Beamspace Channel Estimation via PARAFAC Decomposition for RIS Assisted Millimeter-Wave Multiuser MISO Communications

Xinying Guo, Zongyuan Xie, Jiankang Zhang, *Senior Member, IEEE*, Sheng Chen, *Life Fellow, IEEE*, and Chunhua Zhu

***Abstract*—The reconfigurable intelligent surface (RIS) with massive low-cost passive reflecting elements integrated on a planar surface has the ability to favourably reconfigure the wireless propagation environment, thereby significantly improving the performance of wireless communication networks. In this work, we consider uplink (UL) channel estimation for the RIS assisted millimeter-wave multiuser multiple-input single-output beamspace system where the base station (BS) is equipped with lens antenna array. This channel state information (CSI) estimation task is extremely challenging for two reasons. First, the BS only has limited number of radio frequency chains but the size of beamspace channel is very large. Second, the number of passive components in the RIS is abundance but they lack signal processing capabilities. By exploiting the parallel factor (PARAFAC) decomposition of the received signals, we derive an iterative estimation algorithm, called unitary approximate message passing (UAMP), to accurately estimate the channels between the BS and the RIS as well as the channels between the RIS and the users. To guide the selection of the system parameters, we provide the uniqueness conditions for our PARAFAC decomposition based channel estimation. To theoretically verify the efficiency of our UAMP algorithm, the Cramér-Rao bound (CRB) of the estimation is also derived. Besides, we investigate the achievable downlink (DL) sum rate for the channel estimation obtained by the proposed algorithm by using the maximum power beam selection, the optimized phase shift matrix and the zero forcing precoding. Extensive simulation results demonstrate the excellent mean squared error (MSE) performance of our UAMP estimation algorithm. In particular, for sufficiently high UL signal-to-noise ratio, the MSE of our channel estimation reaches the CRB. Simulation results also show that the DL sum rate achieved by the estimated CSI is very close to that obtained by the perfect CSI. Theoretical analysis and simulation results thus validate the effectiveness and reliability of our beamspace channel estimation approach.**

***Index Terms*—Reconfigurable intelligent surface, lens antenna array, millimeter-wave, beamspace, channel estimation, PARAFAC, Cramér-Rao bound, downlink sum rate.**

This work was supported by the National Natural Science Foundation of China (Grants 61901159), the Cultivation Programme for Young Backbone Teachers in Henan University of Technology (Grants 21420104) and the Natural Science Foundation of Henan (Grants 242300421709). (*Corresponding author: Jiankang Zhang.*)

X. Guo (guoxinying@haut.edu.cn), Z. Xie (xzy999@stu.haut.edu.cn) and C. Zhu (zhuchunhua@haut.edu.cn) are with Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education; and also with College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China.

J. Zhang (jzhang3@bournemouth.ac.uk) is with Department of Computing & Informatics, Bournemouth University, BH12 5BB, UK.

S. Chen (sqc@ecs.soton.ac.uk) is with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK.

I. Introduction

Millimeter wave (mmWave) communication with its abundantly available bandwidth, high data rate and interference immunity is considered as a critical technology for the fifth generation (5G) and future mobile communications [1], [2]. However, severe path loss occurs during signal transmission due to high mmWave frequency. To compensate for this high path loss, antenna array with large number of antenna elements is required to obtain high beamforming gain.

Conventional multiple-input multiple-output (MIMO) technology with a dedicated radio frequency (RF) chain for each antenna is impractical for mmWave MIMO systems, because of substantial hardware costs and extremely high energy consumption [3], [4]. The work [5] proposed the concept of beamspace MIMO, which is capable of reducing the number of RF chains required. A lens antenna array consists of an electromagnetic lens and antennas placed at its focal surface, which focuses signals from different directions on different antennas to convert the spatial channel into a beamspace channel. Owing to the sparse nature of mmWave beamspace channels, a small number of dominant beams exist and they can be selected, to reduce the number of required RF chains and hence the dimensionality of MIMO systems. This results in low-complexity and low-cost mmWave MIMO systems. Extensive researches have been conducted for many topics of beamspace MIMO systems, including beam selection, channel estimation, etc. To address the issues of interference and the waste of RF links in traditional beam selection schemes, the work [6] designed an interference-aware beam selection scheme based on the sum rate maximization. The study [7] designed a phase-shifter-assisted beam selection network, which uses a single RF chain to support multiple focused energy beams so as to mitigate beam skew in wideband mmWave MIMO systems. A support detection based channel estimation scheme was proposed in [8], which is capable of obtaining high estimation accuracy with reliable performance and low pilot overhead. Inspired by the classical successive interference cancellation for multiuser detection, the work [9] investigated a successive support detection based beamspace channel estimation scheme, which does not impose the common support assumption.

Although lens antenna array is capable of reducing the cost and complexity of mmWave communication system, it does not have the ability to address the critical issue that mmWave signal is easily blocked by obstacles. With the recent advance in digital metamaterials, the reconfigurable intelligent

surface (RIS) has been developed into a practical technology with great potential for solving the problem of blind spots in coverage. Moreover, it is also capable of extending signal coverage, overcoming transmission distance limitations and tackling transmission signal blockage difficulties in mmWave band. RIS is made of electromagnetic material and has a unique physical structure. Specifically, it consists of many passive reflection units, integrated on a two-dimensional surface, and each reflection unit is independently controllable. Through joint regulation of the reflection units, the purpose of adjusting phase, amplitude and other parameters of the incident signal can be achieved, and the signal propagation channel can be favourably changed. Therefore, RIS is regarded as a crucial technology in the 5G and the sixth generation (6G) of mobile communications [10]–[12].

Efficient and accurate channel estimation is crucial for practical implementation of RIS assisted communication systems. Although it is revealed that the achievable rate gains of the active RIS would be superior than that of the passive RIS if the power budget and the number of RIS elements satisfy certain conditions [13], it is worth noting that we only consider the channel estimation for the passive RIS instead of the active RIS in this paper. A simple on/off two-state channel estimation protocol was invented in [14]. The protocol adjusts one RIS reflective unit at a time to estimate the cascade channels corresponding to individual reflective elements one by one. However, turning off all but one reflector element leads to a very weak RIS's reflected power, thereby reducing the received signal-to-noise ratio (SNR) and consequently resulting in a poor channel estimation accuracy. To avoid this shortcoming, the authors of [15] designed the RIS phase shift matrix as a discrete Fourier transform (DFT) matrix and proposed the least square (LS) estimator for the cascaded channels. MmWave channels are known to be inherently sparse. Exploiting this property and based on Katri-Rao and Kronecker products, the work [16] converted the cascaded channel estimation into a sparse signal recovery problem. Further applying the orthogonal matching pursuit algorithm for sparse signal recovery, the corresponding cascaded channels can be estimated with relatively low training overhead. In [17], a dual time-scale channel estimation framework and a dual-link guided transmission scheme were proposed, which significantly reduces the training overhead. Based on the innovative uplink (UL) channel estimation protocol, the study [18] presented a method with low pilot overhead by employing the unvarying angle information and the linear correlation of multiuser cascaded paths as well as the partial channel state information (CSI) of the common BS-RIS channel. Distributed machine learning was employed to estimate the downlink (DL) channels for RIS assisted wireless communications [19], which obtains better channel estimation performance compared with the conventional approaches.

Furthermore, the authors of [20] strove to simultaneously acquire both the direct channel from the user to the BS and the cascaded channel by leveraging array signal processing techniques. They proposed a two-stage channel estimation algorithm that begins with the estimation of the angle of departure (AoD) at the RIS using semi-definite programming (SDP). This is followed by the estimation of the direct channel angle and the angle of arrival (AoA) at the BS using an iterative quadratic maximum likelihood approach. However, the separate estimation processes for AoD and AoA necessitate an additional step involving a computationally intensive data association procedure. In [21], two flexible and resilient approaches were introduced for the concurrent estimation of direct and reflective channels in mmWave systems assisted by RIS under dynamic channel conditions. In these systems, the RIS is equipped with more versatile, arbitrarily shaped elements. The authors conceptualized the challenge as a robust low-rank sparse matrix reconstruction issue and crafted an alternating optimization-based solution and a machine learning-based solution.

In addition, some researchers have tackled the channel estimation problem in RIS aided systems through tensor. Tensor as an effective mathematical tool has a wide range of applications in various signal processing problems [22], [23]. The basic principle of tensor decomposition is to decompose a high-dimensional tensor into a linear combination of multiple rank-one matrices, thereby facilitating low-complexity and effective estimation of multiple large MIMO channel matrices [24], [25]. The tensor modeling approach was exploited in [26] to construct the received signal into a three-dimensional tensor. Then based on the parallel factor (PARAFAC) decomposition of the received signal, two channel estimation methods were proposed. The first method solves the rank-1 matrix approximation problem based on the Khatri-Rao factorization to achieve the closed-form cascaded MIMO channel solution, and the second method uses an iterative bilinear alternative LS (ALS) algorithm to find a cascaded MIMO channel estimate. The study [27] considered the UL channel estimation for an RIS aided multiuser multiple-input-single-output (MISO) system. More specifically, a parallel factorization-based channel estimation framework was proposed to unfold the final cascaded channel model, and two iterative algorithms were employed to estimate the cascaded MIMO channel matrices, namely, the channel matrix between the base station (BS) and the RIS as well as the channel matrix between the RIS and the users. The literature [28] manipulated the fourth-order tensors for estimating RIS assisted communication channels for the first time by jointly utilizing tensor structures and low-rank characteristics of mmWave channels. The work [29] considered a dual RIS aided MIMO system and introduced an estimation algorithm based on ALS using Tucker 2 tensor structure, to estimate the cascaded MIMO channels. The work [30] considered the DL RIS assisted mmWave system employing orthogonal frequency division multiplexing transmission. In this work, the received signal is represented as a low-rank third-order tensor by exploring the sparse scattering properties inherent to mmWave channels, to facilitate channel estimation. The study [31] utilized both compressed sensing and tensor decomposition techniques to derive a channel estimation algorithm with low training overhead.

To exploit the advantages of both RIS and lens antenna array simultaneously, this paper considers the RIS aided mmWave multiuser MISO beamspace system where the BS is equipped with lens antenna array. As discussed previously,

RIS has the ability to favourably reconfigure the wireless propagation environment, thereby extending the coverage of signal and improving the quality of the received signal, while lens antenna array contributes to the reduction in the number of RF chains required, thereby reducing costs and saving energy. Consequently, this system improves communication quality at low cost, low energy consumption as well as low complexity [32], [33]. However, channel estimation in this system is an extremely challenging task for two reasons. First, the BS only has limited number of RF chains but the size of beamspace channel is very large. Second, although the number of passive components in the RIS is abundance, it lacks signal processing capabilities. We establish a corresponding relationship between the RIS aided beamspace multiuser MISO communication system model and the tensor model. By utilizing the PARAFAC signal structure, we propose an effective algorithm, called unitary approximate message passing (UAMP), to estimate channels. To the authors' best knowledge, this paper is the first work of using the tensor approach to solve the beamspace channel estimation for the RIS aided mmWave multiuser MISO communication system. We use both analysis and simulation to validate the effectiveness and reliability of our proposed algorithm. The contributions of this paper are as follows.

- By modeling the received pilot signal as a third-order tensor, the channel estimation problem is formulated for the RIS-aided mmWave multiuser MISO beamspace system. By utilizing the PARAFAC decomposition, high dimensional tensor that includes the unknown channels is unfolded into three different terms. Afterward, UAMP is adopted to estimate the channels in an iterative way.
- We analyze the uniqueness conditions for our PARAFAC decomposition based UAMP algorithm, which offers useful guidance for selecting the system parameters. More importantly, the Cramér-Rao Bound (CRB) associated with our algorithm is derived to demonstrate its performance validity and efficiency.
- Utilizing the proposed maximum power (MP) beam selection at the BS, the optimized phase shift matrix at the RIS as well as the zero forcing (ZF) precoding at the BS, we calculate the DL sum rate for the RIS-aided mmWave multiuser MISO system using the channel estimation obtained by our proposed algorithm, thereby verifying its effectiveness.

Table I compares our proposed UAMP algorithm with the existing channel estimation algorithms for RIS assisted communication systems.

It is worth emphasizing that our UAMP algorithm is different from the existing tensor-based methods. This is because the RIS-aided mmWave multiuser MISO beamspace system considered in this paper is totally different from the systems considered in the works [26]–[31]. It is well understood that the signal models constructed for different systems are very distinct, and the way of formulating the tensor in our paper is not the same as those formulated in [26]–[31]. Since the PARAFAC decomposition of the tensor model has the trilinearity property, the three modes of the constructed tensors within the channels that need to be estimated can be achieved. Our UAMP method can then estimate the channels by utilizing the algebraic structure of the PARAFAC model.

The paper is structured as follows. Section II provides a detailed description of the RIS-aided mmWave multiuser MISO beamspace communication system and the tensor signal modeling. Our PARAFAC-based channel estimation algorithm is developed in Section III, where its feasibility conditions are also analyzed. The CRB for the proposed UAMP algorithm is derived in Section IV. Section V is devoted to the achievable DL sum rate of the system using the proposed channel estimate, where the BS adopts the proposed MP beam selection and the ZF precoding, while the RIS employes the optimized phase shift matrix. Section VI presents the simulation results, and finally Section VII offers our conclusions.

II. MmWave Multiuser MISO Beamspace System and Tensor Model

This section first introduces the RIS-aided mmWave multiuser MISO UL beamspace communication system. A PARAFAC decomposition based on the third-order tensor of the received pilot signal is then developed for the cascade channel model of this communication system.

A. Beamspace MIMO Induced by RIS

Fig. 1 illustrates the RIS-aided mmWave multiuser MISO beamspace system, where the BS is equipped with an uniform

TABLE I
Comparison of channel estimation algorithms for RIS Assisted communication systems.

Classification	Year	Paper	Contributions
Non-tensor methods	2019	Mishra and Johansson [14]	Two-state channel estimation protocol
	2020	Jensen and De Carvalho [15]	The RIS phase shift matrix and the LS channel estimator
	2020	Wang *et al.* [16]	A sparse signal recovery problem converted by channel estimation
	2021	Hu *et al.* [17]	A dual time-scale channel estimation framework and a dual-link guided transmission scheme
	2022	Zhou *et al.* [18]	A three-stage uplink channel estimation protocol and a one-dimensional search method
	2022	Dai and Wei [19]	Distributed machine learning-aided channel estimation
	2023	Noh *et al.* [20]	Two-stage channel estimation algorithm
	2024	Chu *et al.* [21]	An alternating optimization-based solution and a machine learning-based solution
Tensor-based methods	2019	de Araújo *et al.* [26]	Two channel estimation methods based on the PARAFAC decomposition
	2021	Wei *et al.* [27]	A parallel factorization based channel estimation framework and two iterative algorithms
	2021	Gherekhloo *et al.* [28]	The fourth-order tensors-aided method
	2022	Ardah *et al.* [29]	An estimation algorithm based on ALS using Tucker 2 tensor structure
	2022	Zheng *et al.* [30]	A structured canonical polyadic decomposition-based method
	2022	Zhang *et al.* [31]	Compressed sensing and tensor decomposition-based channel estimation algorithm
This paper			The first PARAFAC based beamspace channel estimation framework and UAMP algorithm

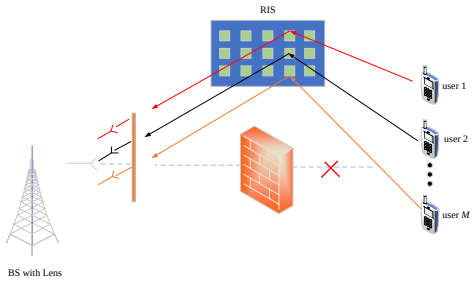

Fig. 1. Illustration of RIS-assisted mmWave multiuser MISO beamspace communication system.

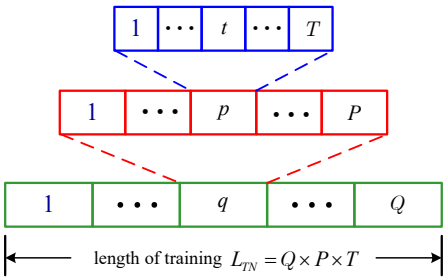

Fig. 2. Pilot overhead requirements for the considered system.

planar array (UPA) of K lens antennas and K_{RF} RF chains, while the RIS has N reflecting elements and can individually adjust its reflection coefficients. The system operates on the time division duplexing mode and supports M single-antenna users on the same time-frequency resource block. It is assumed that there are no line-of-sight or direct links between the BS and the users since the mmWave links are extremely vulnerable to blockage.

The end-to-end cascade communication channel consists of the channel between the RIS and the BS as well as the channel between all the M users and the RIS. The former is represented by the channel matrix $\mathbf{G} \in \mathbb{C}^{K\times N}$, and the latter is represented by the channel matrix $\mathbf{H} = \left[\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_M\right] \in \mathbb{C}^{N\times M}$, where $\mathbf{h}_m \in \mathbb{C}^{N\times 1}$ stands for the channel between the m-th user to the RIS for $1 \le m \le M$. Since the system carrier is mmWave frequency, the well known Saleh-Valenzuela channel model [34] is adopted to represent $\mathbf{G}$ as

$$\mathbf{G} = \sqrt{\frac{KN}{L_G}} \sum_{l_1=1}^{L_G} \alpha_{l_1}^G \mathbf{a}_{\mathrm{BS}}\left(\theta_{l_1}^b, \phi_{l_1}^b\right) \mathbf{a}_{\mathrm{RIS}}\left(\theta_{l_1}^r, \phi_{l_1}^r\right)^{\mathrm{T}}. \quad (1)$$

In (1), L_G is the number of paths in the channel between the RIS and the BS, while $\alpha_{l_1}^G$, $\theta_{l_1}^b$ ($\phi_{l_1}^b$), and $\theta_{l_1}^r$ ($\phi_{l_1}^r$) stand for the complex path gain, the azimuth (elevation) angle at the BS, and the azimuth (elevation) angle at the RIS for the l_1-th path, respectively. Furthermore, the normalized array steering vectors for the BS and the RIS are denoted by $\mathbf{a}_{\mathrm{BS}}\left(\theta^b, \phi^b\right) \in \mathbb{C}^{K\times 1}$ and $\mathbf{a}_{\mathrm{RIS}}\left(\theta^r, \phi^r\right) \in \mathbb{C}^{N\times 1}$, respectively. Given a $K_1 \times K_2$ UPA, where $K = K_1 \times K_2$, its array steering vector $\mathbf{a}(\theta, \phi) \in \mathbb{C}^{K\times 1}$ is given by

$$\mathbf{a}(\theta, \phi) = \mathbf{a}^{az}(\theta) \otimes \mathbf{a}^{el}(\phi)^{\mathrm{T}}, \quad (2)$$

where $\otimes$ denotes the Kronecker product, $\mathbf{a}^{az}(\theta) = \frac{1}{\sqrt{K_1}}\left[e^{\frac{-\mathrm{j}2\pi d \sin(\theta) k_1}{\lambda}}, \forall k_1 \in \mathcal{I}\left(K_1\right)\right] \in \mathbb{C}^{K_1\times 1}$ and $\mathbf{a}^{el}(\phi) = \frac{1}{\sqrt{K_2}}\left[e^{\frac{-\mathrm{j}2\pi d \sin(\phi) k_2}{\lambda}}, \forall k_2 \in \mathcal{I}\left(K_2\right)\right] \in \mathbb{C}^{K_2\times 1}$ are the azimuth and elevation steering vectors, respectively. Here λ is the carrier wavelength, d is the antenna spacing which is set $d = \lambda/2$, and we define $\mathcal{I}(k) = \left\{s - \frac{k-1}{2}, s = 0, 1, \ldots, k-1\right\}$. The array steering vector of the RIS can be obtained in the same way by replacing K with N. Similarly, the channel $\mathbf{h}_m$ is given by

$$\mathbf{h}_m = \sqrt{\frac{N}{L_m}} \sum_{l_2=1}^{L_m} \alpha_{l_2}^m \mathbf{a}_{\mathrm{RIS}}(\theta_{l_2}^m, \phi_{l_2}^m), \quad (3)$$

where the channel between the m-th user and the RIS has L_m paths, $\alpha_{l_2}^m$ and $\theta_{l_2}^m$ ($\phi_{l_2}^m$) are the complex path gain and the azimuth (elevation) angle at the RIS for the l_2-th path.

To reduce the number of RF chains, i.e., achieving $K_{RF} \ll K$, and hence reducing hardware costs and energy consumption, the BS adopts a lens antenna array, which acts as a spatial DFT matrix $\mathbf{U} \in \mathbb{C}^{K\times K}$ on the wireless channel to convert it into a beamspace channel. The matrix $\mathbf{U}$ consists of the array steering vectors of K orthogonal directions (beams) that cover the entire angle space and it is given by

$$\mathbf{U} = \left[\mathbf{a}\left(i/K_1, j/K_2\right), \forall i \in \mathcal{I}\left(K_1\right), \forall j \in \mathcal{I}\left(K_2\right)\right]^{\mathrm{H}}. \quad (4)$$

Assume that the RIS has P feasible phase configurations, and they are collected in a RIS phase shift matrix $\mathbf{\Phi} \in \mathbb{C}^{P\times N}$, where the p-th row of $\mathbf{\Phi}$, denoted as $[\mathbf{\Phi}]_{p,:}$, represents the p-th feasible phase configuration. Obviously, P is no more than N. During the pilot transmission, the BS employs Q combiners $\mathbf{W}_q \in \mathbb{C}^{K_{RF}\times K}$, $q = 1, 2, \cdots, Q$, $Q = K/K_{RF}$, to combine the received UL signal matrices one after another. Further assume that the UL training period corresponding to the p-th feasible phase configuration and the q-th combiner consists of T consecutive time slots, as shown in the blue blocks of Fig. 2, and the transmitted pilot signals are collected in the matrix $\mathbf{X}_{p,q} \in \mathbb{C}^{M\times T}$. Then with the RIS set at its p feasible phase configuration and the BS employing the q-th combiner, the BS's baseband signal matrix $\mathbf{Y}_{p,q} \in \mathbb{C}^{K_{RF}\times T}$ can be expressed as

$$\mathbf{Y}_{p,q} = \mathbf{W}_q \mathbf{U} \mathbf{G} \mathbf{D}_p(\mathbf{\Phi}) \mathbf{H} \mathbf{X}_{p,q} + \mathbf{W}_q \mathbf{Z}_{p,q}, \quad (5)$$

where $\mathbf{D}_p(\mathbf{\Phi}) = \mathrm{diag}\left([\mathbf{\Phi}]_{p,:}\right) \in \mathbb{C}^{N\times N}$ is the diagonal matrix with $[\mathbf{\Phi}]_{p,:}$ as its diagonal entries, and $\mathbf{Z}_{p,q} \in \mathbb{C}^{K\times T}$ is the complex additive white Gaussian noise (AWGN) matrix whose elements have zero mean and variance $\sigma^2/2$ per dimension. To ensure the efficiency of channel estimation, T must be no smaller than M. Typically orthogonal pilot signals are adopted, i.e., $\mathbf{X}_{p,q}\mathbf{X}_{p,q}^{\mathrm{H}} = \mathbf{I}_M$. Overall, the pilot overhead requirements for the considered system is $L_{TN} = Q \times P \times T$, which is depicted in Fig. 2. For convenience, we set $\mathbf{X}_{p,q} = \mathbf{X}$ for $1 \le p \le P$ and $1 \le q \le Q$.

B. Tensor Signal Modeling

After the received UL signal is dealt with the Q combiners, we stack the resulting $\mathbf{Y}_{p,q}$ for $1 \le q \le Q$ in the column

direction to yield

$$\widetilde{\mathbf{Y}}_p = \begin{bmatrix} \mathbf{Y}_{p,1} \\ \vdots \\ \mathbf{Y}_{p,Q} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_Q \end{bmatrix} \mathbf{UGD}_p(\mathbf{\Phi})\mathbf{HX} + \begin{bmatrix} \mathbf{W}_1\mathbf{Z}_{p,1} \\ \vdots \\ \mathbf{W}_Q\mathbf{Z}_{p,Q} \end{bmatrix} = \mathbf{WUGD}_p(\mathbf{\Phi})\mathbf{HX} + \bar{\mathbf{Z}}_p. \tag{6}$$

We design $\mathbf{W} \in \mathbb{C}^{K\times K}$ in (6) as a DFT matrix. Right multiplying $\mathbf{X}^{\mathrm{H}}$ on the both sides of Eq. (6) yields $\widetilde{\mathbf{R}}_p = \widetilde{\mathbf{Y}}_p\mathbf{X}^{\mathrm{H}} \in \mathbb{C}^{K\times M}$ as

$$\widetilde{\mathbf{R}}_p = \mathbf{WUGD}_p(\mathbf{\Phi})\mathbf{H} + \widetilde{\mathbf{Z}}_p, \tag{7}$$

where $\widetilde{\mathbf{Z}}_p = \bar{\mathbf{Z}}_p\mathbf{X}^{\mathrm{H}} \in \mathbb{C}^{K\times M}$. Eq. (7) can be rewritten as

$$\widetilde{\mathbf{R}}_p = \mathbf{FD}_p(\mathbf{\Phi})\mathbf{H} + \widetilde{\mathbf{Z}}_p, \tag{8}$$

where $\mathbf{F} = \mathbf{WUG} \in \mathbb{C}^{K\times N}$. We define the noiseless version of $\widetilde{\mathbf{R}}_p$ as $\mathbf{R}_p = \mathbf{FD}_p(\mathbf{\Phi})\mathbf{H} \in \mathbb{C}^{K\times M}$.

According to the conditions of multi-modal PARAFAC decomposition [35], the matrix $\mathbf{R}_p$ is the p-th frontal matrix slice of a corresponding three-way tensor $\mathcal{R} \in \mathbb{C}^{K\times M\times P}$, and each entry of $\mathbf{R}_p$, denoted as $r_{k,m,p}$, $1 \le k \le K$ and $1 \le m \le M$, is given by

$$r_{k,m,p} = [\mathbf{R}_p]_{k,m} = \sum_{n=1}^{N} [\mathbf{F}]_{k,n} [\mathbf{H}]_{n,m} [\mathbf{\Phi}]_{p,n}, \tag{9}$$

where $[\mathbf{A}]_{i,j}$ denotes the (i,j)-th element of $\mathbf{A}$. Therefore, the noiseless received signal tensor $\mathcal{R}$ has the three PARAFAC factors of $\mathbf{F}$, $\mathbf{H}$ and $\mathbf{\Phi}$, and the PARAFAC decomposition of $\mathcal{R}$ can be expressed as [26]

$$\mathcal{R} = \mathcal{I}_{3,N} \times_1 \mathbf{F} \times_2 \mathbf{H} \times_3 \mathbf{\Phi}. \tag{10}$$

It can easily be seen that the PARAFAC decomposition of $\mathcal{R}$ is trilinear in $\mathbf{F}$, $\mathbf{H}$ and $\mathbf{\Phi}$ and, therefore, the three modes of $\mathcal{R}$ can be derived as [35]

$$\mathbf{R}^1 = \left(\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi}\right)\mathbf{F}^{\mathrm{T}} \in \mathbb{C}^{PM\times K}, \tag{11}$$

$$\mathbf{R}^2 = \left(\mathbf{\Phi} \odot \mathbf{F}\right)\mathbf{H} \in \mathbb{C}^{KP\times M}, \tag{12}$$

$$\mathbf{R}^3 = \left(\mathbf{F} \odot \mathbf{H}^{\mathrm{T}}\right)\mathbf{\Phi}^{\mathrm{T}} \in \mathbb{C}^{MK\times P}, \tag{13}$$

in which $\odot$ represents the Khrtri-Rao product. In the following, by exploiting this algebraic structure of the PARAFAC decomposition, we formulate a method to estimate $\mathbf{G}$ and $\mathbf{H}$.

III. Proposed PARAFAC-Decomposition Based Method

In this section, by utilizing the PARAFAC decomposition formulated in Section II, an iterative estimation algorithm, called UAMP, is derived to estimate the channel matrices $\mathbf{G}$ and $\mathbf{H}$. The feasibility conditions of the proposed UAMP algorithm are also analyzed.

A. UAMP Algorithm for Channel Estimation

Similar to the multi-modal PARAFAC decomposition of the noiseless received signal matrices $\mathbf{R}_p$, $\forall p$, all the P noisy received signal matrices $\widetilde{\mathbf{R}}_p$ in (7) have the corresponding three-way tensor $\widetilde{\mathcal{R}} \in \mathbb{C}^{K\times M\times P}$ given by

$$\widetilde{\mathcal{R}} = \mathcal{R} + \widetilde{\mathcal{Z}}, \tag{14}$$

where $\widetilde{\mathcal{Z}} \in \mathbb{C}^{K\times M\times P}$ is the three-way tensor that corresponds to all the P AWGN matrices $\widetilde{\mathbf{Z}}_p$ in (7). Similar to the unfolded form for $\mathcal{R}$ given in (11)-(13), the noisy version $\widetilde{\mathcal{R}}$ has the unfolding form as follows:

$$\widetilde{\mathbf{R}}^1 = \left(\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi}\right)\mathbf{F}^{\mathrm{T}} + \widetilde{\mathbf{Z}}^1 \in \mathbb{C}^{PM\times K}, \tag{15}$$

$$\widetilde{\mathbf{R}}^2 = \left(\mathbf{\Phi} \odot \mathbf{F}\right)\mathbf{H} + \widetilde{\mathbf{Z}}^2 \in \mathbb{C}^{KP\times M}, \tag{16}$$

$$\widetilde{\mathbf{R}}^3 = \left(\mathbf{F} \odot \mathbf{H}^{\mathrm{T}}\right)\mathbf{\Phi}^{\mathrm{T}} + \widetilde{\mathbf{Z}}^3 \in \mathbb{C}^{MK\times P}. \tag{17}$$

The k-th column of $\widetilde{\mathbf{R}}^1$, denoted as $[\widetilde{\mathbf{R}}^1]_{:,k}$, $1 \le k \le K$, is given by

$$[\widetilde{\mathbf{R}}^1]_{:,k} = \mathbf{A}^1[\mathbf{F}^{\mathrm{T}}]_{:,k} + [\widetilde{\mathbf{Z}}^1]_{:,k}, \tag{18}$$

where $\mathbf{A}^1 = \mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi}$. Similarly, the m-th column of $\widetilde{\mathbf{R}}^2$, denoted as $[\widetilde{\mathbf{R}}^2]_{:,m}$, $1 \le m \le M$, is given by

$$[\widetilde{\mathbf{R}}^2]_{:,m} = \mathbf{A}^2[\mathbf{H}]_{:,m} + [\widetilde{\mathbf{Z}}^2]_{:,m}, \tag{19}$$

where $\mathbf{A}^2 = \mathbf{\Phi} \odot \mathbf{F}$. We regard $\widetilde{\mathbf{R}}^1$ and $\widetilde{\mathbf{R}}^2$ as the observation matrices, while $\mathbf{A}^1$ and $\mathbf{A}^2$ as the measurement matrices.

The variant of approximate message passing based on an unitary transform [36], i.e., UAMP, is leveraged for channel estimation in this paper, which achieves not only low complexity but also high robustness. Denote $\widehat{\mathbf{G}}_{(i)}$ and $\widehat{\mathbf{H}}_{(i)}$ as the estimates of $\mathbf{G}$ and $\mathbf{H}$ for the i-th iteration, respectively. Algorithm 1 presents the step-by-step algorithmic process for the proposed UAMP-based iterative channel estimation approach. More explanations of this UAMP algorithm are provided below.

1) Initial Channel Estimation: To provide a feasible $\mathbf{\Phi}$ for the UAMP algorithm, we first generate an $N \times N$ DFT matrix and then choose its first P rows as $\mathbf{\Phi}$, i.e., $\mathbf{\Phi}$ is a $P \times N$ DFT matrix. We initiate $\widehat{\mathbf{G}}_{(0)}$ and $\widehat{\mathbf{H}}_{(0)}$ through (1) and (3), and then obtain $\widehat{\mathbf{F}}_{(0)} = \mathbf{WU}\widehat{\mathbf{G}}_{(0)}$, $\widehat{\mathbf{A}}^1_{(0)} = \widehat{\mathbf{H}}^{\mathrm{T}}_{(0)} \odot \mathbf{\Phi}$ and $\widehat{\mathbf{A}}^2_{(0)} = \mathbf{\Phi} \odot \widehat{\mathbf{F}}_{(0)}$ according to (8), (18) and (19), respectively. Recalling (1), (3) and (8), it is obvious that the elements of $\mathbf{F}$ and $\mathbf{H}$ obey complex Gaussian distribution with mean 0 and variance 1. Therefore, the prior means $\overline{\gamma}^1_{k,\mathrm{mean}}$ and variances $\overline{\tau}^1_{k,\mathrm{var}}$ of $[\mathbf{F}^{\mathrm{T}}]_{:,k}$, $1 \le k \le K$ are $\mathbf{0}_N$ and $\mathbf{1}_N$, the prior means $\overline{\gamma}^2_{m,\mathrm{mean}}$ and variances $\overline{\tau}^2_{m,\mathrm{var}}$ of $[\mathbf{H}]_{:,m}$, $1 \le m \le M$ are also $\mathbf{0}_N$ and $\mathbf{1}_N$, where $\mathbf{0}_N$ and $\mathbf{1}_N$ represent all-zero and all-one column vectors, respectively. The posterior means $\widetilde{\gamma}^1_{k,\mathrm{mean}}$ and variances $\widetilde{\tau}^1_{k,\mathrm{var}}$ of $[\widehat{\mathbf{F}}^{\mathrm{T}}]_{:,k}$, $1 \le k \le K$ as well as the posterior means $\widetilde{\gamma}^2_{m,\mathrm{mean}}$ and variances $\widetilde{\tau}^2_{m,\mathrm{var}}$ of $[\widehat{\mathbf{H}}]_{:,m}$, $1 \le m \le M$ are initialized the same as the corresponding prior means and variances.

2) UAMP Channel Estimation Update: Based on singular value decomposition (SVD) and unitary transform, $\widehat{\mathbf{F}}$ can be constructed by steps from 2 to 3 in Algorithm 1. In

Algorithm 1 Proposed Channel Estimation Based on UAMP

Input: Give feasible $\mathbf{\Phi}$, observation matrices $\widetilde{\mathbf{R}}^1$ and $\widetilde{\mathbf{R}}^2$, prior means $\overline{\gamma}^1_{k,\text{mean}}$, $\overline{\gamma}^2_{m,\text{mean}}$ and variances $\overline{\tau}^1_{k,\text{var}}$, $\overline{\tau}^2_{m,\text{var}}$ ($1 \le k \le K$, $1 \le m \le M$), noise variance σ^2, termination threshold $\delta = 10^{-5}$, maximum number of UAMP iterations $I_{\max}$, maximum number of updates for posteriors $J_{\max}$;

Initialize $\widehat{\mathbf{H}}_{(0)}$ and $\widehat{\mathbf{F}}_{(0)}$, measurement matrices $\widehat{\mathbf{A}}^1_{(0)}$ and $\widehat{\mathbf{A}}^2_{(0)}$, posterior means $\widetilde{\gamma}^1_{(0),k,\text{mean}}$, $\widetilde{\gamma}^2_{(0),m,\text{mean}}$ and variances $\widetilde{\tau}^1_{(0),k,\text{var}}$, $\widetilde{\tau}^2_{(0),m,\text{var}}$ ($1 \le k \le K$, $1 \le m \le M$);

Set iterative index $i = 0$;

1: **do** $i = i + 1$;

2: Compute SVD $\widehat{\mathbf{A}}^1_{(i-1)} = \mathbf{U}^1\mathbf{\Lambda}^1\mathbf{V}^1$;

3: **for** $1 \le k \le K$; % compute $\widehat{\mathbf{F}}^{\mathrm{T}}_{(i)}$ column by column

3.a: Unitary transform $\bar{\mathbf{r}}^1 = (\mathbf{U}^1)^{\mathrm{H}}[\widetilde{\mathbf{R}}^1]_{:,k}$;

3.b: Set $\mathbf{\Omega}^1 = \mathbf{\Lambda}^1\mathbf{V}^1$, $\mathbf{s}^1_{(-1)} = \mathbf{0}_{PM}$, iterative index $j = 0$;

3.c: **Repeat**; % update posterior means and variances

3.d: Update $\boldsymbol{\tau}^1_t = |\mathbf{\Omega}^1|^2\,\widetilde{\boldsymbol{\tau}}^1_{(j),k,\text{var}}$;

3.e: Update $\mathbf{t}^1 = \mathbf{\Omega}^1\widetilde{\boldsymbol{\gamma}}^1_{(j),k,\text{mean}} - \boldsymbol{\tau}^1_t \circ \mathbf{s}^1_{(j-1)}$;

3.f: Update $\boldsymbol{\tau}^1_s = 1 \circ / \left(\boldsymbol{\tau}^1_t + \sigma^2\mathbf{1}_{PM}\right)$;

3.g: Update $\mathbf{s}^1_{(j)} = \boldsymbol{\tau}^1_s \circ (\bar{\mathbf{r}}^1 - \mathbf{t}^1)$;

3.h: Update $1 \circ / \boldsymbol{\tau}^1_q = |(\mathbf{\Omega}^1)^{\mathrm{H}}|^2\,\boldsymbol{\tau}^1_s$;

3.i: Update $\mathbf{q}^1 = \widetilde{\boldsymbol{\gamma}}^1_{(j),k,\text{mean}} + \boldsymbol{\tau}^1_q \circ \left((\mathbf{\Omega}^1)^{\mathrm{H}}\mathbf{s}^1_{(j)}\right)$;

3.j: Update $\widetilde{\boldsymbol{\tau}}^1_{(j+1),k,\text{var}} = 1 \circ / \left(1 \circ / \overline{\boldsymbol{\tau}}^1_{k,\text{var}} + 1 \circ / \boldsymbol{\tau}^1_q\right)$;

3.k: Update $\widetilde{\boldsymbol{\gamma}}^1_{(j+1),k,\text{mean}} = \widetilde{\boldsymbol{\tau}}^1_{(j+1),k,\text{var}} \circ \left(\mathbf{q}^1 \circ / \boldsymbol{\tau}^1_q + \overline{\boldsymbol{\gamma}}^1_{k,\text{mean}} \circ / \overline{\boldsymbol{\tau}}^1_{k,\text{var}}\right)$;

3.l: Set $j = j + 1$;

3.m: **Until** $j = J_{\max}$;

3.n: Construct $[\widehat{\mathbf{F}}^{\mathrm{T}}_{(i)}]_{:,k} = \widetilde{\boldsymbol{\gamma}}^1_{(j),k,\text{mean}}$;

end for;

4: Update $\widehat{\mathbf{G}}_{(i)}$ by $\widehat{\mathbf{G}}_{(i)} = (\mathbf{WU})^{\dagger}\widehat{\mathbf{F}}_{(i)}$;

5: Compute SVD $\widehat{\mathbf{A}}^2_{(i-1)} = \mathbf{U}^2\mathbf{\Lambda}^2\mathbf{V}^2$;

6: **for** $1 \le m \le M$; % compute $\widehat{\mathbf{H}}_{(i)}$ column by column

6.a: Unitary transform $\bar{\mathbf{r}}^2 = (\mathbf{U}^2)^{\mathrm{H}}[\widetilde{\mathbf{R}}^2]_{:,m}$;

6.b: Set $\mathbf{\Omega}^2 = \mathbf{\Lambda}^2\mathbf{V}^2$, $\mathbf{s}^2_{(-1)} = \mathbf{0}_{KP}$, iterative index $j = 0$;

6.c: **Repeat**; % update posterior means and variances

6.d: Update $\boldsymbol{\tau}^2_t = |\mathbf{\Omega}^2|^2\,\widetilde{\boldsymbol{\tau}}^2_{(j),m,\text{var}}$;

6.e: Update $\mathbf{t}^2 = \mathbf{\Omega}^2\widetilde{\boldsymbol{\gamma}}^2_{(j),m,\text{mean}} - \boldsymbol{\tau}^2_t \circ \mathbf{s}^2_{(j-1)}$;

6.f: Update $\boldsymbol{\tau}^2_s = 1 \circ / \left(\boldsymbol{\tau}^2_t + \sigma^2\mathbf{1}_{KP}\right)$;

6.g: Update $\mathbf{s}^2_{(j)} = \boldsymbol{\tau}^2_s \circ (\bar{\mathbf{r}}^2 - \mathbf{t}^2)$;

6.h: Update $1 \circ / \boldsymbol{\tau}^2_q = |(\mathbf{\Omega}^2)^{\mathrm{H}}|^2\,\boldsymbol{\tau}^2_s$;

6.i: Update $\mathbf{q}^2 = \widetilde{\boldsymbol{\gamma}}^2_{(j),m,\text{mean}} + \boldsymbol{\tau}^2_q \circ \left((\mathbf{\Omega}^2)^{\mathrm{H}}\mathbf{s}^2_{(j)}\right)$;

6.j: Update $\widetilde{\boldsymbol{\tau}}^2_{(j+1),m,\text{var}} = 1 \circ / \left(1 \circ / \overline{\boldsymbol{\tau}}^2_{m,\text{var}} + 1 \circ / \boldsymbol{\tau}^2_q\right)$;

6.k: Update $\widetilde{\boldsymbol{\gamma}}^2_{(j+1),m,\text{mean}} = \widetilde{\boldsymbol{\tau}}^2_{(j+1),m,\text{var}} \circ \left(\mathbf{q}^2 \circ / \boldsymbol{\tau}^2_q + \overline{\boldsymbol{\gamma}}^2_{m,\text{mean}} \circ / \overline{\boldsymbol{\tau}}^2_{m,\text{var}}\right)$;

6.l: Set $j = j + 1$;

6.m: **Until** $j = J_{\max}$;

6.n: Construct $[\widehat{\mathbf{H}}_{(i)}]_{:,m} = \widetilde{\boldsymbol{\gamma}}^2_{(j),m,\text{mean}}$;

end for;

7: Update $\widehat{\mathbf{A}}^1_{(i)} = \widehat{\mathbf{H}}^{\mathrm{T}}_{(i)} \odot \mathbf{\Phi}$;

8: Update $\widehat{\mathbf{A}}^2_{(i)} = \mathbf{\Phi} \odot \widehat{\mathbf{F}}_{(i)}$;

9: **Until** $i = I_{\max}$ or convergence condition (21) is met;

Output: $\widehat{\mathbf{G}}_{(i)}$ and $\widehat{\mathbf{H}}_{(i)}$ are estimates of $\mathbf{G}$ and $\mathbf{H}$.

particular, the inner loop 3.c-3.m involves $J_{\max}$ updates of the posterior means $\widetilde{\gamma}^1_{k,\text{mean}}$ and variances $\widetilde{\tau}^1_{k,\text{var}}$ of $[\widehat{\mathbf{F}}^{\mathrm{T}}]_{:,k}$ for $1 \le k \le K$. Here, $|\mathbf{C}|^2$ represents the elementwise magnitude squared operation on $\mathbf{C}$, and we use $\circ$ and $\circ/$ to denote the elementwise product and elementwise division operations, respectively. Then $\widehat{\mathbf{G}}$ is updated as

$$\widehat{\mathbf{G}} = (\mathbf{WU})^{\dagger}\widehat{\mathbf{F}}, \tag{20}$$

where $\mathbf{C}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{C}$. In a similar way, $\widehat{\mathbf{H}}$ can be constructed by steps from 5 to 6 in Algorithm 1, and similarly the inner loop 6.c-6.m performs $J_{\max}$ updates of the posterior means $\widetilde{\gamma}^2_{m,\text{mean}}$ and variances $\widetilde{\tau}^2_{m,\text{var}}$ of $[\widehat{\mathbf{H}}]_{:,m}$ for $1 \le m \le M$.

3) Iteration Termination Criterion: The UAMP procedure converges either when the channel estimation errors become sufficiently small, specifically, when

$$\begin{cases} \left\|\widehat{\mathbf{H}}_{(i)} - \widehat{\mathbf{H}}_{(i-1)}\right\|^2_F \left\|\widehat{\mathbf{H}}_{(i)}\right\|^{-2}_F \le \delta \\ \text{and } \|\widehat{\mathbf{G}}_{(i)} - \widehat{\mathbf{G}}_{(i-1)}\|^2_F \left\|\widehat{\mathbf{G}}_{(i)}\right\|^{-2}_F \le \delta, \end{cases} \tag{21}$$

where δ is a small positive threshold, e.g., $\delta = 10^{-5}$. Or the algorithm is terminated when a preset maximum number of iterations $I_{\max}$ is exhausted, i.e., when $i = I_{\max}$.

The computational complexity of the proposed UAMP algorithm for estimating the channels $\mathbf{F}$ and $\mathbf{H}$ is next analyzed. It can be seen from Algorithm 1 that the complexity of SVD for $\widehat{\mathbf{A}}^1$ and $\widehat{\mathbf{A}}^2$ are $c_1 = \min\{N(PM)^2, N^2PM\}$ and $c_2 = \min\{N(PK)^2, N^2PK\}$, respectively, and the complexity of the main body for $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{H}}$ are $\mathcal{O}(2N^2PM + 2P^2M^2)$ and $\mathcal{O}(2N^2PK + 2P^2K^2)$, respectively. In summary, the total complexity of the proposed algorithm is $\mathcal{O}(c_1 + c_2 + 2N^2P(M+K) + 2P^2(M^2+K^2))$ per iteration. By contrast, the complexity of the ALS algorithm [27] is $\mathcal{O}(2N^3 + 4N^2P(M+K) + 4NPMK)$ per iteration. Since the number of RIS reflecting elements N is much larger than the other system parameters, the complexity of our proposed algorithm is on the order of N^2 and the complexity of the ALS [27] is on the order of N^3. Therefore our algorithm imposes significantly less computational complexity than the ALS.

B. Uniqueness Condition

The feasibility of our UAMP algorithm depends on the uniqueness of the PARAFAC decomposition. To establish the uniqueness condition for our UAMP algorithm, we employ the uniqueness conditions for the PARAFAC model [37]. Based on the established identifiability condition, we can infer useful system design suggestions. Denote the Kruskal rank or k-rank of $\mathbf{A}$ by $k_{\mathbf{A}}$. By leveraging the identifiability theorem of the PARAFAC model, it can be demonstrated that if

$$k_{\mathbf{F}} + k_{\mathbf{H}} + k_{\mathbf{\Phi}} \ge 2N + 2, \tag{22}$$

the decomposition of $\mathcal{R}$ is unique up to permutation and scaling ambiguities [38]. In other words, if two triples $\left(\mathbf{F}, \mathbf{H}, \mathbf{\Phi}\right)$ and $\left(\overline{\mathbf{F}}, \overline{\mathbf{H}}, \overline{\mathbf{\Phi}}\right)$ both decompose $\mathcal{R}$ into N rank-one arrays, there must exist an $N \times N$ permutation matrix $\mathbf{\Pi}$ and three $N \times N$ diagonal scaling matrices $\mathbf{\Lambda}_i$ with $i = 1, 2, 3$ and

$\mathbf{\Lambda}_1\mathbf{\Lambda}_2\mathbf{\Lambda}_3 = \mathbf{I}_N$, such that $\overline{\mathbf{F}} = \mathbf{F}\mathbf{\Pi}\mathbf{\Lambda}_1$, $\overline{\mathbf{H}} = \mathbf{H}\mathbf{\Pi}\mathbf{\Lambda}_2$ and $\overline{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{\Pi}\mathbf{\Lambda}_3$.

Inequality (22) establishes the sufficient condition for the identifiability of $(\mathbf{F}, \mathbf{H}, \mathbf{\Phi})$. Because the azimuth and elevation angles between the users and the RIS are usually different, $\mathbf{H}$ has full k-rank. As $\mathbf{\Phi}$ is a DFT matrix, it also has full k-rank. Since $\mathbf{W}$ and $\mathbf{U}$ both have full k-rank according to (8), the rank of $\mathbf{F}$ equals to that of $\mathbf{G}$, which is rank deficient. Moreover, the k-rank of $\mathbf{G}$ is equal to the number of paths between the RIS and the BS. Hence, (22) becomes

$$\min\{L_G, N\} + \min\{N, M\} + \min\{P, N\} \geq 2N + 2. \quad (23)$$

The number of RIS elements N in a real deployment is very large, and N is greater than the number of paths L_G between the RIS and the BS or the number of users M. Also it is obvious that $P \leq N$ since $\mathbf{\Phi}$ is the first P rows chosen from an $N \times N$ DFT matrix. Therefore, the uniqueness condition for our PARAFAC decomposition is not easily satisfied.

To resolve this difficulty, we partition the N-element RIS in groups of non-overlapping sub-cells. For example, consider the system with $L_G = 5$, $M = 16$ and an RIS of $N = 128$ elements, and we choose $P = N$. Since $N \gg \max\{L_G, M\}$, the condition (23) is not met. We can partition the 128-element RIS into 8 non-overlapping 16-element sub-RISs. Each 16-element sub-RIS meets the condition (23). The proposed UAMP algorithm can be applied to estimate the channels related to the 8 sub-RISs. Concatenating the 8 channel estimates for all the 8 sub-RISs properly yields the desired channel estimation for the whole 128-element RIS. Overall, the useful system design guidelines can be obtained by (23), in which the system parameters L_G, M, P and N should meet the uniqueness condition for our PARAFAC decomposition.

IV. Cramé-Rao Lower Bound Analysis

For the zero-mean AWGN received signal model (8), the proposed UAMP algorithm provides the maximum likelihood estimations of $\mathbf{G}$ and $\mathbf{H}$, which is asymptotically unbiased [39]. On the other hand, the CRB yields theoretically lower bound on the achieved variances of an unbiased estimate. Therefore, we derive the CRBs for the estimations of $\mathbf{G}$ and $\mathbf{H}$ obtained by our UAMP algorithm, which offers the ultimate achievable estimation accuracy for our UAMP algorithm. As pointed out in Subsection III-B, the trilinear decomposition model suffers from the inherent permutation and scale ambiguity. To remove the scaling ambiguity, thereby assisting the derivation of the CRBs, we fix the first column of $\mathbf{H}$ to $[\mathbf{H}]_{:,1} = \mathbf{1}_N$, where $\mathbf{1}_N$ is the N-dimensional vector with all the elements equal to 1 and $[\mathbf{A}]_{:,j}$ denotes the j-th column of $\mathbf{A}$. Hence, the total number of the unknown complex parameters is reduced to $(K + M - 1)N$.

Consider the noisy version $\widetilde{\mathcal{R}}$ of the $K \times M \times P$ three-way tensor with typical element

$$\tilde{r}_{k,m,p} = \sum_{n=1}^{N} [\mathbf{F}]_{k,n} [\mathbf{H}]_{n,m} [\mathbf{\Phi}]_{p,n} + \tilde{z}_{k,m,p}. \quad (24)$$

Recalling (15)-(17), the likelihood of $\widetilde{\mathcal{R}}$ can be expressed in the following three equivalent ways:

$$\begin{aligned} L(\widetilde{\mathcal{R}}) &= \frac{1}{(\pi\sigma^2)^{KMP}} e^{-\frac{1}{\sigma^2}\sum_{k=1}^{K}\left\|\left[\widetilde{\mathbf{R}}^1\right]_{:,k} - \left(\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi}\right)\mathbf{a}_k\right\|^2} \\ &= \frac{1}{(\pi\sigma^2)^{KMP}} e^{-\frac{1}{\sigma^2}\sum_{m=1}^{M}\left\|\left[\widetilde{\mathbf{R}}^2\right]_{:,m} - (\mathbf{\Phi} \odot \mathbf{F})\mathbf{b}_m\right\|^2} \\ &= \frac{1}{(\pi\sigma^2)^{KMP}} e^{-\frac{1}{\sigma^2}\sum_{p=1}^{P}\left\|\left[\widetilde{\mathbf{R}}^3\right]_{:,p} - \left(\mathbf{F} \odot \mathbf{H}^{\mathrm{T}}\right)\mathbf{c}_p\right\|^2}, \end{aligned} \quad (25)$$

where $\mathbf{a}_k = \left[\mathbf{F}^{\mathrm{T}}\right]_{:,k}$, $\mathbf{b}_m = [\mathbf{H}]_{:,m}$ and $\mathbf{c}_p = \left[\mathbf{\Phi}^{\mathrm{T}}\right]_{:,p}$. Further collect all the unknown parameters $\mathbf{a}_k$, $\forall k$, and $\mathbf{b}_m$, $\forall m$, into the equivalent $1 \times 2(K + M - 1)N$ complex parameter vector

$$\boldsymbol{\vartheta} = \left[\mathbf{a}_1^{\mathrm{T}}, \ldots, \mathbf{a}_K^{\mathrm{T}}, \mathbf{b}_2^{\mathrm{T}}, \ldots, \mathbf{b}_M^{\mathrm{T}}, \mathbf{a}_1^{\mathrm{H}}, \ldots, \mathbf{a}_K^{\mathrm{H}}, \mathbf{b}_2^{\mathrm{H}}, \ldots, \mathbf{b}_M^{\mathrm{H}}\right]. \quad (26)$$

Then we can express the log-likelihood function of $\boldsymbol{\vartheta}$ as

$$\begin{aligned} f(\boldsymbol{\vartheta}) &= -KMP\ln(\pi\sigma^2) - \frac{1}{\sigma^2}\sum_{k=1}^{K}\left\|\left[\widetilde{\mathbf{R}}^1\right]_{:,k} - (\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi})\mathbf{a}_k\right\|^2 \\ &= -KMP\ln\left(\pi\sigma^2\right) - \frac{1}{\sigma^2}\sum_{m=1}^{M}\left\|\left[\widetilde{\mathbf{R}}^2\right]_{:,m} - (\mathbf{\Phi} \odot \mathbf{F})\mathbf{b}_m\right\|^2. \end{aligned} \quad (27)$$

The complex Fisher information matrix (FIM) of $\boldsymbol{\vartheta}$ is defined by

$$\mathbf{\Omega}(\boldsymbol{\vartheta}) = \mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)^{\mathrm{H}}\left(\frac{\partial f(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)\right\}, \quad (28)$$

in which the partial derivatives of $f(\boldsymbol{\vartheta})$ with respect to the unknown parameters are given by

$$\begin{cases} \frac{\partial f(\boldsymbol{\vartheta})}{\partial a_{k,n}} = \frac{1}{\sigma^2}\left(\left[\widetilde{\mathbf{R}}^1\right]_{:,k} - (\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi})\mathbf{a}_k\right)^{\mathrm{H}}(\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi})\mathbf{e}_n, \\ \frac{\partial f(\boldsymbol{\vartheta})}{\partial b_{m,n}} = \frac{1}{\sigma^2}\left(\left[\widetilde{\mathbf{R}}^2\right]_{:,m} - (\mathbf{\Phi} \odot \mathbf{H})\mathbf{b}_m\right)^{\mathrm{H}}(\mathbf{\Phi} \odot \mathbf{H})\mathbf{e}_n, \\ \frac{\partial f(\boldsymbol{\vartheta})}{\partial a_{k,n}^*} = \left(\frac{\partial f(\boldsymbol{\vartheta})}{\partial a_{k,n}}\right)^*, \ \frac{\partial f(\boldsymbol{\vartheta})}{\partial b_{m,n}^*} = \left(\frac{\partial f(\boldsymbol{\vartheta})}{\partial b_{m,n}}\right)^*, \end{cases} \quad (29)$$

where $\mathbf{e}_n$ is the n-th unit coordinate vector.

Because $\tilde{z}_{k,m,p}$ is independent in the dimensions of k, m and p, we have

$$\mathbb{E}\left\{\tilde{z}_{k_1,m_1,p_1}\tilde{z}_{k_2,m_2,p_2}^*\right\} = \sigma^2\delta_{k_1,k_2}\delta_{m_1,m_2}\delta_{p_1,p_2}, \quad (30)$$

$$\mathbb{E}\left\{\tilde{z}_{k_1,m_1,p_1}\tilde{z}_{k_2,m_2,p_2}\right\} = \mathbb{E}\left\{\tilde{z}_{k_1,m_1,p_1}^*\tilde{z}_{k_2,m_2,p_2}^*\right\} = 0. \quad (31)$$

Therefore, it can be shown that the FIM can be written as

$$\mathbf{\Omega}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}^* \end{bmatrix}, \quad (32)$$

where the subblock Hessian matrix $\mathbf{\Psi}$ and the all-zero matrix $\mathbf{0}$ are both of size $(K+M-1)N \times (K+M-1)N$. Based on (29) and noting that $\mathbb{E}\Big\{\left(\left[\widetilde{\mathbf{R}}^1\right]_{:,k} - (\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi})\mathbf{a}_k\right)\left(\left[\widetilde{\mathbf{R}}^1\right]_{:,k} - (\mathbf{H}^{\mathrm{T}} \odot \mathbf{\Phi})\mathbf{a}_k\right)^{\mathrm{H}}\Big\} = \sigma^2\mathbf{I}_N$ and $\mathbb{E}\Big\{\left(\left[\widetilde{\mathbf{R}}^2\right]_{:,m} - (\mathbf{\Phi} \odot \mathbf{F})\mathbf{b}_m\right)\left(\left[\widetilde{\mathbf{R}}^2\right]_{:,m} -$

$(\mathbf{\Phi}\odot\mathbf{F})\mathbf{b}_m)^{\mathrm{H}}\Big\} = \sigma^2\mathbf{I}_N$ as well as (30) and (31), the elements of $\mathbf{\Psi}$ are obtained as

$$\mathbb{E}\left\{\frac{\partial f(\boldsymbol{\vartheta})}{\partial a^*_{k_1,n_1}}\frac{\partial f(\boldsymbol{\vartheta})}{\partial a_{k_2,n_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}^{\mathrm{H}}_{n_1}\left(\mathbf{H}^{\mathrm{T}}\odot\mathbf{\Phi}\right)^{\mathrm{H}}\left(\mathbf{H}^{\mathrm{T}}\odot\mathbf{\Phi}\right)\mathbf{e}_{n_2}\delta_{k_1,k_2}, \tag{33}$$

$$\mathbb{E}\left\{\frac{\partial f(\boldsymbol{\vartheta})}{\partial b^*_{m_1,n_1}}\frac{\partial f(\boldsymbol{\vartheta})}{\partial b_{m_2,n_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}^{\mathrm{H}}_{n_1}\left(\mathbf{\Phi}\odot\mathbf{F}\right)^{\mathrm{H}}\left(\mathbf{\Phi}\odot\mathbf{F}\right)\mathbf{e}_{n_2}\delta_{m_1,m_2}, \tag{34}$$

$$\mathbb{E}\left\{\frac{\partial f(\boldsymbol{\vartheta})}{\partial a^*_{k_1,n_1}}\frac{\partial f(\boldsymbol{\vartheta})}{\partial b_{m,n_2}}\right\} = \frac{1}{\sigma^4}\mathbf{e}^{\mathrm{H}}_{n_1}\left(\mathbf{H}^{\mathrm{T}}\odot\mathbf{\Phi}\right)^{\mathrm{H}}\mathbb{E}\left\{\left[\widetilde{\mathbf{Z}}^1\right]_{:,k}\left[\widetilde{\mathbf{Z}}^2\right]^{\mathrm{H}}_{:,m}\right\}\left(\mathbf{\Phi}\odot\mathbf{F}\right)\mathbf{e}_{n_2}, \tag{35}$$

for $1 \le k_1, k_2 \le K$ and $1 \le n_1, n_2 \le, N$. In (35), the covariance matrix of the weighted channel AWGN is given by

$$\mathbb{E}\left\{\left[\widetilde{\mathbf{Z}}^1\right]_{:,k}\left[\widetilde{\mathbf{Z}}^2\right]^{\mathrm{H}}_{:,m}\right\} = \sigma^2\begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \\ & & & \ddots & & & & & & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \leftarrow(m-1)P+1 \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \leftarrow(m-1)P+2 \\ & & & \ddots & & & & & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \leftarrow(m-1)P+P \\ & & & \ddots & & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \end{bmatrix} \\ \qquad\qquad \uparrow \quad \cdots \quad \uparrow \quad \cdots \quad \uparrow \\ \qquad\qquad k \quad K+k \quad (P-1)K+k \tag{36}$$

Hence, the CRB for the unbiased estimator of the unknown channels $\boldsymbol{\vartheta}$ is expressed as

$$\mathbf{\Omega}^{-1}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathbf{\Psi}^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{\Psi}^{-1}\right)^* \end{bmatrix}, \tag{37}$$

where the inverse of $\mathbf{\Psi}$ can be partitioned into

$$\mathbf{\Psi}^{-1} = \begin{bmatrix} \mathbf{CRB_F} & \mathbf{K} \\ \mathbf{K}^{\mathrm{H}} & \mathbf{CRB_H} \end{bmatrix}, \tag{38}$$

in which $\mathbf{CRB_F} \in \mathbb{C}^{KN\times KN}$ is the CRB for the channel estimation of $\mathbf{F}$, and $\mathbf{CRB_H} \in \mathbb{C}^{(M-1)N\times(M-1)N}$ is the CRB for the channel estimation of $\mathbf{H}$, while the remaining sub-matrix $\mathbf{K} \in \mathbb{C}^{KN\times(M-1)N}$. Further divide the Hermitian matrix $\mathbf{\Psi}$ into

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi}_1 & \mathbf{\Psi}_2 \\ \mathbf{\Psi}_2^{\mathrm{H}} & \mathbf{\Psi}_3 \end{bmatrix}, \tag{39}$$

with $\mathbf{\Psi}_1 \in \mathbb{C}^{KN\times KN}$, $\mathbf{\Psi}_2 \in \mathbb{C}^{(M-1)N\times(M-1)N}$ and $\mathbf{\Psi}_3 \in \mathbb{C}^{KN\times(M-1)N}$. Then the CRBs can be derived derived as

$$\mathbf{CRB_F} = \left(\mathbf{\Psi}_1 - \mathbf{\Psi}_2\mathbf{\Psi}_3^{-1}\mathbf{\Psi}_2^{\mathrm{H}}\right)^{-1}, \tag{40}$$

$$\mathbf{CRB_H} = \left(\mathbf{\Psi}_3 - \mathbf{\Psi}_2^{\mathrm{H}}\mathbf{\Psi}_1^{-1}\mathbf{\Psi}_2\right)^{-1}. \tag{41}$$

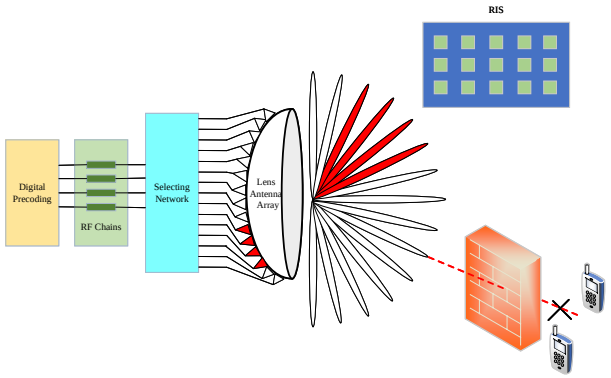

Fig. 3. DL of the RIS-aided mmWave multiuser MISO system with lens antenna array and beam selection at the BS.

According to (8), $\mathbf{F} = \mathbf{WUG}$. Because $\mathbf{W}$ and $\mathbf{U}$ are constant with respect to $\boldsymbol{\vartheta}$, the lower bound of CRB for the estimation of $\mathbf{G}$, denoted by $\mathbf{CRB_G}$, can be derived as

$$\mathbf{CRB_G} = \frac{\left(\mathbf{\Psi}_1 - \mathbf{\Psi}_2\mathbf{\Psi}_3^{-1}\mathbf{\Psi}_2^{\mathrm{H}}\right)^{-1}}{\|\mathbf{WU}\|_F^2}. \tag{42}$$

V. DL Sum Rate Performance Analysis

The DL transmission system is illustrated in Fig. 3. Assume that all the operations are completed within the channel coherence time during which the CSI remains unchanged. Define the RIS phase configuration matrix as $\mathbf{\Theta} \in \mathbb{C}^{N\times N}$ and the binary selecting matrix $\mathbf{B}$ of the size $K \times K_{RF}$ whose elements take the value 0 or 1. Let $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_N]$ and $\mathbf{\Theta} = \mathrm{diag}\left(\left[\beta e^{\mathrm{j}\theta_1}, \cdots, \beta e^{\mathrm{j}\theta_n}, \cdots, \beta e^{\mathrm{j}\theta_N}\right]\right)$, where $\theta_n \in [0,\ 2\pi)$ and $\beta \in [0,\ 1]$.[1] The BS transmits the unit power complex-valued information symbol s_m, chosen from a discrete constellation set, to the m-th user with the precoding vector $\mathbf{p}_m$, for $1 \le m \le M$. Let the BS transmit power be ρ. Then the transmitted signal at the BS is given by $\mathbf{x} = \sqrt{\rho}\sum_{m=1}^{M}\mathbf{p}_m s_m \in \mathbb{C}^{K_{RF}\times 1}$. Therefore, the received signals at the M mobile users, denoted as $\mathbf{y} \in \mathbb{C}^{M\times 1}$, are given by

$$\mathbf{y} = \mathbf{H}^{\mathrm{H}}\mathbf{\Theta}\widetilde{\mathbf{G}}^{\mathrm{H}}\mathbf{Bx} + \mathbf{n}, \tag{43}$$

where $\mathbf{n} \in \mathbb{C}^{M\times 1}$ is the DL channel's complex AWGN vector with zero mean and covariance matrix $\sigma_{DL}^2\mathbf{I}_M$, i.e., $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_{DL}^2\mathbf{I}_M\right)$, and $\widetilde{\mathbf{G}}^{\mathrm{H}} = (\mathbf{UG})^{\mathrm{H}} \in \mathbb{C}^{N\times K}$ defines the DL beamspace channel matrix whose K columns corresponding to K orthogonal beams. It can be seen that the received signals and hence the achievable DL sum rate depends on the beam selection method represented by the beam selection matrix $\mathbf{B}$, the RIS setting represented by it phase configuration matrix $\mathbf{\Theta}$, and the precoding method employed by the BS. To achieve the best attainable DL sum rate for our UAMP channel estimate, we propose a MP beam selection method, optimize the RIS's phase shift matrix and adopt the ZF precoding, using a centralized algorithm [40].

[1]In this paper, each element on the RIS is designed to maximize signal reflection. Thus, we set $\beta = 1$.

A. MP Beam Selection

It is well-known that the beamspace channel $\widetilde{\mathbf{G}}^{\mathrm{H}}$ has a sparse structure because mmWave signal propagation has limited number of dominant scatters. More specifically, only a few elements of $\widetilde{\mathbf{G}}^{\mathrm{H}}$ have dominant values near the line-of-sight direction from the BS to the RIS, as illustrated by the red-coloured beams in Fig. 3. We propose a MP beam selection scheme to select K_{RF} beams by taking advantage of this sparse property of the beamspace channel. It can also easily be seen that in order to match the RIS with K_{RF} available RF chains, thereby fully utilizing the available resources of K_{RF} RF chains, the beam selection matrix $\mathbf{B}$ should satisfy $\left\|[\mathbf{B}]_{k,:}\right\|_0 \leq 1$, $\left\|[\mathbf{B}]_{:,k_{RF}}\right\|_0 \leq 1$ and $\sum_{k=1}^{K}[\mathbf{B}]_{k,k_{RF}} = 1$ for $1 \leq k \leq K$ and $1 \leq k_{RF} \leq K_{RF}$, where $\|\mathbf{b}\|_0$ denotes the zero norm of $\mathbf{b}$ and $[\mathbf{B}]_{i,j}$ represents the (i,j)-th element of $\mathbf{B}$.

Our proposed MP beam selection scheme selects the K_{RF} dominant or MP beams from the K selectable beams for the RIS. Specifically, we first calculate the power of the beamspace channel $\widetilde{\mathbf{G}}^{\mathrm{H}}$ in a column-wise way by $\widetilde{g}_k = \left\|\left[\widetilde{\mathbf{G}}^{\mathrm{H}}\right]_{:,k}\right\|^2$ for $1 \leq k \leq K$. Clearly, $\widetilde{g}_k$ is the sum of squares of the elements of the k-th column vector in $\widetilde{\mathbf{G}}^{\mathrm{H}}$. Then, we sort all the $\widetilde{g}_k$ values in descending order and select the top K_{RF} $\widetilde{g}_k$ which correspond to the best K_{RF} columns of the beamspace channel matrix with the highest power. Finally, the selection matrix $\mathbf{B}$ is designed based on the obtained column indices corresponding to the first K_{RF} $\widetilde{g}_k$. With this MP beam selection matrix $\mathbf{B}$, we obtain the beamspace channel matrix corresponding to the selected beams as $\widetilde{\mathbf{G}}_r = \widetilde{\mathbf{G}}^{\mathrm{H}}\mathbf{B} \in \mathbb{C}^{N\times K_{RF}}$, and the received signals at the M users become

$$\mathbf{y} = \mathbf{H}^{\mathrm{H}}\mathbf{\Theta}\widetilde{\mathbf{G}}_r\mathbf{x} + \mathbf{n}. \tag{44}$$

This MP beam selection algorithm is summarized in Algorithm 2.

Algorithm 2 Proposed MP Beam Selection Method

Input: $\widetilde{\mathbf{G}}^{\mathrm{H}}$, K_{RF};

1: Initialize $\mathbf{B} = \mathrm{zeros}(K, K_{RF})$, $\widetilde{\mathbf{g}} = \mathrm{zeros}(K, 1)$, and $\widetilde{\mathbf{g}}_p = \mathrm{zeros}(K, 1)$;
2: for $k = 1, 2, \cdots, K$ do
3: Calculate channel power $\widetilde{g}_k = \left\|\left[\widetilde{\mathbf{G}}^{\mathrm{H}}\right]_{:,k}\right\|^2$, $\widetilde{\mathbf{g}}(k,1) = \widetilde{g}_k$;
4: Arrange elements of $\widetilde{\mathbf{g}}$ in descending order, while keeping original column indexes corresponding to the rearranged elements in $\widetilde{\mathbf{g}}_p$;
5: end for
6: for $l = 1, 2, \cdots, K_{RF}$ do
7: Arrange elements of $\widetilde{\mathbf{g}}_p$ in ascending order of magnitude;
8: $\mathbf{B}(\widetilde{\mathbf{g}}_p(l), l) = 1$;
9: end for

Output: $\mathbf{B}$.

B. Optimizing RIS Phase Matrix

From the signal model (44), it can be seen that the achievable DL sum rate depends on the RIS's phase matrix $\mathbf{\Theta}$. Therefore we further exploit maximizing the total DL sum rate for all the users by optimizing the reflected beamforming of the RIS's phase shifters, given the CSI of $\mathbf{H}$ and $\widetilde{\mathbf{G}}_r$ as well as the BS precoding matrix. This optimization problem can be formulated as

$$\begin{aligned} \max_{\mathbf{\Theta}} \quad & \sum_{m=1}^{M}\left\|\mathbf{h}_m^{\mathrm{H}}\mathbf{\Theta}\widetilde{\mathbf{G}}_r\right\|^2, \\ \text{s.t.} \quad & 0 \leq \theta_n < 2\pi,\ 1 \leq n \leq N, \end{aligned} \tag{45}$$

assuming that $\mathbf{H}$ and $\widetilde{\mathbf{G}}_r$ are available at the RIS. Let $\mathbf{t} = [t_1, \ldots, t_N]^{\mathrm{H}}$ with $t_n = e^{\mathrm{j}\theta_n}$, $\forall n$. The constraints in (45) are equivalent to $|t_n| = 1$, $\forall n$. With the change of variables $\mathbf{h}_m^{\mathrm{H}}\mathbf{\Theta}\widetilde{\mathbf{G}}_r = \mathbf{t}^{\mathrm{H}}\mathrm{diag}\left(\mathbf{h}_m\right)\widetilde{\mathbf{G}}_r = \mathbf{t}^{\mathrm{H}}\mathbf{C}_m$, where $\mathbf{C}_m = \mathrm{diag}\left(\mathbf{h}_m\right)\widetilde{\mathbf{G}}_r$, the problem (45) is equivalent to

$$\begin{aligned} \max_{\mathbf{t}} \quad & \sum_{m=1}^{M}\mathbf{t}^{\mathrm{H}}\mathbf{C}'_m\mathbf{t}, \\ \text{s.t.} \quad & |t_n| = 1,\ 1 \leq n \leq N, \end{aligned} \tag{46}$$

where $\mathbf{C}'_m = \mathbf{C}_m\mathbf{C}_m^{\mathrm{H}}$. Since $\sum_{m=1}^{M}\mathbf{t}^{\mathrm{H}}\mathbf{C}'_m\mathbf{t} = \mathbf{t}^{\mathrm{H}}\widetilde{\mathbf{C}}\mathbf{t}$, where $\widetilde{\mathbf{C}} = \sum_{m=1}^{M}\mathbf{C}'_m$, the optimization problem (46) can be expressed as

$$\begin{aligned} \max_{\mathbf{t}} \quad & \mathbf{t}^{\mathrm{H}}\widetilde{\mathbf{C}}\mathbf{t}, \\ \text{s.t.} \quad & |t_n| = 1,\ 1 \leq n \leq N. \end{aligned} \tag{47}$$

The objective function of (47) can be expressed as $\mathbf{t}^{\mathrm{H}}\widetilde{\mathbf{C}}\mathbf{t} = \mathrm{tr}\left(\mathbf{t}^{\mathrm{H}}\widetilde{\mathbf{C}}\mathbf{t}\right) = \mathrm{tr}\left(\widetilde{\mathbf{C}}\mathbf{t}\mathbf{t}^{\mathrm{H}}\right)$. Define $\mathbf{T} = \mathbf{t}\mathbf{t}^{\mathrm{H}}$. Then the constraints of (47) become $\mathbf{T} \succeq \mathbf{0}$ and $\mathrm{rank}(\mathbf{T}) = 1$. Since the rank-one constraint is non-convex, it is challenging to directly solve this optimization.

By applying semidefinite relaxation (SDR) technique to relax the non-convex rank-one constraint into $[\mathbf{T}]_{n,n} = 1$ for $1 \leq n \leq N$, we relax the optimization problem (47) into

$$\begin{aligned} \max_{\mathbf{T}} \quad & \mathrm{tr}(\widetilde{\mathbf{C}}\mathbf{T}), \\ \text{s.t.} \quad & [\mathbf{T}]_{n,n} = 1,\ 1 \leq n \leq N, \\ & \mathbf{T} \succeq 0. \end{aligned} \tag{48}$$

The optimization problem (48) is a standard convex SDP, and it can readily be solved by a convex optimization solver, such as CVX.

However, the solution of the problem (48) in general may not be a rank-one solution, i.e., $\mathrm{rank}(\mathbf{T}) \neq 1$, which implies that it is not the solution of the problem (47). Therefore, it is necessary to construct a rank-one problem from the optimal higher-ranking solution $\mathbf{T}$ of the optimization problem (48). This can be achieved by first carrying out the eigenvalue decomposition of $\mathbf{T}$ as $\mathbf{T} = \mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}^{\mathrm{H}}$, where $\mathbf{\Omega} \in \mathbb{C}^{N\times N}$ is a unitary matrix and $\mathbf{\Sigma} = \mathrm{diag}\left(\left[\lambda_1, \cdots, \lambda_N\right]\right)$ is a diagonal matrix with the eigenvalues of $\mathbf{T}$ at its diagonal entries. Then a suboptimal solution to (47) can be obtained as $\mathbf{t}' = \mathbf{\Omega}\mathbf{\Sigma}^{1/2}\mathbf{r}$, where $\mathbf{r} \in \mathbb{C}^{N\times 1}$ is a random vector generated according to the distribution $\mathcal{CN}\left(\mathbf{0}, \mathbf{I}_N\right)$. The objective value of (47) can be approximated by taking the maximum among the best $\mathbf{t}'$ obtained from a set of independently generated Gaussian random vectors $\mathbf{r}$. Finally, the solution $\mathbf{t}$ to the problem (46) can be recovered by $\mathbf{t} = e^{\mathrm{j}\arg\left(\mathbf{t}'\right)}$.

C. Precoding Matrix Design

The BS precoding matrix $\mathbf{P} = \left[\mathbf{p}_1, \cdots, \mathbf{p}_M\right]$ also influences the achievable DL sum rate. By using the notations $\mathbf{V} = \mathbf{H}^{\mathrm{H}}\mathbf{\Theta}(\mathbf{UG})^{\mathrm{H}}\mathbf{B} \in \mathbb{C}^{M\times K_{RF}}$ and $\mathbf{v}_m = [\mathbf{V}^{\mathrm{H}}]_{:,m}$, the received signal at the m-th user, i.e., the m-th element of $\mathbf{y}$, can be written as

$$y_m = \sqrt{\rho}\mathbf{v}_m^{\mathrm{H}}\mathbf{p}_m s_m + \sqrt{\rho}\sum_{i\neq m}^{M}\mathbf{v}_m^{\mathrm{H}}\mathbf{p}_i s_i + n_m, \tag{49}$$

where n_m is the m-th element of $\mathbf{n}$. The last two terms in the right hand side of (49) represent the interference-plus-noise, which is a random variable having zero mean and variance $\rho\sum_{i\neq m}^{M}\left|\mathbf{v}_m^{\mathrm{H}}\mathbf{p}_i\right|^2 + \sigma_{DL}^2$. The interference-plus-noise can be approximated as an complex additive Gaussian noise that is independent of s_m.

Obviously, as with any channel estimator, there exist channel estimation errors for the proposed PARAFAC-based channel estimation technique, which will negatively impact on the achievable DL sum rate. By defining the estimation error matrices for $\mathbf{G}$ and $\mathbf{H}$ as $\mathbf{\Xi}^g \in \mathbb{C}^{K\times N}$ and $\mathbf{\Xi}^h \in \mathbb{C}^{N\times M}$, respectively, the end-to-end channel matrix $\mathbf{V}$ can be expressed as

$$\mathbf{V} = (\widehat{\mathbf{H}} - \mathbf{\Xi}^h)^{\mathrm{H}}\mathbf{\Theta}(\mathbf{U}(\widehat{\mathbf{G}} - \mathbf{\Xi}^g))^{\mathrm{H}}\mathbf{B} = \widehat{\mathbf{H}}^{\mathrm{H}}\mathbf{\Theta}(\mathbf{U}\widehat{\mathbf{G}})^{\mathrm{H}}\mathbf{B} - \mathbf{\Xi}, \tag{50}$$

where $\mathbf{\Xi} = (\mathbf{\Xi}^h)^{\mathrm{H}}\mathbf{\Theta}(\mathbf{U}\widehat{\mathbf{G}})^{\mathrm{H}}\mathbf{B} + \widehat{\mathbf{H}}^{\mathrm{H}}\mathbf{\Theta}(\mathbf{U}\mathbf{\Xi}^g)^{\mathrm{H}}\mathbf{B} - (\mathbf{\Xi}^h)^{\mathrm{H}}\mathbf{\Theta}(\mathbf{U}\mathbf{\Xi}^g)^{\mathrm{H}}\mathbf{B}$. Also the BS precoding vectors can only be calculated based on the estimated channels, which are denoted as $\widehat{\mathbf{p}}_m$ for $1 \leq m \leq M$. Using these expressions, the received signal model with the estimated channels is derived as

$$y_m = \sqrt{\rho}\widehat{\mathbf{v}}_m^{\mathrm{H}}\widehat{\mathbf{p}}_m s_m + \sqrt{\rho}\sum_{i\neq m}^{M}\widehat{\mathbf{v}}_m^{\mathrm{H}}\widehat{\mathbf{p}}_i s_i - \sqrt{\rho}\sum_{j=1}^{M}\boldsymbol{\xi}_j^{\mathrm{H}}\widehat{\mathbf{p}}_j s_j + n_m, \tag{51}$$

where $\widehat{\mathbf{v}}_m = \left[\widehat{\mathbf{V}}^{\mathrm{H}}\right]_{:,m}$, $1 \leq m \leq M$, with $\widehat{\mathbf{V}} = \widehat{\mathbf{H}}^{\mathrm{H}}\mathbf{\Theta}(\mathbf{U}\widehat{\mathbf{G}})^{\mathrm{H}}\mathbf{B}$, and $\boldsymbol{\xi}_j = [\mathbf{\Xi}]_{j,:}$. The first term of (51) is the intended signal and the remaining terms are treated as the interference-plus-noise. The power of the third term in (51) is $\xi = \rho\sum_{j=1}^{M}\left|\boldsymbol{\xi}_j^{\mathrm{H}}\widehat{\mathbf{p}}_j\right|^2$.

The solution of the beam selection matrix $\mathbf{B}$ also depends on the estimated channels, i.e., $\mathbf{B}$ is actually obtained by replacing the real channels with the estimated channels. With the estimated channels, therefor, the achievable DL rate for the m-th user is calculated as

$$\widehat{R}_m = \mathbb{E}\left\{\log_2\left(1 + \frac{\rho\left|\widehat{\mathbf{v}}_m^{\mathrm{H}}\widehat{\mathbf{p}}_m\right|^2}{\rho\sum_{i\neq m}^{M}\left|\widehat{\mathbf{v}}_m^{\mathrm{H}}\widehat{\mathbf{p}}_i\right|^2 + \xi + \sigma_{DL}^2}\right)\right\}. \tag{52}$$

For the perfect CSI, by comparison, the achievable DL rate for the m-th user is given by

$$R_m = \mathbb{E}\left\{\log_2\left(1 + \frac{\rho\left|\mathbf{v}_m^{\mathrm{H}}\mathbf{p}_m\right|^2}{\rho\sum_{i\neq m}^{M}\left|\mathbf{v}_m^{\mathrm{H}}\mathbf{p}_i\right|^2 + \sigma_{DL}^2}\right)\right\}. \tag{53}$$

Given the channel estimation, the ZF precoding scheme sets the precoding matrix as $\widehat{\mathbf{P}} = \widehat{\mathbf{V}}^{\mathrm{H}}(\widehat{\mathbf{V}}\widehat{\mathbf{V}}^{\mathrm{H}})^{-1}$ to eliminate interference among different users. With the ZF precoding, therefore, the achievable DL rate for the m-th user given the channel estimation become

$$\widehat{R}_m^{(\mathrm{ZF})} = \mathbb{E}\left\{\log_2\left(1 + \frac{\rho}{\xi + \sigma_{DL}^2}\right)\right\}. \tag{54}$$

For the perfect CSI, by comparison, the achievable DL rate for the m-th user is given by

$$R_m^{(\mathrm{ZF})} = \log_2\left(1 + \frac{\rho}{\sigma_{DL}^2}\right). \tag{55}$$

VI. Simulation Based Experimental Results

In this section, we conduct the simulation based experiments to evaluate the performance of the proposed UAMP channel estimator for jointly estimating $\mathbf{G}$ and $\mathbf{H}$.

A. Performance Metrics and Default System Settings

In the simulation, the channel estimation accuracy of an estimator can be measured by the normalized mean square error (NMSE) metric calculated as

$$\mathrm{NMSE}\left(\widehat{\mathbf{S}}\right) = \frac{1}{V}\sum_{v=1}^{V}\frac{\left\|\mathbf{S}_v - \widehat{\mathbf{S}}_v\right\|_F^2}{\left\|\mathbf{S}_v\right\|_F^2}, \tag{56}$$

where $\mathbf{S}=\mathbf{H}$ or $\mathbf{G}$, $\widehat{\mathbf{S}}_v$ denotes the estimate of the real channel $\mathbf{S}_v$ at the v-th run, and V is the number of Monte Carlo runs. As we have derived the CRB for our UAMP estimator, it is utilized as the ultimate benchmark for evaluating the performance of our algorithm. We also provide the achievable DL sum rate, which is calculated by averaging over V Monte Carlo runs, to validate the effectiveness of our UAMP method.

All the NMSE and DL sum rate results are averaged over $V = 2000$ independent Monte Carlo trials. The threshold δ is 10^{-5} and the maximum number of iterations $I_{\max}$ is 100 in Algorithm 1. The first column of the channel matrix $\mathbf{H}$ is normalized to remove the scaling ambiguity. We adopt the truncated DFT matrices for the RIS phase matrix $\mathbf{\Phi}$ and assume $K_{RF} = M$. $|\alpha_{l_1}^G| = 10^{-3}d_{\mathrm{BR}}^{-2.2}$, where d_{BR} denotes the distance between the BS and the RIS. $|\alpha_{l_2}^m| = 10^{-3}d_{\mathrm{RU}}^{-2.8}$, where d_{RU} denotes the distance between the RIS and the user [41]. Unless otherwise stated, we set $K=64$, $M=N=T=P=16$ and $L_G=5$.

B. NMSE of Channel Estimation

Fig. 4 depicts the NMSE performance as the function of the UL SNR for our UAMP algorithm. We compare our scheme with the idealized LS estimation and the ALS algorithm of [27]. The idealized scheme calculates the LS estimate of the channel matrix $\mathbf{G}$ between the RIS and the BS given the perfect channel matrix $\mathbf{H}$ between the users and the RIS, and it calculates the LS estimate of $\mathbf{H}$ given the perfect $\mathbf{G}$. Obviously, the idealized LS estimation is impractical and represents potentially the best achievable estimation accuracy. It can be observed from Fig. 4 that the estimation accuracy for $\mathbf{H}$ obtained by our algorithm and the ALS algorithm are almost the same as that obtained by the idealized LS algorithm, as the three NMSE curves basically overlap. On the other hand, the estimation accuracy for $\mathbf{G}$ obtained by our algorithm and

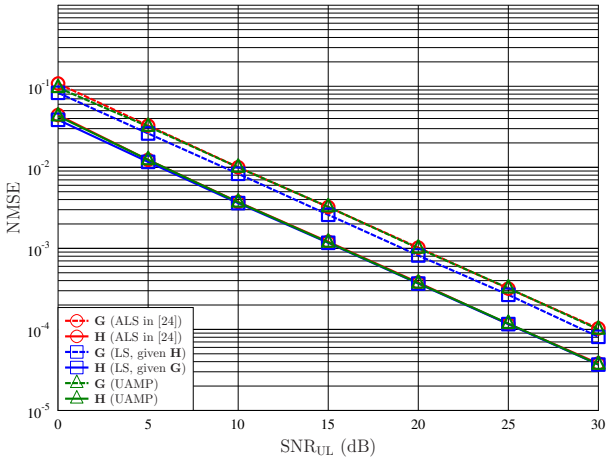

Fig. 4. NMSE performance versus UL SNR comparison for the proposed UAMP estimator, the ALS estimator and the idealized LS scheme.

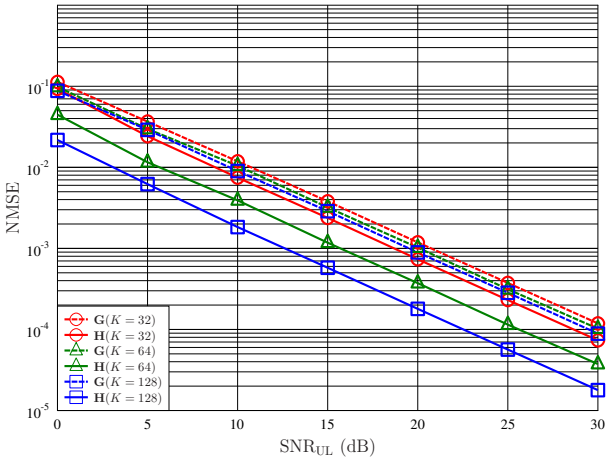

Fig. 6. Estimation accuracy of the proposed UAMP channel estimator, in terms of NMSE versus UL SNR, with different values of $K \in \{32, 64, 128\}$.

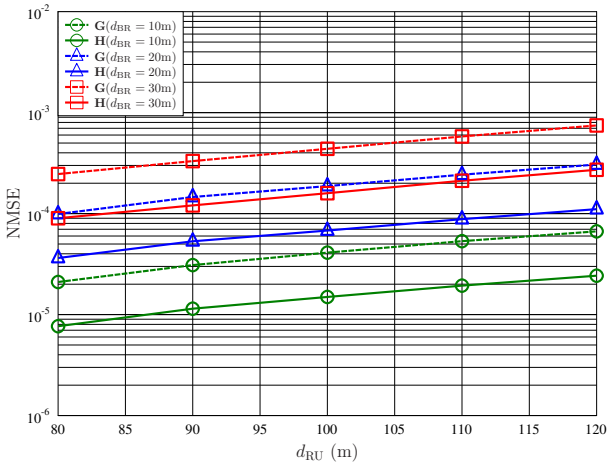

Fig. 5. Estimation accuracy of the proposed UAMP channel estimator, in terms of NMSE versus d_{RU}, with different values of $d_{\mathrm{BR}} \in \{10, 20, 30\}$m.

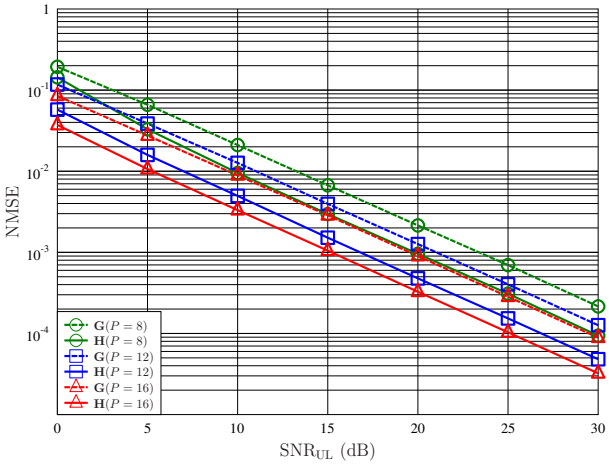

Fig. 7. Estimation accuracy of the proposed UAMP channel estimator, in terms of NMSE versus UL SNR, with $L_G = 10$ and different values of $P \in \{8, 12, 16\}$.

the ALS algorithm are almost the same. While at low SNRs of 0 to 5 dB, the NMSE performances of our algorithm for both **H** and **G** are slightly better than the ALS algorithm. As analyzed in Subsection III-A, our proposed UAMP channel estimation has an additional practical advantage over the ALS channel estimation [27] in that it imposes significantly lower complexity than the ALS scheme. There is only about 1 dB gap in the NMSE between the estimation performance of **G** obtained by our algorithm and that given by the idealized LS estimation. This clearly demonstrates the excellent estimation accuracy achieved by our UAMP estimator.

Fig. 5 portrays the NMSE performance versus d_{RU} of the proposed UAMP estimator with three different values of the distance d_{BR} between the BS and the RIS, given that the sum transmit power constraint of the BS is 0 dBm and the noise variance is -174 dBm/Hz. Observe from Fig. 5 that the NMSE performance of the estimated channels **G** and **H** deteriorate with increasing distance between the BS and the RIS as well as between the RIS and the user. The reason is that increasing distance leads to higher path loss, which in turn degrades the channel estimation performance.

Fig. 6 illustrates the NMSE performance versus the UL SNR of the proposed UAMP channel estimator with three different values for the number of BS antennas K. The results of Fig. 6 indicate that the NMSE performance for the channel estimation becomes better as the number of BS antennas K increases. This is because as K increases, the training overhead also increases, which is beneficial to channel estimation. Also the NMSE performance of the estimated channel **G** does not improve as much as the estimated channel **H** as K increases. This is because the larger K is, the larger the dimension of **G**, but it does not affect the dimension of **H**.

Fig. 7 shows the NMSE performance as the function of UL SNR for our UAMP algorithm given $L_G = 10$ and with three different values of the number of phase shift matrices P. It can be seen that the estimation accuracy improves as P increases, since increasing P implies a longer training sequence.

Fig. 8 illustrates the estimation accuracy of our UAMP channel estimator, in terms of NMSE versus UL SNR, given $K = 32$ and with three different values of the number of paths L_G between the RIS and the BS. The results of Fig. 8 indicate that as L_G increases, the NMSE performance becomes better. This may be explained from the uniqueness condition (23). In this case, we have $L_G < N$. The larger L_G is, the more (23) is satisfied, which lead to better estimation accuracy.

How the number of the RIS elements N influences the

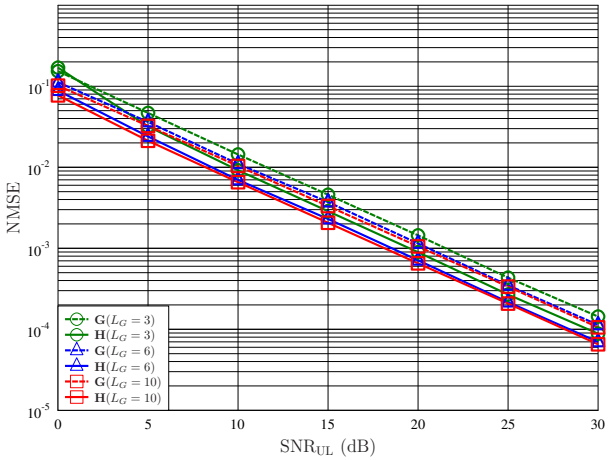

Fig. 8. Estimation accuracy of the proposed UAMP channel estimator, in terms of NMSE versus UL SNR, with $K = 32$ and different values of $L_G \in \{3, 6, 10\}$.

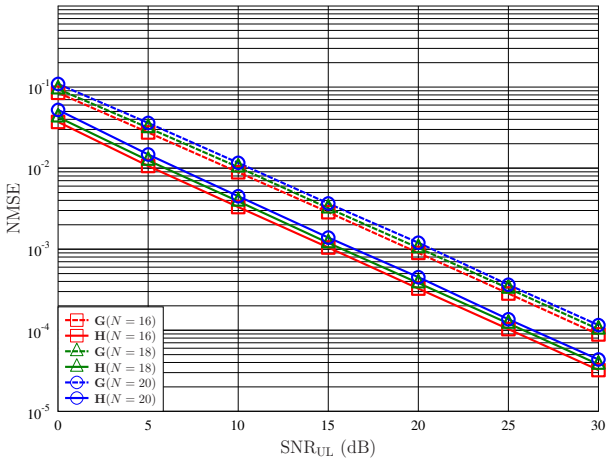

Fig. 9. Estimation accuracy of the proposed UAMP channel estimator, in terms of NMSE versus UL SNR, with $L_G = 10$ and different values of $N \in \{16, 18, 20\}$.

estimation accuracy of our UAMP channel estimator is investigated in Fig. 9. The results of Fig. 9 indicates that the estimation accuracy degrades as N increases. This is because increasing N increases the channel dimension, leading to more channel coefficients in $\mathbf{G}$ and $\mathbf{H}$ to be estimated.

C. CRB of the Proposed UAMP Channel Estimation

We compare the CRB and the NMSE of our UAMP channel estimation in Fig. 10, with $K=32$, $M=N=T=16$, $L_G=10$ and $P\in\{8,16\}$. Observe that over the range of the UL SNRs evaluated, the NMSEs of our UAMP channel estimates are very close to their respective CRBs. In particular, the NMSEs of the two estimated $\mathbf{H}$ reach their respective CRBs when the UL SNR is greater than 10 dB. Furthermore, at low UL SNRs, the NMSE performance with $P = 16$ is closer to the CRB than that with $P=8$. This is because the NMSE performance becomes better as P increases. Additionally, it can be seen that the gap between the NMSE and the corresponding lower bound of CRB is very small for the estimated $\mathbf{G}$. Since the CRB is the ultimate upper bound of achievable estimation accuracy, the results of Fig. 10 offers the clear evidence for the excellent estimation accuracy of our UAMP algorithm.

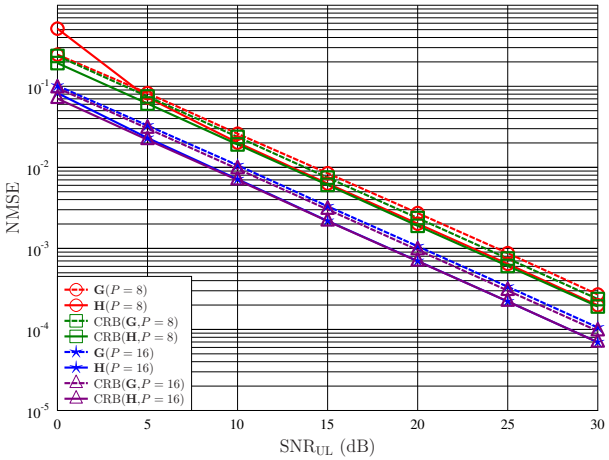

Fig. 10. Comparison of the CRB and the NMSE as the functions of UL SNR for the proposed UAMP channel estimation.

D. DL Sum Rate

In evaluating the DL sum rate achieved with the proposed UAMP channel estimation, unless otherwise specifically stated, we set the system parameters to $K = 16$, $M = N = T = P = 8$ and $L_G = 8$ as well as set the UL SNR to 20 dB.

Fig. 11 illustrates the achievable DL sum rate performance as the function of the DL SNR with different channel estimations, given different UL SNRs. The achievable DL sum rate performance for our proposed UAMP channel estimator and the ALS channel estimator are almost the same, which verifies the NMSE performance in Fig. 4. As expected, when the UL SNR is fixed, the DL sum rate increases with the DL SNR. Similarly, when the DL SNR is fixed, the DL sum rate improves as the UL SNR increases. It can be seen that given a sufficiently high UL training SNR, e.g., 30 dB, the achievable sum rate gap between the estimated CSI and the perfect CSI is very small, particularly, at the poor DL SNR region. At a high DL SNR of 4 dB, the sum rate gap between the estimated CSI and the perfect CSI is only about 1 bits/s/Hz, while the gap is about 3 bits/s/Hz at the DL SNR of 10 dB. The results of Fig. 11 therefore demonstrate the excellent estimation accuracy as well as the effectiveness of our UAMP channel estimator.

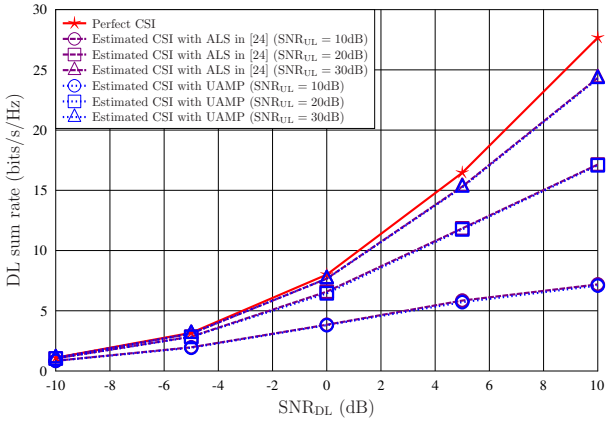

Fig. 11. Achievable DL sum rate versus DL SNR performance with ZF precoding for the proposed UAMP estimation, the ALS estimation and the perfect CSI, given different UL SNRs of 10 dB, 20 dB and 30 dB.

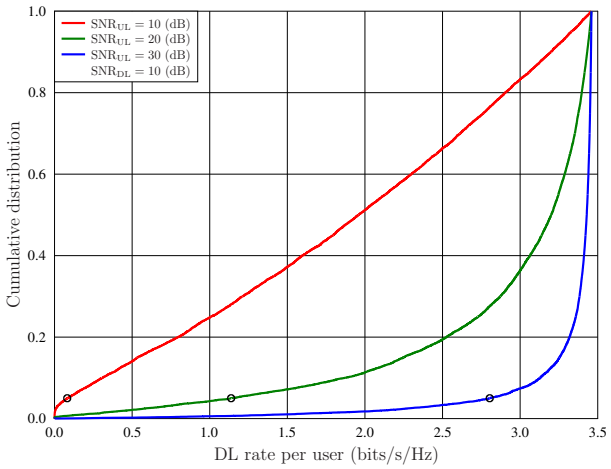

Fig. 12. Cumulative distribution of the achievable DL rate per user for the proposed UAMP channel estimation with ZF precoding, given different UL SNRs of 10 dB, 20 dB and 30 dB as well as DL SNR of 10 dB.

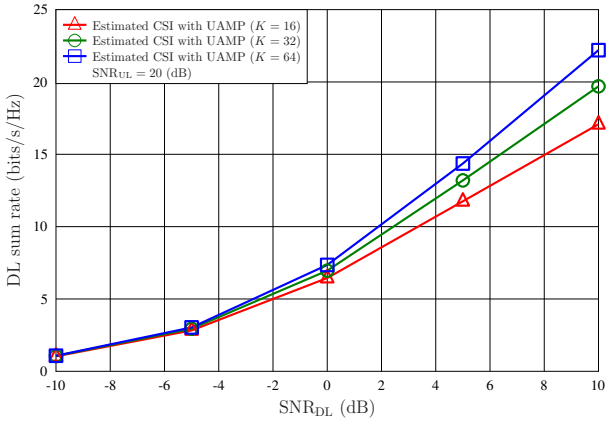

Fig. 13. Achievable DL sum rate versus DL SNR performance for the proposed UAMP channel estimation with ZF precoding, given different values of $K \in \{16, 32, 64\}$.

Fig. 12 depicts the cumulative distribution for the achievable DL rate per user with the proposed UAMP channel estimation, given different UL SNRs of 10 dB, 20 dB and 30 dB as well as DL SNR of 10 dB. The circles indicate the five percent values, i.e., the achievable DL rate per user is greater than or equal to the indicated value with probability 0.95. It can be seen that UL SNR of 20 dB instead of 10 dB increases the achievable DL rate per user by about 1 bits/s/Hz, and UL SNR of 30 dB adds an additional 1.7 bits/s/Hz.

Fig. 13 presents the achievable DL sum rate versus the DL SNR performance for our UAMP channel estimator, given three different values of K. It is evident that the achievable DL sum rate increases with the increases of the number of BS antennas K. This is because the training overhead increases with K, leading to better channel estimation performance for larger K, which can also be seen in Fig. 6.

In Fig. 14, we plot the DL sum rate as the function of the DL SNR for the proposed UAMP algorithm with three different values of P. It can be seen that increasing P results in the enhancement of the achievable DL sum rate. The reason is that an increase in P implies a longer training sequence.

Fig. 15 depicts the achievable DL sum rate performance of the proposed UAMP algorithm under the system settings of $K = 30$, $M = T = P = 10$, $L_G = 10$, $N \in \{10, 12, 14\}$ and UL SNR = 20 dB. Observe from Fig. 15 that at low DL SNRs, the DL sum rates are almost identical given different values of N. However, when the DL SNR is greater than 0 dB, the DL sum rate performance deteriorates as N increases. This is because the number of channel coefficients to be estimated in $\mathbf{G}$ and $\mathbf{H}$ increases with N, leading to poor channel estimation performance and thus poor DL sum rate performance.

Lastly, we verify that the phase shift matrix optimization given in Subsection V-B improves the achievable DL sum rate performance. Fig. 16 compares the DL sum rate performance difference between the random phase shift matrix and the optimized phase shift matrix. As shown in Fig. 16, when the DL SNR is less than 0 dB, the two phase shift matrices exhibit almost the same sum rate performance. However, when the DL SNR is greater than 0 dB, the DL sum rate achieved by the optimized phase shift matrix outperforms that attained by the stochastic DFT phase shift matrix, and this verifies the necessity for the phase shift matrix optimization.

VII. Conclusions

This paper has developed a new channel estimation method for the RIS-aided mmWave multiuser MISO beamspace communication system by exploiting PARAFAC decomposition.

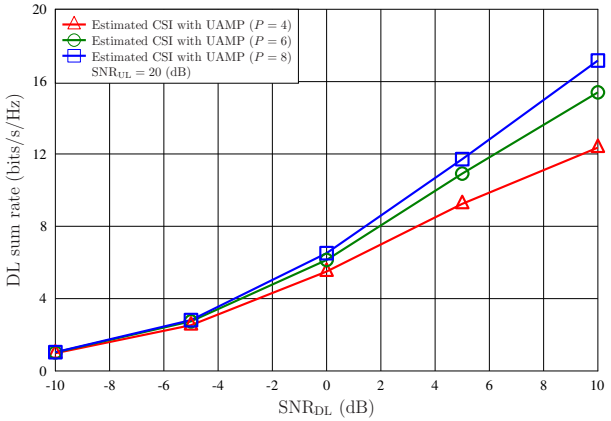

Fig. 14. Achievable DL sum rate versus DL SNR performance for the proposed UAMP channel estimation with ZF precoding, given different values of $P \in \{4, 6, 8\}$.

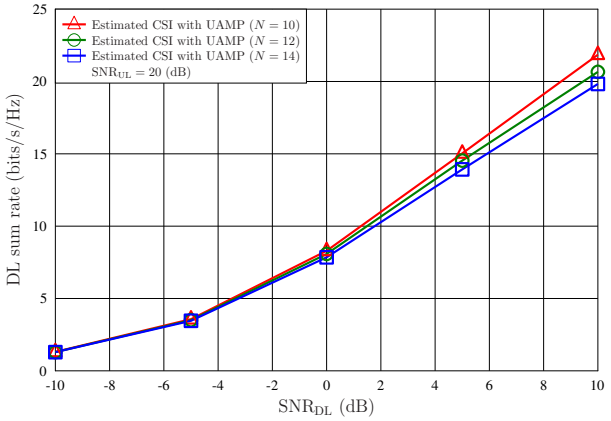

Fig. 15. Achievable DL sum rate versus DL SNR performance for the proposed UAMP channel estimation with ZF precoding, given $K = 30$, $M = T = P = 10$, $L_G = 10$, $N \in \{10, 12, 14\}$, and UL SNR = 20 dB.

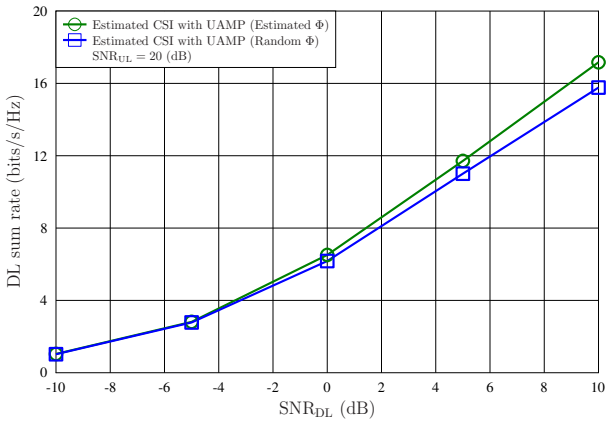

Fig. 16. Achievable DL sum rate versus DL SNR performance of the proposed UAMP CE with ZF precoding given different phase shift matrices.

Specifically, by fully exploiting the tensor structure of the received pilot signals, a novel UAMP channel estimation algorithm has been developed to accurately estimate the cascade beamspace channel matrices. We have presented the feasibility conditions for our UAMP channel estimation algorithm based on which useful system design guidelines can be obtained. Furthermore, we have derived the CRB for the UAMP channel estimation to evaluate the algorithm's performance. With the assistance of our beam selection algorithm and the optimized phase shift matrix obtained using SDR as well as the ZF precoding, we have also calculated the DL sum rate of the RIS-aided mmWave multiuser MISO beamspace system based on the proposed UAMP channel estimation. The excellent estimation accuracy of our proposed UAMP channel estimator has be verified by extensive simulation results. In particular, with sufficiently high UL training SNR, the NMSE of our UAMP channel estimator has been shown to approach the corresponding CRB. In addition, it has been demonstrated that the performance of the proposed UAMP channel estimator is significantly influenced by the number of RIS unit elements and training symbols.

REFERENCES

[1] S. A. Busari, *et al.*, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 836–869, 2nd Quarter 2018.

[2] W. Hao, M. Zeng, Z. Chu and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782-785, Dec. 2017.

[3] X. Gao, *et al.*, "Wideband beamspace channel estimation for millimeter-wave MIMO systems relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4809–4824, Sep. 2019.

[4] X. Guo, *et al.*, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[5] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," in *Proc. Allerton 2010* (Monticello, IL, USA), Sep. 29-Oct. 1, 2010, pp. 1196–1203.

[6] X. Gao, *et al.*, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[7] W. Shen, *et al.*, "Beamspace precoding and beam selection for wideband millimeter-wave MIMO relying on lens antenna arrays," *IEEE Trans. Signal Process*, vol. 67, no. 24, pp. 6301–6313, Dec. 2019.

[8] L. Dai, *et al.*, "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," in *Proc. ICCC 2016* (Chengdu, China), Jul. 27-29, 2016, pp. 1–6.

[9] X. Gao, *et al.*, "Beamspace channel estimation for wideband millimeter-wave MIMO with lens antenna array," in *Proc. ICC 2018* (Kansas City, MO, USA), May 20-24, 2018, pp. 1–6.

[10] C. Pan, *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[11] X. Li, *et al.*, "Exploiting benefits of IRS in wireless powered NOMA networks," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 175–186, Mar. 2022.

[12] X. Li, *et al.*, "Performance analysis of covert communication in IRS-assisted NOMA networks," *Sci. Sin. Inform.*, vol. 54, no. 6, pp. 1502–1515, Jun. 2024.

[13] K. Zhi, *et al.*, "Active RIS versus passive RIS: Which is superior with the same power budget?," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, May 2022.

[14] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. ICASSP 2019* (Brighton, U.K.), May 12-17, 2019, pp. 4659–4663.

[15] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc ICASSP 2020* (Barcelona, Spain), May 4-8, 2020, pp. 5000–5004.

[16] 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, May 2020.

[17] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, Nov. 2021.

[18] G. Zhou, *et al.*, "Channel estimation for RIS-aided multiuser millimeter-Wave systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 1478–1492, Mar. 2022.

[19] L. Dai and X. Wei, "Distributed machine learning based downlink channel estimation for RIS assisted wireless communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4900–4909, Jul. 2022.

[20] S. Noh, *et al.*, "Joint direct and indirect channel estimation for RIS-assisted millimeter-wave systems based on array signal processing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8378-8391, Nov. 2023.

[21] H. Chu, *et al.*, "Adaptive and robust channel estimation for IRS-aided millimeter-wave communications," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 9411-9423, Jul. 2024.

[22] A. Cichocki, *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[23] N. D. Sidiropoulos, *et al.*, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[24] Z. Zhou, *et al.*, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[25] Y. Zniyed, R. Boyer, A. L. F. de Almeida, and G. Favier, "Tensor-train modeling for MIMO-OFDM tensor coding-and-forwarding relay systems," in *Proc. EUSIPCO 2019* (A Coruna, Spain), Sep. 2-6, 2019, pp. 1–5.

[26] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789–802, Apr. 2021.

[27] L. Wei, *et al.*, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.

[28] S. Gherekhloo, K. Ardah, A. L. F. de Almeida, and M. Haardt, "Tensor-based channel estimation and reflection design for RIS-aided millimeter-wave MIMO communication systems," in *Proc. of 55th Asilomar Conf. Signals, System, and Computers* (Pacific Grove, CA, USA), Oct. 31-Nov. 3, 2021, pp. 1683–1689.

[29] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "Double-RIS versus single-RIS aided systems: Tensor-based MIMO channel estimation and design perspectives," in *Proc. ICASSP 2022* (Singapore), May 23-27, 2022, pp. 5183–5187.

[30] X. Zheng, P. Wang, J. Fang, and H. Li, "Compressed channel estimation for IRS-assisted millimeter wave OFDM systems: A low-rank tensor decomposition-based approach," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1258–1262, Jun. 2022.

[31] X. Zhang, *et al.*, "Sparsity-structured tensor-aided channel estimation for RIS-assisted MIMO communications," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2460–2464, Oct. 2022.
[32] P. Liu, *et al.*, "Intelligent reflecting surface aided NOMA for millimeter-wave massive MIMO with lens antenna array," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4419–4434, May 2021.
[33] Y. Wang, *et al.*, "Intelligent reflecting surface-assisted mmWave communication with lens antenna array," *IEEE Trans. Cognit. Commun. and Netw.*, vol. 8, no. 1, pp. 202–215, Mar. 2022.
[34] X. Gao, *et al.*, "Beamspace channel estimation for 3D lens-based millimeter-wave massive MIMO systems," in *Proc. WCSP 2016* (Yangzhou, China), Oct. 13-15, 2016, pp. 1–5.
[35] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
[36] Q. Guo and J. Xi, "Approximate message passing with unitary transformation," *CoRR*, vol. abs/1504.04799, 2015. [Online]. Available: http://arxiv.org/abs/1504.04799
[37] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.
[38] A. Stegeman and N. Sidiropoulos, "On Kruskal's uniqueness condition for the CANDECOMP/PARAFAC decomposition," *Linear Algebra and Its Appl.*, vol. 420, nos. 2-3, pp. 540–552, 2007.
[39] X. Liu and N. Sidiropoulos, "Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.
[40] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
[41] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Dec. 2020.

XINYING GUO received the B.Sc. degree in electronic and information engineering from the North University of China in 2011 and the Ph.D degree in information and communication engineering from Zhengzhou University in 2017. She has been an associate professor with the College of Information Science and Engineering, Henan University of Technology. From 2014 to 2016, she was a Visiting Researcher in School of Electronics and Computer Science, the University of Southampton, U.K. Her research interests are in the areas of wireless communications and signal processing, including channel estimation, reconfigurable intelligent surfaces and unmanned aerial vehicles, etc.

ZONGYUAN XIE received the B.Sc. degree in electronic and information engineering from the Henan University of Technology of China in 2021. He is currently pursuing a master's degree at Henan University of Technology. His research interest includes reconfigurable intelligent surfaces, mm-wave communications, beam selection, signal processing for wireless communications.

JIANKANG ZHANG (IEEE Senior Member) is a Senior Lecturer at Bournemouth University. Prior to joining in Bournemouth University, he was a senior research fellow at University of Southampton, UK. Dr Zhang was a lecturer from 2012 to 2013 and then an associate professor from 2013 to 2014 at Zhengzhou University. His research interests are in the areas of aeronautical communications and networks, evolutionary algorithms, machine learning algorithms and edge computing. He serves as an Associate Editor for IEEE ACCESS.

SHENG CHEN (IEEE Life Fellow) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from City University, London, in 1986, and the higher doctoral (D.Sc.) degree from the University of Southampton, Southampton, U.K., in 2005. From 1986 to 1999, he held research and academic appointments with the University of Sheffield, the University of Edinburgh, and the University of Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he is a Professor of Intelligent Systems and Signal Processing. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimization. Professor Chen has authored over 700 research papers. He have 20,000+ Web of Science citations with h-index 62, and 39,000+ Google Scholar citations with h-index 84. Dr Chen was elected to a fellow of the United Kingdom Royal Academy of Engineering in 2014. He is a fellow of Asia-Pacific Artificial Intelligence Association (FAAIA), a fellow of IET, and an original ISI Highly Cited Researcher in engineering (March 2004).

CHUNHUA ZHU received the B.Sc. degree in communication engineering from the Zhengzhou University of China in 2000 and the Ph.D degree in information and communication engineering from Zhengzhou University in 2013. She has been a professor with the College of Information Science and Engineering, Henan University of Technology. Her research interests are in the areas of wideband wireless communications, including intelligent signal and information processing, advanced detection technology, abnormal recognition, etc.