The following publication R. Wang, S. Zhang, B. Clerckx and L. Liu, "Low-Overhead Channel Estimation Framework for Beyond Diagonal Reconfigurable Intelligent Surface Assisted Multi-User MIMO Communication," in IEEE Transactions on Signal Processing, vol. 73, pp. 4700-4717, 2025 is available at https://doi.org/10.1109/TSP.2025.3628337.

Low-Overhead Channel Estimation Framework for Beyond Diagonal Reconfigurable Intelligent Surface Assisted Multi-User MIMO Communication

Rui Wang, Shuowen Zhang, Bruno Clerckx, and Liang Liu

***Abstract*—Beyond diagonal reconfigurable intelligent surface (BD-RIS) refers to a family of RIS architectures characterized by scattering matrices not limited to being diagonal and enables higher wave manipulation flexibility and large rate performance gain over conventional (diagonal) RIS. However, whether BD-RIS aided network is able to deliver more data volume compared to conventional RIS aided network is still questionable, because far more time may be wasted to estimate the massive channel coefficients associated with the off-diagonal entries of the BD-RIS scattering matrix. Somehow counter intuitively, for the first time in the literature, this paper rigorously proves that the channel estimation overhead in fully connected BD-RIS aided network is actually of the same order as that in the conventional RIS aided network, which was characterized in [1]. This amazing result stems from a key observation: for each user antenna, its cascaded channel matrix associated with one reference BD-RIS element is a scaled version of that associated with any other BD-RIS element due to the common RIS-base station (BS) channel. In other words, the number of independent unknown variables is far less than it would seem at first glance. Building upon this property, this paper manages to characterize the overhead to perfectly estimate all the channels in the ideal case without noise at the BS, and propose a two-phase estimation framework for the practical case with noise at the BS. The main message of this paper is that we can benefit from the non-diagonal scattering matrix design at a channel estimation cost similar to that in conventional RIS aided network.**

***Index Terms*—Beyond diagonal reconfigurable intelligent surface (BD-RIS), channel estimation, low-overhead communication.**

I. Introduction

A. Motivation

Reconfigurable intelligent surfaces (RISs) have emerged as a promising technology for the sixth-generation (6G) cellular network, thanks to their unprecedented capabilities to tune the wireless propagation environment. Conventionally, the scattering matrices of RISs are diagonal matrices such that the RIS elements change the propagation properties independently [3], [4]. Recently, a new technology, i.e., beyond diagonal RIS (BD-RIS), has attracted more and more attention [5], [6]. Specifically, BD-RIS leverages inter-element connections to generate non-diagonal scattering matrices. This non-diagonal structure enables more flexible wave manipulation, enhanced beamforming, and expanded coverage that are critical for 6G networks [5], [7]. Because of the great potential of BD-RIS and the optimality of BD-RIS to achieve the maximum SNR and capacity of wireless channel [6], [8], research has been conducted on multiple fronts: physics-consistent modeling [6], [9], optimal and suboptimal architecture designs [6], [8], [10], transmissive, reflective, hybrid and multi-sector modes [7], [11], performance optimization and analysis [12]–[15], hardware impairments [16], and integration with other technologies and applications such as integrated sensing and communication (ISAC) [17]–[22], physical layer security [23], full duplex systems [24].

The enhanced performance of BD-RIS depends critically on the availability of channel state information (CSI). Without the BD structure, the channel estimation overhead in conventional RIS assisted communication has been characterized in [1]. However, unlike conventional (diagonal) RIS with a single-connected architecture, the interconnected architecture of BD-RIS makes it necessary to estimate far more channel coefficients associated with the non-diagonal elements of the BD-RIS scattering matrix [6], [25]. Moreover, due to the circuit requirement, the non-diagonal scattering matrices of the BD-RIS are unitary matrices [6], [7]. This imposes constraints since all reflecting elements must be adjusted jointly (unlike conventional RIS with independently tunable elements), thereby complicating the design of time-varying scattering strategies for channel estimation. Due to the above reasons, it is widely believed that the channel estimation overhead in BD-RIS assisted communication is significantly larger than that in conventional RIS assisted communication [25]–[29]. Note that if too much time is spent on channel estimation, the true network throughput gain over the conventional RIS aided communication is questionable, despite the rate gain due to the freedom to design non-diagonal scattering matrices.

In this work, somehow counter intuitively, we aim to reveal the fact that the channel estimation overhead of the BD-RIS aided communication and the conventional RIS aided communication are of the same order. Specifically, we consider a fully connected BD-RIS architecture [6] where each reflecting

This paper was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62471421, in part by the Hong Kong Research Grants Council (RGC) General Research Fund under Grant 15230022, Grant 15203222, and Grant 15213322, and in part by the Hong Kong RGC Young Collaborative Research Grant under Grant PolyU C5002-23Y. The materials in this paper have been presented in part at the IEEE International Conference on Wireless Communications and Signal Processing (WCSP), October 2024 [2].

Rui Wang, Shuowen Zhang, and Liang Liu are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China (e-mails: rui-eie.wang@connect.polyu.hk, {shuowen.zhang, liang-eie.liu}@polyu.edu.hk).

Bruno Clerckx is with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. (e-mail: b.clerckx@imperial.ac.uk).

element is connected to all other elements to form a full non-diagonal scattering matrix. While this architecture provides the highest flexibility for wave manipulation, it inherently incurs the highest channel estimation overhead due to the largest number of interconnections compared to other BD-RIS architectures, e.g., group connected BD-RIS [6], tree and forest-connected BD-RIS [10], band-connected and stem-connected BD-RIS [8], whose corresponding graphs have fewer edges (hence interconnections) than fully-connected BD-RIS. By addressing the most challenging fully connected BD-RIS architecture, we will establish a theoretical foundation relating channel estimation overhead to BD-RIS architectural complexity and provide insights that help reducing overhead for other BD-RIS architectures.

B. Prior Work

Under the conventional RIS aided communication, previous research has been conducted to reduce channel estimation overhead by utilizing multi-user channel correlation [1], [30], two-timescale property [31], beamspace channel sparsity [30], [32]–[34], RIS elements grouping [35], etc. However, these methods cannot be applied to BD-RIS aided communication, because massive new channel coefficients associated with the off-diagonal entries in the scattering matrix have to be estimated as well.

Under the BD-RIS assisted communication, most prior works on channel estimation have focused on single-user multiple-input multiple-output (MIMO) scenarios [25]–[28]. [25] and [26] proposed a least square (LS)-based method to derive a closed-form estimation of the cascaded user-RIS-base station (BS) channel and formulated a BD-RIS design problem aimed at minimizing the mean square error (MSE) estimation. [27] developed two tensor decomposition–based methods. One offers a closed-form solution based on the Block Tucker Kronecker factorization (BTKF) algorithm. Given an initial LS estimator of the cascaded channel, it leverages rank-one matrix approximation to extract the individual user-RIS and RIS-BS channels, yielding a noise rejection gain compared to the LS estimation scheme. The other employs an alternating LS (ALS)-based approach to directly decouple the estimation of cascaded channel into individual channel matrices, thereby reducing overhead for estimation. [28] proposed a semi-blind joint channel and symbol estimation scheme based on tensor modeling of the received signals and a trilinear alternating estimation scheme was derived. For the multi-user scenario, [29] adapted the ALS method for multiple single-antenna users by assuming that all users transmit orthogonal pilot sequences such that channels from all users can be simultaneously estimated by the BS.

However, in a fully connected BD-RIS assisted communication system with N BS antennas, M BD-RIS elements, and K users each with U antennas, most of the above methods [25]–[27] require a channel estimation overhead of KUM^2, because all the channel coefficients are estimated independently. Although the ALS-based method proposed in [28], [29] can reduce the channel estimation overhead via 1-mode and 2-mode unfoldings of the received signal tensor, the overhead is still fundamentally high because all the channel coefficients are treated as independent variables. Note that in the conventional RIS assisted communication, it has been shown in [1] that the channel estimation overhead is $M + \lceil M(KU-1)/q \rceil$, where q is the rank of RIS-BS channel. An important question motivating this work arises: *Is it possible to reduce the channel estimation overhead in BD-RIS assisted communication to the same order as that in conventional RIS assisted communication?*

TABLE I
COMPARISON OF CHANNEL ESTIMATION OVERHEAD

Scheme	Channel Estimation Overhead
Our scheme in BD-RIS assisted system	$2M + \left\lceil \frac{M(KU-1)}{q} \right\rceil$
Existing schemes [25]–[27] in BD-RIS assisted system	KUM^2
Our scheme in conventional RIS assisted system [1]	$M + \left\lceil \frac{M(KU-1)}{q} \right\rceil$

C. Main Contributions

In this paper, we study the channel estimation problem in BD-RIS assisted uplink communication consisting of K users each with U antennas, one fully connected BD-RIS with M elements, and one BS with N antennas. Based on the pilot signals from the users, the BS needs to estimate the channels that are necessary for the subsequent BS and BD-RIS beamforming design. The contributions of this paper are summarized as follows.

- We rigorously show that in the ideal case without noise at the BS, $2M + \lceil M(KU-1)/q \rceil$ time instants are sufficient to perfectly estimate all the cascaded channels, where q denotes the rank of the RIS-BS channel matrix. On one hand, the overhead of our approach is an order-of-magnitude lower than that achieved by the existing approaches designed for fully connected BD-RIS [25]–[27], which is KUM^2. Moreover, in the practical case with noise at the BS, we design an efficient algorithm to accurately estimate the cascaded channels with a much smaller MSE compared to that achieved by the existing approaches [25]–[29].
- We achieve the above amazing result by revealing an important hidden property of the cascaded user-RIS-BS channels in BD-RIS aided communication — for each user antenna, its cascaded user-RIS-BS channel matrix associated with any BD-RIS element is a scaled version of that associated with a reference BD-RIS element. This is because all cascaded channels share a common RIS-BS channel. This indicates that the overall cascaded channel is fundamentally determined by a single reference channel matrix and a set of scaling coefficients. This key insight reveals that the number of independent variables to be estimated is drastically smaller than what is assumed in prior works [25]–[27], which treated all channel coefficients independently, leading to larger channel estimation

overhead. To our best knowledge, our work is the first one to reveal this important property in BD-RIS assisted communication.

- Our results mitigate the biggest concern about BD-RIS aided communication — whether its throughput performance outperforms that of conventional RIS aided communication, considering the fact that far more channel coefficients should be estimated. Specifically, the channel estimation overhead in conventional RIS aided communication is $M+\lceil M(KU-1)/q\rceil$ [1]. Therefore, the channel estimation overheads of the fully connected BD-RIS aided communication and of the conventional RIS aided communication are actually of the same order. This indicates that we can reap the rate gain arising from the non-diagonal scattering matrix design with a similar channel estimation overhead to conventional RIS aided communication. Therefore, our results provide solid proof of the effectiveness of BD-RIS in future wireless network. A comparison among the channel estimation overheads achieved by different schemes in different systems is summarized in Table I.

D. Organization

The rest of this paper is organized as follows. Section II introduces the system model. Section III states the channel estimation problem without and with noise at the BS, respectively, and reveals the channel property of the cascaded channels under the BD-RIS assisted communication. Section IV considers the case without noise and characterizes the overhead for perfectly estimating the channels in a special case of a single-antenna user. Section V generalizes the result to a general case of multiple multi-antenna users. Section VI proposes an efficient channel estimation method under the practical case with noise. Section VII provides numerical examples to demonstrate the effectiveness of our proposed scheme. Section VIII concludes the paper.

Notations: For ease of reference, Table II summarizes the main variables which will be used throughout this paper. Furthermore, $\boldsymbol{I}$ and $\boldsymbol{O}$ denote an identify matrix and an all-zero matrix with appropriate dimensions, respectively. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$ denote its transpose and conjugate transpose, respectively. For a square full-rank matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-1}$ denotes its inverse. $\mathrm{vec}(\cdot)$ denotes the vectorization of a matrix, and $\mathrm{unvec}(\cdot)$ denotes the reverse operation of the vectorization. $\boldsymbol{A}(I,:)$ and $\boldsymbol{A}(:,I)$ extracts the subset of rows and columns from a matrix $\boldsymbol{A}$, respectively, where I is a sequence of row-index or column-index. $\|\cdot\|_F$ denotes the Frobenius norm. $\lceil\cdot\rceil$ denotes the ceiling function. $\otimes$ denotes the Kronecker product. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean $\boldsymbol{x}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\boldsymbol{x},\boldsymbol{\Sigma})$. $\mathbb{E}[\cdot]$ denotes the expectation operator.

II. System Model

We consider an uplink communication system consisting of a BS with N antennas, a BD-RIS with M passive reflecting elements, and K users each with U antennas, as shown in Fig. 1. We assume a quasi-static block fading channel model, where the channels remain approximately constant in each coherence block with T time instants. Define $\boldsymbol{D}_k\in\mathbb{C}^{N\times U}$ as the direct baseband equivalent channels from user k to the BS. The baseband equivalent channel from the u-th antenna of user k to the m-th BD-RIS reflecting element and that from the m-th BD-RIS element to the BS are denoted by $r_{k,u,m}\in\mathbb{C}$ and $\boldsymbol{g}_m\in\mathbb{C}^{N\times 1}$, $k=1,\cdots,K$, $u=1,\cdots,U$, $m=1,\cdots,M$, respectively. Define

$$\boldsymbol{R}_k=\begin{bmatrix} r_{k,1,1} & \cdots & r_{k,U,1}\\ \vdots & \ddots & \vdots\\ r_{k,1,M} & \cdots & r_{k,U,M}\end{bmatrix}\in\mathbb{C}^{M\times U},\quad \forall k, \tag{1}$$

as the overall channels from user k to the BD-RIS, and $\boldsymbol{G}=[\boldsymbol{g}_1,\cdots,\boldsymbol{g}_M]\in\mathbb{C}^{N\times M}$ as the overall channels from the BD-RIS to the BS. The rank of $\boldsymbol{G}$ is denoted by q, which is an unknown number. Then, at time instant t, the effective uplink channel from user k to the BS through the BD-RIS, resulting from the direct channel and the RIS reflection channel, is expressed as

$$\boldsymbol{H}_{k,t}=\boldsymbol{D}_k+\boldsymbol{G}\boldsymbol{\Phi}_t\boldsymbol{R}_k,\quad \forall k,t, \tag{2}$$

where $\boldsymbol{\Phi}_t=[\boldsymbol{\phi}_{t,1},\cdots,\boldsymbol{\phi}_{t,M}]\in\mathbb{C}^{M\times M}$ denotes the scattering matrix of the BD-RIS at time instant t, with $\boldsymbol{\phi}_{t,m}$ being

TABLE II
Main Variables and Their Physical Meanings

$\boldsymbol{D}_k$	Direct channel matrix from user k to the BS
$\boldsymbol{R}_k$	Channel matrix from user k to the BD-RIS
$\boldsymbol{G}$	Channel matrix from the BD-RIS to the BS
$\boldsymbol{g}_m$	Channel vector from the m-th RIS element to the BS
$\boldsymbol{H}_{k,t}$	Overall effective channel from user k to the BS via BD-RIS at time instant t
$\boldsymbol{J}_k$	Cascaded user-RIS-BS channel matrix for user k
$\boldsymbol{Q}_{k,u,m}$	Sub-block of $\boldsymbol{J}_k$, representing the cascaded channel associated with the u-th antenna of user k via the m-th RIS element
$\boldsymbol{\Phi}_t$	Scattering matrix of the BD-RIS at time instant t
$\boldsymbol{\phi}_{t,m}$	m-th column of the scattering matrix $\boldsymbol{\Phi}_t$
$\boldsymbol{B}_k$	Matrix of scaling coefficients for user k
$\boldsymbol{\beta}_{k,u}$	Vector of scaling coefficients for the u-th antenna of user k
$\boldsymbol{x}_{k,t}$	Transmit data signal from user k at time instant t
$\boldsymbol{a}_{k,t}$	Transmit pilot signal from user k at time instant t
$\boldsymbol{y}_t$	Received signal at the BS with the direct channel part removed at time instant t
$\boldsymbol{n}_t$	Additive white Gaussian noise (AWGN) vector at the BS at time instant t

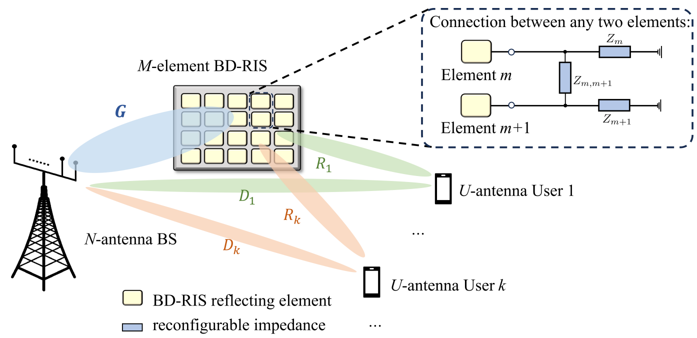


Fig. 1. A BD-RIS assisted MU-MIMO uplink communication system.

its m-th column. Since the BD-RIS has a fully connected architecture where each reflecting element is connected to the other elements, $\boldsymbol{\Phi}_t$ is a full matrix, where the entry on the i-th row and the j-th column denotes the reconfigurable coefficient of the inter-connection between the i-th and the j-th BD-RIS reflecting element. Moreover, $\boldsymbol{\Phi}_t$ is a unitary matrix, i.e.,

$$\boldsymbol{\Phi}_t^H \boldsymbol{\Phi}_t = \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^H = \boldsymbol{I}_M, \quad \forall t, \tag{3}$$

due to the circuit requirement [5]. Then, the received signal of the BS at time instant t is expressed as

$$\begin{aligned} \boldsymbol{y}_t^{\dagger} &= \sum\nolimits_{k=1}^{K} \boldsymbol{H}_{k,t} \sqrt{p} \boldsymbol{x}_{k,t} + \boldsymbol{n}_t \\ &= \sum\nolimits_{k=1}^{K} \left(\boldsymbol{D}_k + \boldsymbol{G}\boldsymbol{\Phi}_t \boldsymbol{R}_k\right) \sqrt{p} \boldsymbol{x}_{k,t} + \boldsymbol{n}_t, \quad t = 1, \cdots, T, \end{aligned} \tag{4}$$

where p denotes the identical transmit power of all users, $\boldsymbol{x}_{k,t} \in \mathbb{C}^{U \times 1}$ denotes the unit-power transmit signal of user k at time instant t, and $\boldsymbol{n}_t \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$ denotes the additive white Gaussian noise (AWGN) of the BS at time instant t.

In this paper, we consider the legacy two-stage transmission protocol for the uplink communications, where each coherence block of length T time instants is divided into the channel estimation stage consisting of $\tau < T$ time instants and data transmission stage consisting of $T - \tau$ time instant. Specifically, in Stage I, each user k transmits a sequence of τ pilot symbols, i.e., $\boldsymbol{x}_{k,t} = \boldsymbol{a}_{k,t} = [a_{k,1,t}, \cdots, a_{k,U,t}]^T$, $k = 1, \cdots, K, t = 1, \cdots, \tau$, to the BS, where $a_{k,u,t}$ is the pilot symbol transmitted by the u-th antenna of user k at time instant t. According to (4), at time instant t of Stage I, the received signal of the BS is

$$\boldsymbol{y}_t^{\dagger} = \sum\nolimits_{k=1}^{K} \left(\boldsymbol{D}_k + \boldsymbol{G}\boldsymbol{\Phi}_t \boldsymbol{R}_k\right) \sqrt{p} \boldsymbol{a}_{k,t} + \boldsymbol{n}_t, \quad t = 1, \cdots, \tau. \tag{5}$$

The task of the BS is then to perform channel estimation based on $\boldsymbol{y}_t^{\dagger}$, $t = 1, \cdots, \tau$.

In Stage II, the scattering matrix of the BD-RIS is fixed over different time instants [25], i.e., $\boldsymbol{\Phi}_t = \bar{\boldsymbol{\Phi}}$, $t = \tau + 1, \cdots, T$. According to (4), at time instant t of Stage II, the received signal of the BS is

$$\boldsymbol{y}_t^{\dagger} = \sum_{k=1}^{K} \left(\boldsymbol{D}_k + \boldsymbol{G}\bar{\boldsymbol{\Phi}} \boldsymbol{R}_k\right) \sqrt{p} \boldsymbol{x}_{k,t} + \boldsymbol{n}_t, \quad t = \tau + 1, \cdots, T, \tag{6}$$

where $\boldsymbol{x}_{k,t} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{S}_k)$ denotes the message signal of user k at time instant t. Then, the channel capacity of user k is

$$C_k = \log_2 \det \left(\boldsymbol{I}_N + \boldsymbol{C}_{y_k} \left(\sum\nolimits_{j \neq k} \boldsymbol{C}_{y_j} + \sigma^2 \boldsymbol{I}_N \right)^{-1} \right), \quad \forall k, \tag{7}$$

where $\boldsymbol{C}_{y_k} = p \bar{\boldsymbol{H}}_k \boldsymbol{S}_k \bar{\boldsymbol{H}}_k^H$, with $\bar{\boldsymbol{H}}_k = \boldsymbol{D}_k + \boldsymbol{G}\bar{\boldsymbol{\Phi}} \boldsymbol{R}_k$ denoting the effective channel between the BS and user k, $\forall k$. By applying the connection between vectorization and Kronecker product [36], we have

$$\boldsymbol{G}\bar{\boldsymbol{\Phi}} \boldsymbol{R}_k = \text{unvec}(\boldsymbol{J}_k \bar{\boldsymbol{\phi}}), \quad \forall k, \tag{8}$$

where $\bar{\boldsymbol{\phi}} = \text{vec}(\bar{\boldsymbol{\Phi}})$ and

$$\boldsymbol{J}_k = \boldsymbol{R}_k^T \otimes \boldsymbol{G} = \begin{bmatrix} \boldsymbol{Q}_{k,1,1} & \cdots & \boldsymbol{Q}_{k,1,M} \\ \vdots & \ddots & \vdots \\ \boldsymbol{Q}_{k,U,1} & \cdots & \boldsymbol{Q}_{k,U,M} \end{bmatrix} \in \mathbb{C}^{UN \times M^2}, \tag{9}$$

with

$$\boldsymbol{Q}_{k,u,m} = r_{k,u,m} \boldsymbol{G} \in \mathbb{C}^{N \times M}, \quad \forall k, u, m, \tag{10}$$

denoting the (u, m)-th sub-block of $\boldsymbol{J}_k$. Note that in the conventional RIS assisted communication, since the scattering matrix $\bar{\boldsymbol{\Phi}}$ is diagonal, for each user k, we need to estimate the channels corresponding to the diagonal elements in $\bar{\boldsymbol{\Phi}}$, i.e., $r_{k,u,1}\boldsymbol{g}_1, \cdots, r_{k,u,M}\boldsymbol{g}_M$, which consists of UNM channel coefficients [1], [30], [37]. In contrast, the non-diagonal scattering matrix of BD-RIS introduces inter-element connections, fundamentally complicating the cascaded channel structure. Specifically, as shown in (9), we need to estimate UNM^2 channel coefficients to estimate in each $\boldsymbol{J}_k$.

III. Problem Statement for Channel Estimation

It is observed from (7) and (8) that to maximize the network throughput by jointly optimizing the transmit covariance matrices $\boldsymbol{S}_k$'s of the users and the scattering matrix $\bar{\boldsymbol{\Phi}}$ at the BD-RIS, the BS has to estimate the user-BS channels $\boldsymbol{D}_k$'s and the user-RIS-BS cascaded channels $\boldsymbol{J}_k$'s in the channel estimation stage. Note that the direct user-BS channels $\boldsymbol{D}_k$'s can be obtained via conventional channel estimation techniques by turning off all the RIS reflecting elements. Therefore, in this paper, we assume that $\boldsymbol{D}_k$'s are perfectly known and focus on low-overhead methods to estimate $\boldsymbol{J}_k$'s. For simplicity, under the channel estimation stage, let us define the effective received signals by removing the contribution made by the user-BS channels as:

$$\begin{aligned} \boldsymbol{y}_t = \boldsymbol{y}_t^{\dagger} - \sum\nolimits_{k=1}^{K} \boldsymbol{D}_k \sqrt{p} \boldsymbol{a}_{k,t} &= \sum\nolimits_{k=1}^{K} \boldsymbol{G}\boldsymbol{\Phi}_t \boldsymbol{R}_k \sqrt{p} \boldsymbol{a}_{k,t} + \boldsymbol{n}_t, \\ &= \sum\nolimits_{k=1}^{K} \sqrt{p} \left(\boldsymbol{a}_{k,t}^T \otimes \boldsymbol{I}_N\right) \boldsymbol{J}_k \bar{\boldsymbol{\phi}}_t + \boldsymbol{n}_t, \quad t = 1, \cdots, \tau, \end{aligned} \tag{11}$$

where $\bar{\boldsymbol{\phi}}_t = \text{vec}(\boldsymbol{\Phi}_t)$. Because the received signals in (11) are linear functions of $\boldsymbol{J}_k$'s, one straightforward way to estimate them is to solve these linear functions. Because there are $KUNM^2$ unknown variables in $\boldsymbol{J}_1, \cdots, \boldsymbol{J}_K$, theoretically speaking, the minimum number of time instants required by the above estimation method is KUM^2, as shown in [25]–[27]. Note that for conventional RIS aided communication, the required overhead is $M + \lceil M(KU-1)/q \rceil$ [1]. The overhead required by the above scheme in BD-RIS aided communication is thus an order-of-magnitude higher than that in conventional RIS aided communication ($\mathcal{O}(M^2)$ versus $\mathcal{O}(M)$), which could potentially overwhelm the beamforming gain brought by BD-RIS. Therefore, in this paper, we consider the following two questions for estimating $\boldsymbol{J}_1, \cdots, \boldsymbol{J}_K$ based on $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_\tau$.

- Q1: In the ideal case without noise, i.e., $\boldsymbol{n}_t = \boldsymbol{0}$, $\forall t$, is it possible to perfectly estimate $\boldsymbol{J}_1, \cdots, \boldsymbol{J}_K$ with the overhead similar to that in conventional RIS aided communication as shown in [1]?

- Q2: In the practical case with noise, how to design efficient algorithms to estimate $\boldsymbol{J}_1, \cdots, \boldsymbol{J}_K$ given signals received by $\tau \geq \bar{\tau}$ time instants with small MSE?

This paper aims to break the limitation of the approaches proposed in [25]–[27] and provide positive answers to the above questions. Note that [25]–[27] treat all the entries in $\boldsymbol{J}_1, \cdots, \boldsymbol{J}_K$ as independent unknown variables. However, these entries are highly correlated. Specifically, according to (9) and (10), different sub-blocks of each $\boldsymbol{J}_k$ follow the relationship

$$\boldsymbol{Q}_{k,u,m} = \beta_{k,u,m} \boldsymbol{Q}_{1,1,1}, \quad \forall (k,u,m) \neq (1,1,1), \tag{12}$$

where

$$\beta_{k,u,m} = \frac{r_{k,u,m}}{r_{1,1,1}}. \tag{13}$$

In other words, after $\boldsymbol{Q}_{1,1,1}$ is estimated, it is sufficient to estimate a scalar $\beta_{k,u,m}$ to reconstruct the whole matrix $\boldsymbol{Q}_{k,u,m}$, $\forall (k,u,m) \neq (1,1,1)$. To summarize, in $\boldsymbol{J}_k$'s, $\forall k$, independent unknown variables are $\boldsymbol{Q}_{1,1,1}$ and $\beta_{k,u,m}$'s, $\forall (k,u,m) \neq (1,1,1)$. Thus, the total number of independent unknown variables to estimate in $\boldsymbol{J}_k$'s is $NM + KMU - 1$. This number is significantly smaller than the total number of unknown variables in $\boldsymbol{J}_k$'s, which is $KUNM^2$. This indicates that it is possible to estimate $\boldsymbol{J}_k$'s with much lower overhead compared to the existing schemes proposed in [25]–[27], which did not exploit the channel property shown in (12).

On the other hand, how to leverage the above channel property to address Q1 and Q2 is challenging. This is because if we treat $\boldsymbol{Q}_{1,1,1}$ and $\beta_{k,u,m}$'s as unknown variables, then the received signals in (11) are no longer linear functions of these variables, which complicates the theoretical analysis and algorithm design. In the rest of the paper, we first answer Q1 by considering a special case of a single-antenna user. This simple case can help shed the light on how the channel property shown in (12) can significantly reduce the channel estimation overhead. Then, based on the insights from this special case, we generalize the theoretical result of Q1 to the case of multiple multi-antenna users. At last, we will design an efficient channel estimation algorithm to address Q2.

IV. Channel Estimation Overhead in Special Case of a Single-Antenna User

Let us first focus on Q1 under the special single-user and single-antenna case, i.e., $K = 1$ and $U = 1$. In this case, we omit the subscripts k and u in the variables defined above. The user-RIS channel reduces to $\boldsymbol{r} = [r_1, \cdots, r_M]^T$. Then, the cascaded channel $\boldsymbol{J}$ can be expressed as

$$\boldsymbol{J} = \boldsymbol{r}^T \otimes \boldsymbol{G} = [\boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_M], \tag{14}$$

where

$$\boldsymbol{Q}_m = r_m \boldsymbol{G}, \quad m = 1, \cdots, M, \tag{15}$$

is the m-th sub-block of $\boldsymbol{J}$. Similar to (12), the correlation among $\boldsymbol{Q}_m$'s can be modeled as:

$$\boldsymbol{Q}_m = \beta_m \boldsymbol{Q}_1, \quad m \neq 1, \tag{16}$$

where

$$\beta_m = \frac{r_m}{r_1}, \tag{17}$$

denotes the scaling coefficient between $\boldsymbol{Q}_m$ and $\boldsymbol{Q}_1$. In the ideal case without noise, the received signal in (11) reduces to:

$$\boldsymbol{y}_t = \sqrt{p} a_t \boldsymbol{G} \boldsymbol{\Phi}_t \boldsymbol{r} = \sqrt{p} a_t \boldsymbol{J} \bar{\boldsymbol{\phi}}_t, \quad t = 1, \cdots, \tau. \tag{18}$$

As discussed in Section III, to reduce the number of estimated variables, we should utilize the channel property shown in (16) and (17) to replace $\boldsymbol{J}$ by $\boldsymbol{Q}_1$ and β_m's, $m = 2, \cdots, M$. Then, the received signal given in (18) reduces to

$$\boldsymbol{y}_t = \sqrt{p} a_t \left(\boldsymbol{Q}_1 \boldsymbol{\phi}_{t,1} + \sum\nolimits_{m=2}^{M} \beta_m \boldsymbol{Q}_1 \boldsymbol{\phi}_{t,m} \right), \quad \forall t. \tag{19}$$

The key challenge to characterize the number of time instants to estimate $\boldsymbol{Q}_1$ and $\beta_2, \cdots, \beta_M$ from the received signals shown in (19) lies in fact that $\boldsymbol{y}_t$'s are non-linear functions of $\boldsymbol{Q}_1$ and $\beta_2, \cdots, \beta_M$, due to the product terms $\beta_m \boldsymbol{Q}_1$, $m = 2, \cdots, M$. Note that without the unitary constraint (3), we can set $\boldsymbol{\phi}_{t,m} = \boldsymbol{0}$, $m = 2, \cdots, M$, in the first few time instants, such that the received signal in (19) is only contributed by the first RIS reflecting element. Therefore, we can estimate $\boldsymbol{Q}_1$ by solving linear functions. Then, in the following time instants, we can set $\boldsymbol{\phi}_{t,m} \neq \boldsymbol{0}$, $m = 2, \cdots, M$. Given $\boldsymbol{Q}_1$, the received signals given in (19) are linear functions of β_m's, and we can estimate them efficiently. However, for BD-RIS, due to the inter-connected circuits to control the reflecting elements, we cannot design a scattering matrix $\boldsymbol{\Phi}_t$ with $\boldsymbol{\phi}_{t,1} \neq \boldsymbol{0}$ and $\boldsymbol{\phi}_{t,m} = \boldsymbol{0}$, $m = 2, \cdots, M$, by shutting down the 2nd to the M-th reflecting elements. Such a scattering matrix does not satisfy the unitary constraint (3).

Although the above approach of shutting down some BD-RIS elements is not possible, the philosophy behind this approach to deal with the product terms $\beta_2 \boldsymbol{Q}_1, \cdots, \beta_M \boldsymbol{Q}_1$, is correct — we can sequentially estimate $\boldsymbol{Q}_1$ and $\beta_2, \cdots, \beta_M$ in two stages, while in each stage, we perform estimation via solving linear equations such that we can characterize the estimation overhead. The question is, when the above on/off strategy is not feasible to BD-RIS, how to design the BD-RIS scattering matrices and the user pilot signals over time so as to first estimate $\boldsymbol{Q}_1$ without the effect of $\beta_m \boldsymbol{Q}_1$, $m = 2, \cdots, M$, and then estimate $\beta_2, \cdots, \beta_M$ given $\boldsymbol{Q}_1$.

Our strategy is described as follows. We divide the overall τ time instants into two parts with the same length δ, i.e., $\tau = 2\delta$. The BS-RIS scattering matrices and the user pilot signals in these two parts are given as follows.

- At time instants $t = 1, \cdots, \delta$ in the first part, the user pilot signals are arbitrary non-zero scalars, i.e.,

$$a_t \neq 0. \tag{20}$$

Moreover, let $\boldsymbol{P} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_M] \in \mathbb{C}^{M \times M}$ be an arbitrary unitary matrix. Then, each m-th column of the scattering matrix at time instant t, i.e., $\boldsymbol{\Phi}_t = [\boldsymbol{\phi}_{t,1}, \cdots, \boldsymbol{\phi}_{t,M}]$, is set as

$$\boldsymbol{\phi}_{t,m} = \boldsymbol{p}_{((m+t-2) \bmod M)+1}, \quad m = 1, \cdots, M, \tag{21}$$

where mod denotes the modulo operation.

- At each time instant $t=\delta+1,...,2\delta$ in the second part, the user pilot signal is set as
$$a_t = a_{t-\delta}. \tag{22}$$
Moreover, the BD-RIS scattering matrix $\mathbf{\Phi}_t = [\boldsymbol{\phi}_{t,1},\cdots,\boldsymbol{\phi}_{t,M}]$ is set as
$$\begin{aligned}&\boldsymbol{\phi}_{t,1} = e^{\mathrm{j}\theta}\boldsymbol{\phi}_{t-\delta,1},\\&\boldsymbol{\phi}_{t,m} = \boldsymbol{\phi}_{t-\delta,m}, \quad m=2,\cdots,M,\end{aligned} \tag{23}$$
where j is the imaginary unit, and $\theta \in (0,2\pi)$ is an arbitrary phase shift. It can be shown that the scattering matrix at time instants $\delta+1,\cdots,2\delta$ also satisfies (3).

Given the above strategy, the received signals over $\tau=2\delta$ time instants can be re-written as
$$\boldsymbol{y}_t = \sqrt{p}a_t\boldsymbol{Q}_1\boldsymbol{\phi}_{t,1} + \sqrt{p}a_t\sum\nolimits_{m=2}^{M}\beta_m\boldsymbol{Q}_1\boldsymbol{\phi}_{t,m}, \tag{24}$$
$$\boldsymbol{y}_{\delta+t} = \sqrt{p}a_te^{\mathrm{j}\theta}\boldsymbol{Q}_1\boldsymbol{\phi}_{t,1} + \sqrt{p}a_t\sum\nolimits_{m=2}^{M}\beta_m\boldsymbol{Q}_1\boldsymbol{\phi}_{t,m}, \quad t=1,\cdots,\delta, \tag{25}$$
where $\boldsymbol{\phi}_{t,m}$'s are given in (21) and (23). Then, we have the following theorem.

Theorem 1. *When $\delta=M$, based on the received signals over $2M$ time instants as given in* (24) *and* (25)*, we can perfectly estimate $\boldsymbol{Q}_1$ and $\beta_2,\cdots,\beta_M$ almost surely, not matter what the rank of $\boldsymbol{G}$, i.e., q, is.*

Proof. We first prove that when $\delta=M$, it is sufficient to perfectly estimate $\boldsymbol{Q}_1$ from (24) and (25). When $\delta=M$, by subtracting $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_M$ from $\boldsymbol{y}_{M+1},\cdots,\boldsymbol{y}_{2M}$, respectively, the signal components contributed by $\beta_2\boldsymbol{Q}_1$ to $\beta_M\boldsymbol{Q}_1$ can be eliminated, obtaining M effective received signals merely contributed by $\boldsymbol{Q}_1$. The t-th effective received signal is
$$\bar{\boldsymbol{y}}_t = \boldsymbol{y}_{M+t} - \boldsymbol{y}_t = \sqrt{p}a_t(e^{\mathrm{j}\theta}-1)\boldsymbol{Q}_1\boldsymbol{\phi}_{t,1}, \quad t=1,\cdots,M. \tag{26}$$
Then, the overall effective received signal is expressed as
$$\bar{\boldsymbol{Y}}_1 = [\bar{\boldsymbol{y}}_1,\cdots,\bar{\boldsymbol{y}}_M] = \sqrt{p}(e^{\mathrm{j}\theta}-1)\boldsymbol{Q}_1\mathbf{\Psi}_1, \tag{27}$$
where $\mathbf{\Psi}_1 = [a_1\boldsymbol{\phi}_{1,1},\cdots,a_M\boldsymbol{\phi}_{M,1}] \in \mathbb{C}^{M\times M}$. $\bar{\boldsymbol{Y}}_1$ is a linear function of $\boldsymbol{Q}_1$. According to (21), it can be shown that
$$\mathbf{\Psi}_1 = [a_1\boldsymbol{\phi}_{1,1},\cdots,a_M\boldsymbol{\phi}_{M,1}] = [a_1\boldsymbol{p}_1,\cdots,a_M\boldsymbol{p}_M]. \tag{28}$$
Because $\boldsymbol{P}$ is an unitary matrix, we have $\mathrm{rank}(\mathbf{\Psi}_1) = \mathrm{rank}(\boldsymbol{P}) = M$. Then, $\boldsymbol{Q}_1$ can be perfectly estimated as
$$\boldsymbol{Q}_1 = (\sqrt{p}(e^{\mathrm{j}\theta}-1))^{-1}\bar{\boldsymbol{Y}}_1\mathbf{\Psi}_1^H(\mathbf{\Psi}_1\mathbf{\Psi}_1^H)^{-1}. \tag{29}$$

After $\boldsymbol{Q}_1$ is known, we then prove that when $\delta=M$, it is sufficient to perfectly $\beta_2,\cdots,\beta_M$ from (24) and (25). We just focus on the signals received in the first M time instants given in (24), because those received in the following M time instants given in (25) contain the same information about β_m's. By removing the signals contributed by the first reflecting element, we obtain the following signals according to (24):
$$\begin{aligned}\tilde{\boldsymbol{y}}_t &= \boldsymbol{y}_t - \sqrt{p}a_t\boldsymbol{Q}_1\boldsymbol{\phi}_{t,1} = \sqrt{p}a_t\sum\nolimits_{m=2}^{M}\beta_m\boldsymbol{Q}_1\boldsymbol{\phi}_{t,m},\\&= \boldsymbol{F}_t\bar{\boldsymbol{\beta}}, \quad t=1,\cdots,M,\end{aligned} \tag{30}$$
where $\boldsymbol{F}_t = \sqrt{p}a_t\boldsymbol{Q}_1[\boldsymbol{\phi}_{t,2},\cdots,\boldsymbol{\phi}_{t,M}] \in \mathbb{C}^{N\times(M-1)}, t=1,\cdots,M$, $\bar{\boldsymbol{\beta}} = [\beta_2,\cdots,\beta_M]^T$. Define $\tilde{\boldsymbol{y}}^{(1)} = [\tilde{\boldsymbol{y}}_1^T,\cdots,\tilde{\boldsymbol{y}}_M^T]^T$. It then follows that
$$\tilde{\boldsymbol{y}}^{(1)} = \mathbf{\Theta}_1\bar{\boldsymbol{\beta}}, \tag{31}$$
where
$$\mathbf{\Theta}_1 = \left[\boldsymbol{F}_1^T,\cdots,\boldsymbol{F}_M^T\right]^T \in \mathbb{C}^{MN\times(M-1)}. \tag{32}$$

Proposition 1. *For $\mathbf{\Theta}_1$ given in* (32)*, we have* $\mathrm{rank}(\mathbf{\Theta}_1) = M-1$ *almost surely.*

Proof. Please refer to Appendix A. ■

According to Proposition 1, we can perfectly estimate $\beta_2,\cdots,\beta_M$ as
$$\bar{\boldsymbol{\beta}} = (\mathbf{\Theta}_1^H\mathbf{\Theta}_1)^{-1}\mathbf{\Theta}_1^H\tilde{\boldsymbol{y}}^{(1)}. \tag{33}$$
This completes the proof of Theorem 1. ■

Theorem 1 indicates that
$$\bar{\tau} = 2M, \tag{34}$$
time instants are sufficient to perfectly estimate $\boldsymbol{J}$. Via this one single-antenna user example, we have shown how to utilize the channel property in (16) to estimate $\boldsymbol{Q}_1$ and $\beta_2,\cdots,\beta_M$, instead of $\boldsymbol{J}$ directly as in other works [25]–[27]. This method can reduce the channel estimation overhead from M^2 time instants as required by [25]–[27] to $2M$. The key reason for this reduction is that instead of estimating all the NM^2 coefficients in $\boldsymbol{J}$ in an independent manner, our scheme leverages the property in (16) to estimate only a $N\times M$ reference channel matrix $\boldsymbol{Q}_1$ and $M-1$ scaling coefficients $\beta_2,\cdots,\beta_M$, which fundamentally reduces the number of variables to be estimated from NM^2 to $NM+M-1$. Although the received signals are not linear functions of $\boldsymbol{Q}_1,\beta_2,\cdots,\beta_M$ as shown in (19) due to the product terms $\beta_m\boldsymbol{Q}_1$'s, we manage to devise a pilot transmission and BD-RIS scattering mechanism based on (20)-(23) such that the order of the channel estimation overhead is reduced from $\mathcal{O}(M^2)$ to $\mathcal{O}(M)$, as shown in Theorem 1.

In the rest of this paper, we first generalize this channel estimation overhead result to the general case with multiple multi-antenna users in Section V. Then, we will propose an efficient channel estimation algorithm for the practical case with noise in Section VI.

Remark 1. *In the case when $\boldsymbol{Q}_1$ is not a full-rank matrix and its rank q is known as prior information, we can potentially use $\tau < 2M$ time instants to estimate the channel, as there are less than MN independent unknowns in $\boldsymbol{Q}_1$. However, before we estimate the channels, we do not know the value of q, and we have to quantify the number of time instants that is required for channel estimation given any value of q, which is $2M$ shown in Theorem 1.*

Remark 2. *Note that under the conventional RIS scenario, the required number of time instants to estimate the cascaded channels for beamforming design in the case of a single-antenna user is M [37]. On one hand, under BD-RIS, our proposed method leads to a channel estimation overhead that is still linear in M, same as that in the conventional RIS* *scenario, although the number of unknown variables in* $\boldsymbol{J}$ *is quadratic in* M*. This is because our scheme can utilize the channel property shown in* (16) *to reduce the number of independent variables to be estimated. On the other hand, the inter-connected circuits to control different reflecting elements raise the channel estimation overhead from* M *to* $2M$*. There are two reasons. First, the number of independent unknown variables is increased from* MN *to* $MN+M-1$*. Second, because of the unitary constraint on scattering matrix, we have less design flexibility. We need to adopt the design rule in* (22) *and* (23) *for creating* M *pair of received signals, each covering two time instants.*

V. Channel Estimation Overhead in General Case of Multiple Multi-Antenna Users

In this section, we consider Q1 under the general case of multiple multi-antenna users, i.e., $K>1$ and $U>1$. In the ideal case without noise, the received signal in (11) reduces to

$$\boldsymbol{y}_t=\sum\nolimits_{k=1}^{K}\boldsymbol{G}\boldsymbol{\Phi}_t\boldsymbol{R}_k\sqrt{p}\boldsymbol{a}_{k,t}=\sum\nolimits_{k=1}^{K}\sqrt{p}(\boldsymbol{a}_{k,t}^T\otimes\boldsymbol{I}_N)\boldsymbol{J}_k\bar{\boldsymbol{\phi}}_t. \tag{35}$$

Similar to the special case in Section IV, to reduce the number of variables to estimate, we should utilize the channel property shown in (12) and (13) to replace $\boldsymbol{J}_k$'s by $\boldsymbol{Q}_{1,1,1}$ and $\beta_{k,u,m}$'s, $(k,u,m)\neq(1,1,1)$. According to the cascaded channel given in (9) and the channel property in (12) and (13), $\boldsymbol{J}_k$ can be re-expressed as

$$\boldsymbol{J}_k=\boldsymbol{B}_k^T\otimes\boldsymbol{Q}_{1,1,1},\quad\forall k, \tag{36}$$

where

$$\boldsymbol{B}_k=[\boldsymbol{\beta}_{k,1},\cdots,\boldsymbol{\beta}_{k,U}]\in\mathbb{C}^{M\times U},\quad\forall k, \tag{37}$$

with $\boldsymbol{\beta}_{k,u}=[\beta_{k,u,1},\cdots,\beta_{k,u,M}]^T$, $\forall u$, and $\beta_{1,1,1}=1$. Then, by substituting (36) into (35), the received pilot signal given in (35) reduces to

$$\boldsymbol{y}_t=\sum\nolimits_{k=1}^{K}\boldsymbol{Q}_{1,1,1}\boldsymbol{\Phi}_t\boldsymbol{B}_k\sqrt{p}\boldsymbol{a}_{k,t},\quad t=1,\cdots,\tau. \tag{38}$$

Similar to the special case in Section IV, the challenge is that $\boldsymbol{y}_t$'s given in (38) are non-linear functions of $\boldsymbol{Q}_{1,1,1}$ and $\boldsymbol{B}_k$'s. Recall that in Section IV, we tackled this nonlinearity by first isolating and estimating the reference cascaded channel $\boldsymbol{Q}_{1,1,1}$ adopting the design of BD-RIS scattering matrices and pilot signals as (20)-(23), and then estimating the corresponding scaling coefficients. By mimicking this approach, we extend it to a two-phase channel estimation protocol for the general case of multiple multi-antenna users such that $\boldsymbol{Q}_{1,1,1}$ and $\boldsymbol{B}_k$'s can be estimated separately and effectively. The overall channel estimation protocol is summarized in Fig. 2, where in Phase I with $\tau_1<\tau$ time instants, the cascaded channel $\boldsymbol{Q}_{1,1,1}$ and scaling coefficients $\beta_{1,1,2},\cdots,\beta_{1,1,M}$ are estimated based on $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{\tau_1}$ for the 1st antenna of user 1; while in Phase II with $\tau_2=\tau-\tau_1$ time instants, the scaling coefficients $\beta_{k,u,m}$'s, $(k,u)\neq(1,1)$, are estimated based on $\boldsymbol{y}_{\tau_1+1},\cdots,\boldsymbol{y}_{\tau_1+\tau_2}$ for the 2nd to U-th antennas of user 1 and all antennas of users 2 to K. In the following, we introduce the implementation details of Phase I and Phase II, respectively.

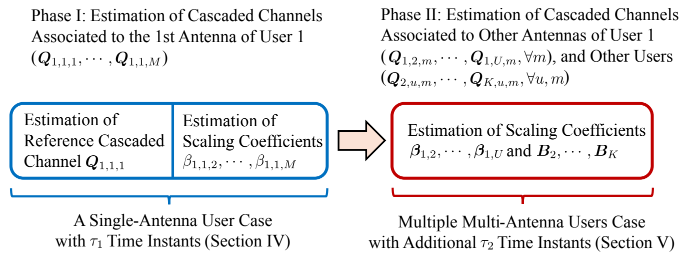


Fig. 2. Illustration of the two-phase channel estimation protocol.

A. Phase I: Estimation of the Cascaded Channel Associated with the 1st Antenna of User 1

Specifically, in Phase I at time instants $t=1,\cdots,\tau_1$, we only allow the 1st antenna of user 1 to transmit non-zero pilot symbols to the BS, while all the other antennas keep silent, i.e.,

$$a_{k,u,t}=\begin{cases}a_t, & \text{if } k=1,u=1,\\ 0, & \text{otherwise},\end{cases}\quad t=1,\cdots,\tau_1. \tag{39}$$

Then, the received signals of the BS in (38) over τ_1 time instants can be re-written as

$$\boldsymbol{y}_t=\sqrt{p}a_t\left(\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,1}+\sum\nolimits_{m=2}^{M}\beta_{1,1,m}\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,m}\right),\quad t=1,\cdots,\tau_1, \tag{40}$$

$\boldsymbol{y}_t$ given in (40) is the same as that in the special single-user and single-antenna case in (19). As a result, we can apply the same method as Section IV to estimate unknown variables of $\boldsymbol{Q}_{1,1,1}$ and $\beta_{1,1,2},\cdots,\beta_{1,1,M}$. Specifically, we divide the overall τ_1 time instants in Phase I into two parts with the same length δ, i.e., $\tau_1=2\delta$. The BD-RIS scattering matrices and the user pilot signals at different time instants follow (20)-(23), According to Theorem 1, we can utilize

$$\bar{\tau}_1=2M, \tag{41}$$

time instants to perfectly estimate $\boldsymbol{Q}_{1,1,1}$ and $\beta_{1,1,2},\cdots,\beta_{1,1,M}$ based on $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{\tau_1}$ in (40) according to (29) and (33), respectively, no matter what the value of q is.

B. Phase II: Estimation of the Scaling Coefficients of the Other Antennas of User 1 and the Other Users

After $\boldsymbol{Q}_{1,1,1}$ and $\beta_{1,1,2},\cdots,\beta_{1,1,M}$ are estimated in Phase I, the remaining unknown variables to estimate reduce to

$$\bar{\boldsymbol{B}}=[\boldsymbol{\beta}_{1,2},\cdots,\boldsymbol{\beta}_{1,U},\boldsymbol{B}_2,\cdots,\boldsymbol{B}_K]\in\mathbb{C}^{M\times(KU-1)}. \tag{42}$$

In the following, we introduce how to estimate $\bar{\boldsymbol{B}}$ in Phase II. Specifically, in Phase II at time instants $t=\tau_1+1,\cdots,\tau_1+\tau_2$, we keep the 1st antenna of user 1 silent, i.e.,

$$a_{1,1,t}=0,\quad t=\tau_1+1,\cdots,\tau_1+\tau_2, \tag{43}$$

while for the other antennas of user 1 and all antennas of users 2 to K, we define the overall transmit pilot symbols as $\bar{\boldsymbol{a}}_t=[a_{1,2,t},\cdots,a_{1,U,t},\boldsymbol{a}_{2,t}^T,\cdots,\boldsymbol{a}_{K,t}^T]^T=$

$[\bar{a}_{1,t},\cdots,\bar{a}_{KU-1,t}]^T\in\mathbb{C}^{(KU-1)\times 1}$. Then, the received signals in (38) over τ_2 time instants can be re-written as

$$\boldsymbol{y}_t=\boldsymbol{Q}_{1,1,1}\boldsymbol{\Phi}_t\bar{\boldsymbol{B}}\sqrt{p}\bar{\boldsymbol{a}}_t=\boldsymbol{T}_t\bar{\boldsymbol{b}},\quad t=\tau_1+1,\cdots,\tau_1+\tau_2, \tag{44}$$

where

$$\boldsymbol{T}_t=\sqrt{p}\bar{\boldsymbol{a}}_t^T\otimes\boldsymbol{Q}_{1,1,1}\boldsymbol{\Phi}_t\in\mathbb{C}^{N\times M(KU-1)}, \tag{45}$$

and $\bar{\boldsymbol{b}}=\mathrm{vec}(\bar{\boldsymbol{B}})$. The overall received signal of the BS over Phase II is given by

$$\boldsymbol{y}^{(2)}=[\boldsymbol{y}_{\tau_1+1}^T,\cdots,\boldsymbol{y}_{\tau_1+\tau_2}^T]^T=\boldsymbol{\Theta}_2\bar{\boldsymbol{b}}, \tag{46}$$

where

$$\boldsymbol{\Theta}_2=\left[\boldsymbol{T}_{\tau_1+1}^T,\cdots,\boldsymbol{T}_{\tau_1+\tau_2}^T\right]^T\in\mathbb{C}^{N\tau_2\times M(KU-1)}. \tag{47}$$

According to (10), $\mathrm{rank}(\boldsymbol{Q}_{1,1,1})=\mathrm{rank}(\boldsymbol{G})=q$. Note that after $\boldsymbol{Q}_{1,1,1}$ is estimated, we can obtain the specific value of q. Then, according to (45) and (47), we have $\mathrm{rank}(\boldsymbol{\Theta}_2)\leq\tau_2 q$. Therefore, to perfectly recover $M(KU-1)$ unknowns in $\bar{\boldsymbol{b}}$ based on the linear equations (46), the number of time instants τ_2 should satisfy $\tau_2\geq\lceil\frac{M(KU-1)}{q}\rceil$. In the following, we design a strategy of BD-RIS scattering matrices and user pilot signals such that the lower bound of the number of time instants can be achieved, which is equivalent to showing that when $\tau_2=\lceil\frac{M(KU-1)}{q}\rceil$, the matrix $\boldsymbol{\Theta}_2$ have a full rank of $M(KU-1)$.

Given $\mathrm{rank}(\boldsymbol{Q}_{1,1,1})=q$, the singular value decomposition (SVD) of $\boldsymbol{Q}_{1,1,1}$ is expressed as $\boldsymbol{Q}_{1,1,1}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are $N\times N$ and $M\times M$ unitary matrices, respectively, and $\boldsymbol{\Sigma}$ is a $N\times M$ rectangular diagonal matrix with q non-zero singular values, denoted by $\sigma_1,\cdots,\sigma_q$. For simplicity, we assume that these singular values are sorted in a descending order, i.e., $\sigma_1\geq\cdots\geq\sigma_q$. Note that

$$\boldsymbol{U}^H\boldsymbol{T}_t=\bar{\boldsymbol{a}}_t^T\otimes\boldsymbol{U}^H\boldsymbol{Q}_{1,1,1}\boldsymbol{\Phi}_t=\bar{\boldsymbol{a}}_t^T\otimes\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_t,\quad\forall t. \tag{48}$$

Therefore, $\boldsymbol{\Theta}_2$ has the same rank as the following matrix

$$\begin{aligned}\tilde{\boldsymbol{\Theta}}_2&=(\boldsymbol{I}_{\tau_2}\otimes\boldsymbol{U}^H)\boldsymbol{\Theta}_2\\&=\begin{bmatrix}\bar{a}_{1,\tau_1+1}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_1+1}&\cdots&\bar{a}_{KU-1,\tau_1+1}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_1+1}\\\vdots&\ddots&\vdots\\\bar{a}_{1,\tau_1+\tau_2}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_1+\tau_2}&\cdots&\bar{a}_{KU-1,\tau_1+\tau_2}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_1+\tau_2}\end{bmatrix}.\end{aligned} \tag{49}$$

In the following, we show that when $\tau_2=\lceil\frac{M(KU-1)}{q}\rceil$, denoted by $\bar{\tau}_2$, such that the number of rows is the minimal integer that is no smaller than the number of columns, we can make $\mathrm{rank}(\tilde{\boldsymbol{\Theta}}_2)=M(KU-1)$.

Define U_o as the smallest integer such that $\frac{MU_o}{q}$ is an integer. Next, define $\eta=\lfloor\frac{KU-1}{U_o}\rfloor$ and $U_\nu=KU-1-\eta U_o$. Our job is to divide the $M(KU-1)$ columns in $\tilde{\boldsymbol{\Theta}}_2$ into $\eta+1$ groups, where in each of the first η groups, there are MU_o columns, while in the last block, there are MU_ν columns (note that when $\frac{KU-1}{U_o}$ is an integer, there are just η groups, because $U_\nu=0$). Our goal is to set many user pilot signals as zero such that $\tilde{\boldsymbol{\Theta}}_2$ is a diagonal block matrix, i.e., the non-zero rows in different blocks are non-overlapping. Then, all the columns tend to be linearly independent.

To achieve this goal, we aim to make $\tilde{\boldsymbol{\Theta}}_2$ consist of $\eta+1$ diagonal blocks, denoted by $\tilde{\boldsymbol{\Theta}}_{2,s}\in\mathbb{C}^{\tau_o q\times MU_o}$, $s=1,\cdots,\eta$, and $\tilde{\boldsymbol{\Theta}}_{2,\eta+1}\in\mathbb{C}^{\tau_\nu q\times MU_\nu}$. If we can make each block full rank, i.e., $\mathrm{rank}(\tilde{\boldsymbol{\Theta}}_{2,s})=MU_o$, $s=1,\cdots,\eta$, and $\mathrm{rank}(\tilde{\boldsymbol{\Theta}}_{2,\eta+1})=MU_\nu$, then each column in $\tilde{\boldsymbol{\Theta}}_{2,s}$, $s=1,\cdots,\eta+1$, cannot be linear combination of others in the same block. Moreover, owing to the diagonal block structure of $\tilde{\boldsymbol{\Theta}}_2$, each column in $\tilde{\boldsymbol{\Theta}}_2$ cannot be linear combination of others. In this case, we have $\mathrm{rank}(\tilde{\boldsymbol{\Theta}}_2)=\sum_{s=1}^{\eta+1}\mathrm{rank}(\tilde{\boldsymbol{\Theta}}_{2,s})=\eta MU_o+MU_\nu=M(KU-1)$. In the following, we set user pilot signals and BD-RIS scattering matrices such that $\tilde{\boldsymbol{\Theta}}_2$ consists of $\eta+1$ diagonal blocks $\tilde{\boldsymbol{\Theta}}_{2,1},\cdots,\tilde{\boldsymbol{\Theta}}_{2,\eta+1}$, and each block $\tilde{\boldsymbol{\Theta}}_{2,s}$, $\forall s$, is full rank.

To achieve the above aims, we divide the overall $\bar{\tau}_2$ time instants into $\eta+1$ non-overlapping groups, i.e.,

$$\begin{cases}\mathcal{T}_s=\{\tau_1+(s-1)\tau_o+1,\cdots,\tau_1+s\tau_o\},\quad s=1,\cdots,\eta,\\\mathcal{T}_{\eta+1}=\{\tau_1+\eta\tau_o+1,\cdots,\tau_1+\eta\tau_o+\tau_\nu\},\end{cases} \tag{50}$$

where $\tau_o=\frac{MU_o}{q}$, and $\tau_\nu=\left\lceil\frac{MU_\nu}{q}\right\rceil$, and divide the overall $KU-1$ user antennas into $\eta+1$ non-overlapping groups, i.e.,

$$\begin{cases}\mathcal{U}_s=\{(s-1)U_o+1,\cdots,sU_o\},\quad s=1,\cdots,\eta,\\\mathcal{U}_{\eta+1}=\{\eta U_o+1,\cdots,\eta U_o+U_\nu\}.\end{cases} \tag{51}$$

Then, the s-th group of user antennas only transmit pilot signals during the s-th group of time instants, $s=1,\cdots,\eta+1$, i.e.,

$$\bar{a}_{j,t}=0,\quad j\in\mathcal{U}_s,\ t\in\mathcal{T}_{s'},\ s\neq s'. \tag{52}$$

In this case, $\tilde{\boldsymbol{\Theta}}_2$ is a diagonal block matrix, where each of the first η diagonal block $\tilde{\boldsymbol{\Theta}}_{2,s}$, consists of $\tau_o\times U_o$ sub-blocks, with the sub-block in the i-th block row and j-th block column being $\bar{a}_{(s-1)U_o+j,\tau_{o,s}+i}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_{o,s}+i}$, where $s=1,\cdots,\eta$, $\tau_{o,s}=\tau_1+(s-1)\tau_o$, $i=1,\cdots,\tau_o$, $(s-1)U_o+j\in\mathcal{U}_s$. The last diagonal block $\tilde{\boldsymbol{\Theta}}_{2,\eta+1}$ consists of $\tau_\nu\times U_\nu$ sub-blocks, with the sub-block in the i-th row block and j-th column block being $\bar{a}_{\eta U_o+j,\tau_{o,\eta+1}+i}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_{o,\eta+1}+i}$, $i=1,\cdots,\tau_\nu$, $\eta U_o+j\in\mathcal{U}_{\eta+1}$.

Next, we set $\bar{a}_{j,t}$'s and $\boldsymbol{\Phi}_t$'s within each of $\tilde{\boldsymbol{\Theta}}_{2,s}$, $s=1,\cdots,\eta+1$, such that it achieves full rank. To achieve this, we let each user antenna transmit pilot signals during ϵ or $\epsilon+1$ consecutive time instants, where $\epsilon=\lceil\frac{M}{q}\rceil$. The set of time instants allocated to the $[(s-1)U_o+j]$-th antenna is defined as

$$\begin{cases}\Upsilon_{s,j}=\{\tau_{o,s}+1,\cdots,\tau_{o,s}+\epsilon\}, & \text{if } j=1,\\\Upsilon_{s,j}=\{\tau_{o,s}+(j-1)(\epsilon-1)+\lceil\frac{(j-1)\rho}{q}\rceil,\cdots\\\qquad\cdots,\tau_{o,s}+j(\epsilon-1)+\lceil\frac{j\rho}{q}\rceil\}, & \text{if } j\geq 2,\end{cases}$$
$$s=1,\cdots,\eta+1,\ (s-1)U_o+j\in\mathcal{U}_s, \tag{53}$$

where $\rho=\mathrm{mod}(\frac{M}{q})$. Then, the user pilot signals during time instants $t=\tau_1+1,\cdots,\tau_1+\bar{\tau}_2$ are set as follows

$$\begin{cases}\bar{a}_{(s-1)U_o+j,t}=c_t, & \text{if } t\in\Upsilon_{s,j},\\\bar{a}_{(s-1)U_o+j,t}=0, & \text{otherwise},\end{cases}$$
$$s=1,\cdots,\eta+1,\ (s-1)U_o+j\in\mathcal{U}_s, \tag{54}$$

where c_t is a non-zero scalar. Moreover, let $\boldsymbol{P}=[\boldsymbol{p}_1,\cdots,\boldsymbol{p}_M]\in\mathbb{C}^{M\times M}$ be an arbitrary unitary matrix, then the BD-RIS scattering matrices $\mathbf{\Phi}_t$'s at time instants $t = \tau_1+1,\cdots,\tau_1+\bar{\tau}_2$ are set as

$$\mathbf{\Phi}_t = \boldsymbol{V}\boldsymbol{P}(S_{M,t},:), \quad t\in\mathcal{T}_s,\ s=1,\cdots,\eta+1, \tag{55}$$

where

$$S_{M,t} = \mathrm{circshift}(S_{M,0},(t-\tau_{o,s}-1)q \bmod M), \tag{56}$$

with circshift denoting the to-the-left circular shifting operation, and $S_{M,0}=\{1,\cdots,M\}$. In this case, each $a_{j,t}\mathbf{\Sigma}\boldsymbol{V}^H\mathbf{\Phi}_t$ is a scaled version of the first q rows of $\boldsymbol{P}(S_{M,t},:), \forall j,t$. With $S_{M,t}$ given in (56), the rows in $\boldsymbol{P}$ are permuted in a circular way along the domain of time instants in $\tilde{\mathbf{\Theta}}_2$. Then, we have the following theorem.

Theorem 2. *When $\tau_2 = \lceil\frac{M(KU-1)}{q}\rceil$, if the user pilot signals are set according to* (54)*, and the BS-RIS scattering matrices are set according to* (55)*, then we can have* $\mathrm{rank}(\tilde{\mathbf{\Theta}}_2) = M(KU-1)$.

Proof. Please refer to Appendix B. ■

According to Theorem 2, it is sufficient to perfectly estimate $\bar{\boldsymbol{b}}$ with

$$\bar{\tau}_2 = \left\lceil\frac{M(KU-1)}{q}\right\rceil, \tag{57}$$

time instants, and the estimator of $\bar{\boldsymbol{b}}$ is given by

$$\bar{\boldsymbol{b}} = (\mathbf{\Theta}_2^H\mathbf{\Theta}_2)^{-1}\mathbf{\Theta}_2^H\boldsymbol{y}^{(2)}. \tag{58}$$

To further explain the above design of user pilot signals and BD-RIS scattering matrices, we provide a simple example as follows.

Example 1. *Consider the case when $M=3$, $K=4$, $U=1$, $N=2$, and $q=2$, while the channel for user* 1*, i.e., $\boldsymbol{Q}_{1,1,1}$, has already been estimated in Phase I using $\tau_1=2M$ time instants. In the following, we show that when $\tau_2=5$,* $\mathrm{rank}(\tilde{\mathbf{\Theta}}_2)=9$. *According to* (53)*, user* 2*,* 3*, and* 4 *transmits pilot signals at time instants $t=2M+1,2M+2$, $t=2M+2,2M+3$, and $t=2M+4,2M+5$, respectively. Moreover, according to* (54)*, the transmitted pilot signals at time instants $t=2M+1,\cdots,2M+5$ are*

$$\begin{aligned}
&[\bar{a}_{2,2M+1},\bar{a}_{3,2M+1},\bar{a}_{4,2M+1}]^T = [c_{2M+1},0,0]^T,\\
&[\bar{a}_{2,2M+2},\bar{a}_{3,2M+2},\bar{a}_{4,2M+2}]^T = [c_{2M+2},c_{2M+2},0]^T,\\
&[\bar{a}_{2,2M+3},\bar{a}_{3,2M+3},\bar{a}_{4,2M+3}]^T = [0,c_{2M+3},0]^T,\\
&[\bar{a}_{2,2M+4},\bar{a}_{3,2M+4},\bar{a}_{4,2M+4}]^T = [0,0,c_{2M+4}]^T,\\
&[\bar{a}_{2,2M+5},\bar{a}_{3,2M+5},\bar{a}_{4,2M+5}]^T = [0,0,0,c_{2M+5}]^T,
\end{aligned} \tag{59}$$

respectively. Then, according to (55)*, the BD-RIS scattering matrix at time instants $t=2M+1,\cdots,2M+5$ are set as*

$$\begin{aligned}
&\mathbf{\Phi}_{2M+1} = \boldsymbol{V}[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3]^T,\ \mathbf{\Phi}_{2M+2} = \boldsymbol{V}[\boldsymbol{p}_3,\boldsymbol{p}_1,\boldsymbol{p}_2]^T,\\
&\mathbf{\Phi}_{2M+3} = \boldsymbol{V}[\boldsymbol{p}_2,\boldsymbol{p}_3,\boldsymbol{p}_1]^T,\ \mathbf{\Phi}_{2M+4} = \boldsymbol{V}[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3]^T,\\
&\mathbf{\Phi}_{2M+5} = \boldsymbol{V}[\boldsymbol{p}_3,\boldsymbol{p}_1,\boldsymbol{p}_2]^T,
\end{aligned} \tag{60}$$

respectively. Thus, based on (49)*, $\tilde{\mathbf{\Theta}}_2$ is given as*

$$\tilde{\mathbf{\Theta}}_2 = \begin{bmatrix}
c_{2M+1}\mathbf{\Sigma}[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3]^T & \boldsymbol{O} & \boldsymbol{O}\\
c_{2M+2}\mathbf{\Sigma}[\boldsymbol{p}_3,\boldsymbol{p}_1,\boldsymbol{p}_2]^T & c_{2M+2}\mathbf{\Sigma}[\boldsymbol{p}_3,\boldsymbol{p}_1,\boldsymbol{p}_2]^T & \boldsymbol{O}\\
\boldsymbol{O} & c_{2M+3}\mathbf{\Sigma}[\boldsymbol{p}_2,\boldsymbol{p}_3,\boldsymbol{p}_2]^T & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & c_{2M+4}\mathbf{\Sigma}[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3]^T\\
\boldsymbol{O} & \boldsymbol{O} & c_{2M+5}\mathbf{\Sigma}[\boldsymbol{p}_3,\boldsymbol{p}_1,\boldsymbol{p}_2]^T
\end{bmatrix}, \tag{61}$$

where $\mathbf{\Sigma} = \begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\end{bmatrix}$ and $[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3]$ is a unitary matrix. By setting $c_{2M+t}=1$, $t=1,\cdots,5$, and eliminating all the zero rows, we can obtain a matrix $\bar{\mathbf{\Theta}}_2$ as follows

$$\bar{\mathbf{\Theta}}_2 = \begin{bmatrix}
\sigma_1\boldsymbol{p}_1^T & \mathbf{0}^T & \mathbf{0}^T\\
\sigma_2\boldsymbol{p}_2^T & \mathbf{0}^T & \mathbf{0}^T\\
\sigma_1\boldsymbol{p}_3^T & \sigma_1\boldsymbol{p}_3^T & \mathbf{0}^T\\
\sigma_2\boldsymbol{p}_1^T & \sigma_2\boldsymbol{p}_1^T & \mathbf{0}^T\\
\mathbf{0}^T & \sigma_1\boldsymbol{p}_2^T & \mathbf{0}^T\\
\mathbf{0}^T & \sigma_2\boldsymbol{p}_3^T & \mathbf{0}^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_1\boldsymbol{p}_1^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_2\boldsymbol{p}_2^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_1\boldsymbol{p}_3^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_2\boldsymbol{p}_1^T
\end{bmatrix}. \tag{62}$$

$\bar{\mathbf{\Theta}}_2$ has the same rank as $\tilde{\mathbf{\Theta}}_2$ because it is built by eliminating the zero-rows in $\tilde{\mathbf{\Theta}}_2$. Then, by setting the 4*th row as the difference between the* 4*th row and the* 1*st row multiplied by $\frac{\sigma_2}{\sigma_1}$, the* 3*rd row as the difference between the* 3*rd row and the* 6*th row multiplied by $\frac{\sigma_1}{\sigma_2}$, and the* 7*th row as the difference between the* 7*th row and the* 10*th row multiplied by $\frac{\sigma_1}{\sigma_2}$, we obtain the following matrix*

$$\mathbf{\Theta}_2^\dagger = \begin{bmatrix}
\sigma_1\boldsymbol{p}_1^T & \mathbf{0}^T & \mathbf{0}^T\\
\sigma_2\boldsymbol{p}_2^T & \mathbf{0}^T & \mathbf{0}^T\\
\sigma_1\boldsymbol{p}_3^T & \mathbf{0}^T & \mathbf{0}^T\\
\mathbf{0}^T & \sigma_2\boldsymbol{p}_1^T & \mathbf{0}^T\\
\mathbf{0}^T & \sigma_1\boldsymbol{p}_2^T & \mathbf{0}^T\\
\mathbf{0}^T & \sigma_2\boldsymbol{p}_3^T & \mathbf{0}^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_1\boldsymbol{p}_1^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_2\boldsymbol{p}_2^T\\
\mathbf{0}^T & \mathbf{0}^T & \sigma_1\boldsymbol{p}_3^T\\
\mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T
\end{bmatrix}. \tag{63}$$

$\mathbf{\Theta}_2^\dagger$ has the same rank as $\bar{\mathbf{\Theta}}_2$, because the former is obtained via elementary row operation of the latter. It can be shown that each of the first 9 *rows in $\mathbf{\Theta}_2^\dagger$ cannot be obtained via linear combination of the other rows. For example, when we look at the* 1*st row, i.e., $\sigma_1\boldsymbol{p}_1^T$, it can not be obtained via linear combination of the first block of the* 2*th row and the first block of the* 3*th row, because $\boldsymbol{p}_1$ is linearly independent with $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$. This indicates that the rank of $\mathbf{\Theta}_2^\dagger$, which is the rank of $\tilde{\mathbf{\Theta}}_2$, is equal to* 9*, the number of columns in $\tilde{\mathbf{\Theta}}_2$.*

To summarize, under the case of multiple multi-antenna users, according to Theorem 1 and Theorem 2, we can utilize

$$\bar{\tau} = \bar{\tau}_1+\bar{\tau}_2 = 2M+\left\lceil\frac{M(KU-1)}{q}\right\rceil, \tag{64}$$

time instants to perfectly estimate $\boldsymbol{Q}_{1,1,1}$ according to (29) and estimate $\boldsymbol{B}_1,\cdots,\boldsymbol{B}_K$ according to (33) and (58). Then, we can recover $\boldsymbol{J}_k$'s based on (12) for beamforming design.

Remark 3. *Compared to the method proposed in [25]–[27], our proposed scheme significantly reduces channel estimation overhead through the channel property indicated in* (12) *from two aspects. First, the channels associated with different RIS reflecting elements for the same single antenna are highly correlated. This correlation allows us to reduce the required number of time instants in Phase I from from M^2 to $2M$ in a single-user and single-antenna case. Second, the channels associated with different antennas and users also exhibit significant correlation. As a result, for a case of multiple multi-antenna users with a total of KU antennas, the scheme in [25]–[27] requires KUM^2 time instants, while our proposed scheme requires only $2M+\lceil\frac{M(KU-1)}{q}\rceil$ time instants.*

VI. Channel Estimation for Case with Noise at the BS

In the previous section, we have shown how to perfectly estimate all the cascaded channels using $\bar{\tau}=2M+\lceil\frac{M(KU-1)}{q}\rceil$ time instants for the ideal case without noise at the BS. In this section, we consider Q2 and introduce how to estimate the cascaded channels under our proposed two-phase scheme for the practical case with noise at the BS, using $\tau\geq\bar{\tau}$ time instants.

A. Phase I: Estimation of the Cascaded Channel Associated With the 1st Antenna of User 1

In Phase I with noise at the BS, with only the 1st antenna of user 1 transmitting non-zero pilot symbols as (39), the received signal of the BS given in (40) is re-expressed as

$$\boldsymbol{y}_t=\sqrt{p}a_t\left(\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,1}+\sum\nolimits_{m=2}^{M}\beta_{1,1,m}\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,m}\right)+\boldsymbol{n}_t,\quad t=1,\cdots,\tau_1. \tag{65}$$

As in Section IV, we divide the overall τ_1 time instants into two parts with the same length δ, i.e., $\tau_1=2\delta$. The BD-RIS scattering matrices and the user pilot signals over various instants of these two parts follow (20) to (23) in Section IV. In this case, the received signal at time instant $t=1,\cdots,\tau_1$ in Phase I given in (24) and (25) is re-expressed as

$$\boldsymbol{y}_t=\sqrt{p}a_t\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,1}+\sqrt{p}a_t\sum\nolimits_{m=2}^{M}\beta_{1,1,m}\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,m}+\boldsymbol{n}_t, \tag{66}$$

and

$$\boldsymbol{y}_{\delta+t}=\sqrt{p}a_te^{\mathrm{j}\theta}\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,1}+\sqrt{p}a_t\sum\nolimits_{m=2}^{M}\beta_{1,1,m}\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,m}+\boldsymbol{n}_{\delta+t},\quad t=1,\cdots,\delta, \tag{67}$$

respectively, where $\delta>M$. Then by subtracting $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_\delta$ from $\boldsymbol{y}_{\delta+1},\cdots,\boldsymbol{y}_{2\delta}$, respectively, we can obtain δ noisy effective received signals as follows:

$$\bar{\boldsymbol{y}}_t=\boldsymbol{y}_{\delta+t}-\boldsymbol{y}_t=\sqrt{p}a_t(e^{\mathrm{j}\theta}-1)\boldsymbol{Q}_{1,1,1}\boldsymbol{\phi}_{t,1}+\boldsymbol{z}_t,\quad t=1,\cdots,\delta, \tag{68}$$

where $\boldsymbol{z}_t=\boldsymbol{n}_{\delta+t}-\boldsymbol{n}_t$ denotes the effective noise, and $\boldsymbol{z}_t\sim\mathcal{CN}(\boldsymbol{0},\sigma_z^2\boldsymbol{I}_N)$ with $\sigma_z^2=2\sigma^2$. The overall effective received signal over τ_1 time instants in Phase I is

$$\bar{\boldsymbol{Y}}_1=[\bar{\boldsymbol{y}}_1,\cdots,\bar{\boldsymbol{y}}_\delta]=\sqrt{p}a_t(e^{\mathrm{j}\theta}-1)\boldsymbol{Q}_{1,1,1}\boldsymbol{\Psi}_1+\boldsymbol{Z}_1, \tag{69}$$

where $\boldsymbol{\Psi}_1=[a_1\boldsymbol{\phi}_{1,1},\ldots,a_\delta\boldsymbol{\phi}_{\delta,1}]$, and $\boldsymbol{Z}_1=[\boldsymbol{z}_1,\cdots,\boldsymbol{z}_\delta]$. Based on (69), we apply the linear minimum MSE (LMMSE) estimator to estimate $\boldsymbol{Q}_{1,1,1}$ as follows:

$$\hat{\boldsymbol{Q}}_{1,1,1}=(\sqrt{p}(e^{\mathrm{j}\theta}-1))^{-1}\bar{\boldsymbol{Y}}_1\left(\boldsymbol{\Psi}_1^H\boldsymbol{C}_Q\boldsymbol{\Psi}_1+\sigma_z^2\boldsymbol{I}_\delta\right)^{-1}\boldsymbol{\Psi}_1^H\boldsymbol{C}_Q, \tag{70}$$

where $\boldsymbol{C}_Q=\mathbb{E}[\boldsymbol{Q}_{1,1,1}^H\boldsymbol{Q}_{1,1,1}]$ denotes the covariance matrix of $\boldsymbol{Q}_{1,1,1}$.

Next, based on the estimation of $\boldsymbol{Q}_{1,1,1}$ given in (70), the noisy version of effective received signal given in (30) is re-expressed as

$$\begin{aligned}\tilde{\boldsymbol{y}}_t&=\boldsymbol{y}_t-\sqrt{p}a_t\hat{\boldsymbol{Q}}_{1,1,1}\boldsymbol{\phi}_{t,1}=\boldsymbol{F}_t\bar{\boldsymbol{\beta}}_{1,1}+\boldsymbol{e}_t+\boldsymbol{n}_t\\&=\hat{\boldsymbol{F}}_t\bar{\boldsymbol{\beta}}_{1,1}+(\boldsymbol{F}_t-\hat{\boldsymbol{F}}_t)\bar{\boldsymbol{\beta}}_{1,1}+\boldsymbol{e}_t+\boldsymbol{n}_t,\quad t=1,\cdots,\delta,\end{aligned} \tag{71}$$

where $\bar{\boldsymbol{\beta}}_{1,1}=[\beta_{1,1,2},\cdots,\beta_{1,1,M}]^T$, $\boldsymbol{e}_t=\sqrt{p}a_t(\boldsymbol{Q}_{1,1,1}-\hat{\boldsymbol{Q}}_{1,1,1})\boldsymbol{\phi}_{t,1}$, and $\hat{\boldsymbol{F}}_t=\sqrt{p}a_t\hat{\boldsymbol{Q}}_{1,1,1}[\boldsymbol{\phi}_{t,2},\cdots,\boldsymbol{\phi}_{t,M}]$. Then, the overall effective received signal used for estimating $\bar{\boldsymbol{\beta}}_{1,1}$ is

$$\begin{aligned}\tilde{\boldsymbol{y}}^{(1)}&=[\tilde{\boldsymbol{y}}_1^T,\cdots,\tilde{\boldsymbol{y}}_\delta^T]^T\\&=\hat{\boldsymbol{\Theta}}_1\bar{\boldsymbol{\beta}}_{1,1}+(\boldsymbol{\Theta}_1-\hat{\boldsymbol{\Theta}}_1)\bar{\boldsymbol{\beta}}_{1,1}+\tilde{\boldsymbol{e}}^{(1)}+\tilde{\boldsymbol{n}}^{(1)},\end{aligned} \tag{72}$$

where $\hat{\boldsymbol{\Theta}}_1=[\hat{\boldsymbol{F}}_1^T,\cdots,\hat{\boldsymbol{F}}_\delta^T]^T$, $\tilde{\boldsymbol{e}}^{(1)}=[\boldsymbol{e}_1^T,\cdots,\boldsymbol{e}_\delta^T]^T$, and $\tilde{\boldsymbol{n}}^{(1)}=[\boldsymbol{n}_1^T,\cdots,\boldsymbol{n}_\delta^T]^T$. In (72), the error propagated from Phase I, i.e., $\boldsymbol{\Theta}_1-\hat{\boldsymbol{\Theta}}_1$ and $\tilde{\boldsymbol{e}}^{(1)}$, makes it hard to obtain the LMMSE estimator of $\bar{\boldsymbol{\beta}}_{1,1}$. In practice, we can increase the pilot sequence length in Phase I, i.e, τ_1, such that $\boldsymbol{\Theta}_1-\hat{\boldsymbol{\Theta}}_1$ and $\tilde{\boldsymbol{e}}^{(1)}$ are sufficiently small. In this case, we assume that $\boldsymbol{\Theta}_1-\hat{\boldsymbol{\Theta}}_1\approx\boldsymbol{0}$ and $\tilde{\boldsymbol{e}}^{(1)}\approx\boldsymbol{0}$. Then, (72) reduces to

$$\tilde{\boldsymbol{y}}^{(1)}\approx\hat{\boldsymbol{\Theta}}_1\bar{\boldsymbol{\beta}}_{1,1}+\tilde{\boldsymbol{n}}^{(1)}. \tag{73}$$

Based on (73), the LMMSE estimator of $\bar{\boldsymbol{\beta}}_{1,1}$ is designed as

$$\hat{\boldsymbol{\beta}}_{1,1}=\boldsymbol{C}_{\beta_1}\hat{\boldsymbol{\Theta}}_1^H\left(\hat{\boldsymbol{\Theta}}_1\boldsymbol{C}_{\beta_1}\hat{\boldsymbol{\Theta}}_1^H+\sigma^2\boldsymbol{I}_{N\delta}\right)^{-1}\tilde{\boldsymbol{y}}^{(1)}, \tag{74}$$

where $\boldsymbol{C}_{\beta_1}$ denotes the covariance matrix of $\bar{\boldsymbol{\beta}}_{1,1}$.

B. Phase II: Estimation of the Scaling Coefficients of the Other Antennas of User 1 and the Other Users

In Phase II with noise at the BS, with the pilot transmission rule in (43) and the estimation of $\boldsymbol{Q}_{1,1,1}$ given in (70), the received signal of the BS given in (44) is re-expressed as

$$\begin{aligned}\boldsymbol{y}_t&=\boldsymbol{T}_t\bar{\boldsymbol{b}}+\boldsymbol{n}_t\\&=\hat{\boldsymbol{T}}_t\bar{\boldsymbol{b}}+(\boldsymbol{T}_t-\hat{\boldsymbol{T}}_t)\bar{\boldsymbol{b}}+\boldsymbol{n}_t,\quad t=\tau_1+1,\cdots,\tau_1+\tau_2,\end{aligned} \tag{75}$$

where $\hat{\boldsymbol{T}}_t=\sqrt{p}\bar{\boldsymbol{a}}_t^T\otimes\hat{\boldsymbol{Q}}_{1,1,1}\boldsymbol{\Phi}_t$. Then, the overall received signal for estimating $\bar{\boldsymbol{b}}$ is re-expressed as

$$\boldsymbol{y}^{(2)}=\boldsymbol{\Theta}_2\bar{\boldsymbol{b}}+\boldsymbol{n}^{(2)}=\hat{\boldsymbol{\Theta}}_2\bar{\boldsymbol{b}}+(\boldsymbol{\Theta}_2-\hat{\boldsymbol{\Theta}}_2)\bar{\boldsymbol{b}}+\boldsymbol{n}^{(2)}, \tag{76}$$

where $\hat{\mathbf{\Theta}}_2 = [\hat{\boldsymbol{T}}_{\tau_1+1}^T, \cdots, \hat{\boldsymbol{T}}_{\tau_1+\tau_2}^T]^T$, and $\boldsymbol{n}^{(2)} = [\boldsymbol{n}_{\tau_1+1}^T, \cdots, \boldsymbol{n}_{\tau_1+\tau_2}^T]^T$. Similar to Phase I, we can increase τ_1 such that the propagated error $\mathbf{\Theta}_2 - \hat{\mathbf{\Theta}}_2$ can be sufficiently small. In this case, we assume that $\mathbf{\Theta}_2 - \hat{\mathbf{\Theta}}_2 \approx \mathbf{0}$. Then, (76) reduces to

$$\boldsymbol{y}^{(2)} \approx \hat{\mathbf{\Theta}}_2 \bar{\boldsymbol{b}} + \boldsymbol{n}^{(2)}. \quad (77)$$

To estimate $\bar{\boldsymbol{b}}$ based on (77), we set the pilot signals and BD-RIS scattering matrices as follows. At time instants $t = \tau_1 + 1, \cdots, \tau_1 + \bar{\tau}_2$, we set the pilot signals as in (54) and BD-RIS scattering matrices as in (55) according to Theorem 2. On the other hand, Theorem 2 characterizes the minimal number of time instants for perfect channel estimation in noiseless case, and does not specify how to design user pilot signals and BD-RIS scattering matrices at time instants larger than this minimal number. In the noisy case, at time instants $t = \tau_1 + \bar{\tau}_2 + 1, \cdots, \tau_1 + \tau_2$, we generate $a_{k,u,t}$'s, $\forall k, u$, based on i.i.d. complex standard normal distribution, and $\mathbf{\Phi}_t$ as a random unitary matrix. Then, based on (77), the LMMSE estimator of $\bar{\boldsymbol{b}}$ can be designed as

$$\hat{\boldsymbol{b}} = \boldsymbol{C}_b \hat{\mathbf{\Theta}}_2^H \left(\hat{\mathbf{\Theta}}_2 \boldsymbol{C}_b \hat{\mathbf{\Theta}}_2^H + \sigma^2 \boldsymbol{I}_{N\tau_2} \right)^{-1} \boldsymbol{y}^{(2)}, \quad (78)$$

where $\boldsymbol{C}_b$ denotes the covariance matrix of $\bar{\boldsymbol{b}}$.

In summary, with the estimation of $\hat{\boldsymbol{Q}}_{1,1,1}$ given in (70) and the estimations of $\beta_{k,u,m}$'s given in (74) and (78), the cascaded channels associated with the m-th BD-RIS reflecting element and the u-th antenna of user k can be estimated as

$$\hat{\boldsymbol{Q}}_{k,u,m} = \hat{\beta}_{k,u,m} \hat{\boldsymbol{Q}}_{1,1,1}, \quad \forall (k,u,m) \neq (1,1,1), \quad (79)$$

based on (12).

At last, we characterize the computational complexity of the LMMSE estimators applied in our proposed scheme. We count the number of complex multiplication (CM) [30] as the measure of complexity. In (70), the estimation of $\boldsymbol{Q}_{1,1,1}$ requires $\Delta_1 = \delta M^2 + \delta^2 M + \delta^3 + N\delta^2 + NM\delta + NM$ CMs; In (74), the estimation of $\boldsymbol{\beta}_{1,1}$ requires $\Delta_2 = N\delta(M-1)^2 + 2(M-1)^3 + N\delta(M-1) + (M-1)^2$ CMs; And in (78), the estimation of $\bar{\boldsymbol{b}}$ requires $\Delta_3 = N\tau_2(M(KU-1))^2 + 2(M(KU-1))^3 + N\tau_2 M(KU-1) + (M(KU-1))^2$ CMs. Note that Δ_1, Δ_2 and Δ_3 asymptotically costs $\mathcal{O}(\delta^3 + N\delta^2)$, $\mathcal{O}(N\delta M^2)$, and $\mathcal{O}(N\tau_2 M^2 K^2 U^2)$, respectively, which all depend dominantly on the number of BD-RIS elements M.

VII. Numerical Results

In this section, we provide numerical results to demonstrate the advantages of our proposed scheme. The channel between the BS and the BD-RIS and that between the u-th antenna of user k and the m-th BD-RIS reflecting element are modeled as $\boldsymbol{G} \sim \mathcal{CN}(\mathbf{0}, \ell^{\text{RB}} M \boldsymbol{I})$ and $r_{k,u,m} \sim \mathcal{CN}(0, \ell^{\text{UR}})$, respectively, where ℓ^{RB} and ℓ^{UR} denote the pass loss and follow the same model as [1], unless otherwise emphasized. In this case, the rank of $\boldsymbol{G}$ is $q = \min\{M, N\}$. The transmit power of users is $p = 33$ dBm. The power spectrum density of the noise at the BS is assumed to be -169 dBm/Hz, and the channel bandwidth is 1 MHz. The normalized mean-squared error (NMSE) is used as the metric to evaluate the performance of channel estimation. Specifically, the overall NMSE for estimating all the users' cascaded channels is defined as

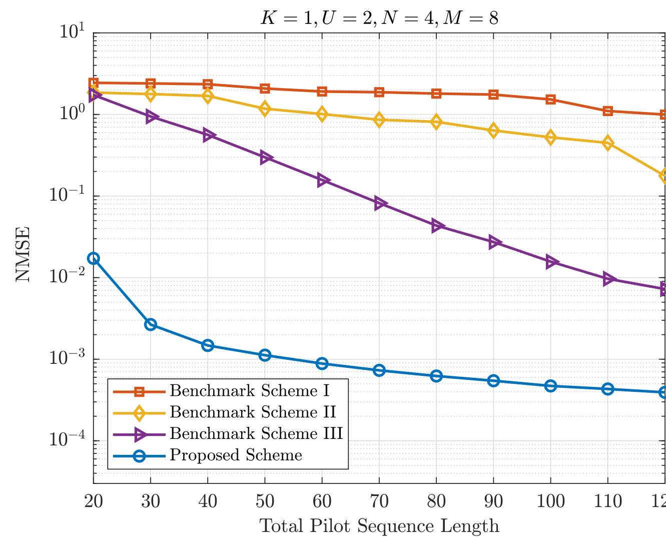


Fig. 3. NMSE performance versus total pilot length when $M = 8, N = 4, U = 2$.

$$\text{NMSE} = \mathbb{E}\left[\frac{1}{K} \sum_{k=1}^{K} \frac{||\hat{\boldsymbol{J}}_k - \boldsymbol{J}_k||_F^2}{||\boldsymbol{J}_k||_F^2} \right], \quad (80)$$

where the overall estimated channel of user k, i.e., $\hat{\boldsymbol{J}}_k$, is given as the same form as $\boldsymbol{J}_k$ in (9), with $\boldsymbol{Q}_{k,u,m}$ replaced by $\hat{\boldsymbol{Q}}_{k,u,m}$, $\forall k, u, m$.

A. Single-User Case

In the literature, the channel estimation problem under BD-RIS assisted communication has also been studied in [25]–[27]. Different from our work which utilizes the hidden channel property given in (12), the above works directly estimate all entries in $\boldsymbol{J}_k$'s based on (11). To show the performance gain of our proposed scheme, we adopt the following benchmark schemes.

- **Benchmark Scheme I**: The LS estimator proposed in [25], [26], which directly applies the LS technique to estimate $\boldsymbol{J}_k$ based on (11).
- **Benchmark Scheme II**: The BTKF algorithm [27], which obtains a LS channel estimate via the 3-mode unfolding of the received pilot tensor and then yields decoupled estimates using rank-one (Kronecker) approximation.
- **Benchmark Scheme III**: The Block Tucker alternating least squares (BTALS) algorithm [27], which iteratively refines decoupled channel estimates via alternating LS on the 1-mode and 2-mode tensor unfoldings of the received pilot tensor.

Figs. 3-6 show the NMSE performance comparison between our proposed scheme and the three benchmark schemes under the single-user case. In Fig. 3, we set the numbers of BD-RIS elements, antennas at the BS and antennas at the user as $M = 8$, $N = 4$ and $U = 2$, respectively, and the total pilot

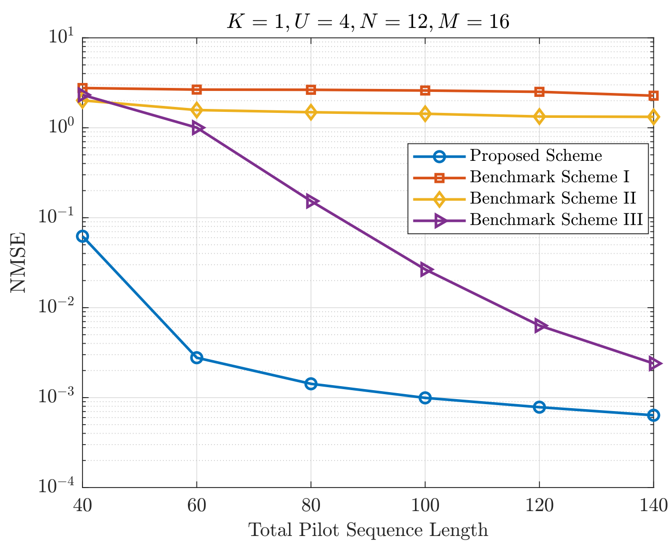


Fig. 4. NMSE performance versus number of BD-RIS elements when $M = 16, N = 12, U = 4$.

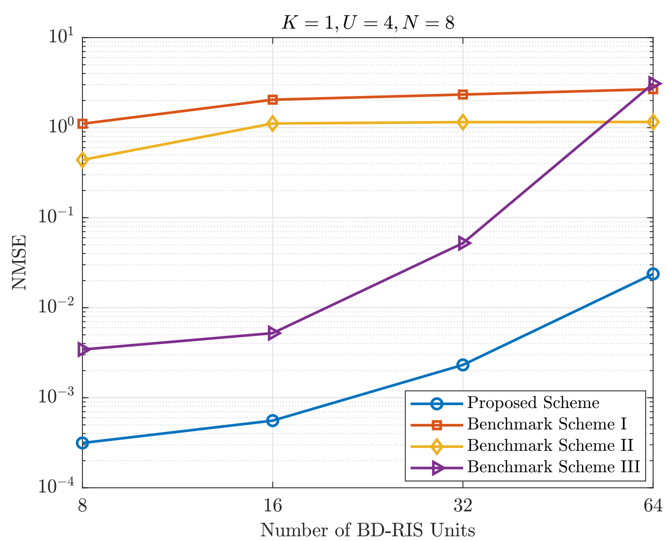


Fig. 5. NMSE performance versus number of BD-RIS elements when $N = 8, U = 4, \tau = 160$.

sequence length ranges from 20 to 120. It is observed that our proposed scheme shows a significant NMSE performance gain compared to the three benchmark schemes. For Benchmark Scheme I and Benchmark Scheme II, this is because the pilot sequence length required by these two schemes is 128 while that required by our proposed scheme is $\bar{\tau} = 18$ due to the utilization of (12). Note that Benchmark Scheme III achieves a lower NMSE than Benchmark Scheme I and II, due to that it already exploits the tensor decomposition structure of the cascaded channel to reduce estimation overhead. However, Benchmark Scheme III still performs worse than our proposed scheme because it treats all the channel coefficients as independent variables.

Fig. 4 shows another numerical example with larger numbers of BD-RIS elements, antennas at the BS and at the user with $M = 16, N = 12$ and $U = 4$. The total pilot sequence length ranges from 40 to 140. Observations similar to Fig. 3 can be obtained. Under this setup, the required pilot sequence length under our proposed scheme is 36 while that under Benchmark Scheme I and Benchmark Scheme II is 1024. Thus, these two schemes can hardly work. Moreover, note that in both Fig. 3 and 4, the performance gain of our proposed scheme compared to Benchmark Scheme III becomes smaller as the pilot length increases. This is because Benchmark Scheme III exploits an iterative alternating LS strategy that gradually refines the channel estimates and averages noise, especially with an abundance of pilot symbols [38], [39]. However, this improved performance comes at the expense of higher pilot transmission overhead due to its not exploiting the correlation in (12), which aggravates its disadvantage of high computational complexity caused by iterative nature.

Fig. 5 shows the NMSE performance versus the number of the BD-RIS elements, with $N = 8$ and $U = 4$. The total pilot sequence length is fixed as 160. It is observed that Benchmark Scheme I and Benchmark Scheme II show a deteriorated NMSE performance, which is caused by the pilot sequence length being insufficient, i.e., less than the required value. Moreover, when the number of BD-RIS elements increases, the NMSE increases more significantly under Benchmark Scheme III than our proposed scheme. This is attributed to that exploiting (12) is more efficient than utilizing tensor decomposition in channel estimation overhead reduction. Specifically, for the single-user and single-antenna case, every additional BD-RIS element increases only 1 unknown variable to estimate under our proposed scheme, while it causes $N+1$ unknown variables to estimate under Benchmark Scheme III.

In the previous numerical results, we assume that the user-RIS channels and RIS-BS channels are all Gaussian distributed. However, we want to emphasize that the results in this paper hold for an arbitrary distribution of the channels. To illustrate this, we show another numerical result in Fig. 6. In this numerical example, the RIS-BS channel follows Rician fading model

$$\boldsymbol{G} = \sqrt{\frac{\kappa}{1+\kappa}}\boldsymbol{G}^{\text{LoS}} + \sqrt{\frac{1}{1+\kappa}}\boldsymbol{G}^{\text{NLoS}}, \tag{81}$$

where κ denotes the Rician factor set as 10dB, $\boldsymbol{G}^{\text{LoS}}$ denotes the LoS component in $\boldsymbol{G}$, and $\boldsymbol{G}^{\text{NLoS}} \sim \mathcal{CN}(\boldsymbol{0}, \ell^{\text{RB}} M \boldsymbol{I})$ denotes the i.i.d. Rayleigh fading component. We set the numbers of BD-RIS elements, antennas at the BS and antennas at the user as $M = 8$, $N = 4$ and $U = 2$, respectively, and the total pilot sequence length ranges from 20 to 120. It can be observed that our proposed scheme still performs much better than the three benchmarks in estimation NMSE.

B. Multi-User Case

In the multi-user and multi-antenna case, we first explore the effect of different allocations between the pilot sequence length of two phases under our proposed scheme. Define $\tau_{\text{res}} = \tau - \bar{\tau}$ as the length of pilot sequence exceeding $\bar{\tau}$ in (64), and ϱ as the proportion of the additional pilot sequence length τ_{res} allocated to Phase I. In Fig. 7, we set the numbers of BD-RIS elements, antennas at the BS, antennas at the user, and

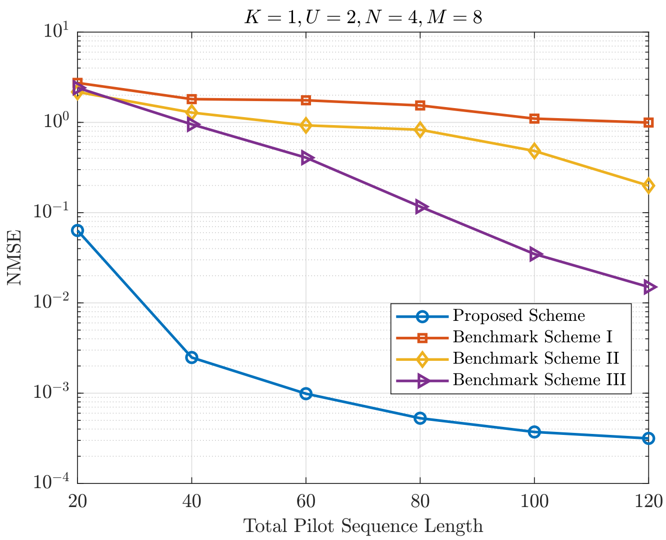


Fig. 6. NMSE performance under Rician fading model of RIS-BS channel $\boldsymbol{G}$.

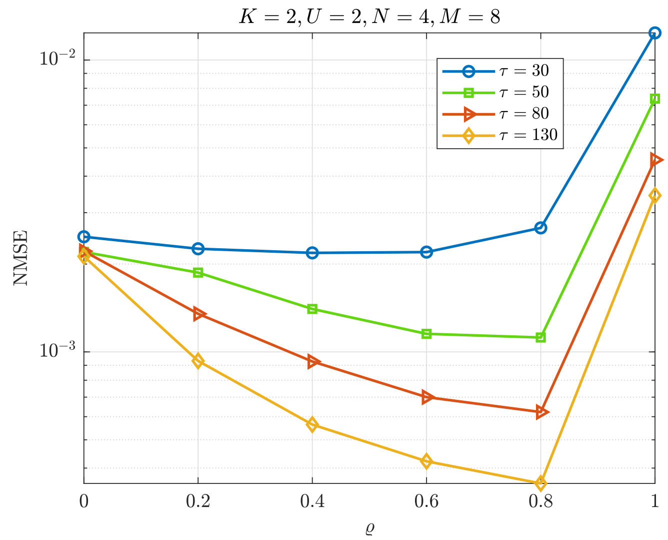


Fig. 7. NMSE performance under different pilot sequence allocations.

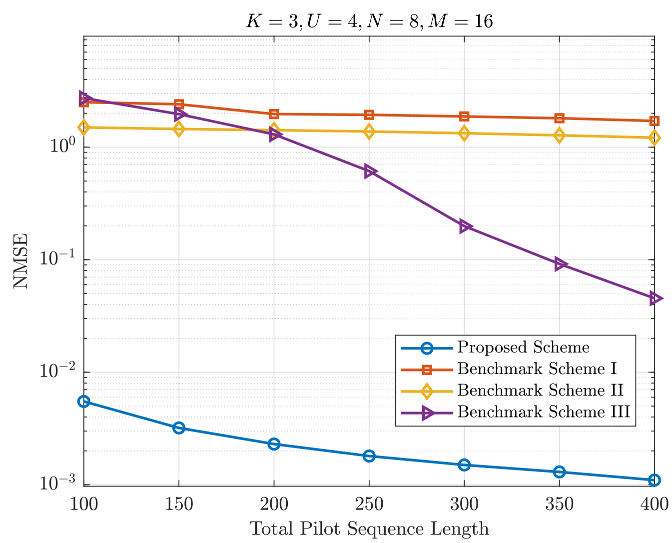


Fig. 8. NMSE performance versus total pilot length when $M = 16, N = 8, U = 4, K = 3$.

the number of users as $M = 8$, $N = 4$, $U = 2$, and $K = 2$, respectively. The total pilot sequence length τ is set as 30, 50, 80 and 130, respectively. We show the NMSE performance of our proposed scheme under different values of ϱ. It is observed that for each set of total pilot sequence length τ, the NMSE shows a "first-drop-then-rise"trend. Note that as ϱ increases, $\boldsymbol{Q}_{1,1,1}$ can be estimated based on more received signals in Phase I, leading to less estimation error propagated to Phase II, which, however, reduces the number of pilot symbols available for Phase II to estimate $\bar{\boldsymbol{B}}$. When the pilot sequence length in Phase I is small, the estimation error propagated to Phase I is the bottleneck to limit the overall NMSE performance, and it is beneficial to increase ϱ. However, when the pilot sequence length in Phase I is sufficiently large, the BS has enough pilot signals to estimate $\boldsymbol{Q}_{1,1,1}$, and it is not a good idea to keep increasing ϱ because this will reduce the pilot signals to estimate $\bar{\boldsymbol{B}}$ in Phase II. This indicates that the pilot sequence length allocation should be carefully designed. In the rest of this section, we always set ϱ such that an optimal NMSE performance can be achieved under our proposed scheme.

In the literature, no work has considered the multi-user and multi-antenna case yet. To show the performance superiority of our proposed scheme, in this subsection, we extend the benchmark schemes in Section VII-A to the multi-user case, where users transmit pilot signals at orthogonal time instants. The benchmark schemes are as follows.

- **Benchmark Scheme I**: The LS estimator is applied to estimate all users' cascaded channels $\boldsymbol{J}_k$'s.
- **Benchmark Scheme II**: The BTKF algorithm is applied to estimate all users' cascaded channels $\boldsymbol{J}_k$'s.
- **Benchmark Scheme III**: The BTALS algorithm is applied to estimate all users' cascaded channels $\boldsymbol{J}_k$'s.

In Fig. 8, the numbers of BD-RIS elements, antennas at the BS and users, and users are set as $M = 16, N = 8, U = 4, K = 3$, respectively. The total pilot sequence length ranges from 100 to 400. It is observed that our proposed scheme shows a significant NMSE performance gain compared to the three benchmark schemes, due to leveraging the channel properties shown in (12). Note that although Benchmark Scheme III achieves a lower NMSE than Benchmark Scheme I and Benchmark Scheme II due to exploiting the built-in tensor decomposition structure of the cascaded channel, it regards the channels of different users as independent and performs much worse than our proposed scheme in the multi-user case.

Fig. 9 shows the NMSE performance versus the number of users, where $M = 16, N = 8, U = 2$, the total pilot length is fixed as $\tau = 64$. It is observed that as the number of users increases, the NMSE performance of all three benchmark schemes degrades markedly. In contrast, our proposed scheme maintains robust NMSE performance even with such a limited pilot sequence length, indicating its superiority in the multi-user case.

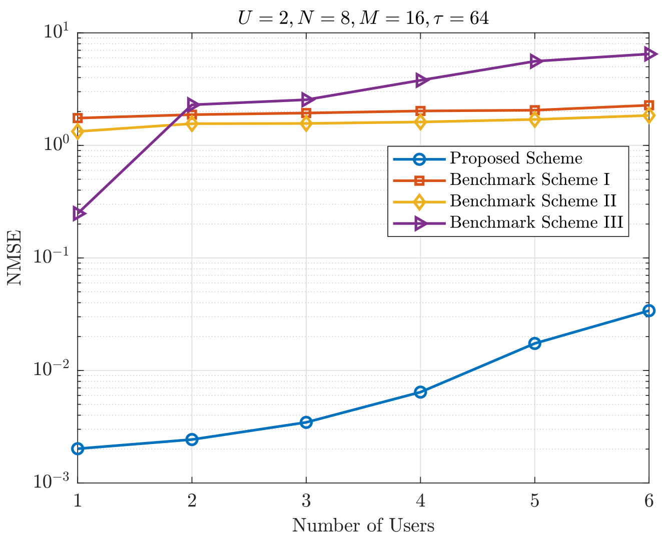


Fig. 9. NMSE performance versus number of users when $M = 16, N = 8, U = 2$.

VIII. Conclusions

In this paper, we revealed an important channel property for BD-RIS aided multi-user MIMO communication. Specifically, arising from the common RIS-BS channel, the cascaded channel associated with one pair of BD-RIS element and BS antenna is a scaled version of the associated with any other pair of BD-RIS element and BS antenna. By exploiting this correlation, we addressed two key challenges: quantifying the pilot transmission overhead for perfect channel estimation in noise-free case and proposing practical estimation design in the case with noise. Interestingly, it is theoretically shown that the overhead for channel estimation in BD-RIS aided communication is in the same order as that in conventional RIS aided communication, although the number of channel coefficients in these two systems are different by orders of magnitude. Numerical results verified that our approach can estimate channels accurately with significantly reduced overhead compared to existing algorithms. It is interesting to characterize the channel estimation overhead in other RIS related systems as future works, such as the group connected BD-RIS aided narrowband system and the BD-RIS aided orthogonal frequency division multiplexing (OFDM) system.

Appendix

A. Proof of Proposition 1

To prove $\text{rank}(\mathbf{\Theta}_1) = M - 1$, it is equivalent to showing the equation $\mathbf{\Theta}_1 \bar{\boldsymbol{\beta}} = \mathbf{0}$ holds for a vector $\bar{\boldsymbol{\beta}} \in \mathbb{C}^{(M-1)\times 1}$, if and only if $\bar{\boldsymbol{\beta}} = \mathbf{0}$. The equation is equivalent to

$$\boldsymbol{Q}_1 [\boldsymbol{\phi}_{t,2}, \dots, \boldsymbol{\phi}_{t,M}] \bar{\boldsymbol{\beta}} = \mathbf{0}, \quad t = 1, \dots, M. \tag{82}$$

Let $\mathcal{S} = \ker(\boldsymbol{Q}_1)$ be the null space of $\boldsymbol{Q}_1$. The dimension of this subspace is $\dim(\mathcal{S}) = M - q$. Define

$$\boldsymbol{w}_t = [\boldsymbol{\phi}_{t,2}, \dots, \boldsymbol{\phi}_{t,M}] \bar{\boldsymbol{\beta}} \in \mathbb{C}^{M\times 1}. \tag{83}$$

Then $\boldsymbol{w}_t$ must lie in the null space $\mathcal{S}$, i.e.,

$$\boldsymbol{w}_t \in \mathcal{S}, \quad t = 1, \dots, M. \tag{84}$$

Define

$$\boldsymbol{b}^{(1)} = [b_1^{(1)}, \cdots, b_M^{(1)}]^T = [0, \beta_2, \dots, \beta_M]^T, \tag{85}$$

then given $\mathbf{\Phi}_t$'s in (21), $\boldsymbol{w}_t$ can be re-expressed as

$$\boldsymbol{w}_t = \boldsymbol{P}\boldsymbol{b}^{(t)}, \tag{86}$$

where $\boldsymbol{b}^{(t)} = [b_1^{(t)}, \cdots, b_M^{(t)}]^T$ with $b_m^{(t)} = b^{(1)}_{(m-t) \bmod M+1}$, $\forall m, t = 2, \cdots, M$.

This proof proceeds by contradiction. We assume there exists a non-zero vector $\bar{\boldsymbol{\beta}} = [\beta_2, \dots, \beta_M]^T \in \mathbb{C}^{(M-1)\times 1}$ such that equations (82) holds. Let $\mathcal{W} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_M\}$ and $\mathcal{B} = \{\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(M)}\}$. Since $\boldsymbol{P}$ is unitary, the dimensions of the subspaces spanned by these two sets are identical:

$$\dim(\text{span}(\mathcal{W})) = \dim(\text{span}(\mathcal{B})). \tag{87}$$

Next, we determine the dimension of $\mathcal{B}$. Let $\boldsymbol{B} = [\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \dots, \boldsymbol{b}^{(M)}]$. The dimension of $\text{span}(\mathcal{B})$ is equal to the rank of $\boldsymbol{B}$. Note that $\boldsymbol{B}$ is actually a circulant matrix. Based on Proposition 1.3 in [40], it can be shown that the rank of $\boldsymbol{B}$ will be M (full rank) with probability 1 given $\bar{\boldsymbol{\beta}}$ in this paper. This implies that the vectors $\boldsymbol{b}^{(1)}, \cdots, \boldsymbol{b}^{(M)}$ are linearly independent. Therefore, for any $\bar{\boldsymbol{\beta}} \neq \mathbf{0}$:

$$\dim(\text{span}(\mathcal{B})) = \text{rank}(\boldsymbol{B}) = M. \tag{88}$$

Based on (87), it follows that $\dim(\text{span}(\mathcal{W})) = M$.

Recall that $\mathbf{\Theta}_1 \bar{\boldsymbol{\beta}} = \mathbf{0}$ for $\bar{\boldsymbol{\beta}} \neq \mathbf{0}$, implies that $\boldsymbol{w}_t \in \mathcal{S}$ for all $t = 1, \dots, M$. This requires the subspace spanned by $\boldsymbol{w}_t$'s to be contained within the null space $\mathcal{S}$, i.e., $\text{span}(\mathcal{W}) \subseteq \mathcal{S}$, which implies the relationship between the dimensions of these two subspaces:

$$\dim(\text{span}(\mathcal{W})) \leq \dim(\mathcal{S}) \Leftrightarrow M \leq M - q \Leftrightarrow q \leq 0. \tag{89}$$

This result is a contradiction, because $q = \text{rank}(\boldsymbol{G})$ must be a positive integer ($q \geq 1$). The only way to resolve this contradiction is to reject the initial assumption that a non-zero $\bar{\boldsymbol{\beta}}$ could satisfy $\mathbf{\Theta}_1 \bar{\boldsymbol{\beta}} = \mathbf{0}$. Therefore, the only solution on (82) and (31) is $\bar{\boldsymbol{\beta}} = \mathbf{0}$, i.e., $\text{nullily}(\mathbf{\Theta}_1) = 0$. Based on Rank-Nullity Theorem, it can be obtained that $\text{rank}(\mathbf{\Theta}_1) = M - 1$.

B. Proof of Theorem 2

Given user pilot signals and BD-RIS scattering matrices as in (54) and (55), respectively, the matrix $\tilde{\mathbf{\Theta}}_2$ is a block-diagonal matrix consisting of $\eta + 1$ diagonal blocks, $\tilde{\mathbf{\Theta}}_{2,s}$, $s = 1, \cdots, \eta + 1$. We show that each $\tilde{\mathbf{\Theta}}_{2,s}$ is full rank.

First let us consider the first η blocks $\tilde{\mathbf{\Theta}}_{2,s}$'s, $s = 1, \cdots, \eta$. Each $\tilde{\mathbf{\Theta}}_{2,s}$, $\forall s$, consists of $\tau_o \times U_o$ sub-blocks, and the sub-block in the i-th block row and j-th block column is denoted by

$$\begin{aligned} &\mathbf{\Xi}_{s,i,j} = \bar{a}_{(s-1)U_o+j, \tau_{o,s}+i} \mathbf{\Sigma} \boldsymbol{V}^H \mathbf{\Phi}_{\tau_{o,s}+i} \in \mathbb{C}^{N\times M}, \\ &s = 1, \cdots, \eta, \ i = 1, \cdots, \tau_o, \ j = 1, \cdots, U_o. \end{aligned} \tag{90}$$

Based on (54) and (55), many of $\boldsymbol{\Xi}_{s,i,j}$'s are zero blocks, and the non-zero sub-blocks are expressed as

$$\begin{aligned}\boldsymbol{\Xi}_{s,i,j} &= c_{\tau_{o,s}+i}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{\Phi}_{\tau_{o,s+i}} \\ &= c_{\tau_{o,s}+i}\boldsymbol{\Sigma}\boldsymbol{P}(S_{M,\tau_{o,s}+i},:), \quad \tau_{o,s}+i \in \Upsilon_{s,j}. \end{aligned} \tag{91}$$

Define a matrix $\tilde{\boldsymbol{\Xi}}_{s,i,j} \in \mathbb{C}^{q\times M}$ with its q'-th row given as

$$\tilde{\boldsymbol{\Xi}}_{s,i,j}(q',:) = \frac{1}{c_{\tau_{o,s}+i}\sigma_{q'}}\boldsymbol{\Xi}_{s,i,j}(q',:), \quad q'=1,\cdots,q, \tag{92}$$

then we obtain

$$\tilde{\boldsymbol{\Xi}}_{s,i,j} = \boldsymbol{P}(S_{M,\tau_{o,s}+i}(\mathcal{Q}),:), \tag{93}$$

where $\mathcal{Q} = \{1,\cdots,q\}$, and given any sequence $\mathcal{B}$, $|\mathcal{B}|$ denotes its length and $\mathcal{B}(I)$ denotes its sub-sequence with $I \subset \{1,\cdots,|\mathcal{B}|\}$. Thus, we can built a matrix $\bar{\boldsymbol{\Theta}}_{2,s} \in \mathbb{C}^{\tau_o q\times MU_o}$ by setting $\tilde{\boldsymbol{\Xi}}_{s,i,j}$ as the block in its i-th block row and j-th block column. Then, $\bar{\boldsymbol{\Theta}}_{2,s}$ has the same rank as $\tilde{\boldsymbol{\Theta}}_{2,s}$ because it is built by eliminating the zero-rows in $\tilde{\boldsymbol{\Theta}}_{2,s}$ and scaling the non-zero rows in $\tilde{\boldsymbol{\Theta}}_{2,s}$. We next show that $\bar{\boldsymbol{\Theta}}_{2,s}$ has a full rank, $\forall s=1,\cdots,\eta$.

Define a sequence obtained by repeating $\{1,\cdots,M\}$ for U_o times, i.e.,

$$\mathcal{J} = \{ \underbrace{1,\cdots,M}_{\text{the 1st } M \text{ elements}}, ..., \underbrace{1,\cdots,M}_{\text{the } U_o\text{-th } M \text{ elements}} \}, \tag{94}$$

and divide it into τ_o parts, i.e.,

$$\mathcal{J} = \{\mathcal{J}_1,\cdots,\mathcal{J}_{\tau_o}\}, \tag{95}$$

where $\mathcal{J}_i$ contains q elements, $i=1,\cdots,\tau_o$. Then, based on (56), we have

$$S_{M,\tau_{o,s}+i}(\mathcal{Q}) = \mathcal{J}_i, \quad \tau_{o,s}+i \in \Upsilon_{s,j}. \tag{96}$$

As a result, $\bar{\boldsymbol{\Theta}}_{2,s}$ can be re-expressed as

$$\bar{\boldsymbol{\Theta}}_{2,s} = \begin{bmatrix} \boldsymbol{P}(\mathcal{J}_1) & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \cdots & \cdots & \cdots & \boldsymbol{O} \\ \boldsymbol{P}(\mathcal{J}_\epsilon) & \boldsymbol{P}(\mathcal{J}_\epsilon) & \cdots & \boldsymbol{O} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{O} & \boldsymbol{P}(\mathcal{J}_{2\epsilon+\lceil\frac{2\rho}{q}\rceil-2}) & \cdots & \boldsymbol{O} \\ & & \ddots & \\ \cdots & \cdots & & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{P}(\mathcal{J}_{(U_o-1)(\epsilon-1)+\lceil\frac{(U_o-1)\rho}{q}\rceil}) \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{P}(\mathcal{J}_{U_o\epsilon+\lceil\frac{U_o\rho}{q}\rceil-U_o}) \end{bmatrix}. \tag{97}$$

In this case, $\bar{\boldsymbol{\Theta}}_{2,s}$ is built by $\boldsymbol{p}_1^T,\cdots,\boldsymbol{p}_M^T$. Moreover, since $|\Upsilon_{s,j}| = \epsilon$ or $\epsilon+1$, $\forall s,j$, it can be obtained that

$$|\Upsilon_{s,j}|q \geq \epsilon q \geq M, \tag{98}$$

and

$$|\Upsilon_{s,j}|q \leq (\epsilon+1)q \leq 3M, \tag{99}$$

because $\epsilon q - 2M = (2-\epsilon)q - 2\rho \leq 0 \Rightarrow (\epsilon+1)q - 3M \leq 0$. Thus, for each $m \in \{1,\cdots,M\}$, the number of $\boldsymbol{p}_m^T$ being used to build the j-th block column in $\bar{\boldsymbol{\Theta}}_{2,s}$, denoted by $\mu_{m,j}$, fulfills

$$1 \leq \mu_{m,j} \leq 3. \tag{100}$$

Moreover, for each pair of j and $j+1$, it should be noted that

$$\Upsilon_{s,j} \cap \Upsilon_{s,j+1} = \{\Upsilon_{s,j}(|\Upsilon_{s,j}|)\} = \{\Upsilon_{s,j+1}(1)\}, \tag{101}$$

which indicates the only one common block row index such that $\tilde{\boldsymbol{\Xi}}_{s,i,j}$ and $\tilde{\boldsymbol{\Xi}}_{s,i,j+1}$, are both non-zero, $j=1,\cdots,U_o-1$. For ease of description, by changing the order of rows such that the rows containing each $\boldsymbol{p}_m^T$ are stacked together, we obtain a matrix $\boldsymbol{\Theta}_{2,s}^\dagger$ as follows

$$\boldsymbol{\Theta}_{2,s}^\dagger = \begin{bmatrix} \boldsymbol{\Theta}_{2,s,1}^\dagger \\ \cdots \\ \boldsymbol{\Theta}_{2,s,M}^\dagger, \end{bmatrix} \tag{102}$$

where $\boldsymbol{\Theta}_{2,s,m}^\dagger \in \mathbb{C}^{U_o\times MU_o}$ is built by $\boldsymbol{p}_m^T$ and $\boldsymbol{0}^T$, $\forall m$. $\boldsymbol{\Theta}_{2,s}^\dagger$ has the same rank as $\bar{\boldsymbol{\Theta}}_{2,s}$ because it is built by changing the order of rows in $\bar{\boldsymbol{\Theta}}_{2,s}$. Based on the above two properties in (100) and (101), $\boldsymbol{\Theta}_{2,s,m}^\dagger$ has the following structure:

- $\boldsymbol{\Theta}_{2,s,m}^\dagger$ consists of $U_o\times U_o$ blocks, where each block is either $\boldsymbol{p}_m^T$ or $\boldsymbol{0}^T$. Denote $\boldsymbol{\Theta}_{2,s,m}^\dagger[j_1,j_2]$ as the block at the j_1-th block row and the j_2-th block column, $j_1,j_2 = 1,\cdots,U_o$;
- According to (100) and (101), we have

$$\begin{aligned} &\boldsymbol{\Theta}_{2,s,m}^\dagger[j_1,j_2] = \boldsymbol{p}_m^T, \quad \forall j_1 = j_2; \\ &\boldsymbol{\Theta}_{2,s,m}^\dagger[j_1,j_2] = \boldsymbol{0}^T, \; j_1 > j_2+1 \text{ and } j_1 < j_2-1; \\ &\boldsymbol{\Theta}_{2,s,m}^\dagger[j_2+1,j_2] = \boldsymbol{p}_m^T, \text{ or } \boldsymbol{\Theta}_{2,s,m}^\dagger[j_2-1,j_2] = \boldsymbol{p}_m^T, \; \forall j_2; \end{aligned} \tag{103}$$

  and $\boldsymbol{\Theta}_{2,s,m}^\dagger[j_2-1,j_2] = \boldsymbol{p}_m^T$ and $\boldsymbol{\Theta}_{2,s,m}^\dagger[j_2,j_2-1] = \boldsymbol{p}_m^T$, $\forall j_2 \geq 2$, never holds at the same time.

Such a matrix can always be equivalent to a diagonal block matrix via elementary row operations, i.e.,

$$\boldsymbol{\Theta}_{2,s,m}^\dagger \Leftrightarrow \begin{bmatrix} \boldsymbol{p}_m^T & \cdots & \boldsymbol{0}^T \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}^T & \cdots & \boldsymbol{p}_m^T \end{bmatrix}, \quad \forall m. \tag{104}$$

As a result, $\boldsymbol{\Theta}_{2,s}^\dagger$ is equivalent to a matrix $\boldsymbol{\Theta}_{2,s}^*$ as follows

$$\boldsymbol{\Theta}_{2,s}^* = \begin{bmatrix} \boldsymbol{p}_1 \cdots \boldsymbol{0} & & \boldsymbol{p}_M \cdots \boldsymbol{0} \\ \vdots \;\ddots\; \vdots & \cdots & \vdots \;\ddots\; \vdots \\ \boldsymbol{0} \cdots \boldsymbol{p}_1 & & \boldsymbol{0} \cdots \boldsymbol{p}_M \end{bmatrix}^T, \tag{105}$$

which has the same rank as $\boldsymbol{\Theta}_{2,s}^\dagger$ because it is built via row operation of $\boldsymbol{\Theta}_{2,s}^\dagger$. It can be shown that each of the rows in $\boldsymbol{\Theta}_{2,s}^*$ cannot be obtained via linear combination of the other rows, because $\boldsymbol{p}_1,\cdots,\boldsymbol{p}_M$ are linearly independent with each other. This indicates that $\text{rank}(\boldsymbol{\Theta}_{2,s}^*) = \text{rank}(\boldsymbol{\Theta}_{2,s}^\dagger) = \text{rank}(\bar{\boldsymbol{\Theta}}_{2,s}) = MU_o$. Therefore, each of the first η diagonal blocks in $\tilde{\boldsymbol{\Theta}}_2$ are all full-rank, i.e., $\text{rank}(\tilde{\boldsymbol{\Theta}}_{2,s}) = \text{rank}(\bar{\boldsymbol{\Theta}}_{2,s}) = MU_o$, $u=1,\cdots,\eta$.

Next, we consider the last diagonal block $\tilde{\boldsymbol{\Theta}}_{2,\eta+1}$ in $\tilde{\boldsymbol{\Theta}}_2$, which consists of $\tau_\nu \times U_\nu$ blocks. Similar to $\bar{\boldsymbol{\Theta}}_{2,u}$'s, we can built a matrix $\bar{\boldsymbol{\Theta}}_{2,\eta+1} \in \mathbb{C}^{\tau_\nu q\times MU_\nu}$ by setting $\tilde{\boldsymbol{\Xi}}_{\eta+1,i,j}$ as the block in its i-th block row and j-th block column, where

$$\tilde{\boldsymbol{\Xi}}_{\eta+1,i,j} = \boldsymbol{P}(S_{M,\tau_{o,\eta+1}+i}(\mathcal{Q})), \; i=1,\cdots,\tau_\nu, \; j=1,\cdots,U_\nu. \tag{106}$$

Then, $\bar{\boldsymbol{\Theta}}_{2,\eta+1}$ has the same rank as $\tilde{\boldsymbol{\Theta}}_{2,\eta+1}$. Then, by changing the order of rows in $\bar{\boldsymbol{\Theta}}_{2,\eta+1}$ such that the rows containing each $\boldsymbol{p}_m^T$ are stacked together, we obtain a matrix

$$\boldsymbol{\Theta}_{2,\eta+1}^{\dagger} = \begin{bmatrix} \boldsymbol{\Theta}_{2,\eta+1,1}^{\dagger} \\ \cdots \\ \boldsymbol{\Theta}_{2,\eta+1,M}^{\dagger}, \end{bmatrix}, \tag{107}$$

having the same rank as $\bar{\boldsymbol{\Theta}}_{2,\eta+1}$, where $\boldsymbol{\Theta}_{2,\eta+1,m}^{\dagger} \in \mathbb{C}^{\lceil \frac{\tau_{\nu} q}{M} \rceil \times MU_v}$, $\forall m$, and has the same structure as $\boldsymbol{\Theta}_{2,s,m}^{\dagger}$'s, $s = 1, \cdots, \eta$. Thus, via elementary row operations, each $\boldsymbol{\Theta}_{2,\eta+1,m}^{\dagger}$ can be equivalent to a diagonal block matrix, where the U_ν blocks on the main diagonal are $\boldsymbol{p}_m^T$. As a result, each of the rows in $\boldsymbol{\Theta}_{2,\eta+1}^{\dagger}$ cannot be obtained via linear combination of the other rows, indicating that $\mathrm{rank}(\boldsymbol{\Theta}_{2,\eta+1}^{\dagger}) = \mathrm{rank}(\bar{\boldsymbol{\Theta}}_{2,\eta+1}) = MU_\nu$, which is the rank of $\tilde{\boldsymbol{\Theta}}_{2,\eta+1}$. At last, the rank of $\tilde{\boldsymbol{\Theta}}_2$ is given as

$$\mathrm{rank}(\tilde{\boldsymbol{\Theta}}_2) = \sum_{u=1}^{\eta+1} \tilde{\boldsymbol{\Theta}}_{2,u} = \eta MU_o + MU_\nu = M(KU-1). \tag{108}$$

References

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

[2] R. Wang, S. Zhang, and L. Liang, "Low-overhead channel estimation for beyond diagonal reconfigurable intelligent surface aided single-user communication," in *Proc. IEEE Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2024.

[3] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, Aug. 2019.

[4] M. Jian, G. C. Alexandropoulos, E. Basar, C. Huang, R. Liu, Y. Liu, and C. Yuen, "Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques," *Intell. Conv. Networks*, vol. 3, no. 1, Mar 2022.

[5] H. Li, S. Shen, M. Nerini, and B. Clerckx, "Reconfigurable intelligent surfaces 2.0: Beyond diagonal phase shift matrices," *IEEE Commun. Mag.*, vol. 62, no. 3, pp. 102–108, Mar. 2024.

[6] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1229–1243, Feb. 2022.

[7] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, Apr. 2023.

[8] Z. Wu and B. Clerckx, "Beyond Diagonal RIS in Multiuser MIMO: Graph Theoretic Modeling and Optimal Architectures with Low Complexity," Feb. 2025. [Online]. Available: https://arxiv.org/pdf/2502.16509

[9] M. Nerini, S. Shen, H. Li, M. Di Renzo, and B. Clerckx, "A universal framework for multiport network analysis of reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14 575–14 590, Jun. 2024.

[10] M. Nerini, S. Shen, H. Li, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces utilizing graph theory: Modeling, architecture design, and optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 9972–9985, Feb. 2024.

[11] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: A multi-sector mode enabling highly directional full-space wireless coverage," *IEEE J. Sel. Areas Commnun.*, vol. 41, no. 8, pp. 2446–2460, Aug. 2023.

[12] M. Nerini, S. Shen, and B. Clerckx, "Closed-form global optimization of beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1037–1051, Feb. 2024.

[13] I. Santamaria, M. Soleymani, E. Jorswieck, and J. Gutiérrez, "MIMO capacity maximization with beyond-diagonal RIS," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Oct. 2024.

[14] M. Samy, H. Al-Hraishawi, A. B. M. Adam, K. Ntontin, S. Chatzinotas, and B. Ottersten, "beyond diagonal RIS-aided networks: performance analysis and sectorization tradeoff," Jun. 2024. [Online]. Available: https://arxiv.org/abs/2406.04009

[15] E. Björnson and Ö. T. Demir, "Capacity maximization for MIMO channels assisted by beyond-diagonal RIS," Jan. 2025. [Online]. Available: https://arxiv.org/abs/2411.18298

[16] M. Nerini, S. Shen, and B. Clerckx, "Discrete-value group and fully connected architectures for beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16 354–16 368, Dec. 2023.

[17] B. Wang, H. Li, S. Shen, Z. Cheng, and B. Clerckx, "A dual-function radar-communication system empowered by beyond diagonal reconfigurable intelligent surface," *IEEE Trans. Commun.*, Aug. 2024.

[18] M. Raeisi, H. Chen, H. Wymeersch, and E. Basar, "Efficient localization with base station-integrated beyond diagonal RIS," Nov. 2024. [Online]. Available: https://arxiv.org/abs/2411.13295

[19] K. Chen and Y. Mao, "Transmitter side beyond-diagonal RIS for mmwave integrated sensing and communications," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sep. 2024, pp. 951–955.

[20] Z. Guang, Y. Liu, Q. Wu, W. Wang, and Q. Shi, "Power minimization for ISAC system using beyond diagonal reconfigurable intelligent surface," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 13 950–13 955, Apr. 2024.

[21] S. Zheng and S. Zhang, "Beyond diagonal intelligent reflecting surface aided integrated sensing and communication," *IEEE Trans. Cogn. Commun. Network.*, early access, Jun. 2025. [Online]. Available: https://arxiv.org/abs/2505.16230

[22] ——, "BD-IRS aided uplink ISAC exploiting prior information: SDMA or TDMA," in *Proc. IEEE Int. Workshop Signal Process. Artificial Intell. Wireless Commun. (SPAWC)*, Jul. 2025.

[23] H. Wang, Z. Han, and A. L. Swindlehurst, "Channel reciprocity attacks using intelligent surfaces with non-diagonal phase shifts," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1469–1485, Feb. 2024.

[24] H. Li and B. Clerckx, "Non-reciprocal beyond diagonal RIS: Multiport network models and performance benefits in full-duplex systems," Nov. 2024. [Online]. Available: https://arxiv.org/abs/2411.04370

[25] H. Li, S. Shen, Y. Zhang, and B. Clerckx, "Channel estimation and beamforming for beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 72, pp. 3318–3332, Jul. 2024.

[26] H. Li, Y. Zhang, and B. Clerckx, "Channel estimation for beyond diagonal reconfigurable intelligent surfaces with group-connected architectures," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sens. Adapt. Process. (CAMSAP)*, Dec. 2023.

[27] A. L. F. de Almeida, B. Sokal, H. Li, and B. Clerckx, "Channel estimation for beyond diagonal RIS via tensor decomposition," *IEEE Trans. Signal Process.*, early access, May 2025.

[28] G. T. de Araújo and A. L. F. de Almeida, "Semi-blind channel estimation for beyond diagonal RIS," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Oct. 2024, pp. 1586–1590.

[29] N. Ginige, A. S. de Sena, N. H. Mahmood, N. Rajatheva, and M. Latva-aho, "Efficient channel prediction for beyond diagonal RIS-assisted MIMO systems with channel aging," *IEEE Trans. Veh. Technol.*, early access, Mar. 2025.

[30] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmwave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6853–6869, Feb. 2023.

[31] 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.

[32] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications: Part II - an improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.

[33] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 1478–1492, Sep. 2022.

[34] Z. Peng, C. Pan, G. Zhou, H. Ren, S. Jin, P. Popovski, R. Schober, and X. You, "Two-stage channel estimation for RIS-aided multiuser mmWave systems with reduced error propagation and pilot overhead," *IEEE Trans. Signal Process.*, vol. 71, pp. 3607–3622, Sep. 2023.

[35] B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec. 2020.

[36] H. V. Henderson and S. R. Searle, “The vec-permutation matrix, the vec operator and kronecker products: A review,” *Linear and Multilinear Algebra*, vol. 9, no. 4, pp. 271–288, 1981.

[37] D. Mishra and H. Johansson, “Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2019.

[38] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

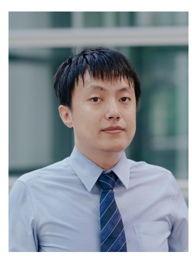
[39] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[40] Z. Chen, “On nonsingularity of circulant matrices,” *Linear Algebra and its Applications*, vol. 612, pp. 162–176, 2021.

Rui Wang (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2017 and the M.S. degree in information and communication engineering from Southeast University, Nanjing, China, in 2020. He is currently working towards the Ph.D. degree at the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong. His research interests include intelligent reflecting surface assisted wireless communication, channel estimation and MIMO systems.

Shuowen Zhang (Senior Member, IEEE) received the B.Eng. degree in information engineering from the Chien-Shiung Wu Honors College, Southeast University, Nanjing, China, in June 2013, and the Ph.D. degree from NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore, in January 2018 under the NGS scholarship. After completing her Ph.D. studies, she served as a Research Fellow at the Department of Electrical and Computer Engineering, National University of Singapore, and then joined The Hong Kong Polytechnic University where she is currently an Assistant Professor with the Department of Electrical and Electronic Engineering. Her research interests include integrated sensing and communication, intelligent surface aided communication, unmanned aerial vehicles, multiple-input multiple-output (MIMO), communication theory, and optimization methods. Dr. Zhang is currently serving as an Editor for IEEE Transactions on Wireless Communications and an Associate Editor for IEEE Transactions on Mobile Computing. She has served as a Guest Editor for various journals such as the IEEE Journal on Selected Areas in Communications. Dr. Zhang is the sole recipient of the 2021 Marconi Society Paul Baran Young Scholar Award, as well as a recipient of the 2022 IEEE Communications Society Young Author Best Paper Award, the 2023 IEEE Communications Society Best Tutorial Paper Award, the 2023 PolyU Young Innovative Researcher Award, and the 2024 IEEE Communications Society Asia-Pacific Outstanding Young Researcher Award.

Bruno Clerckx (Fellow, IEEE) is a Full Professor, the Head of the Communications and Signal Processing Group, and the Head of the Wireless Communications and Signal Processing Lab, within the Electrical and Electronic Engineering Department, Imperial College London, London, U.K. He received the MSc and Ph.D. degrees in Electrical Engineering from Université Catholique de Louvain, Belgium, and the Doctor of Science (DSc) degree from Imperial College London, U.K. He spent many years in industry with Silicon Austria Labs (SAL), Austria, where he was the Chief Technology Officer (CTO) responsible for all research areas of Austria’s top research center for electronic based systems and with Samsung Electronics, South Korea, where he actively contributed to 4G (3GPP LTE/LTE-A and IEEE 802.16m). He has authored two books on “MIMO Wireless Communications” and “MIMO Wireless Networks”, over 350 peer-reviewed international research papers, and 150 standards contributions, and is the inventor of 80 issued or pending patents among which several have been adopted in the specifications of 4G standards and are used by billions of devices worldwide. His research spans the general area of wireless communications and signal processing for wireless networks. He received the prestigious Blondel Medal 2021 from France for exceptional work contributing to the progress of Science and Electrical and Electronic Industries, the 2021 Adolphe Wetrems Prize in mathematical and physical sciences and the 2024 Georges Vanderlinden Prize in Electromagnetism and Telecommunications from Royal Academy of Belgium, multiple awards from Samsung, IEEE best student paper award, and the EURASIP (European Association for Signal Processing) best paper award 2022. He is a Fellow of the IEEE and the IET.



Liang Liu (Senior Member, IEEE) is currently an Associate Professor with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University. He received the B.Eng. degree from Tianjin University, China, in 2010, and the Ph.D. degree from the National University of Singapore, Singapore, in 2014. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, from 2015-2017, a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, from 2017-2019, and an Assistant Professor with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, from 2019-2024. His research interests lie in the next generation cellular technologies such as machine-type communications for the Internet of Things, integrated sensing and communication, etc. He is an IEEE ComSoc Distinguished Lecturer for the class of 2025-2026. He was a recipient of the 2021 IEEE Signal Processing Society Best Paper Award, the 2017 IEEE Signal Processing Society Young Author Best Paper Award, the Best Student Award of 2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the Best Paper Award of the 2011 International Conference on Wireless Communications and Signal Processing. He was recognized by Clarivate Analytics as a Highly Cited Researcher in 2018. He is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was a Leading Guest Editor of IEEE WIRELESS COMMUNICATIONS Special Issue on Massive Machine-Type Communications for IoT. He is a co-author of the book ”Next Generation Multiple Access” published at Wiley-IEEE Press.