Channel Estimation for RIS-Aided Multiuser Millimeter-Wave Massive MIMO Systems

Gui Zhou, Cunhua Pan, Hong Ren, Petar Popovski, *IEEE Fellow*, A. Lee Swindlehurst, *IEEE Fellow*

Abstract

Reconfigurable intelligent surface (RIS) is a promising technique that can reconfigure the electromagnetic propagation environment by adjusting the phase shifts of the reflecting elements on the RIS. However, channel estimation in the RIS-aided massive multiuser multiple-input single-output (MU-MISO) wireless communication systems is challenging due to the passive feature of RIS and the large number of reflecting elements that incur high channel estimation overhead. To address this issue, we propose a novel cascaded channel estimation strategy with low pilot overhead by exploiting the sparsity and the correlation of multiuser cascaded channels in millimeter-wave massive MISO systems. Based on the fact that the phsical positions of the BS, the RIS and users may not change in several or even tens of consecutive channel coherence blocks, we first estimate the full channel state information (CSI) including all the angle and gain information in the first coherence block, and then only re-estimate the channel gains in the remaining coherence blocks with much less pilot overhead. In the first coherence block, we propose a two-phase channel estimation method, in which the cascaded channel of one typical user is estimated in Phase I based on the linear correlation among cascaded paths, while the cascaded channels of other users are estimated in Phase II by utilizing the partial CSI of the common base station (BS)-RIS channel obtained in Phase I. The total theoretical minimum pilot overhead in the first coherence block is $8J-2+(K-1)\lceil(8J-2)/L\rceil$, where K, L and J denote the numbers of users, paths in the BS-RIS channel and paths in the RIS-user channel, respectively. In each of the remaining coherence blocks, the minimum pilot overhead is JK. Moreover, the training phase shift matrices at the RIS are optimized to improve the estimation performance. Simulation results show that the performance

(Corresponding author: Cunhua Pan)

G. Zhou and C. Pan are with the School of Electronic Engineering and Computer Science at Queen Mary University of London, London E1 4NS, U.K. (e-mail: g.zhou, c.pan@qmul.ac.uk). H. Ren is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (hren@seu.edu.cn). Petar Popovski is with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: petarp@es.aau.dk). A. L. Swindlehurst is with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697, USA (e-mail: swindle@uci.edu).

of the proposed method outperforms the existing methods in terms of the estimation accuracy when using the same amount of pilot overhead.

Index Terms

Intelligent reflecting surface (IRS), reconfigurable intelligent surface (RIS), Millimeter wave, massive MIMO, AoA/AoD estimation, channel estimation.

I. INTRODUCTION

Reconfigurable intelligent surface (RIS) is expected to enhance the coverage and capacity in the wireless communication systems with relatively low hardware cost and energy consumption [1]–[5]. In general, RIS is composed of a large number of passive elements, which can assist the wireless communication by reconfiguring the electromagnetic propagation environment between transmitter and receiver. The performance gain provided by the RIS highly relies on the accuracy of the channel state information (CSI). However, it is challenging to acquire the CSI since the reflecting elements at the RIS are passive devices lacking the ability of transmitting, receiving and processing pilot signals.

It is observed that the CSI of the cascaded base station (BS)-IRS-user channel, which is the product of the BS-IRS channel and the IRS-user channel, is sufficient for the transmission design [6], [7]. As a result, most of the existing contributions focused on the cascaded channel estimation [8]–[14]. Specifically, consider a system containing a BS with N antennas, K single-antenna users, and one IRS with M reflecting elements. The authors in [8] proposed least-square (LS)-based estimation method to obtain the cascaded channel estimator which is unbiased in the single-user multiple-input single-output (SU-MISO) system. However, the pilot overhead of the LS-based estimation method is prohibitively high, which scales up with M. To reduce the pilot overhead, [9] divided the elements of RIS into P subgroups, and proposed a transmission protocol to execute channel estimation and phase shift optimization successively with pilot overhead of P. By utilizing the common BS-RIS channel and the linear correlation among the RIS-user channels in multiuser multiple-input single-output (MU-MISO) system, the authors in [10] further proposed a channel estimation strategy, whose pilot overhead is inversely proportional to the number of the antennas at the BS, that is $M+\max(K-1, K\left\lceil\frac{(K-1)M}{N}\right\rceil)$. The estimation method in [10] requires low pilot overhead in a rich scattering communication scenario such that the cascaded channel is full rank, but this method is not applicable in the millimeter-wave (mmWave) massive MISO communication system, in which the mmWave channel is rank-deficient due to the spatial sparsity [15].

To address this issue, the authors in [11]–[14] exploited the sparse property of the cascaded channel matrix in mmWave communication systems and proposed compressed sensing (CS)-based channel estimation methods with low pilot overhead. In particular, [11] directly constructed a sparse signal recovery problem for the cascaded channel estimation and ignored the common parameters of the cascaded channel in SU-MISO system, which leads high power leakage. Thus, the adopted grid-on CS method has high false alarm probability and high estimation error. In order to suppress the effect of power leakage, the atomic norm minimization method as used in [12] to estimate the sparse angles and gains. In the MU-MISO systems, both [13] and [14] investigated the double sparse structure of the cascaded channel and utilized common parameters to jointly estimate the multiuser cascaded channels with low pilot overhead and high estimation accuracy. However, these two papers assumed that the number of BS-RIS channel paths L and the number of RIS-user channel paths J are available, which is difficult to obtain these information. Moreover, the pilot overhead in [13] is proportional to the quotient of the number of RIS elements divided by the number of cascaded spatial paths, i.e., $K\left\lceil\frac{M}{JL}\right\rceil$, which is excessive in large RIS systems with a large number of reflecting elements. Therefore, it motivates us to develop an efficient channel estimation strategy to further reduce the pilot overhead, as well as estimate the sparse level or the number of the spatial paths.

A. Novelty and contributions

Against the above background, this paper proposes a novel uplink cascaded channel estimation strategy in the RIS-aided mmWave massive MU-MISO systems. The proposed estimation strategy has the following appealing features: low pilot overhead, low computational complexity, sparse level (the number of spatial paths) estimation of cascaded channels. These appealing features are achieved based on the following three facts:

Fact 1: The physical positions of the BS and the RIS change much slower than the channel variation [15]. Therefore, it is reasonable to assume that the angle-of-arrival (AoAs) at the BS, and the AoAs and angle-of-departure (AoDs) at the RIS remain unchanged within tens of channel coherence blocks. Once the angle information is estimated in the first channel coherence block, only the cascaded channel gains need to be re-estimated in the following channel coherence blocks. This can greatly reduce the pilot overhead and computational complexity of channel estimation in the remaining channel coherence blocks, because only a few parameters need to be estimated.

Fact 2: The JL cascaded paths are the combination of $J+L$ independent spatial paths. This means that there is a linear correlation among JL cascaded paths, which motivates us to directly estimate $J+L$ sparse paths, rather than estimate JL cascaded sparse paths. Note that the existing contributions in [11]–[14] estimate JL cascaded sparse paths.

Fact 3: All users share a common BS-RIS channel. Based on this fact, [13], [14] utilized the common AoA information of the common BS-RIS channel to simplify the multiuser channel estimation and reduce pilot overhead. In this work, we exploit the AoA, AoD and gain information of the common BS-RIS channel to construct a partial common BS-RIS channel, which enables us to develop a new multiuser channel estimation method with less pilot overhead.

Based on the above discussion, the main contributions of this work are summarized as follows:

- We propose a novel uplink channel estimation protocol in time division duplex (TDD) RIS-aided mmWave massive MU-MISO communication systems, as shown in Fig. 1. Based on **Fact 1**, we consider several or even tens of consecutive channel coherence blocks, where the angle information keeps invariant while the channel gains vary at each channel coherence block. Then, at the first coherence block, we first estimate the full CSI information, including all the angle information and the channel gains. With the estimated angle information, only channel gains need to be estimated in the remaining coherence blocks, which can be obtained by using simple LS method with much less pilot overhead of length JK. Moreover, the training phase shift matrices are optimized to minimize the mutual coherence of the equivalent dictionary for better estimation performance.
- In the first coherence block, we propose a two-phase channel estimation method by making use of **Fact 2** and **Fact 3**. In particular, in Phase I, one typical user sends a sequence of pilots to the BS for cascaded channel estimation. The required theoretical minimum pilot overhead can be down to $8J-2$ by exploiting the linear correlation among the cascaded paths based on **Fact 2**. According to **Fact 3**, we extract the partial CSI of the common BS-RIS channel from Phase I, which can help reduce the pilot overhead for other users. In Phase II, the other users successively transmit pilots to the BS for channel estimation. With the knowledge of the partial common BS-RIS channel, the required theoretical minimum pilot overhead can be reduced to $(K-1)\left\lceil (8J-2)/L \right\rceil$. Therefore, the total theoretical minimum pilot overhead in the first coherence block is $8J-2+(K-1)\left\lceil (8J-2)/L \right\rceil$.

- We demonstrate through numerical results that the proposed cascaded channel estimation strategy outperforms the existing orthogonal matching pursuit (OMP)-based channel estimation algorithm in terms of the mean square error (MSE), the pilot overhead and the computational complexity. Moreover, the MSE performance of the proposed estimation algorithm is close to the performance lower bound at low SNR.

The remainder of this paper is organized as follows. Section II introduces the system model and the cascaded channel sparsity model. The cascaded channel estimation strategy is investigated in Section III. Training phase shift matrices are optimized in Sections V. Section IV compares the pilot overhead and algorithm complexity between the proposed algorithm and the existing algorithms. Finally, Sections VI and VII report the numerical results and conclusions, respectively.

Notations: The following mathematical notations and symbols are used throughout this paper. Vectors and matrices are denoted by boldface lowercase letters and boldface uppercase letters, respectively. The symbols $\mathbf{X}^*$, $\mathbf{X}^{\mathrm{T}}$, $\mathbf{X}^{\mathrm{H}}$, and $||\mathbf{X}||_F$ denote the conjugate, transpose, Hermitian (conjugate transpose), Frobenius norm of matrix $\mathbf{X}$, respectively. The symbol $||\mathbf{x}||_2$ denotes 2-norm of vector $\mathbf{x}$. The symbols $\mathrm{Tr}\{\cdot\}$, $\mathrm{Re}\{\cdot\}$, $|\cdot|$, and $\angle(\cdot)$ denote the trace, real part, modulus, and angle of a complex number, respectively. $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix with the entries of $\mathbf{x}$ on its main diagonal. $[\mathbf{x}]_m$ means the m-th element of the vector $\mathbf{x}$, and $[\mathbf{X}]_{m,n}$ means the (m,n)-th element of the matrix $\mathbf{X}$. $\mathbf{X}_{(:,n)}$ and $\mathbf{X}_{(m,:)}$ mean the n-th column and the m-th row of matrix $\mathbf{X}$. The Kronecker product between two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{X} \otimes \mathbf{Y}$. The Khatri-Rao product between two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{X} \odot \mathbf{Y}$. Additionally, the symbol $\mathbb{C}$ denotes complex field, $\mathbb{R}$ represents real field, and $\mathrm{i} \triangleq \sqrt{-1}$ is the imaginary unit. The inner product $\langle \bullet, \bullet \rangle : \mathbb{C}^{M\times N} \times \mathbb{C}^{M\times N} \to \mathbb{R}$ is defined as $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathbb{R}\{\mathrm{Tr}\{\mathbf{X}^{\mathrm{H}}\mathbf{Y}\}\}$. $\lceil\rceil$ round up to an integer. $\lceil\rfloor$ round towards nearest integer.

II. System Model and Channel Model

A. Signal Model

We consider a narrow-band TDD mmWave massive MISO system where K single-antenna users communicate with an N-antenna BS. To enhance the communication performance, an RIS equipped with M passive reflecting elements, each of which can be dynamically adjusted for electromagnetic wave reconstruction between the BS and users, is deployed to increase the spatial diversity.

In this paper, we consider quasi-static block-fading channels, where each channel remains approximately constant in a channel coherence block with B time slots. Due to channel reciprocity, the CSI of the downlink channel can be obtained by estimating the CSI of the uplink channel. We assume that τ time slots of each coherence block are used for uplink channel estimation and the remaining $B-\tau$ time slots for downlink data transmission. Here, we assume that the knowledge of the direct channels between BS and users are available at the BS due to the fact that the direct BS-user channel estimation in mmWave systems has been extensively studied [16]. Therefore, we only focus on the uplink channel estimation of the user-RIS links and the RIS-BS link.

Let $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ denote the channel from user k to the RIS and $\mathbf{H} \in \mathbb{C}^{N\times M}$ denote the channel from the RIS to the BS. Moreover, denote by $\mathbf{e}_t \in \mathbb{C}^{M\times 1}$ the phase shift vector of the RIS at time slot t in the considered coherence block, which satisfies $|[\mathbf{e}_t]_m|^2 = 1$ for $1 \le m \le M$. Define set $\mathcal{K} = \{1, \ldots, K\}$. Here, we assume that the users transmit the pilot sequences of length τ_k one by one for channel estimation. The received signal from user k at the BS after removing the impact of the direct channel at time slot t, $1 \le t \le \tau_k$, can be expressed as

$$\mathbf{y}_k(t) = \mathbf{H}\mathrm{Diag}(\mathbf{e}_t)\mathbf{h}_k\sqrt{p}s_k(t) + \mathbf{n}_k(t), \forall k \in \mathcal{K}, \tag{1}$$

where $s_k(t)$ and $\mathbf{n}_k(t) \in \mathbb{C}^{N\times 1} \sim \mathcal{CN}(0, \delta^2\mathbf{I})$ denote the transmitted pilot signal of the k-th user and the additive white Gaussian noise (AWGN) with noise power δ^2 at the BS at time slot t, respectively. p is the identical transmit power of each user.

For the convenience of optimization, (1) is rewritten as

$$\mathbf{y}_k(t) = \mathbf{H}\mathrm{Diag}(\mathbf{h}_k)\mathbf{e}_t\sqrt{p}s_k(t) + \mathbf{n}_k(t), \forall k \in \mathcal{K}. \tag{2}$$

In other words, the joint design of the active beamforming at the BS and the passive reflecting beamforming at the RIS only depends on the cascaded user-RIS-BS channels [6], [7]:

$$\mathbf{G}_k = \mathbf{H}\mathrm{Diag}(\mathbf{h}_k) \in \mathbb{C}^{N\times M}, \forall k \in \mathcal{K}. \tag{3}$$

Thus, this work mainly investigates the estimation of the cascaded channels in (3).

In particular, user k sends τ_k pilot signals to the BS. For simplicity, we assume that the pilot symbols satisfy $s_k(t) = 1, 1 \le t \le \tau_k$. The measurement matrix $\mathbf{Y}_k = [\mathbf{y}_k(1), \ldots, \mathbf{y}_k(\tau_k)] \in \mathbb{C}^{N\times\tau_k}$ received at the BS can be expressed as

$$\mathbf{Y}_k = \sqrt{p}\mathbf{G}_k\mathbf{E}_k + \mathbf{N}_k \in \mathbb{C}^{N\times\tau_k}, \tag{4}$$

where

$$\mathbf{E}_k = [\mathbf{e}_1, \ldots, \mathbf{e}_{\tau_k}] \in \mathbb{C}^{M \times \tau_k}, \tag{5a}$$

$$\mathbf{N}_k = [\mathbf{n}_k(1), \ldots, \mathbf{n}_k(\tau_k)] \in \mathbb{C}^{N \times \tau_k}. \tag{5b}$$

According to [8], the LS estimator

$$\mathbf{G}_k^{\mathrm{LS}} = \frac{1}{\sqrt{p}} \mathbf{Y}_k \mathbf{E}_k^{\mathrm{H}} (\mathbf{E}_k \mathbf{E}_k^{\mathrm{H}})^{-1} \tag{6}$$

of $\mathbf{G}_k$ is unbiased when the design of the phase shift matrix $\mathbf{E}_k$ is optimal. However, the required pilot overhead $\tau_k \geq M$ for each user is unacceptable due to the fact that the RIS is generally equipped with a large number of phase shift elements. Therefore, it motivates us to investigate an efficient channel estimation strategy to further reduce the pilot overhead by exploiting the sparsity of the mmWave massive MISO channel.

B. Cascaded Channel Sparsity Model

It is assumed that both BS and RIS are equipped with a uniform linear array (ULA) with antenna space d_{BS} and d_{RIS}, respectively. By applying the geometric channel model in the mmWave system [15], channels $\mathbf{H}$ and $\mathbf{h}_k$ are modeled as

$$\mathbf{H} = \sum_{l=1}^{L} \alpha_l \mathbf{a}_N(\psi_l) \mathbf{a}_M^{\mathrm{H}}(\omega_l), \tag{7}$$

$$\mathbf{h}_k = \sum_{j=1}^{J_k} \beta_{k,j} \mathbf{a}_M(\varphi_{k,j}), \forall k \in \mathcal{K}, \tag{8}$$

where L and J_k denote the number of the propagation paths between the BS and the RIS and that of the propagation paths between the RIS and user k, respectively. The complex gains of the l-th path in the BS-RIS channel and the j-th path in the RIS-user-k channel are represented by α_l and $\beta_{k,j}$, respectively. Denote by $\mathbf{a}_X(x) \in \mathbb{C}^{X \times 1}$ the array steering vector, i.e.,

$$\mathbf{a}_X(x) = [1, e^{-\mathrm{i}2\pi x}, \ldots, e^{-\mathrm{i}2\pi(X-1)x}]^{\mathrm{T}},$$

where $X \in \{M, N\}$ and $x \in \{\omega_l, \psi_l, \varphi_{k,j}\}$. $\omega_l = \frac{d_{\mathrm{RIS}}}{\lambda_c}\cos(\theta_l)$, $\psi_l = \frac{d_{\mathrm{BS}}}{\lambda_c}\cos(\phi_l)$, and $\varphi_{k,j} = \frac{d_{\mathrm{RIS}}}{\lambda_c}\cos(\vartheta_{k,j})$ are the directional cosine with θ_l and ϕ_l denoting the AoD and AoA of the l-th spatial path from RIS to BS, respectively, and $\vartheta_{k,j}$ as the AoA of the j-th spatial path from user k to RIS. λ_c is the carrier wavelength. It should be emphasized here that the channel gains α_l

and $\beta_{k,j}$ change at each channel coherence block, while the angle information vary much slowly than the channel gains, and generally keep invariant during tens of channel coherence blocks.

From (7) and (8), the geometric model of cascaded channels in (3) are formulated as

$$\mathbf{G}_k = \sum_{l=1}^{L} \sum_{j=1}^{J_k} \alpha_l \beta_{k,j} \mathbf{a}_N(\psi_l) \mathbf{a}_M^{\mathrm{H}}(\omega_l - \varphi_{k,j}), \forall k \in \mathcal{K}. \tag{9}$$

Note that $\mathbf{a}_M(\omega_l - \varphi_{k,j})$ is the steering vector of the jl-th cascaded subpath of user k, and the corresponding $\cos(\theta_l) - \cos(\vartheta_{k,j})$ is named as the cosine of cascaded AoD for the jl-th cascaded subpath from user k in the following for simplicity.

The channel model in (9) represents the low rank property and the spatial correlation characteristics of RIS-aided mmWave massive MISO system. Thus, CS-based sparse cascaded channel estimation methods are widely used based on the expression in (9) [11], [13], [14]. In particular, (9) is approximated by using the virtual angular domain (VAD) representation, i.e.,

$$\mathbf{G}_k = \mathbf{A}_R \mathbf{X}_k \mathbf{A}_T^{\mathrm{H}}, \tag{10}$$

where dictionary matrices $\{\mathbf{A}_R, \mathbf{A}_T\}$ can be drawn from the array steering vectors [11], [13] or from the DFT matrix [14]. Besides, $\mathbf{X}_k$ is the angular domain sparse cascaded channel matrix containing $J_k L$ complex channel gains, which exhibits the sparsity. The CS-based estimation methods in [11], [13], [14] need to estimate L AoAs, $J_k L$ cascaded AoDs, and $J_k L$ cascaded complex channel gains. The number of parameters to be estimated in [11], [13], [14] is much less than that in LS estimator in [8], since the number of spatial paths is usually much less than the number of antennas, i.e., $J_k L \ll N$ and $J_k L \ll M$. However, we can further reduce the number of parameters to be estimated by exploiting the structure of the cascaded channel.

Specifically, (7) is reformulated as

$$\mathbf{H} = \mathbf{A}_N \mathbf{\Lambda} \mathbf{A}_M^{\mathrm{H}}, \tag{11}$$

where

$$\mathbf{A}_N = [\mathbf{a}_N(\psi_1), \ldots, \mathbf{a}_N(\psi_L)] \in \mathbb{C}^{N \times L}, \tag{12a}$$

$$\mathbf{\Lambda} = \mathrm{Diag}(\alpha_1, \alpha_2, \ldots, \alpha_L) \in \mathbb{C}^{L \times L}, \tag{12b}$$

$$\mathbf{A}_M = [\mathbf{a}_M(\omega_1), \ldots, \mathbf{a}_M(\omega_L)] \in \mathbb{C}^{M \times L}. \tag{12c}$$

(8) is rewritten as

$$\mathbf{h}_k = \mathbf{A}_{M,k} \boldsymbol{\beta}_k, \forall k \in \mathcal{K}, \tag{13}$$

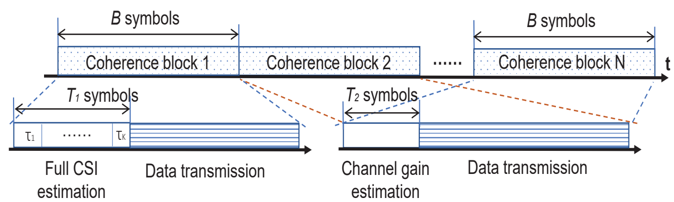

Fig. 1: Channel estimation protocol and frame structure.

where

$$\mathbf{A}_{M,k} = [\mathbf{a}_M(\varphi_{k,1}), \ldots, \mathbf{a}_M(\varphi_{k,J_k})] \in \mathbb{C}^{M\times J_k}, \tag{14a}$$

$$\boldsymbol{\beta}_k = [\beta_{k,1}, \ldots, \beta_{k,J_k}]^{\mathrm{T}} \in \mathbb{C}^{J_k\times 1}. \tag{14b}$$

Hence, (3) can be rewritten as

$$\mathbf{G}_k = \mathbf{A}_N \boldsymbol{\Lambda} \mathbf{A}_M^{\mathrm{H}} \mathrm{Diag}\left(\mathbf{A}_{M,k}\boldsymbol{\beta}_k\right), \forall k \in \mathcal{K}. \tag{15}$$

It is observed from (15) that there are actually only $J_k + L$ complex gains and $2L + J_k$ angles that need to be estimated for each user. In addition, due to the fact that all the users share the common BS-RIS channel $\mathbf{H}$, they share the same L complex gains $\{\alpha_l\}_{l=1}^{L}$ and $2L$ angles $\{\theta_l, \phi_l\}_{l=1}^{L}$. Based on this observation, we develop a novel channel estimation strategy in this work. We remark that the contributions in [13] and [14] only take advantage of the information from common angle $\{\phi_l\}_{l=1}^{L}$ and ignore the information from common gains $\{\alpha_l\}_{l=1}^{L}$ and common angles $\{\theta_l\}_{l=1}^{L}$.

III. Channel Estimation

A. Channel Estimation Protocol

In this section, we develop a novel uplink channel estimation protocol by exploiting the sparsity of the RIS-aided mmWave massive MISO channel, as shown in Fig. 1.

Generally, the BS, the RIS and users may not physically change their positions in a relatively long time duration. Based on this fact, we consider several or even tens of consecutive channel coherence blocks, where the angle information keeps invariant while the channel gains vary at each channel coherence block [15]. Then, at the first coherence block, we first estimate the full CSI information, including all the angle information and the channel gains. With the estimated

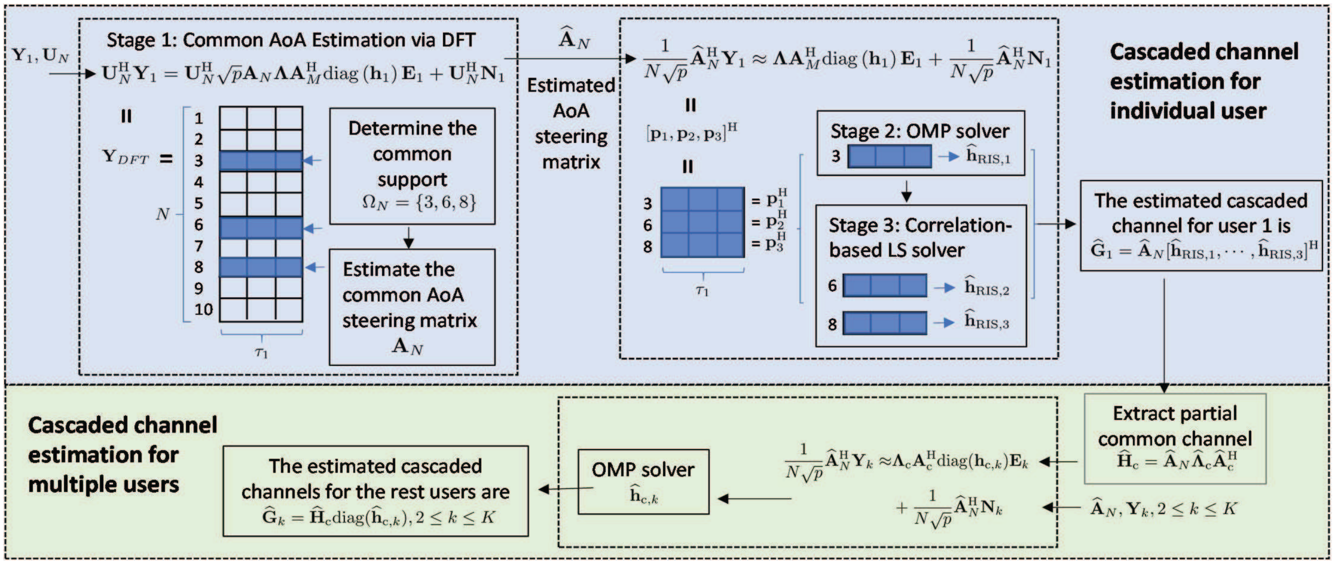

Fig. 2: Cascaded channel estimation strategy for multiple users.

angle information, we only need to estimate the channel gains in the remaining coherence blocks, which can be obtained by using simple LS method with much reduced amount of pilot symbols.

The most difficulty part is the full CSI estimation at the first channel coherence block. The main idea is explained as follows. First, one typical user, denoted by user 1 for convenience, sends a pilot sequence of τ_1 symbols to the BS for the channel estimation by using the DFT and CS techniques. With the knowledge of the estimated AoAs, cascaded AoDs, and cascaded gains of user 1, we construct a partial common BS-RIS channel with known CSI, which can be exploited to reduce the channel estimation overhead associated with user $k, 2 \leq k \leq K$. Then, the remaining users successively transmit $\tau_k, 2 \leq k \leq K$ pilot symbols to the BS for channel estimation. Note that the channel estimation at the first coherence block seems time consuming but will only be performed once at the start of the transmission.

B. Channel Estimation for User 1 at the First Coherence Block

In this subsection, we provide the channel estimation method for user 1 with low pilot overhead by exploiting the property of massive antenna array and the structure of the cascaded channel.

1) Estimation of the common AoAs: Due to the large number of antennas at the BS, the discrete Fourier transform (DFT) approach can be applied efficiently for AoA estimation from

$\mathbf{Y}_1$ in (4). We first present the asymptotic property of $\mathbf{A}_N$ in the following lemmas, which are proved in Appendix A and Appendix B.

Lemma 1 *When $N \to \infty$, the following property holds*

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{a}_N^{\mathrm{H}}(\psi_j)\mathbf{a}_N(\psi_i) = \begin{cases} 1 & \psi_j = \psi_i \\ 0 & \textit{otherwise} \end{cases}, \tag{16}$$

and $\mathbf{A}_N^{\mathrm{H}}\mathbf{A}_N = N\mathbf{I}_L$, where $\mathbf{I}_L$ is the identity matrix of dimension $L \times L$.

Lemma 2 *When $N \to \infty$, if the condition $\frac{d_{\mathrm{BS}}}{\lambda_c} \le 1$ holds, then the DFT of $\mathbf{A}_N$, i.e., $\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N$, is a tall sparse matrix with one nonzero element in each column*

$$\lim_{N\to\infty} [\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N]_{n_l,l} = \mathrm{nonzero}, \ \textit{for} \ \forall l,$$

where $\mathbf{U}_N$ is the normalized DFT matrix with its (n,m)-th entry $[\mathbf{U}_N]_{n,m} = \frac{1}{\sqrt{N}}e^{-\mathrm{i}\frac{2\pi}{N}(n-1)(m-1)}$, and

$$n_l = \begin{cases} N\psi_l + 1, & \psi_l \in [0, \frac{d_{\mathrm{BS}}}{\lambda_c}) \\ N + N\psi_l + 1, & \psi_l \in [-\frac{d_{\mathrm{BS}}}{\lambda_c}, 0) \end{cases}. \tag{17}$$

Based on Lemma 2, any two nonzero elements are not in the same row, i.e., $n_l \ne n_i$ for any $l \ne i$.

Remark 1: It is observed from (17) that when $\psi_l \in [0, \frac{d_{\mathrm{BS}}}{\lambda_c})$, the range of n_l is $n_l \in [1, N\frac{d_{\mathrm{BS}}}{\lambda_c} + 1)$. When $\psi_l \in [-\frac{d_{\mathrm{BS}}}{\lambda_c}, 0)$, we have $n_l \in [N - N\frac{d_{\mathrm{BS}}}{\lambda_c} + 1, N+1)$. In order to avoid ambiguous angle estimation, that is, the same n_l corresponds to two AoAs, we much have $N\frac{d_{\mathrm{BS}}}{\lambda_c} + 1 \le N - N\frac{d_{\mathrm{BS}}}{\lambda_c} + 1$, which leads to $d_{\mathrm{BS}} \le \frac{\lambda_c}{2}$. Therefore, d_{BS} should generally be restricted to be no larger than $\lambda_c/2$ to avoid the AoA ambiguity.

Based on Lemma 2, matrix $\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N$ can be regarded as a row sparse matrix with full column rank. Thus, the DFT of $\mathbf{Y}_1$, i.e., $\mathbf{Y}_{DFT} = \mathbf{U}_N^{\mathrm{H}}\mathbf{Y}_1 = \sqrt{p}\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{\Lambda}\mathbf{A}_M^{\mathrm{H}}\mathrm{Diag}\left(\mathbf{h}_1\right)\mathbf{E}_1 + \mathbf{U}_N^{\mathrm{H}}\mathbf{N}_1$, is an asymptotic row sparse matrix with L nonzero rows, each corresponding to one of the AoAs as shown in Fig. 2. Based on the above discussion, ϕ_l can be immediately estimated from the nonzero rows of $\mathbf{Y}_{DFT}$. However, N is finite in practice, thus $N\psi_l$ is usually not integer. Most power of $\mathbf{Y}_{DFT}$ will concentrate on the $(\lfloor N\psi_l \rceil + 1)$-th or the $(N + \lfloor N\psi_l \rceil + 1)$-th row, while the remaining power leaks to the nearby rows. This is known as the power leakage effect [17]–[19]. Due to the fact that the resolution of the DFT is $1/N$, there exists a mismatch between the discrete estimated angle and the real continuous angle. To improve the angle estimation accuracy, we adopt an angle rotation operation to compensate the mismatch of the DFT [17]–[19].

The angle rotation matrices are defined as

$$\mathbf{\Phi}_N(\triangle\psi_l) = \mathrm{Diag}\{1, e^{\mathrm{i}\triangle\psi_l}, \ldots, e^{\mathrm{i}(N-1)\triangle\psi_l}\}, \forall l,$$

where $\triangle\psi_l \in [-\frac{\pi}{N}, \frac{\pi}{N}]$ for $\forall l$ are the phase rotation parameters. Then, the angle rotation of $\mathbf{Y}_1$ for ϕ_l is defined as

$$\mathbf{Y}_{1,l}^{ro} = \mathbf{\Phi}_N^{\mathrm{H}}(\triangle\psi_l)\mathbf{Y}_1. \tag{18}$$

The aim of the angle rotation in (18) is to rotate $\mathbf{A}_N$ in the angle domain such that there is no power leakage for estimating ψ_l. For better illustration, we take $\mathbf{U}_N^{\mathrm{H}}\mathbf{\Phi}_N^{\mathrm{H}}(\triangle\psi_l)\mathbf{A}_N$ as an example, whose (n,l)-th element is calculated as

$$[\mathbf{U}_N^{\mathrm{H}}\mathbf{\Phi}_N^{\mathrm{H}}(\triangle\psi_l)\mathbf{A}_N]_{n,l} = \sqrt{\frac{1}{N}}\sum_{m=1}^{N} e^{-\mathrm{i}2\pi(m-1)(\psi_l + \frac{\triangle\psi_l}{2\pi} - \frac{n-1}{N})}.$$

It can be readily found that the channel power of ψ_l can concentrate on the n_l-th row without power leakage when the phase rotation parameter satisfies

$$\triangle\psi_l = 2\pi\left(\frac{n_l - 1}{N} - \psi_l\right). \tag{19}$$

For $\mathbf{Y}_1$, the optimal phase rotation parameter for ψ_l can be found based on a one-dimensional searching method by solving the following problem

$$\triangle\psi_l = \arg\max_{\triangle\psi\in[-\frac{\pi}{N},\frac{\pi}{N}]} ||[\mathbf{U}_N]_{:,n_l}^{\mathrm{H}}\mathbf{\Phi}_N^{\mathrm{H}}(\triangle\psi)\mathbf{Y}_1||^2. \tag{20}$$

Fig. 3 is an example of the row sparse characteristic of $\mathbf{Y}_{DFT}$ and the Y-axis is the power of each row of $\mathbf{Y}_{DFT}$. The considered cascaded channel of size $N = M = 100$ contains $L = 1$ paths between the BS and the RIS with $\phi = 14°$. It can be seen from the blue curve that although the beam covers several points because of power leakage, we can locate the power peak point of the beam, which can be utilized for initial AoA estimation. The orange curve demonstrates the effect of the optimal angle rotation for $\phi = 14°$. It is obvious that more power focuses on $\phi = 14°$, which makes the AoA estimation more accurate.

Algorithm 1 summarizes the common AoA estimation. After calculating the sum power of each row of $\mathbf{Y}_{DFT}$ in Step 2, we finds the set of row indexes with power peak in Step 3. $\Gamma(\mathbf{z})$ denotes the operation of finding the indexes with power peak in vector $\mathbf{z}$, $\Omega_N = \{n_l, l = 1, \cdots, \widehat{L}\}$ is a set to collect the indexes of the non-zero rows, and $\widehat{L}$ is the number of non-zero rows. We note

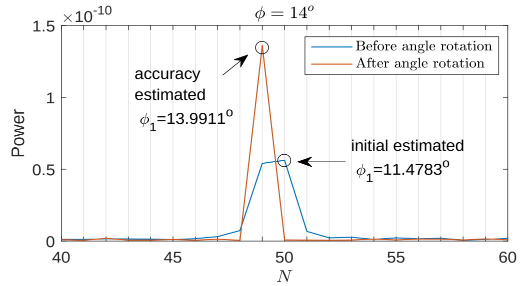

Fig. 3: An example of the row sparse characteristic of $\mathbf{Y}_{DFT}$ and optimal angle rotation, when $L = 1$ and $N = M = 100$.

Algorithm 1 Common AoA Estimation

Input: $\mathbf{Y}_1$.

1: Calculate DFT: $\mathbf{Y}_{DFT} = \mathbf{U}_N^{\mathrm{H}} \mathbf{Y}_1$;

2: Calculate the power of each row: $\mathbf{z}(n) = ||[\mathbf{Y}_{DFT}]_{n,:}||^2, \forall n = 1, 2, \ldots, N$;

3: Find the rows with the power peak: $(\Omega_N, \widehat{L}) = \Gamma(\mathbf{z})$, where $\Omega_N = \{n_l, l = 1, \cdots, \widehat{L}\}$;

4: Calculate the optimal angle rotation parameters $\{\triangle\widehat{\psi}_l\}_{l=1}^{\widehat{L}}$ via (20);

5: Estimate AOAs for $1 \le l \le \widehat{L}$:

$$\widehat{\phi}_l = \begin{cases} \arccos\left(\frac{\lambda_c(n_l-1)}{d_{\mathrm{BS}}N} - \frac{\lambda_c \triangle\psi_l}{2\pi d_{\mathrm{BS}}}\right), & n_l \le N\frac{d_{\mathrm{BS}}}{\lambda_c} \\ \arccos\left(\frac{\lambda_c(n_l-N-1)}{d_{\mathrm{BS}}N} - \frac{\lambda_c \triangle\psi_l}{2\pi d_{\mathrm{BS}}}\right), & n_l > N\frac{d_{\mathrm{BS}}}{\lambda_c} \end{cases}. \tag{21}$$

Output: $\{\widehat{\phi}_l\}_{l=1}^{\widehat{L}}$.

---

that $\widehat{L}$ is the estimated number of the propagation paths between the BS and the RIS, and also the estimated number of common AoAs. For each n_l, Problem (20) is solved to find the optimal angle rotation parameter in Step 4. Finally, the common AOAs are estimated in Step 5.

2) Estimation of the cascaded AoDs and gains: With the estimated AoAs $\{\widehat{\phi}_l\}_{l=1}^{\widehat{L}}$ from Algorithm 1, we obtain the estimated steering matrix $\widehat{\mathbf{A}}_N = [\mathbf{a}_N(\widehat{\psi}_1), \ldots, \mathbf{a}_N(\widehat{\psi}_{\widehat{L}})] \in \mathbb{C}^{N\times\widehat{L}}$. Based on the orthogonality of the massive steering matrix, i.e., $\widehat{\mathbf{A}}_N^{\mathrm{H}} \mathbf{A}_N \approx N\mathbf{I}_L$ due to Lemma 1, the measurement matrix $\mathbf{Y}_1$ can be projected into the common AoA steering matrix subspace as

$$\frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^{\mathrm{H}}\mathbf{Y}_1 \approx \mathbf{\Lambda}\mathbf{A}_M^{\mathrm{H}}\mathrm{Diag}\left(\mathbf{h}_1\right)\mathbf{E}_1 + \frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^{\mathrm{H}}\mathbf{N}_1 = \mathbf{H}_{\mathrm{RIS}}^{\mathrm{H}}\mathbf{E}_1 + \frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^{\mathrm{H}}\mathbf{N}_1, \tag{22}$$

where $\mathbf{H}_{\text{RIS}} = \text{Diag}\left(\mathbf{h}_1^*\right)\mathbf{A}_M\mathbf{\Lambda}^{\text{H}}$. Based on (12b) and (12c), the l-th column of $\mathbf{H}_{\text{RIS}}$ is given by

$$\mathbf{h}_{\text{RIS},l} = \text{Diag}\{\mathbf{h}_1^*\}\mathbf{a}_M(\omega_l)\alpha_l^*, \tag{23}$$

where $\mathbf{H}_{\text{RIS}} = [\mathbf{h}_{\text{RIS},1}, \cdots, \mathbf{h}_{\text{RIS},L}]$. We claim that $\mathbf{h}_{\text{RIS},l}$ can be estimated by transforming each row of (22) into a sparse signal recovery problem. In particular, define $\frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^{\text{H}}\mathbf{Y}_1 = [\mathbf{p}_1, \ldots, \mathbf{p}_L]^{\text{H}}$, where

$$\begin{aligned}
\mathbf{p}_l &= \mathbf{E}_1^{\text{H}}\mathbf{h}_{\text{RIS},l} + \mathbf{n}_{\text{noise}} \in \mathbb{C}^{\tau_1\times 1} \\
&= \mathbf{E}_1^{\text{H}}\text{Diag}\{\mathbf{a}_M(\omega_l)\}\mathbf{h}_1^*\alpha_l^* + \mathbf{n}_{\text{noise}} \\
&= \mathbf{E}_1^{\text{H}}\text{Diag}\{\mathbf{a}_M(\omega_l)\}\mathbf{A}_{M,1}^*\boldsymbol{\beta}_1^*\alpha_l^* + \mathbf{n}_{\text{noise}} \\
&= \mathbf{E}_1^{\text{H}}\begin{bmatrix} \mathbf{a}_M\left(\omega_l - \varphi_{1,1}\right) & \cdots & \mathbf{a}_M\left(\omega_l - \varphi_{1,J_1}\right) \end{bmatrix}\boldsymbol{\beta}_1^*\alpha_l^* + \mathbf{n}_{\text{noise}}
\end{aligned} \tag{24}$$

with $\mathbf{n}_{\text{noise}}$ representing the corresponding noise vector. To extract the cascaded directional cosine $\{\omega_l - \varphi_{1,j}\}_{j=1}^{J_1}$ and gains $\boldsymbol{\beta}_1^*\alpha_l^*$ from $\mathbf{p}_l$, (24) can be approximated by using the VAD representation as

$$\mathbf{p}_l = \mathbf{E}_1^{\text{H}}\mathbf{A}\mathbf{b}_l + \mathbf{n}_{\text{noise}}, \tag{25}$$

where $\mathbf{A} \in \mathbb{C}^{M\times D}(M \ll D)$ is an overcomplete dictionary matrix, each column of which represents the array steering vector of possible values of $\omega_l - \varphi_{1,j}$. Since $\omega_l - \varphi_{1,j} \in [-2\frac{d_{\text{RIS}}}{\lambda_c}, 2\frac{d_{\text{RIS}}}{\lambda_c}]$, $\mathbf{A}$ can be constructed as

$$\mathbf{A} = \left[\mathbf{a}_M(-2\frac{d_{\text{RIS}}}{\lambda_c}), \mathbf{a}_M((-2+\frac{4}{D})\frac{d_{\text{RIS}}}{\lambda_c}), \ldots, \mathbf{a}_M((2-\frac{4}{D})\frac{d_{\text{RIS}}}{\lambda_c})\right]. \tag{26}$$

Recall that $\boldsymbol{\beta}_1 = [\beta_{1,1}, \ldots, \beta_{1,J_1}]^{\text{T}}$ in (14b), $\mathbf{b}_l \in \mathbb{C}^{D\times 1}$ is then a sparse vector with J_1 cascaded gains $\{\alpha_l^*\beta_{1,j}^*\}_{j=1}^{J_1}$ as the nonzero elements. (25) is a sparse signal recovery problem that can be solved by using the CS techniques, such as OMP. Note that the phase shift matrix $\mathbf{E}_1$ in (25) will be designed for better estimation in Section IV. It has been proved that $\tau_1 \geq 8J_1 - 2$ measurements are sufficient to recover a J_1-sparse complex-valued signal vector [20].

However, if OMP is used L times for solving $\mathbf{p}_l(1 \leq l \leq L)$, we need to estimate $J_1 L$ independent sparse variables with high complexity. In order to reduce the complexity, we exploit the following scaling property. Specifically, we observe from (23) that there are angle and gain scaling between the cascaded multipaths formed by different AoDs $\{\omega_l\}_{l=1}^{L}$ on the RIS. That is, there is the following relationship between $\mathbf{h}_{\text{RIS},l}$ and $\mathbf{h}_{\text{RIS},r}$ for $1 \leq l, r \leq L$:

$$\mathbf{h}_{\text{RIS},l} = \text{Diag}\{\mathbf{a}_M(\omega_l - \omega_r)\}\text{Diag}\{\mathbf{h}_1^*\}\mathbf{a}_M(\omega_r)\alpha_r^*\frac{\alpha_l^*}{\alpha_r^*}$$

$$= \mathrm{Diag}\{\mathbf{a}_M(\omega_l - \omega_r)\}\mathbf{h}_{\mathrm{RIS},r}\frac{\alpha_l^*}{\alpha_r^*}. \tag{27}$$

(27) is called the angle-gain scaling property, which implies that $\mathbf{h}_{\mathrm{RIS},l}$ for all l can be represented by one arbitrary $\mathbf{h}_{\mathrm{RIS},r}$. Let

$$\triangle\omega_l = \omega_l - \omega_r, \tag{28a}$$

$$x_l = \frac{\alpha_l^*}{\alpha_r^*}. \tag{28b}$$

(27) is then re-expressed as $\mathbf{h}_{\mathrm{RIS},l} = \mathrm{Diag}\{\mathbf{h}_{\mathrm{RIS},r}\}\mathbf{a}_M(\triangle\omega_l)x_l$. Denote the estimation of $\mathbf{h}_{\mathrm{RIS},r}$ as $\widehat{\mathbf{h}}_{\mathrm{RIS},r}$ obtained from (25) by using OMP. Further defining $\mathbf{z}_l(\triangle\omega_l) = \mathbf{E}_1^{\mathrm{H}}\mathrm{Diag}\{\widehat{\mathbf{h}}_{\mathrm{RIS},r}\}\mathbf{a}_M(\triangle\omega_l)$, (24) can be rewritten as

$$\mathbf{p}_l = \mathbf{z}_l(\triangle\omega_l)x_l + \mathbf{n}_{\mathrm{noise}}. \tag{29}$$

It is observed from (29) that only two variables $\triangle\omega_l$ and x_l need to be estimated. Since $\triangle\omega_l \in [-2\frac{d_{\mathrm{RIS}}}{\lambda_c}, 2\frac{d_{\mathrm{RIS}}}{\lambda_c}]$, $\triangle\omega_l$ can then be estimated via a simple correlation-based scheme

$$\triangle\widehat{\omega}_l = \arg \max_{\triangle\omega\in[-2\frac{d_{\mathrm{RIS}}}{\lambda_c}, 2\frac{d_{\mathrm{RIS}}}{\lambda_c}]} \left|\langle \mathbf{p}_l, \mathbf{z}_l(\triangle\omega)\rangle\right|. \tag{30}$$

x_l is the solution of a LS problem $\min_x||\mathbf{p}_l - \mathbf{z}_l(\triangle\widehat{\omega}_l)x||_2$, that is

$$\widehat{x}_l = (\mathbf{z}_l^{\mathrm{H}}(\triangle\widehat{\omega}_l)\mathbf{z}_l(\triangle\widehat{\omega}_l))^{-1}\mathbf{z}_l^{\mathrm{H}}(\triangle\widehat{\omega}_l)\mathbf{p}_l. \tag{31}$$

Let $\widehat{\mathbf{h}}_{\mathrm{RIS},l} = \mathrm{Diag}\{\widehat{\mathbf{h}}_{\mathrm{RIS},r}\}\mathbf{a}_M(\triangle\widehat{\omega}_l)\widehat{x}_l, (1 \le l \le L, l \ne r)$, the final estimated cascaded channel of user 1 is given by

$$\widehat{\mathbf{G}}_1 = \widehat{\mathbf{A}}_N\widehat{\mathbf{H}}_{\mathrm{RIS}}^{\mathrm{H}}, \tag{32}$$

where $\widehat{\mathbf{H}}_{\mathrm{RIS}} = [\widehat{\mathbf{h}}_{\mathrm{RIS},1}, \cdots, \widehat{\mathbf{h}}_{\mathrm{RIS},L}]$.

Algorithm 2 summarizes the complete estimation of $\mathbf{G}_1$. The common AoA steering matrix $\mathbf{A}_N$ is estimated by using the DFT and the angle rotation techniques in Stage 1. In Stage 2 containing Step 3 to Step 12, OMP is used to estimate $\mathbf{h}_{\mathrm{RIS},r}$. Here, r is determined according to Problem (35) such that the SNR of $\mathbf{p}_r$ is the maximum value among $\{\mathbf{p}_l\}_{l=1}^L$ when they have the same noise power, which is to ensure the better estimation accuracy of the OMP method. The rest $\mathbf{h}_{\mathrm{RIS},l}$ ($1 \le l \le \widehat{L}$ and $l \ne r$) are estimated by using the simple LS method and correlation-based scheme in Stage 3 containing Step 13 to Step 16. Finally, we obtain the estimation $\widehat{\mathbf{G}}_1 = \widehat{\mathbf{A}}_N[\widehat{\mathbf{h}}_{\mathrm{RIS},1}, \cdots, \widehat{\mathbf{h}}_{\mathrm{RIS},\widehat{L}}]^{\mathrm{H}}$. The flow chart of Algorithm 2 is shown in Fig. 2.

We would like to emphasize that the cascaded AoDs and cascaded gains can also been obtained in Algorithm 2, which facilitates the cascaded channel estimation of other users in the next subsection. In particular, the cascaded AoDs and cascaded gains from $\widehat{\mathbf{h}}_{\text{RIS},r}$ in Step 12 are given by

$$[\mathbf{a}_M(\widehat{\omega_r - \varphi_{1,1}}) \cdots \mathbf{a}_M(\widehat{\omega_r - \varphi_{1,\widehat{J}_1}})] = \mathbf{A}_{(:,\Omega_{i-1})}, \tag{33a}$$

$$\widehat{\boldsymbol{\beta}_1^* \alpha_r^*} = \mathbf{b}_{i-1}. \tag{33b}$$

Based on (28) and (33), the cascaded AoDs and cascaded gains from $\widehat{\mathbf{h}}_{\text{RIS},l}$ ($1 \le l \le \widehat{L}$ and $l \neq r$) in Step 16 are given by

$$[\mathbf{a}_M(\widehat{\omega_l - \varphi_{1,1}}) \cdots \mathbf{a}_M(\widehat{\omega_l - \varphi_{1,\widehat{J}_1}})] = \text{Diag}\{\mathbf{a}_M(\triangle\widehat{\omega}_l)\}\mathbf{A}_{(:,\Omega_{i-1})}, \tag{34a}$$

$$\widehat{\boldsymbol{\beta}_1^* \alpha_l^*} = \widehat{\boldsymbol{\beta}_1^* \alpha_r^*} \widehat{x}_l. \tag{34b}$$

Algorithm 2 estimates L AoAs in Stage 1, J_1 cascaded AoDs and J_1 cascaded gains in (33), and $2L-2$ scaling parameters in Step 14 and Step 15. Therefore, Algorithm 2 uses a total of $\tau_1 \geq 8J_1 - 2$ time slots to estimate $3L + 2J_1 - 2$ parameters to recover channel $\mathbf{G}_1$ of dimension $N \times M$. Note that the number of time slots required is not related to L, which evidences the advantage of our proposed estimation method.

C. Channel estimation for Other Users at the First Coherence Block

Algorithm 2 can also be used for the channel estimation of the other users, where Stage 1 can be omitted because all users share the common AoA steering matrix $\widehat{\mathbf{A}}_N$. In addition to $\widehat{\mathbf{A}}_N$, all users also share the common matrices $\boldsymbol{\Lambda}$ and $\mathbf{A}_M$ in their channel matrices $\mathbf{G}_k, \forall k$. Note that the cascaded channel $\mathbf{G}_k = \mathbf{H}\text{Diag}(\mathbf{h}_k), 2 \le k \le K$ in (3), we except that if the common channel $\mathbf{H}$ is known, channel $\mathbf{h}_k$ can be readily estimated from a sparse signal recovery problem. However, it is intractable to obtain $\mathbf{H}$ from the estimated cascaded channel $\widehat{\mathbf{G}}_1$ due to the coupling of angles $\cos(\theta_l) - \cos(\vartheta_{1,j})$ and channel gains $\alpha_l \beta_{1,j}$ with each cascaded subpath of user 1. However, we can construct a substitute for $\mathbf{H}$ (denoted by $\mathbf{H}_\text{c}$) by only using $\widehat{\mathbf{G}}_1$. The $\mathbf{H}_\text{c}$ contains partial information of $\mathbf{H}$. Then, (3) can be rewritten as

$$\mathbf{G}_k = \mathbf{H}_\text{c}\text{Diag}(\mathbf{h}_{\text{c},k}), 2 \le k \le K, \tag{38}$$

where $\mathbf{h}_{\text{c},k}$ is the corresponding partial CSI of $\mathbf{h}_k$. In the following, we first construct $\mathbf{H}_\text{c}$ based on the estimated channel information from Algorithm 2 and then estimate the partial channel information $\mathbf{h}_{\text{c},k}$.

Algorithm 2 DFT-OMP-based Estimation of $\mathbf{G}_1$

Input: $\mathbf{Y}_1$, $\mathbf{A}$.

Stage 1: Return estimated common AoA steering matrix $\widehat{\mathbf{A}}_N$ and $\widehat{L}$ by using Algorithm 1.

Calculate $[\mathbf{p}_1, \ldots, \mathbf{p}_{\widehat{L}}] = \frac{1}{N\sqrt{p}}\mathbf{Y}_1^{\mathrm{H}}\widehat{\mathbf{A}}_N$.

Stage 2: Estimate $\mathbf{h}_{\mathrm{RIS},r}$ from $\mathbf{p}_r$ by using the OMP algorithm, where r is determined according to

$$r = \arg\max_{1\leq i\leq \widehat{L}} ||\mathbf{p}_i||^2. \tag{35}$$

Calculate equivalent dictionary $\mathbf{D} = \mathbf{E}_1^{\mathrm{H}}\mathbf{A}$.

Initialize $\Omega_0 = \emptyset$, $\mathbf{r}_0 = \mathbf{p}_r$, $i = 1$.

repeat

    $d_i = \arg\max_{d=1,2,\ldots,D}|\mathbf{D}_{(:,d)}^{\mathrm{H}}\mathbf{r}_{i-1}|$.

    $\Omega_i = \Omega_{i-1} \cup d_i$.

    LS solution: $\mathbf{b}_i = (\mathbf{D}_{(:,\Omega_i)}^{\mathrm{H}}\mathbf{D}_{(:,\Omega_i)})^{-1}\mathbf{D}_{(:,\Omega_i)}^{\mathrm{H}}\mathbf{p}_r$.

    $\mathbf{r}_i = \mathbf{p}_r - \mathbf{D}_{(:,\Omega_i)}\mathbf{b}_i$.

    $i = i + 1$.

until $||\mathbf{r}_{i-1}||_2 \leq$threshold.

Obtain the estimations:

$$\widehat{J}_1 = i - 1, \tag{36a}$$

$$\widehat{\mathbf{h}}_{\mathrm{RIS},r} = \mathbf{A}_{(:,\Omega_{i-1})}\mathbf{b}_{i-1}. \tag{36b}$$

Stage 3: Estimate $\mathbf{h}_{\mathrm{RIS},l}$ from $\mathbf{p}_l$ for $1 \leq l \leq \widehat{L}$ and $l \neq r$:

Calculate $\triangle\widehat{\omega}_l$ according to (30).

Calculate $\widehat{x}_l$ according to (31).

Obtain the estimations for $1 \leq l \leq \widehat{L}$ and $l \neq r$:

$$\widehat{\mathbf{h}}_{\mathrm{RIS},l} = \mathrm{Diag}\{\widehat{\mathbf{h}}_{\mathrm{RIS},r}\}\mathbf{a}_M(\triangle\widehat{\omega}_l)\widehat{x}_l. \tag{37}$$

Output: $\widehat{\mathbf{G}}_1 = \widehat{\mathbf{A}}_N[\widehat{\mathbf{h}}_{\mathrm{RIS},1}, \cdots, \widehat{\mathbf{h}}_{\mathrm{RIS},\widehat{L}}]^{\mathrm{H}}$.

1) Construction of $\mathbf{H}_{\rm c}$*:* In the following, we show how to construct $\mathbf{H}_{\rm c}$ by exploiting the structure of $\widehat{\mathbf{G}}_1$. In particular, (11) is reformulated as

$$\begin{aligned}\mathbf{H} = \mathbf{A}_N \boldsymbol{\Lambda} \mathbf{A}_M^{\rm H} &= \mathbf{A}_N \frac{1}{\overline{\beta}} \boldsymbol{\Lambda}_{\rm c} \mathbf{A}_{\rm c}^{\rm H} {\rm Diag}(\mathbf{a}_M(\overline{\varphi})) \\ &= \frac{1}{\overline{\beta}} \mathbf{H}_{\rm c} {\rm Diag}\left(\mathbf{a}_M(\overline{\varphi})\right), \end{aligned} \tag{39}$$

with

$$\mathbf{H}_{\rm c} = \mathbf{A}_N \boldsymbol{\Lambda}_{\rm c} \mathbf{A}_{\rm c}^{\rm H}, \tag{40a}$$

$$\boldsymbol{\Lambda}_{\rm c} = \overline{\beta} \boldsymbol{\Lambda}, \tag{40b}$$

$$\mathbf{A}_{\rm c} = {\rm Diag}(\mathbf{a}_M(\overline{\varphi})) \mathbf{A}_M, \tag{40c}$$

$$\overline{\varphi} = -\frac{1}{J_1} \sum_{j=1}^{J_1} \varphi_{1,j}, \tag{40d}$$

$$\overline{\beta} = \frac{1}{J_1} \mathbf{1}_{J_1}^{\rm T} \boldsymbol{\beta}_1, \tag{40e}$$

where $\mathbf{1}_{J_1}$ is an all-one vector with dimension of $J_1 \times 1$, and $\boldsymbol{\beta}_1$ is defined in (14b).

By using (12b), (28b) and (40e), $\boldsymbol{\Lambda}_{\rm c}$ in (40b) can be re-expressed as

$$\begin{aligned}\boldsymbol{\Lambda}_{\rm c} = \overline{\beta} \boldsymbol{\Lambda} &= \overline{\beta} {\rm Diag}(\alpha_1, \ldots, \alpha_L) \\ &= \left[\overline{\beta}^* \alpha_r^* {\rm Diag}(x_1, \ldots, x_L) \right]^* \quad (41a) \\ &= \left[\frac{1}{J_1} \mathbf{1}_{J_1}^{\rm T} \boldsymbol{\beta}_1^* \alpha_r^* {\rm Diag}(x_1, \ldots, x_L) \right]^*, \quad (41b)\end{aligned}$$

where the estimation of $\boldsymbol{\beta}_1^* \alpha_r^*$ is given in (33b), and the estimation of $[x_1, x_2, \ldots, x_L]$ is given in (31). Then, the estimation of $\boldsymbol{\Lambda}_{\rm c}$ is obtained as

$$\widehat{\boldsymbol{\Lambda}}_{\rm c} = \left[\frac{1}{J_1} \mathbf{1}_{J_1}^{\rm T} \widehat{\boldsymbol{\beta}_1^* \alpha_r^*} {\rm Diag}([\widehat{x}_1, \ldots, \widehat{x}_{\widehat{L}}]) \right]^*. \tag{42}$$

For $\mathbf{A}_{\rm c}$, by substituting (12c), (28a) and (40d) into (40c), we have

$$\begin{aligned}\mathbf{A}_{\rm c} &= {\rm Diag}(\mathbf{a}_M(\overline{\varphi})) \mathbf{A}_M \\ &= {\rm Diag}(\mathbf{a}_M(\overline{\varphi})) [\mathbf{a}_M(\omega_1), \ldots, \mathbf{a}_M(\omega_L)] \\ &= {\rm Diag}(\mathbf{a}_M(\omega_r + \overline{\varphi})) [\mathbf{a}_M(\triangle\omega_1), \ldots, \mathbf{a}_M(\triangle\omega_L)] \\ &= {\rm Diag}(\mathbf{a}_M(\omega_r - \frac{1}{J_1} \sum_{j=1}^{J_1} \varphi_{1,j})) [\mathbf{a}_M(\triangle\omega_1), \ldots, \mathbf{a}_M(\triangle\omega_L)]\end{aligned}$$

$$= \mathrm{Diag}(\mathbf{a}_M(\frac{1}{J_1}\sum_{j=1}^{J_1}(\omega_r - \varphi_{1,j})))[\mathbf{a}_M(\triangle\omega_1), \ldots, \mathbf{a}_M(\triangle\omega_L)],$$

where the estimation of $\{\omega_r - \varphi_{1,j}\}_{j=1}^{J_1}$ and $\{\triangle\omega_l\}_{l=1}^{L}$ are given by (30) and (33a), respectively. Then, we can obtain the estimation of $\mathbf{A}_\mathrm{c}$ as

$$\begin{aligned}\widehat{\mathbf{A}}_\mathrm{c} =& \mathrm{Diag}(\mathbf{a}_M(\frac{1}{\widehat{J}_1}\sum_{j=1}^{\widehat{J}_1}(\widehat{\omega_r - \varphi_{1,j}})))[\mathbf{a}_M((\triangle\widehat{\omega}_1), \cdots \mathbf{a}_M(\triangle\widehat{\omega}_{\widehat{L}})] \\ =& \mathrm{Diag}(\mathbf{a}_M(\widehat{\omega_r + \overline{\varphi}}))[\mathbf{a}_M((\triangle\widehat{\omega}_1), \cdots \mathbf{a}_M(\triangle\widehat{\omega}_{\widehat{L}})]. \end{aligned} \tag{44}$$

With $\widehat{\mathbf{A}}_N$, (42) and (44), the estimation of $\mathbf{H}_\mathrm{c}$ is given by $\widehat{\mathbf{H}}_\mathrm{c} = \widehat{\mathbf{A}}_N\widehat{\mathbf{\Lambda}}_\mathrm{c}\widehat{\mathbf{A}}_\mathrm{c}^\mathrm{H}$.

2) Estimation of partial CSI $\mathbf{h}_{\mathrm{c},k}$*:* In this subsection, we discuss how to use the partial common channel $\mathbf{H}_\mathrm{c}$ to help the channel estimation of other users with low pilot overhead. In particular, by substituting $\mathbf{H} = \frac{1}{\beta}\mathbf{H}_\mathrm{c}\mathrm{Diag}(\mathbf{a}_M(\overline{\varphi}))$ in (39) into (3), we have

$$\begin{aligned}\mathbf{G}_k = \mathbf{H}\mathrm{Diag}(\mathbf{h}_k) &= \frac{1}{\beta}\mathbf{H}_\mathrm{c}\mathrm{Diag}\left(\mathbf{a}_M(\overline{\varphi})\right)\mathrm{Diag}(\mathbf{h}_k) \\ &= \frac{1}{\beta}\mathbf{H}_\mathrm{c}\mathrm{Diag}\left(\mathrm{Diag}\left(\mathbf{a}_M(\overline{\varphi})\right)\mathbf{h}_k\right) = \mathbf{H}_\mathrm{c}\mathrm{Diag}(\mathbf{h}_{\mathrm{c},k}),\end{aligned} \tag{45}$$

where

$$\mathbf{h}_{\mathrm{c},k} = \frac{1}{\beta}\mathrm{Diag}\left(\mathbf{a}_M(\overline{\varphi})\right)\mathbf{h}_k \tag{46}$$

contains partial CSI of $\mathbf{h}_k$ and needs to be estimated in this subsection.

Similar to (22), $\mathbf{Y}_k$ in (4) is firstly projected into the common AoA steering matrix subspace as

$$\begin{aligned}\frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^\mathrm{H}\mathbf{Y}_k &= \frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^\mathrm{H}(\sqrt{p}\mathbf{H}_\mathrm{c}\mathrm{Diag}(\mathbf{h}_{\mathrm{c},k})\mathbf{E}_k + \mathbf{N}_k) \\ &\approx \mathbf{\Lambda}_\mathrm{c}\mathbf{A}_\mathrm{c}^\mathrm{H}\mathrm{Diag}(\mathbf{h}_{\mathrm{c},k})\mathbf{E}_k + \frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^\mathrm{H}\mathbf{N}_k.\end{aligned} \tag{47}$$

Recall that $\mathbf{E}_k = [\mathbf{e}_1, \ldots, \mathbf{e}_{\tau_k}]$ in (5a). By vectoring (47) and defining $\mathbf{z}_k = \mathrm{vec}(\frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^\mathrm{H}\mathbf{Y}_k) \in \mathbb{C}^{\tau_k L\times 1}$, we have

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{\Lambda}_\mathrm{c}\mathbf{A}_\mathrm{c}^\mathrm{H}\mathrm{Diag}(\mathbf{h}_{\mathrm{c},k})\mathbf{e}_1 \\ \vdots \\ \mathbf{\Lambda}_\mathrm{c}\mathbf{A}_\mathrm{c}^\mathrm{H}\mathrm{Diag}(\mathbf{h}_{\mathrm{c},k})\mathbf{e}_{\tau_k} \end{bmatrix} + \mathbf{n}_\mathrm{noise} = \mathbf{Z}_k\mathbf{h}_{\mathrm{c},k} + \mathbf{n}_\mathrm{noise}, \tag{48}$$

where $\mathbf{n}_{\text{noise}}$ representing the corresponding noise and

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{\Lambda}_{\text{c}}\mathbf{A}_{\text{c}}^{\text{H}}\text{Diag}\left(\mathbf{e}_1\right) \\ \vdots \\ \mathbf{\Lambda}_{\text{c}}\mathbf{A}_{\text{c}}^{\text{H}}\text{Diag}\left(\mathbf{e}_{\tau_k}\right) \end{bmatrix}. \tag{49}$$

By replacing $\mathbf{h}_k$ with $\mathbf{h}_k = \mathbf{A}_{M,k}\boldsymbol{\beta}_k$ from (13), $\mathbf{h}_{\text{c},k}$ in (46) can be unfolded as

$$\mathbf{h}_{\text{c},k} = \frac{1}{\beta}\text{Diag}\left(\mathbf{a}_M(\overline{\varphi})\right)\mathbf{h}_k = \frac{1}{\beta}[\mathbf{a}_M\left(\varphi_{k,1}+\overline{\varphi}\right)\cdots\mathbf{a}_M\left(\varphi_{k,J_k}+\overline{\varphi}\right)]\boldsymbol{\beta}_k. \tag{50}$$

Since $\varphi_{k,1}+\overline{\varphi} \in [-2\frac{d_{\text{RIS}}}{\lambda_c}, 2\frac{d_{\text{RIS}}}{\lambda_c}]$, (50) can be further approximated by using the VAD representation as

$$\mathbf{h}_{\text{c},k} = \mathbf{A}\mathbf{c}_k, \tag{51}$$

where $\mathbf{A}$ is defined in (26), and $\mathbf{c}_k \in \mathbb{C}^{G\times 1}$ is a sparse vector with J_k gains $\{\frac{1}{\beta}\beta_{k,j}\}_{j=1}^{J_k}$ as the nonzero elements.

With (51), (48) can be approximated as a sparse signal recovery problem

$$\mathbf{z}_k = \mathbf{Z}_k\mathbf{A}\mathbf{c}_k + \mathbf{n}_{\text{noise}}. \tag{52}$$

Note that $\mathbf{Z}_k$ is determined by using (42) and (44). Hence, Problem (52) could be solved by using the CS techniques, such as OMP. Note that the phase shift vectors $\{\mathbf{e}_t\}_{t=1}^{\tau_k}$ in $\mathbf{Z}_k$ will be designed to achieve high estimation accuracy in Section IV.

Algorithm 3 summarizes the OMP-based estimation of $\mathbf{G}_k, 2 \leq k \leq K$. To effectively recover the signal $\mathbf{c}_k$ of J_k-sparse level, the dimension of $\mathbf{z}_k \in \mathbb{C}^{\tau_k L\times 1}$ should satisfy the requirement $\tau_k L \geq 8J_k - 2$ [20]. Thus, the pilot overhead required by user k is $\tau_k \geq (8J_k-2)/L$.

We would like to highlight that the cascaded AoDs can also been obtained after estimation $[\mathbf{a}_M(\widehat{\varphi_{k,1}+\overline{\varphi}})\cdots\mathbf{a}_M(\widehat{\varphi_{k,J_k}+\overline{\varphi}})] = \mathbf{A}_{(:,\Omega_{i-1})}$ from (50) is determined when using OMP, which facilitates the cascaded channel estimation in the subsequent channel coherene blocks in the next subsection. In particular, the cascaded AoDs of user k for $2 \leq k \leq K$ and $1 \leq l \leq L$ are given by

$$\begin{aligned}[\mathbf{a}_M(\widehat{\omega_l-\varphi_{k,1}})\cdots\mathbf{a}_M(\widehat{\omega_l-\varphi_{k,J_k}})] &= \text{Diag}(\mathbf{a}_M(\widehat{\omega_l+\overline{\varphi}}))[\mathbf{a}_M^*(\widehat{\varphi_{k,1}+\overline{\varphi}})\cdots\mathbf{a}_M^*(\widehat{\varphi_{k,J_k}+\overline{\varphi}})] \\ &= \text{Diag}(\widehat{\mathbf{A}}_{\text{c}(:,l)})[\mathbf{a}_M^*(\widehat{\varphi_{k,1}+\overline{\varphi}})\cdots\mathbf{a}_M^*(\widehat{\varphi_{k,J_k}+\overline{\varphi}})],\end{aligned} \tag{54}$$

where $\widehat{\mathbf{A}}_{\text{c}(:,l)}$ is given in (44).

Algorithm 3 Estimation of $\mathbf{G}_k, 2 \le k \le K$

Input: $\mathbf{A}$, $\mathbf{Y}_k, 2 \le k \le K$.

Return $\widehat{\mathbf{A}}_N$ from Algorithm 1.
Construct $\widehat{\mathbf{\Lambda}}_{\mathrm{c}}$ according to (42).
Construct $\widehat{\mathbf{A}}_{\mathrm{c}}$ according to (44).
Calculate equivalent dictionary $\mathbf{R} = \mathbf{Z}_k\mathbf{A}$ according to (49) with $\widehat{\mathbf{\Lambda}}_{\mathrm{c}}$ and $\widehat{\mathbf{A}}_{\mathrm{c}}$.
for $2 \le k \le K$ **do**
    Calculate $\mathbf{z}_k = \mathrm{vec}(\frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^{\mathrm{H}}\mathbf{Y}_k)$.
    Initialize $\Omega_0 = \emptyset$, $\mathbf{r}_0 = \mathbf{z}_k$, $i = 1$.
    repeat
        $d_i = \arg\max_{d=1,2,\ldots,D} |\mathbf{R}^{\mathrm{H}}_{(:,d)}\mathbf{r}_{i-1}|$.
        $\Omega_i = \Omega_{i-1} \cup d_i$.
        LS solution: $\mathbf{b}_i = (\mathbf{R}^{\mathrm{H}}_{(:,\Omega_i)}\mathbf{R}_{(:,\Omega_i)})^{-1}\mathbf{R}^{\mathrm{H}}_{(:,\Omega_i)}\mathbf{p}_r$.
        $\mathbf{r}_i = \mathbf{p}_r - \mathbf{R}_{(:,\Omega_i)}\mathbf{b}_i$.
        $i = i + 1$.
    until $||\mathbf{r}_{i-1}||_2 \le$threshold.
    Calculate the estimated partial common channel $\widehat{\mathbf{H}}_{\mathrm{c}} = \widehat{\mathbf{A}}_N\widehat{\mathbf{\Lambda}}_{\mathrm{c}}\widehat{\mathbf{A}}_{\mathrm{c}}^{\mathrm{H}}$.
    Obtain the estimations:

$$\widehat{\mathbf{h}}_{\mathrm{c},k} = \mathbf{A}_{(:,\Omega_{i-1})}\mathbf{b}_{i-1}, \tag{53a}$$

$$\widehat{\mathbf{G}}_k = \widehat{\mathbf{H}}_{\mathrm{c}}\mathrm{Diag}(\widehat{\mathbf{h}}_{\mathrm{c},k}). \tag{53b}$$

end for

Output: $\widehat{\mathbf{G}}_k, 2 \le k \le K$.

D. Channel Estimation in the Remaining Coherence Blocks

The channel gains need to be re-estimated for the remaining channel coherence blocks as shown in Fig. 1. In general, the physical positions of the BS and the RIS are fixed, and that of a user changes much slower than the channel gains [17]–[19], [21]. Therefore, it is reasonable to assume that the AoAs of the BS and the cascaded AoDs of a user remain unchanged within tens of channel coherence blocks. With the knowledge of the angle information obtained in the first coherence block, only the cascaded channel gains need to be re-estimated.

In the remaining coherence block, the measurement matrix for user k at the BS in (4) is considered again:

$$\mathbf{Y}_k = \sqrt{p}\mathbf{G}_k\mathbf{E}_k + \mathbf{N}_k \in \mathbb{C}^{N\times\tau_k}. \tag{55}$$

Following the same derivations as (22) and (24), we define $\frac{1}{N\sqrt{p}}\widehat{\mathbf{A}}_N^{\mathrm{H}}\mathbf{Y}_k = [\mathbf{q}_{k,1},\ldots,\mathbf{q}_{k,L}]^{\mathrm{H}}$, where

$$\mathbf{q}_{k,l} = \mathbf{E}_k^{\mathrm{H}}\mathbf{B}_{k,l}\boldsymbol{\beta}_k^*\alpha_l^* + \mathbf{n}_{\text{noise}}, \tag{56}$$

$\mathbf{B}_{k,l} = [\mathbf{a}_M(\omega_l - \varphi_{k,1})\cdots\mathbf{a}_M(\omega_l - \varphi_{k,J_k})]$, and $\mathbf{n}_{\text{noise}}$ represents the corresponding noise vector.

Denote the estimation of $\mathbf{B}_{k,l}$ as $\widehat{\mathbf{B}}_{k,l} = [\mathbf{a}_M(\widehat{\omega_l - \varphi_{k,1}})\cdots\mathbf{a}_M(\widehat{\omega_l - \varphi_{k,J_k}})]$ obtained from (33a), (34a) for $k=1$, and from (54) for $2\leq k\leq K$ in the first coherence block, then the LS estimation of $\boldsymbol{\beta}_k^*\alpha_l^*$ is given by

$$\widehat{\boldsymbol{\beta}_k^*\alpha_l^*} = (\widehat{\mathbf{B}}_{k,l}^{\mathrm{H}}\mathbf{E}_k\mathbf{E}_k^{\mathrm{H}}\widehat{\mathbf{B}}_{k,l})^{-1}\widehat{\mathbf{B}}_{k,l}^{\mathrm{H}}\mathbf{E}_k\mathbf{q}_{k,l}, \tag{57}$$

Note that $\mathbf{E}_k^{\mathrm{H}}\widehat{\mathbf{B}}_{k,l}\in\mathbb{C}^{\tau_k\times J_k}$ must be a matrix with full row rank to ensure the feasibility of the pseudo inverse operation in (57), which means the pilot length needs to satisfy $\tau_k\geq J_k$ for user k.

Define $\widehat{\mathbf{H}}_{\text{RIS},k} = [\widehat{\mathbf{B}}_{k,1}\widehat{\boldsymbol{\beta}_k^*\alpha_1^*},\ldots,\widehat{\mathbf{B}}_{k,L}\widehat{\boldsymbol{\beta}_k^*\alpha_L^*}]$. The uplink channel of the k-th user can then be reconstructed by the updated cascaded channel gains obtained in this coherence block and the angle information obtained during the first coherence block as

$$\widehat{\mathbf{G}}_k = \widehat{\mathbf{A}}_N\widehat{\mathbf{H}}_{\text{RIS},k}^{\mathrm{H}}. \tag{58}$$

IV. Training reflection coefficient optimization

In this section, we optimize the training phase shift matrices to improve the channel estimation performance. Specifically, $\mathbf{E}_k,\forall k\in\mathcal{K}$ are designed to improve the ability of CS algorithm such as OMP to recover the sparsest signal $\mathbf{b}_l$ from the sparse signal recovery problem $\mathbf{p}_l = \mathbf{E}_1^{\mathrm{H}}\mathbf{A}\mathbf{b}_l + \mathbf{n}_{\text{noise}}$ in (25) and the sparsest signal $\mathbf{c}_k$ from the problem $\mathbf{z}_k = \mathbf{Z}_k\mathbf{A}\mathbf{c}_k + \mathbf{n}_{\text{noise}}$ in (52), respectively. In the following, we first investigate the design of $\mathbf{E}_1$ in (25), and then extend the solution to the design of $\mathbf{E}_k$ $(2\leq k\leq K)$ from a complex structure $\mathbf{Z}_k$.

Motivated by the theoretical work of [22] that the sparse signal $\mathbf{b}_l$ can be recovered successfully by OMP only when the following condition holds:

$$||\mathbf{b}_l||_0 \leq \frac{1}{2}\left(1+\frac{1}{\mu}\right), \tag{59}$$

where μ is the mutual coherence of the equivalent dictionary $\mathbf{D} = \mathbf{E}_1^{\mathrm{H}}\mathbf{A}$ and defined by

$$\mu = \max_{i \neq j} \frac{|\mathbf{D}_{(:,i)}^{\mathrm{H}} \mathbf{D}_{(:,j)}|}{||\mathbf{D}_{(:,i)}||_2 ||\mathbf{D}_{(:,j)}||_2}. \tag{60}$$

The condition in (59) suggests that $\mathbf{D}$ should be as incoherent (orthogonal) as possible, which leads to the following design problem

$$\begin{aligned}
&\min_{\mathbf{E}_1} ||\mathbf{D}^{\mathrm{H}}\mathbf{D} - \mathbf{I}_G||_F^2 \\
&\text{s.t.} |[\mathbf{E}_1]_{m,n}| = 1, 1 \leq m \leq M, 1 \leq n \leq \tau_1.
\end{aligned} \tag{61}$$

The solution for the unconstrained version of Problem (61) has been investigated in [23], the method in which is extended to solve the constrained Problem (61) in [13]. Based on [13] and [23], we propose a more concise solution in the following. In particular,

$$\begin{aligned}
||\mathbf{D}^{\mathrm{H}}\mathbf{D} - \mathbf{I}_G||_F^2 &= \mathrm{tr}\{\mathbf{D}^{\mathrm{H}}\mathbf{D}\mathbf{D}^{\mathrm{H}}\mathbf{D} - 2\mathbf{D}^{\mathrm{H}}\mathbf{D} + \mathbf{I}_G\} \\
&= \mathrm{tr}\{\mathbf{D}\mathbf{D}^{\mathrm{H}}\mathbf{D}\mathbf{D}^{\mathrm{H}} - 2\mathbf{D}\mathbf{D}^{\mathrm{H}} + \mathbf{I}_{\tau_1}\} + (G - \tau_1) \\
&= ||\mathbf{D}\mathbf{D}^{\mathrm{H}} - \mathbf{I}_{\tau_1}||_F^2 + (G - \tau_1).
\end{aligned} \tag{62}$$

By using (62), Problem (61) reduces to

$$\begin{aligned}
&\min_{\mathbf{E}_1} ||\mathbf{D}\mathbf{D}^{\mathrm{H}} - \mathbf{I}_{\tau_1}||_F^2 = ||\mathbf{E}_1^{\mathrm{H}}\mathbf{A}\mathbf{A}^{\mathrm{H}}\mathbf{E}_1 - \mathbf{I}_{\tau_1}||_F^2 \\
&\text{s.t.} |[\mathbf{E}_1]_{m,n}| = 1, 1 \leq m \leq M, 1 \leq n \leq \tau_1.
\end{aligned} \tag{63}$$

Define the eigenvalue decomposition $\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{U}\boldsymbol{\Upsilon}\mathbf{U}^{\mathrm{H}}$, where $\boldsymbol{\Upsilon}$ is the eigenvalue matrix and $\mathbf{U}$ is a square matrix whose columns are the eigenvectors of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$. Then, construct a matrix $\boldsymbol{\Gamma} \in \mathbb{C}^{\tau_1 \times M}$ with orthogonal rows, i.e., $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{H}} = \mathbf{I}_{\tau_1}$. For example, we can select $\boldsymbol{\Gamma} = [\mathbf{I}_{\tau_1} \mathbf{0}]$. Then, Problem (63) becomes

$$\begin{aligned}
&\min_{\mathbf{E}_1} ||\mathbf{E}_1^{\mathrm{H}}\mathbf{U}\boldsymbol{\Upsilon}^{\frac{1}{2}} - \boldsymbol{\Gamma}||_F^2 \\
&\text{s.t.} |[\mathbf{E}_1]_{m,n}| = 1, 1 \leq m \leq M, 1 \leq n \leq \tau_1.
\end{aligned} \tag{64}$$

The unconstrained LS solution of Problem (64) is $\mathbf{E}_1^{\mathrm{LS}} = (\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}})^{\mathrm{H}}$. By mapping $\mathbf{E}_1^{\mathrm{LS}}$ to the unit-modulus constraint, the final solution to Problem (64) is given by

$$\mathbf{E}_1 = \exp\left(\mathrm{i}\angle(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}})^{\mathrm{H}}\right). \tag{65}$$

For the design of $\mathbf{E}_k$ $(2 \leq k \leq K)$, it is straightforward to formulate a problem similar to Problem (64) as follows

$$\min_{\mathbf{E}_k} ||\mathbf{Z}_k \mathbf{U} \boldsymbol{\Upsilon}^{\frac{1}{2}} - \boldsymbol{\Gamma}||_F^2 \tag{66a}$$

$$\text{s.t.} |[\mathbf{E}_k]_{m,n}| = 1, 1 \leq m \leq M, 1 \leq n \leq \tau_k. \tag{66b}$$

Due to the complex structure of $\mathbf{Z}_k$ in (49), $\boldsymbol{\Gamma}$ should be carefully constructed for a better performance of Problem (66). Here, we proposed an AO method to alternately design $\boldsymbol{\Gamma}$ and $\mathbf{E}_k$.

In particular, with a pre-designed $\mathbf{E}_k$ from a DFT matrix, $\boldsymbol{\Gamma}$ can be constructed by solving the following problem

$$\boldsymbol{\Gamma} = \arg \min_{\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{H}} = \mathbf{I}_{\tau_k}} ||\mathbf{Z}_k \mathbf{U} \boldsymbol{\Upsilon}^{\frac{1}{2}} - \boldsymbol{\Gamma}||_F^2. \tag{67}$$

Problem (67) is an orthogonal Procrustes problem [24]. Define the singular value decomposition of $\mathbf{Z}_k \mathbf{U} \boldsymbol{\Upsilon}^{\frac{1}{2}} = \mathbf{P} \boldsymbol{\Xi} \mathbf{Q}^{\mathrm{H}}$, where $\boldsymbol{\Xi} \in \mathbb{C}^{\tau_k L \times M}$ is a diagonal matrix whose diagonal elements are singular values of $\mathbf{Z}_k \mathbf{U} \boldsymbol{\Upsilon}^{\frac{1}{2}}$, $\mathbf{P} \in \mathbb{C}^{\tau_k L \times \tau_k L}$ and $\mathbf{Q} \in \mathbb{C}^{M \times M}$ are unitary matrices. Then, the optimal solution to Problem (67) is given by $\boldsymbol{\Gamma} = \mathbf{P} \mathbf{Q}^{\mathrm{H}}_{(:,1:\tau_k L)}$ [24].

The complicated structure of (49) is not direct for the design of $\mathbf{E}_k$. To address this difficulty, we reconstruct (66a) through mathematical transformation such that $\mathbf{E}_k$ can be shown in a quadratic form. In particular, denote by $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_1^{\mathrm{T}}, \ldots, \boldsymbol{\Gamma}_{\tau_k}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{\Gamma}_t \in \mathbb{C}^{L \times M}$ for $1 \leq t \leq \tau_k$. With the determined $\boldsymbol{\Gamma}$ and (49), (66a) is equivalent to

$$\sum_{t=1}^{\tau_k} ||\boldsymbol{\Lambda}_{\mathrm{c}} \mathbf{A}_{\mathrm{c}}^{\mathrm{H}} \mathrm{Diag}\left(\mathbf{e}_t\right) \mathbf{U} \boldsymbol{\Upsilon}^{\frac{1}{2}} - \boldsymbol{\Gamma}_t||_F^2 \overset{(a)}{=} \sum_{t=1}^{\tau_k} ||\mathbf{T}\mathbf{e}_t - \mathrm{vec}\left(\boldsymbol{\Gamma}_t\right)||_2^2, \tag{68}$$

where $\mathbf{T} = (\mathbf{U}\boldsymbol{\Upsilon}^{\frac{1}{2}})^{\mathrm{T}} \odot \boldsymbol{\Lambda}_{\mathrm{c}} \mathbf{A}_{\mathrm{c}}^{\mathrm{H}}$. Equation (a) is due to the property $\mathrm{vec}\left(\mathbf{X} \mathrm{Diag}\left(\mathbf{e}_t\right) \mathbf{Y}^{\mathrm{T}}\right) = \left(\mathbf{Y} \odot \mathbf{X}\right) \mathbf{e}_t$ [25]. By parallel stacking $\mathbf{F} = [\mathrm{vec}\left(\boldsymbol{\Gamma}_1\right), \ldots, \mathrm{vec}\left(\boldsymbol{\Gamma}_{\tau_k}\right)]$, (68) is further equivalent to $||\mathbf{T}\mathbf{E}_k - \mathbf{F}||_F^2$. Therefore, Problem (66) is reformulated as

$$\begin{aligned} &\min_{\mathbf{E}_k} ||\mathbf{T}\mathbf{E}_k - \mathbf{F}||_F^2 \\ &\text{s.t.} |[\mathbf{E}_k]_{m,n}| = 1, 1 \leq m \leq M, 1 \leq n \leq \tau_k. \end{aligned} \tag{69}$$

The unconstrained LS solution of Problem (69) is $\mathbf{E}_k^{\mathrm{LS}} = (\mathbf{T}^{\mathrm{H}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{H}}\mathbf{F}$. By mapping $\mathbf{E}_k^{\mathrm{LS}}$ to the unit-modulus constraint, the final solution to Problem (69) is given by

$$\mathbf{E}_k = \exp\left(\mathrm{i}\angle((\mathbf{T}^{\mathrm{H}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{H}}\mathbf{F})\right). \tag{70}$$

Problem (67) and Problem (69) are optimized alternately until the stop criterion is satisfied.

Fig. 4 shows the flow chart of the training phase shift matrix design and the cascaded channel estimation.

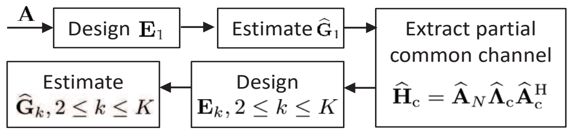

Fig. 4: Flow chart of training reflection matrix design and cascaded channel estimation.

TABLE I: Pilot Overhead and Complexity Comparison of Different Estimation Algorithms.

Algorithm		**Pilot Overhead**	**Complexity**
Proposed algorithm	First coherence block	$8J-2+(K-1)\lceil(8J-2)/L\rceil$	$\mathcal{O}(Ng+8GJ+8KGJ^4)$
	Remaining coherence blocks	JK	$\mathcal{O}(KJ^3)$
Conventional-OMP algorithm [11]		$K\lceil(8J-2)/N\rceil$	$\mathcal{O}(8KMJG^2J^4L^3)$
CS-based algorithm [13]		$K\lceil M/(JL)\rceil$	$\mathcal{O}(N^3+G^3+KM^3)$
DS-OMP algorithm [14]		$K(8J-2)$	$\mathcal{O}(KNG+KMGLJ^3)$

V. Analysis of pilot overhead and computational complexity

In this section, we analyze the pilot overhead and the computational complexity of our proposed channel estimation method. We also compare our results with the other existing algorithms, which are summarized in Table I. In this section we assume $J_1 = J_2 = \cdots = J_K = J$ for simplicity.

A. Pilot Overhead

In the first coherence block, all users need to estimate the full CSI. The theoretical minimum pilot overhead of user 1 is $\tau_1 = 8J-2$, and that of user $k, 2 \le k \le K$, is $\tau_k = \lceil(8J-2)/L\rceil$. Therefore, the total pilot overhead is $8J-2+(K-1)\lceil(8J-2)/L\rceil$. In the remaining channel coherence blocks, each user needs to transmit $\tau_k = J$, $1 \le k \le K$, time slots for the estimation of the cascaded channel gains. Thus, the total pilot overhead in this coherence block is JK.

Compared with the existing estimation algorithms in Table I, the proposed algorithm has a very low pilot overhead for estimating the full CSI in the first coherence block. When the angle information of the cascaded channel is estimated, the pilot overhead is further reduced for the cascaded gains re-estimation in the remaining coherence blocks.

B. Complexity analysis

We first calculate the computational complexity of Algorithm 2 for user 1. The complexity of Stage 1 in Algorithm 2 mainly stems from the angle rotation operation (20) on the complexity order of $\mathcal{O}(Ng)$, where g means the number of search grids within $[-\frac{\pi}{N}, \frac{\pi}{N}]$. For a very large N, a small value of g is good enough for high accuracy and low complexity. The complexity of the OMP algorithm is given by $\mathcal{O}(nml^3)$, where n is the length of the measurement data, m is the length of the sparse signal with l-sparse level [26]. Thus, the complexity of the OMP used in Stage 2 is $\mathcal{O}(8GJ^4)$. Stage 3 could be regarded as an OMP with one sparse signal, thus its complexity is on the order of $\mathcal{O}(8GJ)$. Therefore, the estimation complexity for user 1 is $\mathcal{O}(Ng+8GJ+8GJ^4)$. The computational complexity for user $k, 2 \leq k \leq K$, is from the OMP used to solve Problem (52), this estimation complexity for user $k, 2 \leq k \leq K$, is $\mathcal{O}(8GJ^4)$. Therefore, the total estimation complexity for K users in the first coherence block is given by $\mathcal{O}(Ng + 8GJ + K8GJ^4)$.

In the remaining coherence blocks, only cascaded channel gains need to be updated by using the LS solutions in (56), the computational complexity of which is on the order of $\mathcal{O}(J^3)$. Therefore, the total estimation complexity for K users in this coherence block is on the order of $\mathcal{O}(KJ^3)$.

Since $L \ll N(M)$, $J \ll N(M)$ and $g \ll N$, the complexity of the proposed algorithm in every coherence block is much lower than the other estimation algorithms in the existing literature, as shown in Table I.

VI. Simulation Results

In this section, we present extensive simulation results to validate the effectiveness of the proposed channel estimation method. All results are obtained by averaging over 500 channel realizations. The uplink carrier frequency is set as $f_c = 28$ GHz. The channel complex gains are generated according to $\alpha_l \sim \mathcal{CN}(0, 10^{-3}d_{\text{BR}}^{-2.2})$ and $\beta_{k,j} \sim \mathcal{CN}(0, 10^{-3}d_{\text{RU}}^{-2.8})$, where d_{BR} represents the distance from the BS to the RIS and is assumed to be $d_{\text{BR}} = 100$ m, while d_{RU} denotes the distance between the RIS and users and is set as $d_{\text{RU}} = 10$ m. The SNR is defined as $\text{SNR} = 10\log(10^{-6}d_{\text{BR}}^{-2.2}d_{\text{RU}}^{-2.8}p/\delta^2)$, where the identical transmit power is set as $p = 1$ W. $\{\phi_l, \theta_l, \vartheta_{k,j}\}$ are continuous and uniformly distributed over $[0, \pi)$. The number of users is $K = 4$. The number of the paths in the mmWave channel is up to 4 according to the experiment measurement in dense urban environment [15], thus the number of paths in the cascaded channel are set as $L = 5$ and $J_1 = \cdots = J_K = 4$. The element space at the BS and the RIS are set as $d_{\mathrm{BS}} = \frac{\lambda_c}{2}$ and $d_{\mathrm{RIS}} = \frac{\lambda_c}{4}$, respectively. The normalized mean square error (NMSE) of the cascaded channel matrix is defined as

$$\mathrm{NMSE} = \mathbb{E}\{||\widehat{\mathbf{G}}_k - \mathbf{G}_k||_F^2\}/\mathbb{E}\{||\mathbf{G}_k||_F^2\}.$$

The estimation algorithms considered in the simulation are as follows:

- Proposed-full-CSI: The channels are estimated by using the proposed DFT-OMP-based algorithm in Algorithm 2 in the first coherence block.
- Proposed-gains: When the angle information estimated in the first coherence block is fixed, the channels are estimated by only estimating the cascaded channel gains via the LS method in (56).
- Oracle-LS: The angle information is perfectly known at the BS, and the cascaded channel gains are estimated by (56). This algorithm can be regarded as the performance upper bound.
- LS [8]: The channels are estimated by using the LS estimator (6) with the optimal training phase shift matrix drawn from a DFT matrix.
- Conventional-OMP [11]: After approximating the cascaded channel by using the VAD representations in (10), a sparse signal reconstruction problem is constructed by vectorizing the measurement matrix. Then, the cascaded channels are estimated directly by using the OMP.
- DS-OMP [14]: The double-sparse structure of the angular domain sparse cascaded channel matrix $\mathbf{X}_k$ in (10) is exploited. The cascaded channels are estimated by using OMP for each non-zero row of $\mathbf{X}_k$.

Fig. 5 illustrates the impact of pilot overhead on the estimation performance when SNR is 0 dB. Since the number of time slots allocated to each user for channel estimation in the Proposed-full-CSI algorithm is different, we choose the average number of time slots for each user as the x-axis measurement, denoted as T. It is obvious that more pilot overhead leads to better NMSE performance for all channel estimation algorithms. The Proposed-full-CSI algorithm with $T = 14$ time slots outperforms the LS algorithm with $T = M = 100$ time slots. This is because the cascaded channel estimated by Proposed-full-CSI algorithm exploits the low-rank characteristic of the mmWave channel, while LS algorithm ignores this sparsity feature. When the angle information is estimated, the Proposed-gains algorithm only needs $T = 2J = 8$ time slots to surpass the performance of LS algorithm. In addition, we observe that even if two OMP-

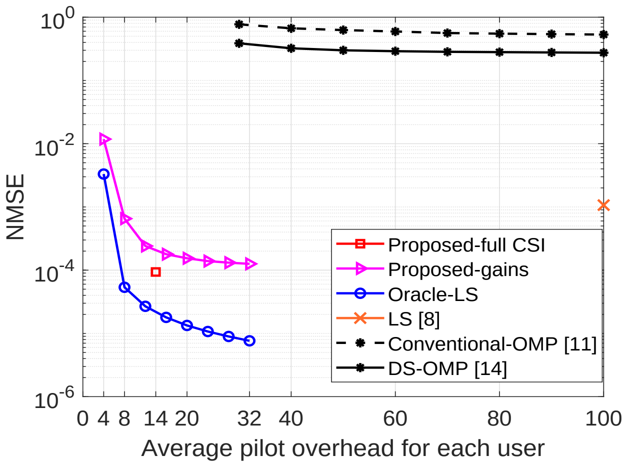

Fig. 5: NMSE versus pilot overhead, when $N = 100$, $M = 100$, $L = 5$, $J = 4$ and SNR=0 dB.

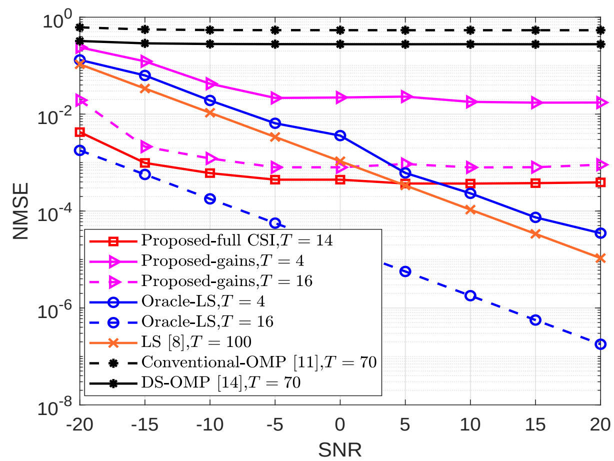

Fig. 6: NMSE versus SNR, when $N = 100$, $M = 100$, $L = 5$ and $J = 4$.

based algorithms in [11] and [14] require much more pilots than their theoretical minimum pilot overhead shown in Table I, it is difficult for them to achieve acceptable good estimation performance. This is because the algorithm in [11] completely ignores the double sparse structure of the cascaded channels, resulting in a lot of false alarm estimates. The algorithm in [14] ignores the impact of power leakage and ideally assumes that the number of multipaths is known, resulting in that the real low-power paths are readily replaced by the virtual high-power paths. The impact of power leakage is addressed in the proposed estimation algorithm by using the angle rotation operation, designing the optimal phase shift matrix, and enlarging the dimension of the dictionary. Fianlly, the proposed channel estimation strategy, including Proposed-full-CSI followed by Proposed-gains, can achieve significant estimation performance with very little pilot overhead, compared with the existing channel estimation algorithms in the RIS-aided mmWave massive MISO communicate systems.

Fig. 6 displays NMSE performances as a function of SNR for difference channel estimation methods. At low SNR, it can be seen that the performance of the proposed algorithms are better than that of LS algorithm. As SNR increases, the estimation accuracy of the proposed algorithms increases but will reach saturation at relatively high SNR. The reasons for the error floor are twofold: one is the slight AoA steering matrix non-orthogonality since N is finite, the other is the mismatch between the estimated cascaded AoDs and the real cascaded AoDs due to the fact that OMP selects the estimation angles from the discrete grid.

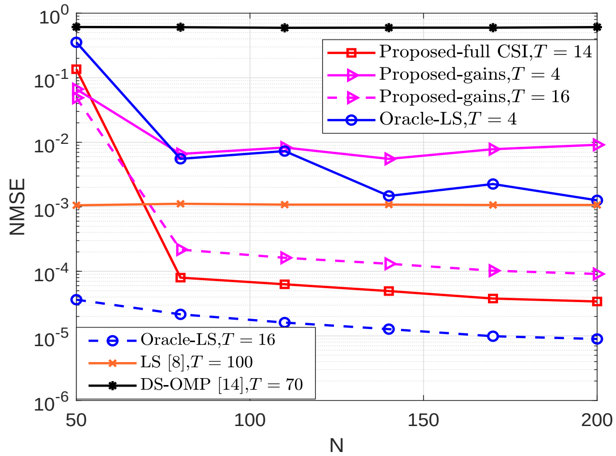

Fig. 7: NMSE versus the number of antennas, when $M=100$, $L=5$, $J=4$ and SNR=0 dB.

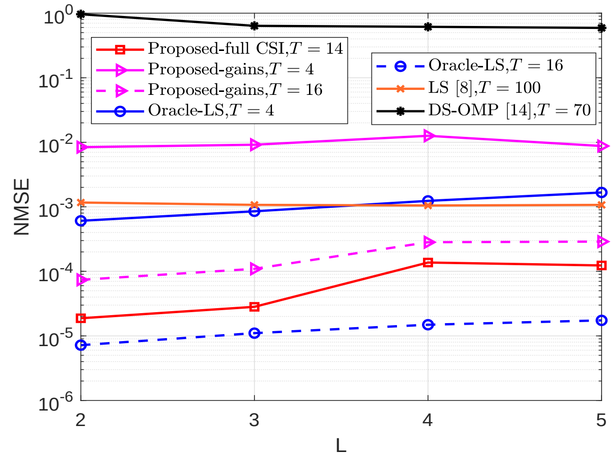

Fig. 8: NMSE versus the number of paths from the BS to the RIS L, when $N=100$, $M=100$, $J=4$ and SNR=0 dB.

We next show that the NMSE performance with various numbers of antennas N when SNR=0 dB in Fig. 7. From the figure, when N increases, the performance of the LS method remains stable and is not affected by the size of the estimated channel, because there are enough time slots to support LS estimation in the complete spatial domain. OMP-based benchmark consistently performs poorly due to its serious power leakage effect. On the other hand, when N is larger than 80, the Proposed-full-CSI algorithm works well, because the resolution of DFT in Algorithm 1 is better with larger N. At the same time, the angle rotation operation in Algorithm 1 can also alleviate the impact of power leakage. In addition, when the pilot overhead increases from $J=4$ to $4J=16$, the performance of the Proposed-gains algorithm increases since more pilot overhead can provide more measurement data diversity in the Proposed-gains algorithm.

Fig. 8 shows the impacts of the number of spatial paths between the BS and the RIS. It is obvious that the number of spatial paths has no effect on LS method. However, the performance of the proposed algorithms decreases when the number of spatial paths increases, which is due to the fact that the number of parameters (sparsity level) to be estimated increases.

VII. Conclusions

In this paper, we developed a cascaded channel estimation method for the RIS-aided uplink mmWave massive MU-MISO systems with much less pilot overhead. Such an interesting result is enabled by utilizing the invariant angle information for a long time, exploiting the linear correlation among cascaded paths, and exploring the partial CSI of the common BS-RIS channel. The theoretical minimum pilot overhead was characterized, and training reflection matrices were designed. Simulation results showed that the NMSE performance of the proposed algorithm outperforms the existing OMP-based algorithms and the pilot overhead required by the proposed algorithm is much less than the existing method.

APPENDIX A
THE PROOF OF LEMMA 1

We calculate

$$\mathbf{a}_N^{\mathrm{H}}(\psi_l)\mathbf{a}_N(\psi_i)=\sum_{m=1}^{N}e^{-\mathrm{i}2\pi(m-1)(\psi_i-\psi_l)}=\frac{1-e^{-\mathrm{i}2\pi N(\psi_i-\psi_l)}}{1-e^{-\mathrm{i}2\pi(\psi_i-\psi_l)}}. \tag{71}$$

$\mathbf{a}_N^{\mathrm{H}}(\psi_l)\mathbf{a}_N(\psi_i)$ is bounded for any $l\neq i$ as $N\to\infty$ and thus $\lim_{N\to\infty}\frac{1}{N}\mathbf{a}_N^{\mathrm{H}}(\psi_l)\mathbf{a}_N(\psi_i)=0$. When $l=i$, direct calculation yields that $\mathbf{a}_N^{\mathrm{H}}(\psi_l)\mathbf{a}_N(\psi_j)=N$ and hence $\lim_{N\to\infty}\frac{1}{N}\mathbf{a}_N^{\mathrm{H}}(\psi_l)\mathbf{a}_N(\psi_i)=1$. Therefore, when $N\to\infty$, the limit of (71) is

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{a}_N^{\mathrm{H}}(\psi_l)\mathbf{a}_N(\psi_i)=\delta\left(\psi_i-\psi_l\right), \tag{72}$$

where $\delta(\cdot)$ is the Dirac delta function.

The proof is completed.

APPENDIX B
THE PROOF OF LEMMA 2

Let us first consider the case $\psi_l\in[0,\frac{d_{\mathrm{BS}}}{\lambda_c})$. Then, the (n,l)-th element of $\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N$ is calculated in

$$\begin{aligned}[\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N]_{n,l}=\left[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)\right]_n&=\sqrt{\frac{1}{N}}\sum_{m=1}^{N}e^{\mathrm{i}\frac{2\pi}{N}(n-1)(m-1)}e^{-\mathrm{i}2\pi(m-1)\psi_l}\\&=\sqrt{\frac{1}{N}}\sum_{m=1}^{N}e^{-\mathrm{i}2\pi(m-1)(\psi_l-\frac{n-1}{N})}=\sqrt{\frac{1}{N}}\frac{1-e^{\mathrm{i}2\pi N(\frac{n-1}{N}-\psi_l)}}{1-e^{\mathrm{i}2\pi(\frac{n-1}{N}-\psi_l)}}.\end{aligned} \tag{73}$$

According to the proof in Appendix A, when $N\to\infty$, the limit of (73) is

$$\lim_{N\to\infty}\left|[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)]_n\right|=\sqrt{N}\delta\left(\frac{n-1}{N}-\psi_l\right). \tag{74}$$

Hence, there always exist some integers $n_l=N\psi_l+1$ such that $|[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)]_{n_l}|=\sqrt{N}$, and the other elements of $\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)$ are zero. In other words, $\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N$ is a sparse matrix with all powers being concentrated on the points (n_l,l) for $\forall l$.

When $\psi_l \in [-\frac{d_{\mathrm{BS}}}{\lambda_c}, 0)$, by using criterion $e^{\mathrm{i}x} = e^{\mathrm{i}(x+2\pi)}$, (73) is equivalent to

$$
\begin{aligned}
[\mathbf{U}_N^{\mathrm{H}}\mathbf{A}_N]_{n,l} = \left[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)\right]_n &= \sqrt{\frac{1}{N}}\sum_{m=1}^{N} e^{-\mathrm{i}\left[2\pi(m-1)(\psi_l-\frac{n-1}{N})+2\pi(m-1)\right]} \\
&= \sqrt{\frac{1}{N}}\sum_{m=1}^{N} e^{-\mathrm{i}2\pi(m-1)(\psi_l-\frac{n-1}{N}+1)}
\end{aligned} \tag{75}
$$

When $N \to \infty$, the limit of (75) is

$$
\lim_{N\to\infty} \left|[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)]_n\right| = \sqrt{N}\delta\left(\psi_l - \frac{n-1}{N} + 1\right). \tag{76}
$$

Hence, there always exist some integers $n_l = N + N\psi_l + 1$ such that $|[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)]_n| = \sqrt{N}$, and the other elements of $[\mathbf{U}_N^{\mathrm{H}}\mathbf{a}_N(\psi_l)]_n$ are zero. Combining (74) and (76), we arrive at (17).

The proof is completed.

References

[1] M. Di Renzo, M. Debbah, D.-T. Phan-Huy *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *J. Wireless Commun. Netw.*, 2019,129(2019).

[2] C. Pan, H. Ren, K. Wang *et al.*, "Reconfigurable intelligent surface for 6G and beyond: Motivations, principles, applications, and research directions," *Accepted by IEEE Communications Magazine*, Jan. 2021.

[3] C. Pan, H. Ren, K. Wang *et al.*, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, 2020.

[4] C. Pan, H. Ren, K. Wang *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020.

[5] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of intelligent reflecting surfaces using scattering parameter network analysis," 2020. [Online]. Available: https://arxiv.org/abs/2011.11362

[6] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast MISO communication systems," *IEEE Trans. Signal Process.*, vol. 68, pp. 3236–3251, 2020.

[7] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," *2019 IEEE GLOBECOM*, pp. 1–6, Dec.

[8] T. Lindstrøm Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5000–5004.

[9] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced ofdm: Channel estimation and reflection optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518–522, Dec. 2019.

[10] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020.

[11] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.

[12] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for ris-aided mmwave mimo systems via atomic norm minimization," *IEEE Transactions on Wireless Communications*, pp. 1–1, Apr. 2021.

[13] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, “Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems,” 2019. [Online]. Available: https://arxiv.org/abs/1912.03619

[14] X. Wei, D. Shen, and L. Dai, “Channel estimation for RIS assisted wireless communications: Part II - an improved solution based on double-structured sparsity (invited paper),” *IEEE Commun. Lett.*, pp. 1–1, 2021.

[15] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter wave channel modeling and cellular capacity evaluation,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[16] M. Wang, F. Gao, S. Jin, and H. Lin, “An overview of enhanced massive MIMO with array signal processing techniques,” *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 886–901, Sept. 2019.

[17] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, “Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sept. 2017.

[18] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, “Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[19] D. Fan, F. Gao, Y. Liu, Y. Deng, G. Wang, Z. Zhong, and A. Nallanathan, “Angle domain channel estimation in hybrid millimeter wave massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.

[20] X. Li and V. Voroninski, “Sparse signal recovery from quadratic measurements via convex programming,” *SIAM J. Math. Anal.*, vol. 45, no. 5, pp. 3019–3033, 2013.

[21] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, “Beam squint and channel estimation for wideband mmwave massive MIMO-OFDM systems,” *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, Dec. 2019.

[22] J. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[23] J. M. Duarte-Carvajalino and G. Sapiro, “Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization,” *IEEE Trans. Signal Process.*, vol. 19, no. 7, pp. 1395–1408, Jul. 2009.

[24] X.-D. Zhang, *Matrix analysis and applications*. Cambridge Univ. Press, 2017.

[25] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, “Tensor decomposition for signal processing and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[26] K. Venugopal, A. Alkhateeb, N. GonzÃ¡lez Prelcic, and R. W. Heath, “Channel estimation for hybrid architecture-based wideband millimeter wave systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sept. 2017.