

Unified Channel Estimation for XL-RIS Assisted mmWave MU-MIMO Systems

Jeongjae Lee
Electronic Engineering
Hanyang University
Seoul, Korea
lyjcje7466@hanyang.ac.kr

Yunseong Cho
Samsung Research America
Plano, Texas, USA
y.cho1@samsung.com

Songnam Hong
Electronic Engineering
Hanyang University
Seoul, Korea
snhong@hanyang.ac.kr

Abstract—Channel estimation is one of the key challenges for the deployment of reconfigurable intelligent surface (RIS) assisted multiple-input multiple-output (MIMO) systems. In this paper, we study the channel estimation problem for extremely large-scale RIS (XL-RIS) assisted MU-MIMO systems with hybrid beamforming structures. For this system, we for the first time propose an *unified* channel estimation method that yields a notable estimation accuracy in the far-field BS-RIS and far-field RIS-User channels (in short, far-far field channels), and far-near field channels. Via experiments and complexity analysis, we verify the effectiveness of the proposed channel estimation method (named CLRA-JO).

Index Terms—Reconfigurable intelligent surface (RIS), XL-MIMO, channel estimation, hybrid beamforming, low-rank approximation.

I. INTRODUCTION

Reconfigurable intelligent surface (RIS) is a promising technology for robust millimeter-wave (mmWave) and terahertz (THz) multiple-input multiple-output (MIMO) systems [1]. A RIS consists of a uniform array with a large number of reflective elements, each of which can control the phase and reflection angle of the incident signal so that the received power of the intended signal is enhanced [2]. The accuracy of a channel estimation plays a crucial role in implementing various RIS-assisted applications [3]. Also, to further enhance the beamforming gain, in future communication systems (e.g., 6G), the number of reflective elements in the RIS tends to be extremely large (XL) [4]. This will lead to a larger Rayleigh distance (RD), the boundary between the near- and far-field channels. As an example, when the RIS is equipped with 256 half-wavelength spacing elements at operating 100GHz, the RD distance is almost 100m [5]. Therefore, in the 6G systems, the near-field effect becomes non-negligible, and accordingly, the RIS will experience *spherical-wavefront* rather than planar-wavefront.

Thus, it is necessary to develop an efficient channel estimation method for near-field XL-RIS assisted MIMO systems.

This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00409492), in part by the Institute of Information & communications Technology Planning & Evaluation (IITP) under the artificial intelligence semiconductor support program to nurture the best talents (IITP-2024-RS-2023-00253914).

To enhance the efficiency of training overhead, the two-phase channel estimation protocol was proposed [6]. Based on this framework, a 3-dimensional multiple measurement vector (MMV) look-ahead OMP (3D-MLAOMP) was developed as an efficient channel estimation method. This method still suffers from the grid-mismatch problem. Moreover, it requires some impractical assumptions such as the number of spatial paths and the angle of line-of-sight (LoS) path being given as a priori. In [7], a fast-sparse-Bayesian-learning (FSBL)-based channel estimation method was proposed by focusing on the limited visual region of the channel between the RIS and each user in XL-RIS systems. This method is impractical due to the expensive computational complexity because a super-resolution (SR) algorithm is used to estimate the common AoD at the BS and its complexity is too expensive.

Motivated by the above, in this paper, we study a channel estimation problem for XL-RIS assisted MU-MIMO systems where every single RIS-User channel can be either a near-field or far-field channel according to the distance between them. Our key idea is that the effective (or cascaded) channels to be estimated can be each factorized as the product of low-rank matrices (i.e., the product of the common (or user-independent) matrix and the user-specific coefficient matrix). Leveraging it, we respectively estimate them via collaborative low-rank approximation and proposed joint optimization among the users thereby efficiently reducing the overall training overhead and computational complexity for operating the proposed channel estimation method. Moreover, this is an *unified* channel estimation method which shows an outstanding performance without any modification both in the i) Far-field BS-RIS and far-field RIS-User channels (i.e., *far-far* field channel); ii) Far-field BS-RIS and near-field RIS-User channels (i.e., *far-near* field channel).

Notations. We use $\mathbf{x}$ and $\mathbf{A}$ to denote a column vector and matrix, respectively. Also, $\mathbf{A}^\dagger$ denotes the Moore-Penrose inverse and $\otimes$ denotes the Kronecker product. Given a $M \times N$ matrix $\mathbf{A}$, let $\mathbf{A}(i, :)$ and $\mathbf{A}(:, j)$ be the i -th row and j -th column of $\mathbf{A}$, respectively, and $\|\mathbf{A}\|_F$ denotes the Frobenius norm of $\mathbf{A}$. Given a vector $\mathbf{v}$, $\text{diag}(\mathbf{v})$ denotes a diagonal matrix whose ℓ -th diagonal element is equal to the ℓ -th element of $\mathbf{v}$ and $\|\mathbf{v}\|_2$ denotes the l_2 norm of $\mathbf{v}$. We let $\mathbf{I}$ and $\mathbf{0}$ denote the identity and all-zero matrices, respectively, where the sizes

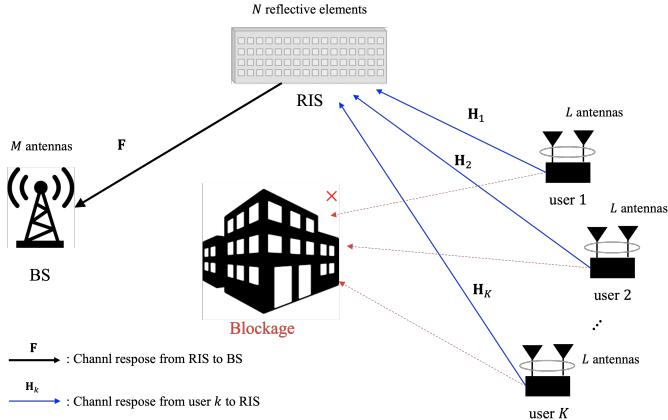


Fig. 1. A XL-RIS assisted MU-MIMO system consisting of the BS with \$M\$ antennas, the RIS with \$N\$ reflective elements, and the \$K\$ users with \$L\$ antennas.

of these matrices are easily obtained from the context.

II. SYSTEM MODEL

We consider an uplink time-division-duplexing (TDD) multi-user multiple-input multiple-output (MU-MIMO) system with uniform planar arrays (UPAs). In this system, a reconfigurable intelligent surface (RIS) with \$N\$ reflective elements assists the communication between the base station (BS) with \$M\$ antennas and the \$K\$ users each having \$L\$ antennas (see Fig. 1). Considering the structure of UPA, we let \$N = N_h \times N_v\$, \$M = M_v \times M_h\$, and \$L = L_v \times L_h\$ throughout the paper. Noticeably, it is assumed that the RIS is equipped with extremely large-scale antenna array (ELAA). For the sake of lower complexity, cost and power consumption [8], the BS is assumed to employ the hybrid beamforming with a limited number of radio frequency (RF) chains, where \$N_{\text{RF}} \leq M\$ denotes the number of RF chains at the BS. As considered in the related works [6], [7], it is assumed that the direct channels between the BS and the \$K\$ users are blocked. The channel responses from the \$k\$-th user to the RIS and from the RIS to the BS are respectively denoted as \$\mathbf{H}_k \in \mathbb{C}^{N \times L}\$ and \$\mathbf{F} \in \mathbb{C}^{M \times N}\$. The reflection vector in the RIS is denoted as \$\mathbf{v} = [v_1, v_2, \dots, v_N]^H\$ with \$v_n = e^{j\vartheta_n}\$, where \$\vartheta_n \in [0, 2\pi)\$ is the phase shift of the \$n\$-th reflective element. Taking this into account, the total channel response from the \$k\$-th user to the BS is defined as

$$\mathbf{F} \text{diag}(\mathbf{v}) \mathbf{H}_k = \mathbf{H}_{[\text{eff}, k]} (\mathbf{I} \otimes \mathbf{v}), \quad (1)$$

where the *effective* (or cascaded) channel is defined as

$$\mathbf{H}_{[\text{eff}, k]} \triangleq [\mathbf{H}_{[\text{eff}, k, 1]} \quad \dots \quad \mathbf{H}_{[\text{eff}, k, L]}] \in \mathbb{C}^{M \times NL}, \quad (2)$$

and where \$\mathbf{H}_{[\text{eff}, k, \ell]} = \mathbf{F} \text{diag}(\mathbf{H}_k(:, \ell)) \in \mathbb{C}^{M \times N}\$.

In the remaining part of this section, we will describe the specific channel models for \$\{\mathbf{H}_k : k \in [K]\}\$ and \$\mathbf{F}\$. We first provide the useful definition:

Definition 1: Given a UPA with \$N = N_h \times N_v\$, a distance between the scatter and the center of array \$d\$, an elevation

angle \$\theta \in [0, 2\pi)\$ and azimuth angle \$\phi \in [0, \pi)\$, we define a *near-field* array response vector [4]:

$$\mathbf{b}_N(d, \theta, \phi) \triangleq \sqrt{\frac{1}{N}} \left[e^{-j\frac{2\pi}{\lambda} \Delta d(1,1)}, \dots, e^{-j\frac{2\pi}{\lambda} \Delta d(N_h, N_v)} \right]^T, \quad (3)$$

Herein, for each \$h \in [N_h]\$ and \$v \in [N_v]\$,

$$\begin{aligned} \Delta d(h, v) = & -\sin \theta \cos \phi x_h - \sin \theta \sin \phi y_v \\ & + \frac{1 - \sin^2 \theta \cos^2 \phi}{2d} x_h^2 + \frac{1 - \sin^2 \theta \sin^2 \phi}{2d} y_v^2 \\ & - \frac{\sin^2 \theta \cos \phi \sin \phi}{d} x_h y_v, \end{aligned} \quad (4)$$

where \$x_h = \frac{\lambda}{2}(h - (N_h + 1)/2)\$ and \$y_v = \frac{\lambda}{2}(v - (N_v + 1)/2)\$. When \$d\$ is sufficiently large, the last three terms in (4) can be negligible. In this case, the array response vector \$\mathbf{b}_N(d, \theta, \phi)\$ is simplified as

$$\begin{aligned} \mathbf{a}_N(\theta, \phi) \triangleq & \sqrt{\frac{1}{N}} \left[1, \dots, e^{j\pi(h \sin \theta \cos \phi + e_v \sin \theta \sin \phi)}, \dots, \right. \\ & \left. e^{j\pi((N_h - 1) \sin \theta \cos \phi + (N_v - 1) \sin \theta \sin \phi)} \right]^T. \end{aligned} \quad (5)$$

Note that this is equivalent to the conventional *far-field* UPA array response vector. ■

Following the *Definition 1*, the channel models are given by

- Near-field \$\mathbf{H}_k\$:

$$\sum_{i=1}^{N_{h_k}} \alpha_{h_k}^i \mathbf{b}_N \left(d_{[h_k, r]}^i, \theta_{[h_k, r]}^i, \phi_{[h_k, r]}^i \right) \mathbf{a}_L^H \left(\theta_{[h_k, t]}^i, \phi_{[h_k, t]}^i \right),$$

- Far-field channel \$\mathbf{H}_k\$:

$$\sum_{i=1}^{N_{h_k}} \alpha_{h_k}^i \mathbf{a}_N \left(\theta_{[h_k, r]}^i, \phi_{[h_k, r]}^i \right) \mathbf{a}_L^H \left(\theta_{[h_k, t]}^i, \phi_{[h_k, t]}^i \right),$$

where \$N_{h_k}\$ is the number of spatial paths, \$\alpha_{h_k}^i\$ is the channel gain for \$i\$-th spatial path, \$d_{[h_k, r]}^i\$ is the distance between \$i\$-th scatter and the RIS, \$\theta_{[h_k, r]}^i\$ (resp. \$\theta_{[h_k, t]}^i\$) and \$\phi_{[h_k, r]}^i\$ (resp. \$\phi_{[h_k, t]}^i\$) are the elevation and azimuth angle of arrival (resp. departure) for the RIS (resp. user \$k\$), respectively.

- Far-field channel \$\mathbf{F}\$:

$$\sum_{j=1}^{N_f} \alpha_f^j \mathbf{a}_M \left(\theta_{[f, r]}^j, \phi_{[f, r]}^j \right) \mathbf{a}_N^H \left(\theta_{[f, t]}^j, \phi_{[f, t]}^j \right),$$

where \$N_f\$ is the number of spatial paths, \$\alpha_f^j\$ is the channel gain for \$j\$-th spatial path, \$\theta_{[f, r]}^j\$ (resp. \$\theta_{[f, t]}^j\$) and \$\phi_{[f, r]}^j\$ (resp. \$\phi_{[f, t]}^j\$) are the elevation and azimuth angle of arrival (resp. departure) for the BS (resp. RIS), respectively.

For the first time, we propose an *unified* channel estimation method in Section III, which can perform in the following categories of wireless channels:

- Far-far field channels: Both \$\mathbf{F}\$ and \$\{\mathbf{H}_k : k \in [K]\}\$ are all far-field channels.
- Far-near field channels: \$\mathbf{F}\$ is a far-field channel and \$\{\mathbf{H}_k : k \in [K]\}\$ are all near-field channels.

III. THE PROPOSED CHANNEL ESTIMATION METHOD

In this section, we present an efficient method to estimate the effective (or cascaded) channels $\{\mathbf{H}_{[\text{eff},k]} : k \in [K]\}$ in (2) from the processed training observations. Since the rank of $\mathbf{F}$ (denoted by $\text{rk}(\mathbf{F})$) is less than equal to N_f , $\mathbf{F}$ is a low-rank matrix in mmWave channels. This is one of the key ingredients for the proposed channel estimation method. Let $\mathbf{S}_{\text{col}}$ denote the $M \times \text{rk}(\mathbf{F})$ matrix whose columns are the basis of the column space of $\mathbf{F}$ (a.k.a. common column space). Thus, each effective channel $\mathbf{H}_{[\text{eff},k,\ell]}$ can be represented as the *low-rank* decomposition:

$$\mathbf{H}_{[\text{eff},k,\ell]} = \mathbf{S}_{\text{col}} \mathbf{T}_{[k,\ell]}, \quad k \in [K], \ell \in [L], \quad (6)$$

where $\mathbf{T}_{[k,\ell]} \in \mathbb{C}^{\text{rk}(\mathbf{F}) \times N}$ denotes a user and antenna specific coefficient matrix. On the basis of the low-rank decomposition, the proposed channel estimation method consists of the two parts. In the first part, the column space of $\mathbf{F}$ (i.e., $\mathbf{S}_{\text{col}}$) is estimated by means of a collaborative low-rank approximation. Then, the coefficient matrices $\mathbf{T}_{[k,\ell]}$'s are jointly optimized. Thus, the proposed channel estimation method is named **Collaborative Low-Rank Approximation and Joint Optimization (CLRA-JO)**.

In the subsequent subsections, we will explain the channel estimation protocol, the proposed channel estimation method and complexity analysis of it. Before describing them, we provide the useful definition below:

Definition 2: Let $\Phi_{[X,Y]}$ be a $X \times Y$ matrix having orthonormal columns. In the following, we will construct the RF combining matrices and the reflection vectors using the matrix $\Phi_{[X,Y]}$. In practice, a DFT matrix can be adopted as $\Phi_{[X,Y]}$. ■

A. Channel Estimation Protocol

The channel estimation protocol performs with J subframes, each of which consists of T symbols (or time slots) with $T \geq KL$. Throughout the paper, J is referred to as the training overhead. In the proposed protocol, the RIS reflection vector $\mathbf{v}_j = [v_{[j,1]}, v_{[j,2]}, \dots, v_{[j,N]}]^T$ and the $M \times N_{\text{RF}}$ RF combining matrix $\mathbf{C}_j$ are unchanged within each subframe $j \in [J]$, but they can be changed across the subframes.

At every subframe $j \in [J]$, each user $k \in [K]$ transmits its orthogonal pilot sequence of the length T to the BS, denoted by $\mathbf{X}_{[j,k]} \in \mathbb{C}^{L \times T}$, such that $\mathbf{X}_{[j,k]} \mathbf{X}_{[j,k]}^H = PT \times \mathbf{I}$, where PT is the energy constraint, P is the transmit power of each user and antenna, and $\mathbf{X}_{[j,k]} \mathbf{X}_{[j,k']}^H = \mathbf{0}$ if $k \neq k'$. This construction is always possible due to the choice of $T \geq KL$. Note that each user k sends the i -th row of $\mathbf{X}_{[j,k]}$ through the i -th transmit antenna during the T time slots. For each subframe $j \in [J]$, the BS then observes the $N_{\text{RF}} \times T$ matrix, given by

$$\mathbf{Y}_j = \mathbf{C}_j^H \left(\sum_{k=1}^K (\mathbf{F} \text{diag}(\mathbf{v}_j) \mathbf{H}_k) \mathbf{X}_{[j,k]} + \mathbf{U}_j \right), \quad (7)$$

where $\mathbf{U}_j$ is the noise matrix whose elements follow independently identically circularly symmetric complex Gaussian

distribution with mean zero and variance σ^2 . Leveraging the orthogonality of the pilot sequences, we can derive the user-specific observation:

$$\begin{aligned} \mathbf{Z}_{[j,k]} &= \frac{1}{PT} \mathbf{Y}_j \mathbf{X}_{[j,k]}^H \in \mathbb{C}^{N_{\text{RF}} \times L} \\ &= \mathbf{C}_j^H (\mathbf{F} \text{diag}(\mathbf{v}_j) \mathbf{H}_k + \mathbf{N}_{[j,k]}), \end{aligned} \quad (8)$$

for $k \in [K]$, where $\mathbf{N}_{[j,k]} \triangleq \frac{1}{PT} \mathbf{U}_j \mathbf{X}_{[j,k]}^H$.

Remark 1: To get suitable observations for each part of the proposed channel estimation method, the minimum subframes are required as follows:

$$J = M_{\text{RF}} B_c + N B_r, \quad (9)$$

where $M_{\text{RF}} \triangleq M/N_{\text{RF}}$ which is assumed as an integer without loss of generality, B_c and B_r denote the hyperparameters which can control the tradeoff between the estimation accuracy and the training overhead. Then, the finally processed observations are respectively determined by

$$\mathbf{M}_{\text{col}} \triangleq \mathbf{F} \mathbf{Q} + \tilde{\mathbf{N}}_{\text{col}}, \quad (10)$$

$$\mathbf{M}_{[\text{row},k,\ell]} \triangleq \Phi_{[M, N_{\text{RF}} B_r]}^H \mathbf{S}_{\text{col}} \mathbf{T}_{[k,\ell]} + \tilde{\mathbf{N}}_{[\text{row},k,\ell]}, \quad (11)$$

where $\mathbf{M}_{\text{col}} \in \mathbb{C}^{M \times B_c K L}$ and $\{\mathbf{M}_{[\text{row},k,\ell]} \in \mathbb{C}^{N_{\text{RF}} B_r \times N} : k \in [K], \ell \in [L]\}$ are used for the first part to estimate the common column space (see section III-B) and the second part estimating the user and antenna specific coefficient matrices (see section III-C), respectively. Herein, the definitions of $\mathbf{Q}$, $\tilde{\mathbf{N}}_{\text{col}}$ and $\tilde{\mathbf{N}}_{[\text{row},k,\ell]}$ are provided with the detail training process for the RIS reflection vector and combining matrix design in the longer version of this paper [9]. ■

B. Estimation of Column Space

Recall that $\mathbf{S}_{\text{col}}$ is the $M \times \text{rk}(\mathbf{F})$ matrix whose columns are the basis of the column space of $\mathbf{F}$. To estimate it, we find the eigenvalue decomposition of the covariance matrix of $\mathbf{M}_{\text{col}}$:

$$\mathbf{M}_{\text{col}} \mathbf{M}_{\text{col}}^H = \tilde{\mathbf{S}}_{\text{col}} \mathbf{\Sigma}_{\text{col}} \tilde{\mathbf{S}}_{\text{col}}^H, \quad (12)$$

where $\tilde{\mathbf{S}}_{\text{col}}$ is the eigenvectors corresponding to the eigenvalue matrix $\mathbf{\Sigma}_{\text{col}} = \text{diag}([\lambda_{[\text{col},1]}, \lambda_{[\text{col},2]}, \dots, \lambda_{[\text{col},M]}])$, where the eigenvalues $\lambda_{[\text{col},\ell]}$'s are ordered in a descending order in magnitude. As suggested in [10], the $\text{rk}(\mathbf{F})$ (i.e., the rank of $\mathbf{F}$ (or $\mathbf{S}_{\text{col}}$)) can be efficiently estimated using the minimum description length (MDL) criterion [11] as

$$\min_{n \in [M]} \left\{ -\log \left(\frac{\prod_{i=n+1}^M \lambda_{[\text{col},i]}^{\frac{1}{M-n}}}{\frac{1}{M-n} \sum_{i=n+1}^M \lambda_{[\text{col},i]}} \right)^{(M-n)B_c T} + \frac{1}{2} n (2M - n) \log(B_c T) \right\}. \quad (13)$$

As long as independent columns are sufficiently sampled, the optimization below can yield a good $\text{rk}(\mathbf{F})$ -dimensional column space of $\mathbf{F}$ [12]:

$$\hat{\mathbf{S}}_{\text{col}} = \arg \min_{\mathbf{S} \in \mathbb{C}^{M \times \text{rk}(\mathbf{F})}, \mathbf{S}^H \mathbf{S} = \mathbf{I}} \left\| (\mathbf{I} - \mathbf{S} \mathbf{S}^H) \mathbf{M}_{\text{col}} \mathbf{M}_{\text{col}}^H \right\|_2^2. \quad (14)$$

The objective function can measure the distance between the column spaces of $\mathbf{S}$ and $\mathbf{M}_{\text{col}}$ since the difference of $\mathbf{S}\mathbf{S}^H\mathbf{M}_{\text{col}}$ (i.e., the projections of the columns of $\mathbf{M}_{\text{col}}$ onto the column space of $\mathbf{S}$) and $\mathbf{M}_{\text{col}}$ should be zero if $\mathbf{S}$ and $\mathbf{M}_{\text{col}}$ has the same column space. From Eckart–Young–Mirsky Theorem [13], the optimal solution to the above problem is derived as $\hat{\mathbf{S}}_{\text{col}} \in \mathbb{C}^{M \times \widehat{\text{rk}}(\mathbf{F})}$ which is given by choosing top $\widehat{\text{rk}}(\mathbf{F})$ eigenvectors of $\tilde{\mathbf{S}}_{\text{col}}$.

C. Estimation of Coefficient Matrices

We propose an iterative algorithm to estimate the coefficient matrices $\{\mathbf{T}_{[k,\ell]} : k \in [K], \ell \in [L]\}$ in (6). We first identify the some relationship among the effective channels due to their special structure:

$$\begin{aligned} \mathbf{H}_{[\text{eff},k,\ell]} &= \mathbf{F} \text{diag}(\mathbf{H}_k(:, \ell)) \\ &= \mathbf{F} [\text{diag}(\mathbf{H}_1(:, 1)) \text{diag}(\mathbf{H}_1(:, 1))^{-1}] \\ &\quad \times \text{diag}(\mathbf{H}_k(:, \ell)) \\ &= \mathbf{F} \text{diag}(\mathbf{H}_1(:, 1)) \mathbf{D}_{[k,\ell]} = \mathbf{H}_{[\text{eff},1,1]} \mathbf{D}_{[k,\ell]}, \end{aligned}$$

where $\mathbf{D}_{[k,\ell]} \triangleq \text{diag}(\mathbf{H}_1(:, 1))^{-1} \text{diag}(\mathbf{H}_k(:, \ell))$ is the $N \times N$ diagonal matrix. Since $\mathbf{H}_{[\text{eff},1,1]} = \mathbf{S}_{\text{col}} \mathbf{T}_{[1,1]}$, each coefficient matrix $\mathbf{T}_{[k,\ell]}$ can be expressed as

$$\mathbf{T}_{[k,\ell]} = \mathbf{T}_{[1,1]} \mathbf{D}_{[k,\ell]}, \quad k \in [K], \ell \in [L]. \quad (15)$$

We thus need to estimate the $\mathbf{T}_{[1,1]}$ and the diagonal matrices $\{\mathbf{D}_{[k,\ell]} : k \in [K], \ell \in [L]\}$.

Letting $\mathbf{P} \triangleq \Phi_{[M, N_{\text{RF}} B_r]}^H \hat{\mathbf{S}}_{\text{col}} \in \mathbb{C}^{N_{\text{RF}} B_r \times \widehat{\text{rk}}(\mathbf{F})}$, we directly derive the solution of the individual least-square (LS) estimation using the pseudo-inverse $\mathbf{P}^\dagger$:

$$\begin{aligned} \hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}} &\triangleq \mathbf{P}^\dagger \mathbf{M}_{[\text{row},k,\ell]} \in \mathbb{C}^{\widehat{\text{rk}}(\mathbf{F}) \times N} \\ &= \mathbf{P}^\dagger \left(\Phi_{[M, N_{\text{RF}} B_r]}^H \mathbf{H}_{[\text{eff},k,\ell]} + \tilde{\mathbf{N}}_{[\text{row},k,\ell]} \right) \\ &= \mathbf{P}^\dagger \left(\Phi_{[M, N_{\text{RF}} B_r]}^H \mathbf{S}_{\text{col}} \mathbf{T}_{[k,\ell]} + \tilde{\mathbf{N}}_{[\text{row},k,\ell]} \right). \quad (16) \end{aligned}$$

Here, the existence of the $\mathbf{P}^\dagger$ can be guaranteed by choosing the hyperparameter B_r such that

$$N_{\text{RF}} B_r \geq \widehat{\text{rk}}(\mathbf{F}). \quad (17)$$

Normally, we have $B_r = 1$ for very sparse channel, i.e., $N_{\text{RF}} \geq \widehat{\text{rk}}(\mathbf{F})$.

The LS method, which is referred as **CLRA-LS** in this paper, would be good in very high SNRs. Whereas, it is well-known that the multiplication of the pseudo-inverse would increase the power of the noise vector $\tilde{\mathbf{N}}_{[\text{row},k,\ell]}$, which can degrade the performance in practical SNRs. Thus, there is a room to enhance the estimation accuracy harnessing the property in (15).

$$\begin{aligned} &\arg \min_{\mathbf{T}_{[1,1]}, \mathbf{D}_{[1,1]}, \dots, \mathbf{D}_{[K,L]}} \mathcal{L}(\mathbf{T}_{[1,1]}, \{\mathbf{D}_{[k,\ell]}\}) \\ &\text{subject to } \{\mathbf{D}_{[k,\ell]}\} \text{ are diagonal matrices,} \quad (18) \end{aligned}$$

where

$$\mathcal{L}(\mathbf{T}_{[1,1]}, \{\mathbf{D}_{[k,\ell]}\}) \triangleq \sum_{k=1}^K \sum_{\ell=1}^L \left\| \hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}} - \mathbf{T}_{[1,1]} \mathbf{D}_{[k,\ell]} \right\|_F^2.$$

We efficiently solve this problem via an alternating optimization (see Algorithm 1).

Iterations. At the t -th iteration, the proposed alternating optimization is performed as follows:

- For the fixed $\hat{\mathbf{T}}_{[1,1]}^{(t-1)}$, our optimization can be formulated as

$$\begin{aligned} \hat{\mathbf{D}}_{[k,\ell]}^{(t)} &= \arg \min_{\mathbf{D}_{[k,\ell]}} \mathcal{L}(\hat{\mathbf{T}}_{[1,1]}^{(t-1)}, \mathbf{D}_{[k,\ell]} | \hat{\mathbf{T}}_{[1,1]}^{(t-1)}) \\ &= \arg \min_{\mathbf{D}_{[k,\ell]}} \left\| \hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}} - \hat{\mathbf{T}}_{[1,1]}^{(t-1)} \mathbf{D}_{[k,\ell]} \right\|_F^2, \quad (19) \end{aligned}$$

for $k \in [K]$, $\ell \in [L]$, where the initial value is determined as $\hat{\mathbf{T}}_{[1,1]}^{(0)} = \hat{\mathbf{T}}_{[1,1]}^{\text{LS}}$. This LS problem is easily solved as

$$\hat{\mathbf{D}}_{[k,\ell]}^{(t)} = \text{diag}(\mathbf{R}(:, \Omega_N)^\dagger \hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}}(\Omega_N)), \quad (20)$$

where $\mathbf{R} = \mathbf{I} \otimes \hat{\mathbf{T}}_{[1,1]}^{(t-1)} \in \mathbb{C}^{N_f N \times N^2}$ and $\Omega_N = \{n + (n-1)N : n \in [N]\}$. Also, vectorization of $\hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}}$ is denoted as $\hat{\mathbf{t}}_{[k,\ell]}^{\text{LS}} = \text{vec}(\hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}}) \in \mathbb{C}^{\widehat{\text{rk}}(\mathbf{F})N \times 1}$.

- For the fixed $\{\hat{\mathbf{D}}_{[k,\ell]}^{(t)} : k \in [K], \ell \in [L]\}$, our optimization can be formulated as a standard LS problem:

$$\begin{aligned} \hat{\mathbf{T}}_{[1,1]}^{(t)} &= \arg \min_{\mathbf{T}} \mathcal{L}(\mathbf{T}, \{\hat{\mathbf{D}}_{[k,\ell]}^{(t)}\} | \{\hat{\mathbf{D}}_{[k,\ell]}^{(t)}\}) \\ &= \arg \min_{\mathbf{T}} \sum_{k=1}^K \sum_{\ell=1}^L \left\| \hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}} - \mathbf{T} \hat{\mathbf{D}}_{[k,\ell]}^{(t)} \right\|_F^2, \quad (21) \end{aligned}$$

where $\hat{\mathbf{D}}_{[1,1]}^{(t)} = \mathbf{I}$ during iterations. The optimal solution is derived as

$$\begin{aligned} \hat{\mathbf{T}}_{[1,1]}^{(t)} &= \left(\sum_{k=1}^K \sum_{\ell=1}^L \hat{\mathbf{T}}_{[k,\ell]}^{\text{LS}} (\hat{\mathbf{D}}_{[k,\ell]}^{(t)})^H \right) \\ &\quad \times \left(\sum_{k=1}^K \sum_{\ell=1}^L \hat{\mathbf{D}}_{[k,\ell]}^{(t)} (\hat{\mathbf{D}}_{[k,\ell]}^{(t)})^H \right)^{-1}. \quad (22) \end{aligned}$$

After the t_{max} iterations, the estimated cascaded effective channels are given by

$$\hat{\mathbf{H}}_{[\text{eff},k,\ell]} = \hat{\mathbf{S}}_{\text{col}} \hat{\mathbf{T}}_{[1,1]}^{(t_{\text{max}})} \hat{\mathbf{D}}_{[k,\ell]}^{(t_{\text{max}})}. \quad (23)$$

D. Complexity Analysis

We derive the computational complexity of the proposed channel estimation method. Following the related work in [10], we measure the computational complexity by counting the number of complex multiplication (CM). Regarding the complexity of the estimation of column space in (14), we need to perform the eigenvalue decomposition, which requires $\mathcal{O}(M^3)$ computational complexity. Also, the dominant complexity of the proposed joint optimization in Algorithm 1 comes from the computations in (16), (20) and (22), which respectively require $\Delta_{\text{LS}} \triangleq N(\widehat{\text{rk}}(\mathbf{F}))^2 KL$, $\Delta_D \triangleq N(3\widehat{\text{rk}}(\mathbf{F}) + 1)KL$ and $\Delta_T \triangleq N(\widehat{\text{rk}}(\mathbf{F}) + 2)KL$. Thus, the overall computational complexity of the proposed CLRA-JO is computed as

$$\Psi_{\text{CLRA-JO}} = \mathcal{O}(M^3 + \Delta_{\text{LS}} + t_{\text{max}}(\Delta_D + \Delta_T)). \quad (24)$$

Remark 2: The coherence time of the BS-RIS channel (denoted by T_f) is commonly much longer than that of the RIS-User channels (denoted by T_h). Accordingly, $\mathbf{S}_{\text{col}}$ and $\mathbf{T}_{[k,\ell]}$'s are changed every T_f and T_h time slots, respectively. Harnessing this fact, we can easily reduce about $M_{\text{RF}}B_c$ subframes and $\mathcal{O}(M^3)$ computational complexity by reusing the estimated common column space for estimating user-specific coefficient matrices during T_f . ■

IV. SIMULATION RESULTS

In this section, we evaluate the performances of the proposed channel estimation method for XL-RIS assisted mmWave MU-MIMO systems with hybrid beamforming structures. We adopt the normalized mean square error (NMSE) for the evaluation of channel estimation accuracy, given by

$$\text{NMSE} \triangleq \mathbb{E} \left[\frac{1}{K} \sum_{k=1}^K \frac{\|\hat{\mathbf{H}}_{[\text{eff},k]} - \mathbf{H}_{[\text{eff},k]}\|_F^2}{\|\mathbf{H}_{[\text{eff},k]}\|_F^2} \right]. \quad (25)$$

The expectation is evaluated by Monte Carlo simulations with 10^3 trials.

For our simulations, we consider the MU-MIMO system with $M = N = 128 \times 2 = 256$, $L = 2 \times 2 = 4$, $N_{\text{RF}} = 8$, the number of spatial paths $N_f = N_{h_k} = 3$, $\forall k \in [K]$. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = 10 \log(P/\sigma^2)$ [dB]. Specifically, we set that $\sigma^2 = 1$ and P is described as a relative value of the noise power. Regarding the modeling of wireless channels, the power is equally divided into each signal path, the i -th normalized complex channel gains of $\{\mathbf{H}_k : k \in [K]\}$ and $\mathbf{F}$ are respectively determined as $\alpha_{h_k}^i = \sqrt{NL/N_{h_k}}\psi_{h_k}^i$ and $\alpha_f^i = \sqrt{MN/N_f}\psi_f^i$, where $\psi_{h_k}^i$ and ψ_f^i are chosen uniformly and randomly from $[0, 2\pi)$. The angle of all AoDs and AoAs are chosen uniformly and randomly from $[0, 2\pi)$.

For comparison, the following benchmark methods are considered:

- **S-MMV:** The channels are estimated by projecting the received signals onto the common column space first, and solved by formulating a multiple measurements vector (MMV) Problem. Here, the MMV problem is addressed via SOMP algorithm. (see Algorithm 1 in [10]). This method is originally constructed based on the far-far field channels
- **S-MJCE:** The channels are estimated by the two-step (subspace) multi-user joint channel estimation procedure. (see Algorithm 2 in [10]). This method is considered as the best-known CS-based method. This method is originally constructed based on the far-far field channels.
- **3D-M-LAOMP:** The channels are estimated by the two-phase multi-user joint channel estimation procedure using 3D MMV-based look ahead OMP (3D-M-LAOMP) in [6]. This method is only applicable in the far-near field channels.

In the benchmark CS-based methods, the resolution of the quantization for the dictionary are set by $G = 512$. In the

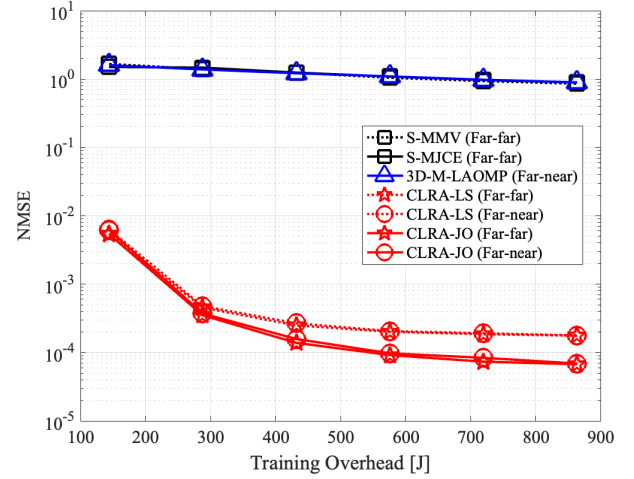


Fig. 2. The impact of the training overhead on the NMSE. SNR = 0 dB and $K = 4$.

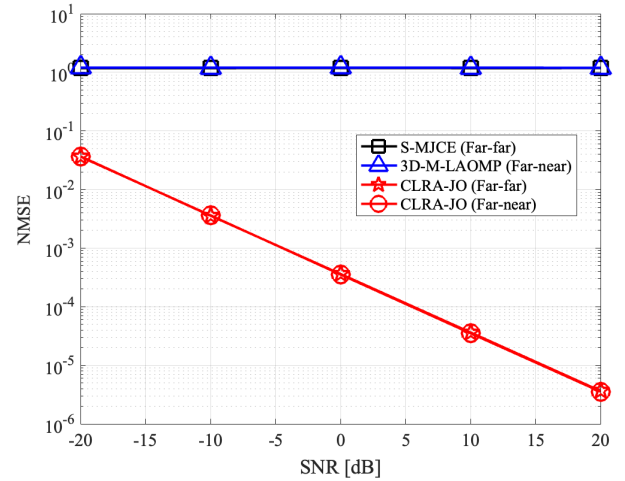


Fig. 3. The impact of SNR on the NMSE. $K = 4$ and $B_c = 10$.

proposed CLRA-JO, we choose the $B_r = 1$ and $t_{\text{max}} = 10$ in Algorithm 1.

Fig. 2 show the impact of training overhead on the NMSE. In the both systems, the proposed CLRA-based methods can outperform the existing CS-based methods in all the categories of wireless channels. From the performance comparison between CLRA-JO and CLRA-LS, we can verify the effectiveness of the proposed joint optimization of user-specific coefficient matrices in Section III-C. Noticeably, the proposed CLRA-JO can achieve higher channel estimation accuracy than the SOTA CS-based methods. Moreover, the proposed CLRA-JO can show the almost similar performances regardless of the categories of wireless channels, which verifies that CLRA-JO is indeed the unified method.

Fig. 3 shows the impact of SNR on the NMSE. We first observe that the performance of CLRA-JO can improve as SNR grows. On the other hand, the performances of the

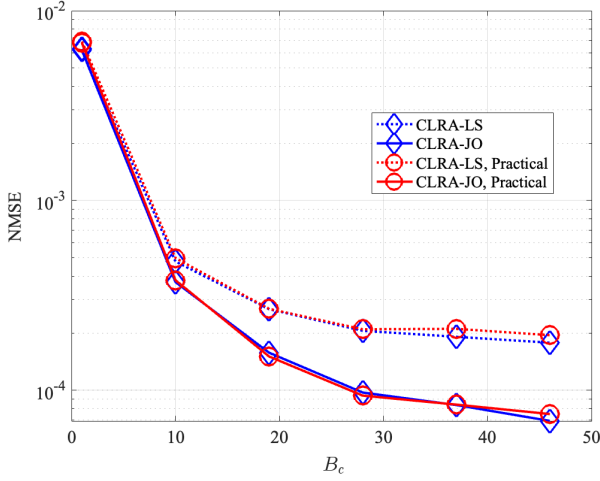


Fig. 4. The impact of the B_c on the NMSE for *practical* systems. SNR = 0 dB and $K = 4$.

CS-based methods are saturated due to the inevitable grid-mismatch problem. Note that such saturation level is determined according to the quantization resolution of steering vectors. Following the *Remark 2*, the proposed CLRA-JO requires only 2.8% computational complexity of S-MMV (i.e., the CS-based method having the lowest computational complexity). Here, the computational complexity of the S-MMV is provided in [10] such as

$$\Psi_{\text{S-MMV}} = \mathcal{O} \left(M^3 + N^2(\text{rk}(\mathbf{F})KL + G) \right).$$

Together with the results in Fig. 3, it is verified that the proposed CLRA-JO is more suitable for practical UPA-equipped systems than the CS-based methods.

Fig. 4 depicts the estimation accuracy of the proposed methods in a more *practical* system where the near-field array response vector is determined by considering the element-wise path loss [14]:

$$\mathbf{b}_N(d, \theta, \phi) \triangleq \frac{\mathbf{u}_N(d, \theta, \phi)}{\|\mathbf{u}_N(d, \theta, \phi)\|_2} \in \mathbb{C}^{N \times 1}, \quad (26)$$

where

$$\mathbf{u}_N(d, \theta, \phi) = \left[\frac{e^{-j\frac{2\pi}{\lambda}\Delta d(1,1)}}{\Delta d(1,1)}, \dots, \frac{e^{-j\frac{2\pi}{\lambda}\Delta d(N_h, N_v)}}{\Delta d(N_h, N_v)} \right]^\top.$$

and the users may experience different channel categories depending on their locations. For this, each RIS-User channel is determined uniformly and randomly from the far and near field channels. Since the benchmark methods cannot perform in this scenario, we only evaluate the performances of the proposed CLRA-LS and CLRA-JO. We observe that the proposed methods can achieve almost the same performance regardless of the categories of wireless channels. This implies that the proposed methods can perform well as long as the low-rankness of the BS-RIS channel (i.e., the low-rankness of $\mathbf{F}$) is guaranteed. These results can support the practicality of the proposed methods.

Based on the above analysis, we can conclude that the proposed CLRA-JO would be a good practical candidate as the channel estimation method for XL-RIS assisted MU-MIMO systems.

V. CONCLUSION

We studied the channel estimation problem for XL-RIS assisted MU-MIMO systems with hybrid beamforming structures. In this system, we proposed the unified channel estimation method (named CLRA-JO) which can perform in the far- and near-field channels without any modification. Whereas, in the existing CS-based methods, dictionary should be designed by taking into account the characteristics of the near- and far-field channels. Via simulations and complexity analysis, it is demonstrated that the proposed CLRA-JO can yield better estimation accuracy than the state-of-the-art CS-based methods while having lower training overhead.

REFERENCES

- [1] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE journal on selected areas in communications*, vol. 38, no. 11, pp. 2450–2525, 2020.
- [2] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active intelligent reflecting surface for simo communications," in *IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.
- [3] H. Wymeersch, J. He, B. Denis, A. Clemente, and M. Juntti, "Radio localization and mapping with reconfigurable intelligent surfaces: Challenges, opportunities, and research directions," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 52–61, 2020.
- [4] M. Cui and L. Dai, "Channel estimation for extremely large-scale mimo: Far-field or near-field?" *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2663–2677, 2022.
- [5] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.
- [6] S. Yang, W. Lyu, Z. Hu, Z. Zhang, and C. Yuen, "Channel estimation for near-field xl-ris-aided mmwave hybrid beamforming architectures," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 8, pp. 11 029–11 034, 2023.
- [7] X. Yu, W. Shen, R. Zhang, C. Xing, and T. Q. S. Quek, "Channel estimation for xl-ris-aided millimeter-wave systems," *IEEE Transactions on Communications*, vol. 71, no. 9, pp. 5519–5533, 2023.
- [8] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive mimo: A survey," *IEEE Communications magazine*, vol. 55, no. 9, pp. 134–141, 2017.
- [9] J. Lee, H. Chung, Y. Cho, S. Kim, and S. Hong, "Near-field channel estimation for xl-ris assisted multi-user xl-mimo systems: Hybrid beamforming architectures," 2024. [Online]. Available: <https://arxiv.org/abs/2401.06966>
- [10] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmwave mimo systems," *IEEE Transactions on Wireless Communications*, no. 10, pp. 6853–6869, 2023.
- [11] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 33, no. 2, pp. 387–392, 1985.
- [12] J. Chae, P. Narayanamurthy, S. Bac, S. M. Sharada, and U. Mitra, "Column-based matrix approximation with quasi-polynomial structure," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.
- [13] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.
- [14] H. Wang, Z. Xiao, and Y. Zeng, "Cramér-rao bounds for near-field sensing with extremely large-scale mimo," *IEEE Transactions on Signal Processing*, vol. 72, pp. 701–717, 2024.