

# Channel Estimation and Passive Beamforming for Pixel-based Reconfigurable Intelligent Surfaces with Non-Separable State Response

Huayan Guo, *Member, IEEE*, Junhui Rao, *Graduate Student Member, IEEE*,  
Alex M. H. Wong, *Senior Member, IEEE*, Ross Murch, *Fellow, IEEE*, and Vincent K. N. Lau, *Fellow, IEEE*,

**Abstract**—Pixel-based reconfigurable intelligent surfaces (RISs) employ a novel design to achieve high reflection gain at a lower hardware cost by eliminating the phase shifters used in traditional RIS. However, this design presents challenges for channel estimation and passive beamforming due to its non-separable state response, rendering existing solutions ineffective. To address this, we first approximate the non-separable RIS response functions using a kernel-based method and a deep neural network, achieving high accuracy while reducing computational and memory complexity. Next, we propose a simplified cascaded channel model that focuses on dominated scattering paths with limited unknown parameters, along with customized algorithms to estimate short-term and long-term parameters separately. Finally, we introduce a low-complexity passive beamforming algorithm to configure the discrete RIS state vector, maximizing the achievable rate. Our simulation results demonstrate that the proposed solution significantly outperforms various baselines across a wide SNR range.

**Index Terms**—Intelligent reflecting surface (IRS), reconfigurable intelligent surface (RIS), non-separable state response, channel estimation, passive beamforming

## I. INTRODUCTION

Reconfigurable intelligent surfaces (RISs) are flat surfaces that alter the propagation environment by manipulating the properties of reflected waves [1]–[7]. Each element of the surface can adjust the reflection responses of incoming waves independently by switching among different RIS states. Through proper configurations using passive beamforming techniques, the RIS can establish a virtual line-of-sight (LoS) propagation path between the transmitter and receiver. This capability offers potential benefits in various areas, including cellular communication [8]–[10], mobile edge computation [11]–[13], and air-ground communication [14]–[16]. The RIS discussed in these papers uses phased array with phase shifters that have 2 or more bits of resolution. However, phase shifters

This work was supported in part by the Hong Kong Research Grants Council Collaborative Research Project Grant C6012-20GF, in part by the National Natural Science Foundation of China under Grant 62101472, and in part by the Research Grants Council under the Areas of Excellence scheme grant AoE/E-101/23-N (Corresponding author: Vincent K. N. Lau.)

Huayan Guo, Junhui Rao, Ross Murch, and Vincent K. N. Lau are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong 999077 (e-mail: eeguohuayan@ust.hk; jraoaa@connect.ust.hk; eermurch@ust.hk; eeknlau@ece.ust.hk).

Alex M. H. Wong is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China, and also with the State Key Laboratory of Terahertz and Millimeter Waves, City University of Hong Kong, Hong Kong SAR, China (e-mail: alex.mh.wong@cityu.edu.hk).

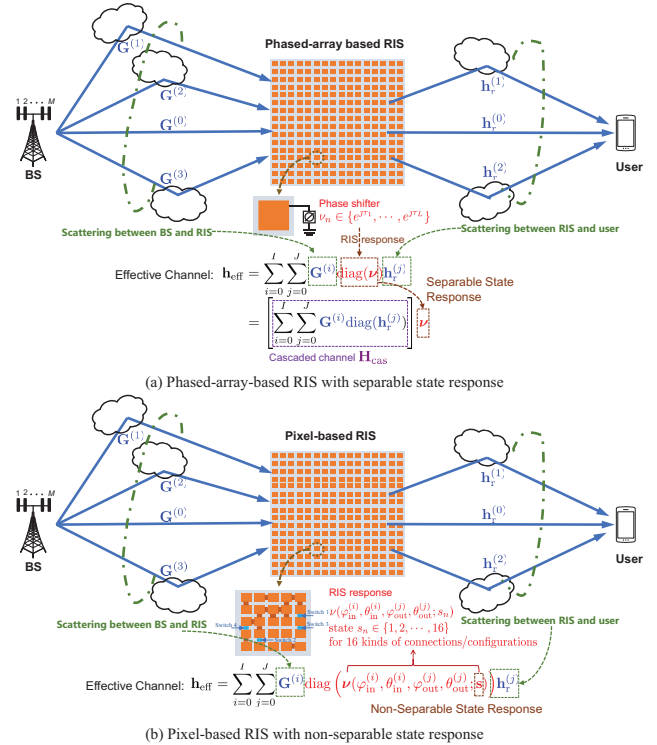


Fig. 1. Illustration of the RIS-assisted downlink MISO system.

are costly in practice and occupy space on the RIS, limiting design flexibility. To tackle this issue, a new design called pixel-based RIS has been introduced in [4]. This design allows flexible control over the radiation elements without needing phase shifters. To fully utilize its benefits, effective channel estimation and passive beamforming techniques for the pixel-based RIS are essential.

For traditional phased-array RIS, designing passive beamforming usually needs accurate channel estimation of the base station (BS)-RIS-user link, as shown in Fig. 1(a). The effective channel between the BS and the user ( $\mathbf{h}_{\text{eff}}$ ) can be modeled as the multiplication of separable components, including the wireless channel from the BS to the RIS ( $\mathbf{G}$ ), the phase shifter network ( $\nu$ ), and the wireless channel from the RIS and the user ( $\mathbf{h}_r$ ). Note that the phase shifter network configuration  $\nu$  can be separated out (see Fig. 1(a)). This allows the effective channel to be represented as the product of a cascaded channel response ( $\mathbf{H}_{\text{cas}}$ ), (which encapsulated

the scattering effects in the BS-RIS and RIS-user links and is independent to the phased array setup), and a *RIS response term*  $\nu$  which depends only on the phased array configuration [17]–[23]. As a result, the conventional phased array RIS shows a *separable channel response*. This property allows the use of standard channel estimation algorithms [24]–[28] to estimate  $\mathbf{H}_{\text{cas}}$  from data collected with different phased array configurations. The estimated  $\mathbf{H}_{\text{cas}}$  can then be used to predict beamforming performance for other configurations, making it useful for beamforming design.

However, the pixel RIS results in a BS-RIS-user channel with a *non-separable state response*. Specifically, the effective BS-RIS-user channel in the pixel-RIS-assisted system,  $\mathbf{h}_{\text{eff}}$ , is a non-separable function of the RIS configuration (i.e., states  $\mathbf{s}$ ) and the angle of departure/arrivals ( $\varphi_{\text{AoD}}, \theta_{\text{AoD}}, \varphi_{\text{AoA}}, \theta_{\text{AoA}}$ ) of the scattering paths in the BS-RIS and RIS-user links, as shown in Fig. 1(b). The RIS responses  $\nu$  will be different for different pairs of incident and reflection rays, and the influence of the RIS state configuration  $\mathbf{s}$  cannot be isolated as a separable term in the effective BS-RIS-user channel. Passive beamforming poses a challenge for RIS with a non-separable state response, as existing channel estimation and passive beamforming strategies become inapplicable. This is because the channel estimation for one RIS configuration cannot be directly used to predict the effective channel response of the pixel-based RIS system at other configurations and hence, cannot be used directly for beamforming design.

Additionally, there is no simple closed-form expression for the 4D array response function for each state of the pixel RIS due to its complicated electromagnetic scattering properties. One method is to create a 4D lookup table and then use curve fitting tools to approximate the response function. However, a lookup table with 1-degree angle resolution would need 3.5 GB of memory, and storing the fitting parameters would need even more space, making this method impractical in terms of memory and time. Therefore, designing a practical approximation for the RIS response function that has low computational and memory complexity is highly desirable for both channel estimation and performance evaluation of the pixel-RIS-assisted wireless system.

Furthermore, passive beamforming in the pixel-RIS-assisted system involves discrete optimization variables to adjust each RIS element to an appropriate state. Continuous relaxation [29]–[32] is employed for traditional phased array RIS, followed by regularization or projection to minimize the relaxation error by leveraging candidate phase values that are uniformly located around a unit circle. However, this property does not hold for the pixel RIS, and a new method is required to tackle the discrete state selection challenge.

More generally, these issues can arise with any RIS that uses elements that change their own structure, instead of adjusting the reflected phase through an external port connected to a phase shifter. This is because reconfiguring the structure alters the element's scattering characteristics, including the magnitude and directionality, resulting in non-separable states. For example, the pixel based elements focused on in this work fall into this category, as the switches between pixels

change the element structure. Furthermore, any element made from sub-elements will potentially also fall into this category, including those elements with switches and elements made up of sub-elements in metasurfaces [5]–[7]. While this work focuses on the pixel based RIS further investigation on other RISs may also reveal non-separable state properties that need to be considered.

In this paper, we investigate passive beamforming and channel estimation for a multiple-user multiple-input-single-output (MU-MISO) system assisted by pixel-based RIS with a non-separable state response. The following summarizes the key contributions of this work.

- Compact Representation of RIS Response Functions for Reduced Computational and Memory Usage:** Accurate and low-overhead representation of the RIS response is essential for channel estimation and passive beamforming design. To achieve this, we first propose a product-Legendre-kernel-function-based approximation, extending existing antenna radiation pattern approximation methods [33], [34] from 2D to 4D functions. Additionally, we introduce a task-specific deep neural network (DNN) to approximate the RIS response function. This DNN employs a product structure to address features from incident and reflection angles, incorporating multiple parallel blocks that mimic kernel functions with the same structure but different weights. The kernel-based solution achieves a normalized mean square error (NMSE) of around 0.001 in most cases, while the DNN-based solution further reduces the NMSE by about 100 times with much lower computational complexity.
- Effective Estimation and Prediction of Cascaded RIS Channel with Non-separable State Response:** To tackle the challenge of channel estimation in the RIS system with a non-separable state response, we propose a simplified channel model focusing on dominated scattering paths with limited unknown parameters. We first introduce a three-step estimation algorithm for instantaneous channel responses given incident/reflection angles. Building on this, we develop an algorithm to estimate these angles using historical channel observations, as they are long-term parameters that vary slowly. Those estimated channel responses and angles are then substituted into the proposed model for channel prediction. Simulations show that the proposed algorithm accurately predicts cascaded channels with new RIS state vectors that have not been observed during channel sounding.
- Low-Complexity Passive Beamforming Algorithm for Discrete RIS State Configuration:** The state configuration in the RIS system corresponds to passive beamforming, modeled as a discrete optimization problem regarding the discrete state vector. To address the combinatorial optimization challenge, we reparameterize each discrete state variable with a one-hot state selection vector, which is relaxed to a continuous-valued vector representing selection probabilities. A sparse regularization constraint is then introduced to encourage the selection probability vector to approach a one-hot configuration. Simulation re-

sults demonstrate that the proposed passive beamforming algorithm significantly outperforms various baselines.

## II. SYSTEM MODEL

### A. RIS with Separable and Non-Separable State Responses

The concept of RIS with separable and non-separable state responses have been described in the introduction and also illustrated in Fig. 1. Further details on how these are commonly implemented are described next.

1) *Phase-Shifting Array-based RIS*: A common implementation of RIS is the phase-shifting array-based RIS, often referred to as a reflectarray, which is an RIS with a separable state response. As shown in the Fig. 2, each element in this type of RIS consists of an antenna loaded with a phase shifter. The antenna receives the signal from the environment, which is then reflected and its phase adjusted by the phase shifter before being re-radiated into free space. This process allows the phase of the reflected signal to be altered without affecting the magnitude, thus separating the reflected phase from the effective channel.

An important electromagnetic property that is often overlooked in the reflectarray approach is that only about half of the incident signals can be controlled by the phase shifter. The total scattered field by the antenna can be divided into two components [35]: the structure scattering term, which is independent of the antenna load, and the antenna mode scattering term, which is tunable by the phase shifter. Typically, antennas with metal grounds, such as patch antennas, are used in RIS applications, resulting in approximately half of the incident power being scattered as the structure scattering term [36]. That is, only approximately half of the scattered signal can be controlled by the RIS and consequently, phase-shifting array-based RIS inherently suffers an approximately 3 dB loss. Furthermore, phase shifter losses can also be high, especially for the common configuration shown in Fig. 2 where the signal passes both forward and backward through the phase shifter, leading to an additional loss of 2 dB or more.

2) *Pixel-based RIS*: As shown in Fig. 3(a), a non-separable state response RIS is the pixel-based RIS first proposed in [4], which consists of a  $5 \times 5$  pixel array with a metal ground and an air gap beneath it. The dimension of each pixel is much smaller than a wavelength while the pixel array is of the scale of a wavelength. The pixels are selectively interconnected, with four RF switches (PIN diodes) strategically placed to create specific resonating structures at the wavelength scale. By changing the states of these diodes, the pattern can be switched to another with different properties. Unlike phase-shifting array-based RIS, pixel-based RIS do not experience the 3 dB loss due to structure scattering, as it considers the total scattered field. In addition the double losses through the external phase shifter are also avoided. However, while the reflected phases can be finely tuned to cover 360 degrees, they depend on both the diode states and the angles of departure/arrival as shown in Fig. 3(b), leading to a non-separable state response.

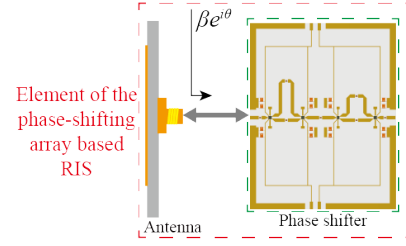


Fig. 2. The architecture of phase-shifting array based RIS element.

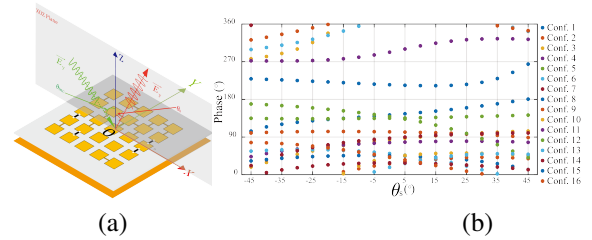


Fig. 3. (a) Details of pixel-based RIS element and coordinate system with incident and reflected angles shown. (b) Phases of the reflected wave are shown by colored filled circles, versus  $\theta_s$  [see (a)] for 16 different RIS configurations (at 2.4 GHz) when excited by vertically incident plane wave.

3) *Motivation for the Development of Pixel-Based RIS*: For flexible control of RIS elements, the phased array needs phase shifters (and for full pattern control elements also ideally need to have amplitude control). However, these items are difficult to prototype and are not cost effective when offering 2 or more bits of resolution. They also take up space on the RIS limiting design flexibility. In addition, they also increase the losses in the RIS. For instance, the signal received by an element must pass through the phase shifter twice—once for reflection and again for reradiation—resulting in a double loss effect that can increase losses by 2-3 dB per element.

In contrast, the pixel antenna approach provides a very flexible and general formulation for controlling radiation element characteristics. Nearly any element geometry can be formed by appropriately connecting the pixels together using electronically controlled switches. This allows the full degrees of freedom within the element area to be exploited and this allows more compact designs compared to phased arrays. The pixels are connected by switches integrated into each element and are therefore compact to include. Pixel-based elements also enable fine control of the radiation characteristics and can provide a wide range of phase shifts, as shown in the published pixel RIS design and prototype [4]. Furthermore, since there are no phase shifters and signals reflect directly off the pixel surface, losses are minimized, eliminating the double loss effect.

In summary, the pixel approach to RIS design does not rely on phase shifters, resulting in reduced losses. With only a few switches needed per element, this method is cost-effective and allows for highly reconfigurable elements that can exhibit a wide variety of characteristics as required. Therefore, the pixel approach offers a competitive alternative to traditional RIS designs that utilize phase shifters.

### B. Scattering Model for BS-RIS and RIS-UE Channels for the Pixel-based RIS

Let  $M$  be the number of BS antennas,  $N$  the number of RIS elements,  $L$  the number of candidate RIS states, and  $K$  the number of users. We consider a propagation environment with limited scatterers with low user mobility. The wireless propagation parameters, such as scattering cluster power, path loss, and arrival/departure angles, remain constant over an extended period, while the small-scale channel coefficients for different signal paths vary across time blocks. Specifically, according to the 3GPP channel model [37, Section 7], the channel from BS to RIS in the  $t$ -th time block is given by

$$\mathbf{G}_t = \mathbf{G}^{(0)} + \sum_{i=1}^I \mathbf{G}_t^{(i)}, \quad (1)$$

where  $\mathbf{G}^{(0)}$  is the LoS component and  $\mathbf{G}_t^{(i)}$  is the NLoS component for the  $i$ -th scattering cluster. Define  $\phi_{\text{BS}} = \{\varphi_{\text{BS}}, \theta_{\text{BS}}\}$  and  $\phi_{\text{in}} = \{\varphi_{\text{in}}, \theta_{\text{in}}\}$ , where  $\varphi_{\text{BS}}$  and  $\theta_{\text{BS}}$  denote the azimuth and elevation angles departing from BS to RIS,  $\varphi_{\text{in}}$  and  $\theta_{\text{in}}$  denote the incident azimuth and elevation angles to the RIS. The components in  $\mathbf{G}_t$  are given by

$$\mathbf{G}^{(0)} = \alpha_{\text{BS}}(\phi_{\text{BS}}^{(0)}) \alpha_{\text{RIS}}^{\text{H}}(\phi_{\text{in}}^{(0)}), \quad (2a)$$

$$\mathbf{G}_t^{(i)} = a_t^{(i)} \sum_{d=1}^{D_i} \mathbf{G}_t^{(i,d)}, \quad (2b)$$

$$\mathbf{G}_t^{(i,d)} = e^{j\epsilon_t^{(i,d)}} \alpha_{\text{BS}}(\phi_{\text{BS}}^{(i,d)}) \alpha_{\text{RIS}}^{\text{H}}(\phi_{\text{in}}^{(i,d)}), \quad (2c)$$

where  $\mathbf{G}_t^{(i,d)}$  indicates the  $d$ -th ray in the  $i$ -th scattering cluster for the NLoS link and  $a_t^{(i)} \in \mathbb{R}$  denotes the path gain,  $\alpha_{\text{BS}} \in \mathbb{C}^{M \times 1}$  and  $\alpha_{\text{RIS}} \in \mathbb{C}^{N \times 1}$  are the steering vectors of BS and RIS arrays, respectively, and  $\epsilon$  denotes the random phase for different rays.

Similarly, the channel from RIS to the  $k$ -th user in the  $t$ -th time block is given by

$$\mathbf{h}_{r,k,t} = \mathbf{h}_{r,k,t}^{(0)} + \sum_{j=1}^J \mathbf{h}_{r,k,t}^{(j)}, \quad (3)$$

where  $\mathbf{h}_{r,k}^{(0)}$  is the LoS component and  $\mathbf{h}_{r,k,t}^{(j)}$  is the NLoS component for the  $j$ -th scattering cluster of user  $k$ . Define  $\phi_{\text{out},k} = \{\varphi_{\text{out},k}, \theta_{\text{out},k}\}$  where  $\varphi_{\text{out},k}$  and  $\theta_{\text{out},k}$  denote the reflection azimuth and elevation angles from RIS to the  $k$ -th user. The components in  $\mathbf{h}_{r,k,t}$  are given by

$$\mathbf{h}_{r,k,t}^{(0)} = b_{k,t}^{(0)} \alpha_{\text{RIS}}(\phi_{\text{out},k}^{(0)}), \quad (4a)$$

$$\mathbf{h}_{r,k,t}^{(j)} = b_{k,t}^{(j)} \sum_{d=1}^{D_j} \mathbf{h}_{r,k,t}^{(j,d)}, \quad (4b)$$

$$\mathbf{h}_{r,k,t}^{(j,d)} = e^{j\epsilon_{k,t}^{(j,d)}} \alpha_{\text{RIS}}(\phi_{\text{out},k}^{(j,d)}), \quad (4c)$$

where  $\mathbf{h}_{r,k,t}^{(j,d)}$  is the  $d$ -th ray in the  $j$ -th scattering cluster of the NLoS link and  $b_{k,t}^{(j)} \in \mathbb{R}$  is the path gain, and  $\epsilon$  denotes the random phase response.

Let  $\nu(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; \ell)$  denote the response of a single element in the pixel-based RIS for the  $i$ -th incident ray and the  $j$ -th reflected ray, when the element is configured to the  $\ell$ -th state.

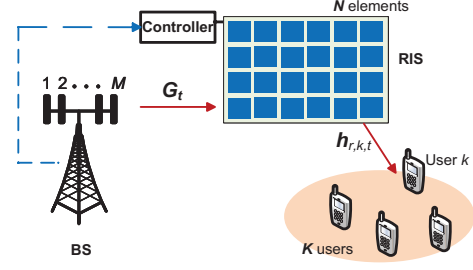


Fig. 4. Illustration of the pixel-based-RIS-assisted MU-MISO system.

Let  $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$  denote the RIS state configuration vector for all the  $N$  elements. The RIS response vector is given by:

$$\boldsymbol{\nu}(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; \mathbf{s}) = \left[ \nu(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; \ell = s_1) \right. \\ \left. \dots, \nu(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; \ell = s_N) \right]^T. \quad (5)$$

Finally, the cascaded channel from BS to the  $k$ -th user at  $t$ -th time block becomes a function of RIS state  $\mathbf{s}_t$ :

$$\mathbf{h}_{k,t}(\mathbf{s}_t) = \sum_{i,j} \mathbf{G}_t^{(i)} \text{diag} \left( \boldsymbol{\nu}(\phi_{\text{in}}^{(i)}, \phi_{\text{out},k}^{(j)}; \mathbf{s}_t) \right) \mathbf{h}_{r,k,t}^{(j)}. \quad (6)$$

From (6), we see that to predict  $\mathbf{h}_{k,t}(\mathbf{s})$  for any state vector  $\mathbf{s}$ , one needs the angles  $\phi_{\text{in}}^{(i)}$  and  $\phi_{\text{out},k}^{(j)}$  for all propagation rays, as well as the RIS response function  $\nu(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; \ell)$  for any states. Note that the RIS response function is determined solely by the structure of the pixel RIS element and is independent of the wireless propagation variables  $\mathbf{G}_t^{(i)}$ ,  $\mathbf{h}_{r,k,t}^{(j)}$ ,  $\phi_{\text{in}}^{(i)}$  and  $\phi_{\text{out},k}^{(j)}$ .

### C. Overall Signal Model for the Pixel-based-RIS-Assisted MU-MISO System

The overall structure of the multiuser-(MU)-MISO communication system assisted by a pixel-based RIS is illustrated in Fig. 4. We assume that the direct link between BS and users are blocked by obstacles and hence, the signals of direct link are ignored for simplicity. In the  $t$ -th time block, the signal received by the  $k$ -th user is given by

$$y_{k,t} = \mathbf{h}_{k,t}(\mathbf{s}_t)^T \sum_{k=1}^K \mathbf{w}_{k,t} x_{k,t} + z_{k,t}, \quad (7)$$

where  $\mathbf{w}_{k,t}$  denotes the precoder for the  $k$ -th user,  $x_{k,t}$  denotes the signal of the  $k$ -th user, and  $z_{k,t} \sim \mathcal{CN}(0, \sigma_z^2)$  denotes the additive white Gaussian noise (AWGN) at the  $k$ -th user's receiver.

To simplify the analysis and focus on the passive beamforming, we utilize a regular zero-forcing (RZF) precoder. Letting  $\mathbf{W}_t = [\mathbf{w}_{1,t}, \mathbf{w}_{2,t}, \dots, \mathbf{w}_{K,t}]$ , and  $\mathbf{H}_t(\mathbf{s}_t) = [\mathbf{h}_{1,t}(\mathbf{s}_t), \mathbf{h}_{2,t}(\mathbf{s}_t), \dots, \mathbf{h}_{K,t}(\mathbf{s}_t)]$ . The precoder is given by

$$\mathbf{W}_t(\mathbf{s}_t) = \sqrt{\gamma(\mathbf{s}_t)} \mathbf{H}_t^*(\mathbf{s}_t) \left( \mathbf{H}_t^T(\mathbf{s}_t) \mathbf{H}_t^*(\mathbf{s}_t) + \varsigma \mathbf{I}_K \right)^{-1}, \quad (8)$$

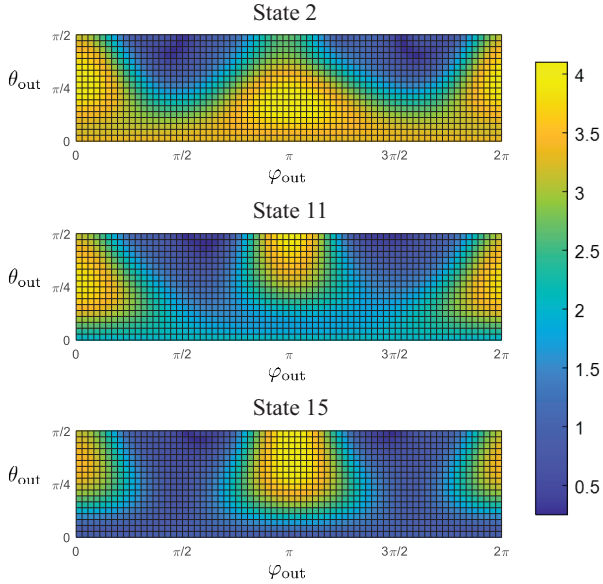


Fig. 5. Illustration of the magnitudes of  $\nu$ , for a single RIS element, for different  $\varphi_{\text{out}}$  and  $\theta_{\text{out}}$  by setting  $\varphi_{\text{in}} = 0$ ,  $\theta_{\text{in}} = \frac{\pi}{4}$  for states 2, 11 and 15.

where  $\varsigma \geq 0$  is a small constant that makes the matrix inverse possible, and  $\gamma(\mathbf{s}_t)$  is used to satisfy the transmit power constraint. The achievable rate of user  $k$  is given by

$$R_{k,t}(\mathbf{s}_t) = \log \left( 1 + \frac{|\mathbf{h}_{k,t}(\mathbf{s}_t)^T \mathbf{w}_{k,t}(\mathbf{s}_t)|^2}{\sum_{i \neq k} |\mathbf{h}_{i,t}(\mathbf{s}_t)^T \mathbf{w}_{i,t}(\mathbf{s}_t)|^2 + \sigma_z^2} \right). \quad (9)$$

The primary objective of this paper is to configure an appropriate RIS state  $\mathbf{s}_t$  to maximize the sum rate:

$$\mathcal{P}(\mathbf{A}) \quad \mathbf{s}_t^* = \underset{\mathbf{s}_t}{\operatorname{argmax}} \sum_{K=1}^K R_{k,t}(\mathbf{s}_t) \quad (10a)$$

$$\text{s.t. } s_{n,t} \in \{1, 2, \dots, L\}, \forall n, \quad (10a)$$

$$\frac{1}{K} \sum_k \|\mathbf{w}_{k,t}(\mathbf{s}_t)\|_2^2 \leq P_T. \quad (10b)$$

### III. APPROXIMATION FOR THE PIXEL-BASED RIS RESPONSE FUNCTION

The pixel RIS prototype in [4] operates at a center frequency of 2.4 GHz, offering a consistent response over about 100 MHz of bandwidth (see [4, Table II]), and its response function is independent of the wireless propagation parameters. In this section, we describe the design of approximate functions for the pixel RIS response ( $\nu(\phi_{\text{in}}, \phi_{\text{out}}; \ell)$ ) for a single RIS element configured in the  $\ell$ -th state. We first propose an approximation by creating a kernel from the product of Legendre functions, followed by a DNN solution that has lower computational complexity but higher approximation accuracy.

#### A. Product-Legendre-Kernel-Based Approximation

The magnitudes of  $\nu(\phi_{\text{in}}, \phi_{\text{out}}; \ell)$  for a single RIS element are illustrated in Fig. 5 by setting  $\varphi_{\text{in}} = 0$  and  $\theta_{\text{in}} = \frac{\pi}{4}$  for  $\ell = 1$  states 2, 11 and 15. It is seen that the response function is

smooth, suggesting that it can be approximated by a weighted sum of a group of kernel functions.

While the approximation of the 4-D RIS response has not been explored, some studies have investigated the approximation of a 2-D antenna radiation pattern using Legendre-based kernel functions [33], [34], demonstrating good performance. In this paper, we generalize the 2-D Legendre-based kernel to a 4-D product-Legendre-based kernel for approximating the RIS response. The approximate function, denoted by  $\bar{\nu}_\ell$ , is defined as follows:

$$\bar{\nu}_\ell(\phi_{\text{in}}, \phi_{\text{out}}; \boldsymbol{\omega}_\ell) = \sum_{c_i, c_o} \omega_{\ell, c_i, c_o} f_{L, c_i}(\phi_{\text{in}}) f_{L, c_o}(\phi_{\text{out}}), \quad (11)$$

where  $\boldsymbol{\omega}_\ell = \{\omega_{\ell, c_i, c_o}, \forall c_i, c_o\}$  is the parameter set to be trained, and the proposed kernel  $f_{L, c_i}(\psi_{\text{in}}) f_{L, c_o}(\psi_{\text{out}})$  are the product of two Legendre-based kernel functions. This product structure arises from the idea that the RIS response can be viewed as a product of the incident and reflection responses. Specifically, the Legendre-based kernel  $f_{L, c}(\phi)$  is given by

$$f_{L, c}(\phi) = Y_b^r(\varphi, \theta) = \begin{cases} (-1)^r \sqrt{2} \kappa_b^r \cos(r\varphi) f_{\text{Leg}, b}^r(\cos(\theta)), & 0 < r \leq b, \\ (-1)^r \sqrt{2} \kappa_b^r \cos(r\varphi) f_{\text{Leg}, b}^{-r}(\cos(\theta)), & -b \leq r < 0, \\ \kappa_b^0 f_{\text{Leg}, c}^r(\cos(\theta)), & r = 0, \end{cases} \quad (12)$$

where  $c = b^2 + b + r + 1$ ,  $0 \leq b \leq B$ ,  $\kappa_b^r = \sqrt{\frac{2b+1}{4\pi} \frac{(b-|r|)!}{(b+|r|)!}}$ , and  $f_{\text{Leg}, b}^r$  represents the Legendre function of degree  $b$  and order  $r$ .

#### B. DNN-Based Approximation

The proposed product-Legendre-kernel-based approximation requires a relatively large  $B$  to achieve high accuracy. However, the number of kernels increases with the square of  $B$ , leading to higher computational complexity. In this subsection, we introduce a DNN-based approximation that relies solely on linear matrix operations and simple activation functions.

1) *Data Pre-Processing*: Inspired by the Legendre-based kernel in (12), we introduce new variables  $\psi_{\text{in}} \in \mathbb{R}^3$  and  $\psi_{\text{out}} \in \mathbb{R}^3$  defined as follows:

$$\psi_{\text{in}} = [\cos(\varphi_{\text{in}}), \sin(\varphi_{\text{in}}), \cos(\theta_{\text{in}})], \quad (13)$$

$$\psi_{\text{out}} = [\cos(\varphi_{\text{out}}), \sin(\varphi_{\text{out}}), \cos(\theta_{\text{out}})]. \quad (14)$$

These variables better capture the periodic nature of the azimuth angles, particularly for the responses of states 11 and 15 shown in Fig. 5.

Same as (11), we denote the approximate function by  $\bar{\nu}_\ell$ , which becomes after replacing  $\{\phi_{\text{in}}, \phi_{\text{out}}\}$  by  $\{\psi_{\text{in}}, \psi_{\text{out}}\}$ :

$$\bar{\nu}_\ell(\phi_{\text{in}}, \phi_{\text{out}}; \boldsymbol{\omega}_\ell) = \bar{\nu}_\ell(\psi_{\text{in}}, \psi_{\text{out}}; \boldsymbol{\omega}_\ell), \quad (15)$$

where  $\boldsymbol{\omega}_\ell$  denotes the DNN parameters to be trained. The remaining task is to design the structure of the DNN function.

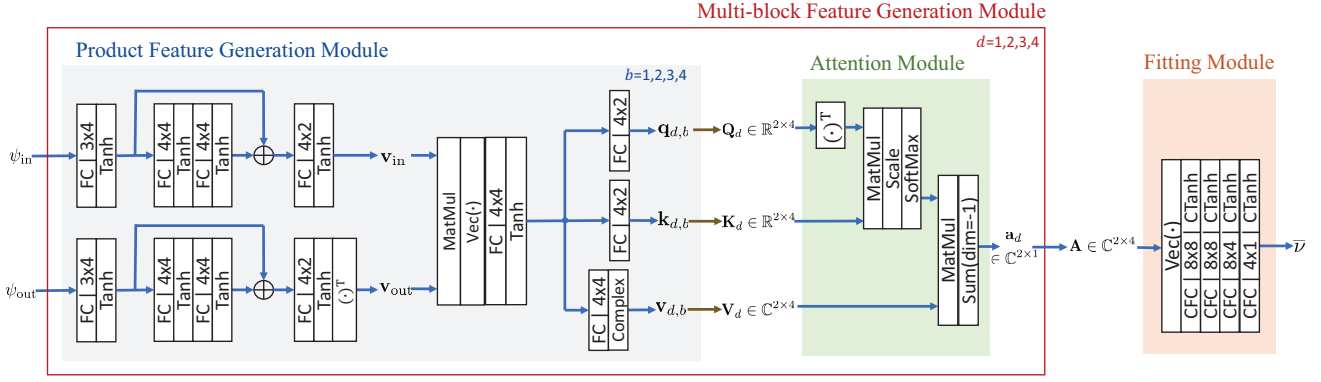


Fig. 6. Architecture of the proposed DNN. Note that “FC 4×2” means a fully-connected layer with 4 input channel and 2 output channel, “CFC 4×2” means a complex fully-connected layer, “CTanh” means a complex activation layer using Tanh(·) function, “MatMul” means matrix multiplication, “(·)<sup>T</sup>” means transpose a matrix, and “Vec(·)” means vectorization.

TABLE I  
NMSE OF THE RIS RESPONSE FUNCTION APPROXIMATIONS.

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8
Kernel	5.98e-3	3.92e-3	2.33e-3	4.63e-3	2.62e-3	3.86e-3	3.25e-3	3.12e-3
DNN	2.46e-5	5.91e-5	2.1e-6	5.8e-6	1.33e-5	1.48e-5	6.0e-6	2.1e-6
	State 9	State 10	State 11	State 12	State 13	State 14	State 15	State 16
Kernel	3.80e-3	6.15e-3	3.16e-3	1.36e-3	5.13e-3	1.22e-2	3.43e-3	1.45e-2
DNN	3.3e-6	6.3e-6	2.04e-5	1.22e-5	2.3e-6	5.3e-6	3.4e-6	1.13e-5

2) *DNN Structure*: The overall DNN design is illustrated in Fig. 6. To reduce complexity, the proposed DNN is designed as a two-tier parallel-block structure.

The inner tier includes a “Product Feature Generation Module” for parallel feature generation and an “Attention Module” to reduce the dimensionality of the stacked features.

- **Product Feature Generation Module**: This module comprises four parallel blocks ( $b = 1, 2, 3, 4$ ) with identical DNN structures but different weights. In each block, the output feature is generated using a product structure for the two data links, with inputs  $\psi_{in}$  and  $\psi_{out}$  processed separately. The DNN function is defined by

$$\mathbf{q}_{d,b}, \mathbf{k}_{d,b}, \mathbf{v}_{d,b} = f_{PF,d,b}(\psi_{in}, \psi_{out}; \omega_{PF,d,b}), \quad (16)$$

where  $\omega_{PF,b}$  is the DNN weights, and the overall weights is  $\omega_{PF,d} = \{\omega_{PF,d,b}, \forall b\}$ .

- **Attention Module**: This module reduces the dimension of the stacked feature  $\mathbf{V}_d$  to a vector  $\mathbf{a}_d$ :

$$\mathbf{a}_d = f_{AT,d}(\mathbf{Q}_d, \mathbf{K}_d, \mathbf{V}_d; \omega_{AT,d}), \quad (17)$$

where  $\omega_{AT,d}$  denotes the DNN weights.

The outer tier includes a “Multi-block Feature Generation Module” for parallel feature generation and a “Fitting Module” to reduce the dimensionality of the stacked features.

- **Multi-block Feature Generation Module**: This module comprises four parallel blocks ( $b = 1, 2, 3, 4$ ) with identical DNN structures but different weights. Each block is designed using to “inner tier” structure introduced above:

$$\mathbf{a}_d = f_{FG,d}(\psi_{in}, \psi_{out}; \omega_{FG,d}), \quad (18)$$

where the weights  $\omega_{FG,d} = \{\omega_{PF,d}, \omega_{AT,d}\}$ , and the overall weights are denoted by  $\omega_{FG} = \{\omega_{FG,d}, \forall d\}$ .

- **Fitting Module**: This module is simply designed as the cascaded of complex fully-connected layers to get the finally output:

$$\bar{\nu} = f_{FG}(\mathbf{A}; \omega_{FM}), \quad (19)$$

where  $\omega_{FM}$  summarizes the DNN weights. The overall DNN weights are  $\omega_{\ell} = \{\omega_{FG}, \omega_{FM}\}$

**Discussion**: Both the “Attention Module” and the “Fitting Module” can reduce feature dimensions; however, the “Attention Module” typically offers better performance but requires additional feature-related inputs. In contrast, the “Fitting Module” has a simpler structure, though its performance may be slightly lower. The proposed two-tier parallel-block structure effectively balances complexity and performance by applying these two types of modules in separate tiers.

### C. Training and Results

For both the product-Legendre-kernel-based solution and the DNN-based solution, we train the weights  $\omega_{\ell}$  using the same loss function to minimize the approximation NMSE:

$$\mathcal{P}(\mathbf{B}) \min_{\omega_{\ell}} \frac{1}{J} \sum_{j=1}^J \frac{|\bar{\nu}(\phi_{in}^{(j)}, \phi_{out}^{(j)}; \omega_{\ell}) - \nu^{(j)}|^2}{|\nu^{(j)}|^2}, \quad (20)$$

where  $J$  is the size of the training dataset. The training dataset is generated by randomly selecting  $10^7$  pairs of incident and reflection angles for each RIS state. Likewise, the validation and testing sets consist of  $10^6$  samples each. ADAM optimizer is employed to solve problem  $\mathcal{P}(\mathbf{B})$  with a learning rate of  $10^{-4}$ . The batch size is defined as 2048, and the maximum number of training epochs is set to be 1000.

TABLE II  
COMPLEXITY OF THE PROPOSED TWO SOLUTIONS.

	Kernel (B=5)	DNN
FLOPs	4530	1432
Parameters	625	550
Memory (KB)	4.6	196

We show the complexity of our two proposed solutions in Table II, detailing the number of DNN parameters, FLOPs, and memory needed for storage. Both solutions exhibit low computational and memory complexity. Specifically, the computational complexity of the DNN is significantly lower, while the Kernel-based solution requires less memory for storage.

We present the approximate NMSE for our proposed solutions in Table I. The product-Legendre-kernel-based solution achieves an NMSE of around 0.001 in most cases but performs poorly for states 14 and 16. In contrast, the DNN solution consistently performs well, reducing the NMSE by about 100 times compared to the product-Legendre-kernel-based solution.

#### IV. MODEL-BASED CHANNEL ESTIMATION FOR PIXEL-BASED-RIS-ASSISTED MU-MISO SYSTEM

##### A. Transmission Frames for RIS Sounding and Data Transmission

We utilize a transmission frame structure that separately supports RIS channel sounding and MU-MISO transmission, as illustrated in Fig. 7. Specifically, each time block is divided into two transmission phases.

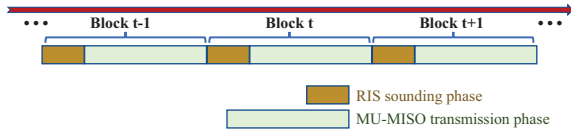


Fig. 7. Illustration of the transmission blocks in a timeline.

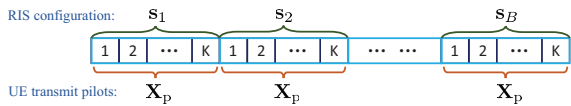


Fig. 8. Illustration of the transmission in the sounding phase.

1) *RIS Sounding Phase*: In this phase, we use a similar uplink channel sounding protocol for traditional RIS cascaded channel estimation [17]–[23], as shown in Fig. 8. The BS receives  $KB$  uplink sounding pilots from users, and the RIS elements are configured into  $B$  different states based on  $s_1, s_2, \dots, s_B$ . The users repeatedly transmit an orthogonal pilot matrix  $\mathbf{X}_p = [\mathbf{x}_{p,1}, \mathbf{x}_{p,2}, \dots, \mathbf{x}_{p,K}]$  a total of  $B$  times. The received signal is given by

$$\begin{aligned} \dot{\mathbf{Y}}_{u,t,b} &= \sum_{k=1}^K \mathbf{h}_{k,t}(s_b) \mathbf{x}_{p,k}^T + \dot{\mathbf{Z}}_{u,t,b} \\ &= \mathbf{H}_t(s_b) \mathbf{X}_p^T + \dot{\mathbf{Z}}_{u,t,b}, \end{aligned} \quad (21)$$

where  $\mathbf{H}_t(s_b) = [\mathbf{h}_{1,t}(s_b), \mathbf{h}_{2,t}(s_b), \dots, \mathbf{h}_{K,t}(s_b)]$  for  $b = 1, 2, \dots, B$ , and  $\dot{\mathbf{Z}}_{u,t,b}$  represents the received AWGN with elements following i.i.d.  $\mathcal{CN}(0, \sigma_u^2)$ . Since  $\mathbf{X}_p$  is an orthogonal matrix, we can remove it from (21), yielding:

$$\mathbf{Y}_{u,t,b} = \mathbf{H}_t(s_b) + \mathbf{Z}_{u,t,b}, \quad (22)$$

where  $\mathbf{Y}_{u,t,b} = \mathbf{x}_{p,k}^* \dot{\mathbf{Y}}_{u,t,b}$  and  $\mathbf{Z}_{u,t,b} = \mathbf{x}_{p,k}^* \dot{\mathbf{Z}}_{u,t,b}$ .

2) *MU-MISO Transmission Phase*: In this phase, the RIS is configured to state  $s_t^*$  by solving  $\mathcal{P}(\mathbf{A})$  mentioned in Section II-C. The downlink received signal at user  $k$  becomes:

$$y_{k,t} = \mathbf{h}_{k,t}(s_t^*)^T \sum_{k=1}^K \mathbf{w}_{k,t} x_{k,t} + z_{k,t}, \quad (23)$$

where the precoder  $\mathbf{w}_{k,t}$  has been defined in (8).

##### B. Estimating and Predicting Instantaneous Channels with an Approximate Model for Dominant Rays

Recall the cascaded channel model in (6), reproduced here:

$$\mathbf{h}_{k,t}(\mathbf{s}) = \sum_{i,j} \mathbf{G}_t^{(i)} \text{diag} \left( \nu(\phi_{\text{in}}^{(i)}, \phi_{\text{out},k}^{(j)}; \mathbf{s}) \right) \mathbf{h}_{r,k,t}^{(j)}. \quad (24)$$

Exactly predicting  $\mathbf{h}_{k,t}(\mathbf{s})$  for any  $\mathbf{s}$  is difficult due to the non-separable state response and the numerous rays in wireless propagation. However, passive beamforming design can tolerate some prediction errors, allowing for an approximate model that enables tractable channel estimation.

1) *Approximate Model for BS-RIS Channel*: The BS-RIS channel model in Section II-B can be simplified for two reasons. First, the BS uses a conventional antenna array, and its response is independent of the RIS configuration. Second, the wireless propagation rays are clustered around the dominant scatterers in the environment. Therefore, we can represent the rays associated with the same scatterer using one or two dominant rays with a small approximation loss. The proposed approximate model is given by

$$\mathbf{G}_t = \sum_{m=1}^M \sum_{i=0}^I a_{t,m,i} \mathbf{v}_m \alpha_{\text{RIS}}^H(\bar{\phi}_{\text{in}}^{(i)}), \quad (25)$$

where  $\mathbf{v}_m$  is the  $m$ -th column of an orthogonal matrix  $\mathbf{V}$ , defined as

$$\mathbf{V} = \text{diag}(\alpha_{\text{BS}}(\phi_{\text{BS}}^{(0)})) (\text{DFT}(M_1) \otimes \text{DFT}(M_2)), \quad (26)$$

for a  $M_1 \times M_2$  UPA, with  $\mathbf{v}_1 = \alpha_{\text{BS}}(\phi_{\text{BS}}^{(0)})$ . Without loss of generality, we set  $a_{t,1,0} = 1$  for all  $t$ . In addition, since the LoS BS-RIS link is known, we have  $\bar{\phi}_{\text{in}}^{(0)} = \phi_{\text{in}}^{(0)}$ , and we assume  $\bar{\phi}_{\text{in}}^{(i)} \in \mathcal{D}$  for all  $i$ , where  $\mathcal{D}$  is a grid-based angle set.

2) *Approximate Model for RIS-UE Channels*: Similarly, the RIS-UE channels can be approximated by dominated rays with angles on the grid  $\mathcal{D}$ :

$$\mathbf{h}_{r,k,t} = \sum_{j=0}^J b_{k,t,j} \alpha_{\text{RIS}}(\bar{\phi}_{\text{out},k}^{(j)}), \quad (27)$$

where  $\bar{\phi}_{\text{out},k}^{(j)} \in \mathcal{D}$  for all  $j$  and  $k$ .

3) *Approximate Cascaded Channels*: Combining (25) and (27), the approximate cascaded channel is expressed as

$$\mathbf{h}_{k,t}(\mathbf{s}) = \sum_{m=1}^M \sum_{i=0}^I \sum_{j=0}^J a_{t,m,i} d_{k,t,j} \mathbf{v}_m \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\bar{\phi}_{\text{in}}^{(i)}) \text{diag}\left(\boldsymbol{\nu}(\bar{\phi}_{\text{in}}^{(i)}, \bar{\phi}_{\text{out},k}^{(j)}; \mathbf{s})\right) \boldsymbol{\alpha}_{\text{RIS}}(\bar{\phi}_{\text{out},k}^{(j)}). \quad (28)$$

To reconstruct  $\mathbf{h}_{k,t}(\mathbf{s})$ , we need to estimate the angles  $\bar{\phi}_{\text{in}}^{(i)}$  and  $\bar{\phi}_{\text{out},k}^{(j)}$  for the dominant rays, along with the time-varying channel responses  $a_{t,m,i}$  and  $d_{k,t,j}$ .

### C. Instantaneous Channel Estimation Given Knowledge of Angles

The cascaded channel in (28) is further written by

$$\mathbf{h}_{k,t}(\mathbf{s}_b) = \sum_{m=1}^M \sum_{i=0}^I \sum_{j=0}^J a_{t,m,i} d_{k,t,j} f_{b,i,j,k} \mathbf{v}_m, \quad (29)$$

where the parameters  $f_{b,m,i,j}$  is defined as

$$f_{b,i,j,k} = \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\bar{\phi}_{\text{in}}^{(i)}) \text{diag}\left(\boldsymbol{\nu}(\bar{\phi}_{\text{in}}^{(i)}, \bar{\phi}_{\text{out},k}^{(j)}; \mathbf{s}_b)\right) \boldsymbol{\alpha}_{\text{RIS}}(\bar{\phi}_{\text{out},k}^{(j)}), \quad (30)$$

We present a three-step method to estimate the instantaneous channel responses  $a_{t,m,i}$  and  $d_{k,t,j}$  given knowledge of all  $f_{b,i,j,k}$ .

- **Step 1** (Estimating LoS Link): Since the LoS ray is generally stronger than each NLoS ray, we first estimate  $d_{k,t,0}$  for all  $k$  given  $a_{t,1,0} = 1$  from observations  $\mathbf{Y}_{u,t,b}$  in (22) by treating all the NLoS components as interference:

$$\mathbf{y}_{u,t,b,k} = d_{k,t,0} f_{b,0,0,k} \mathbf{v}_0 + \mathbf{e}_{t,b,k} + \mathbf{z}_{u,t,b,k}, \quad (31)$$

where  $\mathbf{e}_{t,b,k}$  summarizes all the NLoS components. Then,  $d_{k,t,0}$  is estimated by the least squares (LS) method:

$$d_{k,t,0} = \frac{\sum_b f_{b,0,0,k}^* \mathbf{v}_0^{\text{H}} \mathbf{y}_{u,t,b,k}}{\sum_b |f_{b,0,0,k}|^2}. \quad (32)$$

- **Step 2** (Estimating All  $a_{t,m,i}$ ): Since  $\mathbf{V}$  is orthogonal matrix, we have the observation for the  $m$ -th angular direction at the BS:

$$\bar{\mathbf{y}}_{u,t,b,k,m} = \mathbf{v}_m^{\text{H}} \mathbf{y}_{u,t,b,k} = \sum_{i=0}^I a_{t,m,i} f_{b,i,0,k} + \bar{\mathbf{e}}_{t,b,k} + \bar{\mathbf{z}}_{u,m,t,b,k}, \quad (33)$$

where  $\bar{\mathbf{e}}_{t,b,k}$  summarizes all the components containing the RIS-UE NLoS links, and  $\bar{\mathbf{z}}_{u,m,t,b,k} = \mathbf{v}_m^{\text{H}} \mathbf{z}_{u,t,b,k}$ . Letting  $\bar{\mathbf{y}}_{u,t,m} = [\bar{y}_{u,t,1,1,m}, \bar{y}_{u,t,1,2,m}, \dots, \bar{y}_{u,t,B,K,m}]$ ,  $\mathbf{f}_{i,0} = [f_{1,i,0,1}, f_{1,i,0,2}, \dots, f_{B,i,0,K}]$ , and  $\mathbf{a}_{t,m} = [a_{t,m,0}, a_{t,m,1}, \dots, a_{t,m,I}]$ , we have estimation:

$$\mathbf{a}_{t,m} = \left(\bar{\mathbf{F}}_0^{\text{H}} \bar{\mathbf{F}}_0\right)^{-1} \bar{\mathbf{F}}_0^{\text{H}} \bar{\mathbf{y}}_{u,t,m}, \quad (34)$$

where  $\bar{\mathbf{F}}_0 = [\mathbf{f}_{1,0}, \mathbf{f}_{2,0}, \dots, \mathbf{f}_{I,0}]$ .

### Algorithm 1 Proposed Three-Step Estimation on the Instantaneous Channel

- 1: **Input**: Observation set  $\mathbf{Y}_{u,t,b}$  for all  $b$  and the angles;
- 2: Estimate  $d_{k,t,0}$  by (32) for LoS;
- 3: Estimate  $\mathbf{a}_{t,m}$  for all  $m$  by (34);
- 4: Estimate  $\mathbf{d}_{k,t}$  for all  $k$  by (37);
- 5: **Output**: Predicting  $\mathbf{h}_{k,t}(\mathbf{s})$  based on (29).

- **Step 3** (Estimating All  $d_{k,t,j}$ ): Since all BS-RIS parameters  $a_{t,m,i}$  have been estimated, we define

$$\tilde{\mathbf{v}}_{k,t,j,b} = \sum_{m=1}^M \sum_{i=0}^I a_{t,m,i} f_{b,i,j,k} \mathbf{v}_m, \quad (35)$$

and the received pilot observation is written by

$$\begin{aligned} \mathbf{y}_{u,t,b,k} &= \sum_{j=0}^J \tilde{\mathbf{v}}_{k,t,j,b} d_{k,t,j} + \mathbf{z}_{u,t,b,k} \\ &= \tilde{\mathbf{V}}_{k,t,b} \mathbf{d}_{k,t} + \mathbf{z}_{u,t,b,k}, \end{aligned} \quad (36)$$

where  $\mathbf{d}_{k,t} = [d_{k,t,0}, d_{k,t,1}, \dots, d_{k,t,J}]^{\text{T}}$  and  $\tilde{\mathbf{V}}_{k,t,b} = [\tilde{\mathbf{v}}_{k,t,0,b}, \tilde{\mathbf{v}}_{k,t,1,b}, \dots, \tilde{\mathbf{v}}_{k,t,J,b}]$ . Then, we finally have estimation:

$$\mathbf{d}_{k,t} = \left( \sum_{b=1}^B \tilde{\mathbf{V}}_{k,t,b}^{\text{H}} \tilde{\mathbf{V}}_{k,t,b} \right)^{-1} \left( \sum_{b=1}^B \tilde{\mathbf{V}}_{k,t,b}^{\text{H}} \mathbf{y}_{u,t,b,k} \right). \quad (37)$$

We summarize the estimation method in Algorithm 1. The complexity of Algorithm 1 is  $\mathcal{O}(MI^3 + MI^2BK + KJ^3 + MBKJ^2)$ , which is independent of the angle set size  $|\mathcal{D}|$ . Additionally, to ensure the feasibility of the matrix inverse operations in (34) and (37), we require  $I < BK$  and  $J < BM$  for the proposed approximate model in (28).

### D. Statistical Knowledge Estimation over $T$ Historical Observation Blocks

In this subsection, we focus on estimating the angles  $\bar{\phi}_{\text{in}}^{(i)}$  and  $\bar{\phi}_{\text{out},k}^{(j)}$  for the dominant rays using historical observations  $\mathcal{Y}_t = \{\mathbf{Y}_{u,t,b}, t = 1, 2, \dots, T\}$ . Similar to instantaneous channel estimation, we accomplish this through a three-step method:

- **Step 1** (Estimating LoS Angles  $\bar{\phi}_{\text{out},k}^{(0)}$ ): Since  $\bar{\phi}_{\text{in}}^{(0)}$  is known, we can search the strongest angle in the grid set  $\mathcal{D}$  as the estimation of  $\bar{\phi}_{\text{out},k}^{(0)}$ . Specifically, assuming  $\bar{\phi}_{\text{out},k}^{(0)} = \phi$ , the LoS link is given by

$$\mathbf{h}_{k,t}^{(0)}(\mathbf{s}_b, \phi) = d_{k,t,0}(\phi) \mathbf{v}_0 \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\bar{\phi}_{\text{in}}^{(0)}) \text{diag}\left(\boldsymbol{\nu}(\bar{\phi}_{\text{in}}^{(0)}, \phi; \mathbf{s}_b)\right) \boldsymbol{\alpha}_{\text{RIS}}(\phi), \quad (38)$$

and  $d_{k,t,0}(\phi)$  is defined based on (32):

$$d_{k,t,0}(\phi) = \frac{\sum_b \mathbf{h}_{k,t}^{(0)}(\mathbf{s}_b, \phi)^{\text{H}} \mathbf{y}_{u,t,b,k}}{\sum_b \mathbf{h}_{k,t}^{(0)}(\mathbf{s}_b, \phi)^{\text{H}} \mathbf{h}_{k,t}^{(0)}(\mathbf{s}_b, \phi)}. \quad (39)$$

Then,  $\bar{\phi}_{\text{out},k}^{(0)}$  is estimated by solving:

$$\mathcal{P}(\text{C-1}) \bar{\phi}_{\text{out},k}^{(0)} = \underset{\phi \in \mathcal{D}}{\text{argmin}} \sum_{t,b} \|\mathbf{y}_{u,t,b,k} - \mathbf{h}_{k,t}^{(0)}(\mathbf{s}_b, \phi)\|_2^2.$$

- **Step 2** (Estimating BS-RIS NLoS Angles  $\bar{\phi}_{\text{in}}^{(i)}$ ): We initialize the BS-RIS angle set  $\mathcal{A}_{\text{in}} = \{\bar{\phi}_{\text{in}}^{(0)}\}$ . Given the RIS-UE LoS angles  $\bar{\phi}_{\text{out},k}^{(0)}$  and the response  $d_{k,t,0}$ , we assume the incident angle of the largest NLoS ray is  $\bar{\phi}_{\text{in}}^{(i)} = \phi$ , and the cascaded channel for this ray pair can be expressed as follows

$$\begin{aligned} \mathbf{h}_{k,t}^{(1)}(\mathbf{s}_b, \phi) &= \sum_{m=1}^M a_{t,m}(\phi) d_{k,t,0} \mathbf{v}_m \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\phi) \\ &\quad \text{diag}\left(\boldsymbol{\nu}(\phi, \bar{\phi}_{\text{out},k}^{(0)}; \mathbf{s}_b)\right) \boldsymbol{\alpha}_{\text{RIS}}(\bar{\phi}_{\text{out},k}^{(0)}) \\ &= \sum_{m=1}^M a_{t,m}(\phi) \mathbf{v}_m \bar{f}_{b,k}(\phi), \end{aligned} \quad (41)$$

where  $\bar{f}_{b,k}(\phi)$  is defined as follows

$$\bar{f}_{b,k,t}(\phi) = d_{k,t,0} \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\phi) \text{diag}\left(\boldsymbol{\nu}(\phi, \bar{\phi}_{\text{out},k}^{(0)}; \mathbf{s}_b)\right) \boldsymbol{\alpha}_{\text{RIS}}(\bar{\phi}_{\text{out},k}^{(0)}). \quad (42)$$

Then, based on the estimator in (32), (33) and (34),  $a_{t,m}(\phi)$  in (41) can be estimated by

$$a_{t,m}(\phi) = \frac{\sum_{b,k} \bar{f}_{b,k,t}(\phi) \star \mathbf{v}_m^{\text{H}} \tilde{\mathbf{y}}_{u,t,b,k}}{\sum_{b,k} |\bar{f}_{b,k,t}|^2}, \quad (43)$$

where  $\tilde{\mathbf{y}}_{u,t,b,k}$  is initialized by

$$\tilde{\mathbf{y}}_{u,t,b,k} = \mathbf{y}_{u,t,b,k} - \mathbf{h}_{k,t}^{(1)}(\mathbf{s}_b, \bar{\phi}_{\text{in}}^{(0)}). \quad (44)$$

The strongest incident angle  $\bar{\phi}_{\text{in}}^{(i)}$  for  $\tilde{\mathbf{y}}_{u,t,b,k}$  is estimated by solving:

$$\mathcal{P}(\text{C-2}) \bar{\phi}_{\text{in}}^{(i)} = \underset{\phi \in \mathcal{D}}{\text{argmin}} \sum_{t,b,k} \|\tilde{\mathbf{y}}_{u,t,b,k} - \mathbf{h}_{k,t}^{(1)}(\mathbf{s}_b, \phi)\|_2^2,$$

and we update  $\mathcal{A}_{\text{in}} = \mathcal{A}_{\text{in}} \cup \{\bar{\phi}_{\text{in}}^{(i)}\}$ . We next update the observation by

$$\tilde{\mathbf{y}}_{u,t,b,k} = \tilde{\mathbf{y}}_{u,t,b,k} - \tilde{\mathbf{h}}_{k,t}^{(1)}(\mathbf{s}_b; \mathcal{A}_{\text{in}}), \quad (46)$$

to search for the strongest incident angle that is not in  $\mathcal{A}_{\text{in}}$ , where  $\tilde{\mathbf{h}}_{k,t}^{(1)}(\mathbf{s}_b; \mathcal{A}_{\text{in}})$  is defined by

$$\tilde{\mathbf{h}}_{k,t}^{(1)}(\mathbf{s}_b; \mathcal{A}_{\text{in}}) = \sum_{m=1}^M \sum_{i=0}^{|\mathcal{A}_{\text{in}}|-1} a_{t,m,i} d_{k,t,0} f_{b,i,0,k} \mathbf{v}_m, \quad (47)$$

whose parameters  $\{a_{t,m,i}\}$  are calculated by (34).

- **Step 3** (Estimating RIS-UE Angles  $\bar{\phi}_{\text{out},k}^{(j)}$ ): We estimate the reflection angle  $\bar{\phi}_{\text{out},k}^{(j)}$  in this step given  $\mathcal{A}_{\text{in}}$  and response parameters  $\{a_{t,m,i}\}$ . We initialize the RIS-UE angle set by  $\mathcal{A}_{\text{out},k} = \emptyset$ , and assume the strongest angle

---

### Algorithm 2 Algorithm for Channel Statistical Knowledge Estimation

---

- 1: **Input:** Observation  $\mathcal{Y}_t$ ;
  - 2: Estimate LoS RIS-UE angles  $\bar{\phi}_{\text{out},k}^{(0)}$  by solving  $\mathcal{P}(\text{C-1})$ ;
  - 3: Initialize  $\tilde{\mathbf{y}}_{u,t,b,k}$  by (44) and  $\mathcal{A}_{\text{in}} = \{\bar{\phi}_{\text{in}}^{(0)}\}$ ;
  - 4: **while**  $\mathbb{E}_{t,b,k} \|\tilde{\mathbf{y}}_{u,t,b,k}\|_2^2 > M\sigma_u^2$  and  $|\mathcal{A}_{\text{in}}| < BK$  **do**
  - 5:     Search  $\bar{\phi}_{\text{in}}^{(i)}$  by solving  $\mathcal{P}(\text{C-2})$ ;
  - 6:     Update  $\mathcal{A}_{\text{in}} = \mathcal{A}_{\text{in}} \cup \{\bar{\phi}_{\text{in}}^{(i)}\}$ ;
  - 7:     Update  $\tilde{\mathbf{y}}_{u,t,b,k}$  (46);
  - 8: **end while**
  - 9: Initialize  $\tilde{\mathbf{y}}_{u,t,b,k} = \mathbf{y}_{u,t,b,k}$  and  $\mathcal{A}_{\text{out},k} = \emptyset$ ;
  - 10: **for**  $k = 1 : K$  **do**
  - 11:     **while**  $\mathbb{E}_{t,b,k} \|\tilde{\mathbf{y}}_{u,t,b,k}\|_2^2 > M\sigma_u^2$  and  $|\mathcal{A}_{\text{out},k}| < BM$  **do**
  - 12:         Search  $\bar{\phi}_{\text{out},k}^{(j)}$  by solving  $\mathcal{P}(\text{C-3})$ ;
  - 13:         Update  $\mathcal{A}_{\text{out},k} = \mathcal{A}_{\text{in}} \cup \{\bar{\phi}_{\text{out},k}^{(j)}\}$ ;
  - 14:         Update  $\tilde{\mathbf{y}}_{u,t,b,k}$  (52);
  - 15:     **end while**
  - 16: **end for**
  - 17: **Output:** The estimated angles sets  $\mathcal{A}_{\text{in}}$  and  $\mathcal{A}_{\text{out},k}$  for the dominant rays.
- 

is  $\bar{\phi}_{\text{out},k}^{(j)} = \phi$ . Then, the cascaded channel related to this angle is given by

$$\begin{aligned} \mathbf{h}_{k,t}^{(2)}(\mathbf{s}_b, \phi) &= \sum_{m=1}^M \sum_{i=0}^I a_{t,m,i} d_{k,t}(\phi) \mathbf{v}_m \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\bar{\phi}_{\text{in}}^{(i)}) \\ &\quad \text{diag}\left(\boldsymbol{\nu}(\bar{\phi}_{\text{in}}^{(i)}, \phi; \mathbf{s}_b)\right) \boldsymbol{\alpha}_{\text{RIS}}(\phi) \\ &= d_{k,t}(\phi) \tilde{\mathbf{f}}_{t,b}, \end{aligned} \quad (48)$$

where  $\tilde{\mathbf{f}}_{t,b}$  is defined by

$$\begin{aligned} \tilde{\mathbf{f}}_{t,b} &= \sum_{m=1}^M \sum_{i=0}^I a_{t,m,i} \mathbf{v}_m \boldsymbol{\alpha}_{\text{RIS}}^{\text{H}}(\bar{\phi}_{\text{in}}^{(i)}) \\ &\quad \text{diag}\left(\boldsymbol{\nu}(\bar{\phi}_{\text{in}}^{(i)}, \phi; \mathbf{s}_b)\right) \boldsymbol{\alpha}_{\text{RIS}}(\phi), \end{aligned} \quad (49)$$

and  $d_{k,t}(\phi)$  is given by

$$d_{k,t}(\phi) = \frac{\sum_b \tilde{\mathbf{f}}_{t,b}^{\text{H}} \tilde{\mathbf{y}}_{u,t,b,k}}{\sum_b \tilde{\mathbf{f}}_{t,b}^{\text{H}} \tilde{\mathbf{f}}_{t,b}}, \quad (50)$$

with  $\tilde{\mathbf{y}}_{u,t,b,k}$  initialized by  $\mathbf{y}_{u,t,b,k}$ . Then, we solve  $\bar{\phi}_{\text{out},k}^{(j)}$  by:

$$\mathcal{P}(\text{C-3}) \bar{\phi}_{\text{out},k}^{(j)} = \underset{\phi \in \mathcal{D}}{\text{argmin}} \sum_{t,b} \|\tilde{\mathbf{y}}_{u,t,b,k} - \mathbf{h}_{k,t}^{(2)}(\mathbf{s}_b, \phi)\|_2^2,$$

and we update  $\mathcal{A}_{\text{out},k} = \mathcal{A}_{\text{out},k} \cup \{\bar{\phi}_{\text{out},k}^{(j)}\}$ . We next update the observation by

$$\tilde{\mathbf{y}}_{u,t,b,k} = \tilde{\mathbf{y}}_{u,t,b,k} - \tilde{\mathbf{h}}_{k,t}^{(2)}(\mathbf{s}_b; \mathcal{A}_{\text{out},k}), \quad (52)$$

to search for the strongest reflection angle that is not in  $\mathcal{A}_{\text{out},k}$ , where  $\tilde{\mathbf{h}}_{k,t}^{(2)}(\mathbf{s}_b)$  is defined by

$$\tilde{\mathbf{h}}_{k,t}^{(2)}(\mathbf{s}_b; \mathcal{A}_{\text{out},k}) = \sum_{m,i} \sum_{j=0}^{|\mathcal{A}_{\text{out},k}|-1} a_{t,m,i} d_{k,t,j} f_{b,i,j,k} \mathbf{v}_m, \quad (53)$$

whose parameters  $\{d_{k,t,j}\}$  are calculated by (37).

We summarize the estimation method in Algorithm 2. The complexity of Algorithm 2 is  $\mathcal{O}(BKT M|\mathcal{D}| + MI^3J + MI^2BKJ + IKJ^3 + MIBKJ^2)$ , which is linear in the size of the angle set  $|\mathcal{D}|$ . While a large  $|\mathcal{D}|$  increases complexity, the computation time of Algorithm 2 does not contribute to the processing delay of instantaneous channel estimation and beamforming design. Therefore, Algorithm 2 can be executed with low priority when there are no urgent tasks for the CPU.

## V. PASSIVE BEAMFORMING ALGORITHM

Recall that the passive beamforming problem formulation to maximize the summation of achievable rate:

$$\mathcal{P}(\text{A}) \quad \mathbf{s}_t^* = \underset{\mathbf{s}}{\operatorname{argmax}} \sum_{K=1}^K R_{k,t}(\mathbf{s})$$

$$\text{s.t. } s_n \in \{1, 2, \dots, L\}, \forall n, \quad (54a)$$

$$\frac{1}{K} \sum_k \|\mathbf{w}_{k,t}(\mathbf{s})\|_2^2 \leq P_T, \quad (54b)$$

where the channel and the achievable rate are given by

$$\mathbf{h}_{k,t}(\mathbf{s}) = \sum_{i,j} \mathbf{G}_t^{(i)} \operatorname{diag} \left( \boldsymbol{\nu}(\phi_{\text{in}}^{(i)}, \phi_{\text{out},k}^{(j)}; \mathbf{s}) \right) \mathbf{h}_{r,k,t}^{(j)}, \quad (55)$$

$$R_{k,t}(\mathbf{s}) = \log \left( 1 + \frac{|\mathbf{h}_{k,t}(\mathbf{s})^T \mathbf{w}_{k,t}(\mathbf{s})|^2}{\sum_{i \neq k} |\mathbf{h}_{k,t}(\mathbf{s})^T \mathbf{w}_{i,t}(\mathbf{s})|^2 + \sigma_z^2} \right). \quad (56)$$

In this section, we propose low-complexity algorithm to address the discrete constraint in (54a).

### A. Continuous Approximation for Discrete Variables

We propose tractable approximation to address the discrete constraint on  $s_n$  in  $\mathcal{P}(\text{A})$  by a three-step method.

1) *Reparameterize  $s_{n,t}$  by One-Hot Vectors*: We introduce a one-hot vector  $\bar{\mathbf{s}}_n = [\bar{s}_{n,1}, \bar{s}_{n,2}, \dots, \bar{s}_{n,L}]^T$  for the state of the  $n$ -th RIS element, where  $\bar{s}_{n,\ell} \in \{0, 1\}$  and  $\sum_{\ell} \bar{s}_{n,\ell} = 1$ . Then, the RIS response of the  $n$ -th element for the  $i$ -th incident ray and  $j$ -th reflection ray can be equivalently denoted by a function of  $\bar{\mathbf{s}}_n$ :

$$\boldsymbol{\nu}(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; s_n) = \sum_{\ell=1}^L \bar{s}_{n,\ell} \boldsymbol{\nu}(\phi_{\text{in}}^{(i)}, \phi_{\text{out}}^{(j)}; \ell). \quad (57)$$

Letting  $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \dots, \bar{\mathbf{s}}_N]$  and substituting (57) into (55), the channel  $\mathbf{h}_{k,t}(\mathbf{s})$  is equivalently represented by  $\mathbf{h}_{k,t}(\bar{\mathbf{S}})$ . Consequently, the rate and precoder become  $R_{k,t}(\bar{\mathbf{S}})$  and  $\mathbf{w}_{k,t}(\bar{\mathbf{S}})$ , respectively. Then,  $\mathcal{P}(\text{A})$  is equivalently written by

$$\mathcal{P}(\text{A-1}) \quad \max_{\bar{\mathbf{S}}} \sum_{K=1}^K R_{k,t}(\bar{\mathbf{S}})$$

$$\text{s.t. } \sum_{\ell} \bar{s}_{n,\ell} = 1, \forall n, \quad (58a)$$

$$\bar{s}_{n,\ell} \in \{0, 1\}, \forall \ell, n, \quad (58b)$$

$$\frac{1}{K} \sum_k \|\mathbf{w}_{k,t}(\bar{\mathbf{S}})\|_2^2 \leq P_T. \quad (58c)$$

2) *Continuous Relaxation with Sparse Promoting Constraint*: However, the binary constraint (58b) is still challenging. To address this issue, we relax the constraints in (58a) and (58b) as follows:

$$\sum_{\ell} \sqrt{\bar{s}_{n,\ell}} = 1, \forall n, \quad (59a)$$

$$0 \leq \bar{s}_{n,\ell} \leq 1, \forall \ell, n. \quad (59b)$$

The key intuition is that we prefer large  $\bar{s}_{n,\ell}$  to enhance the RIS response according to (57). However,  $\bar{s}_{n,\ell} \leq \sqrt{\bar{s}_{n,\ell}}$  for  $0 \leq \bar{s}_{n,\ell} \leq 1$ , with equality achieved only when  $\bar{s}_{n,\ell} \in \{0, 1\}$ . Therefore, constraint (59a) promotes the sparsity of  $\bar{\mathbf{s}}_n$ , encouraging  $\bar{s}_{n,\ell}$  to take values close to 0 or 1. Thus we have

$$\mathcal{P}(\text{A-2}) \quad \max_{\bar{\mathbf{S}}} \sum_{K=1}^K R_{k,t}(\bar{\mathbf{S}})$$

$$\text{s.t. } (58c), (59a), (59b).$$

### 3) Equivalent Transformation to Unconstrained Problem:

In practice, an unconstrained problem is preferred, as it can be easily tackled using stochastic gradient descent (SGD)-based algorithms, which are well-implemented in PyTorch functions [38]. To this end, we introduce auxiliary variables  $\tilde{\mathbf{S}} = [\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_N]$  with  $\tilde{\mathbf{s}}_n = [\tilde{s}_{n,1}, \tilde{s}_{n,2}, \dots, \tilde{s}_{n,L}]^T \in \mathbb{R}^L$ . Then,  $\bar{\mathbf{s}}_n$  is represented by  $\tilde{\mathbf{s}}_n$  using softmax function:

$$\bar{\mathbf{s}}_n = (\operatorname{Softmax}(\tilde{\mathbf{s}}_n))^2, \quad (61)$$

which has incorporated constraints (59a) and (59b) implicitly. In addition, we set  $\gamma(\bar{\mathbf{S}})$  in (8) as follows for the RZF precoder:

$$\gamma(\tilde{\mathbf{S}}) = \frac{KP_T}{\left\| \mathbf{H}_t^*(\tilde{\mathbf{S}}) \left( \mathbf{H}_t^T(\tilde{\mathbf{S}}) \mathbf{H}_t^*(\tilde{\mathbf{S}}) + \varsigma \mathbf{I}_K \right)^{-1} \right\|_F^2}, \quad (62)$$

to incorporate the power constraint in (58c). Finally, we have an unconstrained optimization problem

$$\mathcal{P}(\text{A-3}) \quad \max_{\tilde{\mathbf{S}}} \sum_{K=1}^K R_{k,t}(\tilde{\mathbf{S}})$$

We solve  $\mathcal{P}(\text{A-3})$  by utilizing Adam optimizer [39] in PyTorch with learning rate 0.1.

## VI. SIMULATION

### A. Simulation Setup

We examine an indoor femtocell network operating at 2.4 GHz with a bandwidth of 60 kHz, as shown in Fig. 9. Users are randomly distributed in a 10 m  $\times$  8 m area, served by one BS and one RIS. Channel coefficients are generated using the 3GPP ray-tracing model [37, Section 7.5] with parameters for the Indoor-Office scenario [37, Table 7.5-6], and the LoS channel condition is set with a K-factor of 3 dB. The RIS is a 8  $\times$  8 uniform planar array with  $N = 64$  elements. Each element experiences an average reflection loss of approximately 1.5 dB, primarily due to the RF switches. This performance is 3 dB better than that of a reflective antenna of the same size, as it also benefits from structural scattering. To evaluate the advantages of the pixel RIS, we

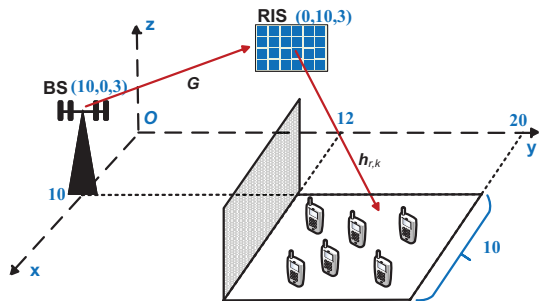


Fig. 9. The simulated RIS-aided  $K$ -user MISO communication scenario comprising of one  $M$ -antenna BS and one  $N$ -element IRS.

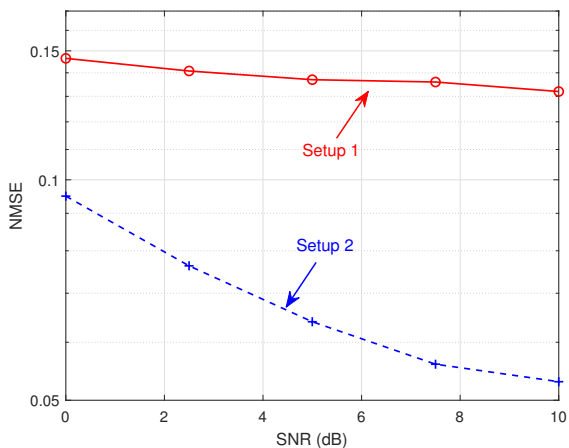


Fig. 10. SNR versus NMSE for different setups.

investigate two setups varying the sizes of the BS array and the number of users:

- **Setup 1:** BS uses a  $4 \times 1$  ULA to serve  $K = 2$  users for spatial multiplexing.
- **Setup 2:** BS uses a  $4 \times 2$  UPA to serve  $K = 4$  users for spatial multiplexing.

For channel estimation, we set  $T = 20$  and  $B = 16$ . We generate 100 snapshots with random locations, and 40 random channel realizations are generated to verify the channel estimation accuracy and the final achievable rate. We set  $P_T = \sigma_z^2 = \sigma_u^2 = 1$ , and the average signal-to-noise ratio (SNR) for wireless propagation is defined by:

$$\text{SNR} \triangleq \mathbb{E}_{t,k} \left[ \frac{\|\mathbf{G}_t\|_F^2 \|\mathbf{h}_{r,k,t}\|_2^2}{MN^2} \right]. \quad (64)$$

To evaluate the effectiveness of the proposed solution, we consider the following baselines:

- **Random:** Each element selects a random state. This baseline does not require channel estimation for RIS configuration.
- **Best Sounding:** This scheme selects the best state among the  $B = 16$  RIS state configurations during channel sounding.
- **Group-Opt [40], [41]:** In this scheme, all the elements select the same state, and the optimal state is selected by exhaustive search given perfect CSI.

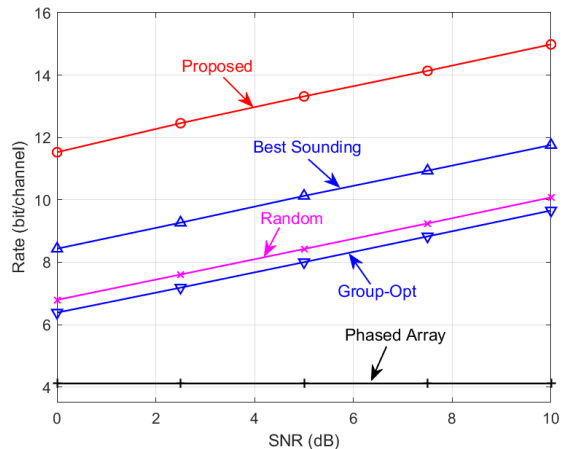


Fig. 11. SNR versus achievable rate for downlink MU-MISO under setup 1 ( $M = 4$ ,  $N = 64$ , and  $K = 2$ ).

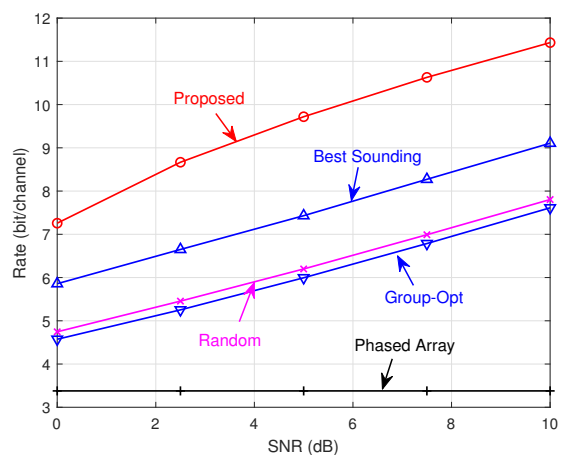


Fig. 12. SNR versus achievable rate for downlink MU-MISO under setup 2 ( $M = 8$ ,  $N = 64$ , and  $K = 4$ ).

- **Phased Array:** This scheme employs traditional RIS with the same size, with phase shifters optimized to maximize the sum rate under perfect CSI assumption, while the BS employs the RZF precoder shown in (8).

## B. Simulation Results

We first evaluate the effectiveness of the proposed channel estimation algorithm in Fig. 10. The estimation NMSE for each wireless realization is assessed using 100 random RIS configurations not observed during the channel sounding phases. The proposed algorithm performs well in both setups, achieving an NMSE below 0.15, with notably better performance in setup 2. This improvement is due to the increased number of BS antennas  $M$  and users  $K$ , which enhances the estimation of dominated rays, as discussed in Section IV-C.

Next, we compare the achievable rates of the proposed solution with baselines in setup 1 (Fig. 11) and setup 2 (Fig. 12). The phased array performs poorly due to the strong BS-RIS LoS link and high channel correlation among multiple users, leading to interference as a key bottleneck. Consequently, the

achievable rate shows minimal improvement with increasing SNR. This is also evidenced by the decrease in achievable rate with more users for spatial multiplexing, despite a proportional increase in BS antennas, as shown in the comparison of the two figures. In contrast, the pixel RIS outperforms the phased array even with random configurations, as its elements can reflect an incident ray in two focused directions (as shown in Fig. 5), which potentially enhances the multiuser channel condition. Furthermore, our proposed solution achieves significant performance gains in all cases, with an SNR improvement of 8 dB in setup 1 and 6 dB in setup 2 compared to the best baseline.

## VII. CONCLUSION

In this paper, we proposed a passive beamforming and channel estimation solution for pixel-based RIS with a non-separable state response. We introduced a DNN to model the RIS response function while ensuring low computational and memory requirements. We then presented a tractable channel estimation algorithm that focuses on the channel responses of dominated scattering rays, with statistical knowledge estimated from historical observations. Finally, we implemented a low-complexity passive beamforming algorithm using a novel continuous approximation to configure the discrete RIS states. Simulations showed that the proposed solution achieved significant gains over various baselines.

## REFERENCES

- [1] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. ElKashlan, M. Chen, M. D. Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, 2021.
- [2] J. Rao, Y. Zhang, S. Tang, Z. Li, C.-Y. Chiu, and R. Murch, "An active reconfigurable intelligent surface utilizing phase-reconfigurable reflection amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 7, pp. 3189–3202, 2023.
- [3] H.-T. Chen, A. J. Taylor, and N. Yu, "A review of metasurfaces: physics and applications," *Rep. Progr. Phys.*, vol. 79, no. 7, p. 076401, 2016.
- [4] J. Rao, Y. Zhang, S. Tang, Z. Li, S. Shen, C. Y. Chiu, and R. Murch, "A novel reconfigurable intelligent surface for wide-angle passive beamforming," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 12, pp. 5427–5439, 2022.
- [5] J. C. Liang, Q. Cheng, Y. Gao, C. Xiao, S. Gao, L. Zhang, S. Jin, and T. J. Cui, "An angle-insensitive 3-bit reconfigurable intelligent surface," *IEEE Trans. Antennas Propag.*, vol. 70, no. 10, pp. 8798–8808, 2022.
- [6] Z. Zhang, J. W. Zhang, J. W. Wu, J. C. Liang, Z. X. Wang, Q. Cheng, Q. S. Cheng, T. J. Cui, H. Q. Yang, G. B. Liu, and S. R. Wang, "Macromodeling of reconfigurable intelligent surface based on microwave network theory," *IEEE Trans. Antennas Propag.*, vol. 70, no. 10, pp. 8707–8717, 2022.
- [7] J. Zhang, J. Xi, P. Li, R. C. C. Cheung, A. M. H. Wong, and J. Li, "Experiment-based deep learning approach for power allocation with a programmable metasurface," *APL Machine Learning*, vol. 1, no. 4, 2023.
- [8] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.
- [9] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy efficiency and spectral efficiency tradeoff in ris-aided multiuser mimo uplink transmission," *IEEE Trans. Signal Process.*, vol. 69, pp. 1407–1421, 2020.
- [10] H. Guo, Y. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, 2020.
- [11] T. Bai, C. Pan, Y. Deng, M. ElKashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, 2020.
- [12] S. Mao, N. Zhang, L. Liu, J. Wu, M. Dong, K. Ota, T. Liu, and D. Wu, "Computation rate maximization for intelligent reflecting surface enhanced wireless powered mobile edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10820–10831, 2021.
- [13] S. Mao, L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, and V. C. Leung, "Reconfigurable intelligent surface-assisted secure mobile edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6647–6660, 2022.
- [14] M. Hua, L. Yang, Q. Wu, C. Pan, C. Li, and A. L. Swindlehurst, "UAV-assisted intelligent reflecting surface symbiotic radio system," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5769–5785, 2021.
- [15] Y. Ge, J. Fan, G. Y. Li, and L.-C. Wang, "Intelligent reflecting surface-enhanced UAV communications: Advances, challenges, and prospects," *IEEE Wireless Commun.*, vol. 30, no. 6, pp. 119–126, 2023.
- [16] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent reflecting surface enhanced multi-UAV NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3051–3066, 2021.
- [17] 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, 2020.
- [18] H. Guo and V. K. N. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 3964–3977, 2022.
- [19] S. Noh, H. Yu, and Y. Sung, "Training signal design for sparse channel estimation in intelligent reflecting surface-assisted millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2399–2413, 2022.
- [20] H. Liu, X. Yuan, and Y. J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, 2020.
- [21] L. Zhou, J. Dai, and W. Xu, "Joint multi-user channel estimation for RIS-assisted massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13993–14006, 2024.
- [22] M. Nakul, A. Rajorija, and R. Budhiraja, "Variational learning algorithms for channel estimation in RIS-assisted mmWave systems," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 222–238, 2024.
- [23] 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, 2023.
- [24] S. J. Lee, "Effect of least square channel estimation errors on achievable rate in MIMO fading channels," *IEEE Commun. Lett.*, vol. 11, no. 11, pp. 862–864, 2007.
- [25] L. Huang, J. W. Bergmans, and F. M. Willems, "Low-complexity LMMSE-based MIMO-OFDM channel estimation via angle-domain processing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5668–5680, 2007.
- [26] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, 2010.
- [27] J. Dai, A. Liu, and V. K. N. Lau, "Joint channel estimation and user grouping for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 622–637, 2019.
- [28] S. Gao, X. Cheng, and L. Yang, "Estimating doubly-selective channels for hybrid mmWave massive MIMO systems: A doubly-sparse approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5703–5715, 2020.
- [29] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts," *IEEE Trans. Commun.*, pp. 1–1, 2021.
- [30] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Low-complexity antenna selection and discrete phase-shifts design in IRS-assisted multiuser massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 3980–3994, 2022.
- [31] H. Gao, K. Cui, C. Huang, and C. Yuen, "Robust beamforming for RIS-assisted wireless communications with discrete phase shifts," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2619–2623, 2021.
- [32] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, 2020.

- [33] K. Ying and Z. Gao, "Precoding design of reconfigurable massive mimo in the electromagnetic domain," in *Proc. IEEE GLOBECOM*, 2023, pp. 5769–5774.
- [34] K. Ying, Z. Gao, Y. Su, T. Qin, M. Matthaiou, and R. Schober, "Reconfigurable massive MIMO: Precoding design and channel estimation in the electromagnetic domain," *IEEE Trans. Commun.*, vol. 73, no. 5, pp. 3423–3440, 2025.
- [35] R. F. Harrington, "Theory of loaded scatterers," in *Proc. Inst. Elect. Eng.*, vol. 111, no. 4. IET, 1964, pp. 617–623.
- [36] R. Hansen, "Relationships between antennas as scatterers and as radiators," *Proc. IEEE*, vol. 77, no. 5, pp. 659–662, 1989.
- [37] *Study on channel model for frequencies from 0.5 to 100 GHz*. 3GPP TS 38.901 version 18.0.0 (Release 18), Apr. 2024.
- [38] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.
- [39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [40] F. Armandoust, E. Tohidi, M. Kasparick, L. Wang, A. H. Gokceoglu, and S. Stanczak, "MIMO systems with reconfigurable antennas: Joint channel estimation and mode selection," in *n Proc. WSA SCC 26th Int. ITG Workshop Smart Antennas 13th Conf. Syst. Commun., Coding*, Feb. 2023, pp. 1–6.
- [41] M. Hasan, I. Bahceci, and B. A. Cetiner, "Downlink multi-user MIMO transmission for radiation pattern reconfigurable antenna systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6448–6463, 2018.