMP, the convergence

Kabuto Arai and Koji Ishibashi are with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo 182-8285, Japan (e-mail: k.arai@awcc.uec.ac.jp, koji@ieee.org)

Hiroki Iimori, Paulo Valente Klaine, and Szabolcs Malomsoky are with Ericsson Research, Ericsson Japan K.K., Yokohama, 220-0012(e-mail: {hiroki.iimori, paulo.valente.klaine, szabolcs.malomsoky}@ericsson.com)

[1]Assuming a bilinear inference problem from the noisy measurement $\mathbf{Y} = \mathbf{HX} + \mathbf{N}$ with objective matrices $\mathbf{H} \in \mathbb{C}^{N \times U}$ and $\mathbf{X} \in \mathbb{C}^{U \times K}$, and a noise matrix $\mathbf{N} \in \mathbb{C}^{N \times K}$, the large system limit means that the dimensions N, U, and K tend to infinity while the rates N/U and K/U remain constant [15].

performance deteriorates significantly when the system size is insufficient, the pilot length is short, or the prior distribution is misspecified [10]. To address these issues, the authors in [10] have proposed bilinear Gaussian belief propagation (BiGaBP), which relaxes the BP update rules of BiGAMP, based on GaBP [16], without heavily relying on the approximation under a large system limit assumption. This relaxation of the approximation leads to performance improvements while maintaining the same complexity order as BiGAMP.

Bilinear inference algorithms that exploit physical model structures, such as channel sparsity in the beam-domain, have been investigated in [12], [13]. In these papers, the channel sparsity is modeled using a Bernoulli Gaussian (BG) prior distribution because this prior is analytically tractable with a closed-form posterior. However, the sparse structure cannot be exactly expressed by the analytically tractable prior, which leads to modeling errors. To tackle this issue, the authors of [11] have integrated a model-based deterministic approach [17] into a Bayesian inference framework [10], referred to as AoA-aided BiGaBP. This deterministic approach rectifies the model mismatch caused by the use of the tractable prior.

However, since this method assumes a far-field model, the model correction via the deterministic approach is insufficient in the near-field, where the channel in the beam-domain exhibits cluster sparsity instead of simple sparsity with a sharp peak, due to energy leakage in the beam-domain. Moreover, AoA-aided BiGaBP relies on a maximum ratio combining (MRC)-based detector, which cannot address the correlation caused by the energy leakage. While the computational complexity of the MRC-based detection is relatively low, the data denoising process in AoA-aided BiGaBP, which relies on prior information about modulation constellations, requires significant computational complexity that scales proportionally not only to the modulation order but also to the number of antennas due to its reliance on the BP update rules. Hence, a channel estimation with a deterministic approach tailored to near-field structure, along with a computationally efficient data detection to address the energy leakage, is essential for the JCDE in the near-field region. Furthermore, an accurate initial channel estimation is crucial for JCDE algorithms because the performance of data and channel estimation heavily depends on the accuracy of the initial estimate, where an inaccurate initial estimate might lead to error propagation in the JCDE iterations.

Within the context outlined above, we propose a JCDE algorithm for multiuser XL-MIMO systems with non-orthogonal pilots. Our contributions are summarized as follows.

- **Initial channel estimation for JCDE**: A novel initialization mechanism for the multiuser near-field channel estimation problem with pilot contamination due to non-orthogonal pilots is proposed, enabling an accurate initial estimate that is then used in the subsequent JCDE algorithm as the starting values. The proposed initial channel estimation algorithm consists of two stages to maintain low computational complexity. In the first stage, angle and distance parameters for all UEs are estimated from the polar grids using the simultaneous orthogonal matching pursuit (SOMP) algorithm without pairing between each estimated path and the corresponding UE. Subsequently, the second stage involves the UE-path pairing using 2D-OMP [18] with a reduced number of grids constructed on the angle and distance parameters derived from the first stage. Owing to the above two-stage procedure, our proposed initial channel estimation outperforms the existing state-of-the-art scheme [9], while maintaining comparable computational complexity.
- **JCDE algorithm with model-based estimation**: A novel bilinear JCDE inference algorithm is proposed, which integrates a model-based deterministic estimation mechanism with a Bayesian inference to address the modeling errors in a prior, as in [10]. In contrast to the state-of-the-art AoA-aided BiGaBP under the assumption of far-field, our algorithm estimates the channels as an aggregation of two distinct quantities: 1) a model-based estimate that captures the near-field channel structure and 2) its modeling error that captures how different the current estimate is from the true channel. The model-based estimate is alternately updated through a matching pursuit algorithm exploiting the near-field model structures, whereas the residual modeling errors and data symbols are jointly estimated by the expectation propagation (EP) algorithm [19], where an approximate posterior is calculated by minimizing the Kullback-Leibler (KL) divergence. To tackle the correlation caused by the energy leakage in the beam-domain while reducing the complexity, we introduce a novel posterior calculation design that enables the implementation of a sub-array-wise linear minimum mean square error (LMMSE)-based filter, allowing parallel computation of the matrix inversion with a smaller dimension than the array size. This design results in lower computational complexity compared to the state-of-the-art method, owing to the modification of an extrinsic value generation that does not rely on BP rules.

Notation: The notation $[\mathbf{A}]_{i,j}$ indicates the (i,j) element of the matrix $\mathbf{A}$. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$ indicate conjugate, transpose, and conjugate transpose, respectively. A block diagonal matrix consisting of matrices $\mathbf{A}_1, \dots \mathbf{A}_N$ is represented as $\mathrm{blkdiag}(\mathbf{A}_1, \dots, \mathbf{A}_N)$. For a random variable x and a probabilistic density function $p(x)$, $\mathbb{E}_{p(x)}[x]$ indicate the expectation of x over $p(x)$. For any function $f(\mathbf{z})$, $\int_{\mathbf{z}/z_i} f(\mathbf{z})$ denotes the integral of $f(\mathbf{z})$ with respect to $\mathbf{z}$ except for z_i. The operator $\otimes$ denotes the Kronecker product. For the index sets $\mathcal{I} = \{1, 2, \dots, I\}$ and $\mathcal{J} = \{1, 2, \dots, J\}$, $\mathcal{I} \times \mathcal{J}$ denotes the cartesian product of $\mathcal{I}$ and $\mathcal{J}$. $\mathcal{I} \setminus i$ represent the set $\{1, \dots, i-1, i+1, \dots, I\}$. The notation $x \in [x_1, x_2]$ denotes that x belongs to the closed interval between x_1 and x_2.

II. System Model

We consider an uplink XL-MIMO system, where a base station (BS) has a uniform linear array (ULA) with N-antennas, serving U single antenna UEs. The ULA is positioned along the y-axis, where $y^{(n)} = (n-1)d - \frac{(N-1)d}{2},\ n = 1, 2, \dots, N$ is the n-th antenna coordinate, and $d = \lambda/2$ is antenna spacing with wavelength λ, as shown in Fig.1.

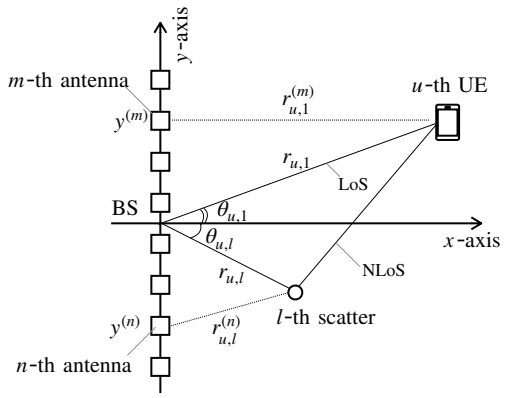


Fig. 1: Near-field channel model. $r_{u,l}^{(n)}$ is the distance between the n-th antenna and the u-th UE ($l = 1$) or scatterers ($l \neq 1$).

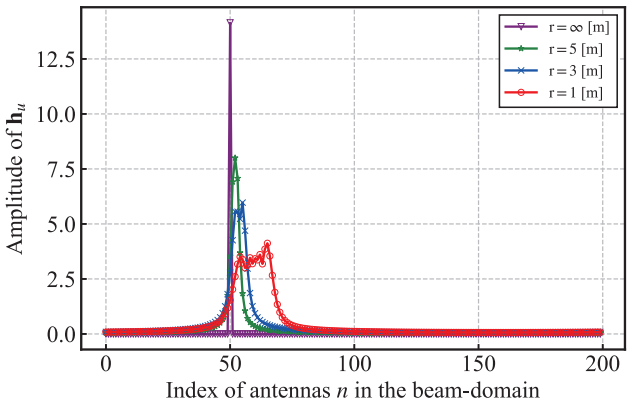


Fig. 2: The amplitude of the channel vector in the beam-domain, $\mathbf{h}_u$ for $N = 200, L = 1, f_c = 100$ GHz. One UE is located r meters away from the BS ($r = \infty$ corresponds to the far-field case).

A. Channel Model

The near-field channel in the spatial-domain between the BS and the u-th UE is modeled as

$$\mathbf{h}_u^{\mathcal{S}} = \sum_{l=1}^{L_u} \mathbf{a}(\theta_{u,l}, r_{u,l}) z_{u,l} = \mathbf{A}(\boldsymbol{\theta}_u, \mathbf{r}_u)\mathbf{z}_u, \tag{1}$$

where $\theta_{u,l} \in \mathbb{R}$, and $z_{u,l} \in \mathbb{C}$ denote the angles of arrival (AoA) and path gain of the l-th path and the u-th UE, and L_u is the number of paths for the u-th UE [20].

Without loss of generality, $l = 1$ represents the line-of-sight (LoS) component and $l \in \{2, \dots L_u\}$ represents the non-line-of-sight (NLoS) components. Let $L \triangleq \sum_{u=1}^{U} L_u$ denote the total number of paths including U-UEs. Accordingly, $r_{u,1} \in \mathbb{R}$ denotes the distance between the BS and the u-th UE, and $r_{u,l} \in \mathbb{R},\ l \neq 1$ is the distance between the BS and the l-th scatterer around the u-th UE. Besides, $\mathbf{a}(\theta_{u,l}, r_{u,l}) \in \mathbb{C}^{N\times 1}$ denotes the array response vector defined as

$$[\mathbf{a}(\theta_{u,l}, r_{u,l})]_n = \exp\left[-j\frac{2\pi}{\lambda}\left(r_{u,l}^{(n)} - r_{u,l}\right)\right], \tag{2}$$

where $r_{u,l}^{(n)} = \sqrt{r_{u,l}^2 + y^{(n)2} - 2r_{u,l}y^{(n)}\sin\theta_{u,l}}$ is the distance between the n-th antenna and the u-th UE ($l = 1$) or scatterers ($l \neq 1$).

For the u-th UE, let us define the collections of AoAs, distances, and path gains as $\boldsymbol{\theta}_u \triangleq \{\theta_{u,l}\}_{l=1}^{L_u}$, $\mathbf{r}_u \triangleq \{r_{u,l}\}_{l=1}^{L_u}$, and $\mathbf{z}_u \triangleq \left[z_{u,1}, \dots, z_{u,L_u}\right]^{\mathrm{T}} \in \mathbb{C}^{L_u\times 1}$, respectively, and the corresponding array response matrix is defined as $\mathbf{A}(\boldsymbol{\theta}_u, \mathbf{r}_u) \triangleq \left[\mathbf{a}(\theta_{u,1}, r_{u,1}), \dots, \mathbf{a}(\theta_{u,L_u}, r_{u,L_u})\right] \in \mathbb{C}^{N\times L_u}$. Then, the channel matrix $\mathbf{H}^{\mathcal{S}} \triangleq \left[\mathbf{h}_1^{\mathcal{S}}, \dots, \mathbf{h}_U^{\mathcal{S}}\right] \in \mathbb{C}^{N\times U}$ is written as

$$\mathbf{H}^{\mathcal{S}} = \mathbf{A}(\boldsymbol{\theta}, \mathbf{r})\mathbf{Z}, \tag{3}$$

where $\mathbf{A}(\boldsymbol{\theta}, \mathbf{r}) \triangleq \left[\mathbf{A}(\boldsymbol{\theta}_1, \mathbf{r}_1), \dots, \mathbf{A}(\boldsymbol{\theta}_U, \mathbf{r}_U)\right] \in \mathbb{C}^{N\times L}$ is the array response matrix consisting of U-UEs with AoAs, distances and path gains defined as $\boldsymbol{\theta} \triangleq \{\boldsymbol{\theta}_u\}_{u=1}^{U}$, $\mathbf{r} \triangleq \{\mathbf{r}_u\}_{u=1}^{U}$, and $\mathbf{Z} \triangleq \mathrm{blkdiag}\left(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_U\right) \in \mathbb{C}^{L\times U}$, respectively.

The array response vector in the far-field region, *i.e.,* when $r_{u,l} \to \infty$ is expressed as $[\mathbf{a}(\theta_{u,l}, \infty)]_n = \exp\left[j\frac{2\pi y^{(n)}}{\lambda}\sin\theta_{u,l}\right]$ from (2). As the far-field array response depends only on the angle, the far-field channel $\mathbf{h}_u^{\mathcal{S}}$ can be converted into a sparse beam-domain channel $\mathbf{h}_u = \mathbf{D}_N\mathbf{h}_u^{\mathcal{S}}$ with the discrete Fourier transform (DFT) matrix $\mathbf{D}_N \in \mathbb{C}^{N\times N}$. In contrast, as the far-field approximation does not hold in the near-field region, the beam-domain near-field channel $\mathbf{h}_u$ exhibits not a simple sparse structure but rather a cluster sparse structure, which is caused by energy leakage due to a model mismatch between the DFT matrix $\mathbf{D}_N$ and the near-field array response $\mathbf{a}(\theta_{u,l}, r_{u,l})$. To illustrate the energy leakage effects, Fig. 2 depicts the amplitude of the beam-domain channel vector in the near-field and far-field regions. It can be seen that the far-field channel possesses a distinct sparse structure with a peaky spike. On the other hand, the near-field channel exhibits a clustered sparsity with flatter peaks due to energy leakage. Hence, conventional channel estimation methods exploiting the beam-domain sparsity [7], [8], [17] encounter significant performance degradation in the near-field.

B. Received Signal Model

To estimate the near-field channel, the u-th UE transmits a pilot sequence $\mathbf{x}_{\mathrm{p},u} \in \mathbb{C}^{K_{\mathrm{p}}\times 1}$ and data symbol $\mathbf{x}_{\mathrm{d},u} \in \mathbb{C}^{K_{\mathrm{d}}\times 1}$ subsequently, where K_{p} and K_{d} are the length of pilots and data symbols[2]. To reduce the pilot overhead for channel estimation, we employ non-orthogonal pilots that satisfy $K_{\mathrm{p}} < U$ and $\mathbf{x}_{\mathrm{p},u}^{\mathrm{T}}\mathbf{x}_{\mathrm{p},u'}^{*} \neq 0, (u' \neq u)$, which results in the pilot contamination between UEs unlike the orthogonal pilots.

Each entry of $\mathbf{x}_{\mathrm{d},u}$ is randomly generated from a Q-quadrature amplitude modulation (QAM) constellation $\mathcal{X} \triangleq \{\mathcal{X}_1, \dots, \mathcal{X}_Q\}$ with average symbol energy E_{s}. Then, the received pilot $\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}} \in \mathbb{C}^{N\times K_{\mathrm{p}}}$ and data $\mathbf{Y}_{\mathrm{d}}^{\mathcal{S}} \in \mathbb{C}^{N\times K_{\mathrm{d}}}$ in the spatial-domain are given by

$$\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}} = \mathbf{H}^{\mathcal{S}}\mathbf{X}_{\mathrm{p}} + \mathbf{N}_{\mathrm{p}}^{\mathcal{S}}, \quad \mathbf{Y}_{\mathrm{d}}^{\mathcal{S}} = \mathbf{H}^{\mathcal{S}}\mathbf{X}_{\mathrm{d}} + \mathbf{N}_{\mathrm{d}}^{\mathcal{S}}, \tag{4}$$

where $\mathbf{X}_{\mathrm{p}} \triangleq \left[\mathbf{x}_{\mathrm{p},1}, \dots, \mathbf{x}_{\mathrm{p},U}\right]^{\mathrm{T}} \in \mathbb{C}^{U\times K_{\mathrm{p}}}$ and $\mathbf{X}_{\mathrm{d}} \triangleq \left[\mathbf{x}_{\mathrm{d},1}, \dots, \mathbf{x}_{\mathrm{d},U}\right]^{\mathrm{T}} \in \mathbb{C}^{U\times K_{\mathrm{d}}}$ are the transmitted pilot matrix and data matrices. $\mathbf{N}_{\mathrm{p}} \in \mathbb{C}^{N\times K_{\mathrm{p}}}$ and $\mathbf{N}_{\mathrm{d}} \in \mathbb{C}^{N\times K_{\mathrm{d}}}$ are the additive white Gaussian noise (AWGN) matrices, whose entries are generated from $\mathcal{CN}(0, \sigma^2)$ with noise variance σ^2. By stacking the received pilot $\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}}$ and data $\mathbf{Y}_{\mathrm{d}}^{\mathcal{S}}$, the

[2]Since the pilots and data symbols are transmitted in distinct time symbols, unlike superimposed pilot transmission schemes [21], contamination between the pilots and data symbols does not occur.

effective received signal $\mathbf{Y}^{\mathcal{S}} \triangleq \left[\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}}, \mathbf{Y}_{\mathrm{d}}^{\mathcal{S}}\right] \in \mathbb{C}^{N\times K}$, with $K \triangleq K_{\mathrm{p}} + K_{\mathrm{d}}$, is formulated as

$$\mathbf{Y}^{\mathcal{S}} = \mathbf{H}^{\mathcal{S}}\mathbf{X} + \mathbf{N}^{\mathcal{S}}, \tag{5}$$

with $\mathbf{N}^{\mathcal{S}} \triangleq \left[\mathbf{N}_{\mathrm{p}}^{\mathcal{S}}, \mathbf{N}_{\mathrm{d}}^{\mathcal{S}}\right] \in \mathbb{C}^{N\times K}$ and $\mathbf{X} \triangleq \left[\mathbf{X}_{\mathrm{p}}, \mathbf{X}_{\mathrm{d}}\right] \in \mathbb{C}^{U\times K}$. For the sake of future convenience, let us define the pilot and data index set as $\mathcal{K} \triangleq \mathcal{K}_{\mathrm{p}} \cup \mathcal{K}_{\mathrm{d}}$ with $\mathcal{K}_{\mathrm{p}} \triangleq \{1, 2, \ldots, K_{\mathrm{p}}\}$ and $\mathcal{K}_{\mathrm{d}} \triangleq \{K_{\mathrm{p}}+1, \ldots, K_{\mathrm{p}}+K_{\mathrm{d}}\}$.

From (3) and (4), the received pilot $\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}}$ is rewritten as

$$\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}} = \mathbf{A}(\boldsymbol{\theta}, \mathbf{r})\mathbf{V} + \mathbf{N}_{\mathrm{p}}^{\mathcal{S}}, \tag{6}$$

where $\mathbf{V} \triangleq \mathbf{Z}\mathbf{X} = \left[\mathbf{x}_1\mathbf{z}_1^{\mathrm{T}}, \mathbf{x}_2\mathbf{z}_2^{\mathrm{T}}, \ldots, \mathbf{x}_U\mathbf{z}_U^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{L\times K_{\mathrm{p}}}$ is the matrix composed of path gains and pilots.

III. Overview of the Proposed Algorithm

This section describes the overview of the proposed algorithm. The overall procedures of the proposed algorithm are illustrated in Fig. 3. As shown in the figure, the proposed algorithm mainly consists of two parts: the initial channel estimation part and subsequent JCDE part. The initial channel estimation part yields an accurate channel estimate to improve the convergence point of the subsequent JCDE algorithm. This initial estimation is composed of two stages to reduce the computational complexity. In the first stage, the angle and distance candidates from large-size polar grids are estimated by utilizing the SOMP algorithm. In the second stage, the pairing between the path candidates obtained in the first stage and corresponding UEs is performed via the 2D-OMP algorithm by using UE specific pilot sequences. The first and second stages for initial channel estimation are described in Section IV-A and Section IV-B, respectively.

In the subsequent JCDE process, the channel and data are jointly estimated via the EP algorithm with a deterministic model-based estimation approach, employing the initial channel estimate as the starting values. To exploit the near-field model structures, the beam-domain channel matrix $\mathbf{H} \in \mathbb{C}^{N\times U}$ is decomposed into a model-based estimate $\hat{\mathbf{S}} \in \mathbb{C}^{N\times U}$ and residual channel error $\mathbf{E} \in \mathbb{C}^{N\times U}$. $\mathbf{E}$ and $\mathbf{X}$ are jointly estimated by the EP algorithm, where the approximate joint posterior for $\mathbf{E}$ and $\mathbf{X}$ is calculated as described in Section V-C and V-D. The model-based estimate $\hat{\mathbf{S}}$ is determined by the initial channel estimate and adaptively updated in the algorithm iterations to further improve estimation performance as described in Section V-F. The performance of data and channel estimation in the JCDE algorithm heavily depends on the accuracy of the initial channel estimate. Therefore, creating an accurate initial channel estimate leads to a better convergence point in the JCDE algorithm, which is the main role of the initial channel estimation algorithm.

IV. Proposed Initial Channel Estimation

A. Angle and Distance Estimation

To leverage the near-field channel sparsity, the virtual channel representation in the polar-domain [7] is utilized with polar grids. The polar grids are designed by spatially quantizing the angle and distance domain into

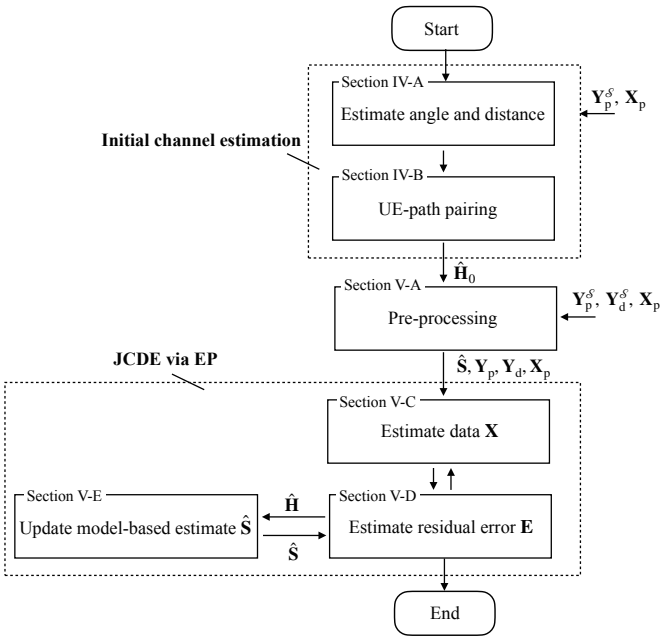


Fig. 3: Overview of the proposed algorithm.

$G_\theta G_r$ grid points as $\tilde{\boldsymbol{\theta}} \triangleq \{\tilde{\theta}_{g_\theta} | g_\theta \in \{1, \ldots, G_\theta\}\}$ and $\tilde{\mathbf{r}} \triangleq \{\tilde{r}_{g_r, g_\theta} | g_r \in \{1, \ldots, G_r\},\ g_\theta \in \{1, \ldots, G_\theta\}\}$ with $\tilde{\theta}_{g_\theta} \in [-\pi/2, \pi/2]$ and $\tilde{r}_{g_r, g_\theta} \in [0, \tilde{r}_{\max}]$[3]. Using the polar grids $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{r}}$, the polar-domain dictionary (*i.e.,* virtual array response matrix) is designed as

$$\tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{r}}) = \Big[\mathbf{a}(\tilde{\theta}_1, \tilde{r}_{1,1}), \ldots \mathbf{a}(\tilde{\theta}_1, \tilde{r}_{G_r,1}), \ldots \\ \ldots, \mathbf{a}(\tilde{\theta}_{G_\theta}, \tilde{r}_{1,G_\theta}), \ldots, \mathbf{a}(\tilde{\theta}_{G_\theta}, \tilde{r}_{G_r,G_\theta})\Big] \in \mathbb{C}^{N\times G_\theta G_r}. \tag{7}$$

From (6) and (7), the received pilot signal $\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}}$ is given by

$$\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}} \simeq \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{r}})\tilde{\mathbf{V}} + \mathbf{N}_{\mathrm{p}}^{\mathcal{S}}, \tag{8}$$

where $\tilde{\mathbf{V}} \in \mathbb{C}^{G_\theta G_r \times K_{\mathrm{p}}}$ is the row sparse matrix such that the number of nonzero rows is only L and other $G_\theta G_r - L$ rows are zero since the channel is composed of a total of L paths defined as in (1), with a sufficiently large number of grids, *i.e.,* $G_\theta G_r \gg L$. Equation (8) exactly holds only if there is no quantization errors in polar grids. In actual environments, however, it approximately holds due to the presence of quantization errors. Therefore, to compensate the quantization errors, we overestimate the number of paths ($\hat{L} > L$)[4]. To estimate $\hat{L}$ path candidates from $G_\theta G_r$ grids, the sparse reconstruction problem for $\tilde{\mathbf{V}}$ is formulated as

$$\begin{aligned} \underset{\tilde{\mathbf{V}}}{\text{minimize}} \quad & \left\|\mathbf{Y}_{\mathrm{p}}^{\mathcal{S}} - \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{r}})\tilde{\mathbf{V}}\right\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \|\tilde{\mathbf{V}}\|_{2,0} = \hat{L}, \end{aligned} \tag{9}$$

[3]In this paper, we design the polar grids using the near-field sampling method proposed in [9]. Using this sampling method, the maximum distance grid is computed as $\tilde{r}_{\max} = 2N^2\lambda\sqrt{\frac{0.001624}{1-\bar{\gamma}_{\mathrm{coh}}}}$ with a desired coherence value $\bar{\gamma}_{\mathrm{coh}}$, which is set to 0.6 in the numerical simulation in Section VI.

[4]This paper assumes that prior knowledge of the number of paths can be obtained from long-term statistics because the temporal variation in the number of paths is significantly slower than that of the channel. This assumption has been widely adopted in various studies [6], [22].

where $\|\tilde{\mathbf{V}}\|_{2,0}$ denotes the number of non-zero rows of $\tilde{\mathbf{V}}$.

The problem in (9) can be approximately solved by a compressed sensing algorithm for multiple measurement vectors (MMV) problems, *e.g.,* SOMP [7], because $\tilde{\mathbf{V}}$ has common support across its columns (*i.e.,* along the pilot dimension), meaning that the index set of nonzero elements is identical for all column vectors of $\tilde{\mathbf{V}}$. The complexity of SOMP at the t-th iteration in a naive implementation is $\mathcal{O}(G_r G_\theta N K_\mathrm{p} + Nt + Nt^2 + t^3)$. Its complexity can be further reduced by using the matrix inversion lemma (MIL) to $\mathcal{O}(G_r G_\theta K_\mathrm{p} t + Nt)$ [23], [24]. Solving the problem (9) yields the angle and distance candidates corresponding to the non-zero rows of $\tilde{\mathbf{V}}$, defined as $\check{\boldsymbol{\theta}} \triangleq \{\check{\theta}_l\}_{l=1}^{\hat{L}}$ and $\check{\mathbf{r}} \triangleq \{\check{r}_l\}_{l=1}^{\hat{L}}$.

B. UE-Path Pairing

The path candidate set $\{\check{\theta}_l, \check{r}_l\}_{l=1}^{\hat{L}}$ obtained in the first stage does not specify the association of individual paths with each UE. To estimate individual channels for each UE, the second stage performs UE-path pairing, where the estimated path candidates are associated with each user using UE-specific non-orthogonal pilot sequences $\{\mathbf{x}_{\mathrm{p},u}\}_{u=1}^{U}$. The usage of limited path candidates $\check{\boldsymbol{\theta}}$, $\check{\mathbf{r}}$, rather than large-size polar grids $\tilde{\boldsymbol{\theta}}$, $\tilde{\mathbf{r}}$ sampling the entire polar-domain, can lead to a complexity reduction. Using the path set $\{\check{\theta}_l, \check{r}_l\}_{l=1}^{\hat{L}}$, the polar-domain dictionary matrix is designed as

$$\check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}}) \triangleq \left[\mathbf{a}(\check{\theta}_1, \check{r}_1), \mathbf{a}(\check{\theta}_2, \check{r}_2), \dots \mathbf{a}(\check{\theta}_{\hat{L}}, \check{r}_{\hat{L}})\right] \in \mathbb{C}^{N \times \hat{L}}. \quad (10)$$

Reducing the size of the polar grids from $G_r G_\theta$ in (7) to $\hat{L} (\ll G_r G_\theta)$ in (10) can effectively lower the complexity in the following compressed sensing algorithm. Then, the channel vector for the u-th UE can be approximated with the polar-domain dictionary $\check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})$ as

$$\mathbf{h}_u^{\mathcal{S}} = \mathbf{A}(\boldsymbol{\theta}_u, \mathbf{r}_u)\mathbf{z}_u \simeq \check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})\check{\mathbf{z}}_u, \quad (11)$$

where $\check{\mathbf{z}}_u \in \mathbb{C}^{\hat{L} \times 1}$ is the virtual path gain vector.

From (11), the received pilot $\mathbf{Y}_\mathrm{p}^{\mathcal{S}}$ is approximated as

$$\mathbf{Y}_\mathrm{p}^{\mathcal{S}} \simeq \sum_{u=1}^{U} \check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})\check{\mathbf{z}}_u \mathbf{x}_{\mathrm{p},u}^{\mathrm{T}} + \mathbf{N}_\mathrm{p} = \check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})\check{\mathbf{Z}}\mathbf{X}_\mathrm{p} + \mathbf{N}_\mathrm{p}^{\mathcal{S}}, \quad (12)$$

with $\check{\mathbf{Z}} \triangleq \left[\check{\mathbf{z}}_1, \check{\mathbf{z}}_2, \dots, \check{\mathbf{z}}_U\right] \in \mathbb{C}^{\hat{L} \times U}$.

The equation (12) can be transformed into a 1D linear equation as $\mathbf{y}_\mathrm{p}^{\mathcal{S}} \simeq \boldsymbol{\Phi}_\mathrm{p}\check{\mathbf{z}}$ with $\mathbf{y}_\mathrm{p}^{\mathcal{S}} \triangleq \mathrm{vec}(\mathbf{Y}_\mathrm{p}^{\mathcal{S}})$, $\check{\mathbf{z}} \triangleq \mathrm{vec}(\check{\mathbf{Z}}) \in \mathbb{C}^{\hat{L}U \times 1}$ and $\boldsymbol{\Phi}_\mathrm{p} \triangleq \left(\mathbf{X}_\mathrm{p}^{\mathrm{T}} \otimes \check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})\right) \in \mathbb{C}^{NK_\mathrm{p} \times \hat{L}U}$. Although the estimation for $\check{\mathbf{z}}$ from the vectorized observation $\mathbf{y}_\mathrm{p}^{\mathcal{S}}$ can be simply addressed by various methods such as OMP [23], this significantly increases the complexity due to the large-size dictionary $\boldsymbol{\Phi}_\mathrm{p}$. Hence, to circumvent the high computational burden, the 2D signal representation in (12) is directly addressed without the vectorized 1D representation. Then, the sparse reconstruction problem for $\check{\mathbf{Z}}$ in (12) is formulated as

$$\begin{aligned} \underset{\check{\mathbf{Z}}}{\text{minimize}} \quad & \left\|\mathbf{Y}_\mathrm{p}^{\mathcal{S}} - \check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})\check{\mathbf{Z}}\mathbf{X}_\mathrm{p}\right\|_2^2 \\ \text{subject to} \quad & \|\check{\mathbf{z}}\|_0 = \hat{L}, \end{aligned} \quad (13)$$

with $\check{\mathbf{z}} \triangleq \mathrm{vec}(\check{\mathbf{Z}}) \in \mathbb{C}^{\hat{L}U \times 1}$.

Algorithm 1 Proposed channel estimation algorithm

Input: $\mathbf{Y}_\mathrm{p}^{\mathcal{S}},\ \mathbf{X}_\mathrm{p},\ \hat{L}$
Output: $\hat{\mathbf{H}}_0^{\mathcal{S}}, \{\hat{\boldsymbol{\theta}}_u, \hat{\mathbf{r}}_u, \hat{\mathbf{z}}_u, \}_{u=1}^{U}$

// First Stage - Angle and distance estimation
1: Calculate polar-domain dictionary $\tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{r}})$ in (7)
2: Estimate $\check{\boldsymbol{\theta}}, \check{\mathbf{r}}$ by solving (9) via SOMP with MIL [23]
// Second Stage - UE-path pairing
3: Calculate polar-domain dictionary $\check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})$ in (10)
4: Estimate $\{\hat{\boldsymbol{\theta}}_u, \hat{\mathbf{r}}_u, \hat{\mathbf{z}}_u\}_{u=1}^{U}$ by solving (13) via 2D-OMP [18]
5: Estimate channel matrix $\hat{\mathbf{H}}_0^{\mathcal{S}}$ in (14)

The optimization problem (13) is solved via a two-dimensional compressed sensing algorithm. The conventional method [9] tackles this problem with the large-size polar dictionary $\tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{r}})$ in (7) instead of $\check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})$ in (10) via the 2D-CoSaMP algorithm, which sacrifices estimation performance for complexity reduction compared to 2D-OMP [18]. In contrast, our proposed method solves the optimization problem (13) via the 2D-OMP algorithm using the small-size polar-domain dictionary $\check{\mathbf{A}}(\check{\boldsymbol{\theta}}, \check{\mathbf{r}})$ constructed by the path candidates $\{\check{\theta}_l, \check{r}_l\}_{l=1}^{\hat{L}}$ in the first stage. As a result, the proposed method possesses the prominent capability to overcome the conventional approach [9] while retaining comparable computational complexity. Detailed discussions regarding the complexity of the proposed algorithm are presented in Section VI.

Solving the problem (13) yields the estimated path gain vector $\hat{\mathbf{z}}_u \in \mathbb{C}^{\hat{L}_u \times 1}$, angle $\hat{\boldsymbol{\theta}}_u \in \mathbb{R}^{\hat{L}_u \times 1}$, and distance $\hat{\mathbf{r}}_u \in \mathbb{R}^{\hat{L}_u \times 1}$ corresponding to the non-zero elements of $\check{\mathbf{z}}_u \in \mathbb{C}^{\hat{L} \times 1}$, where $\hat{L}_u$ is the estimated number of paths for the u-th UE. Given the estimates, the initial channel estimate can be obtained as

$$\hat{\mathbf{H}}_0^{\mathcal{S}} = \left[\hat{\mathbf{h}}_1^{\mathcal{S}}, \dots, \hat{\mathbf{h}}_U^{\mathcal{S}}\right] \in \mathbb{C}^{N \times U}, \text{with } \hat{\mathbf{h}}_u^{\mathcal{S}} = \hat{\mathbf{A}}(\hat{\boldsymbol{\theta}}_u, \hat{\mathbf{r}}_u)\hat{\mathbf{z}}_u, \quad (14)$$

where $\hat{\mathbf{A}}(\hat{\boldsymbol{\theta}}_u, \hat{\mathbf{r}}_u) = \left[\mathbf{a}(\hat{\theta}_{u,1}, \hat{r}_{u,1}), \dots, \mathbf{a}(\hat{\theta}_{u,\hat{L}_u}, \hat{r}_{u,\hat{L}_u})\right] \in \mathbb{C}^{N \times \hat{L}_u}$ is the estimated array response. The proposed initial channel estimation method is summarized in Algorithm 1.

V. Proposed Joint Channel and Data Estimation

Given the initial estimates obtained from Algorithm 1, we aim to improve both the channel and data estimation performance while considering near-field properties. This section elaborates on the proposed JCDE algorithm using the initial channel estimate as the starting values.

A. Pre-processing for Channel and Data Estimation

1) Pre-processing for Channel Estimation: To exploit the channel sparsity, the received signal and channel matrix in the spatial-domain are transformed in the beam-domain as $\mathbf{Y} \triangleq \mathbf{D}_N \mathbf{Y}^{\mathcal{S}} \in \mathbb{C}^{N \times K}$ and $\mathbf{H} \triangleq \mathbf{D}_N \mathbf{H}^{\mathcal{S}} \in \mathbb{C}^{N \times U}$. Since the DFT matrix $\mathbf{D}_N$ is unitary, the transformation between the spatial and beam-domains can be efficiently performed via simple matrix multiplication unlike the polar-domain representation. Therefore, the matrix size remains unchanged after the transformation, resulting in no additional computational complexity in the subsequent iterative channel estimation and data

detection processes. However, as described in Section II-A, the near-field channel in the beam-domain has a cluster sparse structure due to energy leakage. Thus, to tackle this problem, the channel matrix $\mathbf{H}$ is first considered as the aggregation of the model-based estimate $\hat{\mathbf{S}} \in \mathbb{C}^{N\times U}$ and the residual channel estimation error $\mathbf{E} \triangleq \mathbf{H} - \hat{\mathbf{S}} \in \mathbb{C}^{N\times U}$, resulting in

$$\mathbf{H} = \hat{\mathbf{S}} + \mathbf{E}. \tag{15}$$

An initial value for the model-based estimate $\hat{\mathbf{S}}$ is determined with the proposed initial channel estimate $\hat{\mathbf{H}}_0^{\mathcal{S}}$ in (14) as $\hat{\mathbf{S}} = \mathbf{D}_N \hat{\mathbf{H}}_0^{\mathcal{S}}$, and it is adaptively updated based on the near-field model structure as described in Section V-F. As the residual error $\mathbf{E}$ is defined by subtracting the current estimate $\hat{\mathbf{S}}$ from the beam-domain channel $\mathbf{H}$ as in (15), this subtraction results in a sparser domain representation compared to the original beam-domain channel $\mathbf{H}$. The dominant path components are removed from $\mathbf{H}$ by $\hat{\mathbf{S}}$, facilitating the sparse matrix reconstruction by considering $\mathbf{E}$ (instead of $\mathbf{H}$) as the variable to be estimated by a Bayesian inference framework.

2) Pre-processing for Data Estimation: For low-complexity data estimation, the conventional methods based on the far-field assumption, such as [11]–[13], utilize MRC-based detectors, which are ineffective in the near-field scenario because the near-field channel has cluster sparsity due to energy leakage, and the leaked energy is correlated in the beam-domain. Although LMMSE-based detection methods such as [25], [26] are effective to deal with the correlation, these methods require matrix inversion with the size N, which is computationally expensive especially in XL-MIMO systems. To balance the computational complexity and detection performance, the array is virtually divided into multiple sub-arrays, and a sub-array-wise LMMSE-based detector is designed similarly to [27]. In contrast to [27], which assumes perfect channel state information (CSI), the proposed method considers the channel estimation error while jointly estimating data and channel, exploiting the near-field model structures.

Accordingly, the extra-large array with N antennas are partitioned into C sub-arrays, and the sub-array $c \in \mathcal{C} \triangleq \{1, 2, \ldots, C\}$ has N_c antennas satisfying $N = \sum_{c=1}^{C} N_c$. The received signals, residual channel errors, and model-based estimates can be also seen as $\mathbf{Y} = \left[\mathbf{Y}_1^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}}, \ldots, \mathbf{Y}_C^{\mathrm{T}}\right]^{\mathrm{T}}$, $\mathbf{E} = \left[\mathbf{E}_1^{\mathrm{T}}, \mathbf{E}_2^{\mathrm{T}}, \ldots, \mathbf{E}_C^{\mathrm{T}}\right]^{\mathrm{T}}$, and $\hat{\mathbf{S}} = \left[\hat{\mathbf{S}}_1^{\mathrm{T}}, \hat{\mathbf{S}}_2^{\mathrm{T}}, \ldots, \hat{\mathbf{S}}_C^{\mathrm{T}}\right]^{\mathrm{T}}$, with $\mathbf{Y}_c \in \mathbb{C}^{N_c\times K}$, $\mathbf{E}_c \in \mathbb{C}^{N_c\times U}$, and $\hat{\mathbf{S}}_c \in \mathbb{C}^{N_c\times U}$. The received signals $\mathbf{Y}$ and $\mathbf{Y}_c$ can then be rewritten as

$$\mathbf{Y} = \mathbf{E}\mathbf{X} + \hat{\mathbf{S}}\mathbf{X} + \mathbf{N}, \text{ with } \mathbf{Y}_c = \mathbf{E}_c\mathbf{X} + \hat{\mathbf{S}}_c\mathbf{X} + \mathbf{N}_c. \tag{16}$$

For convenience, let us define $\mathcal{N} \triangleq \{1, 2, \ldots, N\}$ as the antenna index set, and $\mathcal{N}_c \triangleq \left\{n_{c(1)}, n_{c(2)}, \ldots, n_{c(Nc)}\right\} \subset \mathcal{N}$ as the antenna index set at the c-th sub-array such that $\mathcal{N}_1 \cup \mathcal{N}_2 \cup \cdots \cup \mathcal{N}_C = \mathcal{N}$ and $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset,\ i \neq j \in \mathcal{C}$.

B. Bayesian Inference Formulation

Based on the linear observation in (16) with the deterministic variable $\hat{\mathbf{S}}$ and random variables $\mathbf{X}$ and $\mathbf{E}$, the likelihood function for $\mathbf{X}$ and $\mathbf{E}$ can be expressed as

$$p(\mathbf{Y}|\mathbf{E}, \mathbf{X}) = \prod_{n\in\mathcal{N}} \prod_{k\in\mathcal{K}} p(y_{n,k}|\bar{\mathbf{x}}_k, \bar{\mathbf{e}}_n), \tag{17}$$

where $p(y_{n,k}|\bar{\mathbf{e}}_n, \bar{\mathbf{x}}_k) = \mathcal{CN}((\bar{\mathbf{e}}_n + \bar{\mathbf{s}}_n)^{\mathrm{T}}\bar{\mathbf{x}}_k,\ \sigma^2)$ with $\bar{\mathbf{e}}_n = [e_{n,1}, \ldots, e_{n,U}]^{\mathrm{T}} \in \mathbb{C}^{U\times 1}$, $\bar{\mathbf{s}}_n = [\hat{s}_{n,1}, \ldots, \hat{s}_{n,U}]^{\mathrm{T}} \in \mathbb{C}^{U\times 1}$, and $\bar{\mathbf{x}}_k = [x_{1,k}, \ldots, x_{U,k}]^{\mathrm{T}} \in \mathbb{C}^{U\times 1}$.

Since each entry of $\mathbf{X}_{\mathrm{d}}$ is randomly selected from the QAM constellation point set $\mathcal{X}$, the prior $p(\mathbf{X})$ can be written as

$$p(\mathbf{X}) = \prod_{u\in\mathcal{U}} \prod_{k\in\mathcal{K}} p(x_{u,k}), \tag{18}$$

with $p(x_{u,k_d}) = \frac{1}{Q}\sum_{\mathcal{X}_i\in\mathcal{X}} \delta(x_{u,k_d} - \mathcal{X}_i),\ \forall k_d \in \mathcal{K}_d$ and $p(x_{u,k_p}) = \delta(x_{u,k_p} - [\mathbf{X}_{\mathrm{p}}]_{u,k_p}),\ \forall k_p \in \mathcal{K}_p$.

Although many conventional methods such as [12], [13] design the i.i.d. sparse prior for the beam-domain channel as $p(\mathbf{H}) = \prod_{n\in\mathcal{N}} \prod_{u\in\mathcal{U}} p(h_{n,u})$ (*e.g.,* BG prior), this modeling causes the model mismatch due to energy leakage effects in the near-field region. Therefore, we design the sparse prior for the residual channel error $\mathbf{E}$ instead of $\mathbf{H}$ as

$$p(\mathbf{E}; \boldsymbol{\Theta}) = \prod_{n\in\mathcal{N}} \prod_{u\in\mathcal{U}} p(e_{n,u}; \boldsymbol{\Theta}), \tag{19}$$

where $p(e_{n,u}; \boldsymbol{\Theta}) = \mathcal{CN}(0,\ \sigma_{n,u}^e)$ is Gaussian prior distribution with zero mean and variance $\sigma_{n,u}^e$, which is widely used for sparse representation in the sparse Bayesian learning (SBL) algorithm [28], where $\boldsymbol{\Theta} \triangleq \{\sigma_{n,u}^e\}_{n\in\mathcal{N},u\in\mathcal{U}}$ is the hyperparameter set to be optimized through the expectation maximization (EM) algorithm [28] as described in Section V-E.

From the likelihood in (17) and priors in (18), (19), the posterior can be written as

$$p(\mathbf{E}, \mathbf{X}|\mathbf{Y}; \boldsymbol{\Theta}) = p(\mathbf{Y}|\mathbf{E}, \mathbf{X})p(\mathbf{X})p(\mathbf{E}; \boldsymbol{\Theta})/p(\mathbf{Y}; \boldsymbol{\Theta}), \tag{20}$$

where $p(\mathbf{Y}; \boldsymbol{\Theta}) = \int_{\mathbf{E},\mathbf{X}} p(\mathbf{Y}, \mathbf{E}, \mathbf{X}; \boldsymbol{\Theta})$ is the marginal likelihood referred to as the evidence for parameter $\boldsymbol{\Theta}$. Our objective is to estimate $\mathbf{E}$, $\mathbf{X}$, and $\boldsymbol{\Theta}$ through the posterior and the evidence.

The estimator for $\boldsymbol{\Theta}$ by the type-II maximum likelihood method [29] is given as

$$\hat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\ p(\mathbf{Y}; \boldsymbol{\Theta}). \tag{21}$$

However, the calculation of the evidence $p(\mathbf{Y}; \boldsymbol{\Theta})$ is intractable due to the multidimensional integral for $\mathbf{X}$ and $\mathbf{E}$. Hence, we utilize the EM algorithm, which maximizes the evidence lower bound (ELBO) in each iteration, instead of directly maximizing the evidence [28]. Given $\boldsymbol{\Theta}^{(t)}$ at the t-th iteration, $\boldsymbol{\Theta}^{(t+1)}$ at the $(t+1)$-th iteration can be obtained as the following E-step and M-step:

$$\begin{aligned} &\text{E-step}: \mathcal{F}(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}) = \mathbb{E}_{p(\mathbf{E},\mathbf{X}|\mathbf{Y};\boldsymbol{\Theta}^{(t)})}[\ln p(\mathbf{Y}, \mathbf{E}, \mathbf{X}; \boldsymbol{\Theta})] + \mathsf{c}_0^{(t)}, \\ &\text{M-step}: \boldsymbol{\Theta}^{(t+1)} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\ \mathcal{F}(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}), \end{aligned} \tag{22}$$

where $\mathcal{F}(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)})$ is the ELBO with the constant value $\mathsf{c}_0^{(t)} = \mathbb{E}_{p(\mathbf{E},\mathbf{X}|\mathbf{Y};\boldsymbol{\Theta}^{(t)})}\left[\ln p(\mathbf{E}, \mathbf{X}|\mathbf{Y}; \boldsymbol{\Theta}^{(t)})\right]$.

Since E-step requires the calculation of a multidimensional integral that is computationally unreasonable, we approximate the posterior by $g^{(t)}(\mathbf{E}, \mathbf{X}|\mathbf{Y}) \simeq p(\mathbf{E}, \mathbf{X}|\mathbf{Y}; \boldsymbol{\Theta}^{(t)})$, using the EP algorithm. After the E-step, the maximization problem in (22) with the approximate posterior $g^{(t)}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ is solved, which is described in detail in Section V-E. The EP procedure continues until it reaches the maximum number of iterations

T. Finally, the last updated parameters at $t = T$ are used as the final estimates as $\hat{\mathbf{\Theta}} \triangleq \mathbf{\Theta}^{(T)}$, $\hat{\mathbf{E}} \triangleq \mathbb{E}_{g^{(T)}(\mathbf{E},\mathbf{X}|\mathbf{Y})}[\mathbf{E}]$, and $\hat{\mathbf{X}} \triangleq \mathbb{E}_{g^{(T)}(\mathbf{E},\mathbf{X}|\mathbf{Y})}[\mathbf{X}]$.

For notation simplicity, let us drop the iteration index t. The approximate posterior $g(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ is derived by KL minimization subject to a Gaussian distribution set $\mathbf{\Phi}$ as

$$\underset{g\in\mathbf{\Phi}}{\text{minimize}}\ \mathrm{KL}\left(p(\mathbf{E}, \mathbf{X}|\mathbf{Y}; \mathbf{\Theta})\|g(\mathbf{E}, \mathbf{X}|\mathbf{Y})\right), \tag{23}$$

where the approximate posterior $g(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ is designed as

$$g(\mathbf{E}, \mathbf{X}|\mathbf{Y}) = Z_g^{-1} Q^x(\mathbf{X})Q^e(\mathbf{E})B^x(\mathbf{X})B^e(\mathbf{E}), \tag{24}$$

where $Z_g = \int_{\mathbf{E},\mathbf{X}} Q^x(\mathbf{X})Q^e(\mathbf{E})B^x(\mathbf{X})B^e(\mathbf{E})$ is a normalizing constant, and $Q^x(\mathbf{X})$, $Q^e(\mathbf{E})$, $B^x(\mathbf{X})$, and $B^e(\mathbf{E})$ are the approximate factors such that $Q^x(\mathbf{X})Q^e(\mathbf{E}) \simeq p(\mathbf{Y}|\mathbf{E}, \mathbf{X})$, $B^x(\mathbf{X}) \simeq p(\mathbf{X})$, and $B^e(\mathbf{E}) \simeq p(\mathbf{E}; \mathbf{\Theta})$ subject to Gaussian distribution set $\mathbf{\Phi}$.

These approximate factors are designed as $Q^x(\mathbf{X}) = \prod_{c,u,k} q^x_{c,u,k}(x_{u,k})$, $Q^e(\mathbf{E}) = \prod_c \prod_{n_c,u,k} q^e_{n_c,u,k}(e_{n_c,u})$, $B^x(\mathbf{X}) = \prod_{u,k} b^x_{u,k}(x_{u,k})$, $B^e(\mathbf{E}) = \prod_c \prod_{n_c,u} b^e_{n_c,u}(e_{n_c,u})$, where $q^x_{c,u,k}(\cdot)$, $q^e_{n_c,u,k}(\cdot)$, $b^x_{u,k}(\cdot)$, and $b^e_{u,k}(\cdot)$ are the parameterized approximate functions defined as

$$q^x_{c,u,k}(x_{u,k}) \triangleq \exp\left(-|x_{u,k} - \hat{x}^q_{c,u,k}|^2/\xi^{q,x}_{c,u,k}\right), \tag{26a}$$

$$q^e_{n_c,u,k}(e_{n_c,u}) \triangleq \exp\left(-|e_{n_c,u} - \hat{e}^q_{n_c,u,k}|^2/\xi^{q,e}_{n_c,u,k}\right), \tag{26b}$$

$$b^x_{u,k}(x_{u,k}) \triangleq \exp\left(-|x_{u,k} - \hat{x}^b_{u,k}|^2/\xi^{b,x}_{u,k}\right), \tag{26c}$$

$$b^e_{n_c,u}(e_{n_c,u}) \triangleq \exp\left(-|e_{n_c,u} - \hat{e}^b_{n_c,u}|^2/\xi^{b,e}_{n_c,u}\right), \tag{26d}$$

where $\boldsymbol{\pi}^{q,x}_{c,u,k} \triangleq \left[\hat{x}^q_{c,u,k}, \xi^{q,x}_{c,u,k}\right]^{\mathrm{T}}$, $\boldsymbol{\pi}^{q,e}_{n_c,u,k} \triangleq \left[\hat{e}^q_{n_c,u,k}, \xi^{q,e}_{n_c,u,k}\right]^{\mathrm{T}}$, $\boldsymbol{\pi}^{b,x}_{u,k} \triangleq \left[\hat{x}^b_{u,k}, \xi^{b,x}_{u,k}\right]^{\mathrm{T}}$, and $\boldsymbol{\pi}^{b,e}_{n_c,u} \triangleq \left[\hat{e}^b_{n_c,u}, \xi^{b,e}_{n_c,u}\right]^{\mathrm{T}}$ are unknown parameters to be optimized by KL minimization.

Since the approximate posterior $g(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ in (24) is designed subject to Gaussian distribution set $\mathbf{\Phi}$, the marginalized approximate posterior $g(x_{u,k}|\mathbf{Y})$ and $g(e_{n_c,u}|\mathbf{Y})$ can be expressed as $g(x_{u,k}|\mathbf{Y}) = \mathcal{CN}(\hat{x}_{u,k}, \xi^x_{u,k})$ and $g(e_{n_c,u}|\mathbf{Y}) = \mathcal{CN}(\hat{e}_{n_c,u}, \xi^e_{n_c,u})$, where $\hat{x}_{u,k}$ and $\hat{e}_{n_c,u}$ are the posterior means, and $\xi^x_{u,k}$ and $\xi^e_{n_c,u}$ are the posterior variances.

Let $\mathbf{\Pi} \triangleq \{\boldsymbol{\pi}^{q,x}_{c,u,k},\ \boldsymbol{\pi}^{b,x}_{u,k},\ \boldsymbol{\pi}^{q,e}_{n_c,u,k},\ \boldsymbol{\pi}^{b,e}_{n_c,u}\}_{c\in\mathcal{C},n_c\in\mathcal{N},u\in\mathcal{U},k\in\mathcal{K}}$ denote an unknown parameter set. The optimal parameter set $\hat{\mathbf{\Pi}}$ is obtained by KL minimization in (23). However, the objective function cannot be expressed in closed-form because $\mathrm{KL}\left(p(\mathbf{E}, \mathbf{X}|\mathbf{Y}; \mathbf{\Theta})\|g(\mathbf{E}, \mathbf{X}|\mathbf{Y})\right)$ includes intractable integral with respect to the true posterior $p(\mathbf{E}, \mathbf{X}|\mathbf{Y}; \mathbf{\Theta})$. To tackle this, we set the target distribution $\hat{p}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ instead of the true posterior into the KL divergence in (23). The target distribution is designed by replacing a part of the true posterior with the approximate functions in (26a)-(26d) as described in the following sections. For the sake of notation for the design of the target distribution, the approximate distribution $l^x_{c,k}(\mathbf{E}_c, \bar{\mathbf{x}}_k) \simeq p(\mathbf{y}_{c,k}|\mathbf{E}_c, \bar{\mathbf{x}}_k)$ and $l^e_{n_c,k}(\bar{\mathbf{e}}_{n_c}, \bar{\mathbf{x}}_k) \simeq p(y_{n_c,k}|\bar{\mathbf{e}}_{n_c}, \bar{\mathbf{x}}_k)$ are expressed using (26a)-(26d) as

$$l^x_{c,k}(\mathbf{E}_c, \bar{\mathbf{x}}_k) \propto \prod_{u\in\mathcal{U}} q^x_{c,u,k}(x_{u,k}) \prod_{u\in\mathcal{U}} \prod_{n_c\in\mathcal{N}_c} q^e_{n_c,u,k}(e_{n_c,u}), \tag{27a}$$

$$l^e_{n_c,k}(\bar{\mathbf{e}}_{n_c}, \bar{\mathbf{x}}_k) \propto \prod_{u\in\mathcal{U}} \left\{q^x_{c,u,k}(x_{u,k})\right\}^{\frac{1}{N_c}} \prod_{u\in\mathcal{U}} q^e_{n_c,u,k}(e_{n_c,u}). \tag{27b}$$

To solve the KL minimization problem, the alternating optimization algorithm [30] is utilized, where a target parameter is optimized while the other parameters are fixed. In what follows, the estimation methods for $\{\boldsymbol{\pi}^{q,x}_{c,u,k}, \boldsymbol{\pi}^{b,x}_{u,k}\}$ and $\{\boldsymbol{\pi}^{b,e}_{n,u}, \boldsymbol{\pi}^{q,e}_{n,u,k}\}$ are described in Section V-C and V-D, respectively.

C. EP for Data Estimation

1) Update $\boldsymbol{\pi}^{q,x}_{c,u,k}$: While the parameter $\boldsymbol{\pi}^{q,x}_{c,u,k}$ in $q^x_{c,u,k}(x_{u,k})$ is updated, the other parameters $\mathbf{\Pi} \setminus \{\boldsymbol{\pi}^{q,x}_{c,u,k}\}$ are fixed as the tentative estimated values, that is, the KL minimization problem for $\boldsymbol{\pi}^{q,x}_{c,u,k}$ is formulated as

$$\underset{\boldsymbol{\pi}^{q,x}_{c,u,k}}{\text{minimize}}\ \mathrm{KL}\left(\hat{p}^{q,x}_{c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y})\|g(\mathbf{E}, \mathbf{X}|\mathbf{Y})\right), \tag{28}$$

where $\hat{p}^{q,x}_{c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ is the target distribution for $\boldsymbol{\pi}^{q,x}_{c,u,k}$, which is designed using $l^x_{c,k}(\mathbf{E}_c, \bar{\mathbf{x}}_k)$ in (27a) as

$$\hat{p}^{q,x}_{c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y}) = C^{q,x}_{c,k}\ p(\mathbf{y}_{c,k}|\mathbf{E}_c, \bar{\mathbf{x}}_k) \prod_{(c',k')\in\mathcal{C}\times\mathcal{K}\setminus(c,k)} \underbrace{l^x_{c',k'}(\mathbf{E}_{c'}, \mathbf{x}_{k'})}_{\simeq p(\mathbf{y}_{c',k}|\mathbf{E}_{c'}, \mathbf{x}_{k'})} \underbrace{B^x(\mathbf{X})B^e(\mathbf{E})}_{\simeq p(\mathbf{X})p(\mathbf{E};\mathbf{\Theta})}, \tag{29}$$

where $C^{q,x}_{c,k}$ is a normalizing constant.

Let $\mathcal{L}^{q,x}_{c,u,k}(\boldsymbol{\pi}^{q,x}_{c,u,k}) \triangleq \mathrm{KL}\left(\hat{p}^{q,x}_{c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y})\|g(\mathbf{E}, \mathbf{X}|\mathbf{Y})\right)$ denote the objective function in (28), resorting to

$$\mathcal{L}^{q,x}_{c,u,k} = \ln Z_g - \mathbb{E}_{\hat{p}^{q,x}_{c,k}(x_{u,k}|\mathbf{Y})}\left[\ln q^x_{c,u,k}(x_{u,k})\right] + \text{const.} \tag{30}$$

Since the objective function $\mathcal{L}^{q,x}_{c,u,k}(\boldsymbol{\pi}^{q,x}_{c,u,k})$ is convex with respect to $\boldsymbol{\pi}^{q,x}_{c,u,k}$, the necessary and sufficient condition for the global optimal , *i.e.,* $\partial\mathcal{L}^{q,x}_{c.u,k}/\partial\boldsymbol{\pi}^{q,x}_{c,u,k} = \mathbf{0}$, is equivalent to

$$g(x_{u,k}|\mathbf{Y}) = \mathrm{proj}_{\mathbf{\Phi}}\left[\hat{p}^{q,x}_{c,k}(x_{u,k}|\mathbf{Y})\right], \tag{31}$$

where $\mathrm{proj}_{\mathbf{\Phi}}\left[p(x)\right] \triangleq \mathcal{CN}\left(\mathbb{E}_{p(x)}[x],\ \mathbb{V}_{p(x)}[x]\right)$ is the projection operator onto Gaussian distribution set $\mathbf{\Phi}$, which indicates the moment matching, *i.e.,* the first and second moments of distribution $p(x)$ matches those of the target distribution.

The marginalized approximate posterior $g(x_{u,k}|\mathbf{Y}) = \int_{\mathbf{E},\mathbf{X}\setminus x_{u,k}} g(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ in (31) is written as

$$g(x_{u,k}|\mathbf{Y}) \propto q^x_{c,u,k}(x_{u,k})v^x_{c,u,k}(x_{u,k}), \tag{32}$$

with $v^x_{c,u,k}(x_{u,k}) \triangleq b^x_{u,k}(x_{u,k}) \prod_{c'\in\mathcal{C}\setminus c} q^x_{c',u,k}(x_{u,k})$, which can also be represented as

$$v^x_{c,u,k}(x_{u,k}) = C^{v,x}_{c,u,k} \exp\left(-|x_{u,k} - \hat{x}^v_{c,u,k}|^2/\xi^{v,x}_{c,u,k}\right), \tag{33}$$

with the normalizing constant $C^{v,x}_{c,u,k}$.

The marginalized target distribution $\hat{p}^{q,x}_{c,k}(x_{u,k}|\mathbf{Y}) = \int_{\mathbf{E},\mathbf{X}\setminus x_{u,k}} \hat{p}^{q,x}_{c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ in (31) is written as

$$\hat{p}^{q,x}_{c,k}(x_{u,k}|\mathbf{Y}) = \bar{p}^x_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})v^x_{c,u,k}(x_{u,k}), \tag{34}$$

where $\bar{p}^x_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})$ is the conditional probability distribution defined in (35) at the top of next page along with $v^e_{n_c,u,k}(e_{n_c,u})$, and $\tilde{v}^e_{c,u,k}(\mathbf{e}_{c,u})$, with $\bar{C}^x_{c,u,k}$ being the normalizing constant and

$$\hat{\mathbf{e}}^v_{c,u,k} = \left[\hat{e}^v_{n_{c(1)},u,k}, \hat{e}^v_{n_{c(2)},u,k}, \ldots, \hat{e}^v_{n_{c(N_c)},u,k}\right]^{\mathrm{T}} \in \mathbb{C}^{N_c\times 1},$$

$$\mathbf{\Xi}^{v,e}_{c,u,k} = \mathrm{diag}\left(\xi^{v,e}_{n_{c(1)},u,k}, \xi^{v,e}_{n_{c(2)},u,k}, \ldots, \xi^{v,e}_{n_{c(N_c)},u,k}\right) \in \mathbb{R}^{N_c\times N_c}.$$

$$\bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k}) \triangleq \bar{C}^{x}_{c,u,k} \int_{\mathbf{E}_c,\bar{\mathbf{x}}_k\setminus x_{u,k}} p(\mathbf{y}_{c,k}|\mathbf{E}_c,\bar{\mathbf{x}}_k) \prod_{u'\in\mathcal{U}\setminus u} v^{x}_{c,u',k}(x_{u',k}) \prod_{u'\in\mathcal{U}} \prod_{n_c\in\mathcal{N}_c} v^{e}_{n_c,u',k}(e_{n_c,u'}), \tag{35}$$

$$v^{e}_{n_c,u,k}(e_{n_c,u}) \triangleq b^{e}_{n_c,u}(e_{n_c,u}) \prod_{k'\in\mathcal{K}\setminus k} q^{e}_{n_c,u,k'}(e_{n_c,u}) = C^{v,e}_{n_c,u,k} \exp\left(-|e_{n_c,u}-\hat{e}^{v}_{n_c,u,k}|^2/\xi^{v,e}_{n_c,u,k}\right), \tag{36}$$

$$\tilde{v}^{e}_{c,u,k}(\mathbf{e}_{c,u}) \triangleq \prod_{n_c\in\mathcal{N}_c} v^{e}_{n_c,u,k}(e_{n_c,u}) \propto \exp\left\{-(\mathbf{e}_{c,u}-\hat{\mathbf{e}}^{v}_{c,u,k})^{\mathrm{H}} \boldsymbol{\Xi}^{e,v-1}_{c,u,k} (\mathbf{e}_{c,u}-\hat{\mathbf{e}}^{v}_{c,u,k})\right\}. \tag{37}$$

From the conditional distribution $\bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})$ in (35), the mean $\tilde{\mathbf{y}}^{x}_{c,u,k} \triangleq \mathbb{E}_{\bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})}[\mathbf{y}_{c,k}]$ and covariance $\boldsymbol{\Omega}^{x}_{c,k} \triangleq \mathbb{E}_{\bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})}\left[(\mathbf{y}_{c,k}-\tilde{\mathbf{y}}^{x}_{c,u,k})(\mathbf{y}_{c,k}-\tilde{\mathbf{y}}^{x}_{c,u,k})^{\mathrm{H}}\right]$, can be calculated as

$$\tilde{\mathbf{y}}^{x}_{c,u,k} = \mathbf{y}_{c,k} - \sum_{u'\in\mathcal{U}\setminus u} \hat{\mathbf{h}}^{v}_{c,u',k}\hat{x}^{v}_{c,u',k}, \tag{38a}$$

$$\boldsymbol{\Omega}^{x}_{c,k} = \sum_{u'\in\mathcal{U}} \Big\{\xi^{v,x}_{c,u',k}\hat{\mathbf{h}}^{v}_{c,u',k}\hat{\mathbf{h}}^{v\mathrm{H}}_{c,u',k} + (\xi^{v,x}_{c,u',k}+|\hat{x}^{v}_{c,u',k}|^2)\boldsymbol{\Xi}^{v,e}_{c,u',k}\Big\} + \sigma^2\mathbf{I}_{N_c}, \tag{38b}$$

with $\hat{\mathbf{h}}^{v}_{c,u',k} \triangleq \hat{\mathbf{e}}^{v}_{c,u',k} + \hat{\mathbf{s}}_{c,u'}$.

Substituting (32) and (34) into (31), the approximate function $q^{x}_{c,u,k}(x_{u,k})$ can be obtained as

$$q^{x}_{c,u,k}(x_{u,k}) \propto \frac{\mathrm{proj}_{\boldsymbol{\Phi}}\left[\bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})v^{x}_{c,u,k}(x_{u,k})\right]}{v^{x}_{c,u,k}(x_{u,k})}. \tag{39}$$

Under large system conditions with CLT, the conditional distribution can be approximated as $\bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k}) \simeq \mathcal{CN}(\tilde{\mathbf{y}}^{x}_{c,u,k},\boldsymbol{\Omega}^{x}_{c,k})$. Thus, the approximate function can be expressed as $q^{x}_{c,u,k}(x_{u,k}) \propto \bar{p}^{x}_{c,u,k}(\mathbf{y}_{c,k}|x_{u,k})$ with the mean and variance calculated as

$$\hat{x}^{q}_{c,u,k} = \hat{\mathbf{h}}^{v\mathrm{H}}_{c,u,k}(\boldsymbol{\Omega}^{x}_{c,k})^{-1}\tilde{\mathbf{y}}^{x}_{c,u,k}/\gamma^{x}_{c,u,k}, \tag{40a}$$

$$\xi^{q,x}_{c,u,k} = 1/\gamma^{x}_{c,u,k} - \xi^{v,x}_{c,u,k}, \tag{40b}$$

with $\gamma^{x}_{c,u,k} = \hat{\mathbf{h}}^{v\mathrm{H}}_{c,u,k}(\boldsymbol{\Omega}^{x}_{c,k})^{-1}\hat{\mathbf{h}}^{v}_{c,u,k}$. The calculation of $\tilde{\mathbf{y}}^{x}_{c,u,k}$ in (38a) corresponds to a soft interference cancellation (Soft-IC) [11] using data replicas $\{\hat{x}^{v}_{c,u',k}\}_{u'\in\mathcal{U}\setminus u}$ and channel replicas $\{\hat{\mathbf{h}}^{v}_{c,u',k}\}_{u'\in\mathcal{U}}$.

Unlike the conventional MRC-based detections [11]–[13], the LMMSE-based detection for each sub-array in (40a) can deal with the correlation between the leaked energy in the beam-domain owing to the whitening operation by $(\boldsymbol{\Omega}^{x}_{c,k})^{-1}$. The correlation within the c-th sub-array, corresponding to the off-diagonal elements of $\boldsymbol{\Omega}^{x}_{c,k}$, increases as the leaked energy increases because the off-diagonal elements is computed by the weighted summation of the outer product of channel replicas $\hat{\mathbf{h}}^{v}_{c,u',k}\hat{\mathbf{h}}^{v\mathrm{H}}_{c,u',k}$ in (38b). As illustrated in Fig. 2, the energy leakage effects become prominent as the distance between the BS and the UE decrease due to the near-field effects. Therefore, when the number of antennas per sub-array N_c is small, including conventional MRC-based detections ($N_c = 1$), each sub-array cannot address the correlation caused by the energy leakage. The number of antennas per sub-array N_c should be selected by considering the trade-off between data detection performance and its computational complexity, which are evaluated through numerical simulations in Section VI.

2) Update $\boldsymbol{\pi}^{b,x}_{u,k}$: The KL minimization problem for $\boldsymbol{\pi}^{b,x}_{u,k}$ is formulated as

$$\underset{\boldsymbol{\pi}^{b,x}_{u,k}}{\mathrm{minimize}}\ \mathrm{KL}\left(\hat{p}^{b,x}_{u,k}(\mathbf{E},\mathbf{X}|\mathbf{Y})\|g(\mathbf{E},\mathbf{X}|\mathbf{Y})\right), \tag{41}$$

where $\hat{p}^{b,x}_{u,k}(\mathbf{E},\mathbf{X}|\mathbf{Y})$ is the target distribution defined as

$$\hat{p}^{b,x}_{u,k}(\mathbf{E},\mathbf{X}|\mathbf{Y}) = C^{b,x}_{u,k}\, p(x_{u,k}) \prod_{(u',k')\in\mathcal{U}\times\mathcal{K}\setminus(u,k)} \underbrace{b^{x}_{u',k'}(x_{u',k'})}_{\simeq p(x_{u',k'})} \underbrace{Q^{x}(\mathbf{X})Q^{e}(\mathbf{E})}_{\simeq p(\mathbf{Y}|\mathbf{E},\mathbf{X})} \underbrace{B^{e}(\mathbf{E})}_{\simeq p(\mathbf{E};\boldsymbol{\Theta})} \tag{42}$$

where $C^{b,x}_{u,k}$ is a normalizing constant.

Similar to the derivation of (31), the optimal condition for $\boldsymbol{\pi}^{b,x}_{u,k}$ is derived as

$$g(x_{u,k}|\mathbf{Y}) = \mathrm{proj}_{\boldsymbol{\Phi}}\left[\hat{p}^{b,x}_{u,k}(x_{u,k}|\mathbf{Y})\right], \tag{43}$$

where $\hat{p}^{b,x}_{u,k}(x_{u,k}|\mathbf{Y}) = \int_{\mathbf{E},\mathbf{X}\setminus x_{u,k}} \hat{p}^{b,x}_{u,k}(\mathbf{E},\mathbf{X}|\mathbf{Y})$ is the marginalized target distribution calculated as

$$\hat{p}^{b,x}_{u,k}(x_{u,k}|\mathbf{Y}) \propto\ p(x_{u,k}) \prod_{c'\in\mathcal{C}} q^{x}_{c',u,k}(x_{u,k}), \tag{44}$$

with the approximate function multiplied over the sub-array direction, $\prod_{c'\in\mathcal{C}} q^{x}_{c',u,k}(x_{u,k})$, calculated as

$$\prod_{c'\in\mathcal{C}} q^{x}_{c',u,k}(x_{u,k}) \propto \exp\left(-|x_{u,k}-\hat{x}^{q}_{u,k}|^2/\xi^{q,x}_{u,k}\right), \tag{45}$$

$$\hat{x}^{q}_{u,k} = \xi^{q,x}_{u,k}\left(\sum_{c'\in\mathcal{C}} \frac{\hat{x}^{q}_{c',u,k}}{\xi^{q,x}_{c',u,k}}\right),\ \xi^{q,x}_{u,k} = \left(\sum_{c'\in\mathcal{C}} \frac{1}{\xi^{q,x}_{c',u,k}}\right)^{-1}. \tag{46}$$

Note that combining the mean $\{\hat{x}^{q}_{c,u,k}\}_{c\in\mathcal{C}}$ and variance $\{\xi^{q,x}_{c,u,k}\}_{c\in\mathcal{C}}$ over the sub-array direction $c \in \mathcal{C}$, as written in (46), leads to further improvements for data detection owing to the spatial diversity. Substituting (44) into (43), the approximate posterior $g(x_{u,k}|\mathbf{Y})$ can be written as

$$g(x_{u,k}|\mathbf{Y}) \propto \mathrm{proj}_{\boldsymbol{\Phi}}\left[p(x_{u,k}) \prod_{c'\in\mathcal{C}} q^{x}_{c',u,k}(x_{u,k})\right]. \tag{47}$$

The approximate posterior mean $\hat{x}_{u,k}$ and variance $\xi^{x}_{u,k}$ of $g(x_{u,k}|\mathbf{Y})$ can be derived using the MMSE denoiser function

$\eta(\cdot)$ [31], which is designed based on the prior for QAM constellation $p(x_{u,k})$ in (18). Then, the posterior mean and variance is expressed as $\hat{x}_{u,k} = \eta(\hat{x}^q_{u,k}, \xi^{q,x}_{u,k}) \triangleq \mathbb{E}_{g(x_{u,k}|\mathbf{Y})}[x_{u,k}]$ and $\xi^x_{u,k} = \xi^{q,x}_{u,k} \frac{\partial \eta(\hat{x}^q_{u,k}, \xi^{q,x}_{u,k})}{\partial \hat{x}^q_{u,k}}$, which can be calculated as

$$\hat{x}_{u,k} = C^{g,x}_{u,k} \sum_{\mathcal{X}_q \in \mathcal{X}} \mathcal{X}_q \exp\left(-|\mathcal{X}_q - \hat{x}^q_{u,k}|^2 / \xi^{q,x}_{u,k}\right), \tag{48a}$$

$$\xi^x_{u,k} = C^{g,x}_{u,k} \sum_{\mathcal{X}_q \in \mathcal{X}} |\mathcal{X}_q|^2 \exp\left(-|\mathcal{X}_q - \hat{x}^q_{u,k}|^2 / \xi^{q,x}_{u,k}\right) - |\hat{x}_{u,k}|^2, \tag{48b}$$

with $(C^{g,x}_{u,k})^{-1} = \sum_{\mathcal{X}_q \in \mathcal{X}} \exp\left(-|\mathcal{X}_q - \hat{x}^q_{u,k}|^2 / \xi^{q,x}_{u,k}\right)$.

From (32), $v^x_{c,u,k}(x_{u,k})$ can be updated as

$$v^x_{c,u,k}(x_{u,k}) \propto \; g(x_{u,k}|\mathbf{Y}) / q^x_{c,u,k}(x_{u,k}), \tag{49}$$

where the associated mean and variance are given by

$$\hat{x}^v_{c,u,k} = \xi^{v,x}_{c,u,k} \left(\frac{\hat{x}_{u,k}}{\xi^x_{u,k}} - \frac{\hat{x}^q_{c,u,k}}{\xi^{q,x}_{c,u,k}} \right), \; \xi^{v,x}_{c,u,k} = \left(\frac{1}{\xi^x_{u,k}} - \frac{1}{\xi^{q,x}_{c,u,k}} \right)^{-1}. \tag{50}$$

As shown in the Soft-IC process in (38a)-(38b), $\hat{x}^v_{c,u,k}$ and $\xi^{v,x}_{c,u,k}$ in (50) are used as soft replicas instead of $\hat{x}_{u,k}$ and $\xi^x_{u,k}$ in (48a)-(48b) in order to suppress the self-noise feedback in the algorithm iterations [32]. In conventional JCDE algorithms [10], [11], the self-feedback suppression is performed before the denoising process in (48a)-(48b) by generating antenna-wise extrinsic values based on BP rules. Hence, the complexity of the denoising process is $\mathcal{O}(QNUK_{\rm d})$. In contrast, the proposed method can reduce the complexity in the denoising process as $\mathcal{O}(QUK_{\rm d})$, since the extrinsic values $\hat{x}^v_{c,u,k}$ and $\xi^{v,x}_{c,u,k}$ are generated after the denoising process in (50).

Finally, from (32), $b^x_{u,k}(x_{u,k})$ can be updated as

$$b^x_{u,k}(x_{u,k}) \; \propto \; g(x_{u,k}|\mathbf{Y}_{u,k}) \Big/ \prod_{c' \in \mathcal{C}} q^x_{c',u,k}(x_{u,k}), \tag{51}$$

with the mean $\hat{x}^b_{u,k}$ and variance $\xi^{b,x}_{u,k}$ of $b^x_{u,k}(x_{u,k})$ being

$$\hat{x}^b_{u,k} = \xi^{b,x}_{u,k} \left(\frac{\hat{x}_{u,k}}{\xi^x_{u,k}} - \frac{\hat{x}^q_{u,k}}{\xi^{q,x}_{u,k}} \right), \; \xi^{b,x}_{u,k} = \left(\frac{1}{\xi^x_{u,k}} - \frac{1}{\xi^{q,x}_{u,k}} \right)^{-1}. \tag{52}$$

D. EP for Residual Channel Error Estimation

1) Update $\boldsymbol{\pi}^{q,e}_{n_c,u,k}$: For $\boldsymbol{\pi}^{q,e}_{n_c,u,k}$, we minimize

$$\underset{\boldsymbol{\pi}^{q,e}_{n_c,u,k}}{\text{minimize}} \; \text{KL}\left(\hat{p}^{q,e}_{n_c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y}) \| g(\mathbf{E}, \mathbf{X}|\mathbf{Y}) \right), \tag{53}$$

where $\hat{p}^{q,e}_{n_c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ is the target distribution designed as

$$\hat{p}^{q,e}_{n_c,k}(\mathbf{E}, \mathbf{X}|\mathbf{Y}) = C^{q,e}_{n_c,k} \; p(y_{n_c,k}|\bar{\mathbf{e}}_{n_c}, \bar{\mathbf{x}}_k) \prod_{(n'_c,k') \in \mathcal{N} \times \mathcal{K} \setminus (n_c,k)} \underbrace{l^e_{n'_c,k'}(\bar{\mathbf{e}}_{n'_c}, \mathbf{x}_{k'})}_{\simeq p(y_{n'_c,k'}|\bar{\mathbf{e}}_{n'_c}, \mathbf{x}_{k'})} \underbrace{B^x(\mathbf{X})}_{\simeq p(\mathbf{X})} \underbrace{B^e(\mathbf{E})}_{\simeq p(\mathbf{E};\boldsymbol{\Theta})}, \tag{54}$$

where $C^{q,e}_{n_c,k}$ is a normalizing constant.

Through the same procedure as the derivation of $q^x_{c,u,k}(x_{u,k})$ in (39), the mean and variance of approximate function $q^e_{n_c,u,k}(e_{n_c,u})$ are obtained as

$$\hat{e}^q_{n_c,u,k} = \frac{\hat{x}^{w*}_{n_c,u,k} \tilde{y}^e_{n_c,u,k}}{|\hat{x}^w_{n_c,u,k}|^2}, \; \xi^{q,e}_{n_c,u,k} = \frac{\phi^e_{n_c,u,k}}{|\hat{x}^w_{n_c,u,k}|^2}, \tag{55}$$

with

$$\tilde{y}^e_{n_c,u,k} = y_{n_c,k} - \sum_{u' \in \mathcal{U} \setminus u} \hat{x}^w_{c,u',k} \hat{e}^v_{n_c,u',k} - \sum_{u' \in \mathcal{U}} \hat{x}^w_{n_c,u',k} \hat{s}_{n_c,u'}, \tag{56a}$$

$$\phi^e_{n_c,u,k} = \sum_{u' \in \mathcal{U}} \left(|\hat{e}^v_{n_c,u',k}|^2 + |\hat{s}_{n_c,u'}|^2 + \xi^{v,e}_{c,u',k} \right) \xi^{w,x}_{n_c,u',k} + \sum_{u' \in \mathcal{U} \setminus u} \xi^{v,e}_{n_c,u',k} |\hat{x}^w_{c,u',k}|^2 + \sigma^2, \tag{56b}$$

$$\hat{x}^w_{c,u,k} = \xi^{w,x}_{c,u,k} \left(\hat{x}^x_{u,k} (\xi^x_{u,k})^{-1} - \hat{x}^q_{c,u,k} (N_c \xi^{q,x}_{c,u,k})^{-1} \right), \tag{56c}$$

$$\xi^{w,x}_{c,u,k} = \left((\xi^x_{u,k})^{-1} - (N_c \xi^{q,x}_{c,u,k})^{-1} \right)^{-1}. \tag{56d}$$

2) Update $\boldsymbol{\pi}^{b,e}_{n_c,u}$: For $\boldsymbol{\pi}^{b,e}_{n_c,u}$, we have

$$\underset{\boldsymbol{\pi}^{b,e}_{n_c,u}}{\text{minimize}} \; \text{KL}\left(\hat{p}^{b,e}_{n_c,u}(\mathbf{E}, \mathbf{X}|\mathbf{Y}) \| g(\mathbf{E}, \mathbf{X}|\mathbf{Y}) \right), \tag{57}$$

where $\hat{p}^{b,e}_{n_c,u}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ is the target distribution designed as

$$\hat{p}^{b,e}_{n_c,u}(\mathbf{E}, \mathbf{X}|\mathbf{Y}) = C^{b,e}_{n_c,u} \; p(e_{n_c,u}; \boldsymbol{\Theta}) \cdot \prod_{(n'_c,u') \in \mathcal{N} \times \mathcal{U} \setminus (n_c,u)} \underbrace{b^e_{n'_c,u'}(e_{n'_c,u'})}_{\simeq p(e_{n_c,u};\boldsymbol{\Theta})} \underbrace{Q^x(\mathbf{X}) Q^e(\mathbf{E})}_{\simeq p(\mathbf{Y}|\mathbf{E},\mathbf{X})} \underbrace{B^x(\mathbf{X})}_{\simeq p(\mathbf{X})}, \tag{58}$$

where $C^{b,e}_{n_c,u}$ is a normalizing constant.

Following the same methodology used to derive $g(x_{u,k}|\mathbf{Y})$ in (47), the approximate posterior $g(e_{n_c,u}|\mathbf{Y})$ are derived as

$$g(e_{n_c,u}|\mathbf{Y}) \propto \text{proj}_{\boldsymbol{\Phi}} \left[p(e_{n_c,u}; \boldsymbol{\Theta}) \prod_{k' \in \mathcal{K}} q^e_{n_c,u,k'}(e_{n_c,u}) \right], \tag{59}$$

where the mean and variance of $g(e_{n_c,u}|\mathbf{Y})$ can be calculated based on the prior distribution $p(e_{n_c,u}; \boldsymbol{\Theta})$ in (19) as

$$\hat{e}_{n_c,u} = \frac{\sigma^e_{n_c,u} \hat{e}^q_{n_c,u}}{\sigma^e_{n_c,u} + \xi^{q,e}_{n_c,u}}, \; \xi^e_{n_c,u} = \left(\frac{1}{\sigma^e_{n_c,u}} + \frac{1}{\xi^{q,e}_{n_c,u}} \right)^{-1}, \tag{60}$$

with

$$\hat{e}^q_{n_c,u} = \xi^{q,e}_{n_c,u} \left(\sum_{k' \in \mathcal{K}} \frac{\hat{e}^q_{n_c,u,k'}}{\xi^{q,e}_{n_v,u,k'}} \right), \xi^{q,e}_{n_c,u} = \left(\sum_{k' \in \mathcal{K}} \frac{1}{\xi^{q,e}_{n_c,u,k'}} \right)^{-1}. \tag{61}$$

Similarly, the approximate function $v^e_{n_c,u,k}(e_{n_c,u})$ can be derived in the same manner as (49):

$$v^e_{n_c,u,k}(e_{n_c,u}) \propto g(e_{n_c,u}|\mathbf{Y}) / q^e_{n_c,u,k}(e_{n_c,u}), \tag{62}$$

from which the mean and variance are respectively given by

$$\hat{e}^v_{n_c,u,k} = \xi^{v,e}_{n_c,u,k} \left((\xi^e_{n_c,u})^{-1} \hat{e}_{n_c,u} - (\xi^{q,e}_{n_c,u,k})^{-1} \hat{e}^q_{n_c,u,k} \right), \tag{63a}$$

$$\xi^{v,e}_{n_c,u,k} = \left((\xi^e_{n_c,u})^{-1} - (\xi^{q,e}_{n_c,u,k})^{-1} \right)^{-1}. \tag{63b}$$

Finally, the approximate function $b^e_{n_c,u}(e_{n_c,u})$ is obtained in a similar way as the derivation of (51) as

$$b^e_{n_c,u}(e_{n_c,u}) \propto g(e_{n_c,u}|\mathbf{Y}) \Big/ \prod_{k' \in \mathcal{K}} q^e_{n_c,u,k'}(e_{n_c,u}), \tag{64}$$

with the mean and variance of $b^e_{n_c,u}(e_{n_c,u})$ being

$$\hat{e}^b_{n_c,u} = \xi^{b,e}_{n_c,u} \left(\frac{\hat{e}_{n_c,u}}{\xi^e_{n_c,u}} - \frac{\hat{e}^q_{n_c,u}}{\xi^{q,e}_{n_c,u}} \right), \; \xi^{b,e}_{n_c,u} = \left(\frac{1}{\xi^e_{n_c,u}} - \frac{1}{\xi^{q,e}_{n_c,u}} \right)^{-1}. \tag{65}$$

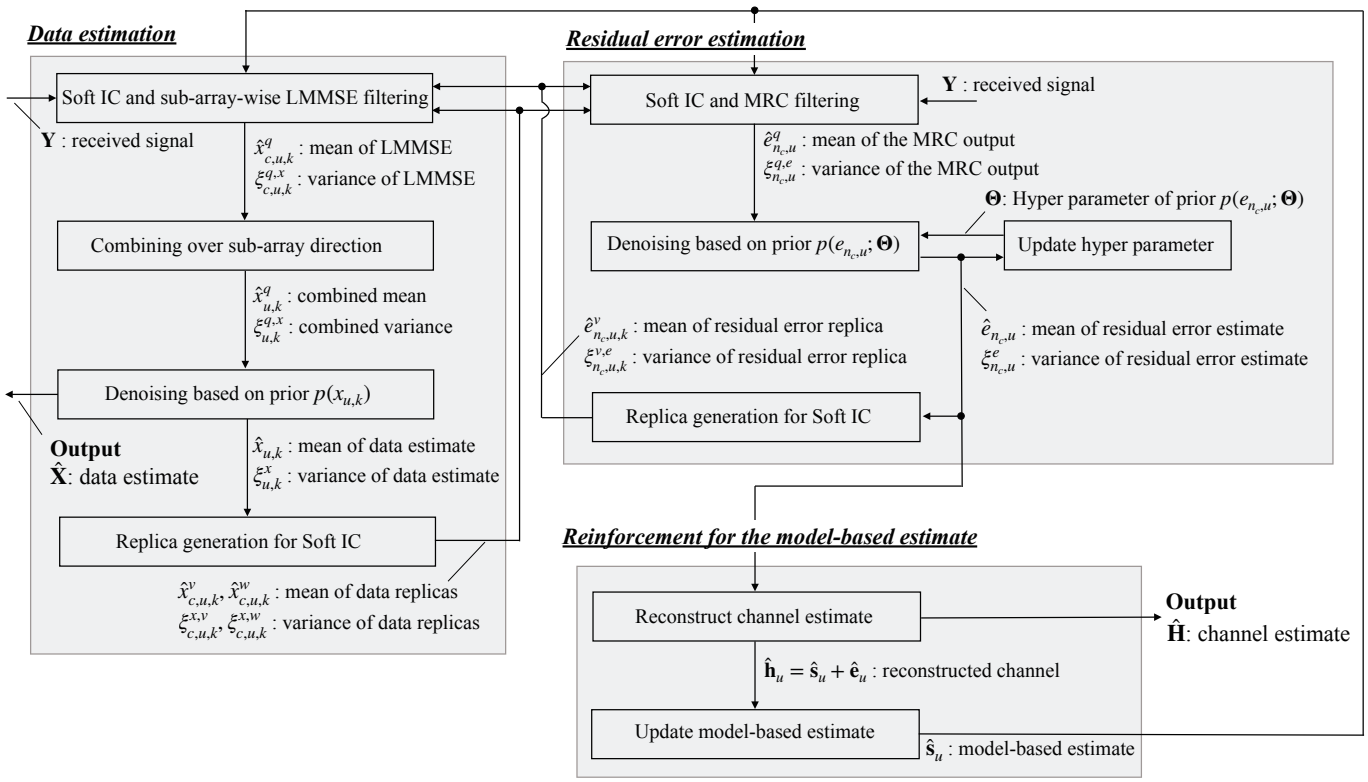


Fig. 4: Framework of the proposed JCDE algorithm.

E. Expectation Maximization for Hyperparameter Learning

In this section, we describe the estimation method for hyperparameter set $\mathbf{\Theta}$ via the EM algorithm corresponding to M-step in (22). Using the approximate posterior $g^{(t)}(\mathbf{E}, \mathbf{X}|\mathbf{Y})$ at the t-th step as descrbed in Section V-C and V-D, the ELBO $\mathcal{F}\left(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}\right)$ in (22) can be approximated as

$$\mathcal{F}(\mathbf{\Theta},\mathbf{\Theta}^{(t)}) \simeq -\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{U}}\Big\{\ln\sigma^e_{n_c,u} + (\sigma^e_{n_c,u})^{-1}\mathbb{E}_{g^{(t)}(e_{n_c,u}|\mathbf{Y})}\left[|e_{n_c,u}|^2\right]\Big\} + \text{const.} \quad (66)$$

Since the ELBO $\mathcal{F}(\mathbf{\Theta}, \mathbf{\Theta}^{(t)})$ is concave for $(\sigma^e_{n,u})^{-1}$, the maximization problem in (22) can be solved by the first-order necessary and sufficient condition $\partial\mathcal{F}(\mathbf{\Theta}, \mathbf{\Theta}^{(t)})/\partial\left(\sigma^e_{n,u}\right)^{-1} = \mathbf{0}$, which derives the optimal variance $\sigma^{e(t+1)}_{n,u}$ at the t-th step as

$$\sigma^{e(t+1)}_{n_c,u} = \mathbb{E}_{g^{(t)}(e_{n_c,u}|\mathbf{Y})}\left[|e_{n_c,u}|^2\right] = |\hat{e}^{(t)}_{n_c,u}|^2 + \xi^{e(t)}_{n_c,u}, \quad (67)$$

where $\hat{e}^{(t)}_{n_c,u}$ and $\xi^{e(t)}_{n_c,u}$ are the approximate posterior mean and variance at the t-th step, calculated in (60).

F. Reinforcement for the Model-Based Estimate

To further improve the convergence performance for the EP algorithm, we update the model-based estimate $\hat{\mathbf{S}}$ in each iteration. Using the estimated residual channel error $\hat{\mathbf{E}}^{(t-1)} \triangleq \mathbb{E}_{g^{(t-1)}(\mathbf{E},\mathbf{X}|\mathbf{Y})}[\mathbf{E}]$ at the $(t-1)$-th iteration, the channel estimate for the u-th UE can be reconstructed as

$$\hat{\mathbf{h}}^{(t-1)}_u = \hat{\mathbf{s}}^{(t-1)}_u + \hat{\mathbf{e}}^{(t-1)}_u. \quad (68)$$

The model-based estimate at the t-th iteration $\hat{\mathbf{s}}^{(t)}_u$ is updated with the channel estimate at the previous iteration $\hat{\mathbf{h}}^{(t-1)}_u$ in (68). To efficiently estimate $\hat{\mathbf{s}}^{(t)}_u$ by leveraging the near-field sparsity, the virtual channel representation with polar grids as described in Section IV-A are utilized. The grids are dynamically designed in the iterations, where the center of the grids is set as the angle and distance estimates at the previous iteration, and the range of grids decreases with the number of iterations. Thus, the angle and distance grids for the u-th UE and l-th path at the t-th iteration are designed as

$$\tilde{\theta}^{(t)}_{u,l,g_\theta} \in \left[\hat{\theta}^{(t-1)}_{u,l} - \sigma^{(t)}_\theta,\ \hat{\theta}^{(t-1)}_{u,l} + \sigma^{(t)}_\theta\right], \quad (69a)$$

$$\tilde{r}^{(t)}_{u,l,g_r} \in \left[\hat{r}^{(t-1)}_{u,l} - \sigma^{(t)}_r,\ \hat{r}^{(t-1)}_{u,l} + \sigma^{(t)}_r\right], \quad (69b)$$

with $g_\theta \in \{1, \ldots, \bar{G}_\theta\}$, $g_r \in \{1, \ldots, \bar{G}_r\}$. $\hat{\theta}^{(t-1)}_{u,l}$ and $\hat{r}^{(t-1)}_{u,l}$ are the angle and distance estimates at the $(t-1)$-th iteration, respectively, and $\sigma^{(t)}_\theta$ and $\sigma^{(t)}_r$ are, respectively, the range of angle and distance grids, where the initial values $\hat{\theta}^{(0)}_{u,l}$ and $\hat{r}^{(0)}_{u,l}$ are determined using the angle and distance estimates obtained by the initial channel estimation as shown in Algorithm 1.

Note that the range of angle and distance grids $\sigma^{(t)}_\theta$ and $\sigma^{(t)}_r$ are respectively designed by a monotonically decreasing function such as $\sigma^{(t)}_\theta = a_\theta \exp(-t/2) + b_\theta$ and $\sigma^{(t)}_r = a_r \exp(-t/2) + b_r$, where the constant values $\{a_\theta, b_\theta, a_r, b_r\}$ are uniquely determined with the desired range $\sigma^{(1)}_\theta$, $\sigma^{(1)}_r$, $\sigma^{(T)}_\theta$, and $\sigma^{(T)}_r$. Accordingly, the sets of angle and distance grids for the u-th UE are defined as $\tilde{\boldsymbol{\theta}}^{(t)}_{u,l} \triangleq \{\tilde{\theta}^{(t)}_{u,l,g_\theta}\}_{g_\theta=1}^{\bar{G}_\theta}$, $\tilde{\boldsymbol{\theta}}^{(t)}_u \triangleq \{\tilde{\boldsymbol{\theta}}^{(t)}_{u,l}\}_{l=1}^{\hat{L}_u}$, $\tilde{\mathbf{r}}^{(t)}_{u,l} \triangleq \{\tilde{r}^{(t)}_{u,l,g_r}\}_{g_r=1}^{\bar{G}_r}$, and $\tilde{\mathbf{r}}^{(t)}_u \triangleq \{\tilde{\mathbf{r}}^{(t)}_{u,l}\}_{l=1}^{\hat{L}_u}$.

Using the angle and distance grids in (69a)-(69b), the polar-domain dictionary matrix for the u-th UE is designed as

$$\tilde{\mathbf{A}}_u(\tilde{\boldsymbol{\theta}}_u^{(t)}, \tilde{\mathbf{r}}_u^{(t)}) = \left[\tilde{\mathbf{A}}_{u,1}(\tilde{\boldsymbol{\theta}}_{u,1}^{(t)}, \tilde{\mathbf{r}}_{u,1}^{(t)}), \ldots, \tilde{\mathbf{A}}_{u,\hat{L}_u}(\tilde{\boldsymbol{\theta}}_{u,\hat{L}_u}^{(t)}, \tilde{\mathbf{r}}_{u,\hat{L}_u}^{(t)})\right], \tag{70}$$

where $\tilde{\mathbf{A}}_{u,l}(\tilde{\boldsymbol{\theta}}_{u,l}^{(t)}, \tilde{\mathbf{r}}_{u,l}^{(t)}) \in \mathbb{C}^{N\times \bar{G}_\theta \bar{G}_r}$ is the virtual array response for the u-th UE and l-th path defined as

$$\tilde{\mathbf{A}}_{u,l}(\tilde{\boldsymbol{\theta}}_{u,l}^{(t)}, \tilde{\mathbf{r}}_{u,l}^{(t)}) = \left[\tilde{\mathbf{a}}(\tilde{\theta}_{u,l,1}^{(t)}, \tilde{r}_{u,l,1}^{(t)}), \ldots, \tilde{\mathbf{a}}(\tilde{\theta}_{u,l,G_\theta}^{(t)}, \tilde{r}_{u,l,G_r}^{(t)})\right].$$

Through the virtual channel representation with the dictionary matrix $\tilde{\mathbf{A}}_u(\tilde{\boldsymbol{\theta}}_u^{(t)}, \tilde{\mathbf{r}}_u^{(t)})$, the near-field channel for the u-th UE can be expressed as

$$\mathbf{h}_u = \sum_{l=1}^{\hat{L}} \tilde{\mathbf{A}}_{u,l}(\boldsymbol{\theta}_{u,l}^{(t)}, \tilde{\mathbf{r}}_{u,l}^{(t)})\tilde{\mathbf{z}}_{u,l} = \tilde{\mathbf{A}}_u(\tilde{\boldsymbol{\theta}}_u^{(t)}, \tilde{\mathbf{r}}_u^{(t)})\tilde{\mathbf{z}}_u, \tag{71}$$

where $\tilde{\mathbf{z}}_{u,l} \in \mathbb{C}^{\bar{G}_\theta \bar{G}_r \times 1}$ is the virtual path gain vector for the l-th path, and $\tilde{\mathbf{z}}_u = [\tilde{\mathbf{z}}_{u,1}^{\mathrm{T}}, \ldots, \tilde{\mathbf{z}}_{u,\hat{L}_u}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{\bar{G}_\theta \bar{G}_r \hat{L}_u \times 1}$ is the virtual path gain vector including all paths.

In light of the near-field model in (71), an update of the model-based estimate $\hat{\mathbf{s}}_u^{(t)}$ can be obtained by

$$\hat{\mathbf{s}}_u^{(t)} = \mathbf{A}(\hat{\boldsymbol{\theta}}_u^{(t)}, \hat{\mathbf{r}}_u^{(t)})\hat{\mathbf{z}}_u^{(t)}, \tag{72}$$

with $\hat{\mathbf{z}}_u^{(t)}$ denoting the path gain estimates, $\mathbf{A}(\hat{\boldsymbol{\theta}}_u^{(t)}, \hat{\mathbf{r}}_u^{(t)})$ being the array responses, which can be computed by solving

$$\begin{aligned} \underset{\tilde{\mathbf{z}}_u}{\text{minimize}} \quad & \left\|\hat{\mathbf{h}}_u^{(t-1)} - \hat{\mathbf{s}}_u^{(t)}\right\|_2^2 \\ \text{subject to} \quad & \hat{\mathbf{s}}_u^{(t)} = \tilde{\mathbf{A}}_u(\tilde{\boldsymbol{\theta}}_u^{(t)}, \tilde{\mathbf{r}}_u^{(t)})\tilde{\mathbf{z}}_u \\ & \|\tilde{\mathbf{z}}_{u,l}\|_0 = 1, \quad \forall l \in \{1, 2, \ldots, \hat{L}_u\}. \end{aligned} \tag{73}$$

Solving the problem (73) yields the path gain estimate $\hat{\mathbf{z}}_u^{(t)}$, angle estimate $\hat{\boldsymbol{\theta}}_u^{(t)}$ and the distance estimate $\hat{\mathbf{r}}_u^{(t)}$, which correspond to the non-zero elements of $\tilde{\mathbf{z}}_u$.

G. Algorithm Description

To illustrate the overall procedure of the proposed JCDE algorithm, the pseudocode is provided in Algorithm 2, and the algorithmic flow is depicted in Fig. 4. As shown in the figure, the algorithm mainly consists of three parts: 1) data estimation, 2) residual error estimation, and 3) reinforcement for the model-based estimate.

In the data estimation, the mean and variance of the tentative data estimates $\{\hat{x}_{u,k}^q, \xi_{u,k}^{q,x}\}$ are generated through the Soft-IC in (38a)-(38b), sub-array-wise LMMSE filtering in (40a)-(40b), and combining processes (46), using the received signal $\mathbf{Y}$ and the soft replicas $\{\hat{x}_{c,u,k}^v, \xi_{c,u,k}^{v,x}\}$ and $\{\hat{e}_{n_c,u,k}^v, \xi_{n_c,u,k}^{v,e}\}$. Then, the tentative estimates $\{\hat{x}_{u,k}^q, \xi_{u,k}^{q,x}\}$ are updated to $\{\hat{x}_{u,k}, \xi_{u,k}^x\}$ through the denoising process in (48a)-(48b) that exploits the prior $p(x_{u,k})$ in (18) based on the QAM constellation. For the Soft-IC process in the next iteration, the soft replicas $\{\hat{x}_{c,u,k}^v, \xi_{c,u,k}^{v,x}\}$ and $\{\hat{x}_{c,u,k}^w, \xi_{c,u,k}^{w,x}\}$ are updated by suppressing the self-noise feedback in (50) and (56c).

In the residual error estimation, similar to the data estimation, the mean and variance of the tentative residual error estimates $\{\hat{e}_{n_c,u}^q, \xi_{n_c,u}^{q,e}\}$ are generated through the Soft-IC in (56a)-(56b) and MRC filtering processes in (61), using the received signal $\mathbf{Y}$ and the soft replicas $\{\hat{x}_{c,u,k}^w, \xi_{c,u,k}^{w,x}\}$ and $\{\hat{e}_{n_c,u,k}^v, \xi_{n_c,u,k}^{v,e}\}$. Then, the tentative estimates $\{\hat{e}_{n_c,u}^q, \xi_{n_c,u}^{q,e}\}$ are updated to $\{\hat{e}_{n_c,u}, \xi_{n_c,u}^e\}$ through the denoising process in (60) that exploits the prior $p(e_{n_c,u}; \boldsymbol{\Theta})$ in (19) reflecting the sparsity of the residual errors. For the Soft-IC process in the next iteration, the soft replicas $\{\hat{e}_{n_c,u,k}^v, \xi_{n_c,u,k}^{v,e}\}$ are updated by suppressing the self-noise feedback in (63a)-(63b). The hyperparameter $\boldsymbol{\Theta}$ incorporated in the prior $p(e_{n_c,u}; \boldsymbol{\Theta})$ is updated through the M-step of the EM algorithm in (67). Based on the updated residual error estimate $\hat{\mathbf{e}}_u$, the channel estimate is reconstructed as $\hat{\mathbf{h}}_u = \hat{\mathbf{s}}_u + \hat{\mathbf{e}}_u$, and the model-based estimate $\hat{\mathbf{s}}_u$ is subsequently refined by solving the optimization problem (73) via OMP [23].

After repeating the above procedure for a fixed number of iterations T, $\hat{\mathbf{X}}$ and $\hat{\mathbf{H}}$ are adopted as the estimate of the data and channel. To enhance convergence performance, a damping scheme [15] is introduced in line 6, 7, 12, and 19 of Algorithm 2. The damping factor in the proposed method is determined through several simulations by considering the trade-off between estimation performance and convergence rate as in [10], [26]. Note that the proposed initial channel estimation in Algorithm 1 and the JCDE in Algorithm 2 does not require distinguishing between the LoS and NLoS paths.

VI. Simulation Results

This section evaluates the performance of the proposed initial channel estimation and subsequent JCDE algorithms under the following setup. The carrier frequency is 100 GHz, the number of BS antennas N is 200, the number of UEs U is 50, the modulation order Q is 64-QAM, and the length of pilots K_{p} and data K_{d} are 25 and 100, respectively. The non-orthogonal pilot $\mathbf{X}_{\mathrm{p}} \in \mathbb{C}^{50\times 25}$ is designed by the frame design method in [14]. The near-field channel is composed of $L_u = 3$ paths, *i.e.,* 1 LoS path and 2 NLoS paths. The total number of paths is $L = 50\times 3 = 150$, and the corresponding oversampling quantity used in Algorithm 1 is set to $\hat{L} = 250$. The AoAs and distances are uniformly randomly generated in the range $\theta_{u,l} \in [-60°, 60°]$ and $r_{u,l} \in [1, 10]$ m, respectively. The path gain is generated as $z_{u,l} \sim \mathcal{CN}(0, K_{\mathrm{f}}/(K_{\mathrm{f}}+1))$, $(l = 1)$ and $z_{u,l} \sim \mathcal{CN}(0, 1/((K_{\mathrm{f}}+1)(L-1)))$, $(l \neq 1)$ with a Rician K-factor $K_{\mathrm{f}} = 10$ dB. The polar-domain dictionary $\tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{r}})$ in (7) is designed with $G_r = 7$, $G_\theta = 395$ and desired coherence $\bar{\gamma}_{\mathrm{coh}} = 0.6$ in [9]. The performance is evaluated by the normalized mean-squared error (NMSE) and bit error rate (BER) under various signal-to-noise ratio (SNR). NMSE and SNR are defined as $\mathrm{NMSE}(\mathbf{H}) \triangleq \mathbb{E}\left[\|\mathbf{H} - \hat{\mathbf{H}}\|_{\mathrm{F}}^2 / \|\mathbf{H}\|_{\mathrm{F}}^2\right]$, and $\mathrm{SNR} \triangleq \mathbb{E}\left[\|\mathbf{HX}\|_{\mathrm{F}}^2\right] / \mathbb{E}\left[\|\mathbf{N}\|_{\mathrm{F}}^2\right]$. In what follows, the initial channel estimation and JCDE performance are evaluated in Section VI-A and VI-B, respectively.

A. Initial Channel Estimation Performance

To evaluate the initial channel estimation performance, the following estimation methods are compared: (a) LS: a classical least squares-based channel estimation, (b) P-SOMP [6]: a near-field channel estimation without considering the non-orthogonality of pilots. (c) 2D-CoSaMP [9]: a near-field

Algorithm 2 Proposed JCDE algorithm

Input: $\mathbf{Y}$, $\mathbf{X}_\mathrm{p}$, $\hat{\mathbf{H}}_0^{\mathcal{S}}$, $\{\hat{\boldsymbol{\theta}}_u, \hat{\mathbf{r}}_u\}_{u\in\mathcal{U}}$, $\{\hat{L}_u\}_{u=1}^{U}$, T, $\bar{G}_\theta$, $\bar{G}_r$, $\sigma_\theta^{(1)}$, $\sigma_\theta^{(T)}$, $\sigma_r^{(1)}$, $\sigma_r^{(T)}$
Output: $\hat{\mathbf{X}}$, $\hat{\mathbf{H}}$, $\{\hat{\boldsymbol{\theta}}_u,\ \hat{\mathbf{r}}_u\}_{u\in\mathcal{U}}$

// Initialization
$\hat{\mathbf{S}} = \mathbf{D}_N\hat{\mathbf{H}}_0^{\mathcal{S}}$ from Algorithm 1
$\hat{x}^v_{c,u,k_\mathrm{d}} = \hat{x}^w_{c,u,k_\mathrm{d}} = 0$, $\xi^{v,x}_{c,u,k_\mathrm{d}} = \xi^{w,x}_{c,u,k_\mathrm{d}} = E_s$, $\forall k_\mathrm{d}\in\mathcal{K}_\mathrm{d}$
$\hat{x}^v_{c,u,k_\mathrm{p}} = \hat{x}^w_{c,u,k_\mathrm{p}} = [\mathbf{X}_\mathrm{p}]_{u,k_\mathrm{p}}$, $\xi^{v,x}_{c,u,k_\mathrm{p}} = \xi^{w,x}_{c,u,k_\mathrm{p}} = 0$, $\forall k_\mathrm{p}\in\mathcal{K}_\mathrm{p}$
$\hat{e}^v_{n_c,u,k} = 0$, $\xi^{v,e}_{n_c,u,k} = |[\mathbf{D}_N\hat{\mathbf{H}}_0^{\mathcal{S}}]_{n_c,u}|^2$, $\forall k\in\mathcal{K}$
for $t = 1, 2, \ldots, T$ **do**
  // EP for data estimation
  Calculate $\tilde{\mathbf{y}}^x_{c,u,k}$, $\boldsymbol{\Omega}^x_{c,k}$ from (38a)-(38b)
  Calculate $\hat{x}^q_{c,u,k}$, $\xi^{q,x}_{c,u,k}$ from (40a)-(40b)
  Calculate $\hat{x}^q_{u,k}$, $\xi^{q,x}_{u,k}$ from (46)
  Calculate $\hat{x}_{u,k}$, $\xi^x_{u,k}$ from (48a)-(48b)
  Calculate $\hat{x}^v_{c,u,k}$, $\xi^{v,x}_{c,u,k}$ from (50) with damping
  Calculate $\hat{x}^w_{c,u,k}$, $\xi^{w,x}_{c,u,k}$ from (56c) with damping
  // EP for residual error estimation
  Calculate $\tilde{y}^e_{n_c,u,k}$, $\phi^e_{n_c,u,k}$ from (56a)-(56b)
  Calculate $\hat{e}^q_{n_c,u,k}$, $\xi^{q,e}_{n_c,u,k}$ from (55)
  Calculate $\hat{e}^q_{n_c,u}$, $\xi^{q,e}_{n_c,u}$ from (61)
  Calculate $\hat{e}_{n_c,u}$, $\xi^e_{n_c,u}$ from (60)
  Calculate $\hat{e}^v_{n_c,u,k}$, $\xi^{v,e}_{n_c,u,k}$ from (63a)-(63b) with damping
  Update channel estimate $\hat{\mathbf{h}}_u = \hat{\mathbf{s}}_u + \hat{\mathbf{e}}_u$ from (68)
  // EM algorithm for hyperparameter learning
  Update $\sigma^e_{n_c,u}$ from (67)
  // Reinforcement for the model-based estimate
  Generate the grids $\tilde{\theta}_{u,l,g_\theta}$, $\tilde{r}_{u,l,g_r}$ from (69a)-(69b)
  Design the dictionary $\tilde{\mathbf{A}}_u(\tilde{\boldsymbol{\theta}}_u, \tilde{\mathbf{r}}_u)$ from (70)
  Obtain $\hat{\boldsymbol{\theta}}_u, \hat{\mathbf{r}}_u, \hat{\mathbf{z}}_u$ by solving (73)
  Calculate $\hat{\mathbf{s}}_u$ from (72) with damping
end for

channel estimation considering non-orthogonality, and (d) the proposed initial channel estimation method in Algorithm 1.

Fig. 5a shows the NMSE against SNR. The P-SOMP exhibits limited improvement with an increase in SNR due to pilot contamination stemming from non-orthogonal pilots, whereas 2D-CoSaMP demonstrates a performance enhancement compared to P-SOMP. The proposed method surpasses these conventional methods by mitigating noise amplification through the utilization of 2D-OMP in the second stage associated with UE-path pairing, resulting in superior channel estimation. Fig. 5b and Table I show the computational complexity evaluated by floating point operations (FLOPs). As depicted in the figure, the FLOPs of the proposed method are comparable to 2D-CoSaMP, owing to the two-stage procedure separating angle-distance estimation and UE-path pairing.

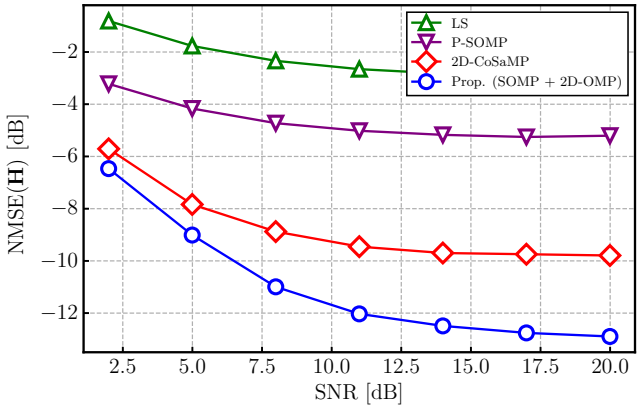
B. JCDE Performance

In this subsection, we evaluate the proposed JCDE algorithm. As for JCDE algorithm parameters, the damping factor is set to 0.5, the number of iterations is $T = 30$, the number of grids are $\bar{G}_\theta = 5, \bar{G}_r = 5$, the grid ranges are $\sigma_\theta^{(1)} = 5°$, $\sigma_\theta^{(T)} = 0.1°$, $\sigma_r^{(1)} = 5$ m, and $\sigma_r^{(T)} = 1$ m, respectively. To

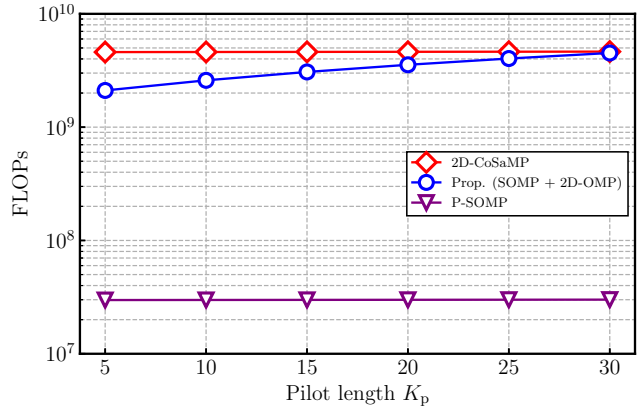
TABLE I: Computational complexity of initial channel estimation

Algotrithm	FLOPs
P-SOMP [6]	$\mathcal{O}\left(NU(G_rG_\theta + \hat{L}) + \hat{L}^2U(N + G_rG_\theta)\right)$
2D-CoSaMP [9]	$\mathcal{O}\left(T_\mathrm{iter}\left(NU(K_\mathrm{p} + G_rG_\theta) + \hat{L}NK_\mathrm{p} + \hat{L}^3\right)\right)$
Proposed	$\mathcal{O}\Big(NK_\mathrm{p}(G_rG_\theta + \hat{L} + NU)$ $+\hat{L}^2(G_rG_\theta K_\mathrm{p} + NU + NK_\mathrm{p}) + \hat{L}^3K_\mathrm{p} + \hat{L}^4\Big)$

Note: T_iter is the iteration number of 2D-CoSaMP, determined by [9].



(a) NMSE versus SNR with the pilot length $K_\mathrm{p} = 25$.

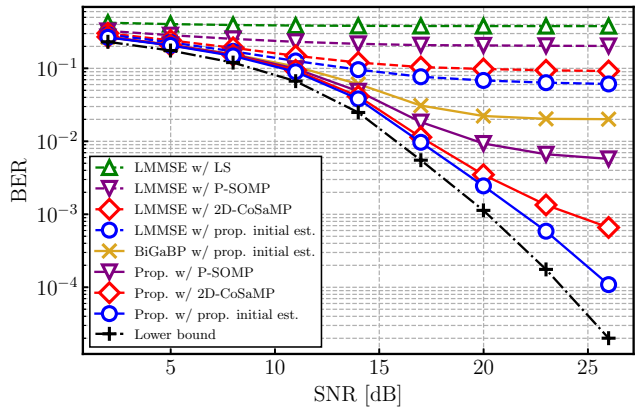


(b) FLOPs versus the pilot length K_p.

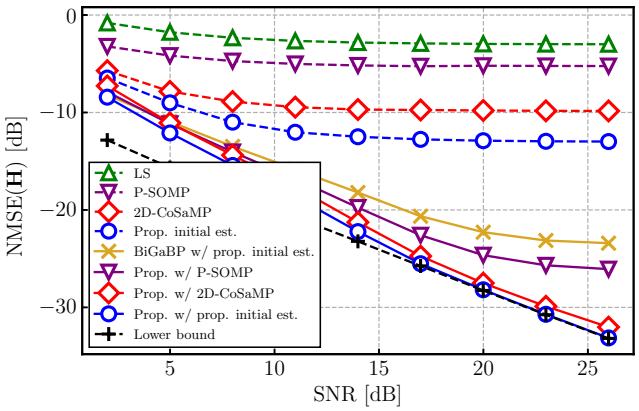
Fig. 5: NMSE and FLOPs performance.

consider the simplicity of implementation and the complexity of the LMMSE-based filter in (40a), the large array with N antennas is evenly divided into C sub-arrays with $N_c = N/C$, $\forall c \in \mathcal{C}$ antennas. Since the complexity order of the LMMSE-based filter is $\mathcal{O}\left(K_\mathrm{d}\sum_{c=1}^{C}N_c^3 + UK_\mathrm{d}\sum_{c=1}^{C}N_c^2\right)$, even sub-array partitioning is optimal in terms of minimizing computational complexity given C. Unless otherwise specified, these parameters are set to $N = 200$, $C = 4$, $N_c = 50$. For comparison, AoA-aided BiGaBP [11] are employed as a benchmark, which is a state-of-the-art JCDE algorithm. Besides, we consider a Genie-aided case with perfect CSI or data corresponding to the lower bound of the proposed method.

1) JCDE Performance with Initial Channel Estimation: This subsection reveals the NMSE and BER performance of the JCDE algorithms with various initial channel estimation methods, including P-SOMP, 2D-CoSaMP, and the proposed initial channel estimation method. To evaluate the data detec-

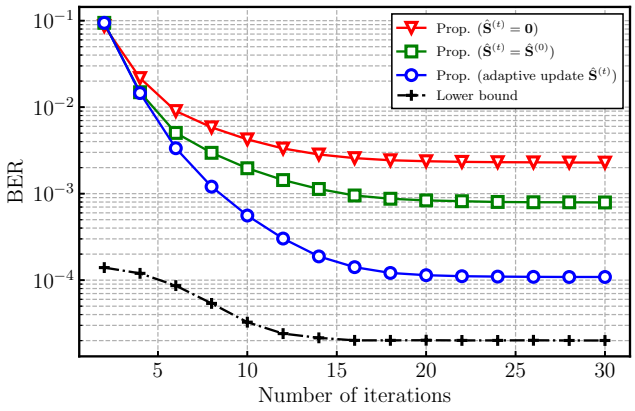


(a) BER versus SNR with the pilot length $K_{\mathrm{p}} = 25$.

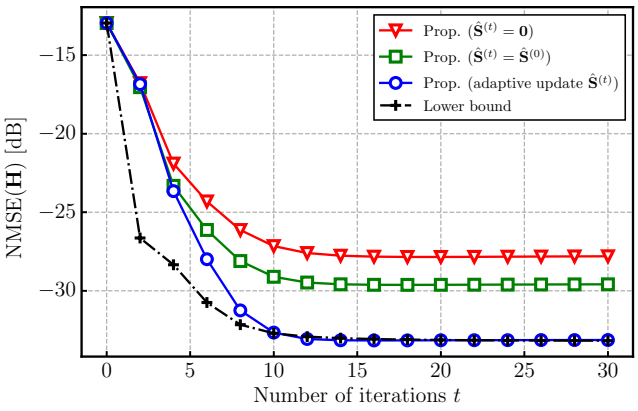


(b) NMSE versus SNR with the pilot length $K_{\mathrm{p}} = 25$.

Fig. 6: BER and NMSE performance.



(a) BER against the number of algorithmic iterations.



(b) NMSE against the number of algorithmic iterations.

Fig. 7: Convergence behavior in term of BER and NMSE performance with $K_{\mathrm{p}} = 25$, SNR $= 26$ dB.

tion capability of the above initial channel estimation methods, the LMMSE detector is used for data estimation.

Fig. 6 shows the BER and NMSE performance. As shown in the figures, while LMMSE with LS, which cannot take advantage of the near-field model structures, exhibits poor BER performance, LMMSE with the other initial estimation approaches considering the near-field model structure achieve a slight performance improvement. However, there remains high-level error floors due to the non-orthogonal pilots. In contrast, the JCDE algorithms boost BER performance due to utilizing both pilot and consecutive data. In particular, the proposed JCDE algorithm with the proposed initial channel estimation demonstrates a significant performance gain, approaching the lower bound of perfect CSI or perfect data.

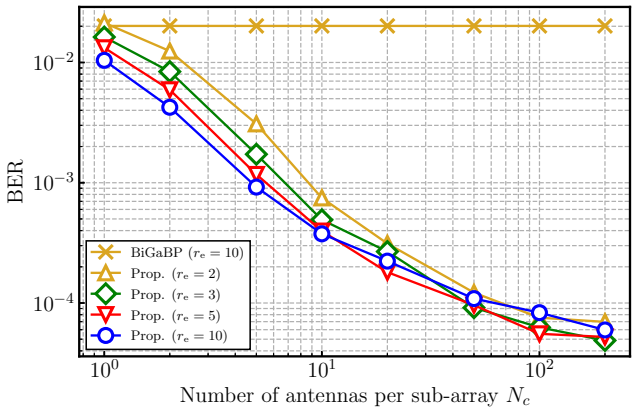
Moreover, the proposed JCDE algorithm demonstrates a notable BER performance compared to the state-of-the-art AoA-aided BiGaBP [11]. The performance improvement can be attributed to two primary factors. The first factor is that the proposed algorithm can leverage the near-field model-based estimation described in Section V-F, whereas BiGaBP relies on the far-field assumption. The second factor is that the proposed sub-array-wise LMMSE-based detection in (40a) is capable of addressing the correlation between the leaked energy in the beam-domain, whereas BiGaBP is incapable of doing so because of its MRC. To reveal the aforementioned two factors, in Section VI-B2, we show the convergence analysis with and without near-field model information. Besides, we evaluate in Section VI-B3 the performance of the sub-array-wise LMMSE-based detection and its complexity across various numbers of antennas per sub-array N_c.

2) Convergence Analysis: To clarify the advantages gained by leveraging the near-field structure, we evaluate the proposed JCDE algorithm with and without the model-based estimation process explained in Section V-F. Fig. 7 illustrates the BER and NMSE convergence behavior with respect to the number of algorithmic iterations. In the figure, the red triangle marker corresponds to the proposed JCDE algorithm without the model-based estimate, *i.e.,* $\hat{\mathbf{S}}^{(t)} = \mathbf{0}$, where the prior distribution is designed i.i.d. for each element of $\mathbf{H}$ instead of $\mathbf{E}$, akin to [12], [13]. The green square marker corresponds to the proposed JCDE algorithm with the initial model-based estimate but without updating in iterations, *i.e.,* $\hat{\mathbf{S}}^{(t)} = \hat{\mathbf{S}}^{(0)}$. Comparing the red triangle maker and green square marker, we can verify the performance improvement owing to the use of the near-filed model through the decomposition of $\mathbf{H}$ into $\hat{\mathbf{S}}$ and $\mathbf{E}$ in (15). Furthermore, in comparison to the proposed algorithm with adaptive update, it can be seen that the adaptive updating of the model-based estimate $\hat{\mathbf{S}}^{(t)}$ enhances the BER and NMSE by exploiting the near-field model.

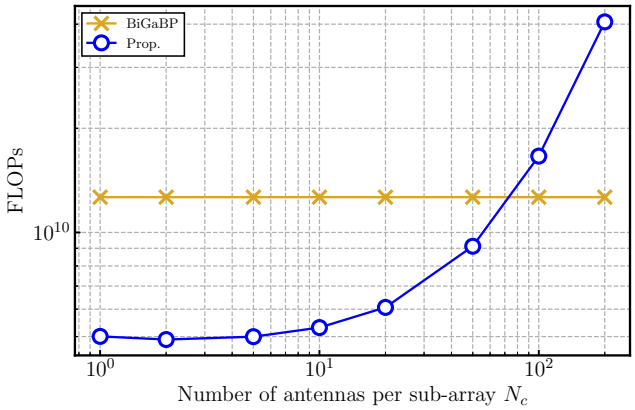
TABLE II: Computational complexity of JCDE algorithms

Algotrithm	FLOPs
BiGaBP [11]	$\mathcal{O}\Big(UK_{\mathrm{d}}NQ + UKN + U\hat{L}_u^2N^2 + UN^2\hat{L}_u\Big)$
Proposed	$\mathcal{O}\Big(CUK_{\mathrm{d}}N_c^2 + CK_{\mathrm{d}}N_c^3 + UK_{\mathrm{d}}Q + UKN + UN\bar{G}_\theta\bar{G}_r + U\hat{L}_u^2N + U\hat{L}_u\bar{G}_\theta\bar{G}_r\Big)$

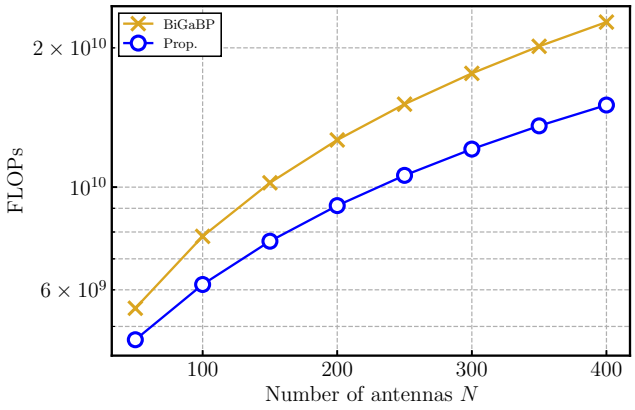
3) Performance Against the Number of Antennas per Sub-array: To analyze the impact of the number of antennas per sub-arrays N_c on the performance of the JCDE algorithm employing the sub-array-wise LMMSE-based detection, we offer in Fig. 8 the BER and FLOPs with respect to various numbers of antennas per sub-array N_c, where $N_c = N = 200$ corresponds to the full-array LMMSE-based detection and $N_c = 1$ corresponds to the MRC-based detection. As depicted in the figure, the BER decreases as the number of antennas per sub-array N_c increases (*i.e.,* the number of sub-arrays C decreases) because each sub-array effectively address the correlation in the beam-domain caused by energy leakage as N_c increases. In particular, the MRC-based detection corresponding to $N_c = 1$ exhibits poor performance. To analyze the energy leakage effects under various intensities of near-field effects, the BS-UE distance r_{e} is varied across different values, where the distances between the BS and the u-th UE are uniformly randomly generated within the range $r_{u,l} \in [1, r_{\mathrm{e}}]$ [m]. As shown in the figure, when N_c is small, the BER performance deteriorates as the BS-UE distance decreases. This is because each sub-array with a small number of antennas cannot capture the leaked energy that grows more prominent at shorter distances, as illustrated in Fig 2. In contrast, a decrease in the number of antennas per sub-array leads to a reduction in FLOPs attributed to the decreased size of the inverse matrix associated with the LMMSE-based detection in (40a). Despite relying on the LMMSE-based detector, the proposed algorithm can achieve lower FLOPs when $N_c \leq 50$ $(C \geq 4)$ compared to BiGaBP, which relies on an MRC-based detector, since the proposed method suppresses self-feedback in (50) after the denoising process as in (48a)-(48b) with FLOPs $\mathcal{O}(UK_dQ)$, whereas BiGaBP suppresses self-feedback before the denoising process [11], [32] with FLOPs $\mathcal{O}(NUK_dQ)$ that is dominant complexity throughout the entire process as shown in Table II. As a result, the proposed method achieves lower computational complexity across various numbers of antennas N compared to the conventional method, as shown in Fig. 8c. From the above results, it is evident that the proposed method outperforms the conventional method in both data detection and complexity.



(a) BER against the number of antennas per sub-array N_c. To investigate the impact of energy leakage effects, the distances between the BS and the u-th UE are uniformly randomly generated within the range $r_{u,l} \in [1, r_{\mathrm{e}}]$ m.



(b) FLOPs against the number of antennas per sub-array N_c with $N = 200$.



(c) FLOPs against the number of antennas N with $N_c = 50$.

Fig. 8: BER and FLOPs performance with $K_{\mathrm{p}} = 25$, SNR $= 26$ dB.

VII. Conclusion

This paper proposed an initial channel estimation algorithm and subsequent JCDE algorithm for multiuser XL-MIMO systems with non-orthogonal pilots. The initial channel estimation is performed by an efficient two-stage compressed sensing algorithm exploiting the polar-domain sparsity. Furthermore, the initial channel estimates are refined by jointly utilizing both non-orthogonal pilots and data via the EP algorithm. To improve channel estimation accuracy, the model-based deterministic approach is integrated into a Bayesian inference framework. In addition, to address the near-field specific correlation in the beam-domain, a sub-array-wise LMMSE filter is designed considering the correlation and channel estimation errors for data detection. Computer simulations validated that the proposed method is superior to existing approaches in terms of channel estimation, data detection, and complexity.

For future work, a flexible sub-array partitioning scheme, where each sub-array has a different number of antennas,

should be explored to more effectively mitigate the leaked energy in the beam-domain.

REFERENCES

[1] K. Arai, K. Ishibashi, H. Iimori, P. V. Klaine, and S. Malomsoky, "Near-field channel estimation for uplink multiuser XL-MIMO systems," in *Proc. IEEE 25th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2024, pp. 256–260.

[2] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wirel. Commun.*, vol. 27, no. 4, pp. 74–80, 2020.

[3] Z. Wang *et al.*, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1560–1605, 2024.

[4] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, 2023.

[5] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1999–2049, 2023.

[6] 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, 2022.

[7] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 5, pp. 2946–2960, 2018.

[8] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, 2018.

[9] X. Xie, Y. Wu, J. An, D. W. K. Ng, C. Xing, and W. Zhang, "Massive unsourced random access for near-field communications," *IEEE Trans. Commun.*, vol. 72, no. 6, pp. 3256–3272, 2024.

[10] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear gaussian belief propagation for massive MIMO detection with non-prthogonal pilots," *IEEE Trans. Commun.*, vol. 72, no. 2, pp. 1045–1061, 2024.

[11] K. Ito, T. Takahashi, K. Igarashi, S. Ibi, and S. Sampei, "AoA estimation-aided Bayesian receiver design via bilinear inference for mmWave massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 6474–6479.

[12] W. Yan and X. Yuan, "Semi-blind channel-and-signal estimation for uplink massive MIMO with channel sparsity," *IEEE Access*, vol. 7, pp. 95 008–95 020, 2019.

[13] L. Chen and X. Yuan, "Blind multiuser detection in massive MIMO channels with clustered sparsity," *IEEE Wirel. Commun. Lett.*, vol. 8, no. 4, pp. 1052–1055, 2019.

[14] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 11, pp. 7694–7710, 2021.

[15] 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, 2014.

[16] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A, Math. Gen.*, vol. 36, no. 43, 2003.

[17] D. Fan *et al.*, "Angle domain channel estimation in hybrid millimeter wave massive MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 12, pp. 8165–8179, 2018.

[18] Y. Fang, J. Wu, and B. Huang, "2D sparse signal recovery via 2D orthogonal matching pursuit," *Sci. China Inf. Sci.*, vol. 55, pp. 889–897, 2012.

[19] T. P. Minka, "Expectation propagation for approximate Bayesian inference," *Proc. 17th Conf. Uncertainty Artif*, pp. 362–369, 2001.

[20] H. Iimori, G. T. F. de Abreu, O. Taghizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, 2020.

[21] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, 2004.

[22] H. Lei, J. Zhang, Z. Wang, B. Ai, and D. W. Kwan Ng, "Hybrid-field channel estimation for XL-MIMO systems with stochastic gradient pursuit algorithm," *IEEE Transa. Signal Process.*, vol. 72, pp. 2998–3012, 2024.

[23] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput*, 1993, pp. 40–44 vol.1.

[24] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, 2012, pp. 220–224.

[25] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[26] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[27] H. Wang, A. Kosasih, C.-K. Wen, S. Jin, and W. Hardjawana, "Expectation propagation detector for extra-large scale massive MIMO," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 3, pp. 2036–2051, 2020.

[28] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Germany: Springer-Verlag, 2006.

[29] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, no. 2, pp. 211–244, 2002.

[30] P. Jain, P. Kar *et al.*, "Non-convex optimization for machine learning," *Found. Trends Mach. Learn.*, vol. 10, no. 3-4, pp. 142–363, 2017.

[31] Q. Zou and H. Yang, "A concise tutorial on approximate message passing," *arXiv:2201.07487*, 2022.

[32] R. Tamaki, K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Suppression of self-noise feedback in GAMP for highly correlated large MIMO detection," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022, pp. 1300–1305.



Kabuto Arai (Graduate Student Member, IEEE) received B.E. and M.E. degrees in engineering from Chiba University, Japan, in 2019, and 2021, respectively. From 2021 to 2023, he was with the Research and Development Center, Toshiba Corporation. Currently, he is pursuing Ph.D. with Advanced Wireless & Communication Research Center (AWCC), The University of Electro-Communications, Tokyo, Japan. His current research interests include signal processing, millimeter-wave communication and MIMO communication. He was a recipient of the IEICE RCS Active Research Award in 2024.



Koji Ishibashi (Senior Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from Yokohama National University, Yokohama, Japan, in 2007. From 2007 to 2012, he was an Assistant Professor at the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. Since April 2012, he has been with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo, Japan where he is currently a Professor. From 2010 to 2012, he was a Visiting Scholar at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA. He is a senior member of IEICE and IEEE. He is a recipient of Takayanagi Research Encouragement Award in 2009 and KDDI Foundation Award in 2023. He was certified as an Exemplary Reviewer of IEEE Wireless Communications Letters in 2015 and awarded by the Telecommunication Technology Committee (TTC) for his devotion to standardization activities in 2020. He has been an Associate Editor for IEEE Internet of Things Journal and IEEE Wireless Communications Letters since 2025. He in the past served as an Associate Editor for IEICE Transactions on Communications, an Associate Editor for IEEE Journal on Selected Areas in Communications (JSAC) and a Guest Editor for IEEE Open Journal of the Communications Society. His current research interests are near-field communications, grant-free access, energy-harvesting, compressed sensing, coding, and MIMO technologies.



Hiroki Iimori (S'18–M'22) is a Senior Researcher at the Ericsson Research. Prior to this, he was a visiting scholar in the Department of Electrical and Computer Engineering at the University of Toronto, Canada, in 2020. He earned his Ph.D. with special distinction (Auszeichnung, equivalent to Summa Cum Laude) from Jacobs University Bremen, Germany, in 2022. He also holds B.Eng. and M.Eng. (Hons.) degrees in Electrical and Electronic Engineering from Ritsumeikan University, Kyoto, Japan, awarded in 2017 and 2019, respectively. His research interests include optimization theory, wireless communications, and signal processing. He is a recipient of the YKK Doctoral Fellowship from the Yoshida Scholarship Foundation and the IEICE Young Researcher of the Year Award from the IEICE Smart Radio Committee, among other honors.



Paulo Valente Klaine received his B. Eng. degree in electrical and electronic engineering from the Federal University of Technology - Parana (UTFPR), Brazil in 2014, the MSc. degree from the University of Surrey, U.K., in mobile communications systems in 2015, and the PhD degree in electrical and Electronics engineering from the University of Glasgow, U.K., in 2019. He received the IEEE ICEICT'21 Best Paper award. His research interests include self organizing networks, machine learning in wireless networks,V2X communications, wireless blockchain and massive MIMO. Since February of 2022 he has been working as an experienced researcher at Ericsson K. K. Japan.



Szabolcs Malomsoky is the head of the Ericsson Research Japan. Earlier he was leading research units in Hungary, Sweden and the US. He received a Ph.D. degree from the Budapest University of Technology and Economics in 2003. Szabolcs worked with strategy setting and technical leadership in research areas including wireless technologies, network analytics, cloud computing, network management and programmable networks.