

# Decision-Directed Hybrid RIS Channel Estimation With Minimal Pilot Overhead

Ly V. Nguyen<sup>ID</sup>, *Member, IEEE*, and A. Lee Swindlehurst<sup>ID</sup>, *Fellow, IEEE*

**Abstract**—To reap the benefits of reconfigurable intelligent surfaces (RIS), channel state information (CSI) is generally required. However, CSI acquisition in RIS systems is challenging and often results in very large pilot overhead, especially in unstructured channel environments. Consequently, the RIS channel estimation problem has attracted a lot of interest and also been a subject of intense study in recent years. In this paper, we propose a decision-directed RIS channel estimation framework for general unstructured channel models. The employed RIS contains some hybrid elements that can simultaneously reflect and sense the incoming signal. We show that with the help of the hybrid RIS elements, it is possible to accurately recover the CSI with a pilot overhead proportional to the number of users. Therefore, the proposed framework substantially improves the system spectral efficiency compared to systems with passive RIS arrays since the pilot overhead in passive RIS systems is proportional to the number of RIS elements times the number of users. We also perform a detailed spectral efficiency analysis for both the pilot-directed and decision-directed frameworks. Our analysis takes into account both the channel estimation and data detection errors at both the RIS and the BS. Finally, we present numerous simulation results to verify the accuracy of the analysis as well as to show the benefits of the proposed decision-directed framework.

**Index Terms**—Reconfigurable intelligent surfaces, channel estimation, sensing, decision-directed, spectral efficiency analysis.

## I. INTRODUCTION

RECONFIGURABLE intelligent surfaces (RIS) are a novel technology that has changed the conventional long-standing perspective that wireless channels are an uncontrollable part of the environment. RISs are planar arrays composed of elements whose electromagnetic reflection coefficients can be adaptively configured to shape the wireless channel in beneficial ways. As such, they can be deployed to improve the system throughput, network coverage, or energy efficiency [1], [2]. However, the exploitation of this channel-shaping ability generally requires RIS-related channel state information (CSI), which is challenging to obtain since the number of RIS elements can be very large, and the RIS elements are often constructed as passive devices without active radio-frequency (RF) chains or computational resources. Therefore, the RIS channel estimation problem has been a

subject of intense study in the last few years [3]. The literature of RIS channel estimation can be divided into two categories including structured and unstructured channel estimations. While structured channel estimation considers models that are parameterized by the angles of arrival (AoAs), angles of departure (AoDs), and complex gains of the propagation paths, unstructured channel estimation methods assume more generic channels described by arbitrary complex coefficients.

Numerous results on structured RIS channel estimation have been reported, for example in [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], and [18], where the sparsity property of high-frequency (e.g., millimeter-wave, or “mmWave”) channels are exploited to reduce the pilot overhead. For example, the studies in [4], [5], [6], [7], [8], [9], [10], [11], and [12] formulated the cascaded mmWave channel estimation problem as a sparse signal recovery problem so that various compressive sensing techniques can be exploited to recover the channel parameters, e.g., distributed orthogonal matching pursuit (OMP) [4], iterative atom pruning based subspace pursuit (IAP-SP) [5], atomic norm minimization [7], Newtonized orthogonal matching pursuit [8], alternating direction method of multipliers (ADMM) [10], and the hybrid multi-objective evolutionary paradigm [12]. Several other system scenarios and designs were investigated in [11], [13], [14], [16], and [17]. More specifically, the work in [11] considers low-precision analog-to-digital converters (ADCs) at the BS and derives a linear channel estimator. The authors in [14] exploited the sparse structure of mmWave channels to derive a Cramér-Rao lower bound (CRB) for the channel parameters, which is then optimized to design an RIS reflection pattern. The effect of beam squint was taken into account in [17] and a twin-stage orthogonal matching pursuit (TS-OMP) algorithm was developed to estimate the channel parameters. The double-structured angular sparsity of cascaded channels was exploited in [13] and [16] to both reduce the pilot overhead and improve the estimation performance. The work in [18] developed a maximum likelihood (ML) channel estimation framework for estimating the line-of-sight (LoS) user-RIS channel. Exploiting the fact that the channel angles vary much slower than the channel gains, the authors in [15] proposed a two-timescale parametric estimation strategy which estimates all the channel angles and gains in the first coherence block, and then only re-estimates the channel gains in the remaining coherence blocks.

Unlike the aforementioned works where all the RIS elements are assumed to be passive, some other structured channel estimation studies in [19], [20], [21], [22], [23], [24], and [25] assume that the RIS contains a small number of

Manuscript received 12 September 2023; revised 30 January 2024 and 19 March 2024; accepted 20 April 2024. Date of publication 30 April 2024; date of current version 18 October 2024. This work was supported by the U.S. National Science Foundation Grants CNS-2107182 and ECCS-2030029. The associate editor coordinating the review of this article and approving it for publication was L. Yang. (Corresponding author: Ly V. Nguyen.)

The authors are with the Center for Pervasive Communications and Computing, Henry Samueli School of Engineering, University of California at Irvine, Irvine, CA 92697 USA (e-mail: vanln1@uci.edu; swindle@uci.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCOMM.2024.3395333>.

Digital Object Identifier 10.1109/TCOMM.2024.3395333

0090-6778 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

active elements that can operate in sensing mode to estimate partial CSI, which is then exploited together with the sparsity structure of mmWave channels to reconstruct the full CSI. While compressed sensing (CS) methods were used in [19] and [20], some other techniques were employed in [21], [22], and [23], e.g., signal parameters via rotational invariance technique (ESPRIT) and multiple signal classification (MUSIC) in [21] and [23] and deep residual networks in [22]. Unlike the methods in [19], [20], [21], [22], and [23] that require both uplink and downlink training signals, the work in [24] developed a variational inference-sparse Bayesian learning channel estimator that uses only the uplink training signals and exploits the received signals at both the RIS and the BS. In [25], a single-active element, referred to as a wireless beacon, is used at the RIS to obtain partial CSI based on which a hybrid structured sparsity expectation-maximization (EM) algorithm was developed to make the formulated CS problem tractable and efficiently achieve the sparse channel recovery.

On the other hand, unstructured RIS channel estimation has also been rigorously investigated in many works, e.g., literature [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], and [38]. For single-user systems, the works in [26] and [27] used a binary reflection strategy where only one reflecting element is turned on in each time slot. It was then shown in [28] and [29] that turning on all the RIS elements at the same time and using a discrete Fourier transform (DFT) matrix as the reflecting pattern provides better performance compared to the binary reflection strategy. Similar results were also reported for the case of multiple users in [30]. Additionally, the study in [31] examines the reflecting pattern design problem while imposing the restriction that the phase shifts are limited to a finite set of discrete values. For multi-user systems, the work in [32] exploits known spatial correlation at both the BS and the RIS as well as other statistical characteristics of multi-specular fading to derive Bayesian channel estimators. The work in [36] assumes a low-rank RIS-BS channel and develops a two-stage algorithm based on matrix factorization and matrix completion. Some other methods such as matrix-calibration-based factorization and parallel factor tensor decomposition were used in [33], [34], and [35], respectively. More general channel models were considered in [37] and [38] where two- and three-phase estimation approaches were proposed, respectively. While both of these latter approaches require the same pilot overhead, the two-phase approach outperforms the other thanks to the alleviation of error propagation. Unlike the common approach in which the cascaded CSI is estimated to optimize the RIS configuration, the authors in [39] and [40] proposed a codebook-based approach where the RIS configuration is selected from a pre-defined codebook. For each codebook configuration, only the effective channel is estimated to evaluate the performance of that configuration. The configuration giving the best performance is then selected for the data transmission phase. Although the training overhead of this approach is independent of the number of RIS elements, it scales with the codebook size.

All of the aforementioned methods fall into the class of pilot-directed (PD) channel estimation, i.e., only pilot signals are used for channel estimation. In this paper, we focus on a

different class, referred to as decision-directed (DD) channel estimation, where both pilot and data signals are exploited for the channel estimation task. Compared to PD, the DD approach can help significantly reduce the pilot overhead. Joint channel estimation and data detection was studied in [41], [42], [43], and [44] where sparsity-structured and unstructured channels were considered in [41], [42], [43], and [44], respectively. However, these works all assumed passive RISs. In this paper, we consider a recent hybrid RIS structure [45], [46], [47] in which the RIS elements can simultaneously reflect and sense the incoming signal, and we develop a DD channel estimator that can be used for unstructured channels where AoA/AoD information cannot be exploited. The novelty of the approach lies in the application of hybrid RIS for unstructured channel estimation, and the use of DD to reduce the pilot overhead. It should be noted that the hybrid RIS structure in [19], [20], [21], [22], [23], and [24] can only operate in either reflecting or receiving mode, while the hybrid RIS structure in [45], [46], and [47] can simultaneously reflect and absorb the incoming signal. The contributions of our paper are summarized as follows:

- Based on the hybrid RIS structure, we first develop a two-phase pilot-directed (PD) channel estimation approach. The estimation strategy is similar to that in [38] but we show that the pilot overhead is lower for multiuser systems.
- Next, we propose a two-phase DD channel estimation framework and we show that with the help of the hybrid RIS elements, it is possible to accurately recover the CSI with a pilot overhead only proportional to the number of users. Therefore, the proposed DD framework substantially improves the system spectral efficiency (SE). More specifically, in the channel estimation stage, the users transmit a sequence including both pilot and data symbols where the number of pilot symbols is the same as the number of users. The RIS uses some sensor elements with RF chains to recover the data symbols, and then forwards the detected data symbols to the BS for cascaded channel estimation. For the BS to accurately estimate the CSI, the RIS phase shifts must be varied. We point out that changing the RIS phase shifts does not affect data detection by the sensing RIS elements, and thus both data recovery at the RIS and channel estimation at the BS are guaranteed. We also explain why accurate CSI recovery is not guaranteed when the DD approach is applied at the BS and the RIS has no sensing elements.
- We then perform a detailed spectral efficiency (SE) analysis for both the PD and DD frameworks for single-user systems. Our analysis takes into account both the channel estimation and data detection errors at the RIS and the BS, and thus accurately reflects the uncertainty of RIS-assisted data detection in the DD framework. It is observed that there is often a crossing point at which the DD framework outperforms the PD one, and so the analysis can be used to decide when the PD or DD approach should be used. Finally, we present numerous simulation results to verify the accuracy of the SE analysis as well as to show the benefits of the proposed DD framework.

The rest of this paper is organized as follows: Section II presents the considered system model. The pilot directed and decision-directed channel estimation frameworks are presented in Section III and Section IV, respectively. We perform the spectral efficiency analysis in Section V. Section VI shows simulation results and finally Section VII concludes the paper.

*Notation:* Upper-case and lower-case boldface letters denote matrices and column vectors, respectively. Scalars  $x_{ij}$  and  $[X]_{ij}$  both denote the element at the  $i$ th row and  $j$ th column of a matrix  $\mathbf{X}$ . Vectors  $\mathbf{x}_i$  and  $\mathbf{X}_{:,i}$  both denote the  $i$ th column of a matrix  $\mathbf{X}$ , while  $\mathbf{X}_{k,:}$  denotes the  $k$ -th row of  $\mathbf{X}$ . The notation  $\mathbf{X}_{i:j,k:\ell}$  represents the sub-matrix of  $\mathbf{X}$  that includes rows  $i$  to  $j$  and columns  $k$  to  $\ell$ . The expectation, variance, and covariance of random quantities are denoted by  $\mathbb{E}[\cdot]$ ,  $\text{Var}[\cdot]$ , and  $\text{Cov}[\cdot]$ , respectively. Depending on the context, the operator  $|\cdot|$  is used to denote the absolute value of a number, or the cardinality of a set. The  $\ell_2$ -norm of a vector is represented by  $\|\cdot\|$ . The transpose and conjugate transpose are denoted by  $[\cdot]^T$  and  $[\cdot]^H$ , respectively,  $j$  is the unit imaginary number satisfying  $j^2 = -1$ ,  $\mathcal{N}(\cdot, \cdot)$  and  $\mathcal{CN}(\cdot, \cdot)$  represent the real and the complex normal distributions respectively, where the first argument is the mean and the second argument is the variance or the covariance matrix. The  $i$ -th element of the set  $\mathcal{A}$  is indicated by  $\mathcal{A}(i)$ . The Q-function that quantifies the tail distribution of a standard normal random variable is given by  $Q(\cdot)$ .

## II. SYSTEM MODEL

We consider an uplink RIS-assisted MIMO system in which a BS with  $M$  antennas serves  $K$  single-antenna users under the assistance of an  $N$ -element RIS. Let  $\mathbf{H}_d \in \mathbb{C}^{M \times K}$ ,  $\mathbf{H} \in \mathbb{C}^{M \times N}$ , and  $\mathbf{G} \in \mathbb{C}^{N \times K}$  denote the direct channel from the users to the BS, the channel from the RIS to the BS, and the channel from the users to the RIS, respectively. The RIS contains a number of sensing elements equipped with radio-frequency (RF) chains as illustrated in Fig. 1. These sensing elements are able to simultaneously reflect and sense the impinging signal. Let  $\mathcal{A}$  denote the index set of the sensing elements, so that  $\mathcal{A} \subset \{1, \dots, N\}$ , and let  $N_{\mathcal{A}}$  be the number of sensing elements, i.e.,  $N_{\mathcal{A}} = |\mathcal{A}|$ , where it is assumed that  $K \leq N_{\mathcal{A}} \ll N$ . We also let  $\mathcal{B}$  denote the index set of the regular elements without sensing capability. This implies  $\mathcal{A} \cap \mathcal{B} = \emptyset$  and  $\mathcal{A} \cup \mathcal{B} = \{1, \dots, N\}$ . Without loss of generality, we assume that the indices in  $\mathcal{A}$  and  $\mathcal{B}$  are sorted in an ascending order, i.e.,  $\mathcal{A}(1) < \dots < \mathcal{A}(N_{\mathcal{A}})$  and  $\mathcal{B}(1) < \dots < \mathcal{B}(N - N_{\mathcal{A}})$ .

Define the channel matrices  $\mathbf{H}_d \triangleq [\mathbf{h}_{d,1}, \dots, \mathbf{h}_{d,K}]$  and  $\mathbf{G} \triangleq [\mathbf{g}_1, \dots, \mathbf{g}_K]$ , so that the received signal at the BS is modeled as<sup>1</sup>

$$\mathbf{y}^{\text{BS}} = \sqrt{P} \sum_{k=1}^K \mathbf{H}_{c,k} \text{diag} \left( \begin{bmatrix} [0.8]1 \\ \boldsymbol{\rho} \end{bmatrix} \right) \begin{bmatrix} [0.8]1 \\ \boldsymbol{\phi} \end{bmatrix} s_k + \mathbf{n}^{\text{BS}} \quad (1)$$

where  $\boldsymbol{\phi} = [\phi_1, \dots, \phi_N]^T$  is the phase shift vector of the RIS,  $\mathbf{H}_{c,k} = [\mathbf{h}_{d,k}, \mathbf{H} \text{diag}(\mathbf{g}_k)] \in \mathbb{C}^{M \times (N+1)}$  is the cascaded

<sup>1</sup>In this paper, we focus on a narrowband system model. However, the proposed approach can be directly applied to OFDM systems by adopting our channel estimator separately to each subcarrier.

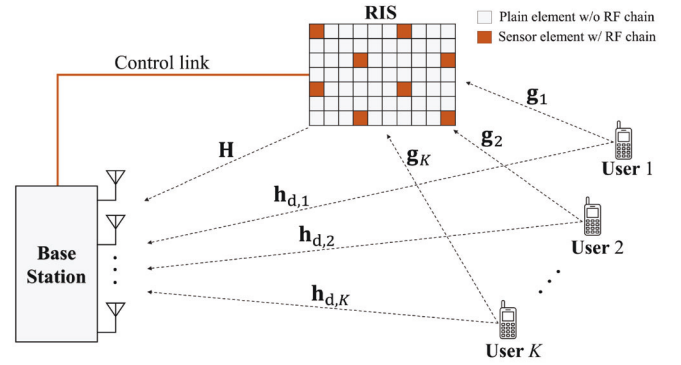


Fig. 1. Sensing-RIS-assisted multi-user MIMO system.

channel of the  $k$ -th user,  $P$  is the transmit power, and  $\boldsymbol{\rho} \triangleq [\rho_1, \dots, \rho_N]^T$  with  $0 \leq \rho_n \leq 1$  if  $n \in \mathcal{A}$ , otherwise  $\rho_n = 1$ . Hence,  $\rho_n^2$  is the portion of the power of the impinging signal that is reflected by the  $n$ -th RIS element. For convenience, we use the notation  $\boldsymbol{\rho}^{\mathcal{A}} = [\rho_1^{\mathcal{A}}, \dots, \rho_{N_{\mathcal{A}}}^{\mathcal{A}}]^T$  where  $\rho_i^{\mathcal{A}} \triangleq \rho_{\mathcal{A}(i)}$  for  $i = 1, \dots, N_{\mathcal{A}}$ , and  $\boldsymbol{\eta}^{\mathcal{A}} = [\eta_1^{\mathcal{A}}, \dots, \eta_{N_{\mathcal{A}}}^{\mathcal{A}}]^T$  where  $\eta_i^{\mathcal{A}} = \sqrt{1 - (\rho_i^{\mathcal{A}})^2}$ . Hence,  $(\eta_i^{\mathcal{A}})^2$  represents the amount of signal power absorbed by the RIS element  $\mathcal{A}(i)$  for the purpose of sensing.

With  $N_{\mathcal{A}}$  sensing elements at the RIS, the received signal at the RIS is given as

$$\mathbf{y}^{\text{RIS}} = \sqrt{P} \text{diag}(\boldsymbol{\eta}^{\mathcal{A}}) \text{diag}(\boldsymbol{\phi}^{\mathcal{A}}) \sum_{k=1}^K \mathbf{g}_k^{\mathcal{A}} s_k + \mathbf{n}^{\text{RIS}}, \quad (2)$$

where  $\mathbf{g}_k^{\mathcal{A}} = [g_{k,1}^{\mathcal{A}}, \dots, g_{k,N_{\mathcal{A}}}^{\mathcal{A}}]^T$  with  $g_{k,i}^{\mathcal{A}} \triangleq g_{k,\mathcal{A}(i)}$ . In this paper, it is important to note that the superscripts  $(\cdot)^{\mathcal{A}}$  and  $(\cdot)^{\mathcal{B}}$  are used to imply variables that are associated with the sensing and reflecting RIS elements, respectively.

We assume an uplink communication protocol with two stages including a channel estimation stage followed a data transmission stage. After the channel estimation stage, the RIS phase shifts are optimized and configured before the data transmission stage begins. It should be noted that during the channel estimation stage, data detection occurs at the RIS since the users transmit both pilot and data symbols during this stage. During the data detection stage, in order to minimize power consumption at the RIS, the sensing function of the hybrid RIS elements is turned off and the incoming signal is completely reflected.

## III. PILOT-DIRECTED CHANNEL ESTIMATION

In this section, we present a two-phase pilot-directed approach for estimating the cascaded channel matrices  $\mathbf{H}_{c,1}, \dots, \mathbf{H}_{c,K}$ . Since all the users experience the same RIS-BS channel, i.e. the same channel matrix  $\mathbf{H}$ , the total number of channel elements to be estimated is  $M(K + N) + N(K - 1)$  [37], [38]. Let  $\mathbf{A}_k = \mathbf{H} \text{diag}(\mathbf{g}_k)$ , then we have  $\mathbf{A}_k = \mathbf{A}_1 \text{diag}(\boldsymbol{\lambda}_k)$  where  $\lambda_{k,n} = g_{k,n}/g_{1,n}$  for  $k = 2, \dots, K$  and  $n = 1, \dots, N$ . Note that  $\boldsymbol{\lambda}_1 = \mathbf{1}_N$ . Therefore, it suffices to estimate  $\mathbf{H}_d$ ,  $\mathbf{A}_1$ , and  $\boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_K$ .

Our two-phase estimation strategy is similar to that in [38] where  $\mathbf{H}_{c,1} = [\mathbf{h}_{d,1}, \mathbf{A}_1]$  is estimated in phase 1 and  $\mathbf{h}_{d,2}, \dots, \mathbf{h}_{d,K}$  and  $\boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_K$  are estimated in phase 2.



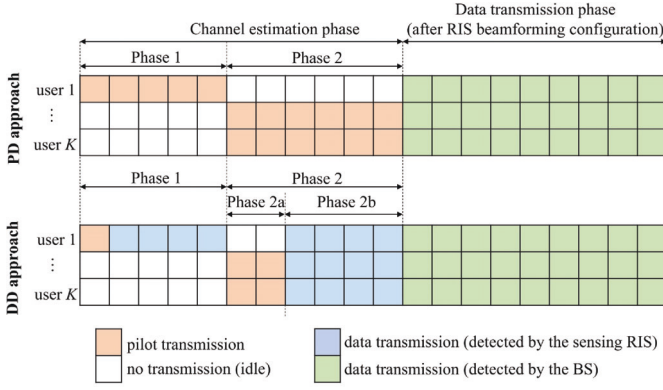


Fig. 2. Illustration of the proposed channel estimation framework. The upper and lower diagrams are for the PD and DD approaches, respectively.

However, unlike the work in [38], which considers an RIS with passive elements only, our work here considers a hybrid RIS structure as presented above. In this section, we assume that only pilot signals are used for the channel estimation. For notational convenience, let  $\mathcal{T}_1 = \{1, \dots, \tau_1\}$  and  $\mathcal{T}_2 = \{\tau_1 + 1, \dots, \tau_1 + \tau_2\}$  where  $\tau_1$  and  $\tau_2$  are the length of phase 1 and phase 2, respectively. An illustration of the PD approach is given in the upper diagram of Fig. 2.

#### A. Phase 1

In this phase, we estimate  $\mathbf{H}_{c,1} = [\mathbf{h}_{d,1}, \mathbf{A}_1]$ . One selected user transmits a pilot vector of length  $N + 1$ , while the other users remain idle. Without loss of generality, we set the index of the typical user to 1. The received signal at the BS in this phase is given as

$$\mathbf{y}_t^{\text{BS}} = \sqrt{P} \mathbf{H}_{c,1} \text{diag} \left( \begin{bmatrix} [0.8]1 \\ \boldsymbol{\rho} \end{bmatrix} \right) \begin{bmatrix} [0.8]1 \\ \boldsymbol{\phi}_t \end{bmatrix} s_{1,t} + \mathbf{n}_t^{\text{BS}}.$$

Since  $\mathbf{H}_{c,1}$  contains  $N + 1$  columns, we need at least  $\tau_1 = N + 1$  time slots to accurately estimate  $\mathbf{H}_{c,1}$ . For simplicity, we can set the pilot vector as  $\mathbf{S}_{1,\tau_1} = \mathbf{1}_{\tau_1}^T$  and the RIS phase shift matrix  $\boldsymbol{\Phi} = [\phi_1, \dots, \phi_{\tau_1}]$  is chosen so that  $[\mathbf{1}_{\tau_1}, \boldsymbol{\Phi}^T]^T = \mathbf{V}_{\tau_1}$  where  $\mathbf{V}_{\tau_1}$  is the DFT matrix of size  $\tau_1 \times \tau_1$ . This means  $[1, \phi_t^T]^T$  is the  $t$ -th column of  $\mathbf{V}_{\tau_1}$ . Then, the cascaded channel  $\mathbf{H}_{c,1}$  can be estimated via standard methods, such as for example the least-squares (LS)<sup>2</sup>:

$$\begin{aligned} \hat{\mathbf{H}}_{c,1} &= \frac{1}{\sqrt{P\tau_1}} \mathbf{Y}_{:, \tau_1}^{\text{BS}} \boldsymbol{\Phi}_{\tau_1}^H \text{diag} \left( \begin{bmatrix} [0.8]1 \\ \boldsymbol{\rho} \end{bmatrix} \right)^{-1} \\ &= \mathbf{H}_{c,1} + \frac{1}{\sqrt{P\tau_1}} \mathbf{N}_{:, \tau_1}^{\text{BS}} \boldsymbol{\Phi}_{\tau_1}^H \text{diag} \left( \begin{bmatrix} [0.8]1 \\ \boldsymbol{\rho} \end{bmatrix} \right)^{-1}. \end{aligned} \quad (3)$$

The received signal at the sensing elements of the RIS is

$$\mathbf{y}_t^{\text{RIS}} = \sqrt{P} \text{diag}(\boldsymbol{\eta}^A) \text{diag}(\boldsymbol{\phi}_t^A) \mathbf{g}_1^A + \mathbf{n}_t^{\text{RIS}} \quad (4)$$

and so the sensed portion of the channel  $\mathbf{g}_1^A$  can be estimated as

$$\hat{\mathbf{g}}_1^A = \frac{1}{\sqrt{P\tau_1}} \text{diag}(\boldsymbol{\eta}^A)^{-1} \sum_{t=1}^{\tau_1} \psi_t, \quad (5)$$

where  $\psi_t = \text{diag}(\boldsymbol{\phi}_t^A)^{-1} \mathbf{y}_t^{\text{RIS}}$  is the received signal compensated by the phase rotation at the sensing elements of the RIS.

<sup>2</sup>For simplicity, we assume the LS approach in this paper, but our proposed channel estimation framework is general and not restricted to a specific type of channel estimation approach. A different channel estimator can easily be adopted instead of the LS approach.

Note that since a passive RIS is assumed in [38], there is no estimation of  $\mathbf{g}_1^A$ . Here, we can obtain an estimate of  $\mathbf{g}_1^A$  as in (5) which will be exploited in phase 2 for estimating the cascaded channel of the other users.

#### B. Phase 2

In this phase, the typical user remains idle while the other users transmit pilot sequences. Since  $\mathbf{h}_{d,1}$  and  $\mathbf{A}_1$  have been estimated in phase 1, we will estimate  $\mathbf{h}_{d,2}, \dots, \mathbf{h}_{d,K}$  and  $\boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_K$  during phase 2. The received signal at the BS can be decomposed as

$$\begin{aligned} \mathbf{y}_t^{\text{BS}} &= \sqrt{P} \sum_{k=2}^K (\mathbf{h}_{d,k} + \mathbf{A}_1^{\text{B}} \text{diag}(\boldsymbol{\phi}_t^{\text{B}}) \boldsymbol{\lambda}_k^{\text{B}} \\ &\quad + \mathbf{A}_1^{\text{A}} \text{diag}(\boldsymbol{\rho}^A) \text{diag}(\boldsymbol{\phi}_t^A) \boldsymbol{\lambda}_k^A) s_{k,t} + \mathbf{n}_t^{\text{BS}} \end{aligned} \quad (6)$$

where  $\mathbf{A}_1^{\text{A}}$  and  $\mathbf{A}_1^{\text{B}}$  are matrices whose columns are drawn from  $\mathbf{A}_1$  with indices  $\mathcal{A}$  and  $\mathcal{B}$ , respectively,  $\boldsymbol{\lambda}_k^{\text{A}} = [\lambda_{k,1}^{\text{A}}, \dots, \lambda_{k,N_{\text{A}}}^{\text{A}}]^T$  and  $\boldsymbol{\lambda}_k^{\text{B}} = [\lambda_{k,1}^{\text{B}}, \dots, \lambda_{k,N_{\text{B}}}^{\text{B}}]^T$  where  $\lambda_{k,i}^{\text{A}} \triangleq \lambda_{k,\mathcal{A}(i)}$  and  $\lambda_{k,i}^{\text{B}} \triangleq \lambda_{k,\mathcal{B}(i)}$ .

1) *Estimating  $\boldsymbol{\lambda}_2^{\text{A}}, \dots, \boldsymbol{\lambda}_K^{\text{A}}$* : This is done at the RIS. Since  $\lambda_{k,i}^{\text{A}} = g_{k,i}^{\text{A}}/g_{1,i}^{\text{A}}$ , the parameters  $\boldsymbol{\lambda}_2^{\text{A}}, \dots, \boldsymbol{\lambda}_K^{\text{A}}$  are defined as long as  $\mathbf{G}^{\text{A}} = [\mathbf{g}_1^{\text{A}}, \dots, \mathbf{g}_K^{\text{A}}]$  is known. Note that the first column of  $\mathbf{G}^{\text{A}}$  has been estimated using (5) in phase 1. The signal received at the RIS in phase 2 is given as

$$\mathbf{y}_t^{\text{RIS}} = \sqrt{P} \text{diag}(\boldsymbol{\eta}^A) \text{diag}(\boldsymbol{\phi}_t^A) \mathbf{G}_{:,2:K}^{\text{A}} \mathbf{S}_{2:K,t} + \mathbf{n}_t^{\text{RIS}}. \quad (7)$$

The sub-matrix  $\mathbf{G}_{:,2:K}^{\text{A}}$  can be estimated by the RIS as follows:

$$\hat{\mathbf{G}}_{:,2:K}^{\text{A}} = \frac{1}{\sqrt{P}} \text{diag}(\boldsymbol{\eta}^A)^{-1} \boldsymbol{\Psi}_{\mathcal{T}_2} \mathbf{S}_{2:K,\mathcal{T}_2}^H (\mathbf{S}_{2:K,\mathcal{T}_2} \mathbf{S}_{2:K,\mathcal{T}_2}^H)^{-1} \quad (8)$$

where  $\boldsymbol{\Psi}_{\mathcal{T}_2} = [\psi_{\tau_1+1}, \dots, \psi_{\tau_1+\tau_2}]$ . In (8), to obtain an estimate of  $\mathbf{G}_{:,2:K}^{\text{A}}$ , we left- and right-multiply  $\boldsymbol{\Psi}_{\mathcal{T}_2}$  by  $\frac{1}{\sqrt{P}} \text{diag}(\boldsymbol{\eta}^A)^{-1}$  and  $\mathbf{S}_{2:K,\mathcal{T}_2}^H (\mathbf{S}_{2:K,\mathcal{T}_2} \mathbf{S}_{2:K,\mathcal{T}_2}^H)^{-1}$  to cancel the effect of the received power scaling and pilot signal, respectively. Thus, an estimate of  $\lambda_{k,i}^{\text{A}}$  can be obtained as

$$\hat{\lambda}_{k,i}^{\text{A}} = \hat{g}_{k,i}^{\text{A}} / \hat{g}_{1,i}^{\text{A}}, \quad (9)$$

where  $\hat{g}_{1,i}^{\text{A}}$  has been estimated in phase 1 as in (5).

2) *Estimating  $\mathbf{h}_{d,2}, \dots, \mathbf{h}_{d,K}$  and  $\boldsymbol{\lambda}_2^{\text{B}}, \dots, \boldsymbol{\lambda}_K^{\text{B}}$* : This is accomplished at the BS. Let

$$\begin{aligned} \mathbf{B}_t &= [\mathbf{I}_M, \mathbf{A}_1^{\text{B}} \text{diag}(\boldsymbol{\phi}_t^{\text{B}})], \\ \mathbf{v}_k &= [\mathbf{h}_{d,k}^T, (\boldsymbol{\lambda}_k^{\text{B}})^T]^T, \\ \mathbf{f}_{k,t}^{\text{A}} &= \mathbf{A}_1^{\text{A}} \text{diag}(\boldsymbol{\rho}^A) \text{diag}(\boldsymbol{\phi}_t^A) \boldsymbol{\lambda}_k^{\text{A}}, \end{aligned}$$

then the received signal at the BS in (6) can be written in the following form:

$$\begin{aligned} \mathbf{y}_t^{\text{BS}} &= \sqrt{P} \sum_{k=2}^K (\mathbf{B}_t \mathbf{v}_k + \mathbf{f}_{k,t}^{\text{A}}) s_{k,t} + \mathbf{n}_t^{\text{BS}} \\ &= \sqrt{P} ((\mathbf{S}_{2:K,t}^T \otimes \mathbf{B}_t) \mathbf{v} + \mathbf{F}_t^{\text{A}} \mathbf{S}_{2:K,t}) + \mathbf{n}_t^{\text{BS}} \\ &= \sqrt{P} (\mathbf{Q}_t \mathbf{v} + \tilde{\mathbf{y}}_t^{\text{BS}}) + \mathbf{n}_t^{\text{BS}} \end{aligned} \quad (10)$$

where  $\mathbf{v} = [\mathbf{v}_2^T, \dots, \mathbf{v}_K^T]^T$ ,  $\mathbf{Q}_t = \mathbf{S}_{2:K,t}^T \otimes \mathbf{B}_t$ ,  $\mathbf{F}_t^A = [\mathbf{f}_{2,t}^A, \dots, \mathbf{f}_{K,t}^A]$ , and  $\tilde{\mathbf{y}}_t^{\text{BS}} = \mathbf{F}_t^A \mathbf{S}_{2:K,t}$ . Stacking the received signals  $\{\mathbf{y}_t^{\text{BS}}\}$  in (10) with  $t \in \mathcal{T}_2$  on top of each other, we have the following

$$\text{vec}(\mathbf{Y}_{:, \mathcal{T}_2}^{\text{BS}}) - \sqrt{P} \text{vec}(\tilde{\mathbf{Y}}_{:, \mathcal{T}_2}^{\text{BS}}) = \sqrt{P} \mathbf{Q} \mathbf{v} + \mathbf{n}^{\text{BS}} \quad (11)$$

where  $\mathbf{Q} = [\mathbf{Q}_{\tau_1+1}^T, \dots, \mathbf{Q}_{\tau_1+\tau_2}^T]^T$ . Note that  $\mathbf{v}$  is the vector we need to estimate and the size of  $\mathbf{Q}$  is  $M\tau_2 \times (K-1)(M+N-N_A)$ . Therefore, in order to accurately recover  $\mathbf{v}$ , two conditions should be satisfied:  $M\tau_2 \geq (K-1)(M+N-N_A)$  and  $\text{rank}(\mathbf{Q}) = M+N-N_A$ . An estimate of  $\mathbf{v}$  can be then obtained as

$$\hat{\mathbf{v}} = \frac{1}{\sqrt{P}} \mathbf{Q}^\dagger \left( \text{vec}(\mathbf{Y}_{:, \mathcal{T}_2}^{\text{BS}}) - \sqrt{P} \text{vec}(\tilde{\mathbf{Y}}_{:, \mathcal{T}_2}^{\text{BS}}) \right). \quad (12)$$

If  $M \geq N - N_A$ , we need at least  $\tau_2 = 2(K-1)$  time slots to recover  $\mathbf{v}$  and if  $M < N - N_A$ , we need at least  $\tau_2 = K-1 + \left\lceil \frac{(K-1)(N-N_A)}{M} \right\rceil$  time slots to recover  $\mathbf{v}$ .

When  $M \geq N - N_A$ , in the two time slots  $\tau_1 + 2(k-2) + 1$  and  $\tau_1 + 2(k-2) + 2$ , only user- $k$  transmits while the other users stay idle. The matrix  $\mathbf{Q}$  becomes a block-diagonal matrix and so the solution in (12) can be decomposed into  $K-1$  separate expressions as follows:

$$\begin{aligned} \hat{\mathbf{v}}_k &= \frac{1}{\sqrt{P}} \left[ \begin{array}{c} [1.15] s_{k, \tau_1+2(k-2)+1} \mathbf{B}_{\tau_1+2(k-2)+1} \\ s_{k, \tau_1+2(k-2)+2} \mathbf{B}_{\tau_1+2(k-2)+2} \end{array} \right]^\dagger \\ &\times \left( \left[ \begin{array}{c} [1.15] \mathbf{y}_{\tau_1+2(k-2)+1}^{\text{BS}} \\ \mathbf{y}_{\tau_1+2(k-2)+2}^{\text{BS}} \end{array} \right] - \sqrt{P} \left[ \begin{array}{c} [1.15] \tilde{\mathbf{y}}_{\tau_1+2(k-2)+1}^{\text{BS}} \\ \tilde{\mathbf{y}}_{\tau_1+2(k-2)+2}^{\text{BS}} \end{array} \right] \right). \end{aligned} \quad (13)$$

The scaling parameters  $\lambda_2^A, \dots, \lambda_K^A$  and  $\lambda_2^B, \dots, \lambda_K^B$  associated with the sensing and reflection-only elements are sequentially estimated in (9) and (12), respectively. The estimation of  $\lambda_2^B, \dots, \lambda_K^B$  depends on the estimates  $\hat{\lambda}_2^A, \dots, \hat{\lambda}_K^A$  as the term  $\text{vec}(\tilde{\mathbf{Y}}_{:, \mathcal{T}_2}^{\text{BS}})$  constructed from  $\hat{\lambda}_2^A, \dots, \hat{\lambda}_K^A$  is subtracted from the received signal  $\text{vec}(\mathbf{Y}_{:, \mathcal{T}_2}^{\text{BS}})$  in (12). The work in [38] does not follow this procedure but directly estimates all scaling parameters from the received signal  $\text{vec}(\mathbf{Y}_{:, \mathcal{T}_2}^{\text{BS}})$  since a passive RIS structure was assumed. In the following, we show that the pilot overhead of our sensing-based PD approach is generally lower than that of the PD approach in [38].

### C. Overall Training Overhead

Since the pilot overhead in phase 1 is  $\tau_1 = N+1$  and in phase 2 is  $\tau_2 = 2(K-1)$  if  $M \geq N - N_A$ , or  $\tau_2 = K-1 + \left\lceil \frac{(K-1)(N-N_A)}{M} \right\rceil$  otherwise, the total pilot overhead of the proposed sensing-based PD approach will be

$$\begin{aligned} \tau_p &= \tau_1 + \tau_2 \\ &= \begin{cases} N + 2K - 1, & \text{if } M \geq N - N_A, \\ N + K + \left\lceil \frac{(K-1)(N-N_A)}{M} \right\rceil, & \text{otherwise.} \end{cases} \end{aligned} \quad (14)$$

The pilot overhead of the approach in [38] is also  $N + 2K - 1$  when  $M \geq N$ . However, when  $M < N$ , it is  $N + K + \left\lceil \frac{(K-1)N}{M} \right\rceil$ , which is clearly no less than  $N + K + \left\lceil \frac{(K-1)(N-N_A)}{M} \right\rceil$  in (14) since  $N_A > 0$ .

Although the pilot overhead in (14) is less than  $K(N+1)$  as required by the generic LS channel estimator at the BS, it is still proportional to the number of RIS elements  $N$ , which can be excessively large. In the following section, we propose a DD channel estimation approach whose pilot overhead is only proportional to the number of users  $K$ .

## IV. DECISION-DIRECTED CHANNEL ESTIMATION

In this section, we propose a two-stage DD channel estimation approach to substantially reduce the pilot overhead, and thus improve the system SE. An illustration of this DD approach is given in the lower diagram of Fig. 2. As in the PD approach above, there are also two phases as detailed in what follows.

### A. Phase 1

Similar to the PD approach, here we also use  $\tau_1 = N+1$  time slots to estimate  $\mathbf{H}_{c,1}$  with the typical user active and the other users remaining idle. However, unlike the PD approach where all of the  $N+1$  time slots are used for pilot signals, the DD approach uses only the first time slot to transmit a pilot symbol and the remaining  $N$  time slots are for transmitting data symbols.

1) *Estimating  $\mathbf{g}_1^A$* : The received signal at the RIS in the first time slot with the pilot symbol  $s_{1,1} = 1$  is given in (4), and so an estimate of  $\mathbf{g}_1^A$  can be obtained as

$$\hat{\mathbf{g}}_1^A = \frac{1}{\sqrt{P}} \text{diag}(\boldsymbol{\eta}^A)^{-1} \text{diag}(\boldsymbol{\phi}_1^A)^{-1} \mathbf{y}_1^{\text{RIS}}. \quad (15)$$

2) *Data Detection for User 1 in Time Slots 2,  $\dots$ ,  $\tau_1$* : The received signal is

$$\mathbf{y}_t^{\text{RIS}} = \sqrt{P} \text{diag}(\boldsymbol{\eta}^A) \text{diag}(\boldsymbol{\phi}_t^A) \mathbf{g}_1^A s_{1,t} + \mathbf{n}_t^{\text{RIS}}. \quad (16)$$

From (16), it can be seen that the effect of the phase shift vector  $\boldsymbol{\phi}_t^A$  is merely a phase rotation of the noiseless received signal. Since  $\boldsymbol{\phi}_t^A$  is known at the RIS, it can be used by the RIS to detect the data symbols  $s_{1,t}$  as follows:

$$\hat{s}_{1,t} = \arg \min_{s \in \mathcal{S}} \left\| \mathbf{y}_t^{\text{RIS}} - \sqrt{P} \text{diag}(\boldsymbol{\eta}^A) \text{diag}(\boldsymbol{\phi}_t^A) \hat{\mathbf{g}}_1^A s \right\|^2 \quad (17)$$

where  $\mathcal{S}$  denotes the set of constellation points. Thus, even when the phase shift vector  $\boldsymbol{\phi}_t^A$  varies in different time slots, it is still possible for the RIS to accurately recover the data symbols since the effect of  $\boldsymbol{\phi}_t^A$  can be easily taken into account as in (17). As will be explained later in Section IV-D, this accurate data recovery is not possible when the RIS has no sensing elements and the DD strategy is applied at the BS.

3) *Estimating  $\mathbf{H}_{c,1}$* : The detected data symbols  $\{\hat{s}_{1,t}\}$  in (17) will be forwarded<sup>3</sup> by the RIS to the BS through a control link so that the BS can use these detected symbols to estimate the cascaded channel matrix  $\mathbf{H}_{c,1}$  as follows:

$$\hat{\mathbf{H}}_{c,1} = \frac{1}{\sqrt{P\tau_1}} \mathbf{Y}_{:, \tau_1}^{\text{BS}} \text{diag}(\hat{\mathbf{s}}_{\tau_1})^{-1} \Phi_{\tau_1}^H \text{diag}\left(\begin{bmatrix} [0.8]1 \\ \rho \end{bmatrix}\right)^{-1}. \quad (18)$$

where  $\hat{\mathbf{s}}_{\tau_1} = [1, \hat{s}_{1,2}, \dots, \hat{s}_{1,\tau_1}]$ . Thus, in phase 1, with the help of the sensing elements, we use only one time slot for pilot signalling, i.e.,  $\tau_1 = 1$ , while the other  $N$  time slots are used for data transmission. The BS is still able to accurately recover the cascaded channel matrix  $\mathbf{H}_{c,1}$  as long as the data symbols are correctly detected by the RIS.

### B. Phase 2

We divide phase 2 into two sub-phases that we refer to as 2a and 2b. Sub-phase 2a is associated with the time frame  $\mathcal{T}_{2a} = \tau_1 + 1, \dots, \tau_1 + K - 1$  where user 1 is idle and users 2 through  $K$  transmit their pilot signals. Sub-phase 2b is associated with the time frame  $\mathcal{T}_{2b} = \tau_1 + K, \dots, \tau_1 + \tau_2$  where all the users transmit data symbols.

1) *Estimating  $\mathbf{g}_2^A, \dots, \mathbf{g}_K^A$  and  $\lambda_2^A, \dots, \lambda_K^A$* : Pilot signals are transmitted in the first  $K - 1$  time slots of phase 2, from  $\tau_1 + 1$  to  $\tau_1 + K - 1$ , so the sub-matrix  $\mathbf{G}_{:,2:K}^A = [\mathbf{g}_2^A, \dots, \mathbf{g}_K^A]$  can be estimated by the RIS as follows:

$$\hat{\mathbf{G}}_{:,2:K}^A = \frac{\text{diag}(\eta^A)^{-1} \Psi_{\mathcal{T}_{2a}} \mathbf{S}_{2:K, \mathcal{T}_{2a}}^H (\mathbf{S}_{2:K, \mathcal{T}_{2a}} \mathbf{S}_{2:K, \mathcal{T}_{2a}}^H)^{-1}}{\sqrt{P}}, \quad (19)$$

where  $\Psi_{\mathcal{T}_{2a}} = [\psi_{\tau_1+1}, \dots, \psi_{\tau_1+K-1}]$ . Similar to (8), to obtain an estimate of  $\hat{\mathbf{G}}_{:,2:K}^A$  in (19) we left- and right-multiply  $\Psi_{\mathcal{T}_{2a}}$  by  $\frac{1}{\sqrt{P}} \text{diag}(\eta^A)^{-1}$  and  $\mathbf{S}_{2:K, \mathcal{T}_{2a}}^H (\mathbf{S}_{2:K, \mathcal{T}_{2a}} \mathbf{S}_{2:K, \mathcal{T}_{2a}}^H)^{-1}$  to cancel the effect of the received power scaling and pilot signal, respectively. Furthermore, an estimate of  $\lambda_{k,i}^A$  can be obtained as  $\hat{\lambda}_{k,i}^A = \hat{g}_{k,i}^A / \hat{g}_{1,i}^A$ .

2) *Detecting Data*: For the remaining time slots from  $\tau_1 + K$  to  $\tau_1 + \tau_2$  in sub-phase 2b, all  $K$  users can transmit data, and the received signal at the RIS is

$$\mathbf{y}_t^{\text{RIS}} = \sqrt{P} \text{diag}(\eta^A) \text{diag}(\phi_t^A) \mathbf{G}^A \mathbf{s}_t + \mathbf{n}_t^{\text{RIS}}.$$

The RIS can use  $\mathbf{y}_t^{\text{RIS}}$  and  $\hat{\mathbf{G}}^A = [\hat{\mathbf{g}}_1^A, \dots, \hat{\mathbf{g}}_K^A]$  to detect the users' transmitted data  $\mathbf{s}_t$ , which is a conventional MIMO data detection problem. Similarly, the effect of  $\phi_t^A$  is merely the phase rotation of the noiseless received signal, and so it is feasible for the RIS to accurately recover the data symbols as  $\phi_t^A$  varies in time.

3) *Estimating  $\mathbf{h}_{d,2}, \dots, \mathbf{h}_{d,K}$  and  $\lambda_2^B, \dots, \lambda_K^B$* : Since the typical user also transmits data during sub-phase 2b, the received signal at the BS can be re-written in the following form:

$$\begin{aligned} \mathbf{y}_t^{\text{BS}} - \mathbf{H}_{c,1} \text{diag}\left(\begin{bmatrix} [0.8]1 \\ \rho \end{bmatrix}\right) \begin{bmatrix} [0.8]1 \\ \phi_t \end{bmatrix} s_{1,t} \\ = \sqrt{P} \sum_{k=2}^K (\mathbf{B}_t \mathbf{v}_k + \mathbf{f}_{k,t}^A) s_{k,t} + \mathbf{n}_t^{\text{BS}} \end{aligned}$$

$$\begin{aligned} &= \sqrt{P} ((\mathbf{S}_{2:K,t}^T \otimes \mathbf{B}_t) \mathbf{v} + \mathbf{F}_t^A \mathbf{S}_{2:K,t}) + \mathbf{n}_t^{\text{BS}} \\ &= \sqrt{P} (\mathbf{Q}_t \mathbf{v} + \tilde{\mathbf{y}}_t^{\text{BS}}) + \mathbf{n}_t^{\text{BS}}, \end{aligned} \quad (20)$$

where  $t \in \mathcal{T}_2$ . Note that  $s_{1,t} = 0$  for  $t \in \mathcal{T}_{2a}$  since user 1 is idle during sub-phase 2a. Then, we can use a similar technique as in the PD approach for estimating  $\mathbf{v}$ , but we need to replace  $s_{k,t}$  with  $\hat{s}_{k,t}$  for  $t \in \mathcal{T}_{2b}$ .

### C. Overall Training Overhead

The overall training overhead for the proposed decision-directed approach is  $\tau_p = K$  since phase 1 and phase 2 require only 1 and  $K - 1$  time slots for pilot signalling, respectively. The BS is guaranteed to accurately recover the channel matrices as long as the data symbols are correctly detected by the RIS. Although only the typical user transmits in phase 1 and thus the spectral efficiency will not be as large as if all the users were transmitting, we will show in the numerical results that the proposed DD approach can still result in an increase in the spectral efficiency compared with the PD approach.

### D. Comparison With a Passive RIS and DD at the BS

In the proposed DD method, the BS can accurately recover the cascaded channel matrices when the data symbols are correctly detected by the RIS. Here, we explain why an alternative scenario in which the RIS has no sensing elements and the DD strategy is applied at the BS cannot guarantee accurate CSI estimation. To show this, it is enough to consider the case with only one user, where the received signal at the BS is given as

$$\mathbf{y}_t^{\text{BS}} = \mathbf{H}_{c,1} \phi_t s_t + \mathbf{n}_t^{\text{BS}}.$$

To accurately recover  $\mathbf{H}_{c,1}$ , the phase shift vector  $\phi_t$  must vary for different time slots  $t$  in order to make  $\Phi_{\mathcal{T}_1} = [\phi_1, \dots, \phi_{\tau_1}]$  full-rank. However, if we change  $\phi_t$ , the effective channel  $\mathbf{f}_t = \mathbf{H}_{c,1} \phi_t$  changes as well. A pilot signal is sent in the first time slot, and the BS can only estimate the effective channel  $\mathbf{f}_1$ . With an estimate of  $\mathbf{f}_1$ , the BS cannot guarantee correct detection of data symbols in subsequent time slots  $2, \dots, \tau_1$  since the effective channels in these time slots are different from  $\mathbf{f}_1$  due to the change in  $\phi_t$ . For the effective channel to remain unchanged, the phase shift vector  $\phi_t$  must be time-invariant, but in this case the matrix  $\Phi_{\mathcal{T}_1}$  would be rank-1, which prevents recovery of the cascaded channel  $\mathbf{H}_{c,1}$ . This is unlike the proposed DD channel estimation framework presented above where the data symbols can be accurately recovered by the RIS even when the phase shift vector of the RIS changes in time.

## V. SPECTRAL EFFICIENCY ANALYSIS

In this section, we perform a spectral efficiency analysis for both the PD and DD approaches. We consider an RIS-aided system where the BS has one antenna serving one user without a direct channel. We assume  $D$ -PSK data signalling, i.e.,  $s \in \mathcal{S} = \{\exp(j\pi \frac{2\ell+1}{D})\}$  for  $\ell \in \{0, \dots, D-1\}$  with the Gray code mapping data bits to data symbols. We also assume that the data symbols are equally likely. Let  $a_i = h_i g_i$  be the cascaded channel where  $g_i$  and  $h_i$  are the channels from the

<sup>3</sup>In fact, it suffices for the RIS to forward only the indices of the detected data symbols  $\{\hat{s}_{1,t}\}$ .

user and the BS to the  $i$ -th element of the RIS, respectively. It is assumed that  $g_i \sim \mathcal{CN}(0, \sigma_g^2)$  and  $h_i \sim \mathcal{CN}(0, \sigma_h^2)$  are independent of each other. Let  $\mathbf{a} = [a_1, \dots, a_N]^T$  so that the received signal at the BS can be written as

$$y^{\text{BS}} = \sqrt{P} \mathbf{a}^H \boldsymbol{\phi} s + n^{\text{BS}}. \quad (21)$$

Let  $\hat{\mathbf{a}} = \mathbf{a} + \boldsymbol{\epsilon}$  be the estimated cascaded channel where  $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_N]^T$  is the channel estimation error. Given the  $\hat{\mathbf{a}}$ , the RIS coefficients  $\boldsymbol{\phi}$  in the data transmission phase are chosen to maximize the effective channel strength, i.e.,

$$\underset{\{\boldsymbol{\phi}\}}{\text{maximize}} \quad |\hat{\mathbf{a}}^H \boldsymbol{\phi}|^2 \quad \text{subject to} \quad |\phi_i| \leq 1 \quad \forall i = 1, \dots, N,$$

which has the optimal solution

$$\phi_i^* = e^{j\angle(\hat{a}_i)}. \quad (22)$$

The received signal at the BS in the data transmission phase will then be

$$y^{\text{BS}} = \sqrt{P} \sum_{i=1}^N a_i^* e^{j\angle(a_i + \epsilon_i)} s + n^{\text{BS}} = \sqrt{P} \sum_{i=1}^N z_i s + n^{\text{BS}} \quad (23)$$

where  $z_i \triangleq a_i^* e^{j\angle(a_i + \epsilon_i)}$ . Thus we have

$$\begin{aligned} z_{i,\mathcal{R}} &\triangleq \Re\{z_i\} = \frac{|a_i|^2 + \Re\{a_i \epsilon_i^*\}}{\sqrt{|a_i|^2 + 2\Re\{a_i \epsilon_i^*\} + |\epsilon_i|^2}}, \\ z_{i,\mathcal{I}} &\triangleq \Im\{z_i\} = \frac{\Im\{a_i \epsilon_i^*\}}{\sqrt{|a_i|^2 + 2\Re\{a_i \epsilon_i^*\} + |\epsilon_i|^2}}. \end{aligned}$$

#### A. Pilot-Directed

For the pilot-directed method, the SE is given as [48], [49]

$$\text{SE}_{\text{PD}} = \frac{\tau_c - \tau_p}{\tau_c} (1 - \text{BER}_{\text{PD}}) \log_2(D), \quad (24)$$

where  $\tau_c$  and  $\tau_p$  are the lengths of the coherence block and the pilot sequence, respectively. In the PD approach, the first  $N$  time slots are used for channel estimation (i.e.,  $\tau_p = N$ ), and we assume without loss of generality that the pilot signal is an all-ones vector. Assuming a DFT matrix of size  $N$  is used to configure the RIS phase shifts during the channel estimation phase, we have that  $\epsilon_i \sim \mathcal{CN}(0, \frac{N_0^{\text{BS}}}{PN})$ . The CSI errors  $\{\epsilon_i\}$  are also i.i.d. and uncorrelated with  $a_i$ . We will compute the PD bit error rate  $\text{BER}_{\text{PD}}$ , which requires the distribution of the effective channel  $f = \sum_{i=1}^N z_i$ .

We first obtain the following approximate means

$$\begin{aligned} \mu_{z_{i,\mathcal{R}}} &\triangleq \mathbb{E}[z_{i,\mathcal{R}}] \stackrel{(a)}{\approx} \frac{\mathbb{E}[|a_i|^2] + \Re\{\mathbb{E}[a_i \epsilon_i^*]\}}{\sqrt{\mathbb{E}[|a_i|^2] + 2\Re\{\mathbb{E}[a_i \epsilon_i^*]\} + \mathbb{E}[|\epsilon_i|^2]}} \\ &\stackrel{(b)}{=} \frac{\sigma_a^2}{\sqrt{\sigma_a^2 + \sigma_\epsilon^2}} \end{aligned} \quad (25)$$

$$\begin{aligned} \mu_{z_{i,\mathcal{I}}} &\triangleq \mathbb{E}[z_{i,\mathcal{I}}] \stackrel{(c)}{\approx} \frac{\Im\{\mathbb{E}[a_i \epsilon_i^*]\}}{\sqrt{\mathbb{E}[|a_i|^2] + 2\Re\{\mathbb{E}[a_i \epsilon_i^*]\} + \mathbb{E}[|\epsilon_i|^2]}} \\ &\stackrel{(d)}{=} 0 \end{aligned} \quad (26)$$

where the approximations (a) in (25) and (c) in (26) are obtained by applying a first-order Taylor expansion around the means  $\mathbb{E}[|a_i|^2]$ ,  $\mathbb{E}[|\epsilon_i|^2]$ , and  $\mathbb{E}[a_i \epsilon_i^*]$ . Similar approximations

were also used in [50]. The results (b) in (25) and (d) in (26) hold since  $\mathbb{E}[a_i \epsilon_i^*] = 0$ . Note here that we denote  $\sigma_a^2 \triangleq \mathbb{E}[|a_i|^2]$  and  $\sigma_\epsilon^2 \triangleq \mathbb{E}[|\epsilon_i|^2]$ . Similarly, we can obtain the following variances

$$\sigma_{z_{i,\mathcal{R}}}^2 \triangleq \text{Var}[z_{i,\mathcal{R}}] = \mathbb{E}[z_{i,\mathcal{R}}^2] - \mathbb{E}[z_{i,\mathcal{R}}]^2 \approx \frac{7\sigma_a^4 + \sigma_a^2 \sigma_\epsilon^2}{2(\sigma_a^2 + \sigma_\epsilon^2)} \quad (27)$$

$$\sigma_{z_{i,\mathcal{I}}}^2 \triangleq \text{Var}[z_{i,\mathcal{I}}] = \mathbb{E}[z_{i,\mathcal{I}}^2] - \mathbb{E}[z_{i,\mathcal{I}}]^2 \approx \frac{\sigma_a^2 \sigma_\epsilon^2}{2(\sigma_a^2 + \sigma_\epsilon^2)} \quad (28)$$

where  $\sigma_a^2 = \sigma_g^2 \sigma_h^2$  and we have used the following results:

$$\mathbb{E}[a_{i,\mathcal{R}}^2] = \mathbb{E}[(h_{i,\mathcal{R}} g_{i,\mathcal{R}} - h_{i,\mathcal{I}} g_{i,\mathcal{I}})^2] = \frac{1}{2} \sigma_h^2 \sigma_g^2 = \frac{1}{2} \sigma_a^2,$$

$$\mathbb{E}[a_{i,\mathcal{R}}^4] = \mathbb{E}[(h_{i,\mathcal{R}} g_{i,\mathcal{R}} - h_{i,\mathcal{I}} g_{i,\mathcal{I}})^4] = \frac{3}{2} \sigma_h^4 \sigma_g^4 = \frac{3}{2} \sigma_a^4.$$

It can be seen that the means and variances  $\mu_{z_{i,\mathcal{R}}}$ ,  $\mu_{z_{i,\mathcal{I}}}$ ,  $\sigma_{z_{i,\mathcal{R}}}^2$ , and  $\sigma_{z_{i,\mathcal{I}}}^2$  above are the same for different indices  $i$ . Therefore, for convenience in the rest of the PD-SE analysis, we drop the index  $i$  for these values.

Since the  $\{a_i\}$  and  $\{\epsilon_i\}$  are i.i.d., then the  $\{z_i\}$  are also i.i.d.. Using the central-limit theorem, for large  $N$  we have  $f_{\mathcal{R}} = \sum_{i=1}^N z_{i,\mathcal{R}} \sim \mathcal{N}(N\mu_{z_{\mathcal{R}}}, N\sigma_{z_{\mathcal{R}}}^2)$  and  $f_{\mathcal{I}} = \sum_{i=1}^N z_{i,\mathcal{I}} \sim \mathcal{N}(N\mu_{z_{\mathcal{I}}}, N\sigma_{z_{\mathcal{I}}}^2)$ . Note that we have  $\text{Cov}[f_{\mathcal{R}}, f_{\mathcal{I}}] = 0$  since the  $\{z_i\}$  are i.i.d., and  $\text{Cov}[z_{i,\mathcal{R}}, z_{i,\mathcal{I}}] = 0$ .

Let  $\tilde{y} = y s^* = \sqrt{P} f + \tilde{n}$  be the rotated received signal, and define  $r_{\tilde{y}} = \sqrt{\tilde{y}_{\mathcal{R}}^2 + \tilde{y}_{\mathcal{I}}^2}$  and  $\theta_{\tilde{y}} = \angle(\tilde{y}) = \arctan(\tilde{y}_{\mathcal{I}}/\tilde{y}_{\mathcal{R}})$ . Then the joint pdf of  $r_{\tilde{y}}$  and  $\theta_{\tilde{y}}$  is given as

$$\begin{aligned} p(r_{\tilde{y}}, \theta_{\tilde{y}}) &= \frac{r_{\tilde{y}}}{2\pi \sqrt{(PN\sigma_{z_{\mathcal{R}}}^2 + N_0/2)(PN\sigma_{z_{\mathcal{I}}}^2 + N_0/2)}} \\ &\times \exp \left\{ -\frac{1}{2} \left[ \frac{(r_{\tilde{y}} \cos \theta_{\tilde{y}} - \sqrt{PN} \mu_{z_{\mathcal{R}}})^2}{PN\sigma_{z_{\mathcal{R}}}^2 + N_0/2} + \frac{(r_{\tilde{y}} \sin \theta_{\tilde{y}})^2}{PN\sigma_{z_{\mathcal{I}}}^2 + N_0/2} \right] \right\}. \end{aligned} \quad (29)$$

Since the data symbols are equally likely, to compute the BER, we can assume that any one of the data symbols was transmitted, and we choose  $\mathcal{S}(0)$ . The probability that the detected symbol is  $\mathcal{S}(\ell)$  given  $\mathcal{S}(0)$  was transmitted is

$$\begin{aligned} p_\ell^{\text{PD}} &\triangleq \mathbb{P}[\hat{s} = \mathcal{S}(\ell) \mid s = \mathcal{S}(0)] \\ &= \int_0^\infty \int_{\frac{(2\ell-1)\pi}{D}}^{\frac{(2\ell+1)\pi}{D}} p(r_{\tilde{y}}, \theta_{\tilde{y}}) d\theta_{\tilde{y}} dr_{\tilde{y}}. \end{aligned} \quad (30)$$

Thus, the BER is given as

$$\text{BER}_{\text{PD}} = \frac{1}{\log_2(D)} \sum_{\ell=1}^D p_\ell^{\text{PD}} e_{\text{bit}}(0, \ell), \quad (31)$$

where  $e_{\text{bit}}(0, \ell)$  is the number of bit differences between symbols  $\mathcal{S}(0)$  and  $\mathcal{S}(\ell)$ .

*Theorem 1:* At high SNR,  $\text{BER}_{\text{PD}}$  can be approximated as

$$\begin{aligned} \text{BER}_{\text{PD}} &\approx \frac{2}{\log_2(D)} Q \left( \frac{\pi \sqrt{PN} \sigma_a \tan \theta}{4 \sqrt{(PN(1 - \frac{\pi^2}{16}) \sigma_a^2 + \frac{N_0}{2}) \tan^2 \theta + \frac{N_0}{2}}} \right). \end{aligned} \quad (32)$$

*Proof:* See Appendix A.  $\square$



### B. Decision-Directed

The SE of the decision-directed approach is given as

$$\text{SE}_{\text{DD}} = \left( \frac{\tau_{\text{d},1}(1 - \text{BER}_{\text{DD1}}) + \tau_{\text{d},2}(1 - \text{BER}_{\text{DD2}})}{\tau_{\text{c}}} \right) \log_2(D) \quad (33)$$

where  $\tau_{\text{d},1}$  and  $\text{BER}_{\text{DD1}}$  are the data transmission length and the BER in the channel estimation stage. Similarly,  $\tau_{\text{d},2}$  and  $\text{BER}_{\text{DD2}}$  are the data transmission length and the BER in the data transmission stage. Thus, for the DD spectral analysis, we need to compute  $\text{BER}_{\text{DD1}}$  and  $\text{BER}_{\text{DD2}}$  to obtain  $\text{SE}_{\text{DD}}$ . While  $\text{BER}_{\text{DD1}}$  is simple to obtain and can be computed exactly, obtaining an exact value  $\text{BER}_{\text{DD2}}$  is much more challenging and thus we provide an accurate approximation.

To simplify the analysis, we assume the RIS has only one active receiver element, which we take to be element  $N$ . We further assume that this element completely absorbs the incoming signal power during the channel estimation stage and is then turned off in the data transmission stage. In the first time slot, a pilot signal  $s_1 = 1$  is transmitted to generate the following received signal at the RIS:  $y_1^{\text{RIS}} = \sqrt{P}g_N + n_1^{\text{RIS}}$ , and so an estimate of  $g_N$  can be obtained as  $\hat{g}_N = y_1^{\text{RIS}}/\sqrt{P} = g_N + \frac{1}{\sqrt{P}}n_1^{\text{RIS}}$ . From the second time slot to the  $(N-1)$ -th time slot, data symbols are transmitted and the received signal at the RIS is  $y_t^{\text{RIS}} = \sqrt{P}g_N s_t + n_t^{\text{RIS}}$ ,  $t = 2, \dots, N-1$ . An equalizer based on  $\hat{g}_N$  is used by the RIS to detect the symbols as

$$\tilde{s}_t = \frac{y_t^{\text{RIS}}}{\sqrt{P}\hat{g}_N} = \frac{\sqrt{P}g_N s_t + n_t^{\text{RIS}}}{\sqrt{P}g_N + n_1^{\text{RIS}}}, \quad t = 2, \dots, N-1 \quad (34)$$

To evaluate the BER of (34), we consider the rotated signal

$$\bar{s}_t = \tilde{s}_t s_t^* = \frac{\sqrt{P}g_N + \tilde{n}_t^{\text{RIS}}}{\sqrt{P}g_N + n_1^{\text{RIS}}}. \quad (35)$$

Since the error rates are the same for different time indices  $t$ , we drop the subscript  $t$  for convenience. Let  $r_{\bar{s}} = \sqrt{\bar{s}_{\Re}^2 + \bar{s}_{\Im}^2}$  and  $\theta_{\bar{s}} = \arctan(\bar{s}_{\Im}/\bar{s}_{\Re})$ , whose joint pdf can be obtained as follows [51]:

$$p(r_{\bar{s}}, \theta_{\bar{s}}) = \frac{r_{\bar{s}}(\sigma_{y^{\text{RIS}}}^4 - P^2\sigma_g^4)}{\pi\sigma_{y^{\text{RIS}}}^8} \left( \frac{1 + r_{\bar{s}}^2}{\sigma_{y^{\text{RIS}}}^2} - \frac{2P\sigma_g^2 r_{\bar{s}} \cos \theta_{\bar{s}}}{\sigma_{y^{\text{RIS}}}^4} \right)^{-2},$$

where  $\sigma_{y^{\text{RIS}}}^2 = P\sigma_g^2 + N_0^{\text{RIS}}$ . The probability that the detected symbol is  $\mathcal{S}(\ell)$  given that  $\mathcal{S}(0)$  was transmitted is

$$p_{\ell}^{\text{DD1}} = \int_0^{\infty} \int_{\frac{(2\ell-1)\pi}{D}}^{\frac{(2\ell+1)\pi}{D}} p(r_{\bar{s}}, \theta_{\bar{s}}) d\theta_{\bar{s}} dr_{\bar{s}}. \quad (36)$$

Similar to (31), we can compute  $\text{BER}_{\text{DD1}}$  using  $p_{\ell}^{\text{DD1}}$  in (36).

Next we compute  $\text{BER}_{\text{DD2}}$ . Let  $\hat{\mathbf{s}} = [1, \hat{s}_2, \dots, \hat{s}_{N-1}]^T$  be the detected data symbol vector forwarded to the BS by the RIS, where  $\hat{s}_t = \arg \min_{s \in \mathcal{S}} |s - \tilde{s}_t|^2$ . Let  $\Phi = \mathbf{V}_{N-1}$  be the phase shift matrix of the RIS during the channel estimation stage. The BS uses  $\hat{\mathbf{s}}$  to obtain an estimate of the cascaded channel as follows:

$$\hat{a}_i^* = \frac{1}{N-1} \mathbf{a}^H \Phi \text{diag}(\mathbf{s}) \text{diag}(\hat{\mathbf{s}})^{-1} \Phi_{i,:}^H + \tilde{n}_i^* \quad (37)$$

$$= a_i^* + \underbrace{\frac{1}{1-N} \mathbf{a}^H \Phi \text{diag}(\boldsymbol{\xi}) \Phi_{i,:}^H}_{\epsilon_i^*} + \tilde{n}_i^* \quad (38)$$

where  $\boldsymbol{\xi} = [\xi_1, \dots, \xi_{N-1}]^T$  with  $\xi_t = 1 - s_t \hat{s}_t^*$ . Note that  $\xi_1 = 0$  since  $s_1 = 1$  is a known pilot signal. The noise terms  $\tilde{n}_i^* = \frac{1}{\sqrt{P(N-1)}} \mathbf{n}^T \text{diag}(\hat{\mathbf{s}})^{-1} \Phi_{i,:}^H$  are i.i.d. for different indices  $i$  and the channel estimation error term is  $\epsilon_i^* = \delta_i^* + \tilde{n}_i^*$  where

$$\delta_i^* = \frac{1}{1-N} \mathbf{a}^H \Phi \text{diag}(\boldsymbol{\xi}) \Phi_{i,:}^H. \quad (39)$$

From (39), it can be seen that the channel estimation error depends on the data symbols detected at the RIS.

Our DD-SE analysis will rely on the results in Lemma 1 presented next.

*Lemma 1:* Let  $\mu_{\xi} \triangleq \mathbb{E}[\xi_t]$ ,  $\mu_{|\xi|^2} \triangleq \mathbb{E}[|\xi_t|^2]$ , and  $\mu_{\xi^2} \triangleq \mathbb{E}[(\xi_t)^2]$ , then we have

$$\mu_{\xi} = 1 - p_0^{\text{DD1}} + p_{\frac{D}{2}}^{\text{DD1}} - 2 \sum_{d=1}^{\frac{D}{2}-1} p_d^{\text{DD1}} \cos\left(\frac{2\pi d}{D}\right), \quad (40)$$

$$\mu_{\xi^2} = 1 - p_0^{\text{DD1}} + 3p_{\frac{D}{2}}^{\text{DD1}} + 2 \sum_{d=1}^{\frac{D}{2}-1} p_d^{\text{DD1}} \left[ \cos\left(\frac{4\pi d}{D}\right) - 2 \cos\left(\frac{2\pi d}{D}\right) \right], \quad (41)$$

$$\mu_{|\xi|^2} = 4p_{\frac{D}{2}}^{\text{DD1}} + 4 \sum_{d=1}^{\frac{D}{2}-1} p_d^{\text{DD1}} \left[ 1 - \cos\left(\frac{2\pi d}{D}\right) \right]. \quad (42)$$

*Proof:* See Appendix B.  $\square$

To obtain the  $\text{BER}_{\text{DD2}}$ , we will compute the means and variances of the real and imaginary parts of the effective channel  $f = \sum_{i=1}^{N-1} z_i$  during the data transmission stage, which are  $\mu_{f_{\Re}} \triangleq \mathbb{E}[f_{\Re}]$  and  $\mu_{f_{\Im}} \triangleq \mathbb{E}[f_{\Im}]$ ,  $\sigma_{f_{\Re}}^2 \triangleq \text{Var}[f_{\Re}]$  and  $\sigma_{f_{\Im}}^2 \triangleq \text{Var}[f_{\Im}]$ .

1) *Computing the Means  $\mu_{f_{\Re}}$  and  $\mu_{f_{\Im}}$ :* Since  $\mu_{f_{\Re}} = \sum_{i=1}^{N-1} \mu_{z_{i,\Re}}$  and  $\mu_{f_{\Im}} = \sum_{i=1}^{N-1} \mu_{z_{i,\Im}}$ , we need to compute  $\mu_{z_{i,\Re}}$  and  $\mu_{z_{i,\Im}}$ . Similar to (25) and (26), we use a first-order Taylor expansion to obtain an approximation of  $\mu_{z_{i,\Re}}$  and  $\mu_{z_{i,\Im}}$  as follows:

$$\mu_{z_{i,\Re}} \approx \frac{\mathbb{E}[|a_i|^2] + \Re\{\mathbb{E}[a_i \epsilon_i^*]\}}{\sqrt{\mathbb{E}[|a_i|^2] + 2\Re\{\mathbb{E}[a_i \epsilon_i^*]\} + \mathbb{E}[|\epsilon_i|^2]}}, \quad (43)$$

$$\mu_{z_{i,\Im}} \approx \frac{\Im\{\mathbb{E}[a_i \epsilon_i^*]\}}{\sqrt{\mathbb{E}[|a_i|^2] + 2\Re\{\mathbb{E}[a_i \epsilon_i^*]\} + \mathbb{E}[|\epsilon_i|^2]}}. \quad (44)$$

Since

$$\begin{aligned} \mathbb{E}[a_i \epsilon_i^*] &= \mathbb{E}[a_i \delta_i^*] = \frac{1}{1-N} \mathbb{E} \left[ \sum_{n=1}^{N-1} a_i a_n^* \Phi_{n,:} \text{diag}(\boldsymbol{\xi}) \Phi_{i,:}^H \right] \\ &\stackrel{(a)}{=} \frac{1}{1-N} \mathbb{E} \left[ |a_i|^2 \Phi_{i,:} \text{diag}(\boldsymbol{\xi}) \Phi_{i,:}^H \right] \\ &= \frac{\sigma_a^2}{1-N} \sum_{t=2}^{N-1} \mathbb{E}[\xi_t] = \frac{N-2}{1-N} \sigma_a^2 \mu_{\xi}, \end{aligned} \quad (45)$$

we have  $\mathbb{E}[\Re\{a_i \epsilon_i^*\}] = \frac{N-2}{1-N} \sigma_a^2 \mu_{\xi}$  and  $\mathbb{E}[\Im\{a_i \epsilon_i^*\}] = 0$ . The equality (a) in (45) holds because  $\mathbb{E}[a_i a_n^*] = 0$  for all  $i \neq n$ .



Substituting  $\mathbb{E}[\Re\{a_i \epsilon_i^*\}] = \frac{N-2}{1-N} \sigma_a^2 \mu_\xi$  and  $\mathbb{E}[\Im\{a_i \epsilon_i^*\}] = 0$  into (43) and (44) gives us the mean of  $z_{i,\Re}$  and  $z_{i,\Im}$  as follows:

$$\mu_{z_{i,\Re}} \approx \frac{\sigma_a^2 + \frac{N-2}{1-N} \sigma_a^2 \mu_\xi}{\sqrt{\sigma_a^2 + 2 \frac{N-2}{1-N} \sigma_a^2 \mu_\xi + \sigma_\epsilon^2}} \text{ and } \mu_{z_{i,\Im}} \approx 0, \quad (46)$$

where  $\sigma_a^2 \triangleq \mathbb{E}[|a|^2]$  and  $\sigma_\epsilon^2 \triangleq \mathbb{E}[|\epsilon_i|^2] = \sigma_{\delta_i}^2 + \sigma_{\tilde{n}_i}^2$ . To obtain  $\sigma_\epsilon^2$ , we need to find  $\sigma_{\delta_i}^2$  and  $\sigma_{\tilde{n}_i}^2$ , which represent the variance of  $\delta_i$  and  $\tilde{n}_i$ , respectively. The variances  $\sigma_{\delta_i}^2$  and  $\sigma_{\tilde{n}_i}^2$  are given as follows:

$$\begin{aligned} \sigma_{\delta_i}^2 &= \frac{\sigma_a^2}{(N-1)} \text{tr} \left( \mathbb{E} \left[ \text{diag}(\boldsymbol{\xi}) \boldsymbol{\Phi}_{i,:}^H \boldsymbol{\Phi}_{i,:} \text{diag}(\boldsymbol{\xi}^*) \right] \right) \\ &= \frac{\sigma_a^2}{(N-1)} \sum_{t=2}^{N-1} \mathbb{E}[|\xi_t|^2] = \frac{N-2}{(N-1)} \sigma_a^2 \mu_{|\xi|^2} \end{aligned}$$

and  $\sigma_{\tilde{n}_i}^2 = \frac{N_0^{\text{BS}}}{P(N-1)}$ . Hence, the variance of  $\epsilon$  is

$$\sigma_\epsilon^2 = \frac{P(N-2) \sigma_a^2 \mu_{|\xi|^2} + N_0^{\text{BS}}}{P(N-1)}. \quad (47)$$

2) *Compute the Variances  $\sigma_{f_{\Re}}^2$  and  $\sigma_{f_{\Im}}^2$* : The variance of  $f_{\Re}$  and  $f_{\Im}$  are

$$\sigma_{f_{\Re}}^2 = \sum_{i=1}^{N-1} \text{Var}[z_{i,\Re}] + 2 \sum_{i < t} \text{Cov}[z_{i,\Re}, z_{t,\Re}], \quad (48)$$

$$\sigma_{f_{\Im}}^2 = \sum_{i=1}^{N-1} \text{Var}[z_{i,\Im}] + 2 \sum_{i < t} \text{Cov}[z_{i,\Im}, z_{t,\Im}]. \quad (49)$$

Since  $\text{Var}[z_{i,\Re}] = \mathbb{E}[z_{i,\Re}^2] - \mathbb{E}[z_{i,\Re}]^2$  and  $\text{Var}[z_{i,\Im}] = \mathbb{E}[z_{i,\Im}^2] - \mathbb{E}[z_{i,\Im}]^2$ , and since  $\mathbb{E}[z_{i,\Re}]$  and  $\mathbb{E}[z_{i,\Im}]$  are already given in (46), we now need to obtain  $\mathbb{E}[z_{i,\Re}^2]$  and  $\mathbb{E}[z_{i,\Im}^2]$ . Similar to (25) and (26), we use a first-order Taylor expansion to obtain an approximation of  $\mathbb{E}[z_{i,\Re}^2]$  and  $\mathbb{E}[z_{i,\Im}^2]$  as follows:

$$\mathbb{E}[z_{i,\Re}^2] \approx \frac{\mathbb{E}[|a_i|^4] + 2\mathbb{E}[|a_i|^2 \Re\{a_i \epsilon_i^*\}] + \mathbb{E}[\Re\{a_i \epsilon_i^*\}^2]}{\mathbb{E}[|a_i|^2] + 2\Re\{\mathbb{E}[a_i \epsilon_i^*]\} + \mathbb{E}[|\epsilon_i|^2]}, \quad (50)$$

$$\mathbb{E}[z_{i,\Im}^2] \approx \frac{\mathbb{E}[\Im\{a_i \epsilon_i^*\}^2]}{\mathbb{E}[|a_i|^2] + 2\Re\{\mathbb{E}[a_i \epsilon_i^*]\} + \mathbb{E}[|\epsilon_i|^2]}. \quad (51)$$

The first term in the numerator of (50) is  $\mathbb{E}[|a_i|^4] = 4\sigma_a^4$ . The second term in the numerator of (50) can be obtained by computing  $\mathbb{E}[|a_i|^2 a_i \epsilon_i^*]$  since  $\mathbb{E}[|a_i|^2 \Re\{a_i \epsilon_i^*\}] = \Re\{\mathbb{E}[|a_i|^2 a_i \epsilon_i^*]\}$ . We have

$$\begin{aligned} \mathbb{E}[|a_i|^2 a_i \epsilon_i^*] &= \mathbb{E}[|a_i|^2 a_i \delta_i^*] \\ &= \frac{1}{1-N} \mathbb{E} \left[ \sum_{n=1}^{N-1} |a_i|^2 a_n^* \boldsymbol{\Phi}_{n,:} \text{diag}(\boldsymbol{\xi}) \boldsymbol{\Phi}_{i,:}^H \right] \\ &\stackrel{(a)}{=} \frac{1}{1-N} \mathbb{E} \left[ |a_i|^4 \boldsymbol{\Phi}_{i,:} \text{diag}(\boldsymbol{\xi}) \boldsymbol{\Phi}_{i,:}^H \right] \\ &= \frac{1}{1-N} 4\sigma_a^4 \sum_{t=2}^{N-1} \mathbb{E}[\xi_t] = \frac{N-2}{1-N} 4\sigma_a^4 \mu_\xi, \end{aligned} \quad (52)$$

which is a real number, and so  $\mathbb{E}[|a_i|^2 \Re\{a_i \epsilon_i^*\}] = \frac{N-2}{1-N} 4\sigma_a^4 \mu_\xi$ . The equality (a) in (52) holds because  $\mathbb{E}[|a_i|^2 a_n^*] = 0$  for all  $i \neq n$ .

We exploit the following relations:

$$|a_i \epsilon_i^*|^2 = \Re\{a_i \epsilon_i^*\}^2 + \Im\{a_i \epsilon_i^*\}^2, \quad (53)$$

$$\Re\{(a_i \epsilon_i^*)^2\} = \Re\{a_i \epsilon_i^*\}^2 - \Im\{a_i \epsilon_i^*\}^2, \quad (54)$$

to obtain  $\mathbb{E}[\Re\{a_i \epsilon_i^*\}^2]$  (the last term in the numerator of (50)) and  $\mathbb{E}[\Im\{a_i \epsilon_i^*\}^2]$  (the numerator term of (51)) as follows:

$$\mathbb{E}[\Re\{a_i \epsilon_i^*\}^2] = \frac{\mathbb{E}[|a_i \epsilon_i^*|^2] + \Re\{\mathbb{E}[(a_i \epsilon_i^*)^2]\}}{2}, \quad (55)$$

$$\mathbb{E}[\Im\{a_i \epsilon_i^*\}^2] = \frac{\mathbb{E}[|a_i \epsilon_i^*|^2] - \Re\{\mathbb{E}[(a_i \epsilon_i^*)^2]\}}{2}. \quad (56)$$

To retrieve  $\mathbb{E}[\Re\{a_i \epsilon_i^*\}^2]$  and  $\mathbb{E}[\Im\{a_i \epsilon_i^*\}^2]$  in (55) and (56), we need to find  $\mathbb{E}[|a_i \epsilon_i^*|^2]$  and  $\mathbb{E}[(a_i \epsilon_i^*)^2]$ . The term  $\mathbb{E}[|a_i \epsilon_i^*|^2]$  is given by

$$\begin{aligned} \mathbb{E}[|a_i \epsilon_i^*|^2] &= \mathbb{E}[|a_i \delta_i^*|^2] + \mathbb{E}[|a_i \tilde{n}_i^*|^2] \\ &= \mathbb{E}[|a_i \delta_i^*|^2] + \frac{\sigma_a^2 N_0^{\text{BS}}}{P(N-1)} \end{aligned} \quad (57)$$

where

$$\begin{aligned} \mathbb{E}[|a_i \delta_i^*|^2] &= \frac{\mathbb{E} \left[ |a_i \mathbf{a}^H \boldsymbol{\Phi} \text{diag}(\boldsymbol{\xi}) \boldsymbol{\Phi}_{i,:}^H|^2 \right]}{(N-1)^2} \\ &= \frac{\mathbb{E} \left[ \text{tr} \left\{ |a_i|^2 \boldsymbol{\Phi}^H \mathbf{a} \mathbf{a}^H \boldsymbol{\Phi} \text{diag}(\boldsymbol{\Phi}_{i,:}^H) \boldsymbol{\xi} \boldsymbol{\xi}^H \text{diag}(\boldsymbol{\Phi}_{i,:}) \right\} \right]}{(N-1)^2} \\ &\stackrel{(a)}{=} \frac{\text{tr} \left\{ \boldsymbol{\Phi}^H \text{diag}(\boldsymbol{\alpha}_i) \boldsymbol{\Phi} \text{diag}(\boldsymbol{\Phi}_{i,:}^H) \mathbf{R}_{\boldsymbol{\xi} \boldsymbol{\xi}^H} \text{diag}(\boldsymbol{\Phi}_{i,:}) \right\}}{(N-1)^2} \\ &\stackrel{(b)}{=} \frac{N-2}{(N-1)^2} \sigma_a^4 ((N+2) \mu_{|\xi|^2} + 3(N-3) \mu_\xi^2). \end{aligned} \quad (58)$$

The equality (a) in (58) is obtained by defining  $\mathbf{R}_{\boldsymbol{\xi} \boldsymbol{\xi}^H} = \mathbb{E}[\boldsymbol{\xi} \boldsymbol{\xi}^H]$  and letting  $\boldsymbol{\alpha}_i$  be a vector whose  $i$ th-element is  $4\sigma_a^4$  and whose other elements are  $\sigma_a^4$ . The equality (b) in (58) is obtained by simply calculating the trace in the equality (a) by noting that the diagonal elements of  $\mathbf{R}_{\boldsymbol{\xi} \boldsymbol{\xi}^H}$  are  $\mu_{|\xi|^2}$  and the off-diagonal elements are approximately  $\mu_\xi$ .

The term  $\mathbb{E}[(a_i \epsilon_i^*)^2]$  is given by

$$\begin{aligned} \mathbb{E}[(a_i \epsilon_i^*)^2] &= \mathbb{E}[(a_i \delta_i^*)^2] \\ &= \frac{\mathbb{E} \left[ \left( a_i \mathbf{a}^H \boldsymbol{\Phi} \text{diag}(\boldsymbol{\xi}) \boldsymbol{\Phi}_{i,:}^H \right)^2 \right]}{(N-1)^2} \\ &= \frac{\mathbb{E} \left[ \text{tr} \left\{ a_i^2 \text{diag}(\boldsymbol{\Phi}_{i,:}^*) \boldsymbol{\Phi}^T \mathbf{a}^* \mathbf{a}^H \boldsymbol{\Phi} \text{diag}(\boldsymbol{\Phi}_{i,:}^H) \boldsymbol{\xi} \boldsymbol{\xi}^T \right\} \right]}{(N-1)^2} \\ &\stackrel{(a)}{=} \frac{\text{tr} \left\{ 4\sigma_a^4 \text{diag}(\boldsymbol{\Phi}_{i,:}^*) \boldsymbol{\Phi}^T \text{diag}(\mathbf{e}_i) \boldsymbol{\Phi} \text{diag}(\boldsymbol{\Phi}_{i,:}^H) \mathbf{R}_{\boldsymbol{\xi} \boldsymbol{\xi}^T} \right\}}{(N-1)^2} \\ &\stackrel{(b)}{=} \frac{N-2}{(N-1)^2} 4\sigma_a^4 (\mu_{\xi^2} + (N-3) \mu_\xi^2). \end{aligned} \quad (59)$$

Similar to (58), the equality (a) in (59) is obtained by defining  $\mathbf{R}_{\boldsymbol{\xi} \boldsymbol{\xi}^T} = \mathbb{E}[\boldsymbol{\xi} \boldsymbol{\xi}^T]$  and letting  $\mathbf{e}_i$  be a one-hot vector whose  $i$ th-element is 1 and whose other elements are zero. The equality (b) in (59) is obtained by simply calculating the trace in the equality (a) by noting that the diagonal

elements of  $\mathbf{R}_{\xi\xi^H}$  are  $\mu_{\xi^2}$  and the off-diagonal elements are approximately  $\mu_{\xi^2}$ .

We now need to compute the covariances  $\text{Cov}[z_{i,\mathcal{R}}, z_{t,\mathcal{R}}]$  and  $\text{Cov}[z_{i,\mathcal{S}}, z_{t,\mathcal{S}}]$ , which are given by

$$\text{Cov}[z_{i,\mathcal{R}}, z_{t,\mathcal{R}}] = \mathbb{E}[z_{i,\mathcal{R}} z_{t,\mathcal{R}}] - \mathbb{E}[z_{i,\mathcal{R}}] \mathbb{E}[z_{t,\mathcal{R}}]$$

$$\text{Cov}[z_{i,\mathcal{S}}, z_{t,\mathcal{S}}] = \mathbb{E}[z_{i,\mathcal{S}} z_{t,\mathcal{S}}] - \mathbb{E}[z_{i,\mathcal{S}}] \mathbb{E}[z_{t,\mathcal{S}}].$$

Since  $\mathbb{E}[z_{i,\mathcal{R}}]$  and  $\mathbb{E}[z_{i,\mathcal{S}}]$  have been computed earlier, we need to find  $\mathbb{E}[z_{i,\mathcal{R}} z_{t,\mathcal{R}}]$  and  $\mathbb{E}[z_{i,\mathcal{S}} z_{t,\mathcal{S}}]$  which can be approximated as follows:

$$\mathbb{E}[z_{i,\mathcal{R}} z_{t,\mathcal{R}}] \approx \frac{\mathbb{E}[(|a_i|^2 + \Re\{a_i \epsilon_i^*\})(|a_t|^2 + \Re\{a_t \epsilon_t^*\})]}{\sqrt{\mathbb{E}[\mathcal{X}_{i,t}]}} \quad (61)$$

$$\mathbb{E}[z_{i,\mathcal{S}} z_{t,\mathcal{S}}] \approx \frac{\mathbb{E}[\Im\{a_i \epsilon_i^*\} \Im\{a_t \epsilon_t^*\}]}{\sqrt{\mathbb{E}[\mathcal{X}_{i,t}]}} \quad (62)$$

where  $\mathcal{X}_{i,t}$  is given in (60), as shown at the bottom of the page. The expansion of  $\mathcal{X}_{i,t}$  also includes the numerator terms in (61) and (62), thus to obtain  $\mathbb{E}[z_{i,\mathcal{R}} z_{t,\mathcal{R}}]$  and  $\mathbb{E}[z_{i,\mathcal{S}} z_{t,\mathcal{S}}]$ , it suffices to compute the expectation of all the terms in (60), which are given in detail in Appendix B.

Finally, we approximate the distribution of  $f_{\mathcal{R}}$  and  $f_{\mathcal{S}}$  as Gaussian with the above approximate means and variances, and so  $\text{BER}_{\text{DD2}}$  can be obtained in the same manner as in (29) – (31).

## VI. NUMERICAL RESULTS

In this section, we present various numerical results to verify our SE analysis and to show the benefits of the proposed DD channel estimation framework. We use a general channel model with  $\mathbf{h}_{d,k} = \sqrt{\beta_k^{\text{UB}}} \tilde{\mathbf{h}}_{d,k}$ ,  $\mathbf{g}_k = \sqrt{\beta_k^{\text{UR}}} \tilde{\mathbf{g}}_k$ ,  $\mathbf{H} = \sqrt{\beta_k^{\text{RB}}} \tilde{\mathbf{H}}$  where  $\tilde{\mathbf{h}}_{d,k} \sim \mathcal{CN}(\mathbf{0}, \Sigma_k^{\text{UB}})$ ,  $\tilde{\mathbf{g}}_k \sim \mathcal{CN}(\mathbf{0}, \Sigma_k^{\text{UR}})$ , and  $\tilde{\mathbf{H}} = (\Sigma^{\text{B}})^{1/2} \tilde{\mathbf{H}} (\Sigma^{\text{R}})^{1/2}$ . In our simulations, we employ the simple exponential spatial correlation model [52]; more general correlation models can be found in [53]. The elements of  $\tilde{\mathbf{H}}$  are i.i.d. and normally distributed as  $\mathcal{CN}(0,1)$ . The large-scale fading coefficients are  $\beta_k^{\text{UB}} = \beta_0 (d_k^{\text{UB}}/d_0)^{-\alpha^{\text{UB}}}$ ,  $\beta_k^{\text{UR}} = \beta_0 (d_k^{\text{UR}}/d_0)^{-\alpha^{\text{UR}}}$ ,  $\beta_k^{\text{RB}} = \beta_0 (d_k^{\text{RB}}/d_0)^{-\alpha^{\text{RB}}}$  where  $\alpha^{\text{UB}}$ ,  $\alpha^{\text{UR}}$ ,  $\alpha^{\text{RB}}$  are the respective path loss exponents, and  $d_k^{\text{UB}}$ ,  $d_k^{\text{UR}}$ ,  $d_k^{\text{RB}}$  are the respective distances between user- $k$  and the BS, user- $k$  and the RIS, and the RIS and BS. We set  $\beta_0 = -20$  dB as the path loss at the reference distance  $d_0 = 1$  m,  $\alpha^{\text{UB}} = 5$ ,  $\alpha^{\text{UR}} = \alpha^{\text{RB}} = 2.2$ , and the coherence block length at  $\tau_c = 500$  symbols. If not specifically stated, the noise power is set to  $-169$  dBm/Hz and a bandwidth of 1 MHz is assumed. The number of elements with active receivers at the RIS is taken to be equal to the number of users, and the sensing elements are assumed to be uniformly distributed over the surface.<sup>4</sup>

<sup>4</sup>Given a number of sensing elements much less than the total number of the surface elements, it would be interesting topic for future work to find an optimal strategy for distributing the sensing elements based for example on the distribution of the channel.

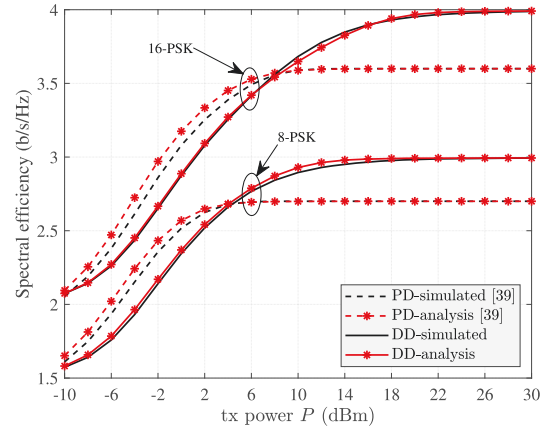


Fig. 3. SE comparison with  $K = M = 1$  and  $N = 50$ .

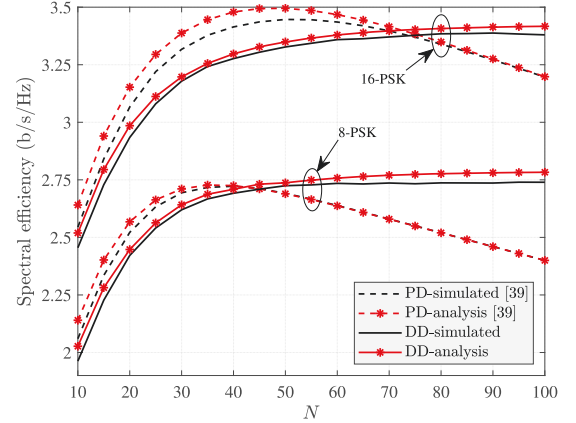


Fig. 4. SE comparison with  $K = M = 1$ ,  $N$  varies, and  $P = 5$  dBm.

It is important to note that when there is only one user, the proposed hybrid PD channel estimation approach coincides with the passive PD approach in [38] because only phase 1 is required for the channel estimation and we should set  $\rho_i^A = 1$ . Therefore, in case of  $K = 1$ , the PD performance presented in this section represents both the proposed hybrid PD approach and the passive one in [38]. However, for the case of multiple users, phase 2 of the two approaches is different and will lead to different performance as will be shown later.

First, we numerically validate our derived analytical SE expressions in Figs. 3 and 4. Here, we set both the user-RIS and RIS-BS distances to 100 m. In Fig. 3, the SE is evaluated for 50 RIS elements versus the transmit power for the cases of 8-PSK and 16-PSK modulation. It is observed that when the transmit power is low, i.e., at low signal-to-noise ratios (SNR), the SE of the PD approach is higher than for DD, but the situation reverses as the SNR grows. This is because the channel estimation performance of the DD approach strongly depends on the data detection performance at the RIS; at higher SNRs the data symbols detected by the RIS are more reliable and this results in better channel estimation and higher spectral efficiency. There is a critical SNR point at which the

$$\begin{aligned} \mathcal{X}_{i,t} = & |a_i|^2 |a_t|^2 + 2|a_i|^2 \Re\{a_t \epsilon_t^*\} + |a_i|^2 |\epsilon_t|^2 + 2|a_t|^2 \Re\{a_i \epsilon_i^*\} + 4\Re\{a_i \epsilon_i^*\} \Re\{a_t \epsilon_t^*\} \\ & + 2\Re\{a_i \epsilon_i^*\} |\epsilon_t|^2 + |\epsilon_i|^2 |a_t|^2 + 2|\epsilon_i|^2 \Re\{a_t \epsilon_t^*\} + |\epsilon_i|^2 |\epsilon_t|^2. \end{aligned} \quad (60)$$

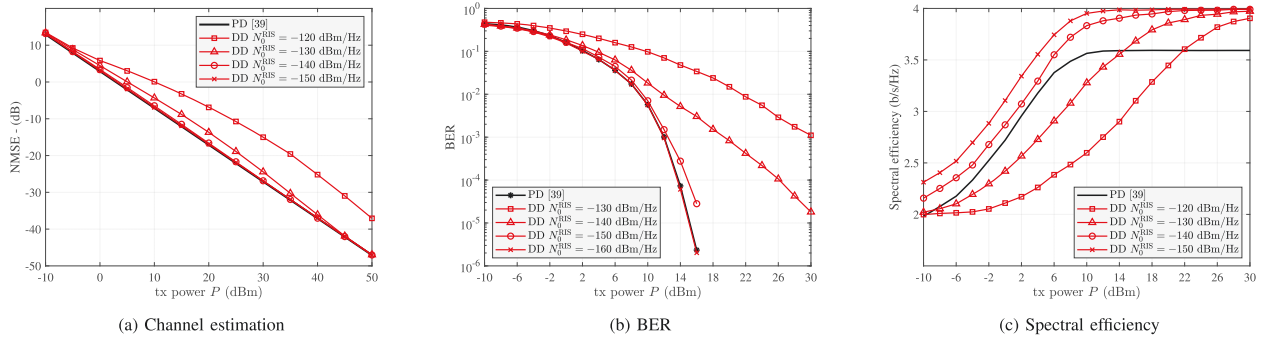


Fig. 5. Performance comparison for  $K = 1$ ,  $M = 8$ ,  $N = 50$ ,  $\rho_i^A = 0.5$ , and 16-PSK.

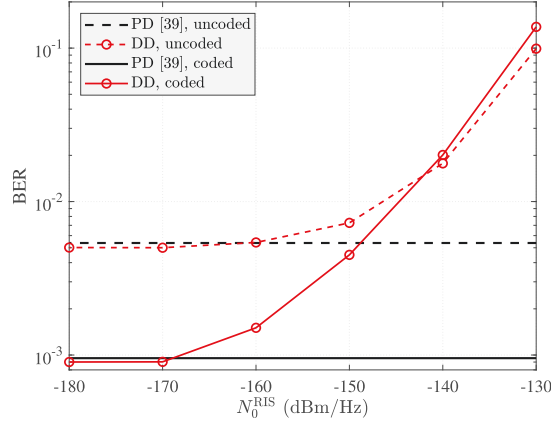


Fig. 6. Coded BER performance with  $P = 10$  dBm,  $K = 1$ ,  $M = 8$ ,  $N = 50$ , and 16-PSK signalling.

DD approach begins to perform better than PD, which in this scenario is about 4-dBm and 8-dBm for 8-PSK and 16-PSK, respectively. It can also be seen from Fig. 3 that our analytical SE approximations match well with the numerical results and accurately predict the performance crossover point. Thus the analytical SE can be used in the system design to determine the crossing point and decide whether the PD or DD approach should be used.

We evaluate the SE of the PD and DD approaches as the number of RIS elements  $N$  increases in Fig. 4, where the transmit power is fixed at 5-dBm. It is interesting to observe that increasing the number of RIS elements can actually lead to a reduction in SE for the PD framework, since the pilot overhead of the PD approach grows proportionally with  $N$  leading to a reduction in the number of time slots available for data transmission. On the other hand, the pilot overhead of the DD framework does not depend on the number of RIS elements, and thus the DD approach does not suffer from SE reduction as  $N$  increases. Again, our analysis accurately predicts the performance crossover point, which is an important factor for the system design.

In Figs. 5, 6, and 7 we also consider a single user scenario but the BS is equipped with multiple antennas. After the channel estimation stage, the phase shift vector  $\phi$  of the RIS is optimized to maximize the effective channel strength  $\|\hat{\mathbf{h}}_{d,1} + \hat{\mathbf{A}}_1 \phi\|^2$ , which is solved by semi-definite relaxation (SDR). Fig. 5 shows the channel estimation, bit error rate (BER), and spectral efficiency performance for  $M = 8$ ,  $N = 50$ ,  $\rho_i^A = 0.5$ , and 16-PSK signalling with different noise power

levels and user transmit powers. The normalized mean-squared error (NMSE) is computed as  $\mathbb{E}[\|\hat{\mathbf{H}}_{c,1} - \mathbf{H}_{c,1}\|_F^2 / \|\mathbf{H}_{c,1}\|_F^2]$ . The results in Fig. 5a show that the PD method achieves a better channel estimate and BER than DD, but this gain is offset by the increased training overhead for either higher transmit power or a lower noise figure when the DD method can reliably decode the data at the RIS.

In Fig. 6 we present a BER comparison between the PD and DD approaches considering both uncoded and coded systems for different noise levels at the sensing RIS. The coded system employs a rate 1/2 convolutional code. It can be seen that as the noise power at the sensing RIS decreases, the DD coded BER is significantly improved compared to the DD uncoded BER. They are also both slightly lower than the BER of the PD method since the data detected by the sensing RIS is more reliable in this noise regime.

Fig. 7 illustrates that there is a trade-off in the choice of the fraction of the incident power  $\rho_i^A$  that is reflected by the RIS elements with active receivers. A larger  $\rho_i^A$  means more signal power is reflected and less is sensed by the RIS. When the amount of signal power sensed by the RIS is too small, the noise at the RIS may dominate the received signal and cause data detection errors, which in turn leads to lower channel estimation accuracy and SE. On the other hand, if the amount of signal power sensed by the RIS is large so as to efficiently recover the data symbols at the RIS, the signals reflected from the RIS to the BS will be weaker, which can lead to less accurate channel estimation at the BS and a reduction in SE as well. The trade-off is not too serious to handle for small noise levels, but becomes more important as the SNR decreases. For the cases considered in this example, a relatively small value such as  $\rho_i^A = 0.2$  appears to provide the best system performance.

To study the case of multiple users, we position the RIS and BS at the locations  $(x, y) = (50, 50)$  and  $(x, y) = (100, 0)$ , respectively, and we locate the users randomly within a square whose side length is 20m and is centered at the origin. Simulation results for a scenario with  $K = 4$ ,  $M = 8$ ,  $N = 200$ ,  $\rho_i^A = 0.5$ , and 16-PSK signalling are given in Fig. 8. In the sub-phase 2b, we employ the conventional zero-forcing (ZF) detector for recovering data symbols at the RIS. To configure the RIS phase shift after the channel estimation stage, we find the  $\phi$  that maximizes the minimum signal-to-interference plus noise ratio (SINR) using the SDR approach as in [54]. An SE comparison is given in Fig. 8a

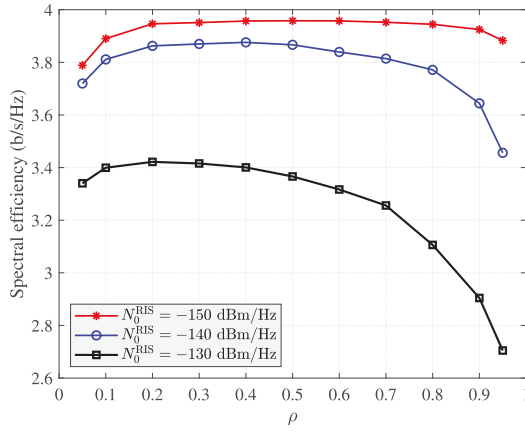
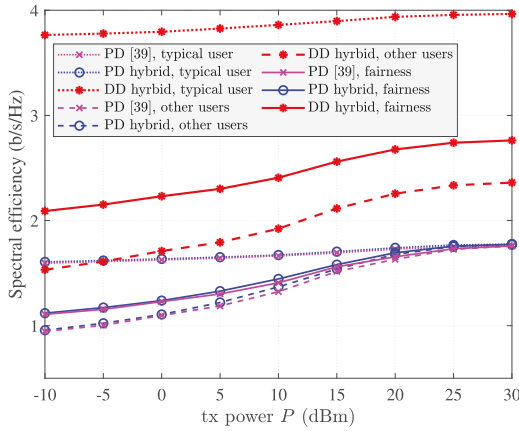
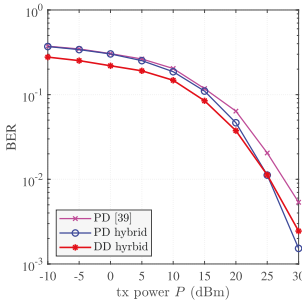


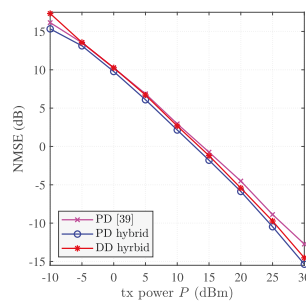
Fig. 7. Spectral efficiency of the DD approach as  $\rho_i^A$  varies,  $P = 10$  dBm,  $K = 1$ ,  $M = 8$ ,  $N = 50$ , and 16-PSK signalling.



(a) Spectral efficiency



(b) Fairness BER



(c) Fairness NMSE

Fig. 8. Performance comparison with  $K = 4$ ,  $M = 8$ ,  $N = 200$ ,  $\rho_i^A = 0.5$ , and 16-PSK.

and the corresponding BER and NMSE performance is given in Figs. 8b and 8c, respectively. It is seen from Fig. 8a that the SE of the DD method is much higher than that for PD because data is also sent in the channel estimation stage of the DD method while only pilot signals are used in the PD approaches.

It is also observed in Fig. 8a that the SE of the typical user is much higher than that of the other users since the typical user has more data transmission slots, as can be seen from the illustration in Fig. 2. This creates a fairness issue that can be addressed in a number of ways. For example, the solid curves show the result obtained by rotating the role of the typical user among all the users over different coherence blocks, e.g., user  $k = b - K \lfloor (b-1)/K \rfloor$  is assigned to be the typical user

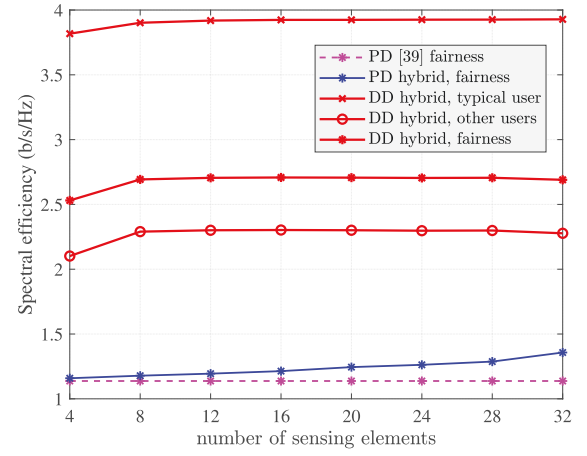


Fig. 9. Spectral efficiency comparison versus number of sensing elements  $N_A$  with  $K = 4$ ,  $M = 4$ ,  $N = 200$ ,  $\rho_i^A = 0.5$ , 16-PSK signalling, transmit power  $P = 30$  dBm.

in coherence block  $b$  for  $b = 1, 2, \dots$ . In this approach, the average SE of all the users will be the same and is given as

$$SE_{\text{fair}} = \frac{SE_{\text{typ}} + (K-1)SE_{\text{oth}}}{K} \quad (63)$$

where  $SE_{\text{fair}}$ ,  $SE_{\text{typ}}$ , and  $SE_{\text{oth}}$  denote the “fair” SE, the SE of the typical, and the SE of the other users, respectively. The SE of the typical and other users  $SE_{\text{typ}}$  and  $SE_{\text{oth}}$  are computed as

$$SE_U = \frac{\tau_c - \tau_{t,U}}{\tau_c} (1 - \text{BER}_U) \log_2(D), \quad (64)$$

where  $U \in \{\text{typ}, \text{oth}\}$ ,  $\tau_{t,U}$  is the training overhead of the corresponding user, and  $\text{BER}_U$  is obtained numerically. We see from Fig. 8 that an improvement compared with the unbalanced case is obtained. In particular, the fair DD approach yields approximately a 60% improvement in SE performance compared to the PD approaches. In terms of BER, the performance of the proposed PD and DD approaches is better than that of the passive PD in [38]. This is because the DD method obtains reliable data detection at the sensing RIS. In addition, the channel estimate of the proposed PD and DD methods is also observed to be more accurate than the passive PD approach in [38] as can be seen in Fig. 8c since in phase 2 of the proposed hybrid PD and DD methods, the channels associated with the RIS sensing elements are estimated directly at the RIS. Note that the fair BER and NMSE in Figs. 8b and 8c are computed in a manner similar to (63).

Finally, we study the effect of the number of sensing elements  $N_A$  on the spectral efficiency in Fig. 9. The noise power at the RIS is set to  $-120$  dBm/Hz and the transmit power  $P$  is 30 dBm. Interestingly, increasing the number of sensing elements  $N_A$  only slightly improves the spectral efficiency of the DD approach, indicating that very few sensing elements at the RIS are necessary to achieve the benefit of decision direction. The SE improves more with increasing  $N_A$  for the proposed hybrid PD approach, since unlike the passive PD algorithm in [38], increasing  $N_A$  results in a reduction in the pilot overhead of the hybrid PD approach.



## VII. CONCLUSION

In this paper, we have proposed a decision-directed channel estimation framework for general unstructured RIS channel models. It has been shown that with the help of some RIS elements with active receivers, it is possible to accurately estimate the CSI with a pilot overhead only proportional to the number of users and thus significantly improve the spectral efficiency compared to systems with passive RIS arrays. We also performed an intensive spectral efficiency analysis to verify the efficiency of the proposed DD framework. Our analysis takes into account both the channel estimation and data detection errors at both the RIS and the BS, and thus accurately reflects the data detection uncertainty inherent in the decision directed approach.

### APPENDIX A PROOF OF THEOREM 1

The symbol error rate (SER) can be approximated as

$$\text{SER} \approx \mathbb{P}[\tilde{y}_{\mathcal{R}} \tan \theta - \tilde{y}_{\mathcal{S}} \leq 0] + \mathbb{P}[\tilde{y}_{\mathcal{R}} \tan \theta + \tilde{y}_{\mathcal{S}} \leq 0] \quad (65)$$

where  $\tilde{y}_{\mathcal{R}} \tan \theta - \tilde{y}_{\mathcal{S}} = 0$  and  $\tilde{y}_{\mathcal{R}} \tan \theta + \tilde{y}_{\mathcal{S}} = 0$  define the rotated decision boundaries. We have  $(\tilde{y}_{\mathcal{R}} \tan \theta - \tilde{y}_{\mathcal{S}}) \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$  and  $(\tilde{y}_{\mathcal{R}} \tan \theta + \tilde{y}_{\mathcal{S}}) \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$  where

$$\begin{aligned} \tilde{\mu} &= \sqrt{P} N \mu_{z_{\mathcal{R}}} \tan \theta, \\ \tilde{\sigma}^2 &= \left( P N \sigma_{z_{\mathcal{R}}}^2 + \frac{N_0}{2} \right) \tan^2 \theta + P N \sigma_{z_{\mathcal{S}}}^2 + \frac{N_0}{2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{P}[\tilde{y}_{\mathcal{R}} \tan \theta - \tilde{y}_{\mathcal{S}} \leq 0] \\ &= \mathbb{P}[\tilde{y}_{\mathcal{R}} \tan \theta + \tilde{y}_{\mathcal{S}} \leq 0] \\ &= Q \left( \frac{\sqrt{P} N \mu_{z_{\mathcal{R}}} \tan \theta}{\sqrt{(P N \sigma_{z_{\mathcal{R}}}^2 + \frac{N_0}{2}) \tan^2 \theta + P N \sigma_{z_{\mathcal{S}}}^2 + \frac{N_0}{2}}} \right), \quad (66) \end{aligned}$$

which means the SER can be approximated as

$$\text{SER} \approx 2Q \left( \frac{\sqrt{P} N \mu_{z_{\mathcal{R}}} \tan \theta}{\sqrt{(P N \sigma_{z_{\mathcal{R}}}^2 + \frac{N_0}{2}) \tan^2 \theta + P N \sigma_{z_{\mathcal{S}}}^2 + \frac{N_0}{2}}} \right). \quad (67)$$

At high SNRs,  $\epsilon$  is small, and we have

$$\mu_{z_{\mathcal{R}}} = \mathbb{E}[z_{i,\mathcal{R}}] \approx \mathbb{E}[|a_i|] = \mathbb{E}[|h_i|] \mathbb{E}[|g_i|] = \frac{\pi}{4} \sigma_a, \quad (68)$$

$$\begin{aligned} \sigma_{z_{\mathcal{R}}}^2 &= \text{Var}[z_{i,\mathcal{R}}] = \mathbb{E}[z_{i,\mathcal{R}}^2] - |\mathbb{E}[z_{i,\mathcal{R}}]|^2 \\ &\approx \mathbb{E}[|a_i|^2] - \mathbb{E}[|a_i|]^2 = \left( 1 - \frac{\pi^2}{16} \right) \sigma_a^2. \end{aligned} \quad (69)$$

Substituting (68) and (69) into (67) and using the result that  $\text{BER} \approx \text{SER}/\log_2(D)$  for a Gray code at high SNRs gives us the approximated BER in (32).

### APPENDIX B PROOF OF LEMMA 1

We have

$$\mathbb{E}[\xi_t] = 1 - \mathbb{E}[s_t \hat{s}_t^*], \quad (70)$$

$$\mathbb{E}[\xi_t^2] = 1 - 2\mathbb{E}[s_t \hat{s}_t^*] + \mathbb{E}[(s_t \hat{s}_t^*)^2]. \quad (71)$$

Thus, to obtain  $\mathbb{E}[\xi_t]$  and  $\mathbb{E}[\xi_t^2]$ , we need to compute  $\mathbb{E}[s_t \hat{s}_t^*]$  and  $\mathbb{E}[(s_t \hat{s}_t^*)^2]$ , which are given as follows:

$$\begin{aligned} \mathbb{E}[s_t \hat{s}_t^*] &= \sum_{d=0}^{D-1} p_d^{\text{DD1}} \mathcal{S}(0) \mathcal{S}(d)^* \\ &= p_0^{\text{DD1}} - p_{\frac{D}{2}}^{\text{DD1}} + 2 \sum_{d=1}^{\frac{D}{2}-1} p_d^{\text{DD1}} \cos\left(\frac{2\pi d}{D}\right), \quad (72) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[(s_t \hat{s}_t^*)^2] &= \sum_{d=0}^{D-1} p_d^{\text{DD1}} (\mathcal{S}(0) \mathcal{S}(d)^*)^2 \\ &= p_0^{\text{DD1}} + p_{\frac{D}{2}}^{\text{DD1}} + 2 \sum_{d=1}^{\frac{D}{2}-1} p_d^{\text{DD1}} \cos\left(\frac{4\pi d}{D}\right). \quad (73) \end{aligned}$$

Substituting (72) and (73) into (70) and (71), we obtain  $\mu_\xi$  and  $\mu_{\xi^2}$  as in (40) and (41), respectively. The expectation of  $|\xi_t|^2$  is given as

$$\begin{aligned} \mathbb{E}[|\xi_t|^2] &= \mathbb{E}[|\hat{s}_t - s_t|^2] = \sum_{d=0}^{D-1} |\mathcal{S}(d) - \mathcal{S}(0)|^2 p_d^{\text{DD1}} \\ &= 4p_{\frac{D}{2}}^{\text{DD1}} + 4 \sum_{d=1}^{\frac{D}{2}-1} \left[ 1 - \cos\left(\frac{2\pi d}{D}\right) \right] p_d^{\text{DD1}}. \quad (74) \end{aligned}$$

Note that in (72), (73), and (74) we have used the following results:  $p_d^{\text{DD1}} = p_{d+\frac{D}{2}}^{\text{DD1}}$ ,  $\mathcal{S}(0) \mathcal{S}(d)^* = \mathcal{S}(0) \mathcal{S}(d + D/2)^* = \cos(4\pi d/D)$ , and  $|\mathcal{S}(d) - \mathcal{S}(0)|^2 = |\mathcal{S}(d + D/2) - \mathcal{S}(0)|^2 = 4(1 - \cos(2\pi d/D))$  for  $d = 1, \dots, D/2 - 1$ .

### APPENDIX C CALCULATION OF THE TERMS IN (60)

First, we have  $\mathbb{E}[|a_i|^2 |a_t|^2] = \sigma_a^4$ . Using the same approach as in (52), we obtain

$$\mathbb{E}[|a_i|^2 a_t \epsilon_t^*] = \mathbb{E}[|a_t|^2 a_i \epsilon_i^*] = \frac{N-2}{1-N} \sigma_a^4 \mu_\xi. \quad (75)$$

The two terms  $\mathbb{E}[|a_i \epsilon_t|^2]$  and  $\mathbb{E}[|a_t \epsilon_i|^2]$  are equal and given by

$$\begin{aligned} \mathbb{E}[|a_i \epsilon_t|^2] &= \mathbb{E}[|a_i \delta_t|^2] + \mathbb{E}[|a_i \tilde{n}_t|^2] \\ &= \mathbb{E}[|a_i \delta_t|^2] + \frac{\sigma_a^2 N_0^{\text{BS}}}{P(N-1)}, \end{aligned} \quad (76)$$

where

$$\begin{aligned} &\mathbb{E}[|a_i \delta_t|^2] \\ &= \frac{\mathbb{E}\left[|a_i \mathbf{a}^H \mathbf{\Phi} \text{diag}(\boldsymbol{\xi}) \mathbf{\Phi}^H|_{t,:}^2\right]}{(N-1)^2} \\ &= \frac{\mathbb{E}\left[\text{tr}\left\{|a_i|^2 \mathbf{\Phi}^H \mathbf{a} \mathbf{a}^H \mathbf{\Phi} \text{diag}(\mathbf{\Phi}_{t,:}^H) \boldsymbol{\xi} \boldsymbol{\xi}^H \text{diag}(\mathbf{\Phi}_{t,:})\right\}\right]}{(N-1)^2} \\ &= \frac{\text{tr}\left\{\mathbf{\Phi}^H \text{diag}(\boldsymbol{\alpha}_i) \mathbf{\Phi} \text{diag}(\mathbf{\Phi}_{t,:}^H) \mathbf{R}_{\boldsymbol{\xi} \boldsymbol{\xi}^H} \text{diag}(\mathbf{\Phi}_{t,:})\right\}}{(N-1)^2}. \end{aligned} \quad (77)$$

The term  $\mathbb{E}[a_i \epsilon_i^* a_t \epsilon_t^*]$  is given as follows:

$$\begin{aligned} \mathbb{E}[a_i \epsilon_i^* a_t \epsilon_t^*] &= \mathbb{E}[a_i a_t \delta_i^* \delta_t^*] \\ &= \frac{\mathbb{E}[a_i a_t \mathbf{a}^H \Phi \text{diag}(\xi) \Phi_{i,:}^H \mathbf{a}^H \Phi \text{diag}(\xi) \Phi_{t,:}^H]}{(N-1)^2} \\ &= \frac{\mathbb{E}[\text{tr}\{\Phi^T a_i a_t \mathbf{a}^H \Phi \text{diag}(\xi) \Phi_{i,:}^H \Phi_{t,:}^* \text{diag}(\xi)\}]}{(N-1)^2} \\ &= \frac{\text{tr}\{\Phi^T \Sigma \Phi \text{diag}(\Phi_{i,:}^H) \mathbf{R}_{\xi \xi^T} \text{diag}(\Phi_{t,:}^*)\}}{(N-1)^2}, \end{aligned} \quad (78)$$

where  $\Sigma$  is a matrix with  $\Sigma_{i,t} = \Sigma_{t,i} = \sigma_a^4$  and zeroes elsewhere. The term  $\mathbb{E}[a_i \epsilon_i^* a_t^* \epsilon_t]$  is obtained as follows:

$$\begin{aligned} \mathbb{E}[a_i \epsilon_i^* a_t^* \epsilon_t] &= \mathbb{E}[a_i a_t^* \delta_i^* \delta_t] \\ &= \frac{\mathbb{E}[a_i a_t^* \mathbf{a}^H \Phi \text{diag}(\xi) \Phi_{i,:}^H \mathbf{a}^T \Phi^* \text{diag}(\xi^*) \Phi_{t,:}^T]}{(N-1)^2} \\ &= \frac{\mathbb{E}[\text{tr}\{\Phi^H a_i a_t^* \mathbf{a}^H \Phi \text{diag}(\xi) \Phi_{i,:}^H \Phi_{t,:}^* \text{diag}(\xi^*)\}]}{(N-1)^2} \\ &= \frac{\text{tr}\{\Phi^H \Omega \Phi \text{diag}(\Phi_{i,:}^H) \mathbf{R}_{\xi \xi^H} \text{diag}(\Phi_{t,:}^*)\}}{(N-1)^2}, \end{aligned} \quad (79)$$

where  $\Omega$  is a matrix with  $\Omega_{t,i} = \sigma_a^4$ , and zeroes elsewhere. Note that we have used the same technique in (58) for deriving (77) and the same technique in (59) for deriving (78) and (79).

The two terms  $\mathbb{E}[a_i \epsilon_i^* |\epsilon_t|^2]$  and  $\mathbb{E}[a_i \epsilon_i^* |\epsilon_t|^2]$  are also equal and given by

$$\begin{aligned} \mathbb{E}[a_i \epsilon_i^* |\epsilon_t|^2] &= \mathbb{E}[a_i \delta_i^* |\delta_t|^2] + \mathbb{E}[a_i \delta_i^* |\tilde{n}_t|^2] \\ &\approx \mathbb{E}[a_i \delta_i^* |\tilde{n}_t|^2] = -\frac{(N-2)N_0^{\text{BS}}}{P(N-1)^2} \sigma_a^2 \mu_\xi. \end{aligned} \quad (80)$$

Finally, the term  $\mathbb{E}[|\epsilon_i|^2 |\epsilon_t|^2]$  is approximated as

$$\begin{aligned} \mathbb{E}[|\epsilon_i|^2 |\epsilon_t|^2] &= \mathbb{E}[|\delta_i|^2 |\delta_t|^2] + 2\mathbb{E}[|\delta_i|^2 |\tilde{n}_t|^2] + \mathbb{E}[|\tilde{n}_i|^2 |\tilde{n}_t|^2] \\ &\approx 2\mathbb{E}[|\delta_i|^2 |\tilde{n}_t|^2] + \mathbb{E}[|\tilde{n}_i|^2 |\tilde{n}_t|^2] \\ &= \frac{2(N-2)N_0^{\text{BS}}}{P(N-1)^2} \sigma_a^2 \mu_{|\xi|^2} + \frac{(N_0^{\text{BS}})^2}{P^2(N-1)^2}. \end{aligned} \quad (81)$$

The quantities  $\mathbb{E}[a_i \delta_i^* |\delta_t|^2]$  in (80) and  $\mathbb{E}[|\delta_i|^2 |\delta_t|^2]$  in (81) are ignored in our approximation since they make a negligible contribution to the result.

## REFERENCES

- [1] 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. 116753–116773, 2019.
- [2] M. Di Renzo et al., "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [3] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfaces—A general framework," *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, Sep. 2022.
- [4] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, Jun. 2020, pp. 1–6.
- [5] X. Ma et al., "Joint channel estimation and data rate maximization for intelligent reflecting surface assisted terahertz MIMO communication systems," *IEEE Access*, vol. 8, pp. 99565–99581, 2020.
- [6] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.
- [7] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5786–5797, Sep. 2021.
- [8] Y. Liu, S. Zhang, F. Gao, J. Tang, and O. A. Dobre, "Cascaded channel estimation for RIS assisted mmWave MIMO transmissions," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 2065–2069, Sep. 2021.
- [9] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Process. Lett.*, vol. 28, pp. 513–517, 2021.
- [10] H. Liu, J. Zhang, Q. Wu, H. Xiao, and B. Ai, "ADMM based channel estimation for RISs aided millimeter wave communications," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2894–2898, Sep. 2021.
- [11] W. Zhang, J. Xu, W. Xu, D. W. K. Ng, and H. Sun, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593–597, Feb. 2021.
- [12] Z. Chen et al., "Hybrid evolutionary-based sparse channel estimation for IRS-assisted mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1586–1601, Mar. 2022.
- [13] 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, Oct. 2023, doi: 10.1109/TWC.2023.3246264.
- [14] 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, Apr. 2022.
- [15] 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, 2022.
- [16] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.
- [17] S. Ma, W. Shen, J. An, and L. Hanzo, "Wideband channel estimation for IRS-aided systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6240–6253, Oct. 2021.
- [18] E. Björnson and P. Ramezani, "Maximum likelihood channel estimation for RIS-aided communications with LOS channels," in *Proc. 56th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, 2022, pp. 403–407.
- [19] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.
- [20] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.
- [21] X. Chen, J. Shi, Z. Yang, and L. Wu, "Low-complexity channel estimation for intelligent reflecting surface-enhanced massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 996–1000, May 2021.
- [22] Y. Jin, J. Zhang, X. Zhang, H. Xiao, B. Ai, and D. W. K. Ng, "Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11083–11088, Oct. 2021.
- [23] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1132–1142, Feb. 2022.
- [24] I.-S. Kim, M. Bennis, J. Oh, J. Chung, and J. Choi, "Bayesian channel estimation for intelligent reflecting surface-aided mmWave massive MIMO systems with semi-passive elements," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9732–9745, May 2023.
- [25] X. Guo, Y. Chen, and Y. Wang, "Wireless beacon enabled hybrid sparse channel estimation for RIS-aided mmWave communications," *IEEE Trans. Commun.*, vol. 71, no. 5, pp. 3144–3160, May 2023.

- [26] 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)*, Brighton, U.K., May 2019, pp. 4659–4663.
- [27] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.
- [28] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 5000–5004.
- [29] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.
- [30] Q. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, 2020.
- [31] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, Nov. 2020.
- [32] Ö. T. Demir and E. Björnson, "Is channel estimation necessary to select phase-shifts for RIS-assisted massive MIMO?" *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9537–9552, Nov. 2022.
- [33] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789–802, Apr. 2021.
- [34] 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, Nov. 2020.
- [35] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.
- [36] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.
- [37] 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.
- [38] Y. Wei, M.-M. Zhao, M.-J. Zhao, and Y. Cai, "Channel estimation for IRS-aided multiuser communications with reduced error propagation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2725–2741, Apr. 2022.
- [39] 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.*, vol. 70, no. 2, pp. 1245–1260, Feb. 2022.
- [40] J. An et al., "Codebook-based solutions for reconfigurable intelligent surfaces and their open challenges," *IEEE Wireless Commun.*, vol. 31, no. 2, Apr. 2024, doi: [10.1109/MWC.010.2200312](https://doi.org/10.1109/MWC.010.2200312).
- [41] L. Wei et al., "Joint channel estimation and signal recovery for RIS-empowered multiuser communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4640–4655, Jul. 2022.
- [42] J. Du, X. Luo, X. Li, M. Zhu, K. M. Rabie, and F. Kara, "Semi-blind joint channel estimation and symbol detection for RIS-empowered multiuser mmWave systems," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 362–366, Jan. 2023.
- [43] C. Huang, J. Xu, W. Zhang, W. Xu, and D. W. K. Ng, "Semi-blind channel estimation for RIS-assisted MISO systems using expectation maximization," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10173–10178, Sep. 2022.
- [44] A. S. Alwakeel and A. Elzanaty, "Semi-blind channel estimation for intelligent reflecting surfaces in massive MIMO systems," *IEEE Access*, vol. 10, pp. 127783–127797, 2022.
- [45] G. C. Alexandropoulos, N. Shlezinger, I. Alamzadeh, M. F. Imani, H. Zhang, and Y. C. Eldar, "Hybrid reconfigurable intelligent metasurfaces: Enabling simultaneous tunable reflections and sensing for 6G wireless communications," *IEEE Veh. Technol. Mag.*, vol. 19, no. 1, Mar. 2024, doi: [10.1109/MVT.2023.3332580](https://doi.org/10.1109/MVT.2023.3332580).
- [46] I. Alamzadeh, G. C. Alexandropoulos, N. Shlezinger, and M. F. Imani, "A reconfigurable intelligent surface with integrated sensing capability," *Sci. Rep.*, vol. 11, no. 1, p. 20737, Oct. 2021.
- [47] H. Zhang et al., "Channel estimation with hybrid reconfigurable intelligent metasurfaces," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2441–2456, Apr. 2023.
- [48] N. O. Garzón, H. C. Mora, F. A. García, and C. D. Altamirano, "On the bit error probability and the spectral efficiency of opportunistic wireless transmissions in Rician fading channels," *IEEE Access*, vol. 9, pp. 49267–49280, 2021.
- [49] L. V. Nguyen, D. T. Ngo, N. H. Tran, A. L. Swindlehurst, and D. H. N. Nguyen, "Supervised and semi-supervised learning for MIMO blind detection with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2427–2442, Apr. 2020.
- [50] J. An, C. Xu, L. Wang, Y. Liu, L. Gan, and L. Hanzo, "Joint training of the superimposed direct and reflected links in reconfigurable intelligent surface assisted multiuser communications," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 739–754, Jun. 2022.
- [51] R. J. Baxley, B. T. Walkenhorst, and G. Acosta-Marum, "Complex Gaussian ratio distribution with applications for error rate calculation in fading channels with imperfect CSI," in *Proc. IEEE Global Telecommun. Conf.*, Miami, FL, USA, Dec. 2010, pp. 1–5.
- [52] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.
- [53] J. An, C. Yuen, C. Huang, M. Debbah, H. Vincent Poor, and L. Hanzo, "A tutorial on holographic MIMO communications—Part I: Channel modeling and channel estimation," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1664–1668, Jul. 2023.
- [54] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.



**Ly V. Nguyen** (Member, IEEE) received the B.Eng. degree in electronics and telecommunications from the University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam, in 2014, the M.Sc. degree in advanced wireless communications systems from CentraleSupélec, Paris-Saclay University, France, in 2016, and the joint Ph.D. degree from San Diego State University and University of California at Irvine, Irvine, CA, USA, in 2022, in the Joint Doctoral Program in Computational Science. His research interests include wireless communications, signal processing, and machine learning. He received the Best Paper Award at the 2020 IEEE International Conference on Communications (ICC).



**A. Lee Swindlehurst** (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Brigham Young University (BYU), Provo, UT, USA, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1991. He was with the Department of Electrical and Computer Engineering, BYU, from 1990 to 2007. From 1996 to 1997, he held a joint appointment as a Visiting Scholar with Uppsala University, Uppsala, Sweden, and the Royal Institute of Technology, Stockholm, Sweden.

From 2006 to 2007, he was on leave working as a Vice President of Research with ArrayComm LLC, San Jose, CA, USA. Since 2007, he has been a Professor with the Electrical Engineering and Computer Science Department, University of California at Irvine, Irvine, CA, USA. From 2014 to 2017, he was also a Hans Fischer Senior Fellow with the Institute for Advanced Studies, Technical University of Munich, Munich, Germany. In 2016, he was elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA). His research interests include array signal processing for radar, wireless communications, and biomedical applications. He has more than 350 publications in these areas. He was the inaugural Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He was a recipient of the 2000 IEEE W.R.G. Baker Prize Paper Award, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, the 2006, 2010, and 2021 IEEE Signal Processing Society Best Paper Awards, the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award, the Best Paper award at the 2020 IEEE International Conference on Communications, and the 2022 Claude Shannon-Harry Nyquist Technical Achievement Award from the IEEE Signal Processing Society.