Channel Estimation with Reconfigurable Intelligent Surfaces – A General Framework

A. Lee Swindlehurst *Fellow, IEEE*[1], Gui Zhou *Student Member, IEEE*[2], Rang Liu *Student Member, IEEE*[3],
Cunhua Pan *Member, IEEE*[2], and Ming Li, *Senior Member, IEEE*[3]

[1]Center for Pervasive Communications & Computing, University of California, Irvine, USA
[2]School of Electronic Engineering & Computer Science, Queen Mary University of London, England
[3]School of Information & Communication Engineering, Dalian University of Technology, China

Abstract—**Optimally extracting the advantages available from reconfigurable intelligent surfaces (RISs) in wireless communications systems requires estimation of the channels to and from the RIS. The process of determining these channels is complicated by the fact that the RIS is typically composed of passive elements without any data processing capabilities, and thus the channels must be estimated indirectly by a non-colocated device, typically a controlling base station. In this article, we examine channel estimation for RIS-based systems from a fundamental viewpoint. We study various possible channel models and the identifiability of the models as a function of the available pilot data and behavior of the RIS during training. In particular, we will consider situations with and without line-of-sight propagation, single- and multiple-antenna configurations for the users and base station, correlated and sparse channel models, single-carrier and wideband OFDM scenarios, availability of direct links between the users and base station, exploitation of prior information, as well as a number of other special cases. We further conduct numerical comparisons of achievable performance for various channel models using the relevant Cramér-Rao bounds.**

Index Terms—**Channel estimation, DOA estimation, reconfigurable intelligent surface, intelligent reflecting surface, massive MIMO**

I. INTRODUCTION

There has been an explosion of interest in the use of reconfigurable metasurfaces for wireless communication systems in the last few years. Such reconfigurable intelligent surfaces (RIS) provide tunable degrees-of-freedom for adjusting the propagation characteristics of problematic channels (e.g., sparse channels with frequent blockages) that make them a valuable resource for maintaining and enhancing the quality of service (QoS) for users (UEs) in the network. However, most techniques that exploit this ability require channel state information (CSI) to and from the elements of the RIS, which is a challenge since the number of RIS elements may be very large, and more importantly, they are usually constructed only as passive devices without active transceivers or computational resources. Consequently, channel estimation for RIS-based systems has been a subject of intense study.

Because the RIS is passive, the CSI must be estimated by devices – most often a basestation (BS) or access point – that are not co-located with the RIS. For example, training signals transmitted by the UEs are received by the BS after reflection from the RIS, and possibly also over a direct path to the

BS, and these known signals are exploited for CSI estimation. In order to estimate the RIS-based channel components, the reflection coefficients of the RIS must be varied as well, at least during a portion of the training period. However, even with variable training from the UEs and RIS, the fact that the impact of the RIS is only indirectly viewed in the data means that the complete structure of the channel is not identifiable. In particular, while the *cascaded* or *composite* channel from the UEs to the BS can be determined, the individual components of the channel involving the RIS cannot. Fortunately, this is typically not a problem for designing beamforming algorithms at the BS or optimizing the RIS reflection properties, since ultimately the QoS only depends on the composite channel.

A large amount of published work on CSI estimation for RIS-based systems has appeared recently. Initially, this work focused on estimating *unstructured* models, where the channels are simply described using complex gains [1–11]. Such models are simple and lead to straightforward algorithms, but the required training overhead is very large and may render such approaches impractical. Methods for reducing the training overhead, for example based on grouping the RIS elements or exploiting the common BS-RIS channel among the users, have been proposed, but larger reductions are possible when the channels are sparse if parametric or geometric channel models are used instead [12–25]. In these models, the channels are parameterized by the angles of arrival (AoAs), angles of departure (AoDs) and complex gains of each propagation path. As long as the number of multipaths is not large, then the total number of parameters to be estimated can be 1-2 orders of magnitude smaller than in the unstructured case, and the amount of training can be correspondingly reduced. On the other hand, geometric models require knowledge of the array calibration and RIS element responses, as well as the model order; errors in the modeling assumptions will degrade some of this advantage. In addition, we will see later that the algorithms for estimating the geometric channel parameters can in general be quite complex.

Many CSI estimation techniques have been proposed under a wide array of assumptions, from Rayleigh fading to line-of-sight (LoS) propagation, single- to multi-antenna configurations, single- and multi-carrier modulation, scenarios with and without a direct link between BS and UEs, and a variety of other special cases. In this paper, we take a systematic approach to the problem and organize the various approaches that have been

proposed – as well as some that have not – under a common framework. In this way, the advantages and disadvantages of different assumptions and solution approaches become clearer, and avenues for future work are elucidated.

After stating our general assumptions and notational conventions in Section II, we begin with a discussion of CSI estimation for unstructured channels in Section III. We will first consider the narrowband single user MIMO case and the corresponding least-squares (LS) and minimum mean squared error (MMSE) solutions, and then we will examine extensions to the wideband and multi-user cases, as well as special cases involving a single antenna BS and UEs and methods for reducing the training overhead. Then in Section IV we focus on estimation of geometric channel models, and we follow the same format of beginning with the narrowband single user MIMO case and then considering the same generalizations and special cases as in the previous section. Numerical examples involving the Cramér-Rao bound (CRB) will be given in Section V to illustrate the main conclusions. Several additional topics will be briefly considered in Section VI, including the use of some active transceivers at the RIS, scenarios with more than one RIS, machine learning approaches, etc. Finally some conclusions and suggestions for future research are offered in Section VII.

II. General Assumptions and Notation

In this paper, we primarily consider scenarios with a single basestation (BS), a single RIS, and potentially multiple co-channel UEs. Various assumptions are made about the number of antennas at the BS and UEs, and the number of UEs that are active. We assume the BS and UEs employ fully digital rather than hybrid digital/analog architectures. We also assume a standard time-division duplex protocol in which pilot symbols transmitted by the UEs in the uplink are exploited by the BS to obtain a channel estimate, which is then used for downlink beamforming or multiplexing. This assumes reciprocal uplink and downlink channels between the BS, RIS and users, which in turn typically requires some type of RF transceiver calibration and RIS elements whose behavior is independent of the angle of incidence. Pilots could also be embedded in the downlink for channel estimation at the UEs, but this is similar to the uplink problem and thus is not explicitly considered.

Matrices and vectors are denoted by boldface capital and lowercase letters, respectively. In some cases, the k-th column or row of a matrix $\mathbf{A}$ will be denoted by $\mathbf{A}_{:k}$ or $\mathbf{A}_{k:}$, respectively. The transpose, conjugate transpose, and conjugate are denoted by $(\cdot)^T, (\cdot)^H$, and $(\cdot)^*$, respectively. The Kronecker, Khatri-Rao, and Hadamard products of two matrices are indicated by $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$, $\mathbf{C} = \mathbf{A} \diamond \mathbf{B}$ and $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$, respectively. An $N \times N$ identity matrix is represented as $\mathbf{I}_N$, and $N \times 1$ vectors composed of all ones or zeros are denoted by $\mathbf{1}_N$ and $\mathbf{0}_N$, respectively. A circular complex multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{R}$ is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$. The function vec($\mathbf{A}$) creates a vector from matrix $\mathbf{A}$ by stacking its columns. The function $\lfloor a \rfloor$ creates an integer from real number a by truncating its decimal part, and $a \bmod b$ is the modulo operator that returns the integer remainder of a/b. A diagonal matrix with elements of vector $\mathbf{c}$ on the diagonal is indicated by diag($\mathbf{c}$), and a block diagonal matrix with block
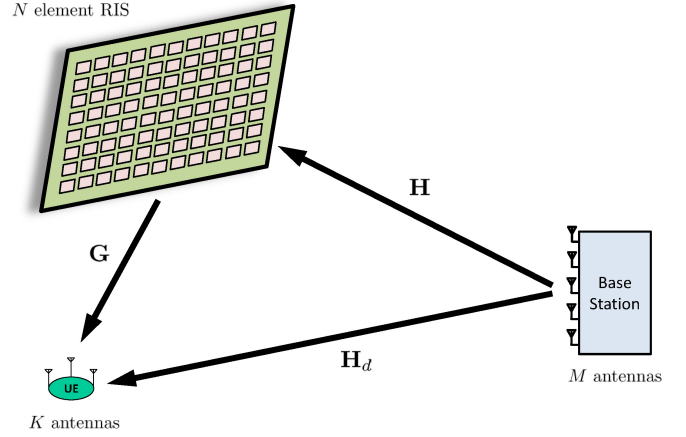

Fig. 1: A generic scenario involving an M-antenna basestation, an N-element RIS, and a K-element user.

entries $\mathbf{C}_1, \mathbf{C}_2, \ldots$ is written blkdiag $([\mathbf{C}_1 \quad \mathbf{C}_2 \quad \ldots])$.

The reflective properties of an RIS with N elements is described by the $N \times N$ diagonal matrix $\boldsymbol{\Phi} = \{\text{diag}(\boldsymbol{\phi})\}$, where $\boldsymbol{\phi} = [\beta_1 e^{j\alpha_1} \quad \cdots \quad \beta_N e^{j\alpha_N}]^T$. There are a number of practical issues associated with $\boldsymbol{\phi}$ that are important for RIS performance optimization, such as the dependence of the gains β on the phases α, the fact that the phases are typically discrete and frequency dependent, etc. For the most part, these issues are not directly relevant to the generic channel estimation problem, which only requires that $\boldsymbol{\phi}$ be known and sufficiently controllable. However, certain simplifying assumptions about $\boldsymbol{\phi}$ are made below for performance analyses or purposes of illustration.

III. Estimation of Unstructured Channel Models

We begin with models where the channel between individual network elements is described by a complex coefficient in the case of a narrowband single carrier signal, or a complex-valued impulse response for wideband transmission. Such unstructured or nonparametric channel models are appropriate for situations with rich multipath scattering (e.g., sub-6GHz systems), where it is difficult to describe the aggregate characteristics of the propagation environment. We initially focus on the narrowband single-user scenario, and then examine cases involving wideband signals or multiple users. As will become clear, the limiting factor with unstructured CSI estimation is the large training overhead that is required. Approaches for reducing the training overhead are discussed at the end of the section.

A. Narrowband Single User MIMO

The scenario assumed here is as depicted in Fig. 1, with an M-antenna BS, an N-element RIS, and a single UE with K antennas. The geometries of the RIS and the arrays at the BS and UE are arbitrary. If the UE transmits the $K \times 1$ vector $\mathbf{x}_t$ at time t, the signal received at the BS is given by

$$\mathbf{y}_t = \sqrt{P} \left(\mathbf{H}_d + \mathbf{H} \boldsymbol{\Phi}_t \mathbf{G}^H \right) \mathbf{x}_t + \mathbf{n}_t , \qquad (1)$$

where $\mathbf{H}_d, \mathbf{H}, \mathbf{G}$ are respectively the channels between the BS and UE, the BS and RIS, and the RIS and UEs, and $\mathbf{n}_t$ denotes additive noise or interference. Assuming $\mathcal{E}\{\mathbf{x}_t \mathbf{x}_t^H\} = \mathbf{I}_K$ and $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$, P represents the transmit power, and the

signal-to-noise ratio (SNR) is defined as P/σ^2. The channels are all assumed to be block flat fading and constant over a coherence interval sufficiently long to permit channel estimation and subsequent data transmission. On the other hand, the reflection coefficients of the RIS, $\boldsymbol{\Phi}_t$, can vary synchronously with the UE uplink transmission. Some prior work ignores the direct channel component $\mathbf{H}_d$, assuming that it is either not present (e.g., due to a blockage), or that it was estimated in a previous step and its contribution has been removed from the received data, *i.e.*, $\mathbf{y}_t \longrightarrow \mathbf{y}_t - (1/\sqrt{P})\mathbf{H}_d\mathbf{x}_t$.

It is important to note that not all of the components of the channel-related term $\mathbf{H}_d + \mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}^H$ are individually identifiable. In particular, for any invertible $N \times N$ diagonal matrix $\boldsymbol{\Lambda}$, we have

$$\mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}^H = \mathbf{H}\boldsymbol{\Lambda}\boldsymbol{\Phi}_t\boldsymbol{\Lambda}^{-1}\mathbf{G}^H = \tilde{\mathbf{H}}\boldsymbol{\Phi}_t\tilde{\mathbf{G}}^H , \qquad (2)$$

where $\tilde{\mathbf{H}} = \mathbf{H}\boldsymbol{\Lambda}$ and $\tilde{\mathbf{G}} = \mathbf{G}(\boldsymbol{\Lambda}^*)^{-1}$. Thus there is a scaling ambiguity between each pair of the N columns $\{\mathbf{h}_k, \mathbf{g}_k\}$ of $\mathbf{H}$ and $\mathbf{G}$ that cannot be resolved using data obtained as in (1). Most methods for beamforming, precoding or RIS reflection optimization do not require this ambiguity to be resolved, although as briefly discussed later, with certain additional information the individual channel components can be identified. For this reason, channel estimation in the context of RIS-aided communication systems focuses primarily on determination of the $MK \times (N+1)$ *composite* or *cascaded* channel $\mathbf{H}_c$, defined using properties of the Khatri-Rao product:

$$\mathrm{vec}\left(\mathbf{H}_d + \mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}^H\right) = [\mathbf{h}_d \ \ \mathbf{G}^* \diamond \mathbf{H}] \begin{bmatrix} 1 \\ \boldsymbol{\phi}_t \end{bmatrix} \equiv \mathbf{H}_c\tilde{\boldsymbol{\phi}}_t , \tag{3}$$

with $\mathbf{h}_d = \mathrm{vec}(\mathbf{H}_d)$. Eq. (3) together with further use of the Kronecker product allows us to rewrite (1) in a compact form:

$$\mathbf{y}_t = \sqrt{P}\left(\mathbf{x}_t^T \otimes \mathbf{I}_M\right)\mathbf{H}_c\tilde{\boldsymbol{\phi}}_t + \mathbf{n}_t \tag{4a}$$

$$= \sqrt{P}\left[\tilde{\boldsymbol{\phi}}_t^T \otimes \mathbf{x}_t^T \otimes \mathbf{I}_M\right]\mathbf{h}_c + \mathbf{n}_t \tag{4b}$$

$$\equiv \sqrt{P}\mathbf{Z}_t\mathbf{h}_c + \mathbf{n}_t , \tag{4c}$$

where $\mathbf{h}_c = \mathrm{vec}(\mathbf{H}_c)$ and the $M \times MK(N+1)$ matrix $\mathbf{Z}_t$ is implicitly defined.

The composite channel $\mathbf{h}_c$ is clearly underdetermined in Eq. (4), and thus multiple pilot symbols must be transmitted in order for it to be uniquely estimated. Combining the data from T such pilots together, we have

$$\mathfrak{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \sqrt{P} \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_T \end{bmatrix} \mathbf{h}_c + \mathfrak{n} \equiv \sqrt{P}\mathbf{Z}\mathbf{h}_c + \mathfrak{n} . \tag{5}$$

Provided that $T \geq K(N+1)$ and $\mathbf{Z}$ is full rank, there are two common ways to estimate $\mathbf{h}_c$, as discussed below.

1) Least Squares

The simplest approach for estimating $\mathbf{h}_c$ is to use the standard deterministic least-squares (LS) method,

$$\hat{\mathbf{h}}_{c,LS} = \arg\min_{\mathbf{h}_c} \|\mathfrak{y} - \sqrt{P}\mathbf{Z}\mathbf{h}_c\|^2 , \tag{6}$$

whose solution is given by

$$\hat{\mathbf{h}}_{c,LS} = \frac{1}{\sqrt{P}}\mathbf{Z}^\dagger\mathfrak{y} , \tag{7}$$

where $\mathbf{Z}^\dagger = \left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H$. Assuming again that $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ and that the noise is temporally uncorrelated, the LS channel estimate is unbiased and equivalent to the maximum likelihood (ML) estimate, and its covariance matrix corresponds to the CRB:

$$\mathbf{R}_{\hat{\mathbf{h}}_{c,LS}} = \mathcal{E}\left\{\left(\hat{\mathbf{h}}_{c,LS} - \mathbf{h}_c\right)\left(\hat{\mathbf{h}}_{c,LS} - \mathbf{h}_c\right)^H\right\} \tag{8a}$$

$$= \frac{1}{P}\mathcal{E}\left\{\mathbf{Z}^\dagger\mathfrak{n}\mathfrak{n}^H(\mathbf{Z}^\dagger)^H\right\} = \frac{\sigma^2}{P}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1} . \tag{8b}$$

Ideally, $\mathbf{x}_t$ and $\boldsymbol{\phi}_t$ should be designed to optimize the CSI estimation performance. While such an optimization is generally intractable, a good choice can be found [1] by noting that for any positive definite matrix $\mathbf{B}$, we have

$$\left[\mathbf{B}^{-1}\right]_{ii} \geq \frac{1}{\mathbf{B}_{ii}} , \tag{9}$$

with equality for all i only if $\mathbf{B}$ is diagonal. Thus, a good choice for $\mathbf{Z}$ would make (8) diagonal. Such a choice may not be optimal in general, but a diagonal covariance matrix also greatly simplifies the computation of $\hat{\mathbf{h}}_{c,LS}$ in (7).

The most common training approach that meets the above design goals breaks the training interval T into T/K subblocks of length K, where T/K is assumed to be an integer. For each subblocks, $b = 1, \cdots, T/K$, $\boldsymbol{\phi}_t = \bar{\boldsymbol{\phi}}_b$ is held constant, while the pilots $\mathbf{x}_t$ are chosen as an orthonormal sequence that repeats itself for each subblock. For example, the subblock sequence for the UE is $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_K]$, where $\mathbf{X}\mathbf{X}^H = K\mathbf{I}_K$, which is then repeated T/K times:

$$\mathbf{x}_t \text{ pilots} = [\ \underbrace{\mathbf{X}\ \mathbf{X}\ \cdots\ \mathbf{X}}_{\text{repeated } T/K \text{ times}}\ ] \tag{10a}$$

$$\boldsymbol{\phi}_t \text{ pilots} = [\ \underbrace{\bar{\boldsymbol{\phi}}_1 \cdots \bar{\boldsymbol{\phi}}_1}_{\text{repeated } K \text{ times}} \cdots \underbrace{\bar{\boldsymbol{\phi}}_{\frac{T}{K}} \cdots \bar{\boldsymbol{\phi}}_{\frac{T}{K}}}_{\text{repeated } K \text{ times}}] \tag{10b}$$

Using this approach, we have

$$\mathbf{Z}^H\mathbf{Z} = \sum_{t=1}^T \left[\tilde{\boldsymbol{\phi}}_t^*\tilde{\boldsymbol{\phi}}_t^T \otimes \mathbf{x}_t^*\mathbf{x}_t^T \otimes \mathbf{I}_M\right] \tag{11a}$$

$$= \left(\sum_{t=1}^T \left[\tilde{\boldsymbol{\phi}}_t^*\tilde{\boldsymbol{\phi}}_t^T \otimes \mathbf{x}_t^*\mathbf{x}_t^T\right]\right) \otimes \mathbf{I}_M \tag{11b}$$

$$= \left(\sum_{b=1}^{T/K} \tilde{\bar{\boldsymbol{\phi}}}_b^*\tilde{\bar{\boldsymbol{\phi}}}_b^T\right) \otimes (\mathbf{X}\mathbf{X}^H)^* \otimes \mathbf{I}_M \tag{11c}$$

$$= K\left(\boldsymbol{\Psi}^H\boldsymbol{\Psi}\right)^* \otimes \mathbf{I}_{MK} , \tag{11d}$$

where $\tilde{\bar{\boldsymbol{\phi}}}_b^H = [1 \ \bar{\boldsymbol{\phi}}_b^H]$ and

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & \bar{\boldsymbol{\phi}}_1^H \\ \vdots & \vdots \\ 1 & \bar{\boldsymbol{\phi}}_{T/K}^H \end{bmatrix} . \tag{12}$$

To achieve a diagonal $\boldsymbol{\Psi}^H \boldsymbol{\Psi}$, the columns of the $\frac{T}{K} \times (N+1)$ matrix $\boldsymbol{\Psi}$ must be orthogonal, with $\frac{T}{K} \geq N+1$. If $\boldsymbol{\Psi}^H \boldsymbol{\Psi}$ can be made proportional to an identity matrix, then $\mathbf{Z}^H \mathbf{Z}$ is also a scaled identity matrix.

For the above training protocol, the general solution in (7) is implemented by taking data from the b-th pilot subblock,

$$\mathbf{Y}_b = \sqrt{P} \left(\mathbf{H}_d + \mathbf{H} \bar{\boldsymbol{\Phi}}_b \mathbf{G}^H \right) \mathbf{X} + \mathbf{N}_b \,, \tag{13}$$

and multiplying on the right by $\mathbf{X}^H / (K\sqrt{P})$ to obtain

$$\mathfrak{y}_b \equiv \frac{1}{K\sqrt{P}} \mathrm{vec} \left(\mathbf{Y}_b \mathbf{X}^H \right) = \mathbf{H}_c \tilde{\bar{\phi}}_b + \bar{\mathbf{n}}_b \,. \tag{14}$$

where $\bar{\bar{\boldsymbol{\Phi}}}_b = \mathrm{diag}(\bar{\phi}_b)$ and $\bar{\mathbf{n}}_b = \mathrm{vec} \left(\mathbf{N}_b \mathbf{X}^H \right) / (K\sqrt{P})$. The result $\mathfrak{y}_b$ from each of the T/K subblocks then forms a column of the following combined equation:

$$\mathfrak{Y}_c = \mathbf{H}_c \left[\tilde{\bar{\phi}}_1 \; \cdots \; \tilde{\bar{\phi}}_{\frac{T}{K}} \right] + \bar{\mathbf{N}} = \mathbf{H}_c \boldsymbol{\Psi}^H + \bar{\mathbf{N}} \,, \tag{15}$$

where $\bar{\mathbf{N}} = [\bar{\mathbf{n}}_1 \; \cdots \; \bar{\mathbf{n}}_{T/K}]$, from which an estimate of the composite channel is obtained by multiplying $\mathfrak{Y}_c$ by $\boldsymbol{\Psi}$ on the right, assuming $\boldsymbol{\Psi}^H \boldsymbol{\Psi} \propto \mathbf{I}_{N+1}$.

Several methods have been proposed to choose the RIS training sequence to satisfy $\boldsymbol{\Psi}^H \boldsymbol{\Psi} \propto \mathbf{I}_{N+1}$:

- When the direct path is absent (the first column of $\boldsymbol{\Psi}$ is removed), a simple approach is to set $\frac{T}{K} = N$ and "turn on" one RIS element at a time for each K-sample pilot subblock, with all other elements "turned off"[1] [5], [26]. This results in $\boldsymbol{\Psi}^H \boldsymbol{\Psi} = \mathrm{diag}\{\beta_1^2, \cdots, \beta_N^2\}$. If each (identical) RIS element when active is tuned to the same phase, it is reasonable to assume that $\beta_i = \beta$, which results in $\mathbf{Z}^H \mathbf{Z} = \beta^2 K \mathbf{I}_{MKN}$ and an estimate variance of $\sigma^2/(\beta^2 PK)$ for each element of $\mathbf{h}_c$.
- Better performance is achieved by activating all RIS elements over the entire training interval, in order to benefit from the RIS array gain. One approach for doing so assigns the RIS phase shifts such that the $N+1$ columns of $\boldsymbol{\Psi}$ equal the columns of the $\frac{T}{K} \times \frac{T}{K}$ matrix that defines the $\frac{T}{K}$-point Discrete Fourier Transform (DFT) [1], [5]:

$$[\boldsymbol{\Psi}]_{mn} = e^{j2\pi(m-1)(n-1)/(T/K)} \tag{16}$$

for $m = 1, \cdots, \frac{T}{K}$ and $n = 1, \cdots, N+1$. If the RIS gains are assumed to be phase-independent and satisfy $\beta_i = \beta$, then this leads to $\boldsymbol{\Psi}^H \boldsymbol{\Psi} = \frac{T\beta^2}{K} \mathbf{I}_{N+1}$ and the variance of the channel coefficient estimates is $\sigma^2/(\beta^2 PT)$, a factor of $T/K \geq N+1$ smaller than in the first approach. In addition to the need for phase-independent RIS element gains, which is difficult to achieve in practice, the RIS phase shifts should be tunable with at least $\log_2(T/K)$ bits of resolution, which may be problematic for large N.
- An alternative that achieves the same performance is to choose the columns of $\boldsymbol{\Psi}$ from among the columns of a T/K-dimensional Hadamard matrix, whose entries are constrained to be ± 1 [4], [27]. This achieves orthogonality for $\boldsymbol{\Psi}$, and has the advantage of requiring only two phase states for each RIS element (one bit of resolution). In

addition, a diagonal $\boldsymbol{\Psi}^H \boldsymbol{\Psi}$ only requires that the RIS gains be equal at these two phase values. In this approach, T/K must be a multiple of 4 for the Hadamard matrix to exist, but this is not a significant issue for large N.

2) Linear Minimum Mean Squared Error

The LS approach assumes a deterministic channel with no prior information. On the other hand, the minimum mean-squared error (MMSE) estimator assumes a stochastic model for $\{\mathbf{H}_d, \mathbf{G}, \mathbf{H}\}$, usually in terms of correlated Rayleigh fading with prior information of the second-order statistics. However, the composite channel is composed of products of the Gaussian elements in $\mathbf{H}$ and $\mathbf{G}$, which makes the MMSE estimate $\mathcal{E}\{\mathbf{h}_c|\mathfrak{y}\}$ difficult to compute, although message-passing algorithms have been proposed for this problem [28–30]. Instead, the linear MMSE, or LMMSE, estimate given by $\hat{\mathbf{h}}_{c,LM} = \mathbf{W}\mathfrak{y}$ can be found by solving [31], [32]

$$\mathbf{W} = \arg\min_{\tilde{\mathbf{W}}} \mathcal{E} \left\{ \|\tilde{\mathbf{W}}\mathfrak{y} - \mathbf{h}_c\|^2 \right\} \,. \tag{17}$$

Assuming spatially and temporally white Gaussian noise uncorrelated with $\mathbf{h}_c$, the LMMSE estimate is given by

$$\hat{\mathbf{h}}_{c,LM} = \sqrt{P} \mathbf{R}_{\mathbf{h}_c} \mathbf{Z}^H \left(P \mathbf{Z} \mathbf{R}_{\mathbf{h}_c} \mathbf{Z}^H + \sigma^2 \mathbf{I}_{MT} \right)^{-1} \mathfrak{y} \,, \tag{18}$$

where $\mathbf{R}_{\mathbf{h}_c} = \mathcal{E}\{\mathbf{h}_c \mathbf{h}_c^H\}$ and we have assumed $\mathcal{E}\{\mathbf{h}_c\} = \mathbf{0}$.

Using orthogonal pilot and RIS reflection sequences like those discussed above also simplifies computation of the LMMSE estimate. For example, let $\mathbf{I} = \mathbf{I}_{MK(N+1)}$ and assume the Hadamard reflection pattern so that $\mathbf{Z}^H \mathbf{Z} = T\mathbf{I}$. Then the LMMSE estimate simplifies to

$$\hat{\mathbf{h}}_{c,LM} = \frac{1}{\sqrt{PT}} \mathbf{R}_{\mathbf{h}_c} \left(\mathbf{R}_{\mathbf{h}_c} + \frac{\sigma^2}{PT} \mathbf{I} \right)^{-1} \mathbf{Z}^H \mathfrak{y} \,. \tag{19}$$

The matrices in (19) involving $\mathbf{R}_{\mathbf{h}_c}$ are data independent, and can be computed and stored offline since $\mathbf{R}_{\mathbf{h}_c}$ changes relatively slowly. The resulting error covariance is given by

$$\mathbf{R}_{e,LM} = \mathbf{R}_{\mathbf{h}_c} - \mathbf{R}_{\mathbf{h}_c} \left(\mathbf{R}_{\mathbf{h}_c} + \frac{\sigma^2}{PT} \mathbf{I} \right)^{-1} \mathbf{R}_{\mathbf{h}_c} \,. \tag{20}$$

For the above training protocol, $\mathbf{R}_{e,LM}$ converges to $\mathbf{R}_{e,LS} = (\sigma^2/(PT))\mathbf{I}$ for high SNR (i.e., $\sigma^2/P \to 0$) or long training intervals ($T \to \infty$).

A bigger issue than the computational complexity of (19) is how to determine the composite channel covariance $\mathbf{R}_{\mathbf{h}_c}$. In theory, the covariance could be estimated using simulations involving detailed propagation models of the environment, or by taking sample statistics of channel estimates obtained over a long period of time. However, the size of $\mathbf{h}_c$ means that such procedures would require a large amount of data. Instead, a more reasonable approach is to determine $\mathbf{R}_{\mathbf{h}_c}$ based on covariance information about its constituent parts. For MIMO channels, it is commonly assumed that the multipath scattering at the source is uncorrelated with the scattering at the destination, which leads to the following descriptions:

$$\mathbf{H} = \mathbf{R}_{HB}^{\frac{1}{2}} \tilde{\mathbf{H}} \mathbf{R}_{HR}^{\frac{H}{2}} \tag{21a}$$

$$\mathbf{G} = \mathbf{R}_{GU}^{\frac{1}{2}} \tilde{\mathbf{G}} \mathbf{R}_{GR}^{\frac{H}{2}} \tag{21b}$$

$$\mathbf{H}_d = \mathbf{R}_{H_dB}^{\frac{1}{2}} \tilde{\mathbf{H}}_d \mathbf{R}_{H_dU}^{\frac{H}{2}} \,, \tag{21c}$$

---

[1] "Turning off" an RIS element assumes it becomes a perfect absorber of RF energy, which in practice is not possible. Thus, such elements will still reflect a small amount of energy and thus degrade the orthogonality assumption.

where the subscripts $\{B, R, U\}$ respectively correspond to BS, RIS, and UE, and indicate which side of the link the correlation matrix is associated with (*e.g.*, $\mathbf{R}_{HB}$ is the correlation matrix for the BS-side of the channel $\mathbf{H}$). The matrices $\{\tilde{\mathbf{H}}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}}_d\}$ are of the same dimensions as $\{\mathbf{H}, \mathbf{G}, \mathbf{H}_d\}$ respectively, and are composed of uncorrelated $\mathcal{CN}(0, 1)$ elements. Under this model, it can be shown that the composite channel covariance matrix has the following form:

$$\mathbf{R}_{\mathbf{h}_c} = \left[\begin{array}{cc} \mathbf{R}_{H_dU}^T \otimes \mathbf{R}_{H_dB} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{R}_R \otimes \mathbf{R}_{GU}^T \otimes \mathbf{R}_{HB} \end{array} \right] , \quad (22)$$

where we define $\mathbf{R}_R = \mathbf{R}_{GR} \odot \mathbf{R}_{HR}^T$.

Estimating the correlation matrices $\mathbf{R}_{H_dU}, \mathbf{R}_{H_dB}, \mathbf{R}_{GU}$ and $\mathbf{R}_{HB}$ is relatively straightforward since the BS and UEs have active transceivers that can collect and process data. However, determining the RIS-side correlation matrices $\mathbf{R}_{GR}$ and $\mathbf{R}_{HR}$ is problematic since the RIS is typically passive. Various assumptions can be made to further simplify $\mathbf{R}_{\mathbf{h}_c}$. With uncorrelated scattering at the RIS, $\mathbf{R}_{GR}$ and $\mathbf{R}_{HR}$ can be taken as identity matrices, and $\mathbf{R}_{\mathbf{h}_c}$ is block diagonal with identical block entries except for the block associated with $\mathbf{H}_d$. This greatly simplifies computation of (19). If we go a step further and assume all channels exhibit uncorrelated Rayleigh fading, then the LMMSE estimate and error simplify to

$$\hat{\mathbf{h}}_{c,LM} = \frac{1}{\sqrt{PT}} \left[\begin{array}{cc} \nu_{H_d}\mathbf{I}_{MK} & \mathbf{0} \\ \mathbf{0} & \nu_{GH}\mathbf{I}_{MNK} \end{array} \right] \mathbf{Z}^H \mathfrak{y} \quad (23a)$$

$$\mathbf{R}_{e,LM} = \frac{\sigma^2}{PT} \left[\begin{array}{cc} \nu_{H_d}\mathbf{I}_{MK} & \mathbf{0} \\ \mathbf{0} & \nu_{GH}\mathbf{I}_{MNK} \end{array} \right] , \quad (23b)$$

where

$$\nu_{H_d} = \frac{PT\sigma_{H_d}^2}{PT\sigma_{H_d}^2 + \sigma^2} < 1 \quad (24a)$$

$$\nu_{GH} = \frac{PT\sigma_G^2\sigma_H^2}{PT\sigma_G^2\sigma_H^2 + \sigma^2} < 1 , \quad (24b)$$

and $\sigma_{H_d}^2, \sigma_H^2, \sigma_G^2$ represent the variances of the channels $\mathbf{H}_d, \mathbf{H}, \mathbf{G}$, respectively. Since ν_{H_d} and ν_{GH} are less than one, the LMMSE estimates have a smaller error than for LS, which is due to the exploitation of the prior statistical information. However, assumptions of uncorrelated fading are hard to justify in RIS-aided wireless systems, which are typically motivated by propagation environments with sparse propagation paths and frequent blockages. In these environments, the BS and RIS installations are envisioned to be in elevated positions away from nearby RF scatterers. This leads to low-rank channel correlation matrices and consideration of geometric models, as discussed in Section IV.

B. Wideband Single User MIMO

In wideband scenarios where the channel is frequency selective, we assume the UE transmits an OFDM signal composed of N_c subcarriers from each of its k antennas. The symbols are given by the rows of the $K \times N_c$ matrix $\mathbf{X}_t^F = \left[\mathbf{x}_{t,1}^F \cdots \mathbf{x}_{t,N_c}^F \right]$ in the frequency domain, where here t is the OFDM symbol index. Prior to transmission, the data $\mathbf{X}_t^F$ is first converted to the time domain using the $N_c \times N_c$ matrix $\mathbf{F}^H$ that denotes the N_c-point inverse DFT: $\mathbf{X}_t = \mathbf{X}_t^F \mathbf{F}^H$, and

then is appended with a cyclic prefix of length L_{cp} that is longer than the maximum delay spread of the channel, L. At the BS, the cyclic prefix is removed, and the data are converted back to the frequency domain through multiplication by the DFT matrix $\mathbf{F}$. This generates a model essentially identical to (1) for each subcarrier n:

$$\mathbf{y}_{t,n}^F = \sqrt{P} \left(\mathbf{H}_{d,n}^F + \mathbf{H}_n^F \boldsymbol{\Phi}_{t,n} \mathbf{G}_n^{FH} \right) \mathbf{x}_{t,n}^F + \mathbf{n}_{t,n}^F , \quad (25)$$

where $\{\mathbf{H}_{d,n}^F, \mathbf{H}_n^F, \mathbf{G}_n^F\}$ represent the DFT at subcarrier n for the UE-BS, RIS-BS, and UE-RIS channel impulse responses, respectively. Thus, one can employ the same estimation methods discussed above on a per-subcarrier basis, although to exploit the channel correlation in frequency and reduce the training overhead, pilot data is normally transmitted only on a subset of the subcarriers, and interpolation used to construct channel estimates for others [2]. An alternative approach proposed in [11] is to use shorter OFDM symbols during the training period.

Note that most prior work on RIS channel estimation with OFDM signals has assumed that the RIS reflection properties are frequency independent, *i.e.*, $\boldsymbol{\Phi}_{t,n} = \boldsymbol{\Phi}_t$, but this is generally true only for relatively narrow bandwidths [33], [34]. If one sets $\boldsymbol{\phi}_{t,n}$ to have desirable properties (*e.g.*, $\mathbf{Z}$ with orthogonal columns) at a particular subcarrier n, then in general those properties will not be inherited at other subcarriers. This issue motivates the design of RIS circuit architectures that have invariant properties across wider frequency bands.

An alternative to estimating the channels in the frequency domain and using interpolation is to directly estimate the channel impulse response. In the time domain, we represent the data received for sample s of OFDM symbol t as

$$\mathbf{y}_{t,s} = \sqrt{P} \sum_{k=0}^{L-1} \left(\mathbf{H}_d(k) + \mathbf{H}(k)\boldsymbol{\Phi}_{t,s-k}\mathbf{G}^H(k) \right) \mathbf{x}_{t,s-k} + \mathbf{n}_{t,s} , \quad (26)$$

where $\{\mathbf{H}_d(k), \mathbf{H}(k), \mathbf{G}(k)\}_{k=0}^{L-1}$ represent the channel impulse responses and L is the maximum number of taps. Defining $\mathbf{h}_d(k) = \text{vec}(\mathbf{H}_d(k))$ and $\mathbf{h}_c(k) = \text{vec}\left([\mathbf{h}_d(k) \ \mathbf{G}^*(k) \diamond \mathbf{H}(k)] \right)$, after removal of the cyclic prefix we can write

$$\mathbf{y}_{t,s} = \sqrt{P} \sum_{k=0}^{L-1} \left[\tilde{\boldsymbol{\phi}}_{t,s-k}^T \otimes \mathbf{x}_{t,s-k}^T \otimes \mathbf{I}_M \right] \mathbf{h}_c(k) + \mathbf{n}_{t,s} \quad (27a)$$

$$= \sqrt{P} \sum_{k=0}^{L-1} \mathbf{Z}_{t,s-k}\mathbf{h}_c(k) + \mathbf{n}_{t,s} \quad (27b)$$

$$= \sqrt{P} \left[\mathbf{Z}_{t,s} \ \mathbf{Z}_{t,s-1} \ \cdots \ \mathbf{Z}_{t,s-L+1} \right] \mathfrak{h}_c + \mathbf{n}_{t,s} \quad (27c)$$

$$\mathfrak{y}_t = \left[\begin{array}{c} \mathbf{y}_{t,1} \\ \vdots \\ \mathbf{y}_{t,N_c} \end{array} \right] = \sqrt{P} \mathfrak{Z}_t \mathfrak{h}_c + \mathfrak{n}_t , \quad (27d)$$

where $\mathfrak{h}_c = \left[\mathbf{h}_c^T(0) \ \cdots \ \mathbf{h}_c^T(L-1) \right]^T$ is the $LMK(N+1) \times 1$ vector containing all unknown channel coefficients, and $\mathfrak{Z}_t$ is an $MN_c \times LMK(N+1)$ block-circulant matrix with first

block row $[\mathbf{Z}_{t,1}\ \mathbf{Z}_{t,N_c}\ \cdots\ \mathbf{Z}_{t,N_c-L+2}]$. Finally, assuming the channel is stationary over T_o total OFDM symbols, we have

$$\mathfrak{y} = \sqrt{P}\begin{bmatrix} \mathfrak{Z}_1 \\ \vdots \\ \mathfrak{Z}_{T_o} \end{bmatrix}\mathfrak{h}_c + \mathfrak{n} = \sqrt{P}\mathfrak{Z}\mathfrak{h}_c + \mathfrak{n}\ . \qquad (28)$$

The time-domain approach assumes only pilot data is transmitted first, followed by payload data. The total number of pilot symbols required is $T = T_o N_c \geq KL(N+1)$. While more OFDM symbols are likely required for the frequency domain method to obtain the same channel estimation accuracy, this is offset by the fact that data and pilots can be transmitted together.

C. Single Antenna Scenarios

1) Single Antenna UE

The single-antenna UE case is often considered in the literature, since it simplifies the notation and reduces the algorithm complexity, but there is fundamentally little difference with the general multi-antenna UE case described above. The channel $\mathbf{G}$ becomes a $1 \times N$ row vector that we denote by $\mathbf{g}^T$, while the direct channel $\mathbf{H}_d$ is simply an $M \times 1$ vector $\mathbf{h}_d$. The pilot data received at the BS is given by

$$\mathbf{y}_t = \sqrt{P}\left(\mathbf{h}_d + \mathbf{H}\,\mathrm{diag}(\mathbf{g}^*)\boldsymbol{\phi}_t\right)x_t + \mathbf{n}_t\ , \qquad (29)$$

where the composite channel is now $\mathbf{H}_c = \mathbf{g}^H \diamond \mathbf{H} = \mathbf{H}\,\mathrm{diag}(\mathbf{g}^*)$. The training overhead in this case is reduced to $N + 1$ samples.

2) Single Antenna BS and UE

When both the BS and UE have only a single antenna, we denote the RIS-BS channel as the $N \times 1$ row vector $\mathbf{h}^T$, and write the BS output and composite channel as

$$y_t = \sqrt{P}\mathbf{h}_c^T\tilde{\boldsymbol{\phi}}_t x_t + n_t \qquad (30\text{a})$$

$$\mathbf{h}_c^T = \begin{bmatrix} h_d & \mathbf{g}^H \odot \mathbf{h}^T \end{bmatrix} = \begin{bmatrix} h_d & \bar{\mathbf{h}}_c^T \end{bmatrix}\ , \qquad (30\text{b})$$

where only $\bar{\mathbf{h}}_c$ is identifiable.

D. Multiple User Scenarios

The models and approaches discussed above are easily generalized to the multiple UE case. Assuming UE u has K_u antennas for $u = 1, \cdots, U$, then the model in (1) holds if we simply set $K = \sum_u K_u$ and all UE antennas transmit orthogonal pilot sequences. Some prior work has proposed that the users take turns transmitting pilots, in which case there is no change to the algorithms described above, but this only makes sense if one exploits the fact that each user's composite channel shares a common RIS-BS component $\mathbf{H}$. This idea will be explored further in the next subsection. For multicarrier signals, a scheme is required to allocate the pilot subcarriers to the UEs, but otherwise the channel estimation is the same. One implication for the LMMSE approach is that, assuming the channels for different UEs are uncorrelated, the matrices $\mathbf{R}_{GU}$ and $\mathbf{R}_{H_cU}$ will be block-diagonal.

E. Reducing the Complexity and Training Overhead

As noted already above, one of the key hurdles to overcome in CSI estimation for RIS-aided systems is the large required training overhead. Consequently, recent work has focused on

a variety of methods to reduce this overhead, some of which is described below. The use of geometric channel models to reduce pilot overhead is reserved for Section IV.

1) RIS Element Grouping

A simple approach to reduce the number of pilots and estimation complexity is to assign identical phases to RIS elements with highly correlated channels [2], [4]. High channel correlation occurs when adjacent RIS elements are closely spaced; retaining the flexibility of arbitrary phase shifts for such elements provides minimal additional beamforming gains. Suppose groups of size J are identifed, and assume for simplicity that $N' = N/J$ is an integer and no direct channel $\mathbf{H}_d$ is present. Then we define $\boldsymbol{\phi}_t = \boldsymbol{\phi}'_t \otimes \mathbf{1}_J$, where $\boldsymbol{\phi}'_t$ is $N' \times 1$, and write

$$\mathbf{H}_c\boldsymbol{\phi}_t = \mathbf{H}_c(\boldsymbol{\phi}'_t \otimes \mathbf{1}_J) = \mathbf{H}_c(\mathbf{I}_{N'} \otimes \mathbf{1}_J)\boldsymbol{\phi}'_t = \mathbf{H}'_c\boldsymbol{\phi}'_t\ , \quad (31)$$

where the effective composite channel $\mathbf{H}'_c$ is now $MK \times N'$. Each column of $\mathbf{H}'_c$ is thus a unit-coefficient linear combination of the columns of $\mathbf{H}_c$ corresponding to a given group of RIS elements. The revised model is identical in form to the general case, and thus the methods described above can be implemented to estimate $\mathbf{H}'_c$ with a reduction in the required training overhead by a factor of J. A generalization of this idea presented in [4] successively reduces the size of the groups over multiple blocks of pilot and payload data in order to eventually resolve the channels for all of the RIS elements.

2) Low-Rank Channel Covariance

We see from the noise-free part of (5), $\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{h}_c$, that in the general case, the $MT \times MK(N+1)$ data matrix $\mathbf{Z}$ should be full rank $MK(N+1)$, since otherwise components of $\mathbf{h}_c$ in the nullspace of $\mathbf{Z}$ could not be identified. Like the LS approach, this requires $T \geq K(N+1)$ training samples. However, if $\mathbf{R}_{\mathbf{h}_c}$ is rank deficient, then it would be enough for the column span of $\mathbf{R}_{\mathbf{h}_c}$ to lie within the column span of $\mathbf{Z}^T$. In particular, suppose $\mathbf{R}_{\mathbf{h}_c}$ is rank $r < MK(N+1)$, and thus can be factored as $\mathbf{R}_{\mathbf{h}_c} = \mathbf{U}\mathbf{U}^H$, where $\mathbf{U}$ has r columns. Then in principle it would be sufficient to choose

$$\mathbf{Z}^T = \mathbf{U}\mathbf{V} \qquad (32)$$

for some full rank $r \times MT$ matrix $\mathbf{V}$, and thus theoretically it would be sufficient that $T \geq r/M$. Unfortunately, due to constraints on the possible values for $\boldsymbol{\phi}_t$, finding a $\mathbf{V}$ that exactly satisfies (32) is generally not possible if $T < K(N+1)$. It may however be possible to approximately solve (32) for larger values of T that are still much smaller than $K(N+1)$, provided that r is not too large. In addition to reducing the training overhead, the low rank channel covariance can be exploited to significantly reduce the cost of computing the LMMSE solution in (18), since only an $r \times r$ inverse rather than an $MT \times MT$ inverse is required:

$$\hat{\mathbf{h}}_c = \frac{\sqrt{P}}{\sigma^2}\mathbf{U}\left[\mathbf{I}_r - \mathbf{W}\left(\mathbf{W} + \frac{\sigma^2}{P}\mathbf{I}_r\right)^{-1}\right]\mathbf{U}^H\mathbf{Z}^H\mathfrak{y}, \quad (33)$$

where $\mathbf{W} = \mathbf{U}^H\mathbf{Z}^H\mathbf{Z}\mathbf{U}$.

3) Exploiting Common Channels

The LS method in Section III-A1 ignores the Kronecker product structure of the composite channel, which can be exploited to reduce the training overhead. The key observation is that, in the uplink, the composite channel for each user shares the same RIS-BS channel $\mathbf{H}$ [8]. To explain how this information can be exploited, assume without loss of generality a scenario with K single-antenna users. The approach is divided into two steps [7], [10]. In the first, one of the users is selected and the composite channel for this user is estimated in the normal way, while the other users do not transmit. Then, in the second step, the other users transmit and the estimate of the RIS-BS channel obtained in the first step is exploited to reduce the training required for the remaining channels.

Assume the users are ordered such that the user corresponding to the first row of $\mathbf{G}$, denoted by $\mathbf{g}_1^T$, is the one selected for the first step. The LS method is used to estimate the composite channel $\mathbf{H}\,\mathrm{diag}(\mathbf{g}_1)$ and the direct channel $\mathbf{h}_{d,1}$, which requires at least $T_1 = N + 1$ training samples. Recall that only the product $\mathbf{H}\,\mathrm{diag}(\mathbf{g}_1)$ is estimated and not the individual terms $\mathbf{H}$ and $\mathbf{g}_1$. In fact, we can treat $\mathrm{diag}(\mathbf{g}_1)$ as $\mathbf{\Lambda}$ in (2), so step 1 provides us with an estimate of $\hat{\mathbf{H}}$, and we can set the first row of $\tilde{\mathbf{G}}^H$ to $\mathbf{1}_K$. With the estimate $\hat{\hat{\mathbf{H}}}$, during step 2 the training data model is approximately given by

$$\mathbf{y}_t \simeq \sqrt{P}\left(\tilde{\mathbf{H}}_d + \hat{\hat{\mathbf{H}}}\mathbf{\Phi}_t\tilde{\mathbf{G}}^H\right)\mathbf{x}_t + \mathbf{n}_t \qquad (34\mathrm{a})$$

$$\simeq \sqrt{P}\,\mathbf{x}_t^T \otimes \underbrace{\left[\mathbf{I}_M \quad \hat{\hat{\mathbf{H}}}\mathbf{\Phi}_t\right]}_{M \times (K-1)(M+N)}\begin{bmatrix}\tilde{\mathbf{h}}_d \\ \tilde{\mathbf{g}}^*\end{bmatrix} + \mathbf{n}_t \qquad (34\mathrm{b})$$

$$\simeq \sqrt{P}\tilde{\mathbf{Z}}_t\tilde{\mathbf{h}}_c + \mathbf{n}_t \,, \qquad (34\mathrm{c})$$

where we drop the first column of $\mathbf{H}_d$ to create $\tilde{\mathbf{H}}_d$, and we drop the first row of ones in $\tilde{\mathbf{G}}$, since UE 1 does not transmit. We also have defined $\tilde{\mathbf{g}} = \mathrm{vec}(\tilde{\mathbf{G}}^T)$ and $\tilde{\mathbf{h}}_c^T = \begin{bmatrix}\tilde{\mathbf{h}}_d^T & \tilde{\mathbf{g}}^H\end{bmatrix}$. Stacking T_2 of these training vectors together, we get an equation analogous to (5), where in this case $\mathbf{Z}$ is $MT_2 \times (M + N)(K - 1)$. Assuming linearly independent pilots $\mathbf{x}_t$ and RIS reflection vectors $\boldsymbol{\phi}_t$ are chosen, we can solve for the remaining channel parameters using $\hat{\tilde{\mathbf{h}}}_c = \mathbf{Z}^\dagger\mathfrak{y}/\sqrt{P}$ provided that $MT_2 \geq (K - 1)(M + N)$, or equivalently, $T_2 \geq (K - 1)(\frac{N}{M} + 1)$. Given the $N + 1$ samples needed for step 1, the minimum required training time is thus

$$T_{min} = (N + 1) + \left(\frac{N}{M} + 1\right)(K - 1)\,, \qquad (35)$$

which for large M is significantly less than the value $K(N+1)$ required by the standard LS method.

IV. ESTIMATION OF STRUCTURED CHANNELS

The large training overhead required for unstructured channel estimation motivates the consideration of channel models that are described by fewer parameters. Such models are often used in millimeter wave or higher frequency bands, where multipath scattering is sparse and propagation is often dominated by strong specular components. In such cases, the channels can be described by a small number of propagation paths defined by path gains, angles of arrival (AoAs), and angles of departure

(AoDs)[2]. The resulting number of parameters is often more 1-2 orders of magnitude less than that required in the unstructured case, and the training overhead is correspondingly reduced.

Parametric channels are described by the array response or "steering" vectors associated with the angle of an incoming (AoA) or outgoing (AoD) signal. For example, the response of an M_x-element uniform linear array (ULA) to a signal arriving with azimuth angle θ_{az} is described by the Vandermonde vector

$$\mathbf{a}_x(\omega_x) = [1 \; e^{j\omega_x} \; e^{j2\omega_x} \; \cdots \; e^{j(M_x-1)\omega_x}]^T \,, \qquad (36)$$

where the spatial frequency ω_x is defined by $\omega_x = 2\pi\Delta_x\sin(\theta_{az})$, and Δ_x is the distance in wavelengths between the antennas[3]. For an $M_x \times M_y$ uniform rectangular array (URA) with antenna separations of Δ_x and Δ_y in the x and y directions, the array response vector can be written as

$$\mathbf{a}(\boldsymbol{\omega}) = \mathbf{a}_x(\omega_x) \otimes \mathbf{a}_y(\omega_y)\,, \qquad (37)$$

where the vertical array response component is similar to (36),

$$\mathbf{a}_y(\omega_y) = [1 \; e^{j\omega_2} \; e^{j2\omega_2} \; \cdots \; e^{j(M_y'-1)\omega_2}]^T \,, \qquad (38)$$

but defined by $\omega_y = 2\pi\Delta_y\sin(\theta_{el})\cos(\theta_{az})$ with elevation AoA θ_{el}. The vector $\boldsymbol{\omega} = [\omega_x \; \omega_y]^T$ corresponds to a 2D spatial frequency. For either a ULA or URA, there is a one-to-one correspondence between the angles and spatial frequencies as long as $\{\Delta_x, \Delta_y\}$ are no more than one-half wavelength. This is important for applications involving localization, since the angles provide useful information for locating a signal source. However, from the viewpoint of channel estimation, it is enough to know $\boldsymbol{\omega}$, and any ambiguities in determining the angles need not be resolved.

In this section we focus on estimation of structured or geometric channel models. To simplify the discussion, we assume that the direct UE-BS channel is absent. This assumption is typical for scenarios with low-rank near-specular propagation at high frequencies, where blockages are common. The extra steps and pilot data required to estimate $\mathbf{H}_d$ when it is present generate minimal additional overhead. We will further assume that the BS and the UEs (when they have multiple antennas) employ ULAs, so that their array response depends on a single angle/spatial frequency, and we assume that the RIS elements are arranged as a URA, so its spatial response depends on two spatial frequencies. Generalizations to arbitrary array geometries are straightforward.

A. Parametric Estimation

The structured channel estimators that we will consider assume parametric channel models of the following form, which we describe first for the RIS-BS channel:

$$\mathbf{H} = \sum_{k=1}^{d_H} \gamma_{H,k}\mathbf{a}_B(\boldsymbol{\omega}_{BH,k})\mathbf{a}_U^H(\boldsymbol{\omega}_{RH,k}) \qquad (39\mathrm{a})$$

$$= \mathbf{A}_B(\boldsymbol{\omega}_{BH})\mathbf{\Gamma}_H\mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\,, \qquad (39\mathrm{b})$$

---

[2]For very large RIS, where the BS or UEs are in the Fresnel region of the RIS, the channel parameterization must also include range or the 3-D coordinates of the various devices, and the large scale fading becomes antenna-dependent. However, here we focus on the more common far-field scenario.

[3]Note that we assume a narrowband propagation model here where time delays can be represented by phase shifts. For large arrays, ignoring the frequency dependence of the model leads to the beam-squint effect [22].

where the columns of

$$\mathbf{A}_B(\boldsymbol{\omega}_{BH}) = [\mathbf{a}_B(\omega_{BH,1}) \quad \cdots \quad \mathbf{a}_B(\omega_{BH,d_H})] \quad (39c)$$

$$\mathbf{A}_R(\boldsymbol{\omega}_{RH}) = [\mathbf{a}_R(\omega_{RH,1}) \quad \cdots \quad \mathbf{a}_R(\omega_{RH,d_H})] \quad (39d)$$

respectively represent the steering vectors for the propagation paths with AoAs $\boldsymbol{\omega}_{BH} = [\omega_{BH,1} \cdots \omega_{BH,d_H}]^T$ at the BS and AoDs $\boldsymbol{\omega}_{RH} = [\boldsymbol{\omega}_{RH,1}^T \cdots \boldsymbol{\omega}_{RH,d_H}]^T$ from the RIS. The diagonal matrix $\boldsymbol{\Gamma}_H = \text{diag}\{\boldsymbol{\gamma}_H\} = \text{diag}\{\gamma_{H,1} \cdots \gamma_{H,d_H}\}$ contains the complex path gains $\boldsymbol{\gamma}_H = [\gamma_{H,1} \cdots \gamma_{H,d_H}]^T$. The RIS AoDs for path k, denoted by $\boldsymbol{\omega}_{RH,k}$, are written as vectors since the RIS spatial frequencies are two-dimensional:

$$\boldsymbol{\omega}_{RH,k} = \begin{bmatrix} \omega_{RH,k,x} \\ \omega_{RH,k,y} \end{bmatrix} . \quad (40)$$

Parametric models like (39) are usually employed when the number of paths d_H is smaller than the array dimensions M and N, and thus the channel $\mathbf{H}$ is low-rank.

Parametric CSI estimation involves finding the spatial frequencies of signals collected by an array. For example, suppose n observations are available from an arbitrary M'-element array receiving signals from d directions:

$$\mathbf{Y}' = \mathbf{A}(\boldsymbol{\omega}')\mathbf{S}' + \mathbf{N}' , \quad (41)$$

where $\mathbf{Y}'$ is $M' \times n$, $\mathbf{S}'$ is $d \times n$, $\mathbf{N}'$ is noise, $\mathbf{A} = [\mathbf{a}(\boldsymbol{\omega}_1') \quad \cdots \quad \mathbf{a}(\boldsymbol{\omega}_d')]$ is the $M' \times d$ array response matrix, and $M' > d$. The matrix $\mathbf{S}'$ is not typically assumed to be known. This is the classical model assumed for AoA estimation, and many methods have been developed to estimate $\boldsymbol{\omega}' = [\boldsymbol{\omega}_1' \cdots \boldsymbol{\omega}_d']^T$ from $\mathbf{Y}'$. The simplest method is based on (matched filter) beamforming, which involves searching for d peaks in the spectrum

$$p_B(\boldsymbol{\omega}) = \mathbf{a}^H(\boldsymbol{\omega})\mathbf{R}_{Y'}\mathbf{a}(\boldsymbol{\omega}) , \quad (42)$$

where $\mathbf{R}_{Y'}$ is the sample covariance matrix

$$\mathbf{R}_{Y'} = \frac{1}{n}\mathbf{Y}'\mathbf{Y}'^H . \quad (43)$$

Alternatively, one can employ higher resolution algorithms such as the MUSIC [35] or ESPRIT [36], which require computation of the eigendecompositon of $\mathbf{R}_{Y'}$. The beamforming and MUSIC spectra are either one- or two-dimensional functions, depending on whether the steering vectors depend on one- or two-dimensional spatial frequencies. If $\mathbf{N}'$ is spatially and temporally white, the (deterministic) maximimum likelihood (ML) method [37] finds the AoA estimates from the d-dimensional (or $2d$-dimensional for azimuth/elevation angles) problem

$$\hat{\boldsymbol{\omega}}_{ML}' = \arg\min_{\boldsymbol{\omega}} \text{trace}\left(\mathbf{P}_A^{\perp}(\boldsymbol{\omega})\mathbf{R}_{Y'}\right) , \quad (44)$$

where $\mathbf{P}_A^{\perp}(\boldsymbol{\omega}) = \mathbf{I}_{M'} - \mathbf{A}(\boldsymbol{\omega})\left[\mathbf{A}^H(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega})\right]^{-1}\mathbf{A}^H(\boldsymbol{\omega})$. The corresponding ML estimate of $\mathbf{S}'$ is given by $\hat{\mathbf{S}}' = \mathbf{A}^{\dagger}(\hat{\boldsymbol{\omega}}_{ML}')\mathbf{Y}'$, where $(\cdot)^{\dagger}$ represents the pseudo-inverse, although estimates of $\hat{\boldsymbol{\omega}}'$ from other algorithms can be substituted for $\boldsymbol{\omega}_{ML}'$ to estimate $\mathbf{S}'$. MUSIC and ESPRIT require that $\text{rank}(\mathbf{S}') = d$ and thus that $n \geq d$, but the beamforming and ML methods are theoretically viable for any value of n, including $n = 1$. For a massive antenna array where M' is

large, all of the above methods provide asymptotically efficient AoA estimates [38], and thus a simple technique such as beamforming is preferred due to its low computational load and minimal assumptions. A key requirement for all of the above methods is that the value of d be known or estimated from the data.

B. Compressive Sensing

Compressive sensing (CS) formulations of the geometric CSI estimation problem are also possible, using sparse representations from an overcomplete dictionary [39–42]. For example, we can represent the array response vectors for the BS side as

$$\mathbf{A}_B(\boldsymbol{\omega}_{BH}) = \mathbf{A}_{BD}\mathbf{Q}_{BH} , \quad (45)$$

where $\mathbf{A}_{BD}$ is an $M \times N_{BD}$ matrix whose columns are BS array response vectors sampled on an N_{BD} grid of frequencies corresponding to the possible BS AoAs, and $\mathbf{Q}_{BH}$ is an $N_{BD} \times d_H$ matrix whose k-th column has a single 1 in the position corresponding to $\omega_{BH,k}$, assuming it is one of the grid points. If $\omega_{BH,k}$ is not on the grid, then the model in (45) is an approximation. While the error can be made small by making N_{BD} large, increasing the coherence of the dictionary eventually leads to computational and numerical issues. More will be said on this topic below. The RIS also has a similar overcomplete representation:

$$\mathbf{A}_R(\boldsymbol{\omega}_{RH}) = \mathbf{A}_{RD}\mathbf{Q}_{RH} , \quad (46)$$

where $\mathbf{A}_{RD}$ is $N \times N_{RD}$ and $\mathbf{Q}_{RH}$ is $N_{RD} \times d_H$. Because the RIS AoDs are two dimensional, a two-dimensional grid is necessary to specify the AoD pairs, and hence $N_{RD} = O(N_{BD}^2)$. Substituting (45) and (46) into (39), we can write

$$\mathbf{H} = \mathbf{A}_{BD}\mathbf{Q}_{BH}\boldsymbol{\Gamma}_H\mathbf{Q}_{RH}^T\mathbf{A}_{RD}^H \quad (47a)$$

$$\mathfrak{h} = \text{vec}(\mathbf{H}) = (\mathbf{A}_{RD}^* \otimes \mathbf{A}_{BD})\boldsymbol{\gamma}_{HD} , \quad (47b)$$

where $\boldsymbol{\gamma}_{HD} = \text{vec}(\mathbf{Q}_{BH}\boldsymbol{\Gamma}_H\mathbf{Q}_{RH}^T)$ is a d_H-sparse vector of length $N_{BD}N_{RD}$ whose d_H non-zero elements are equal to $\boldsymbol{\gamma}_H$. The $MN \times N_{BD}N_{RD}$ matrix $\mathbf{A}_{RD}^* \otimes \mathbf{A}_{BD}$ can be thought of as an overcomplete dictionary for the vectorized channel $\mathfrak{h}$.

Consider an overcomplete representation of the data in (41):

$$\mathbf{Y}' = \mathbf{A}_D\mathbf{Q}'\mathbf{S}' + \mathbf{N}' = \mathbf{A}_D\mathbf{C}' + \mathbf{N}' , \quad (48)$$

where the dictionary $\mathbf{A}_D$ is $M' \times N_D$ and the matrix $\mathbf{Q}'$ is $N_D \times d$, with one non-zero element per column equal to one. As before $\mathbf{S}'$ is $d \times n$. The matrix $\mathbf{C}' = \mathbf{Q}'\mathbf{S}'$ exhibits common row sparsity; only d rows have non-zero elements, and these rows are equal in some order to the rows of $\mathbf{S}'$. If $\mathbf{C}'$ can be estimated, then the location of its non-zero rows would correspond to the AoAs, and the entries in those rows to the rows of $\mathbf{S}'$. The problem can be solved, for example, by means of a convex LASSO-type approach of the form

$$\hat{\mathbf{C}}' = \arg\min_{\mathbf{C}} \frac{1}{2}\|\mathbf{Y}' - \mathbf{A}_D\mathbf{C}\|_2^2 + \lambda\|\mathbf{C}\|_{2,1} , \quad (49)$$

where $\mathbf{C}^T = [\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_{N_D}]$ and the regularization

$$\|\mathbf{C}\|_{2,1} = \sum_{k=1}^{N_D} \|\mathbf{c}_k\|_2 \quad (50)$$

promotes row sparsity. As mentioned above, the accuracy of approaches like (49) is limited by the resolution of the sampled grid, which cannot be made infinitely fine due to numerical and computational issues. Off-grid AoAs create a basis mismatch that leads to leakage of energy into adjacent rows of $\mathbf{C}'$. One approach to refine the AoA estimates obtained by (49) is to apply a small angular rotation to each selected column of $\mathbf{A}_D$ that maximizes the correlation with the received data [18]:

$$\boldsymbol{\omega}'_k = \arg\max_{\boldsymbol{\omega}} \left\| \mathrm{diag}\left(\mathbf{a}(\boldsymbol{\omega})\right) \mathbf{a}_{D,k} \mathbf{Y}'^H \right\|_2 \qquad (51)$$

for $k = 1, \cdots, d$, where $\mathbf{a}_{D,k}$ is the k-th column of $\mathbf{A}_D \hat{\mathbf{Q}}'$. The final estimate of $\boldsymbol{\omega}_k$ is the spatial frequency corresponding to $\mathbf{a}_{D,k}$ plus the estimated rotation $\boldsymbol{\omega}'_k$. A more fundamental approach to solving the basis mismatch problem for one-dimensional spatial frequencies is to recast the sparse recovery problem using the atomic norm [19], [43].

C. Single User MIMO Single Carrier

In this section, we consider structured channel estimation for the case involving a single multi-antenna UE. As with the RIS-BS channel, we can define parametric and overcomplete representations of the RIS-UE channel as follows:

$$\mathbf{G} = \mathbf{A}_U(\boldsymbol{\omega}_{UG})\boldsymbol{\Gamma}_G \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}) \qquad (52a)$$
$$= \mathbf{A}_{UD}\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G \mathbf{Q}_{RG}^T \mathbf{A}_{RD}^H \qquad (52b)$$
$$\mathfrak{g} = \mathrm{vec}\left(\mathbf{G}\right) = \left(\mathbf{A}_{RD}^* \otimes \mathbf{A}_{UD}\right)\boldsymbol{\gamma}_{GD}, \qquad (52c)$$

where we assume d_G paths with gains $\boldsymbol{\gamma}_G = [\gamma_{G,1} \cdots \gamma_{G,d_G}]^T$ and $\boldsymbol{\Gamma}_G = \mathrm{diag}\{\boldsymbol{\gamma}_G\}$. In this case, $\boldsymbol{\gamma}_{GD} = \mathrm{vec}(\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G \mathbf{Q}_{RG}^T)$ is a d_G-sparse vector of length $N_{UD}N_{RD}$ whose non-zero elements correspond to $\boldsymbol{\gamma}_G$.

Recall the general model in (5), where the BS data from the T training samples is stacked together in a single vector:

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{h}_c + \mathfrak{n}, \qquad (53)$$

where

$$\mathbf{Z} = \begin{bmatrix} \boldsymbol{\phi}_1^T \otimes \mathbf{x}_1^T \otimes \mathbf{I}_M \\ \vdots \\ \boldsymbol{\phi}_T^T \otimes \mathbf{x}_T^T \otimes \mathbf{I}_M \end{bmatrix} \qquad (54)$$

is $MT \times MKN$ and involves $\boldsymbol{\phi}_t$ instead of $\tilde{\boldsymbol{\phi}}_t$ since we are assuming no direct UE-BS channel is present, which also implies that $\mathbf{h}_c = \mathrm{vec}(\mathbf{H}_c) = \mathrm{vec}\left(\mathbf{G}^* \diamond \mathbf{H}\right)$. Using various properties of the Kronecker and Khatri-Rao products, the composite channel can be decomposed using either a parametric approach or via overcomplete dictionaries as follows:

$$\mathbf{H}_c = \underbrace{\left(\mathbf{A}_U^*(\boldsymbol{\omega}_{UG}) \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH})\right)}_{MK \times d_H d_G} \underbrace{\boldsymbol{\Gamma}_{GH}}_{d_H d_G \times d_H d_G}$$
$$\times \underbrace{\left(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\right)}_{d_H d_G \times N} \qquad (55a)$$
$$= \underbrace{\left(\mathbf{A}_{UD}^* \otimes \mathbf{A}_{BD}\right)}_{MK \times N_{UD}N_{BD}} \underbrace{\mathbf{Q}_{GH}}_{N_{UD}N_{BD} \times N_{RD}^2} \underbrace{\left(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H\right)}_{N_{RD}^2 \times N}, \qquad (55b)$$

VARIABLE	PARAMETERS	DESCRIPTION
$\boldsymbol{\omega}_{BH}$	d_H	AoA frequencies at BS
$\boldsymbol{\omega}_{RH}$	$2(d_H - 1)$	AoD frequencies at RIS
$\boldsymbol{\gamma}_H$	$2(d_H - 1)$	complex gains for BS-RIS channel
$\boldsymbol{\omega}_{RG}$	$2d_G$	AoA frequencies at RIS
$\boldsymbol{\omega}_{UG}$	d_G	AoD frequencies at UE
$\boldsymbol{\gamma}_G$	$2d_G$	complex path gains for RIS-UE channel

TABLE I: Composite Channel Parameterization for Single User MIMO Single Carrier Case. The elements of these vectors form the elements of the parameter vector $\boldsymbol{\eta}$ for geometric channel models.

where

$$\boldsymbol{\Gamma}_{GH} = \boldsymbol{\Gamma}_G^* \otimes \boldsymbol{\Gamma}_H \qquad (56a)$$
$$\mathbf{Q}_{GH} = \underbrace{\left(\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G^* \mathbf{Q}_{RG}^T\right)}_{N_{UD} \times N_{RD}} \otimes \underbrace{\left(\mathbf{Q}_{BH}\boldsymbol{\Gamma}_H \mathbf{Q}_{RH}^T\right)}_{N_{BD} \times N_{RD}} \qquad (56b)$$
$$= \underbrace{\left(\mathbf{Q}_{UG} \otimes \mathbf{Q}_{BH}\right)}_{N_{UD}N_{BD} \times d_H d_G} \boldsymbol{\Gamma}_{GH} \underbrace{\left(\mathbf{Q}_{RG} \otimes \mathbf{Q}_{RH}\right)^T}_{d_H d_G \times N_{RD}^2}. \qquad (56c)$$

As in the non-parametric case, not all of the parameters or decompositions shown in (55) are identifiable. To see this, let

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(\mathbf{a}_R(\boldsymbol{\omega}')\right) \qquad (57)$$

be a diagonal matrix formed from an arbitrary RIS array response vector for 2×1 DOA $\boldsymbol{\omega}'$. Then,

$$\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) = \mathbf{A}_R^T(\boldsymbol{\omega}_{RG})\boldsymbol{\Lambda} \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\boldsymbol{\Lambda}^*$$
$$= \mathbf{A}_R^T(\boldsymbol{\omega}'_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}'_{RH}) \qquad (58)$$

where the elements of $\boldsymbol{\omega}'_{RH}$ and $\boldsymbol{\omega}'_{RG}$ are defined as $\omega'_{RH,k} = \omega_{RH,k} + \omega'$ and $\omega'_{RG,k} = \omega_{RG,k} + \omega'$. Thus, if the RIS AoA and AoD spatial frequencies are shifted by the same amount, there is no change to the composite channel response. In addition, since the channel gains for $\mathbf{H}$ and $\mathbf{G}$ always appear together as $\boldsymbol{\Gamma}_G^* \boldsymbol{\Gamma}_H$, there is a scaling ambiguity; in particular $\boldsymbol{\gamma}_H$ and $\boldsymbol{\gamma}_G$ yield the same composite channel as $\alpha\boldsymbol{\gamma}_H$ and $\boldsymbol{\gamma}_G/\alpha^*$. To obtain an identifiable parameterization $\boldsymbol{\eta}$ for the composite channel in (55), one could for example set $\boldsymbol{\omega}_{RH,1} = [0 \ 0]^T$ so that $\mathbf{a}_R(\boldsymbol{\omega}_{RH,1})$ is a vector of ones, and set $\gamma_{H,1} = 1$. With these assumptions, the set of unique parameters $\boldsymbol{\eta}$ that describe the composite channel matrix are given in Table I. The total number of parameters is thus $\dim(\boldsymbol{\eta}) = 5d_H + 5d_G - 4$, which for typical values of d_H and d_G is much smaller than the number of parameters $2MKN$ that must be estimated for a non-parametric $\mathbf{H}_c$, which could be in the thousands.

1) Channel Estimation for the General Case

Using (55), the composite channel is given by

$$\mathbf{h}_c = \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\gamma}_{GH} \qquad (59a)$$
$$= \mathbf{A}_D \mathbf{q}_{GH}, \qquad (59b)$$

where $\boldsymbol{\gamma}_{GH} = \boldsymbol{\gamma}_G^* \otimes \boldsymbol{\gamma}_H$, $\mathbf{q}_{GH} = \mathrm{vec}(\mathbf{Q}_{GH})$,

$$\mathbf{A}(\boldsymbol{\omega}) = \left(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\right)^T$$
$$\diamond \left(\mathbf{A}_U^*(\boldsymbol{\omega}_{UG}) \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH})\right) \qquad (60a)$$
$$\mathbf{A}_D = \left(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H\right)^T \otimes \left(\mathbf{A}_{UD}^* \otimes \mathbf{A}_{BD}\right). \qquad (60b)$$

and $\boldsymbol{\omega}^T = \begin{bmatrix} \boldsymbol{\omega}_{BH}^T & \boldsymbol{\omega}_{RH}^T & \boldsymbol{\omega}_{RG}^T & \boldsymbol{\omega}_{UG}^T \end{bmatrix}$. Similarly, there are two forms of the general data model in (53):

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{A}(\boldsymbol{\omega})\boldsymbol{\gamma}_{GH} + \mathfrak{n} \tag{61a}$$

$$= \sqrt{P}\mathbf{Z}\mathbf{A}_D\mathbf{q}_{GH} + \mathfrak{n} \ . \tag{61b}$$

At first glance, Eq. (61a) has the standard form assumed in AoA estimation problems, where $\mathbf{A}(\boldsymbol{\omega}') = \mathbf{Z}\mathbf{A}(\boldsymbol{\omega})$, but there are some caveats. First, $\boldsymbol{\gamma}_{GH}$ is not arbitrary, but is instead a nonlinear function of $\boldsymbol{\gamma}_H$ and $\boldsymbol{\gamma}_G$. Second, note that the k-th column of $\mathbf{A}(\boldsymbol{\omega})$ can be expressed as

$$[\mathbf{A}(\boldsymbol{\omega})]_{:k} = \left[\mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})\right]_{:k} \otimes \mathbf{a}_U^*(\omega_{UG,\ell}) \otimes \mathbf{a}_B(\omega_{BH,p}), \tag{62}$$

where $\ell = \lfloor k/d_H \rfloor$, $p = \mathrm{mod}_{d_H}(k)$, and

$$\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH}) = \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \ . \tag{63}$$

Each column of $\mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$ only depends on a pair of 2D angles, one each from $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\omega}_{RH}$ [18]:

$$\left[\mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})\right]_{:k} = \left[\left(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\right)_{k:}\right]^T \tag{64a}$$

$$= \left[\mathbf{a}_R^T(\boldsymbol{\omega}_{RG,\ell}) \odot \mathbf{a}_R^H(\boldsymbol{\omega}_{RH,p})\right]^T \tag{64b}$$

$$= \mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) \ . \tag{64c}$$

Thus, from (62), we see that the k-th composite steering vector in (61a) is parameterized by different pairs of entries from $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\omega}_{RH}$, and from $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}$:

$$[\mathbf{A}(\boldsymbol{\omega})]_{:k} = \mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) \otimes \mathbf{a}_U^*(\omega_{UG,\ell}) \otimes \mathbf{a}_B(\omega_{BH,p}) \ . \tag{65}$$

The fact that $\mathbf{A}(\boldsymbol{\omega})$ only depends on the differences between the elements of $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\omega}_{RH}$ is a direct consequence of the fact that they are not separately identifiable, as mentioned above. An AoA estimation algorithm that takes the special structure of $\boldsymbol{\gamma}_{GH}$ into account is difficult to formulate, and one-dimensional methods such as beamforming and MUSIC cannot exploit the inherent relationship between the columns of $\mathbf{A}(\boldsymbol{\omega})$. However, assuming $\mathbf{A}(\boldsymbol{\omega})$ is full rank[4] for all possible $\boldsymbol{\omega}$, one could ignore the structure of $\boldsymbol{\gamma}_{GH}$ and use the deterministic ML (DML) criterion to estimate $\boldsymbol{\omega}$, setting $\boldsymbol{\omega}_{RH,1} = [0 \ 0]^T$ to make the model identifiable:

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \ \mathfrak{y}^H \mathbf{P}_{\mathbf{Z}\mathbf{A}(\boldsymbol{\omega})}^\perp \mathfrak{y} \ , \tag{66}$$

where $\mathbf{P}_{\mathbf{Z}\mathbf{A}(\boldsymbol{\omega})}^\perp$ is the projection orthogonal to the columns of the effective array response $\mathbf{Z}\mathbf{A}(\boldsymbol{\omega})$. This would require a non-convex optimization over the $3(d_G + d_H) - 2$ spatial frequencies in $\boldsymbol{\omega}$.

One special case worth mentioning occurs when $d_H = d_G = 1$, or when the UE-RIS and RIS-BS channels are LoS. In this case, (61a) simplifies to $\mathbf{h}_c = \gamma \mathbf{a}(\boldsymbol{\omega})$, where

$$\mathbf{a}(\boldsymbol{\omega}) = \mathbf{a}_R(\boldsymbol{\omega}_{RG}) \otimes \mathbf{a}_U^*(\omega_{UG}) \otimes \mathbf{a}_B(\omega_{BH}) \ , \tag{67}$$

---

[4]The matrix $\mathbf{A}(\boldsymbol{\omega})$ will generically be full rank as long as the BS, RIS, and UE arrays have unambiguous array manifolds (*e.g.*, elements spaced no more than $\lambda/2$ apart for a ULA or URA). However, there are pathological cases where $\mathbf{A}(\boldsymbol{\omega})$ can drop rank. This could occur for example, if any RIS angle differences are repeated, *i.e.*, $\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p} = \boldsymbol{\omega}_{RG,\ell'} - \boldsymbol{\omega}_{RH,p'}$ for $\ell \neq \ell'$ or $p \neq p'$, or if the BS and UE arrays have identical manifolds and share an angle between the vectors $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}$. Although such cases occur with probability zero, $\mathbf{A}(\boldsymbol{\omega})$ will become ill-conditioned if they are approximately true, and numerical problems would ensue.

with 4 angle parameters of interest: $\omega_{BH}, \omega_{UG}, \boldsymbol{\omega}_{RG}$. This results in a standard single-snapshot AoA estimation problem, and the vector $\boldsymbol{\omega}$ can be determined either by maximizing the beamforming criterion $|\mathfrak{y}^H \mathbf{Z}\mathbf{a}(\boldsymbol{\omega})|^2$ or minimizing the DML criterion $\mathfrak{y}^H \mathbf{P}_{\mathbf{Z}\mathbf{a}(\boldsymbol{\omega})}^\perp \mathfrak{y}$, and in either case setting $\hat{\gamma} = \mathbf{a}^\dagger(\hat{\boldsymbol{\omega}})\mathfrak{y}/\sqrt{P}$. Another way that beamforming can be applied in the general case with arbitrary d_H and d_G is to ignore the interdependence of the columns of $\mathbf{A}(\boldsymbol{\omega})$ on different combinations of the elements of $\boldsymbol{\omega}$, and just treat the angle parameters of each column as if they were independent variables. This reparameterizes the model in (61a) and (65) as

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{A}(\boldsymbol{\omega}')\boldsymbol{\gamma} + \mathfrak{n} \tag{68a}$$

$$\mathbf{A}(\boldsymbol{\omega}') = \begin{bmatrix} \mathbf{a}(\boldsymbol{\omega}_1') & \cdots & \mathbf{a}(\boldsymbol{\omega}_{d_H d_G}') \end{bmatrix} \tag{68b}$$

$$\mathbf{a}(\boldsymbol{\omega}_k') = \mathbf{a}_R(\boldsymbol{\omega}_{k,1}') \otimes \mathbf{a}_U^*(\boldsymbol{\omega}_{k,2}') \otimes \mathbf{a}_B(\boldsymbol{\omega}_{k,3}') \ , \tag{68c}$$

where $\boldsymbol{\gamma}$ is an arbitrary $d_H d_G$ vector, and $\boldsymbol{\omega}'$ has 4 elements, one each for $\boldsymbol{\omega}_{k,2}'$ and $\boldsymbol{\omega}_{k,3}'$ since they are 1D spatial frequencies, and two for $\boldsymbol{\omega}_{k,1}'$ since it is 2D. While this increases the number of angle parameters that must be estimated to $4d_H d_G$, the beamforming criterion $\|\mathfrak{y}^H \mathbf{Z}\mathbf{a}(\boldsymbol{\omega}')\|_2$ can be applied since the columns of $\mathbf{A}(\boldsymbol{\omega}')$ are identically parameterized. This would require a search for $d_H d_G$ local maxima in a 4-dimensional space. In this case, $\hat{\boldsymbol{\gamma}}' = \mathbf{A}^\dagger(\hat{\boldsymbol{\omega}}')\mathfrak{y}/\sqrt{P}$.

CSI estimation for the dictionary-based model in (61b) also requires an unconventional approach, due to the sparsity pattern in $\mathbf{q}_{GH}$ [12], [18], [20]. The first issue is again due to the ambiguity in specifying the spatial frequencies of the RIS AoAs and AoDs; any circular shift in the columns of $\mathbf{A}_{RD}$ or the rows of $\mathbf{Q}_{RG}$ and $\mathbf{Q}_{RH}$ will leave $\mathbf{A}_D$ unchanged in (60b). As before, this ambiguity can be rectified by forcing the first column of $\mathbf{Q}_{RH}$ to have its non-zero element in the position corresponding to $\boldsymbol{\omega}_{RH} = [0 \ 0]^T$. The second issue is most easily understood in via Eq. (56b). The matrix $\mathbf{Q}_1 = \left(\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G^*\mathbf{Q}_{RG}^T\right)$ has d_G non-zero elements at the row/column coordinates corresponding to the AoA/AoD pairs of $\mathbf{G}$, while $\mathbf{Q}_2 = \left(\mathbf{Q}_{BH}\boldsymbol{\Gamma}_H\mathbf{Q}_{RH}^T\right)$ has d_H non-zero elements at positions corresponding to the AoA/AoD pairs of $\mathbf{H}$. The Kronecker product of $\mathbf{Q}_1$ with $\mathbf{Q}_2$ then repeats the sparse structure of $\mathbf{Q}_2$ at the block rows/columns in $\mathbf{Q}_{GH}$ corresponding to the non-zero elements of $\mathbf{Q}_2$. In addition, the non-zero values in these repeating blocks are scaled versions of one another; the d_H non-zero elements in $\mathbf{Q}_2$ correspond to $\boldsymbol{\gamma}_H$, and each time they are repeated in $\mathbf{Q}_{GH}$ they are multiplied by a different element of $\boldsymbol{\gamma}_G$. This structure in $\mathbf{Q}_{GH}$ in turn creates a corresponding repeated pattern in $\mathbf{q}_{GH}$. Thus, $\mathbf{q}_{GH}$ is $d_H d_G$-sparse, but it has only $d_H + d_G - 1$ unique complex entries. While cumbersome, imposing the required structured sparsity constraint on $\mathbf{q}_{GH}$ is possible [12], [20].

A bigger issue with the CS approach in the general case is the size of the dictionary in (61b), which has $N_{BD} N_{UD} N_{RD}^2$ elements. If we assume 100 grid points for each of the spatial frequency dimensions, this results in 10^{12} elements! A smaller dimensional problem can be formulated based on the parameterization in (68), such that

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{A}_D'\mathbf{q}' + \mathfrak{n} \ , \tag{69}$$

where $\mathbf{q}'$ is $d_H d_G$-sparse and unstructured, and the dictionary $\mathbf{Z}\mathbf{A}'_D$ is composed of $N_{RD}N_{UD}N_{BD}$ entries defined by

$$\mathbf{Z}\mathbf{A}'_D = \mathbf{Z}\left(\mathbf{A}_{RD} \otimes \mathbf{A}^*_{UD} \otimes \mathbf{A}_{BD}\right) . \tag{70}$$

While the dictionary is now a factor of N_{RD} smaller, the resulting problem is likely still intractable. Assuming LoS propagation only serves to reduce the sparsity level, without reducing the dimension of the dictionary. Consequently, a more tractable approach is needed, as described next.

2) A Simpler Decoupled Approach

In this approach, the channel estimation is decoupled into two stages; in the first, the BS and UE components of the channel are determined from an initial set of pilot data and then removed from the composite channel, and in the second additional pilot data are used to estimate the remainder of the channel [19], [21]. In particular, stage 1 assumes that the UE transmits a $K \times T_1$ matrix of orthogonal ($T_1 \geq K$) pilot data $\mathbf{X}_1$ while the RIS holds a fixed reflection pattern ϕ, which results in the following received signal at the BS:

$$\mathbf{Y}_1 = \sqrt{P}\mathbf{H}\bar{\boldsymbol{\Phi}}\mathbf{G}^H\mathbf{X}_1 + \mathbf{N}_1 \tag{71a}$$

$$= \sqrt{P}\mathbf{A}_B(\boldsymbol{\omega}_{BH})\boldsymbol{\Xi}\mathbf{A}^H_U(\boldsymbol{\omega}_{UG})\mathbf{X}_1 + \mathbf{N}_1 \tag{71b}$$

$$= \sqrt{P}\mathbf{A}_{BD}\mathbf{Q}_{BH}\boldsymbol{\Xi}\mathbf{Q}^T_{UG}\mathbf{A}^H_{UD}\mathbf{X}_1 + \mathbf{N}_1 , \tag{71c}$$

where

$$\boldsymbol{\Xi} = \boldsymbol{\Gamma}_H\mathbf{A}^H_R(\boldsymbol{\omega}_{RH})\boldsymbol{\Phi}\mathbf{A}_R(\boldsymbol{\omega}_{RG})\boldsymbol{\Gamma}^*_G . \tag{72}$$

We see immediately that, assuming $M > d_H$, (71b)-(71c) are in the form of a standard AoA estimation problem as in (41), and thus $\boldsymbol{\omega}_{BH}$ could be estimated from $\mathbf{Y}_1$ using any AoA estimation or compressive sensing algorithm[5]. Assuming $K \geq d_G$, estimation of $\boldsymbol{\omega}_{UG}$ can also be performed separately based on the following equation:

$$\frac{\mathbf{X}_1\mathbf{Y}^H_1}{T_1\sqrt{P}} = \mathbf{A}_U(\boldsymbol{\omega}_{UG})\boldsymbol{\Xi}^H\mathbf{A}^H_B(\boldsymbol{\omega}_{BH}) + \frac{\mathbf{X}_1\mathbf{N}^H_1}{T_1\sqrt{P}} \tag{73a}$$

$$= \mathbf{A}_U(\boldsymbol{\omega}_{UG})\mathbf{S} + \frac{1}{T_1\sqrt{P}}\mathbf{X}_1\mathbf{N}^H_1 . \tag{73b}$$

For the second stage, the UE transmits T_2 pilots $\mathbf{x}_t, t = 1, \cdots, T_2$, and the RIS reflection pattern also changes at the symbol rate: $\boldsymbol{\Psi}^H = \begin{bmatrix}\boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_{T_2}\end{bmatrix}$. The training received at the BS data at time t can be represented as follows:

$$\mathbf{Y}_2 = \begin{bmatrix}\mathbf{y}_1 & \cdots & \mathbf{y}_{T_2}\end{bmatrix} \tag{74a}$$

$$\mathbf{y}_t = \mathbf{B}_t\boldsymbol{\Gamma}_{GH}\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})\boldsymbol{\phi}_t + \mathbf{n}_t , \tag{74b}$$

where

$$\mathbf{B}_t = \sqrt{P}\left(\mathbf{x}^T_t \otimes \mathbf{I}_M\right)\left(\mathbf{A}^*_U(\boldsymbol{\omega}_{UG}) \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH})\right) , \tag{75}$$

[5]Note that in some work it is assumed that the BS-RIS channel changes slowly since the BS and RIS are in fixed locations. In such cases $\boldsymbol{\omega}_{RH}$ can be estimated infrequently and thus may already be known, and hence the above estimation step may not always be necessary [3], [6], [14], [44].

and $\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$ is defined in (63). With sufficiently accurate estimates of $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}$ from stage 1, one can eliminate $\mathbf{B}_t$ from (74) and form the matrix

$$\mathbf{Y}_B = \begin{bmatrix}\mathbf{B}^\dagger_1\mathbf{y}_1 & \cdots & \mathbf{B}^\dagger_{T_2}\mathbf{y}_{T_2}\end{bmatrix}^T \tag{76a}$$

$$\simeq \boldsymbol{\Psi}^*\left(\mathbf{A}^T_R(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}^H_R(\boldsymbol{\omega}_{RH})\right)^T\boldsymbol{\Gamma}_{GH} + \mathbf{N}_B \tag{76b}$$

$$\simeq \boldsymbol{\Psi}^*\left(\mathbf{A}^T_{RD} \diamond \mathbf{A}^H_{RD}\right)^T(\mathbf{Q}_{RG} \otimes \mathbf{Q}_{RH})\boldsymbol{\Gamma}_{GH} + \mathbf{N}_B , \tag{76c}$$

where $\mathbf{N}_B$ is defined similarly to $\mathbf{Y}_B$.

A sparse estimation problem could be set up for $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\omega}_{RH}$ based on vectorizing (76c), and while the resulting dictionary is significantly smaller than in (61b), it still has N^2_{RD} elements. Instead, a much simpler solution can be found by noting the structure of $\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$ described in (64), which indicates that each of the $d_G d_H$ columns of (76b) depends only on a pair of 2D angles, one from $\boldsymbol{\omega}_{RG}$ and one from $\boldsymbol{\omega}_{RH}$. In particular, for the k-th column,

$$\mathbf{y}_{B,k} \simeq \gamma_k\boldsymbol{\Psi}^*\mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) + \mathbf{n}_{B,k} , \tag{77}$$

where γ_k is the k-th element of $\boldsymbol{\gamma}^*_G \otimes \boldsymbol{\gamma}_H$ and $k = (\ell - 1)d_H + p$ for $\ell = 1, \cdots, d_H$ and $p = 1, \cdots, d_G$. Each of the $d_H d_G$ columns of $\mathbf{Y}_B$ is thus approximately equivalent to a single snapshot from a T_2-element "array" with a single 2D spatial frequency. As before, the gains and 2D frequencies of the columns are interrelated, but if one ignores this fact, the unknown part of the channel in (74) can be reconstructed in a suboptimal way by solving a series of $d_H d_G$ one-dimensional AoA estimation problems on each of the columns of $\mathbf{Y}_B$. In particular, if $\mathbf{c}^T_k$ represents the k-th row of $\boldsymbol{\Gamma}_{GH}\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$, then $\mathbf{c}^T_k$ is estimated using the estimates of γ_k and $\boldsymbol{\omega}_k = \boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}$ obtained for the k-th column of $\mathbf{Y}_B$:

$$\hat{\mathbf{c}}_k = \hat{\gamma}_k\mathbf{a}_R(\hat{\boldsymbol{\omega}}_k) . \tag{78}$$

For LoS channels with $d_H = d_G = 1$, only a single sparse estimation problem needs to be solved.

D. Wideband Single User MIMO

As in Section III-B, for geometric channel models with OFDM signals, either time- or frequency domain models can be employed. We will focus on the time-domain approach here as it is a bit more straightforward. Our starting point is Eq. (28), which differs from (61) in that the wideband channel parameter vector $\mathfrak{h}_c$ is composed of L single-tap terms like $\mathbf{h}_c$ stacked together. Partition the columns of $\mathfrak{Z}$ into L blocks, each block corresponding to one of the taps $\mathbf{h}_c(k)$, $k = 1, \cdots, L$, in $\mathfrak{h}_c$:

$$\mathfrak{Z} = \begin{bmatrix}\mathfrak{Z}(0) & \mathfrak{Z}(1) & \cdots & \mathfrak{Z}(L-1)\end{bmatrix} . \tag{79}$$

Using (59), we can write

$$\mathfrak{y} = \sqrt{P}\mathfrak{A}(\boldsymbol{\Omega})\begin{bmatrix}\boldsymbol{\gamma}_{GH}(0) \\ \vdots \\ \boldsymbol{\gamma}_{GH}(L-1)\end{bmatrix} + \mathfrak{n} \tag{80a}$$

$$= \sqrt{P}\mathfrak{A}_D\begin{bmatrix}\mathbf{q}_{GH}(0) \\ \vdots \\ \mathbf{q}_{GH}(L-1)\end{bmatrix} + \mathfrak{n} , \tag{80b}$$

where

$$\mathfrak{A}(\mathbf{\Omega}) = \begin{bmatrix} \mathfrak{Z}(0)\mathbf{A}(\boldsymbol{\omega}(0)) & \cdots & \mathfrak{Z}(L-1)\mathbf{A}(\boldsymbol{\omega}(L-1)) \end{bmatrix} \tag{81a}$$

$$\mathfrak{A}_D = \begin{bmatrix} \mathfrak{Z}(0)\mathbf{A}_D & \cdots & \mathfrak{Z}(L-1)\mathbf{A}_D \end{bmatrix}, \tag{81b}$$

$\mathbf{\Omega} = \begin{bmatrix} \boldsymbol{\omega}(0)^T & \cdots & \boldsymbol{\omega}^T(L-1) \end{bmatrix}^T$, and where $\boldsymbol{\omega}(k), \boldsymbol{\gamma}_{GH}(k)$ and $\mathbf{q}_{GH}(k)$ are the spatial frequencies, path gains and sparse vectors associated with the channel for tap k. Note that we have assumed the general case where the AoAs/AoDs are potentially different for each tap. Consequently, while we see from (80) that the full-scale parameterization of the problem is similar to that in (61), the dictionary size and parameter dimensions are all a factor of L larger.

The decoupled approach described in Section IV-C2 can be exploited to significantly reduce the required complexity. As before, we ignore the direct channel and assume an initial training period of $T_1 = T_{o1}N_c$ samples from T_{o1} OFDM symbols where the RIS reflection state is fixed at ϕ. Using (26), after removal of the cyclic prefix we can collect all N_c samples from OFDM symbol t in the matrix $\mathbf{Y}_{1,t}$ as follows:

$$\mathbf{Y}_{1,t} = \sqrt{P} \begin{bmatrix} \mathbf{H}(0)\mathbf{\Phi}\mathbf{G}^H(0) & \cdots & \mathbf{H}(L-1)\mathbf{\Phi}\mathbf{G}^H(L-1) \end{bmatrix}$$
$$\times \mathbf{X}_{1,t} + \mathbf{N}_{1,t} \tag{82a}$$
$$= \sqrt{P}\mathbf{A}_B(\mathbf{\Omega}_{BH})\mathbf{\Pi}\mathbf{X}_{1,t} + \mathbf{N}_{1,t} \tag{82b}$$

where $\mathbf{X}_{1,t}$ is block circulant with first block row $[\mathbf{x}_{t,1} \cdots \mathbf{x}_{t,N_c}]$, $\mathbf{\Omega}_{BH} = \begin{bmatrix} \boldsymbol{\omega}_{BH}(0)^T & \cdots & \boldsymbol{\omega}_{BH}^T(L-1) \end{bmatrix}^T$, and

$$\mathbf{A}_B(\mathbf{\Omega}_{BH}) = \begin{bmatrix} \mathbf{A}_B(\boldsymbol{\omega}_{BH}(0)) & \cdots & \mathbf{A}_B(\boldsymbol{\omega}_{BH}(L-1)) \end{bmatrix} \tag{83a}$$

$$\mathbf{\Pi} = \mathrm{blkdiag}\left(\left\{ \mathbf{\Xi}(0)\mathbf{A}_U^H(\boldsymbol{\omega}_{UG}(0)) \right\}_{k=0}^{L-1} \right), \tag{83b}$$

where $\mathbf{\Xi}(k)$ is the matrix corresponding to (72) for tap k. Concatenating data from the T_{o1} OFDM symbols yields

$$\mathbf{Y}_1 = \sqrt{P} \begin{bmatrix} \mathbf{Y}_{1,1} & \cdots & \mathbf{Y}_{1,T_{o1}} \end{bmatrix} \tag{84a}$$
$$= \sqrt{P} \underbrace{\mathbf{A}_B(\mathbf{\Omega}_{BH})}_{M \times Ld_H} \underbrace{\mathbf{\Pi}\mathbf{X}_1}_{Ld_H \times T_1} + \mathbf{N}_1, \tag{84b}$$

where $\mathbf{X}_1$ and $\mathbf{N}_1$ are defined like $\mathbf{Y}_1$. Assuming $M > Ld_H$, the BS AoAs for each tap of the impulse response can be estimated using standard approaches. The AoDs at the UE can be found by noting that

$$\frac{(\mathbf{Y}_1\mathbf{X}_1^\dagger)^H}{\sqrt{P}T_1} = \begin{bmatrix} \mathbf{A}_U(\boldsymbol{\omega}_{UG}(0))\mathbf{\Xi}(0)\mathbf{A}_B^H(\boldsymbol{\omega}_{BH}(0)) \\ \vdots \\ \mathbf{A}_U(\boldsymbol{\omega}_{UG}(L-1))\mathbf{\Xi}(0)\mathbf{A}_B^H(\boldsymbol{\omega}_{BH}(L-1)) \end{bmatrix}$$
$$+ \frac{(\mathbf{N}_1\mathbf{X}_1^\dagger)^H}{\sqrt{P}T_1}. \tag{85}$$

Assuming $K > d_G$, the UE angles are found by solving AoA estimation problems on the L blocks in (85).

In Stage 2, T_{o2} additional OFDM training symbols are transmitted, for a total of $T_2 = T_{o2}N_c$ samples. As in (74)-(75),

$$\mathbf{y}_{t,s} = \sqrt{P} \sum_{k=0}^{L-1} \mathbf{B}_{t,s-k}\mathbf{\Gamma}_{GH}(k)\mathbf{C}(k)\phi_{t,s-k} + \mathbf{n}_{t,s} \tag{86a}$$

$$= \begin{bmatrix} \mathbf{B}_{t,s} & \cdots & \mathbf{B}_{t,s-L+1} \end{bmatrix}$$
$$\times \begin{bmatrix} \mathbf{\Gamma}_{GH}(0)\mathbf{C}(0)\phi_{t,s} \\ \vdots \\ \mathbf{\Gamma}_{GH}(L-1)\mathbf{C}(L-1)\phi_{t,s-L+1} \end{bmatrix} + \mathbf{n}_{t,s}$$

$$\mathbf{B}_{t,s-k} = \sqrt{P}\left(\mathbf{x}_{t,s-k}^T \otimes \mathbf{I}_M \right)$$
$$\times \left(\mathbf{A}_U^*(\boldsymbol{\omega}_{UG}(k)) \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH}(k)) \right), \tag{86b}$$

where $\mathbf{C}(k) = \mathbf{C}(\boldsymbol{\omega}_{RG}(k), \boldsymbol{\omega}_{RH}(k))$. Replacing $\boldsymbol{\omega}_{UG}(k)$ and $\boldsymbol{\omega}_{BH}(k)$ with their estimates from Stage 1, and assuming $M \geq Ld_H d_G$, we multiply each $\mathbf{y}_{t,s}$ on the left by the estimate of the pseudo-inverse of $\begin{bmatrix} \mathbf{B}_{t,s} & \cdots & \mathbf{B}_{t,s-L+1} \end{bmatrix}$, stack them together and transpose as in (76):

$$\underbrace{\mathbf{Y}_{B,t}}_{N_c \times Ld_H d_G} = \begin{bmatrix} \underbrace{\mathbf{Y}_{B,t,0}}_{N_c \times d_H d_G} & \cdots & \mathbf{Y}_{B,t,L-1} \end{bmatrix} \tag{87a}$$

$$\mathbf{Y}_{B,t,k} \simeq \underbrace{\mathbf{\Psi}_{t,k}^*}_{N_c \times N} \underbrace{\mathbf{C}^T(k)\mathbf{\Gamma}_{GH}(k)}_{N \times d_H d_G} + \mathbf{N}_{B,t,k} \tag{87b}$$

$$\mathbf{\Psi}_t^* = \begin{bmatrix} \mathbf{\Psi}_{t,0}^* & \cdots & \mathbf{\Psi}_{t,L-1}^* \end{bmatrix}, \tag{87c}$$

where $\mathbf{\Psi}_t^*$ is block circulant with first set of rows defined by $\mathbf{\Psi}_{t,0}^T = \begin{bmatrix} \phi_{t,1} & \cdots & \phi_{t,N_c} \end{bmatrix}$. Stacking the result from all T_{o2} training symbols $\mathbf{Y}_B^T = \begin{bmatrix} \mathbf{Y}_{B,1}^T & \cdots & \mathbf{Y}_{B,T_{o2}}^T \end{bmatrix}$ and partitioning them into L blocks of $d_H d_G$ columns each, we have

$$\bar{\mathbf{Y}}_{B,k} \simeq \begin{bmatrix} \mathbf{\Psi}_{1,k}^* \\ \vdots \\ \mathbf{\Psi}_{N_c,k}^* \end{bmatrix} \mathbf{C}^T(k)\mathbf{\Gamma}_{GH}(k) + \mathbf{N}_{B,k}, \tag{88}$$

where $\bar{\mathbf{Y}}_{B,k}$ holds columns $k+1$ through $k+L$ of $\mathbf{Y}_B$. This equation is equivalent in form to (76b), and thus the methods discussed previously can be used to solve for the remaining channel parameters for path k. The process is then repeated for all L paths, $k = 0, \cdots, L-1$.

E. Single Antenna Scenarios
1) Single Antenna UE

When the UE has only a single antenna, $\boldsymbol{\omega}_{UG} = \varnothing$ is the empty set and $\mathbf{A}_U(\boldsymbol{\omega}_{UG}) = \mathbf{1}_{d_G}^T$. The matrix $\mathbf{A}(\boldsymbol{\omega})$ in (61a) still has $d_H d_G$ columns, now given by

$$[\mathbf{A}(\boldsymbol{\omega})]_{:,k} = \mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) \otimes \mathbf{a}_B(\omega_{BH,p}), \tag{89}$$

where $\ell = \lfloor k/d_H \rfloor$ and $p = \mathrm{mod}_{d_H}(k)$. We assume without loss of generality that $x_t = 1, \forall\, t$, so (54) simplifies to

$$\mathbf{Z} = \begin{bmatrix} \phi_1^T \otimes \mathbf{I}_M \\ \vdots \\ \phi_T^T \otimes \mathbf{I}_M \end{bmatrix}. \tag{90}$$

Ignoring the structure of $\boldsymbol{\gamma}_{GH}$ in (61a), the DML criterion in (66) can be applied to estimate the $3d_h + 2d_G - 2$ spatial frequencies in $\boldsymbol{\omega}$, which represents only a slight savings compared with the multi-antenna UE case. The simpler beamforming criterion can be used if one ignores the relationship between the columns of $\mathbf{A}(\boldsymbol{\omega})$ as in (68), treating them as independent vectors that are a function of three frequency variables, one for the BS and two for the RIS. This results in a search for $d_H d_G$ local maxima in a 3D space. In the LoS case, the problem is solved by optimizing a function of three frequency variables in either the DML or beamforming approach.

The CS-based model in (61b) for single-antenna UEs can also be approached in two ways. The first retains the full geometric structure of the channel, with

$$\mathbf{A}_D = \left(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H \right)^T \otimes \mathbf{A}_{BD} \tag{91a}$$

$$\mathbf{Q}_{GH} = \underbrace{\left(\boldsymbol{\gamma}_G^H \mathbf{Q}_{RG}^T \right)}_{1 \times N_{RD}} \otimes \underbrace{\left(\mathbf{Q}_{BH} \boldsymbol{\Gamma}_H \mathbf{Q}_{RH}^T \right)}_{N_{BD} \times N_{RD}} . \tag{91b}$$

In this case, the $d_H d_G$-sparse vector $\mathbf{q}_{GH} = \text{vec}(\mathbf{Q}_{GH})$ has a similar structure as before, and the dictionary has $N_{BD} N_{RD}^2$ elements. The second approach ignores the sparse structure as in (69), except in this case the dictionary $\mathbf{Z}\mathbf{A}_D' = \mathbf{Z}(\mathbf{A}_{RD} \otimes \mathbf{A}_{BD})$ has only $N_{RD} N_{BD}$ terms.

Further complexity reduction is possible using the decoupled approach described in Section IV-C2. In stage 1, $\boldsymbol{\omega}_{BH}$ is estimated from $\mathbf{Y}_1$ as before, but estimation of $\boldsymbol{\omega}_{UG}$ is not required. Stage 2 proceeds as before, but the solution is obtained in a different way. In particular, in this case we define

$$\mathbf{y}_t = \sqrt{P} \mathbf{A}_B(\boldsymbol{\omega}_{BH}) \left(\boldsymbol{\gamma}_G^H \otimes \boldsymbol{\Gamma}_H \right) \mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH}) \boldsymbol{\phi}_t + \mathbf{n}_t$$

$$\mathbf{Y}_B = \frac{1}{\sqrt{P}} \left[\mathbf{A}_B^\dagger(\hat{\boldsymbol{\omega}}_{BH}) \mathbf{Y}_2 \right]^T \tag{92a}$$

$$\simeq \boldsymbol{\Psi}^* \mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH}) \left(\boldsymbol{\gamma}_G^* \otimes \boldsymbol{\Gamma}_H \right) + \mathbf{N}_B , \tag{92b}$$

and note that the d_H columns of $\mathbf{Y}_B$ are now linear combinations of RIS array response vectors:

$$\mathbf{y}_{B,k} \simeq \gamma_{H,k} \boldsymbol{\Psi}^* \sum_{n=1}^{d_G} \gamma_{G,n}^* \mathbf{a}_R(\boldsymbol{\omega}_{RG,n} - \boldsymbol{\omega}_{RH,k}) + \mathbf{n}_{B,k} \tag{93a}$$

$$\simeq \gamma_{H,K} \boldsymbol{\Psi}^* \text{diag}\left(\mathbf{A}_R(\boldsymbol{\omega}_{RG}) \boldsymbol{\gamma}_G^* \right) \mathbf{a}_R(-\boldsymbol{\omega}_{RH,k}) + \mathbf{n}_{B,k}, \tag{93b}$$

This special structure allows for a simpler solution than that required in the multi-antenna UE case considered in Section IV-C2 [18]. To see this, note that because of our identifiability conditions $\gamma_{H,1} = 1$ and $\boldsymbol{\omega}_{RH,1} = \mathbf{0}$, the first column of $\mathbf{Y}_B$ is given by

$$\mathbf{y}_{B,1} \simeq \boldsymbol{\Psi}^* \mathbf{A}_R(\boldsymbol{\omega}_{RG}) \boldsymbol{\gamma}_G^* + \mathbf{n}_{B,1} . \tag{94}$$

It is clear that the parameters $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\gamma}_G$ can be estimated from $\mathbf{y}_{B,1}$ using an AoA estimation such as beamforming or DML, or using a d_G-sparse CS algorithm. Once estimated, these parameters can be substituted into (93), and $\mathbf{y}_{B,k}$ can be used to estimate $\gamma_{H,k}$ and $\boldsymbol{\omega}_{RH,k}$ for $k = 2, \cdots, d_H$. Thus, instead of solving $d_H d_G$ 1-sparse AoA estimation problems that

ignore the underlying structure of the data as in Section IV-C2, for single-antenna UEs we can estimate the channel with one d_G-sparse estimation, followed by $d_H - 1$ 1-sparse problems, that when combined provide estimates of $\boldsymbol{\omega}_{RH}, \boldsymbol{\omega}_{RG}, \boldsymbol{\gamma}_H, \boldsymbol{\gamma}_G$.

2) Single Antenna UE and BS

Here there is no need for a first stage, as there are no angles to estimate at the BS or UE. Instead, we simply collect T observations at the BS and stack them together in a $T \times 1$ vector $\mathbf{y} = [y_1 \ \cdots \ y_T]^T$, which yields

$$\mathbf{y} = \boldsymbol{\Psi}^* \mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})(\boldsymbol{\gamma}_G^* \otimes \boldsymbol{\gamma}_H) + \mathbf{n} \tag{95a}$$

$$= \boldsymbol{\Psi}^* \sum_{k=1}^{d_G} \sum_{n=1}^{d_H} \gamma_{G,k}^* \gamma_{H,n} \mathbf{a}_R \left(\boldsymbol{\omega}_{RG,k} - \boldsymbol{\omega}_{RH,n} \right) + \mathbf{n} . \tag{95b}$$

We have an equivalent single-snapshot AoA estimation problem, although with the nonlinear Kronecker structure for the gains. Ignoring this as before, the spatial frequencies can be determined using a $2d_H + 2d_G - 2$ dimensional DML search, or a single $d_H d_G$-sparse CS problem with or without application of the special row structure induced by the model.

F. Multi-User Scenarios

Multiple users in the geometric model can be accounted for by redefining $\mathbf{G}$ as follows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_U(\boldsymbol{\omega}_{UG}^1) \boldsymbol{\Gamma}_{G^1} \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^1) \\ \vdots \\ \mathbf{A}_U(\boldsymbol{\omega}_{UG}^U) \boldsymbol{\Gamma}_{G^U} \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^U) \end{bmatrix} \equiv \begin{bmatrix} \mathbf{G}^1 \\ \vdots \\ \mathbf{G}^U \end{bmatrix} , \tag{96}$$

where the superscripts indicate the user index. UE u is assumed to have K_u antennas and the corresponding channel has d_{G^u} propagation paths, so that $d_G = \sum_u d_{G^u}$. The composite channel and its vectorized form is given by

$$\mathbf{H}_c = \begin{bmatrix} \mathbf{G}^1 \diamond \mathbf{H} \\ \vdots \\ \mathbf{G}^U \diamond \mathbf{H} \end{bmatrix} \tag{97a}$$

$$\mathbf{h}_c = \begin{bmatrix} \mathbf{h}_c^1 \\ \vdots \\ \mathbf{h}_c^U \end{bmatrix} = \mathbf{J} \text{vec}(\mathbf{H}_c) , \tag{97b}$$

where the permutation matrix $\mathbf{J}$ is implicitly defined to group the blocks of $\mathbf{h}_c$ by user rather than by RIS elements. Compact expressions for the multi-user case can be found for the general

data model that parallel those in (61):

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{J}\begin{bmatrix} \mathbf{A}(\boldsymbol{\omega}^1)\boldsymbol{\gamma}_{G^1H} \\ \vdots \\ \mathbf{A}(\boldsymbol{\omega}^U)\boldsymbol{\gamma}_{G^UH} \end{bmatrix} + \mathfrak{n} \tag{98a}$$

$$= \sqrt{P}\,\mathbf{Z}\underbrace{[\mathbf{J}_1\mathbf{A}(\boldsymbol{\omega}^1) \;\cdots\; \mathbf{J}_U\mathbf{A}(\boldsymbol{\omega}^U)]}_{\mathbf{A}(\boldsymbol{\omega})}\underbrace{\begin{bmatrix} \boldsymbol{\gamma}_{G^1H} \\ \vdots \\ \boldsymbol{\gamma}_{G^UH} \end{bmatrix}}_{\boldsymbol{\gamma}_{GH}} + \mathfrak{n} \tag{98b}$$

$$= \sqrt{P}\,\mathbf{Z}\underbrace{[\mathbf{J}_1\mathbf{A}_D \;\cdots\; \mathbf{J}_U\mathbf{A}_D]}_{\text{effective dictionary}}\underbrace{\begin{bmatrix} \mathbf{q}_{G^1H} \\ \vdots \\ \mathbf{q}_{G^UH} \end{bmatrix}}_{\mathbf{q}_{GH}} + \mathfrak{n}, \tag{98c}$$

where $\{\boldsymbol{\omega}^u, \boldsymbol{\gamma}_{G^uH}, \mathbf{q}_{G^uH}\}$ are the angle parameters, channel gains, and sparse vectors for UE u as in (59). The block column $\mathbf{J}_u$ of the permutation matrix $\mathbf{J} = [\mathbf{J}_1 \;\cdots\; \mathbf{J}_U]$ is of dimension $MKN \times MK_uN$. General methods can be developed for simultaneous estimation of all the channel parameters based on these equations and the approaches discussed in Section IV-C, but as before the dimension of the resulting optimization problems is likely prohibitive except for certain simple cases.

The decoupled approach of Section IV-C2 can be applied to reduce the estimation complexity. Provided that $M \geq d_H d_G$, the BS AoAs $\boldsymbol{\omega}_{BH}$ are estimated as before, while the equation for estimating the UE AoDs is slightly different than (73):

$$\frac{\mathbf{X}_1\mathbf{Y}_1^H}{T_1\sqrt{P}} = \begin{bmatrix} \mathbf{A}_U(\boldsymbol{\omega}_{UG}^1) & & \\ & \ddots & \\ & & \mathbf{A}_U(\boldsymbol{\omega}_{UG}^U) \end{bmatrix}\mathbf{S} + \frac{\mathbf{X}_1\mathbf{N}_1^H}{T_1\sqrt{P}} \tag{99}$$

where $\mathbf{S} = \boldsymbol{\Xi}^H\mathbf{A}_B^H(\boldsymbol{\omega}_{BH})$ and

$$\boldsymbol{\Xi}^H = \boldsymbol{\Gamma}_G\begin{bmatrix} \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^1) \\ \vdots \\ \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^U) \end{bmatrix}\boldsymbol{\Phi}^*\mathbf{A}_R(\boldsymbol{\omega}_{RH})\boldsymbol{\Gamma}_H^* . \tag{100}$$

Assuming $K_u > d_{G^u}$, UE AoD estimates can be found separately for each user by processing different block rows of (99). Once estimates of $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}^u$ are determined, we collect additional training data and proceed as in (76) where

$$\mathbf{B}_t = \sqrt{P}\left(\mathbf{x}_t^T \otimes \mathbf{I}_M\right)$$
$$\times \left(\begin{bmatrix} \mathbf{A}_U(\boldsymbol{\omega}_{UG}^1) & & \\ & \ddots & \\ & & \mathbf{A}_U(\boldsymbol{\omega}_{UG}^U) \end{bmatrix} \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH})\right). \tag{101}$$

This leads to

$$\mathbf{Y}_B \simeq \boldsymbol{\Psi}^*\underbrace{\begin{bmatrix} \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}^1) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \\ \vdots \\ \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}^U) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \end{bmatrix}^T}_{\mathbf{C}^T(\boldsymbol{\omega}_{RG}^1,\cdots,\boldsymbol{\omega}_{RG}^U,\boldsymbol{\omega}_{RH})}\boldsymbol{\Gamma}_{GH} + \mathbf{N}_B . \tag{102}$$

As before, each column of (102) involves only the difference between only one spatial frequency from $\{\boldsymbol{\omega}_{RG}^1,\cdots,\boldsymbol{\omega}_{RG}^U\}$ and one from $\boldsymbol{\omega}_{RH}$. There are $d_H d_G$ such combinations, and thus the remainder of the channel parameters can be found by solving $d_H d_G$ single spatial frequency estimation problems. Note also that when the UEs have only a single antenna, the simplification discussed in Section IV-E1 holds, where only one d_G-dimensional AoA estimation followed by $d_H - 1$ one-dimensional AoA estimation problems are necessary.

G. Reducing the Complexity and Training Overhead

The methods discussed in Section III-E can be used to further reduce the training required even for geometric channel models. The availability of prior knowledge of low-rank channel covariance matrices is useful for geometric models, although less for reducing the amount of training than for designing the pilot symbols $\mathbf{x}_t$ and RIS training $\boldsymbol{\phi}_t$ to improve the received SNR. Also, the common channel $\mathbf{H}$ associated with multiple UE antennas can be exploited as before to reduce the algorithm complexity. For geometric channel models, this approach could be implemented as follows [18]:

1) Choose one antenna from one of the UEs, and transmit training data while the other antennas are silent to estimate the channel $\mathbf{H}\text{diag}(\mathbf{g}_1)$.
2) Transmit training data from the remaining UE antennas, and estimate the UE AoDs $\boldsymbol{\omega}_{UG}^k$ as in (73) if there is only one user with multiple antennas, or as in (99) if there are multiple multi-antennas users.
3) Set the RIS reflection vector to a fixed value $\bar{\boldsymbol{\phi}}$ and send at least d_G additional training symbols $\mathbf{x}_t$ to obtain

$$\mathbf{Y} = \mathbf{H}\bar{\boldsymbol{\Phi}}\mathbf{A}_R(\boldsymbol{\omega}_{RG})\boldsymbol{\Gamma}_G\mathbf{A}_U^H(\boldsymbol{\omega}_{UG})\mathbf{X} + \mathbf{N} . \tag{103}$$

Then, multiply on the right by $\left(\mathbf{A}_U^H(\hat{\boldsymbol{\omega}}_{UG})\mathbf{X}\right)^\dagger$. Since we have an estimate of $\mathbf{H}$, the resulting matrix is approximately in the standard form for AoA estimation, using either steering vectors drawn from $\hat{\mathbf{H}}\bar{\boldsymbol{\Phi}}\mathbf{A}_R(\boldsymbol{\omega}_{RG})$, or an overcomplete dictionary $\hat{\mathbf{H}}\bar{\boldsymbol{\Phi}}\mathbf{A}_{RD}$.

Use of the geometric model offers a further opportunity for dramatic reductions in both computational complexity and training overhead. The AoAs and AoDs in the geometric model change very slowly, and can be considered to be stationary over multiple coherence blocks. Only the complex gains $\boldsymbol{\gamma}_G$ and $\boldsymbol{\gamma}_H$ change substantially from block to block. This suggests that, once estimated, the spatial frequencies for subsequent blocks can be considered constant, and only the gains need to be re-estimated [18]. For CS-based methods, this means that the support of the sparse solution is already known. Consequently, with at least $d_H d_G$ training samples, the channel gains in equations (61), (80), or (98) can be estimated using least-squares; *e.g.,* for (61a) we have

$$\hat{\boldsymbol{\gamma}}_{GH} = \frac{1}{\sqrt{P}}\left(\mathbf{Z}\mathbf{A}(\hat{\boldsymbol{\omega}})\right)^\dagger\mathfrak{y} . \tag{104}$$

V. NUMERICAL EXAMPLES

To illustrate the CSI estimation performance for typical RIS-based scenarios, we will use the CRB rather than plotting the results of individual algorithms (for which there are too

many examples to fairly consider). In particular, we will show numerical results for the CRB of $\check{\mathbf{h}}_c^T = \begin{bmatrix} \text{Re}(\mathbf{h}_c)^T & \text{Im}(\mathbf{h}_c)^T \end{bmatrix}$, where $\text{Re}(\mathbf{h}_c)$ and $\text{Im}(\mathbf{h}_c)$ respectively represent the real and imaginary parts of $\mathbf{h}_c$.

In the unstructured case, the parameters are the elements of the channel itself: $\boldsymbol{\eta} = \check{\mathbf{h}}_c$. Assuming temporally and spatially white Gaussian noise with variance σ^2 as in (1) and (5), the log-likelihood is given by

$$f_u(\check{\mathbf{h}}_c) = -MT\ln(\pi\sigma^2) - \frac{1}{\sigma^2}\left\| \mathfrak{y} - \sqrt{P}\mathbf{Z}\mathbf{h}_c \right\|^2 , \quad (105)$$

where the subscript u denotes "unstructured." The CRB is defined in terms of the Fisher Information Matrix (FIM):

$$\text{CRB}_u(\check{\mathbf{h}}_c) = \text{FIM}_u^{-1}(\check{\mathbf{h}}_c) \quad (106)$$

$$\text{FIM}_u(\check{\mathbf{h}}_c) = \mathbb{E}\left\{ \frac{\partial f_u(\check{\mathbf{h}}_c)}{\partial \check{\mathbf{h}}_c} \left(\frac{\partial f_u(\check{\mathbf{h}}_c)}{\partial \check{\mathbf{h}}_c} \right)^T \right\} , \quad (107)$$

where $\mathbb{E}(\cdot)$ denotes expectation with respect to the noise distribution. It is straightforward to show that the CRB for the unstructured model is given by

$$\text{CRB}_u(\check{\mathbf{h}}_c) = \frac{\sigma^2}{2P}\left(\check{\mathbf{Z}}^T \check{\mathbf{Z}} \right)^{-1} , \quad (108)$$

where

$$\check{\mathbf{Z}} = \begin{bmatrix} \text{Re}(\mathbf{Z}) & -\text{Im}(\mathbf{Z}) \\ \text{Im}(\mathbf{Z}) & \text{Re}(\mathbf{Z}) \end{bmatrix} \quad (109)$$

and $\mathbf{Z}$ is as defined in (5). If $\mathbf{Z}$ is designed to be orthogonal as in Section III-A1, then the CRB simplifies to

$$\text{CRB}_u(\check{\mathbf{h}}_c) = \frac{\sigma^2}{2PT}\mathbf{I}_{2MK(N+1)} , \quad (110)$$

signifying that the lower bound is identical for every element of $\check{\mathbf{h}}_c$ (note that we have assumed the RIS element gains are $\beta = 1$). Note that the CRB in the unstructured case is independent of the number of BS antennas, RIS elements, and UEs (although $T \geq K(N+1)$ must hold for the model to be identifiable, and hence for the FIM to be invertible).

The CRB for the geometric channel model depends on the parameter vector $\boldsymbol{\eta}$ defined by the angles and gains listed in Table I. The CRB for $\boldsymbol{\eta}$ is more difficult to compute, requiring a large number of tedious derivative calculations which we do not include here. When one forms an estimate of the composite channel $\mathbf{h}_c$ from the estimate of the parameters in $\boldsymbol{\eta}$, then the resulting CRB for the channel is given by

$$\text{CRB}_s(\check{\mathbf{h}}_c) = \frac{\partial \check{\mathbf{h}}_c}{\partial \boldsymbol{\eta}}\text{CRB}(\boldsymbol{\eta})\left(\frac{\partial \check{\mathbf{h}}_c}{\partial \boldsymbol{\eta}} \right)^T , \quad (111)$$

where here the subscript s is for "structured." The examples below compare $\text{CRB}_u(\check{\mathbf{h}}_c)$ and $\text{CRB}_s(\check{\mathbf{h}}_c)$ for several different scenarios. The performance metric adopted in all numerical examples is the average of the diagonal elements of the CRB over a large number of different (geometric) channel realizations with random AoAs/AoDs and path gains.

The first case considers a single-UE with $K \in \{1, 2\}$ antennas, orthogonal pilot signals, and uniformly distributed RIS phases during training. The AoAs/AoDs were generated using a uniform distribution over $[-90°, 90°]$ for azimuth and

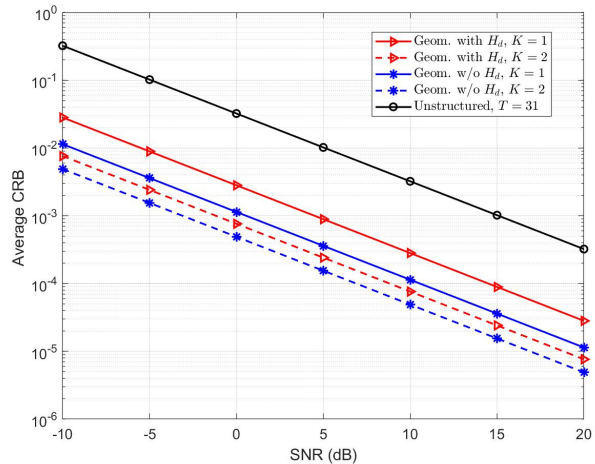Fig. 2: CRB for various channel models vs. SNR. Channel parameters are $M = 30, N = 30, T = 31, K = 2, d_F = 5, d_G = 5, d_H = 2$.

$[0°, 90°]$ for elevation. The path gains were generated as unit-variance Rayleigh random variables. Since the path gains are independent of the number of paths, a channel with more paths will generally make a larger contribution to the overall SNR. The path gain distribution is the same for all three channels $\mathbf{H}_d, \mathbf{H}, \mathbf{G}$, which corresponds to a case where the BS, RIS and UE are approximately located at the vertices of an equilateral triangle with similar propagation characteristics. Also, this assumption means that, even though during channel estimation the RIS phases $\boldsymbol{\Phi}_t$ are not chosen to maximize the coherent gain offered by the RIS, the RIS provides sufficient combining gain such that $\|\mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}\| \gg \|\mathbf{H}_d\|$. The SNR in the examples below is defined as $\mathbf{P}/\sigma^2$, but the effective SNR will be a function of the channel gain $\|\mathbf{H}_d + \mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}\|$, which will change as a function of the number of antennas, RIS elements, and path gains.

Figs. 2-4 assume a BS with $M = 30$ antennas and an RIS with $N = 30$ elements. Fig. 2 plots the CRB as a function of SNR for a case with $d_H = 2$, $d_F = d_G = 5$, and $T = 31$ training samples, which is the minimum required for the case of a single-antenna UE ($K = 1$). The CRB for the unstructured model and the geometric model with and without the direct BS-UE channel $\mathbf{H}_d$ are included. The advantage of the geometric model is clear in this example, providing 10-15dB of gain over the unstructured case, although a few dB of this gain would be lost due to imprecise BS and RIS array calibration [45].

The advantage of the geometric model in terms of training data is apparent from Fig. 3 for the same scenario with an SNR of 5dB; we see that the same performance can be achieved with more than order of magnitude fewer training samples. Note that the plots show a degradation in geometric channel estimation performance when the direct channel is present; this is due to the observation above that for the assumed model, we have $\|\mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}\| \gg \|\mathbf{H}_d\|$. In effect, when $\mathbf{H}_d$ is present, we are faced with the task of estimating additional parameters that are observed more weakly in the data.

Fig. 4 plots the CRB as a function of the number of paths present in the geometric channel model, again assuming $M = N = 30$ and the minimum number of $T = K(N+1) = $
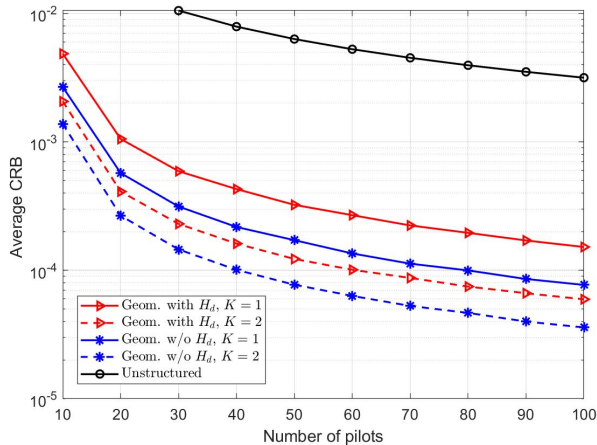
Fig. 3: CRB for various channel models vs. number of training samples T. Channel parameters are $M = 30, N = 30, \text{SNR} = 5\text{dB}, K = 2, d_F = 5, d_G = 5, d_H = 2$.
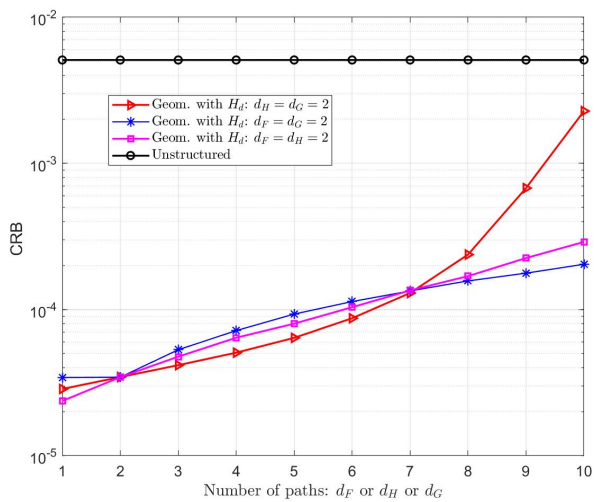

Fig. 4: CRB for various models vs. number of propagation paths. Channel parameters are $M = 30, N = 30, T = 62, K = 2, \text{SNR} = 5\text{dB}$.

Fig. 5: CRB for various channel models vs. number of RIS elements N, where $M + N = 70$. Channel parameters are $\text{SNR} = 5\text{dB}, K = 2, d_F = 5, d_G = 5, d_H = 2$. For the unstructured model and the dashed blue and red curves (geometric model), we have $T = K(N+1)$. For the solid blue and red curves (geometric model), $T = 31$.

Fig. 6: CRB for various channel models vs. SNR. Channel parameters are $M = 6, N = 64, T = 260, U = 2, K = 4, d_F = 2, d_G = 3, d_H = 2$.

62 training samples. The red curve assumes $d_H = d_G = 2$ and shows an increase in the CRB as d_F increases, the blue curve assumes $d_F = d_G = 2$ and plots versus d_H, while the magenta curve is for $d_F = d_H = 2$ versus d_G. While the presence of more paths increases the power of the received signals in this example, this gain is offset by the fact that more parameters must be estimated, and hence the geometric CRB increases with the number of paths. The increase due to larger values of d_F causes the most dramatic increase for the reason noted in the previous paragraph. For the common case where the direct channel is blocked, it is clear that the geometric model can include a relatively large number of paths before its performance degrades to the level of the unstructured case.

The last example for the first scenario is depicted in Fig. 5, which shows the achievable channel estimation performance as a function of N assuming that $M + N = 70$, with the other parameters set as before ($K = 2, d_F = 5, d_G = 5, d_H = 2, \text{SNR} = 5\text{dB}$). The curve for the unstructured model assumes that T is increasing with N according to $T = K(N+1)$, which explains why the performance improves as the size of the RIS
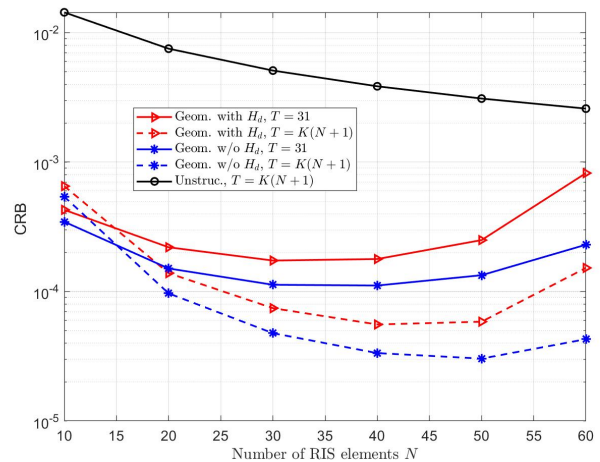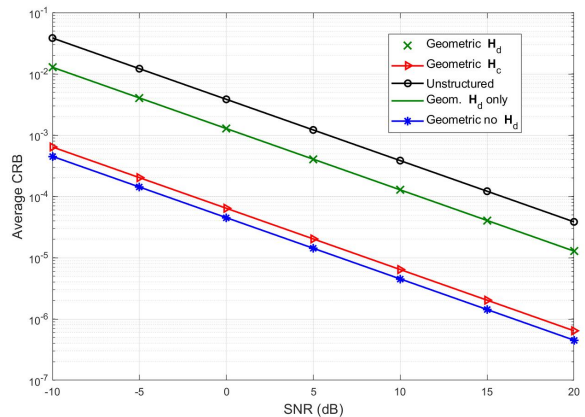
increases. The dashed curves for the geometric model assume the same increasing values for T, while the solid lines assume a fixed value of $T = 31$. For fixed T, the best performance is achieved when $M \simeq N$, while larger values of T favor the use of a larger RIS and a smaller number of BS antennas.

The second scenario is different from the first in a number of ways, but the conclusions are essentially the same. Unlike the previous case, there are $U = 2$ users with 2 antennas each ($K = 4$), the RIS training sequences are not random but rather chosen such that the diagonal CRB in (110) holds, the BS has $M = 6$ antennas, and there are $d_H = d_F = 2$ paths in the BS-RIS and BS-UE channels. Except for Fig. 8, $d_G = 3$. In addition, the Rayleigh distributed path gains in this example have a variance equal to the reciprocal of the number of paths. Figs. 6-8 assume a larger RIS with $N = 64$ elements, and thus the LS channel estimator requires at least $T = K(N+1) = 260$ training samples in order to obtain a unique estimate. This value for T will be assumed, except for the case where performance is plotted versus T.

Figs. 6-9 each show the CRB for five different cases: (1) "Geometric $\mathbf{H}_d$" - geometric model for $\mathbf{H}_d$ channel with the

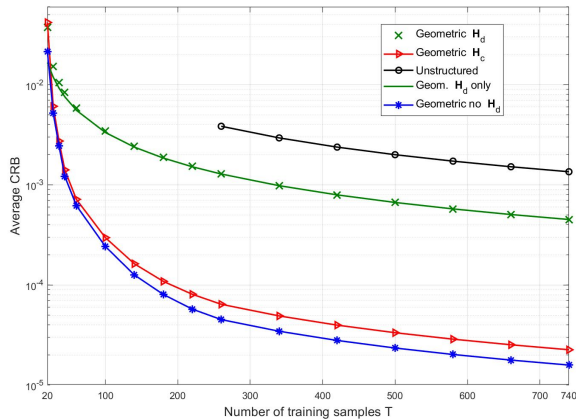Fig. 7: CRB for various channel models vs. number of training samples T. Channel parameters are $M = 6, N = 64, \mathrm{SNR} = 0\mathrm{dB}, U = 2, K = 4, d_F = 2, d_G = 3, d_H = 2$.
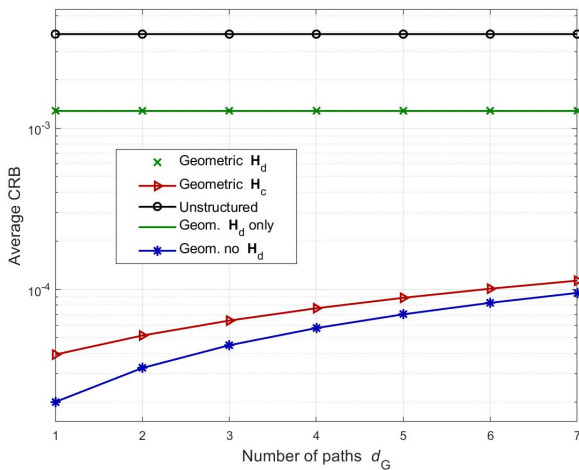

Fig. 8: CRB for various models vs. number of propagation paths d_G. Channel parameters are $M = 6, N = 64, T = 260, U = 2, K = 4, \mathrm{SNR} = 0\mathrm{dB}, d_F = 2, d_H = 2$.

RIS channel present, (2) "Geometric $\mathbf{H}_c$" - geometric model for $\mathbf{H}_c$ with $\mathbf{H}_d$ included, (3) "Unstructured" - unstructured model (independent of $\mathbf{H}_d$), (4) "Geom. $\mathbf{H}_d$ only" - geometric model when only $\mathbf{H}_d$ is present, and (5) "Geometric no $\mathbf{H}_d$" - geometric model for $\mathbf{H}_c$ without $\mathbf{H}_d$ present. In all cases, the results for "Geometric $\mathbf{H}_d$" and "Geom. $\mathbf{H}_d$ only" are essentially identical, indicating that the presence or absence of the RIS should not impact the quality of the estimate of $\mathbf{H}_d$.

Fig. 6 shows the CRB versus the SNR, and as in the previous case we see an estimation gain of about 15dB for the geometric model in the ideal case. Achievable performance versus T is illustrated in Fig. 7 for 0dB SNR. The curve for the unstructured case is only shown for $T \geq 260$ where the channel is identifiable, and as before there is at least an order of magnitude reduction in training data required for the geometric model. Fig. 8 provides results versus d_G for 0dB SNR, and we still see a considerable gap between the achievable performance of the unstructured and geometric models even as the number of propagation paths grows larger. The CRB versus the number of RIS elements $N = N_x^2$ for 0dB SNR is shown in Fig. 9 with $T = K(N+1) = 788$ in order for the unstructured model
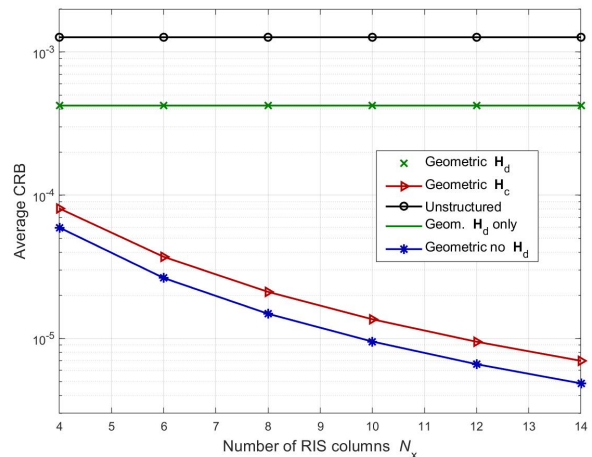

Fig. 9: CRB for various channel models vs. number of columns N_x in a square RIS with $N = N_x^2$ elements. Channel parameters are $M = 6, T = 788, \mathrm{SNR} = 0\mathrm{dB}, U = 2, K = 4, d_F = 2, d_G = 3, d_H = 2$.

to be identifiable for the case of the largest RIS.

VI. ADDITIONAL TOPICS

A. Active RIS Elements

The methods discussed above have assumed a purely passive RIS with elements that are capable of applying only a controllable phase shift to the impinging signal, together with a phase- and frequency-dependent attenuation factor. Recently, several research teams have studied the added flexibility that can be provided if the RIS is equipped with a few active receivers and local baseband processing, suggesting that there are benefits that outweigh the resulting increase in power consumption at the RIS. Most of this work has assumed only active receiving and not active transmission at the RIS, hence the additional required power is still significantly less than a similarly-sized active relay. Instead of the lower SNR data at the BS that has propagated from the UEs through the RIS, the data collected by the RIS has higher SNR due to lower path loss and will lead to higher fidelity channel estimates. In addition, the channels $\mathbf{H}$ and $\mathbf{G}$ can be separately identified, which as we have seen above is useful in cases where one component varies less rapidly than the other, and need not be estimated as frequently.

The active RIS elements are used to estimate the gains and angles of the geometric channel model, and these estimates are then used to infer the full $\mathbf{H}$ and $\mathbf{G}$ matrices using the known RIS geometry [46–48] or a deep neural network (DNN) [49]. This matrix completion approach is similar to methods proposed for channel estimation in hybrid digital/analog systems. A tensor completion approach is proposed in [50], and also extended to the wideband OFDM case. The active elements can either be embedded in fixed positions with a separate RF chain for each as in the methods above, or different linear combinations of the RIS outputs can be combined together through a single RF chain over the training interval to achieve a similar result [51]. We also mention here the related work of [52] which, instead of active RIS elements, proposes deploying active anchor nodes in known locations near the UEs to reduce the CSI estimation complexity as discussed in Section IV-G

and resolve the ambiguity of the composite channel.

B. Double RIS Systems

We have only considered communication links between a BS and UEs that include a single RIS, but channel estimation for scenarios with two RIS have been studied for certain special cases. In [53], a double RIS scenario is considered where a BS communicates with a single UE over an indirect channel that passes from the BS to RIS 1 to RIS 2 to the UE, assuming all other channels are blocked. They propose an unstructured LS channel estimator that requires at least $N_1 N_2$ training samples, assuming the two RIS have N_1 and N_2 elements respectively. They also consider the special case of an LoS channel between the two RIS, which reduces the training overhead to $N_1 + N_2$. A more general scenario is considered in [54], in which both RIS have unblocked channels with the BS and UEs, in addition to the direct link between the two RIS. A two-stage unstructured LS approach is proposed where, in the first, the phase shifts of RIS 1 are held fixed while those of RIS 2 are varied, and the composite single-bounce (UE-IRS 2-BS) and double-bounce (UE-IRS 2-IRS 1-BS) channels are estimated. In stage 2, the phase of both RIS are varied in order to estimate the corresponding components of RIS 1, which are superimposed on the composite channels estimated in stage 1. Assuming K single-antenna UEs, the required training overhead for this approach is of the following order:

$$T = O\Big(\frac{K N_1 N_2}{M} + K(N_1 + N_2)\Big), \tag{112}$$

which can be large when the number of BS antennas M is small. An earlier approach to the double-RIS problem in [55] assumed LoS propagation between two RIS that both possess active receivers, which allowed the individual channel components to be determined. It is unclear if a geometric model could be used for a purely passive double-RIS scenario to achieve an identifiable parameterization.

C. Learning-Based Methods

The application of machine learning to RIS systems and channel estimation has also been growing recently. For example, the approach in [56] designs a DNN that directly takes the received training data and produces the channel estimate. The DNN is trained using synthetic uplink data generated assuming the direct and composite channels are Rayleigh fading. As has been noted in other work, the performance improves if not only the real and imaginary parts of the input data are provided, but also a third component related to the data, in this case its magnitude. Other work has employed the phase of the data as the third component [57]. However, [57] uses a federated learning approach in the downlink, in which the UEs generate local channel estimates using a learning-based optimization, and the gradients of the local networks are fed back to the BS to update the global model.

A common learning-based approach for RIS systems involves denoising an initial channel estimate with a DNN. The initial estimate can be found using the methods described above, such as LS [31], [58] or compressive sensing in the geometric model [59]. Multi-stage DNN denoising networks are proposed in [60], [61], where separate DNNs are used to first denoise the direct channel (if present) with the RIS switched off, then to denoise a reduced-dimension version of the composite channel with only a subset of the RIS elements active, and finally to map the sampled composite channel to its full dimensions. A similar technique is proposed in [49] that uses a DNN to map the channel estimated with a few active RIS elements to the full RIS array. The approach in [47] also assumes the availability of active RIS receivers, but instead of attempting to estimate the channel with the full RIS, the sampled channels from the active RIS subarray are used as "environment descriptors" during training to find the best codebook of RIS phases that yields the highest sum rate.

The primary advantage of the above learning-based methods is that, once the neural networks are trained, the channel estimates are obtained with relatively little computation. However, since the training is performed using synthetic data, the ability of the network to handle non-idealities not present in the simulated training data is still an open question.

VII. CONCLUSIONS

Wireless channel estimation for RIS-based systems provides a rich source of interesting research problems. We have highlighted some of the solutions to these problems for two general classes of channel models: unstructured models that make no assumptions about the propagation environment or the RIS or array geometries, and structured or geometric models that rely on the assumption of sparse propagation paths and calibrated antenna arrays and RIS element responses. Algorithms for estimating unstructured channels are conceptually simple and robust, but for RIS-based systems they require a large training overhead and their achievable accuracy is limited due to the large number of channel coefficients to be estimated. On the other hand, geometric channel models lead to estimation of many fewer parameters and hence have a much smaller training burden and can achieve dramatically better performance. However, this improvement comes at the cost of increased algorithm complexity, a requirement to determine the model order (*i.e.,* number of propagation paths), and some of the performance gain will be lost in practice due to inevitable modeling errors. We have highlighted these issues using both theoretical derivations and numerical CRB examples, and we have also briefly discussed on-going research related to RIS with active elements, double-RIS systems, and machine learning algorithms. In addition to these areas, many open research problems remain, including how to estimate channels for multiple RIS when they are visible to multiple basestations and their reflections interact, how to take into account more realistic models of RIS behavior (coupled dependence of gains and phases over frequency, mutual coupling, etc.), how to reduce the complexity of compressive-sensing based approaches that normally require huge dictionaries for RIS-based CSI estimation, how to exploit available side information such as knowledge of the local propagation environment or known channel statistics for optimal design of pilots and RIS phases during training, etc.

REFERENCES

[1] T. Lindstrøm Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance

unbiased estimator," in *Proc. IEEE ICASSP*, 2020, pp. 5000–5004.

[2] 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, Dec. 2019.

[3] 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.

[4] 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, Jul. 2020.

[5] Q.-U.-A. 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. of the Commun. Soc.*, vol. 1, pp. 661–680, May 2020.

[6] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, Apr. 2021.

[7] 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.

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

[9] 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.

[10] Y. Wei, M.-M. Zhao, M.-J. Zhao, and Y. Cai, "Channel estimation for IRS-aided multiuser communications with reduced error propagation," Jan. 2021. [Online]. Available: https://arxiv.org/abs/2101.03314

[11] B. Zheng, C. You, and R. Zhang, "Fast channel estimation for IRS-assisted OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 580–584, 2021.

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

[13] G. T. de Araujo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," vol. 15, no. 3, Feb. 2021, pp. 789–802.

[14] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmwave massive MIMO systems," in *Proc. IEEE ICC*, Jul. 2020.

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

[16] S. Noh, H. Yu, and Y. Sung, "Training signal design for sparse channel estimation in intelligent reflecting surface-assisted millimeter-wave communication," Dec. 2020. [Online]. Available: https://arxiv.org/abs/2012.14221

[17] X. Ma, Z. Chen, W. Chen, Z. Li, Y. Chi, C. Han, and S. Li, "Joint channel estimation and data rate maximization for intelligent reflecting surface assisted terahertz MIMO communication systems," *IEEE Access*, vol. 8, pp. 99 565–99 581, 2020.

[18] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," June 2021. [Online]. Available: https://arxiv.org/abs/2106.14792

[19] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmwave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, Apr. 2021.

[20] 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.

[21] 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 Sig. Proc. Lett.*, vol. 28, pp. 513–517, Feb. 2021.

[22] S. Ma, W. Shen, J. An, and L. Hanzo, "Wideband channel estimation for IRS-aided systems in the face of beam squint," Jun. 2021. [Online]. Available: https://arxiv.org/abs/2106.02883

[23] Z. Chen, J. Tang, X. Y. Zhang, D. K. C. So, S. Jin, and K.-K. Wong, "Hybrid evolutionary-based sparse channel estimation for IRS-assisted mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, 2021.

[24] 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.*, 2021.

[25] H. Liu, J. Zhang, Q. Wu, H. Xiao, and B. Ai, "ADMM based channel estimation for RISs aided millimeter wave communications," *IEEE Commun. Lett.*, 2021.

[26] S. Kim, H. Lee, J. Cha, S. Kim, J. Park, and J. Choi, "Practical channel estimation and phase shift design for intelligent reflecting surface empowered MIMO systems," Apr. 2021. [Online]. Available: https://arxiv.org/abs/2104.14161

[27] E. Björnson, "Optimizing a binary intelligent reflecting surface for OFDM communications under mutual coupling," Jun. 2021. [Online]. Available: https://arxiv.org/abs/2106.04280

[28] Z.-Q. He, H. Liu, X. Yuan, Y.-J. A. Zhang, and Y.-C. Liang, "Semi-blind cascaded channel estimation for reconfigurable intelligent surface aided massive MIMO," Feb. 2021. [Online]. Available: https://arxiv.org/abs/2101.07315

[29] 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, 2020.

[30] J. Mirza and B. Ali, "Channel estimation method and phase shift design for reconfigurable intelligent surface assisted MIMO networks," *IEEE Trans. Cognitive Commun. & Net.*, vol. 7, no. 2, pp. 441–451, 2021.

[31] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided MISO communications: From LMMSE to deep learning solutions," *IEEE Open J. of the Commun. Soc.*, vol. 2, pp. 471–487, 2021.

[32] S. Sumanthiran, D. Kudathanthirige, K. T. Hemachandra, T. Samarasinghe, and G. A. Baduge, "Rank-1 matrix approximation-based channel estimation for intelligent reflecting surface-aided multi-user MISO communications," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2589–2593, 2021.

[33] W. Cai, H. Li, M. Li, and Q. Liu, "Practical modeling and beamforming for intelligent reflecting surface aided wideband systems," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1568–1571, 2020.

[34] W. Yang, H. Li, M. Li, Y. Liu, and Q. Liu, "Channel estimation for practical IRS-assisted OFDM systems," in *Proc. IEEE WCNC Workshops*, 2021.

[35] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Ant. & Prop.*, vol. 34, no. 3, pp. 276–280, 1986.

[36] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT - A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoustics, Speech, & Sig. Proc.*, vol. 34, no. 5, pp. 1340–1342, 1986.

[37] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoustics, Speech, & Sig. Proc.*, vol. 38, no. 7, pp. 1132–1143, 1990.

[38] M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Sig. Proc.*, vol. 43, no. 2, pp. 469–477, 1995.

[39] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Info. Theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[40] I. Bilik, "Spatial compressive sensing for direction-of-arrival estimation of multiple sources using dynamic sensor arrays," *IEEE Trans. Aero. & Elec. Sys.*, vol. 47, no. 3, pp. 1754–1769, 2011.

[41] A. C. Gurbuz, V. Cevher, and J. H. Mcclellan, "Bearing estimation via spatial sparsity using compressive sensing," *IEEE Trans. Aero. & Elec. Sys.*, vol. 48, no. 2, pp. 1358–1369, 2012.

[42] Q. Shen, W. Liu, W. Cui, and S. Wu, "Underdetermined DOA estimation under the compressive sensing framework: A review," *IEEE Access*, vol. 4, pp. 8865–8878, 2016.

[43] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Trans. Sig. Proc.*, vol. 63, no. 12, pp. 3139–3153, 2015.

[44] C. You, B. Zheng, and R. Zhang, "Fast beam training for IRS-assisted multiuser communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1845–1849, 2020.

[45] S. Rao, A. Mezghani, and A. L. Swindlehurst, "Channel estimation in one-bit massive MIMO systems: Angular versus unstructured models," *IEEE J. Sel. Topics Sig. Proc.*, vol. 13, no. 5, pp. 1017–1031, 2019.

[46] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Trans. Wireless Commun.*, 2021.

[47] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, 2021.

[48] 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, 2021.

[49] 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. Tech.*, 2021.

[50] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based algebraic channel estimation for hybrid IRS-assisted MIMO-OFDM," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3770–3784, 2021.

[51] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *Proc. IEEE ICASSP*, 2020, pp. 9175–9179.

[52] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," Feb. 2021. [Online]. Available: https://arxiv.org/abs/2102.10886

[53] C. You, B. Zheng, and R. Zhang, "Wireless communication via double IRS: Channel estimation and passive beamforming designs," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 431–435, Feb. 2021.

[54] B. Zheng, C. You, and R. Zhang, "Efficient channel estimation for double-IRS aided multi-user MIMO system," vol. 69, no. 6, Mar. 2021, pp. 3818–3832.

[55] Y. Han, S. Zhang, L. Duan, and R. Zhang, "Cooperative double-IRS aided communication: Beamforming design and power scaling," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1206–1210, 2020.

[56] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, 2020.

[57] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and IRS-assisted massive MIMO," Aug. 2020. [Online]. Available: https://arxiv.org/abs/2008.10846

[58] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, 2021.

[59] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmwave intelligent reflecting surfaces," *IEEE Trans. Veh. Tech.*, vol. 69, no. 8, pp. 9223–9228, 2020.

[60] Y. Wang, H. Lu, and H. Sun, "Channel estimation in IRS-enhanced mmWave system with super-resolution network," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2599–2603, 2021.

[61] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep multi-stage CSI acquisition for reconfigurable intelligent surface aided MIMO systems," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2024–2028, 2021.

Gui Zhou (Graduate Student Member, IEEE) received the B.S. and M.E. degrees from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2015 and 2019, respectively. She is currently pursuing the Ph.D. degree at the School of electronic Engineering and Computer Science, Queen Mary University of London, U.K. Her major research interests include intelligent reflection surface (IRS) and signal processing.

Rang Liu (Graduate Student Member, IEEE) received the B.S. degree in 2018 in electronics information engineering from the Dalian University of Technology, Dalian, China, where she is currently working toward the Ph.D. degree with the School of Information and Communication Engineering. Her current research interests include signal processing, mmWave communications, massive MIMO systems, and reconfigurable intelligent surfaces.

Cunhua Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he was a Research Associate at the University of Kent, U.K. He held a post-doctoral position at Queen Mary University of London, U.K., from 2016 and 2019, where he is currently a Lecturer. His research interests mainly include intelligent reflection surface (IRS), machine learning, UAV, Internet of Things (IoTs), and mobile edge computing. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019. He also serves as an Editor of IEEE Wireless Communication Letters, IEEE Communication Letters and IEEE Access.

Ming Li (Senior Member, IEEE) received the M.S. and Ph.D. degrees in Electrical Engineering from the State University of New York at Buffalo (SUNY-Buffalo), Buffalo, in 2005 and 2010, respectively. From January 2011 to August 2013, he was a Post-Doctoral Research Associate with the Signals, Communications, and Networking Research Group, Department of Electrical Engineering, SUNY-Buffalo. From August 2013 to June 2014, he joined Qualcomm Technologies Inc., as a Senior Engineer. Since June 2014, he has been with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, where he is currently an Associate Professor. His current research interests include the general areas of communication theory and signal processing with applications to mmWave communications, secure wireless communications, cognitive radios and networks, data hiding, and steganography.

A. Lee Swindlehurst (Fellow, IEEE) received the B.S. (1985) and M.S. (1986) degrees in Electrical Engineering from Brigham Young University (BYU), and the PhD (1991) degree in Electrical Engineering from Stanford University. He was with the Department of Electrical and Computer Engineering at BYU from 1990-2007.During 1996-97, he held a joint appointment as a visiting scholar at Uppsala University and the Royal Institute of Technology in Sweden. From 2006-07, he was on leave working as Vice President of Research for ArrayComm LLC in San Jose, California. Since 2007 he has been a Professor in the Electrical Engineering and Computer Science Department at the University of California Irvine. During 2014-17 he was also a Hans Fischer Senior Fellow in the Institute for Advanced Studies at the Technical University of Munich. In 2016, he was elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA). His research focuses on array signal processing for radar, wireless communications, and biomedical applications, and he has over 350 publications in these areas. Dr. Swindlehurst is a Fellow of the IEEE and was the inaugural Editor-in-Chief of the IEEE Journal of Selected Topics in Signal Processing. He received 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 and 2010 IEEE Signal Processing Society's Best Paper Awards, the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award, and a Best Paper award at the 2020 IEEE International Conference on Communications.