

Mutual Coupling Aware Channel Estimation for RIS-Aided Multi-User mmWave Systems

Tian Qiu, Ruidong Li, Cunhua Pan, *Senior Member, IEEE*, Taihao Zhang, Dongnan Xia, Changhong Wang, and Hong Ren, *Member, IEEE*

Abstract—This paper proposes a three-stage uplink channel estimation protocol for reconfigurable intelligent surface (RIS)-aided multi-user (MU) millimeter-wave (mmWave) multiple-input single-output (MISO) systems, where both the base station (BS) and the RIS are equipped with uniform planar arrays (UPAs). The proposed approach explicitly accounts for the mutual coupling (MC) effect, modeled via scattering parameter multipoint network theory. In Stage I, a dimension-reduced subspace-based method is proposed to estimate the common angle of arrival (AoA) at the BS using the received signals across all users. In Stage II, MC-aware cascaded channel estimation is performed for a typical user. The equivalent measurement vectors for each cascaded path are extracted and the reference column is reconstructed using a compressed sensing (CS)-based approach. By leveraging the structure of the cascaded channel, the reference column is rearranged to estimate the AoA at the RIS, thereby reducing the computational complexity associated with estimating other columns. Additionally, the common angle of departure (AoD) at the RIS is also obtained in this stage, which significantly reduces the pilot overhead for estimating the cascaded channels of other users in Stage III. The RIS phase shift training matrix is designed to optimize performance in the presence of MC and outperforms random phase scheme. Simulation results validate that the proposed method yields better performance than the MC-unaware and existing approaches in terms of estimation accuracy and pilot efficiency.

Index Terms—Reconfigurable intelligent surface, mutual coupling, millimeter wave, channel estimation

I. INTRODUCTION

RECONFIGURABLE intelligent surface (RIS) has emerged as a promising technology for 6G-and-beyond communication systems [1], [2]. An RIS comprises a large number of low-cost, deployment-convenient reflective elements. By appropriately tuning the phase shifts, the reflected signals can be constructively combined at the intended receiver to enhance received signal power, while undesired reflections can be suppressed through destructive interference, thereby mitigating co-channel interference [3], [4]. These advantages makes RIS particularly attractive for millimeter-wave (mmWave) multiple-input multiple-output (MIMO) systems. However, fully exploiting the potential of RIS requires accurate channel state information (CSI), which remains challenging to obtain. The difficulty arises from the large number of antennas at the base station (BS) and

reflective elements at the RIS, necessitating efficient pilot overhead reduction strategies. Furthermore, RIS elements are typically passive, lacking active radio frequency (RF) chains and signal processing capabilities, rendering conventional channel estimation techniques inapplicable [5], [6]. More critically, emerging RIS hardware designs introduce additional impairments, most notably the mutual coupling (MC) effect between RIS elements, which significantly alters the cascaded channel characteristics [7]–[10]. Consequently, channel estimation for RIS-aided systems has become a focal point of intense research.

Several studies have addressed channel estimation under sparse structured channel models, which exploit the inherent sparsity of high-frequency mmWave channels to reduce pilot overhead and improve estimation accuracy by leveraging compressed sensing (CS) techniques [11], direction-of-arrival (DOA) estimation methods [12], and sparse Bayesian learning (SBL) frameworks [13]. The authors of [14] proposed a CS-based channel estimation method by exploiting the inherent sparse structure of the user-RIS-BS cascaded channel. In [15], the authors exploited the fact that the sparse channel matrices of the cascaded channels of all users share a common row-column-block sparsity structure arising from the shared RIS-BS channel, and proposed a multi-user (MU) joint iterative channel estimator based on this observation. Inspired by the common column-block sparsity property, the double-structured sparsity of the angular cascaded channels among users was exploited in [16] to propose a double-structured orthogonal matching pursuit (DS-OMP) algorithm, thereby reducing pilot overhead. A significant reduction in pilot overhead for MU mmWave systems was achieved in [17] by fully exploiting the correlation among different cascaded channels. The authors of [18] extended the channel strategy in [17] to the uniform planar array (UPA)-type MU-MIMO case.

Nevertheless, these works are based on the conventional RIS-aided channel model, which assumes that each RIS element radiates electromagnetic (EM) waves independently and neglects the nonlinear coupling between elements. Notable efforts have been dedicated to accurately modeling MC in RIS-aided communication systems. For instance, microwave multipoint network theory extends fundamental circuit and network concepts to facilitate complex microwave analysis in modeling MC. A circuit-based communication model for RIS-aided wireless systems based on the mutual impedances between all radiating elements was introduced in [7]. The authors of [8] were the first to derive an MC-aware beyond-diagonal (BD) RIS-aided wireless communication model using scattering (S -parameter) and impedance (Z -parameter) representations, and

Tian Qiu, Cunhua Pan, Taihao Zhang, Dongnan xia and Hong Ren are with National Mobile Communications Research Laboratory, Southeast University, Nanjing, China (e-mail: tianqiu, cpan, taihao, dnxia, hren@seu.edu.cn).

Ruidong Li is the co-first author. Ruidong Li and Changhong Wang are with Shandong Yunhai Guochuang Innovative Technology Co., Ltd., Jinan, China (e-mail: lird, wangchh01@inspur.com).

Corresponding author: Cunhua Pan.

proved their equivalence. The authors of [9] proposed new optimization algorithms based on the S -parameter multiport network model and discussed its advantages compared with its counterpart based on the Z -parameter representation. The authors of [10] employed scattering parameter network analysis to derive a physically and EM compliant yet straightforward and tractable RIS-aided communication model that fully accounts for the effects of impedance mismatching and mutual coupling at the transmitter, RIS, and receiver. Despite the importance of considering the MC effect in RIS-aided channel model, few studies have addressed channel estimation for RIS-aided systems in the presence of MC. The authors of [19] proposed a two-stage MC-aware channel estimator that performs effectively under strong MC conditions by employing CS and dictionary reduction (DR) techniques for an active RIS-aided multiple-input single-output (MISO) system. However, this work assumed only a single user, and is therefore unsuitable for practical multi-user scenarios as well as the ubiquitous connectivity envisioned in 6G systems.

Against the above background, we propose an MC-aware three-stage channel estimation protocol for RIS-aided MU-MISO mmWave communication systems with single-antenna users, where both the BS and the RIS are equipped with UPAs. This work advances RIS-aided MU-MISO channel estimation by explicitly incorporating the MC effect among RIS elements, which enables accurate characterization of cascaded channel under practical RIS deployment conditions. The main contributions of this paper are summarized as follows:

- We propose an MC-aware three-stage channel estimation protocol for RIS-aided MU-MISO systems based on the S -parameter multiport network model. In Stage I, the common AoA at the BS is estimated using the received signals from all users. In Stage II, complete CSI estimation is performed for a typical user. After eliminating the common AoA influence, the equivalent measurement vectors for each cascaded path are extracted, and the reference column is reformulated using an OMP-based approach. By exploiting the structure of the cascaded channel, the elements of the reference column are rearranged to estimate the AoA at the RIS for the typical user. This step reduces the computational complexity of estimating the remaining columns. Additionally, the common AoD at the RIS is obtained in Stage II, thereby significantly reducing the pilot overhead for estimating the cascaded channels of other users in Stage III.
- In Stage I of the protocol, we propose an effective dimension-reduced subspace-based method to estimate common AoA at the BS by processing the received signals from all users. The use of a UPA configuration in this paper necessitates extending subspace-based methods to two-dimensional (2-D) versions, which substantially increases computational complexity. To address this challenge, we propose a dimension-reduced approach that reduces computational burden while maintaining compatibility with existing one-dimensional (1-D) subspace-based algorithms by exploiting the UPA structure and leveraging the properties of the Kronecker product.

Specifically, we evaluate the performance of root multiple signal classification (Root-MUSIC) and total least squares estimation of signal parameters via rotational invariance techniques (TLS-ESPRIT) algorithms for estimation in this stage.

- We optimize the equivalent RIS phase shift training matrices to enforce orthogonality among the columns of the equivalent dictionary and to enhance the performance of the OMP-based channel estimation, since MC inherently reduces the orthogonality of the equivalent dictionary. To this end, an efficient Riemannian manifold optimization framework is developed, which achieves performance gains compared to the random phase scheme. Furthermore, we analyze the pilot overhead and computational complexity of the proposed estimation protocol.
- We present a channel estimation scheme tailored to the conventional RIS-aided cascaded channel model that neglects the MC effect. This scheme is applied to received signals affected by MC and serves as a baseline for comparison with the proposed and existing MC-aware algorithms. This comparative analysis rigorously characterizes the impact of MC on channel estimation performance and highlights the advantages of the proposed MC-aware estimation protocol.

The remainder of this paper is organized as follows. Section II introduces the system model. Section III details the proposed MC-aware three-stage channel estimation protocol, which includes full CSI estimation for the typical user in Stages I and II, and for other users in Stage III. This section also presents the MC-unaware channel estimation scheme and an analysis of the pilot overhead and computational complexity. The design of phase shift training matrices is presented in Section IV. Section V provides simulation results. Finally, Section VI concludes the paper.

Notations: Vectors and matrices are denoted by boldface lowercase and uppercase letters, respectively. For a matrix $\mathbf{X}$ of arbitrary size, the symbols $\mathbf{X}^*$, $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathbf{X}^\dagger$ represent the conjugate, transpose, Hermitian, and pseudo-inverse of matrix $\mathbf{X}$, respectively. For a square full-rank matrix $\mathbf{X}$, $\mathbf{X}^{-1}$ denotes its inverse. The Frobenius norm of matrix $\mathbf{X}$ is denoted by $\|\mathbf{X}\|_F$, and the Euclidean norm of vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_2$. The modulus of a scalar is denoted by $|\cdot|$, the norm of a vector by $\|\cdot\|$. $\text{Diag}(\mathbf{x})$ denotes a diagonal matrix with the entries of vector $\mathbf{x}$ on its main diagonal. The vectorization operator $\text{vec}(\mathbf{X})$ stacks the columns of $\mathbf{X}$ into a column vector, whereas $\text{mat}(\mathbf{X})$ reshapes $\mathbf{X}$ into a matrix of specified dimensions following column-wise ordering. The expectation operator is denoted by $\mathbb{E}[\cdot]$. $[\mathbf{x}]_m$ denotes the m -th element of vector $\mathbf{x}$, and $[\mathbf{X}]_{m,n}$ denotes the (m,n) -th element of matrix $\mathbf{X}$. The m -th row and n -th column of matrix $\mathbf{X}$ are represented by $\mathbf{X}_{m,:}$ and $\mathbf{X}_{:,n}$, respectively. Additionally, the Kronecker product, Hadamard product, Khatri-Rao product, and transposed Khatri-Rao product between two matrices $\mathbf{X}$ and $\mathbf{Y}$ are denoted by $\mathbf{X} \otimes \mathbf{Y}$, $\mathbf{X} \odot \mathbf{Y}$, $\mathbf{X} \diamond \mathbf{Y}$ and $\mathbf{X} \bullet \mathbf{Y}$, respectively. $\lfloor x \rfloor$ rounds down to the nearest integer.

II. SYSTEM MODEL

A narrow-band time-division duplex (TDD) mmWave system is considered, where K single-antenna users communicate with a BS equipped with an UPA comprising $N = N_h \times N_v$ antennas, where N_h and N_v denote the numbers of horizontal and vertical elements. To improve communication efficiency, an RIS is deployed, which comprises a passive reflecting UPA with $M = M_h \times M_v$ elements (M_h horizontal elements and M_v vertical elements). In this model, it is assumed that the direct channels between the users and the BS are blocked. In the following subsections, the RIS-BS and user-RIS subchannels are characterized and then the cascaded channel model is introduced while accounting for the MC effects among the elements of RIS.

A. The RIS-BS Subchannel

The narrow-band frequency-domain mmWave channel from the RIS to the BS, denoted as $\mathbf{H} \in \mathbb{C}^{N \times M}$, is given by

$$\mathbf{H} = \sum_{l=1}^L \alpha_l \mathbf{a}_N(\psi_l, \nu_l) \mathbf{a}_M^H(\omega_l, \mu_l), \quad (1)$$

where L denotes the number of propagation paths between the RIS and the BS, and the parameters α_l , (ψ_l, ν_l) , and (ω_l, μ_l) denote the complex channel gain, the angle of arrival (AoA) at the BS, and the angle of departure (AoD) at the RIS for the l -th path in the RIS-BS channel, respectively.

The Saleh-Valenzuela (SV) model is employed [20] to characterize the channel by exploiting the limited scattering characteristics of mmWave propagation. Define the array response vector (ARV) under a UPA configuration as $\mathbf{a}_X(y, z) \in \mathbb{C}^{X \times 1}$, where $X = X_h \times X_v$, i.e.,

$$\mathbf{a}_X(z, y) = \mathbf{a}_{X_v}(z) \otimes \mathbf{a}_{X_h}(y), \quad (2)$$

where $\mathbf{a}_{X_v}(z) = [1, e^{-i2\pi z}, \dots, e^{-i2\pi(X_v-1)z}]^T$ and $\mathbf{a}_{X_h}(y) = [1, e^{-i2\pi y}, \dots, e^{-i2\pi(X_h-1)y}]^T$, denoting the steering vectors for the vertical (z -axis) and horizontal (y -axis) UPA dimensions, respectively. The variables z and y represent the equivalent spatial frequencies in the vertical and horizontal planes of the UPA, respectively.

Define the elevation angle $\varrho \in [-90^\circ, 90^\circ]$ and the azimuth angle $\varsigma \in [-180^\circ, 180^\circ]$ of the incoming signal. The spatial frequency pair (z, y) is related to the physical angle pair (ϱ, ς) as follows, $z = \frac{d}{\lambda_c} \sin(\varrho)$ and $y = \frac{d}{\lambda_c} \sin(\varsigma) \cos(\varrho)$, where λ_c is the carrier wavelength and d is the inter-element spacing of the UPA. If $d \leq \frac{\lambda_c}{2}$, a one-to-one mapping exists between spatial frequencies and physical angles on one side of the UPA [17]. We assume this relationship holds throughout the paper, and refer to the steering vector arguments interchangeably as either physical angles or spatial frequencies.

Moreover, the RIS-BS channel in Eq. (1) can be written in a more compact way as

$$\mathbf{H} = \mathbf{A}_N \mathbf{A} \mathbf{A}_M^H, \quad (3)$$

where

$$\mathbf{A}_N = [\mathbf{a}_N(\psi_1, \nu_1), \dots, \mathbf{a}_N(\psi_L, \nu_L)] \in \mathbb{C}^{N \times L}, \quad (4a)$$

$$\mathbf{A} = \text{Diag}(\alpha_1, \dots, \alpha_L) \in \mathbb{C}^{L \times L}, \quad (4b)$$

$$\mathbf{A}_M = [\mathbf{a}_M(\omega_1, \mu_1), \dots, \mathbf{a}_M(\omega_L, \mu_L)] \in \mathbb{C}^{M \times L}. \quad (4c)$$

B. The User-RIS Subchannel

Similarly, the frequency-domain channel from user k to the RIS, denoted as $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$, can be represented as

$$\mathbf{h}_k = \sum_{j=1}^{J_k} \beta_{k,j} \mathbf{a}_M(\varphi_{k,j}, \theta_{k,j}), \forall k \in \mathcal{K}, \quad (5)$$

where J_k denotes the number of propagation paths between user k and the RIS, and $\beta_{k,j}$ is the complex channel gain of the j -th path in the user-RIS channel. The pair $(\varphi_{k,j}, \theta_{k,j})$ specifies the AoA of the j -th path from user k to the RIS. The set of users is defined as $\mathcal{K} = \{1, 2, \dots, K\}$. The k -th user-RIS channel in Eq. (5) can also be represented in a more compact form as

$$\mathbf{h}_k = \mathbf{A}_{M,k} \boldsymbol{\beta}_k, \quad (6)$$

where

$$\mathbf{A}_{M,k} = [\mathbf{a}_M(\varphi_{k,1}, \theta_{k,1}), \dots, \mathbf{a}_M(\varphi_{k,J_k}, \theta_{k,J_k})] \in \mathbb{C}^{M \times J_k}, \quad (7a)$$

$$\boldsymbol{\beta}_k = [\beta_{k,1}, \dots, \beta_{k,J_k}]^T \in \mathbb{C}^{J_k \times 1}. \quad (7b)$$

C. Cascaded Channel Model

Since direct channels between the users and BS are assumed to be blocked, hence, the conventional RIS-aided channel model without MC between elements of the RIS is given by

$$\mathbf{h}_{cv,k} = \mathbf{H} \boldsymbol{\Gamma} \mathbf{h}_k, \quad (8)$$

where $\boldsymbol{\Gamma} = \text{Diag}(\boldsymbol{\gamma})$, with $\boldsymbol{\gamma} \in \mathbb{C}^{M \times 1}$ denoting the RIS phase shift vector. Eq. (8) reveals that the received signal at the BS is a linear combination of reflections from each element. This model assumes that each RIS element reflects incident EM waves independently, neglecting any MC effects. When strong MC occurs among the elements of RIS, the assumption can cause model mismatch and hinder tasks such as channel estimation, thereby degrading the overall communication performance. Based on the S -parameter multiport network theory, an MC-aware communication model has been recently derived and validated [8], [9], [19]. Incorporating MC into the RIS response, the channel model in Eq. (8) is reformulated as:

$$\mathbf{h}_{mc,k} = \mathbf{H} (\boldsymbol{\Gamma}^{-1} - \mathbf{S})^{-1} \mathbf{h}_k, \quad (9)$$

where $\mathbf{S} \in \mathbb{C}^{M \times M}$ denotes the RIS scattering matrix capturing MC among the elements of RIS. The scattering and impedance matrices are related by $\mathbf{S} = (\mathbf{Z} + Z_0 \mathbf{I}_M)^{-1} (\mathbf{Z} - Z_0 \mathbf{I}_M)$ [8], where $\mathbf{Z} \in \mathbb{C}^{M \times M}$ is the RIS impedance matrix and Z_0 is the characteristic impedance. The explicit expression of $\mathbf{Z}$ is provided in Eq. (10), in which Z_{qp} denotes the MC effect of the RIS radiating element p on element q , while the definitions of the remaining parameters are provided in [7], [21]. Unless

$$Z_{qp} = \frac{j\eta_0}{4\pi k_0} \int_{-l_q/2}^{l_q/2} \int_{-l_p/2}^{l_p/2} \frac{e^{-jk_0 R(\xi, z)} \sin(k_0(l_p/2 - |\xi|)) \sin(k_0(l_q/2 - |z|))}{R(\xi, z) \sin(k_0 l_p/2) \sin(k_0 l_q/2)} \left(k_0^2 - \frac{jk_0}{R(\xi, z)} - \frac{k_0^2(z - \xi + \rho_2)^2 + 1}{R^2(\xi, z)} + \frac{3jk_0(z - \xi + \rho_2)^2}{R^3(\xi, z)} + \frac{3(z - \xi + \rho_2)^2}{R^4(\xi, z)} \right) d\xi dz. \quad (10)$$

otherwise specified, channel parameters implicitly include MC and will be denoted without the subscript mc.

Comparing Eq. (8) with Eq. (9), it can be observed that each RIS element no longer radiates EM waves independently when the MC effect is taken into account and exhibits non-linear coupling between elements. Furthermore, the MC effect influences signal processing and degrades the orthogonality of the equivalent dictionary in OMP-based approaches, as shown in the subsequent analysis, thereby introducing additional challenges for channel estimation.

D. The Received Signal

Denote $s_k(t)$ as the pilot signal of the k -th user. To eliminate inter-user interference, the users employ a one-by-one transmission scheme. Accordingly, during the uplink transmission, the received signal from user k in time slot t , where $1 \leq t \leq \tau_k$, can be expressed as

$$\begin{aligned} \mathbf{y}_k(t) &= \mathbf{H}(\mathbf{\Gamma}_t^{-1} - \mathbf{S})^{-1} (\sqrt{p} \mathbf{h}_k s_k(t) + \mathbf{n}_{2,k}(t)) + \mathbf{n}_{1,k}(t) \\ &\triangleq \sqrt{p} \mathbf{H}(\mathbf{\Gamma}_t^{-1} - \mathbf{S})^{-1} \mathbf{h}_k s_k(t) + \mathbf{n}_k(t), \end{aligned} \quad (11)$$

where $\mathbf{n}_k(t) = \mathbf{H}(\mathbf{\Gamma}_t^{-1} - \mathbf{S})^{-1} \mathbf{n}_{2,k}(t) + \mathbf{n}_{1,k}(t)$ and p represents transmit power of each user. $\mathbf{n}_{1,k}(t) \in \mathbb{C}^{N \times 1}$ and $\mathbf{n}_{2,k}(t) \in \mathbb{C}^{M \times 1}$ model the thermal noise at the BS and RIS, respectively, which are distributed as $\mathbf{n}_{1,k}(t) \sim \mathcal{CN}(0, \sigma_1^2 \mathbf{I}_N)$ and $\mathbf{n}_{2,k}(t) \sim \mathcal{CN}(0, \sigma_2^2 \mathbf{I}_M)$. In most existing works, $\mathbf{n}_{2,k}(t)$ is neglected because its contribution is comparatively small relative to the noise power at the BS after reflection and going through backward channel. However, due to MC, the noise introduced at the RIS cannot be ignored. Furthermore, incorporating this component enables a more accurate quantification of signal-to-noise ratio (SNR) degradation in RIS-aided systems by capturing correlated noise effects at the RIS.

Assuming that the pilot symbols satisfy $s_k(t) = 1$ for $1 \leq t \leq \tau_k$ and defining $\mathbf{B}_t \triangleq (\mathbf{\Gamma}_t^{-1} - \mathbf{S})^{-1}$, the received signal can be rewritten as

$$\begin{aligned} \mathbf{y}_k(t) &= \sqrt{p} \mathbf{H} \mathbf{B}_t \mathbf{h}_k + \mathbf{n}_k(t) \\ &= \sqrt{p} \text{vec}(\mathbf{A}_N \mathbf{\Lambda} \mathbf{A}_M^H \mathbf{B}_t \mathbf{A}_{M,k} \beta_k) + \mathbf{n}_k(t) \\ &= \sqrt{p} (\beta_k^T \otimes \mathbf{A}_N \mathbf{\Lambda}) \text{vec}(\mathbf{A}_M^H \mathbf{B}_t \mathbf{A}_{M,k}) + \mathbf{n}_k(t) \\ &= \sqrt{p} (1 \otimes \mathbf{A}_N) (\beta_k^T \otimes \mathbf{\Lambda}) (\mathbf{A}_{M,k}^T \otimes \mathbf{A}_M^H) \text{vec}(\mathbf{B}_t) + \mathbf{n}_k(t) \\ &\triangleq \sqrt{p} \mathbf{A}_N (\beta_k^T \otimes \mathbf{\Lambda}) (\mathbf{A}_{M,k}^T \otimes \mathbf{A}_M^H) \boldsymbol{\eta}_t + \mathbf{n}_k(t), \end{aligned} \quad (12)$$

where $\boldsymbol{\eta}_t = \text{vec}(\mathbf{B}_t)$ is the RIS phase shift vector incorporating the MC effect. Stacking the τ_k received signal vectors across time slots yields the received signal matrix $\mathbf{Y}_k \in \mathbb{C}^{N \times \tau_k}$, which can be written as

$$\mathbf{Y}_k = [\mathbf{y}_k(1), \dots, \mathbf{y}_k(\tau_k)] \triangleq \sqrt{p} \mathbf{G}_k \boldsymbol{\Theta}_k + \mathbf{N}_k, \quad (13)$$

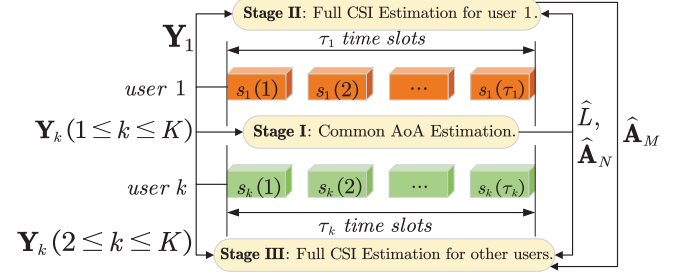


Fig. 1. The proposed MC-aware three-stage channel estimation protocol.

where $\mathbf{G}_k = \mathbf{A}_N (\beta_k^T \otimes \mathbf{\Lambda}) (\mathbf{A}_{M,k}^T \otimes \mathbf{A}_M^H)$ defines the equivalent MC-aware cascaded user-RIS-BS channel for user k , which we aim to estimate. Moreover, $\boldsymbol{\Theta}_k = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{\tau_k}] \in \mathbb{C}^{M^2 \times \tau_k}$ denotes the MC-aware phase shift training matrix of the RIS and $\mathbf{N}_k = [\mathbf{n}_k(1), \dots, \mathbf{n}_k(\tau_k)] \in \mathbb{C}^{N \times \tau_k}$ represents the noise matrix for user k . The high dimensionality of $\boldsymbol{\Theta}_k$, together with the consideration of MC, will degrade the orthogonality of the equivalent dictionary.

III. MC-AWARE CHANNEL ESTIMATION

A. Channel Estimation Protocol

Fig. 1 illustrates the proposed MC-aware three-stage uplink channel estimation protocol. In Stage I, we employ a dimension-reduced subspace-based method to estimate the common AoA at the BS, thereby reducing computational complexity while maintaining compatibility with various one-dimensional subspace-based estimation algorithms. Moreover, we estimate the common AoA with the received signals of all users, effectively mitigating error propagation across stages. In Stage II, we perform MC-aware cascaded channel estimation for user 1, designated as the typical user. Specifically, the user located closest to the RIS is selected as the typical user due to its lower path loss in the reflecting link, leading to a strong received signal at the BS and enabling more accurate channel estimation. With the common AoA components eliminated, the equivalent measurement vectors for each cascaded path are extracted, and the reference column is reconstructed using an OMP-based approach. By exploiting the structure of the cascaded channel and leveraging the properties of the Kronecker product, we rearrange the elements of the reference column to estimate the AoA at the RIS for user 1. This reduces the computational complexity of estimating the other columns (cascaded paths). Additionally, the common AoD at the RIS can be obtained in Stage II. Finally, in Stage III, we exploit the structural characteristics of the common RIS-BS channel shared by all users to substantially reduce both

the pilot overhead and computational complexity involved in estimating the MC-aware cascaded channels for other users. The following subsections present a detailed elaboration of the proposed MC-aware channel estimation protocol.

B. Stages I and II: Channel Estimation for Typical User

This subsection outlines a comprehensive procedure for estimating the full CSI of the typical user, denoted as user 1. In Stage I, a dimension-reduced subspace-based method is employed to estimate the common AoA at the BS by processing the received signal of all users, thereby mitigating error propagation. In Stage II, we estimate the MC-aware cascaded channel for user 1 by reformulating the equivalent measurement vector for each cascaded path. The channel is then reconstructed using an OMP-based approach. This stage also yields the estimate of the common AoD at the RIS, which subsequently reduces the pilot overhead for the other users.

1) *Estimation of Common AoA at the BS:* From Eq. (13), the received signal matrix $\mathbf{Y}_k \in \mathbb{C}^{N \times \tau_k}$ can be expressed as

$$\mathbf{Y}_k = \sqrt{p} \mathbf{G}_k \mathbf{\Theta}_k + \mathbf{N}_k \triangleq \sqrt{p} \mathbf{A}_N \tilde{\mathbf{S}}_k + \mathbf{N}_k, \quad (14)$$

where $\tilde{\mathbf{S}}_k = (\boldsymbol{\beta}_k^T \otimes \mathbf{\Lambda}) (\mathbf{A}_{M,k}^T \otimes \mathbf{A}_M^H) \mathbf{\Theta}_k \in \mathbb{C}^{L \times \tau_k}$ represents the equivalent transmitted signal matrix for user k . Consequently, subspace-based methods exhibit superior performance under this standard signal models. However, the UPA configuration in this paper necessitates their extension to the 2-D version. Such an extension substantially increases the algorithm's computational complexity. For instance, the 2-D Root-MUSIC algorithm requires evaluating determinants whose dimensions scale with the array size [22]. To address this, a dimension-reduced subspace-based method is proposed to reduce computational burden while remaining compatible with existing 1-D estimation algorithms.

Exploiting the structure of the UPA and leveraging the properties of the Kronecker product, we derive a dimension-reduced received signal matrix $\mathbf{Y}_{k,h} \in \mathbb{C}^{N_h \times \tau_k N_v}$ for the horizontal dimension to estimate the horizontal angles $\{\nu_l\}_{l=1}^L$, which is achieved via row extraction as follows

$$\begin{aligned} [\mathbf{Y}_{k,h}]_{:,1+(v-1)\tau_k:v\tau_k} &= [\mathbf{Y}_k]_{1+(v-1)N_h:vN_h,:} \\ &= \sqrt{p} [\mathbf{A}_N]_{1+(v-1)N_h:vN_h,:} \tilde{\mathbf{S}}_k + [\mathbf{N}_k]_{1+(v-1)N_h:vN_h,:} \\ &\triangleq \sqrt{p} [\zeta_h(v,1) \mathbf{a}_{N_h}(\nu_1), \dots, \zeta_h(v,L) \mathbf{a}_{N_h}(\nu_L)] \tilde{\mathbf{S}}_k + \mathbf{N}_{k,h}^v \\ &= \sqrt{p} \mathbf{A}_{N_h} \text{Diag}(\zeta_h(v,1), \dots, \zeta_h(v,L)) \tilde{\mathbf{S}}_k + \mathbf{N}_{k,h}^v, \end{aligned} \quad (15)$$

where $\zeta_h(v,l) = e^{-i2\pi(v-1)\psi_l}$ with $v = 1, 2, \dots, N_v$ denoting the row index, $\mathbf{A}_{N_h} \triangleq [\mathbf{a}_{N_h}(\nu_1), \dots, \mathbf{a}_{N_h}(\nu_L)]$ denotes the horizontal AoA steering matrix, and $\mathbf{N}_{k,h} = [\mathbf{N}_{k,h}^1, \dots, \mathbf{N}_{k,h}^{N_v}]$ represents the corresponding noise. Defining $\tilde{\mathbf{S}}_{k,h}^v = \text{Diag}(\zeta_h(v,1), \dots, \zeta_h(v,L)) \tilde{\mathbf{S}}_k$ yields $\tilde{\mathbf{S}}_{k,h} = [\tilde{\mathbf{S}}_{k,h}^1, \dots, \tilde{\mathbf{S}}_{k,h}^{N_v}] \in \mathbb{C}^{L \times N_v \tau_k}$. Thus, the 1-D standard signal model is given by $\mathbf{Y}_{k,h} = \sqrt{p} \mathbf{A}_{N_h} \tilde{\mathbf{S}}_{k,h} + \mathbf{N}_{k,h}$. To enhance subspace-based methods by eliminating error propagation and increasing the equivalent received signal snapshots, we concatenate the dimension-reduced received matrices of all users

into $\mathbf{Y}_h = [\mathbf{Y}_{1,h}, \dots, \mathbf{Y}_{K,h}] \in \mathbb{C}^{N_h \times N_v \sum_{k=1}^K \tau_k}$, which is then used to compute the received signal covariance matrix

$$\mathbf{R}_h = \mathbb{E}[\mathbf{Y}_h \mathbf{Y}_h^H] = p \mathbf{A}_{N_h} \mathbb{E}[\tilde{\mathbf{S}}_h \tilde{\mathbf{S}}_h^H] \mathbf{A}_{N_h}^H + \mathbb{E}[\mathbf{N}_h \mathbf{N}_h^H], \quad (16)$$

where $\tilde{\mathbf{S}}_h = [\tilde{\mathbf{S}}_{1,h}, \dots, \tilde{\mathbf{S}}_{K,h}] \in \mathbb{C}^{L \times N_v \sum_{k=1}^K \tau_k}$ and $\mathbf{N}_h = [\mathbf{N}_{1,h}, \dots, \mathbf{N}_{K,h}] \in \mathbb{C}^{N_h \times N_v \sum_{k=1}^K \tau_k}$ represents the equivalent transmitted signal and noise across all users, respectively.

Since the actual received signal matrix has finite length, eigenvalue decomposition is typically performed using the maximum likelihood estimation (MLE) of the signal covariance matrix, given by

$$\mathbf{R}_h = \frac{1}{N_v \sum_{k=1}^K \tau_k} \mathbf{Y}_h \mathbf{Y}_h^H. \quad (17)$$

By performing eigenvalue decomposition of $\mathbf{R}_h$, we identify the signal subspace spanned by the eigenvectors associated with the L largest eigenvalues. The value L also corresponds to the number of common AoAs, which equals the number of propagation paths between the RIS and the BS¹. The same procedure is used to estimate the steering vectors for the vertical (z -axis) UPA dimension. It enables estimation of the common AoA steering matrix $\hat{\mathbf{A}}_N = \hat{\mathbf{A}}_{N_v} \diamond \hat{\mathbf{A}}_{N_h}$ shared by all users. Two subspace-based estimation algorithms, Root-MUSIC and TLS-ESPRIT, are considered. Algorithm 1 outlines the common AoA estimation procedure while omitting the specific subspace-based algorithm procedures [23], [24].

2) *Estimation of Common AoD at the RIS:* Let $\Delta \mathbf{A}_N \triangleq \hat{\mathbf{A}}_N - \mathbf{A}_N$ denotes the estimation error associated with the common AoA estimates. Exploiting the asymptotic orthogonality of the steering matrix for large-scale arrays [17], we use the identity $\hat{\mathbf{A}}_N^H \mathbf{A}_N = N \mathbf{I}_L + (\Delta \mathbf{A}_N)^H \mathbf{A}_N$. Thus, the linear transformation $\frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H$ can be applied to the received signal $\mathbf{Y}_k$ to eliminate the effect of the common AoA components:

$$\frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H \mathbf{Y}_k = (\boldsymbol{\beta}_k^T \otimes \mathbf{\Lambda}) (\mathbf{A}_{M,k}^T \otimes \mathbf{A}_M^H) \mathbf{\Theta}_k + \check{\mathbf{N}}_k^H, \quad (19)$$

where $\check{\mathbf{N}}_k^H \triangleq \frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H \mathbf{N}_k + \frac{1}{N} (\Delta \mathbf{A}_N)^H \mathbf{A}_N (\boldsymbol{\beta}_k^T \otimes \mathbf{\Lambda}) (\mathbf{A}_{M,k}^T \otimes \mathbf{A}_M^H) \mathbf{\Theta}_k$ denotes the corresponding noise and the second term characterizes the residual error propagation introduced by common AoA estimates. Now, we define the transpose of $\frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H \mathbf{Y}_k$ as the equivalent measurement matrix $\check{\mathbf{Y}}_k \in \mathbb{C}^{\tau_k \times L}$ for user k , given by

$$\begin{aligned} \check{\mathbf{Y}}_k &\triangleq \left(\frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H \mathbf{Y}_k \right)^H \\ &= \mathbf{\Theta}_k^H (\mathbf{A}_{M,k}^* \otimes \mathbf{A}_M) (\boldsymbol{\beta}_k^* \otimes \mathbf{\Lambda}^H) + \check{\mathbf{N}}_k \\ &= \mathbf{\Theta}_k^H \mathbf{H}_{\text{RIS}}^k + \check{\mathbf{N}}_k, \end{aligned} \quad (20)$$

where $\mathbf{H}_{\text{RIS}}^k \triangleq (\mathbf{A}_{M,k}^* \otimes \mathbf{A}_M) (\boldsymbol{\beta}_k^* \otimes \mathbf{\Lambda}^H)$. By exploiting the structure of $\mathbf{H}_{\text{RIS}}^1$ associated with user 1, we extract the r -th

¹While the eigenvalues corresponding to the signal subspace are typically much larger than those of the noise subspace, information-theoretic criteria such as the minimum description length (MDL) and Akaike information criterion (AIC) can be employed to more accurately determine the number of signal sources.

Algorithm 1 Dimension-reduced Subspace-Based Common AoA Estimation

Input: $\mathbf{Y}_k$, $1 \leq k \leq K$.

- 1: Calculate the horizontal dimension-reduced signal matrix $\mathbf{Y}_{k,h}$ from $\mathbf{Y}_k$ ($1 \leq k \leq K$) according to Eq. (15).
- 2: Concatenate the dimension-reduced covariance matrices across all users to obtain $\mathbf{Y}_h = [\mathbf{Y}_{1,h}, \dots, \mathbf{Y}_{K,h}]$.
- 3: Estimate $\mathbf{R}_h$ of $\mathbf{Y}_h$ according to Eq. (17).
- 4: Perform the eigenvalue decomposition of $\mathbf{R}_h$

$$\mathbf{R}_h = \mathbf{U}_{S,h} \boldsymbol{\Sigma}_{S,h} \mathbf{U}_{S,h}^H + \mathbf{U}_{N,h} \boldsymbol{\Sigma}_{N,h} \mathbf{U}_{N,h}^H, \quad (18)$$

where $\mathbf{U}_{S,h} \boldsymbol{\Sigma}_{S,h} \mathbf{U}_{S,h}^H$ represents the signal portion and $\mathbf{U}_{N,h} \boldsymbol{\Sigma}_{N,h} \mathbf{U}_{N,h}^H$ represents the noise portion.

- 5: Sort the eigenvalues and determine the number of propagation paths in the RIS-BS channel, denoted as $\hat{L}$.
- 6: Apply the subspace-based algorithm to estimate the common AoAs in the horizontal dimension $\{\hat{\nu}_l\}_{l=1}^{\hat{L}}$ and construct the corresponding steering matrix $\hat{\mathbf{A}}_{N_h} = [\mathbf{a}_{N_h}(\hat{\nu}_1), \dots, \mathbf{a}_{N_h}(\hat{\nu}_{\hat{L}})]$.
- 7: Repeat the similar procedure for the vertical dimension to estimate $\{\hat{\psi}_l\}_{l=1}^{\hat{L}}$ and construct the steering matrix $\hat{\mathbf{A}}_{N_v} = [\mathbf{a}_{N_v}(\hat{\psi}_1), \dots, \mathbf{a}_{N_v}(\hat{\psi}_{\hat{L}})]$.

Output: $\hat{\mathbf{A}}_N = \hat{\mathbf{A}}_{N_v} \diamond \hat{\mathbf{A}}_{N_h}$.

column of the equivalent measurement matrix $\check{\mathbf{Y}}_1$, denoted as $\check{\mathbf{y}}_{1,r}$, which is given by

$$\begin{aligned} \check{\mathbf{y}}_{1,r} &= \boldsymbol{\Theta}_1^H (\mathbf{A}_{M,1}^* \otimes \mathbf{A}_M) [(\boldsymbol{\beta}_1^* \otimes \mathbf{A}^H)]_{:,r} + \check{\mathbf{n}}_{1,r} \\ &= \boldsymbol{\Theta}_1^H (\mathbf{A}_{M,1}^* \otimes \mathbf{a}_M(\omega_r, \mu_r)) (\alpha_r^* \boldsymbol{\beta}_1^*) + \check{\mathbf{n}}_{1,r}, \end{aligned} \quad (21)$$

where $\mathbf{A}_{M,1}^* \otimes \mathbf{a}_M(\omega_r, \mu_r) = [\mathbf{a}_M^*(\varphi_{1,1}, \theta_{1,1}) \otimes \mathbf{a}_M(\omega_r, \mu_r), \dots, \mathbf{a}_M^*(\varphi_{1,J_1}, \theta_{1,J_1}) \otimes \mathbf{a}_M(\omega_r, \mu_r)] \in \mathbb{C}^{M^2 \times J_1}$ and $\check{\mathbf{n}}_r$ denotes the r -th column of $\check{\mathbf{N}}_1$. This result is derived by observing that $[(\boldsymbol{\beta}_1^* \otimes \mathbf{A}^H)]_{:,r}$ contains non-zero elements at every r -th position within blocks of L rows, forming the sequence $\{\alpha_r^* \boldsymbol{\beta}_{1,j}^*\}_{j=1}^{J_1}$. This structure allows the extraction and concatenation of the r -th column from each block of L columns of $\mathbf{A}_{M,1}^* \otimes \mathbf{A}_M$, resulting in the term $\mathbf{h}_{\text{RIS},r}^1 \triangleq (\mathbf{A}_{M,1}^* \otimes \mathbf{a}_M(\omega_r, \mu_r)) (\alpha_r^* \boldsymbol{\beta}_1^*)$, which corresponds to the r -th column of $\mathbf{H}_{\text{RIS}}^1$.

To estimate $\mathbf{h}_{\text{RIS},r}^1$, Eq. (21) can be approximated by using virtual angular domain (VAD) representation as

$$\check{\mathbf{y}}_{1,r} = \boldsymbol{\Theta}_1^H (\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M) \mathbf{b}_{1,r} + \check{\mathbf{n}}_{1,r}, \quad (22)$$

where $\tilde{\mathbf{A}}_{M,1} \in \mathbb{C}^{M \times D_1}$ and $\tilde{\mathbf{A}}_M \in \mathbb{C}^{M \times D}$ denote over-complete dictionary matrices. The columns of $\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M$ contain possible values of $\{\mathbf{a}_M(\varphi_{1,j}, \theta_{1,j})^* \otimes \mathbf{a}_M(\omega_r, \mu_r)\}_{j=1}^{J_1}$, evaluated over the discretized angle domains. As an example, $\tilde{\mathbf{A}}_M$ is constructed as

$$\tilde{\mathbf{A}}_M = \tilde{\mathbf{A}}_{M_v} \otimes \tilde{\mathbf{A}}_{M_h}, \quad (23a)$$

$$[\tilde{\mathbf{A}}_{M_v}]_{:,g_v} = \mathbf{a}_{M_v} \left(\left(-1 + \frac{2}{D_v} g_v \right) \frac{d_{\text{RIS}}}{\lambda_c} \right), \quad (23b)$$

$$[\tilde{\mathbf{A}}_{M_h}]_{:,g_h} = \mathbf{a}_{M_h} \left(\left(-1 + \frac{2}{D_h} g_h \right) \frac{d_{\text{RIS}}}{\lambda_c} \right), \quad (23c)$$

where $g_v = 0, 1, \dots, D_v - 1$ (with $D_v \geq M_v$) and $g_h = 0, 1, \dots, D_h - 1$ (with $D_h \geq M_h$), and $D = D_v \times D_h$. The atoms in $\tilde{\mathbf{A}}_{M_v}$ and $\tilde{\mathbf{A}}_{M_h}$ span the intervals $[-\frac{d_{\text{RIS}}}{\lambda_c}, (1 - \frac{2}{D_v}) \frac{d_{\text{RIS}}}{\lambda_c}]$ and $[-\frac{d_{\text{RIS}}}{\lambda_c}, (1 - \frac{2}{D_h}) \frac{d_{\text{RIS}}}{\lambda_c}]$, with resolutions $2/D_v$ and $2/D_h$, respectively. Furthermore, $\mathbf{b}_{1,r} \in \mathbb{C}^{DD_1 \times 1}$ is a sparse vector with J_1 non-zero entries corresponding to the cascaded channel path gains $\{\alpha_r^* \boldsymbol{\beta}_{1,j}^*\}_{j=1}^{J_1}$ ². To ensure optimal CS performance, the MC-aware RIS phase shift training matrix $\boldsymbol{\Theta}_1$ must be designed such that the columns of the equivalent dictionary $\boldsymbol{\Theta}_1^H (\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M)$ are orthogonal. Designing the RIS phase shift training matrix simultaneously addresses the MC effect. The detailed design procedure is discussed in Section IV.

Using CS, the estimated value $\hat{\mathbf{h}}_{\text{RIS},r}^1$ is obtained according to Eq. (21). The remaining columns of $\mathbf{H}_{\text{RIS}}^1$, denoted as $\{\hat{\mathbf{h}}_{\text{RIS},l}^1\}_{l \neq r}^L$, can be estimated by recovering the AoAs at the RIS for user 1, i.e., the matrix $\mathbf{A}_{M,1}$, thereby reducing computational complexity. Specifically, the term $\mathbf{a}_M^*(\varphi_{1,j}, \theta_{1,j}) \otimes \mathbf{a}_M(\omega_r, \mu_r) = (\mathbf{a}_{M_v}^*(\varphi_{1,j}) \otimes \mathbf{a}_{M_h}^*(\theta_{1,j})) \otimes (\mathbf{a}_{M_v}(\omega_r) \otimes \mathbf{a}_{M_h}(\mu_r))$ represents a four-dimensional (4-D) angular product. A significant reduction in dimensionality can be achieved by applying a similar procedure as described in Eq. (15), thereby enabling accurate angle estimation with reduced computational burden. The dimension-reduced received signal matrix $\mathbf{Y}_{\text{AoA}_h}^1 \in \mathbb{C}^{M_h \times MM_v}$ is obtained via element extraction, as defined by

$$[\mathbf{Y}_{\text{AoA}_h}^1]_{m_h, (m_v-1)M+m} = [\mathbf{h}_{\text{RIS},r}^1]_{i_h(m_h, m_v, m)}, \quad (24)$$

where $i_h(m_h, m_v, m) = (m_h - 1)M + (m_v - 1)MM_h + m$ is an indexing function, and the indices m_h , m_v , and m take values from the sets $\{1, \dots, M_h\}$, $\{1, \dots, M_v\}$, and $\{1, \dots, M\}$, respectively. By omitting the row index indicator m_h , the element-wise concatenated result can be expressed as

$$\begin{aligned} [\mathbf{Y}_{\text{AoA}_h}^1]_{:, (m_v-1)M+m} &= [\kappa_h(1, m, m_v) \mathbf{a}_{M_h}^*(\theta_{1,1}), \dots, \kappa_h(J_1, m, m_v) \mathbf{a}_{M_h}^*(\theta_{1,J_1})] (\alpha_r^* \boldsymbol{\beta}_1^*) \\ &= \mathbf{A}_{M_h,1}^* \text{Diag}(\kappa_h(1, m, m_v), \dots, \kappa_h(J_1, m, m_v)) (\alpha_r^* \boldsymbol{\beta}_1^*), \end{aligned} \quad (25)$$

where $\frac{\ln(\kappa_h(j, m, m_v))}{-i2\pi} = (m - \lfloor \frac{m-1}{M_h} \rfloor M_h - 1) \mu_l + \lfloor \frac{m-1}{M_h} \rfloor \omega_l - (m_v - 1) \varphi_{1,j}$ and $\mathbf{A}_{M_h,1} \triangleq [\mathbf{a}_{M_h}(\theta_{1,1}), \dots, \mathbf{a}_{M_h}(\theta_{1,J_1})]$ denotes the horizontal steering matrix for the y -axis dimension of the UPA corresponding to the AoA at the RIS for user 1. After element-wise concatenation, a formulation suitable for sparse signal recovery is obtained. Analogously, the dimension-reduced received signal in the vertical dimension can also be derived via element extraction as follows

$$[\mathbf{Y}_{\text{AoA}_v}^1]_{m_v, m^*} = [\mathbf{h}_{\text{RIS},r}^1]_{i_v(m_v, m^*)}, \quad (26)$$

where $i_v(m_v, m^*) = (m_v - 1)MM_h + m^*$ denotes the indexing function, and m^* takes values from the set

²The number of propagation paths between user 1 and the RIS, denoted by J_1 , determines the sparsity level in the recovery problems associated with Eqs. (22) and (28). In Stage II of the proposed protocol, OMP is used as the recovery algorithm. In this case, the stopping criterion is based on the residual power. Specifically, the algorithm terminates when the residual energy falls below a predefined threshold. The number of iterations is therefore treated as an estimate of J_1 .

$\{1, \dots, MM_h\}$. Then, by omitting the row index indicator m_v , the element-wise concatenated result can be obtained as:

$$\begin{aligned} [\mathbf{Y}_{\text{AoA}_v}^1]_{:,m^*} &= \\ [\kappa_v(1, m^*) \mathbf{a}_{M_v}^*(\varphi_{1,1}), \dots, \kappa_v(J_1, m^*) \mathbf{a}_{M_v}^*(\varphi_{1,J_1})] (\alpha_r^* \beta_1^*) \\ &= \mathbf{A}_{M_v,1}^* \text{Diag}(\kappa_v(1, m^*), \dots, \kappa_v(J_1, m^*)) (\alpha_r^* \beta_1^*), \quad (27) \end{aligned}$$

where $\frac{\ln(\kappa_v(j, m^*))}{-i2\pi} = \left(m^* - \lfloor \frac{m^* - 1}{M_h} \rfloor M_h - 1 \right) \mu_l + \lfloor \frac{m^* - 1}{M_h} \rfloor \omega_l - \lfloor \frac{m^* - 1}{M} \rfloor \theta_{1,j}$ and $\mathbf{A}_{M_v,1} \triangleq [\mathbf{a}_{M_v}(\varphi_{1,1}), \dots, \mathbf{a}_{M_v}(\varphi_{1,J_1})]$ denotes the corresponding vertical steering matrix. Remarkably, Eqs. (25) and (27) are compatible with both subspace-based and CS techniques. We demonstrate using the CS method as an example in Algorithm 2. After executing the respective angle estimation procedures, the complete matrix $\mathbf{A}_{M,1}$ can be obtained as $\hat{\mathbf{A}}_{M,1} = \hat{\mathbf{A}}_{M_v,1} \diamond \hat{\mathbf{A}}_{M_h,1}$. The remaining columns of $\mathbf{H}_{\text{RIS}}^1$, i.e., $\{\mathbf{h}_{\text{RIS},l}^1\}_{l \neq r}^{\hat{L}}$, can then be estimated via a J_1 -sparse signal recovery problem, thereby reducing computational complexity:

$$\check{\mathbf{y}}_{1,l} = \Theta_1^H \left(\hat{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M \right) \mathbf{b}_{1,l} + \check{\mathbf{n}}_{1,l}, \quad (28)$$

where $\tilde{\mathbf{A}}_M$ is defined in Eq. (23a). Therefore, the remaining columns $\mathbf{h}_{\text{RIS},l}^1$, for $1 \leq l \leq \hat{L}, l \neq r$, can be obtained. Finally, the estimated cascaded channel for user 1 is expressed as

$$\hat{\mathbf{G}}_1 = \hat{\mathbf{A}}_N \left[\hat{\mathbf{h}}_{\text{RIS},1}^1, \dots, \hat{\mathbf{h}}_{\text{RIS},\hat{L}}^1 \right]^H. \quad (29)$$

The overall estimation of $\mathbf{G}_1$ is summarized in Algorithm 2.

Notably, the OMP algorithm facilitates the extraction of the set $\{\hat{\mathbf{A}}_{M,1}^* \otimes \hat{\mathbf{a}}_M(\omega_l, \mu_l)\}_{l=1}^{\hat{L}}$ from the equivalent dictionaries $\mathbf{D}_1$ and Ξ_1 . By leveraging the properties of the Kronecker product and the structure in Eq. (21), these components are concatenated to construct the matrix $\hat{\mathbf{A}}_{M,1}^* \otimes \hat{\mathbf{A}}_M$. Considering that $[\hat{\mathbf{A}}_{M,1}^*]_{1,1} = 1$, the common AoD steering matrix at the RIS can be directly obtained as $\hat{\mathbf{A}}_M = [\hat{\mathbf{A}}_{M,1}^* \otimes \hat{\mathbf{A}}_M]_{1:M,1:\hat{L}}$.

C. Stage III: Channel Estimation for Other Users

In this subsection, we exploit the fact that all users share the common RIS-BS channel to reduce the pilot overhead in channel estimation. Building upon the common AoA steering matrix $\hat{\mathbf{A}}_N$ at the BS, as derived in Section III-B1, the equivalent measurement matrix for the k -th user can be constructed according to Eq. (20). This results in a measurement model analogous to that in Eq. (21), expressed as $\check{\mathbf{y}}_{k,r} = \Theta_k^H \left(\hat{\mathbf{A}}_{M,k}^* \otimes \mathbf{a}_M(\omega_r, \mu_r) \right) (\alpha_r^* \beta_k^*) + \check{\mathbf{n}}_{k,r}$. Leveraging the common AoD steering matrix $\hat{\mathbf{A}}_M$ obtained in Section III-B2, the J_k -sparse signal recovery problem for other users can be reformulated to reduce computational complexity as follows

$$\check{\mathbf{y}}_{k,r} = \Theta_k^H \left(\hat{\mathbf{A}}_{M,k}^* \otimes \hat{\mathbf{A}}_M \right) \mathbf{b}_{k,r} + \check{\mathbf{n}}_{k,r}, \quad (31)$$

where the construction of $\hat{\mathbf{A}}_{M,k}$ follows the same procedure as in Eq. (23a). Similarly, after obtaining the estimate $\hat{\mathbf{h}}_{\text{RIS},r}^k$ via Eq. (31), we exploit its 4-D angular product structure, $\mathbf{a}_{M_v}^*(\varphi_{k,j}, \theta_{k,j}) \otimes \mathbf{a}_M(\omega_r, \mu_r) = (\mathbf{a}_{M_v}^*(\varphi_{k,j}) \otimes \mathbf{a}_{M_h}^*(\theta_{k,j})) \otimes$

Algorithm 2 Estimation of MC-aware cascaded channel for Typical User

Input: $\mathbf{Y}_1$

- 1: Obtain estimated number of paths in RIS-BS channel $\hat{L}$ and common AoA steering matrix $\hat{\mathbf{A}}_N$ from Algorithm 1.
- 2: Compute the equivalent measurement matrix $[\check{\mathbf{y}}_{1,1}, \dots, \check{\mathbf{y}}_{1,\hat{L}}] = \left(\frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H \mathbf{Y}_1 \right)^H$.
- 3: Determine the reference column index r as:

$$r = \arg \max_{1 \leq i \leq \hat{L}} \|\check{\mathbf{y}}_{1,i}\|_2^2. \quad (30)$$

- 4: Construct the dictionary $\mathbf{D}_1 = \Theta_1^H \left(\hat{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M \right)$.
- 5: Estimate $\hat{\mathbf{h}}_{\text{RIS},r}^1$ from $\check{\mathbf{y}}_{1,r}$ using $\mathbf{D}_1$ and the OMP algorithm with Eq. (22).
- 6: Estimate $\hat{\mathbf{A}}_{M_h,1}$ from $\hat{\mathbf{h}}_{\text{RIS},r}^1$ using Eq. (24) via element extraction and the OMP algorithm with dictionary $\Xi_{1,h} = \tilde{\mathbf{A}}_{M_h,1}^*$, constructed similarly to Eq. (23c).
- 7: Estimate $\hat{\mathbf{A}}_{M_v,1}$ from $\hat{\mathbf{h}}_{\text{RIS},r}^1$ using Eq. (26) via element extraction and the OMP algorithm with dictionary $\Xi_{1,v} = \tilde{\mathbf{A}}_{M_v,1}^*$, constructed similarly to Eq. (23b).
- 8: Recover the AoA steering matrix at the RIS for user 1 as $\hat{\mathbf{A}}_{M,1} = \hat{\mathbf{A}}_{M_v,1} \diamond \hat{\mathbf{A}}_{M_h,1}$.
- 9: Construct the dictionary $\Xi_1 = \Theta_1^H \left(\hat{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M \right)$.
- 10: **for** $1 \leq l \leq \hat{L}, l \neq r$ **do**
- 11: Estimate $\hat{\mathbf{h}}_{\text{RIS},l}^1$ from $\check{\mathbf{y}}_{1,l}$ using Ξ_1 and the OMP algorithm with Eq. (28).
- 12: **end for**

Output: $\hat{\mathbf{G}}_1 = \hat{\mathbf{A}}_N \left[\hat{\mathbf{h}}_{\text{RIS},1}^1, \dots, \hat{\mathbf{h}}_{\text{RIS},\hat{L}}^1 \right]^H$.

$(\mathbf{a}_{M_v}(\omega_r) \otimes \mathbf{a}_{M_h}(\mu_r))$ to recover the horizontal and vertical AoA steering matrices $\mathbf{A}_{M_h,k}$ and $\mathbf{A}_{M_v,k}$ via Eqs. (24) and (26), respectively. We then derive the full AoA steering matrix at the RIS for user k as $\hat{\mathbf{A}}_{M,k} = \hat{\mathbf{A}}_{M_v,k} \diamond \hat{\mathbf{A}}_{M_h,k}$, which reduces the computational complexity associated with estimating the remaining columns of $\mathbf{H}_{\text{RIS}}^k$, i.e., $\{\hat{\mathbf{h}}_{\text{RIS},l}^k\}_{l \neq r}^{\hat{L}}$. The specific pilot overhead and computational complexity reduction achieved through the above operations will be discussed in Section III-E.

D. MC-unaware Channel Estimation

This subsection presents a channel estimation scheme tailored to the conventional RIS-aided cascaded channel model as described in Eq. (8). In Section V, the scheme is applied to received signals under the MC effect and compared with the proposed and existing MC-aware algorithms. This comparative analysis rigorously characterizes the impact of MC on estimation performance while highlighting the advantages of the proposed MC-aware estimation protocol.

Similar to Eq. (11), the received signal that ignores the MC effect can be expressed as

$$\begin{aligned} \mathbf{y}_{\text{cv},k}(t) &= \sqrt{p} \mathbf{h}_{\text{cv},k} + \mathbf{n}_{\text{cv},k}(t) \\ &= \sqrt{p} \mathbf{A}_N \mathbf{A}_M^H \mathbf{A}_M \mathbf{A}_{M,k} \beta_k + \mathbf{n}_{\text{cv},k}(t) \\ &= \sqrt{p} \mathbf{A}_N \mathbf{A}_M^H \text{Diag}(\mathbf{A}_{M,k} \beta_k) \boldsymbol{\gamma} + \mathbf{n}_{\text{cv},k}(t). \quad (32) \end{aligned}$$

Similar to Eq. (13), stacking the τ_k received signal vectors across time slots yields the matrix $\mathbf{Y}_{cv,k} \in \mathbb{C}^{N \times \tau_k}$ as follows

$$\mathbf{Y}_{cv,k} = [\mathbf{y}_{cv,k}(1), \dots, \mathbf{y}_{cv,k}(\tau_k)] \triangleq \sqrt{p} \mathbf{G}_{cv,k} \mathbf{\Theta}_{cv,k} + \mathbf{N}_{cv,k}, \quad (33)$$

where $\mathbf{G}_{cv,k} = \mathbf{A}_N \mathbf{\Lambda} \mathbf{A}_M^H \text{Diag}(\mathbf{A}_{M,k} \beta_k)$ defines the MC-unaware cascaded user-RIS-BS channel for user k . In addition, $\mathbf{\Theta}_{cv,k} = [\gamma_1, \dots, \gamma_{\tau_k}] \in \mathbb{C}^{M \times \tau_k}$ denotes the RIS phase shift training matrix for user k , and $\mathbf{N}_{cv,k} = [\mathbf{n}_{cv,k}(1), \dots, \mathbf{n}_{cv,k}(\tau_k)] \in \mathbb{C}^{N \times \tau_k}$ represents the corresponding noise matrix.

By examining the structure of $\mathbf{G}_{cv,k}$, it can be observed that the estimation procedure for the common AoA at the BS is identical to that described in Section III-B1. After eliminating the common AoA influence, the equivalent MC-unaware measurement signal can be obtained as

$$\begin{aligned} \check{\mathbf{Y}}_{cv,k} &\triangleq \left(\frac{1}{N\sqrt{p}} \hat{\mathbf{A}}_N^H \mathbf{Y}_{cv,k} \right)^H \\ &= \mathbf{\Theta}_{cv,k}^H \text{Diag}(\mathbf{h}_k^*) \mathbf{A}_M \mathbf{\Lambda}^* + \check{\mathbf{N}}_{cv,k} \\ &= \mathbf{\Theta}_{cv,k}^H \mathbf{H}_{\text{RIS}}^{cv,k} + \check{\mathbf{N}}_{cv,k}, \end{aligned} \quad (34)$$

where $\mathbf{H}_{\text{RIS}}^{cv,k} \triangleq \text{Diag}(\mathbf{h}_k^*) \mathbf{A}_M \mathbf{\Lambda}^*$ and $\check{\mathbf{N}}_{cv,k}$ is the corresponding noise. By exploiting the structure of $\mathbf{H}_{\text{RIS}}^{cv,k}$, we obtain

$$\begin{aligned} \mathbf{H}_{\text{RIS}}^{cv,k} &= \mathbf{h}_k^* \bullet (\mathbf{A}_M \mathbf{\Lambda}^*) = (\mathbf{A}_{M,k} \beta_k)^* \bullet (\mathbf{A}_M \mathbf{\Lambda}^*) \\ &= (\mathbf{A}_{M,k}^* \bullet \mathbf{A}_M) (\beta_k^* \otimes \mathbf{\Lambda}^*), \end{aligned} \quad (35)$$

where $\mathbf{A}_{M,k}^* \bullet \mathbf{A}_M = [\mathbf{a}_M(\omega_1 - \varphi_{k,1}, \mu_1 - \theta_{k,1}), \dots, \mathbf{a}_M(\omega_L - \varphi_{k,J_k}, \mu_L - \theta_{k,J_k})] \in \mathbb{C}^{M \times L J_k}$. The final equality follows from the identity $(\mathbf{A} \bullet \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \bullet (\mathbf{BD})$ [25]. Similar to Eq. (21), channel estimation can be performed via column extraction as follows

$$\begin{aligned} \check{\mathbf{y}}_{k,l} &= \mathbf{\Theta}_{cv,k}^H \text{Diag}(\mathbf{h}_k^*) [\mathbf{A}_M \mathbf{\Lambda}^*]_{:,l} + \check{\mathbf{n}}_{k,l} \\ &= \mathbf{\Theta}_{cv,k}^H \mathbf{h}_k^* \bullet (\alpha_l^* \mathbf{a}_M(\omega_l, \mu_l)) + \check{\mathbf{n}}_{k,l} \\ &= \mathbf{\Theta}_{cv,k}^H (\mathbf{A}_{M,k}^* \bullet \mathbf{a}_M(\omega_l, \mu_l)) \alpha_l^* \beta_k + \check{\mathbf{n}}_{k,l}, \end{aligned} \quad (36)$$

where $\mathbf{A}_{M,k}^* \bullet \mathbf{a}_M(\omega_l, \mu_l) = [\mathbf{a}_M(\omega_l - \varphi_{k,1}, \mu_l - \theta_{k,1}), \dots, \mathbf{a}_M(\omega_l - \varphi_{k,J_k}, \mu_l - \theta_{k,J_k})] \in \mathbb{C}^{M \times J_k}$, and $\check{\mathbf{n}}_{k,l}$ denotes the l -th column of $\check{\mathbf{N}}_{cv,k}$. Note that $\text{Diag}(\mathbf{h}_k^*) [\mathbf{A}_M \mathbf{\Lambda}^*]_{:,l}$ corresponds to the l -th column of $\mathbf{H}_{\text{RIS}}^{cv,k}$, denoted as $\mathbf{h}_{\text{RIS},l}^{cv,k}$. Given that all angles lie within the interval $[-\frac{d_{\text{RIS}}}{\lambda_c}, \frac{d_{\text{RIS}}}{\lambda_c}]$, Eq. (36) can be reformulated as a J_k -sparse signal recovery problem, which can be efficiently solved using CS techniques. By jointly processing all $\hat{L}$ columns for each of the K users, the complete set of $\{\mathbf{H}_{\text{RIS}}^{cv,k}\}_{k=1}^K$ can be achieved. Consequently, the MC-unaware equivalent cascaded channel estimates for all users are obtained as $\hat{\mathbf{G}}_{cv,k} = \hat{\mathbf{A}}_N [\hat{\mathbf{h}}_{\text{RIS},1}^{cv,k}, \dots, \hat{\mathbf{h}}_{\text{RIS},\hat{L}}^{cv,k}]^H$.

E. Pilot Overhead and Computational Complexity Analysis

In this subsection, we analyze the pilot overhead and computational complexity of the proposed MC-aware three-stage channel estimation protocol. For simplicity, we assume that $J_1 = \dots = J_K = J$.

1) *Pilot Overhead Analysis*: The pilot overhead τ_1 for the typical user and τ_k for user k ($2 \leq k \leq K$) are associated with different stages of the proposed protocol. Specifically, τ_1 governs the measurement dimension for the J_1 -sparse signal recovery problem in Eq. (22) in Stage II, while τ_k determines the required pilot overhead for the J_k -sparse signal recovery problem in Eq. (31) in Stage III. Both τ_1 and τ_k jointly influence the common AoA estimation in Stage I.

According to compressive sensing theory [26], recovering a l -sparse complex signal of dimension n requires the number of measurements m on the order of $\mathcal{O}(l \log n)$. For the typical user, the sensing matrix in Eq. (22) is given by $\mathbf{\Theta}_1^H (\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M) \in \mathbb{C}^{\tau_1 \times DD_1}$, indicating that the required pilot overhead satisfies $\tau_1 \geq \mathcal{O}(J \log(DD_1))$. In the simulation setup of Section V, the dictionary resolution is set to twice the dimension of the corresponding basis matrix, e.g., $D_h = 2M_h$, resulting in $\tau_1 \geq \mathcal{O}(J \log(16M^2))$. For other users, the sensing matrix in Eq. (31) is $\mathbf{\Theta}_k^H (\tilde{\mathbf{A}}_{M,k}^* \otimes \tilde{\mathbf{A}}_M) \in \mathbb{C}^{\tau_k \times LD_k}$, and the required pilot overhead satisfies $\tau_k \geq \mathcal{O}(J \log(LD_k))$. Given the same dictionary resolution assumptions, this becomes $\tau_k \geq \mathcal{O}(J \log(4ML))$. Combining both results, the total pilot overhead required for all K users is $\mathcal{O}(J \log(16M^2) + (K-1)J \log(4ML))$.

2) *Computational Complexity Analysis*: For estimating the common AoA at the BS in Stage I, as described in Algorithm 1, the computational complexity primarily arises from the MLE of the covariance matrix and the eigenvalue decomposition, with respective complexities of $\mathcal{O}(N_v N_h^2 \sum_{k=1}^K \tau_k) = \mathcal{O}(N N_h \sum_{k=1}^K \tau_k)$ and $\mathcal{O}(N_h^3)$ in the horizontal dimension. The overall computational complexity in Stage I is $\mathcal{O}(N(N_h + N_v) \sum_{k=1}^K \tau_k + N_h^3 + N_v^3)$. For estimating the remaining parameters of the typical user in Stage II, as outlined in Algorithm 2, the computational complexity primarily arises from the OMP-based method used for estimating $\mathbf{h}_{\text{RIS},r}^1$ in Step 5, $\mathbf{A}_{M,1}$ in Steps 6 and 7, and $\mathbf{h}_{\text{RIS},l}^1$ ($1 \leq l \leq \hat{L}$, $l \neq r$) in Step 11. Since the dominant computational complexity of OMP is $\mathcal{O}(mnl)$ [17], where m denotes the number of measurements, n the length of the sparse signal with sparsity level l , the computational complexities of these steps are $\mathcal{O}(\tau_1 DD_1 J)$, $\mathcal{O}(M_h D_{1,h} J)$, $\mathcal{O}(M_v D_{1,v} J)$, and $\mathcal{O}((L-1)\tau_1 D J^2)$, respectively. The overall computational complexity in Stage II is $\mathcal{O}(\tau_1 D J (D_1 + (L-1)J) + J(M_h D_{k,h} + M_v D_{k,v}))$. From the above analysis, it can be seen that estimating the vertical and horizontal directions separately transforms the complexity from a multiplicative form to an additive form.

Then, for estimating the remaining parameters of the other users in Stage III, the computational complexity also arises from the OMP-based method used for estimating $\mathbf{h}_{\text{RIS},r}^k$, $\mathbf{A}_{M,k}$, and $\mathbf{h}_{\text{RIS},l}^k$ ($1 \leq l \leq \hat{L}$, $l \neq r$). Similarly, for user k ($2 \leq k \leq K$), the corresponding computational complexities of these steps are $\mathcal{O}(\tau_k D_k L J)$, $\mathcal{O}(M_h D_{k,h} J)$, $\mathcal{O}(M_v D_{k,v} J)$, and $\mathcal{O}((L-1)\tau_k L J^2)$, respectively. The overall computational complexity in Stage III is $\mathcal{O}(\sum_{k=2}^K (\tau_k L J (D_k + (L-1)J) + J(M_h D_{1,h} + M_v D_{1,v})))$. As observed above, estimating $\hat{\mathbf{A}}_M$ significantly reduces the

computational complexity for other users, with the reduction specifically reflected in replacing D with L .

IV. PHASE SHIFT MATRIX DESIGN

The performance of the OMP-based channel estimation is strongly influenced by the orthogonality of the equivalent dictionary. To this end, we optimize the RIS phase shift training matrices to generate approximately orthogonal dictionaries. Specifically, $\Theta_{cv,k} \in \mathbb{C}^{M \times \tau_k}$ ($\forall k \in \mathcal{K}$) are designed to enhance the ability of OMP and mitigate the influence of MC effect to recover the sparse vectors $\mathbf{b}_{k,l}$ in problems defined by Eqs. (22) and (31). Without loss of generality, we focus on the design of $\Theta_{cv,1}$ as a representative case.

Our approach is inspired by the theoretical result in [27], which states that successful recovery of the sparse signal $\mathbf{b}_{1,l}$ via OMP is guaranteed when the following condition holds:

$$\|\mathbf{b}_{1,l}\|_0 \leq \frac{1}{2} \left(1 + \frac{1}{v} \right), \quad (37)$$

where v denotes the mutual coherence of equivalent dictionary $\mathbf{D}_1 = \Theta_{mc,1}^H (\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M) \in \mathbb{C}^{\tau_1 \times DD_1}$, given by

$$v = \max_{i \neq j} \frac{|\mathbf{D}_{1(:,i)}^H \mathbf{D}_{1(:,j)}|}{\|\mathbf{D}_{1(:,i)}\|_2 \|\mathbf{D}_{1(:,j)}\|_2}. \quad (38)$$

The condition in Eq. (38) indicates that the dictionary $\mathbf{D}_1$ should exhibit low mutual coherence, i.e., its columns should be approximately orthogonal. This leads to the following optimization problem:

$$\begin{aligned} \min_{\Theta_{cv,1}} & \|\mathbf{D}_1^H \mathbf{D}_1 - \mathbf{I}_{DD_1}\|_F^2 \\ \text{s.t.} & |[\Theta_{cv,1}]_{m,t}| = 1, 1 \leq m \leq M, 1 \leq t \leq \tau_1. \end{aligned} \quad (39)$$

The unconstrained version of Problem (39) was previously studied in [28], and the approach therein was extended to handle the unit-modulus constraint in [15]. Building on these works, a more concise formulation for the objective is adopted. Observe that

$$\begin{aligned} & \|\mathbf{D}_1^H \mathbf{D}_1 - \mathbf{I}_{DD_1}\|_F^2 \\ &= \text{tr}(\mathbf{D}_1^H \mathbf{D}_1 \mathbf{D}_1^H \mathbf{D}_1 - 2\mathbf{D}_1^H \mathbf{D}_1 + \mathbf{I}_{DD_1}) \\ &= \text{tr}(\mathbf{D}_1 \mathbf{D}_1^H \mathbf{D}_1 \mathbf{D}_1^H - 2\mathbf{D}_1 \mathbf{D}_1^H + \mathbf{I}_{\tau_1}) + (DD_1 - \tau_1) \\ &= \|\mathbf{D}_1 \mathbf{D}_1^H - \mathbf{I}_{\tau_1}\|_F^2 + (DD_1 - \tau_1). \end{aligned} \quad (40)$$

Considering the structure of $\mathbf{D}_1$ and properties of the Kronecker product, Eq. (40) can be simplified as:

$$\begin{aligned} \mathbf{D}_1 \mathbf{D}_1^H &= \Theta_{mc,1}^H (\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M) (\tilde{\mathbf{A}}_{M,1}^* \otimes \tilde{\mathbf{A}}_M)^H \Theta_{mc,1} \\ &= \Theta_{mc,1}^H (\tilde{\mathbf{A}}_{M,1}^* \tilde{\mathbf{A}}_{M,1}^T \otimes \tilde{\mathbf{A}}_M \tilde{\mathbf{A}}_M^H) \Theta_{mc,1}. \end{aligned} \quad (41)$$

Based on the over-complete dictionary construction in Eqs. (23a)–(23c) and the asymptotic orthogonality of steering vector matrices established in [17], we have:

$$\begin{aligned} \tilde{\mathbf{A}}_M \tilde{\mathbf{A}}_M^H &= (\tilde{\mathbf{A}}_{M_v} \otimes \tilde{\mathbf{A}}_{M_h}) (\tilde{\mathbf{A}}_{M_v} \otimes \tilde{\mathbf{A}}_{M_h})^H \\ &= \tilde{\mathbf{A}}_{M_v} \tilde{\mathbf{A}}_{M_v}^H \otimes \tilde{\mathbf{A}}_{M_h} \tilde{\mathbf{A}}_{M_h}^H \\ &\approx D_v \mathbf{I}_{M_v} \otimes D_h \mathbf{I}_{M_h} = D \mathbf{I}_M. \end{aligned} \quad (42)$$

Following the reasoning in Eq. (42), we also have $\tilde{\mathbf{A}}_{M,1}^* \tilde{\mathbf{A}}_{M,1}^T \approx D_1 \mathbf{I}_M$. Substituting this and Eq. (42) into Eq. (41), we obtain $\mathbf{D}_1 \mathbf{D}_1^H \approx \Theta_{mc,1}^H (D_1 \mathbf{I}_M \otimes D \mathbf{I}_M) \Theta_{mc,1} = DD_1 \Theta_{mc,1}^H \Theta_{mc,1}$. Hence, by applying the same derivation steps as in Eq. (40), the original optimization problem in Eq. (39) simplifies to

$$\begin{aligned} \min_{\Theta_{cv,1}} & \|\Theta_{mc,1} \Theta_{mc,1}^H - \mathbf{I}_{M^2}\|_F^2 \\ \text{s.t.} & |[\Theta_{cv,1}]_{m,t}| = 1, 1 \leq m \leq M, 1 \leq t \leq \tau. \end{aligned} \quad (43)$$

This reformulation not only ensures approximate orthogonality of the equivalent dictionary but also facilitates the application of Eq. (20) in subspace-based algorithms with reduced pilot overhead. Notably, since $\Theta_{mc,k}^H \in \mathbb{C}^{\tau_k \times M^2}$, the condition $\tau_k \geq M^2$ is required to guarantee the existence of its left pseudo-inverse.

To tackle the non-convex optimization problem in Eq. (43) subject to unit-modulus constraints on complex variables, an efficient Riemannian manifold optimization framework is developed. Let $f = \|\Theta_{mc,1} \Theta_{mc,1}^H - \mathbf{I}_{M^2}\|_F^2$. By exploiting the linearity of the trace operator, the objective function can be expressed as

$$\begin{aligned} f &= \text{tr}((\Theta_{mc,1} \Theta_{mc,1}^H - \mathbf{I}_{M^2})^H (\Theta_{mc,1} \Theta_{mc,1}^H - \mathbf{I}_{M^2})) \\ &= \text{tr}((\Theta_{mc,1} \Theta_{mc,1}^H)^2) - 2\text{tr}(\Theta_{mc,1} \Theta_{mc,1}^H) + M^2. \end{aligned} \quad (44)$$

The Euclidean gradient of the objective function f with respect to $\Theta_{mc,1}$ is given by

$$\frac{\partial f}{\partial \Theta_{mc,1}} = 4\Theta_{mc,1} (\Theta_{mc,1}^H \Theta_{mc,1} - \mathbf{I}_{\tau_1}). \quad (45)$$

Accordingly, the partial derivative with respect to η_t is $\frac{\partial f}{\partial \eta_t} = 4\Theta_{mc,1} [\Theta_{mc,1}^H \Theta_{mc,1} - \mathbf{I}_{\tau_1}]_{:,t}$. To compute the element-wise derivatives in Eq. (45), the partial derivative of $\mathbf{B}_t = (\Gamma_t^{-1} - \mathbf{S})^{-1}$ with respect to $[\gamma_t]_m$ is evaluated first. Using the matrix inverse differentiation identity yields

$$\frac{\partial \Gamma_t^{-1}}{\partial [\gamma_t]_m} = -\Gamma_t^{-1} \left(\frac{\partial \Gamma_t}{\partial [\gamma_t]_m} \right) \Gamma_t^{-1} = -[\gamma_t]_m^{-2} \mathbf{E}_m, \quad (46a)$$

$$\frac{\partial \mathbf{B}_t}{\partial [\gamma_t]_m} = -\mathbf{B}_t \left(\frac{\partial (\Gamma_t^{-1} - \mathbf{S})}{\partial [\gamma_t]_m} \right) \mathbf{B}_t = [\gamma_t]_m^{-2} \mathbf{B}_t \mathbf{E}_m \mathbf{B}_t, \quad (46b)$$

where $\mathbf{E}_m = \text{Diag}(\mathbf{e}_m) \in \mathbb{R}^{M \times M}$, and $\mathbf{e}_m \in \mathbb{R}^{M \times 1}$ is the standard basis vector whose m -th entry is 1 and all others are zero. By applying the vectorization identity, we derive

$$\frac{\partial \eta_t}{\partial [\gamma_t]_m} = [\gamma_t]_m^{-2} (\mathbf{B}_t^T \otimes \mathbf{B}_t) \text{vec}(\mathbf{E}_m). \quad (47)$$

Since $\text{vec}(\mathbf{E}_m) \in \mathbb{R}^{M^2 \times 1}$ contains only a single nonzero entry at position $(m-1)M + m$, we apply the chain rule and matrix calculus identities to compute the partial derivative of f with respect to $[\gamma_t]_m^*$ as

$$\mathbf{G}_{m,t}^E \triangleq \frac{\partial f}{\partial [\gamma_t]_m^*} = 4[\gamma_t]_m^{-2} [\mathbf{B}_t^* \odot \mathbf{R}_t]_{m,m}, \quad (48)$$

where $\mathbf{R}_t = \text{mat} \left(\Theta_{\text{mc},1} \left[\Theta_{\text{mc},1}^H \Theta_{\text{mc},1} - \mathbf{I}_{\tau_1} \right]_{:,t} \right) \in \mathbb{C}^{M \times M}$, and $\mathbf{G}^E \in \mathbb{C}^{M \times \tau_1}$ denotes the Euclidean gradient matrix. To enforce the unit-modulus constraint $[\gamma_t]_m \in \mathbb{S}_\mathbb{C}^1$, the Euclidean gradient is projected onto the tangent space of the complex circle manifold, yielding the Riemannian gradient matrix $\mathbf{G}^R \in \mathbb{C}^{M \times \tau_1}$ as follows

$$\mathbf{G}_{:,t}^R = \mathbf{G}_{:,t}^E - \text{Re} \{ \mathbf{G}_{:,t}^E \odot \gamma_t^* \} \odot \gamma_t. \quad (49)$$

The resulting optimization problem is solved using the Polak–Ribière conjugate gradient method on the Riemannian manifold. The initial value of $\Theta_{\text{cv},k}$ ($\forall k \in \mathcal{K}$) is set to the random Bernoulli matrix, with each element independently drawn from $\{-1, +1\}$ with equal probability [16].

V. SIMULATION RESULTS

In this section, we present simulation results to evaluate the performance of the proposed MC-aware three-stage channel estimation protocol for a RIS-aided MU-MISO system. The RIS is centered at $(0, 0, 0)$ facing the positive X-axis. The BS is located at $100 \text{ m} \times (\sin \frac{5\pi}{6}, \cos \frac{5\pi}{6}, 0)$, and each user is placed at $10 \text{ m} \times (\sin \frac{\pi}{3}, \cos \frac{\pi}{3}, 0)$. The complex channel gains are modeled as $\alpha_l \sim \mathcal{CN}(0, 10^{-3} d_{\text{BR}}^{-2.2})$ and $\beta_{k,j} \sim \mathcal{CN}(0, 10^{-3} d_{\text{RU}}^{-2.8})$, where d_{BR} and d_{RU} denote the distances between the BS and RIS, and between the RIS and the users, respectively. The antenna spacing at the BS is set to $d_{\text{BS}} = \lambda_c/2$. The thermal noise power at both the BS and the RIS is fixed at $\sigma_1^2 = \sigma_2^2 = -80 \text{ dBm}$. The number of users is $K = 4$, and the carrier frequency is set to $f = 28 \text{ GHz}$. The RIS passive elements are modeled as cylindrical thin wires made of perfectly conducting material, each having a length $l = \lambda_c/32$ and a radius $a = \lambda_c/500$, where $a \ll l$ to satisfy the thin-wire approximation. The characteristic impedance is assumed to be $Z_0 = 50\Omega$. Unless otherwise specified, the transmit power is set to $p = 25 \text{ dBm}$. The inter-element spacing at the RIS is $d_{\text{RIS}} = \lambda_c/2$. The dimensions of the UPAs deployed at the BS and the RIS are $N_h \times N_v = 8 \times 8$ and $M_h \times M_v = 4 \times 4$, respectively. The number of propagation paths is set to $L = 3$ for the RIS-BS link and $J_1 = \dots = J_K = J = 2$ for all user-RIS links. To evaluate the performance of the proposed channel estimation protocol, we adopt the normalized mean square error (NMSE) as the primary metric, defined as

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\sum_{k=1}^K \|\hat{\mathbf{Y}}_{i,k} - \bar{\mathbf{Y}}_k\|_F^2}{\sum_{k=1}^K \|\bar{\mathbf{Y}}_k\|_F^2} \right\},$$

where $\bar{\mathbf{Y}}_k$ denotes the noise-free received signal for user k , and the subscript $i \in \{\text{mc}, \text{cv}\}$ indicates whether the MC-aware or conventional channel model is employed. Specifically, the reconstructed signals are given by $\hat{\mathbf{Y}}_{\text{mc},k} = \sqrt{p} \hat{\mathbf{G}}_{\text{mc},k} \Theta_{\text{mc},k}$ and $\hat{\mathbf{Y}}_{\text{cv},k} = \sqrt{p} \hat{\mathbf{G}}_{\text{cv},k} \Theta_{\text{cv},k}$.

A comprehensive performance evaluation is conducted to compare several algorithms integrated into the proposed MC-aware three-stage channel estimation protocol. The following summarizes the key features of each algorithm.

- **Proposed1** : Proposed1 adopts the Root-MUSIC algorithm as the dimension-reduced subspace-based common AoA estimation, as implemented in Algorithm 1 during

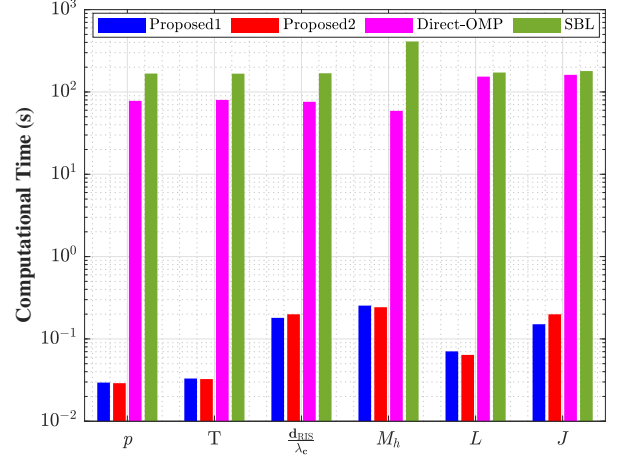


Fig. 2. Computational time of the algorithms under different variables.

Stage I. Root-MUSIC identifies the roots of a polynomial that lie closest to the unit circle in the complex plane, where the angular positions of these roots directly correspond to the DOAs [23].

- **Proposed2** : Proposed2 employs the TLS-ESPRIT algorithm as the dimension-reduced subspace-based common AoA estimation, also implemented in Algorithm 1 during Stage I. TLS-ESPRIT refines the solution to the rotational invariance equation using the TLS criterion, which accounts for perturbations in both the signal subspace and the observed data [24].
- **MC-unaware** : The MC-unaware algorithm employs the Root-MUSIC algorithm for common AoA estimation and uses the estimation strategy described in Section III-D for estimating the remaining component $\mathbf{H}_{\text{RIS}}^{\text{cv},k}$.
- **Direct-OMP** : Direct-OMP formulates the vectorized received signal as $\text{vec}(\mathbf{Y}_k) = \sqrt{p}(\Theta_k^T \otimes \mathbf{I}_N) \text{vec}(\mathbf{G}_k) + \text{vec}(\mathbf{N}_k)$, based on direct vectorization of Eq. (13) [19]. The estimation problem is then solved via $L^2 J_k$ -sparse signal recovery problem, which incurs both high computational complexity and significant pilot overhead³.
- **SBL** : Sparse Bayesian learning can be employed to solve the sparse signal recovery problem. The equivalent measurement matrix $\check{\mathbf{Y}}_k$ in Eq. (20) is recovered using SBL, after eliminating the common AoA effect via the Root-MUSIC algorithm. The algorithm is executed with the maximum number of iterations set to 50 and the convergence threshold set to 10^{-3} .

To facilitate a direct comparison of computational complexity, Fig. 2 presents the computational time per channel simulation under varying parameters. As shown in the figure, the proposed methods exhibit substantially lower computational complexity compared to the other algorithms.

³ $\text{vec}(\mathbf{G}_k) = \left((\mathbf{A}_{M,k}^T \otimes \mathbf{A}_{M,k}^H)^T \otimes \mathbf{A}_N \right) \text{vec}(\beta_k^T \otimes \mathbf{A})$. Moreover, by Kronecker product properties, optimizing the RIS phase shift training matrix as described in Problem (43) enhances the orthogonality of the equivalent dictionary matrix in this formulation.

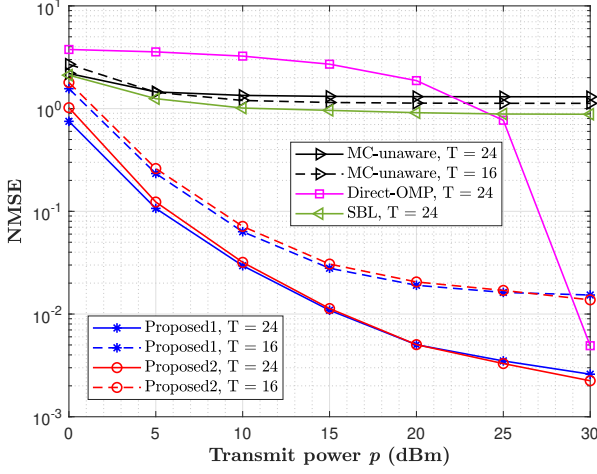


Fig. 3. NMSEs vs. Transmit power p .

Fig. 3 illustrates the NMSE performance as a function of the transmit power, evaluated under two average pilot overhead settings: $T = 24$ and $T = 16$. In the case of $T = 24$, the typical user (user 1) is allocated $\tau_1 = 30$ pilots, while each other user is assigned $\tau_k = 22$; for $T = 16$, $\tau_1 = 22$ and $\tau_k = 14$. The larger allocation to the typical user is intended to mitigate error propagation in the three-stage estimation process. The results show that both Proposed1 and Proposed2 achieve a substantial NMSE reduction as p increases. At lower transmit powers, Proposed1 slightly outperforms Proposed2. In contrast, the Direct-OMP algorithm suffers from high NMSE at low power due to its high-dimensional sparse recovery formulation. It only reaches comparable performance to the proposed methods when the transmit power reaches 30 dBm. Therefore, to ensure fair comparisons in subsequent simulations, the transmit power is fixed at $p = 25$ dBm. The SBL method maintains high NMSE across the transmit power range. Although it outperforms the MC-unaware baseline, its performance is hindered by the difficulty of accurately estimating noise covariance in RIS-aided system model. Meanwhile, the MC-unaware methods consistently underperform, further highlighting the importance of modeling MC-aware in RIS-aided models.

Fig. 4 presents the NMSE performance as a function of the average pilot overhead, with two transmit power levels $p = 25$ dBm and $p = 15$ dBm for the proposed methods. The figure demonstrates that the NMSEs of Proposed1 and Proposed2 decrease monotonically as the pilot overhead increases, indicating improved estimation accuracy with more pilot resources. Notably, Proposed1 achieves slightly better performance than Proposed2 in low pilot regimes. Both Direct-OMP and SBL exhibit gradual improvements as T increases. However, their estimation accuracy remains significantly inferior to that of the proposed approaches, particularly at low transmit power ($p = 15$ dBm), underscoring their higher pilot and power requirements. Direct-OMP's poor performance stems from its reliance on full vectorization and large-scale sparse recovery, which imposes stringent demands on pilot

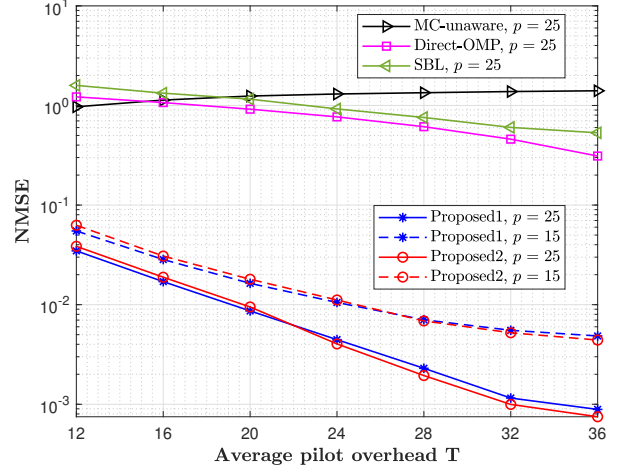


Fig. 4. NMSEs vs. Average pilot overhead T .

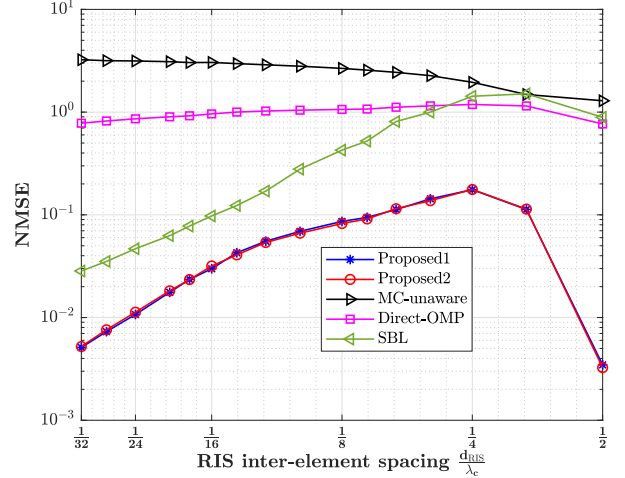


Fig. 5. NMSEs vs. RIS inter-element spacing $\frac{d_{\text{RIS}}}{\lambda_c}$ with $p = 25$ dBm.

overhead. Interestingly, the MC-unaware method's NMSEs worsen as pilot overhead grows. This phenomenon, also observed in Fig. 3, is attributed to the model mismatch that becomes increasingly pronounced as pilot dimensions expand. Specifically, the divergence between the conventional RIS channel model and the MC-aware formulation grows with the dimensionality of the training matrices Θ_{cv} and Θ_{mc} , leading to increased estimation errors.

Fig. 5 illustrates the NMSE performance as a function of the RIS inter-element spacing, which controls the severity of MC effects. As the inter-element spacing decreases, the proposed MC-aware methods exhibit a non-monotonic NMSE trend with decreasing RIS inter-element spacing, initially increasing before declining. In particular, NMSE peaks around $\frac{d_{\text{RIS}}}{\lambda_c} = \frac{1}{6}$ to $\frac{d_{\text{RIS}}}{\lambda_c} = \frac{1}{3}$. Nevertheless, both Proposed1 and Proposed2 consistently outperform SBL and Direct-OMP across all spacing levels. Notably, SBL performs relatively well under small RIS inter-element spacing conditions. The MC-unaware method displays a steadily degrading trend as inter-element spacing

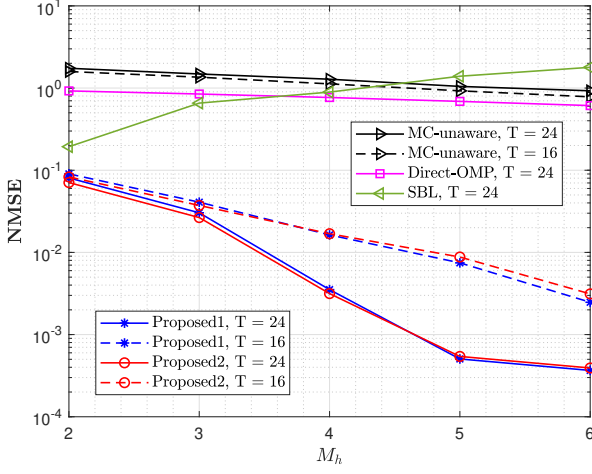


Fig. 6. NMSEs vs. Number of elements at the RIS $M = M_h \times M_v$ with $p = 25$ dBm.

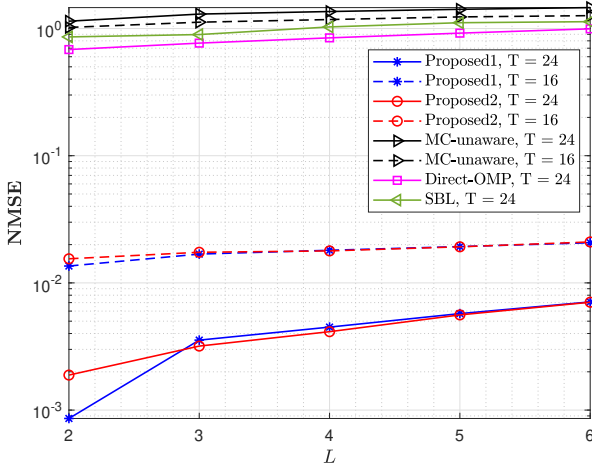


Fig. 7. NMSEs vs. Numbers of propagation paths in the RIS-BS link L with $p = 25$ dBm.

decreases, underscoring its vulnerability to model mismatch under increasing MC severity. This highlights the importance of incorporating MC-aware modeling, particularly when dense RIS deployments are considered.

Fig. 6 presents the NMSE performance as a function of the number of RIS elements along the horizontal dimension, under two average pilot overhead settings: $T = 24$ and $T = 16$. Increasing M_h expands the RIS aperture while maintaining fixed inter-element spacing, thereby reducing the MC effect. This structural change explains the observed trend: the NMSEs of proposed methods gradually decrease with increasing M_h . In low M_h regimes Proposed2 slightly outperforms Proposed1. The Direct-OMP and MC-unaware methods continue to exhibit high NMSE values across all settings. Notably, the performance of SBL worsens as M_h grows. This degradation is attributed to the increasing dimensionality of the equivalent dictionary matrix $\mathbf{D}_k$, which complicates convergence and elevates estimation error in iteration.

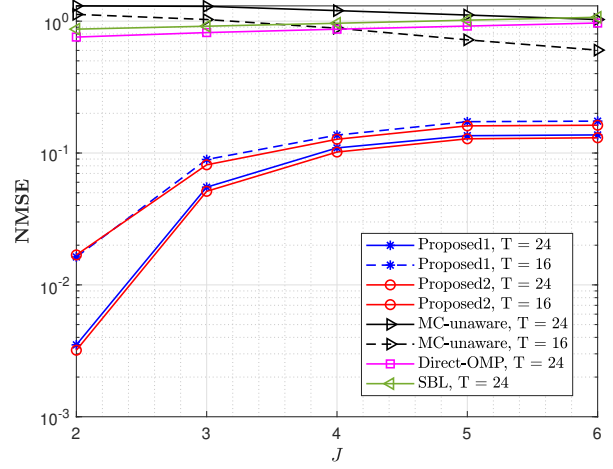


Fig. 8. NMSEs vs. Numbers of propagation paths in the user-RIS links J with $p = 25$ dBm.

Fig. 7 depicts the impact of the numbers of propagation paths in the RIS-BS link on the NMSE performance, under two average pilot overhead settings: $T = 24$ and $T = 16$. As L increases, a consistent performance degradation is observed across all methods. This degradation is primarily attributed to two factors: (i) an increase in the number of unknown channel parameters (e.g., AoAs at the BS, AoDs at the RIS, and path gains), which exacerbates the estimation difficulty under fixed pilot overhead; and (ii) the need to estimate the number of paths itself, introducing additional model uncertainty. Despite the performance drop at larger L , the proposed methods consistently outperform the other methods across all settings.

Fig. 8 illustrates the impact of the numbers of propagation paths in the user-RIS links on the NMSE performance, under two average pilot overhead settings: $T = 24$ and $T = 16$. As J increases, the NMSEs show a rising trend, reflecting increased estimation difficulty as explained in analysis presented in Fig. 7. This performance degradation is more pronounced compared to L due to the direct influence of J on the sparsity level in sparse signal recovery problems. In contrast, the MC-unaware method exhibits performance improvement as J increases. This is because a larger number of paths enhances angular diversity, facilitating better equivalent dictionary matching while maintaining fixed resolution. However, its performance remains significantly inferior to the MC-aware proposed methods, especially in low pilot overhead regimes.

Fig. 9 examines the impact of optimizing the RIS phase shift training matrix on estimation performance. Specifically, it shows the NMSE performance as a function of transmit power in a single-user scenario ($K = 1$), evaluated under two pilot overhead settings, $T = 28$ and $T = 12$. The Non-optimized scheme employs the Root-MUSIC algorithm in Stage I but does not employ optimized RIS phase shift training matrix; instead, it relies on the random phase shift matrix. The performance gap between the Non-optimized scheme and Proposed1 demonstrates the performance gains achieved through RIS phase shift training matrix optimization. Moreover, the trends of the other algorithms with respect to

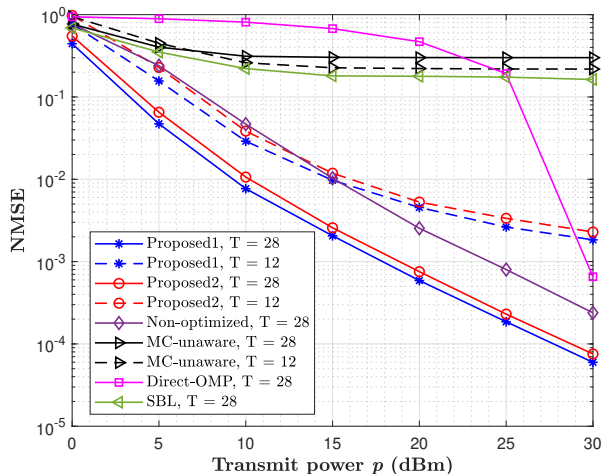


Fig. 9. NMSEs vs. Transmit power p .

transmit power are consistent with those observed in Fig. 3.

VI. CONCLUSION

In this paper, we proposed an MC-aware three-stage channel estimation protocol for RIS-aided MU-MISO mmWave systems, which accurately captures the cascaded channel characteristics under practical RIS deployment conditions. We leveraged a dimension-reduced subspace-based method to achieve accurate estimation of common AoA at the BS in Stage I. Then an OMP-based approach was developed to enable low-complexity estimation of the cascaded channel for the typical user, while simultaneously estimating common AoD at the RIS. We further reduced pilot overhead for other users by leveraging the fact that all users share the common RIS-BS channel. Additionally, we proposed a phase shift training matrix design strategy for RIS that accounts for MC effect and achieves performance gains compared to random phase scheme. Simulation results validate that proposed method significantly outperforms MC-unaware and existing approaches in terms of both estimation accuracy and pilot efficiency.

REFERENCES

- [1] M. Di Renzo *et al.*, “Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [2] C. Pan *et al.*, “Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions,” *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, June 2021.
- [3] K. Zhi, C. Pan, H. Ren, and K. Wang, “Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI,” *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3558–3574, May 2022.
- [4] C. Pan *et al.*, “Multicell MIMO communications relying on intelligent reflecting surfaces,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.
- [5] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, “Channel estimation with reconfigurable intelligent surfaces—a general framework,” *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, Sept. 2022.
- [6] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, “A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels,” *IEEE Trans. Signal Processing*, vol. 68, pp. 5092–5106, Aug. 2020.

- [7] G. Gradoni and M. Di Renzo, “End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 938–942, May 2021.
- [8] H. Li, S. Shen, M. Nerini, M. Di Renzo, and B. Clerckx, “Beyond diagonal reconfigurable intelligent surfaces with mutual coupling: Modeling and optimization,” *IEEE Commun. Lett.*, vol. 28, no. 4, pp. 937–941, Apr. 2024.
- [9] A. Abrardo, A. Toccafondi, and M. Di Renzo, “Design of reconfigurable intelligent surfaces by using S-parameter multiport network theory—optimization and full-wave validation,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 17 084–17 102, Nov. 2024.
- [10] S. Shen, B. Clerckx, and R. Murch, “Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1229–1243, Feb. 2022.
- [11] Z. Tan, P. Yang, and A. Nehorai, “Joint sparse recovery method for compressed sensing with structured dictionary mismatches,” *IEEE Trans. Signal Processing*, vol. 62, no. 19, pp. 4997–5008, Oct. 2014.
- [12] D. Fan *et al.*, “Angle domain channel estimation in hybrid millimeter wave massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.
- [13] X. Wu, S. Ma, X. Yang, and G. Yang, “Clustered sparse Bayesian learning based channel estimation for millimeter-wave massive MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 12 749–12 764, Dec. 2022.
- [14] P. Wang, J. Fang, H. Duan, and H. Li, “Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems,” *IEEE Signal Processing Lett.*, vol. 27, pp. 905–909, May 2020.
- [15] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, “Channel estimation for reconfigurable intelligent surface aided multi-user mmwave MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6853–6869, Oct. 2023.
- [16] X. Wei, D. Shen, and L. Dai, “Channel estimation for RIS assisted wireless communications—part ii: An improved solution based on double-structured sparsity,” *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.
- [17] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, “Channel estimation for RIS-aided multiuser millimeter-wave systems,” *IEEE Trans. Signal Processing*, vol. 70, pp. 1478–1492, Mar. 2022.
- [18] Z. Peng *et al.*, “Channel estimation for RIS-aided multi-user mmwave systems with uniform planar arrays,” *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8105–8122, Dec. 2022.
- [19] P. Zheng, S. Tarboush, H. Saeeddeen, and T. Y. Al-Naffouri, “Mutual coupling-aware channel estimation and beamforming for RIS-assisted communications,” *arXiv preprint arXiv:2410.04110*, 2024.
- [20] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhatieb, and R. W. Heath, “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
- [21] P. Zheng, X. Ma, and T. Y. Al-Naffouri, “On the impact of mutual coupling on RIS-assisted channel estimation,” *IEEE Wireless Commun. Lett.*, vol. 13, no. 5, pp. 1275–1279, May 2024.
- [22] Y. Changbo, Z. Beizuo, L. Baobao, and Z. Xiaofei, “Computationally efficient 2D-DOA estimation for uniform planar arrays: RD-ROOT-MUSIC algorithm,” *Trans. Nanjing Univ. Aeronaut. Astronaut.*, vol. 38, no. 4, Aug. 2021.
- [23] B. Rao and K. Hari, “Performance analysis of Root-Music,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.
- [24] B. Ottersten, M. Viberg, and T. Kailath, “Performance analysis of the total least squares ESPRIT algorithm,” *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.
- [25] V. Slyusar, “A family of face products of matrices and its properties,” *Cybern. Syst. Anal.*, vol. 35, no. 3, pp. 379–384, May 1999.
- [26] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [27] J. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [28] J. M. Duarte-Carvajalino and G. Sapiro, “Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization,” *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1395–1408, July 2009.