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*[4], and Ming Li *Senior Member, IEEE*[3]

[1]Center for Pervasive Communications & Computing, University of California, Irvine, CA, USA
[2]School of Electronic Engineering & Computer Science, Queen Mary University of London, England
[3]School of Information & Communication Engineering, Dalian University of Technology, Dalian, China
[4]National Mobile Communications Research Laboratory, Southeast University, Nanjing, 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 when the RIS is composed of passive elements without any sensing or 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 passive 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 multi-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 simulations of representative algorithms and comparisons of their 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 [1], [2]. 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 often constructed only as passive devices without active RF receivers or computational resources. Consequently, channel estimation for RIS-based systems has been a subject of intense study.

When the RIS is composed of passive elements, the CSI must be estimated by devices – most often a basestation (BS) or access point – that are not colocated 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 work on CSI estimation for passive RIS-based systems has been published recently. Initially, this work focused on estimating *unstructured* models, where the channels are simply described using complex gains [3–14]. 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 [15–29]. 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.

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. Consistent with most work that has been published to date, we will focus on channel estimation for scenarios involving a single passive RIS, although a brief discussion about extensions to cases with multiple RIS or RIS with active RF and digital processing will be provided at the end.

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. We summarize some of the trade-offs associated with channel estimation for unstructured and geometric models in Section V, and we include some representative numerical examples in Section VI to illustrate their relative performance. Several additional topics will be briefly considered in Section VII, 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 VIII.

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, which to date is the most common case considered in the literature (see [20], [24], [25] for approaches using a hybrid RF network). 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 $\text{vec}(\mathbf{A})$ 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 $\text{diag}(\mathbf{c})$, and a block diagonal matrix with block entries $\mathbf{C}_1, \mathbf{C}_2, \ldots$ is written $\text{blkdiag}\left([\mathbf{C}_1 \;\; \mathbf{C}_2 \;\; \ldots]\right)$.

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 $\boldsymbol{\phi} = [\beta_1 e^{j\alpha_1} \;\; \cdots \;\; \beta_N e^{j\alpha_N}]^T$. 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

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

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 UE, 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

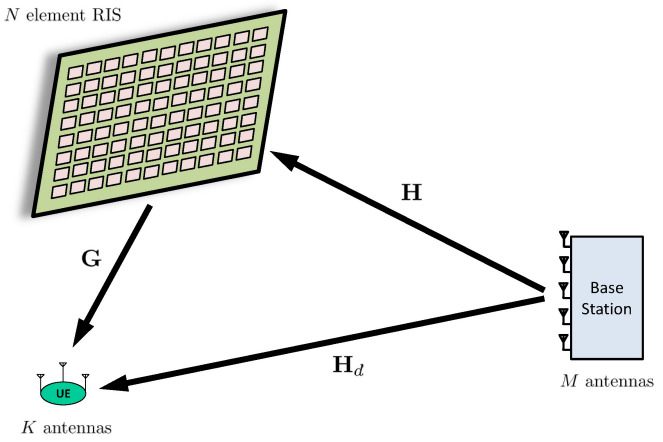

Fig. 1: A generic scenario involving an M-antenna basestation, an N-element RIS, and a K-element user.

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.,* $\mathbf{y}_t \longrightarrow \mathbf{y}_t - \sqrt{P}\mathbf{H}_d\mathbf{x}_t$.

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

$$\mathbf{H}\boldsymbol{\Phi}_t\mathbf{G}^H = \mathbf{H}\boldsymbol{\Lambda}\boldsymbol{\Phi}_t\boldsymbol{\Lambda}^{-1}\mathbf{G}^H = \tilde{\mathbf{H}}\boldsymbol{\Phi}_t\tilde{\mathbf{G}}^H \;, \tag{2}$$

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:

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

where $\mathbf{h}_d = \mathrm{vec}(\mathbf{H}_d)$, $\boldsymbol{\phi}_t$ is a vector containing the diagonal elements of $\boldsymbol{\Phi}_t$, and $\tilde{\boldsymbol{\phi}}_t$ is defined implicitly in the last equality.

Given an estimate of $\mathbf{H}_c$, if one wishes to obtain a separate estimate of $\mathbf{H}$ and $\mathbf{G}$, then a method is needed to resolve the ambiguity in (2). For example, $\mathbf{H}_c$ could be uniquely decomposed if one fixes the first row of $\mathbf{H}$ or $\mathbf{G}$ to be some known vector (*e.g.,* all ones). Another method to resolve the ambiguity will be discussed later in the context of geometric channel models. Eq. (3) together with further use of the Kronecker product allows us to rewrite (1) in a compact form:

$$\begin{aligned} \mathbf{y}_t &= \sqrt{P}\left(\mathbf{x}_t^T \otimes \mathbf{I}_M\right)\mathbf{H}_c\tilde{\boldsymbol{\phi}}_t + \mathbf{n}_t && \text{(4a)} \\ &= \sqrt{P}\left[\tilde{\boldsymbol{\phi}}_t^T \otimes \mathbf{x}_t^T \otimes \mathbf{I}_M\right]\mathbf{h}_c + \mathbf{n}_t && \text{(4b)} \\ &\equiv \sqrt{P}\mathbf{Z}_t\mathbf{h}_c + \mathbf{n}_t \;, && \text{(4c)} \end{aligned}$$

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

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

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,

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

whose solution is given by

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

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 the Cramér-Rao bound (CRB):

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{h}}_{c,LS}} &= \mathcal{E}\left\{\left(\hat{\mathbf{h}}_{c,LS} - \mathbf{h}_c\right)\left(\hat{\mathbf{h}}_{c,LS} - \mathbf{h}_c\right)^H\right\} && \text{(8a)} \\ &= \frac{1}{P}\mathcal{E}\left\{\mathbf{Z}^\dagger\boldsymbol{\mathfrak{n}}\boldsymbol{\mathfrak{n}}^H(\mathbf{Z}^\dagger)^H\right\} = \frac{\sigma^2}{P}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1} \;. && \text{(8b)} \end{aligned}$$

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 [3] 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 subblock, $b = 1, \cdots, T/K$, $\boldsymbol{\phi}_t = \bar{\boldsymbol{\phi}}_b$ 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

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

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

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

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 (*i.e.*, 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] [7], [30]. This results in $\boldsymbol{\Psi}^H\boldsymbol{\Psi} = \text{diag}\{\beta_1^2, \cdots, \beta_N^2\}$. If the RIS elements are identical, and each active element is tuned to the same phase, then $\beta_i = \beta$, which results in $\mathbf{Z}^H\mathbf{Z} = \beta^2 K\mathbf{I}_{MKN}$. This in turn leads to 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) [3], [7]:
$$[\boldsymbol{\Psi}]_{mn} = e^{j2\pi(m-1)(n-1)/(T/K)} \tag{16}$$
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 would have to 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 [6], [31]. 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.

[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.

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 [5], [32–34]. Instead, the linear MMSE, or LMMSE, estimate given by $\hat{\mathbf{h}}_{c,LM} = \mathbf{W}\mathfrak{y}$ can be found by solving [35], [36]

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

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

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

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 by taking sample statistics of $\hat{\mathbf{h}}_c$ obtained over a long period of time, or using simulations involving detailed propagation models of the environment. However, the size of $\mathbf{h}_c$ means that such procedures would require a large amount of data. At first glance, a more reasonable approach might be 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 would normally lead to the following descriptions:

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

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

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

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 would have the following form:

$$\mathbf{R}_{\mathbf{h}_c} = \begin{bmatrix} \mathbf{R}_{H_dU}^T \otimes \mathbf{R}_{H_dB} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{R}_R \otimes \mathbf{R}_{GU}^T \otimes \mathbf{R}_{HB} \end{bmatrix} \ , \tag{22}$$

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

However, there are problems with the use of the so-called Kronecker product channel correlation model for $\mathbf{H}$ and $\mathbf{G}$ in (21a)-(21b). First, it has been shown that the model is not accurate for RIS channels, especially when N is large or the spacing between RIS elements is small [37]. Second, the inherent ambiguity illustrated in Eq. (2) poses a difficulty for describing $\mathbf{H}$ or $\mathbf{G}$ as in (21a)-(21b). Even if the actual channels followed a Gaussian model described by (21a)-(21b), we do not have access to the actual channels (or even estimates thereof), only versions for which the ambiguity in (2) has been removed. For example, resolving the ambiguity by fixing the first row of $\mathbf{H}$ to be a vector of ones corresponds to dividing each column of the BS-RIS channel by its first element, rendering an $\mathbf{H}$ – and also a $\mathbf{G}$ – that will no longer be Gaussian. Thus, estimation of the covariance matrices in (21a)-(21b) is problematic. Various assumptions could be made to simplify direct calculation of $\mathbf{R}_{\mathbf{h}_c}$, such as assuming for example that the multipath scattering at the RIS is isotropic, in which case $\mathbf{R}_{GR}$ and $\mathbf{R}_{HR}$ could be taken as identity matrices, and $\mathbf{R}_{\mathbf{h}_c}$ would be block diagonal with identical block entries except for the block associated with $\mathbf{H}_d$. This would greatly simplify computation of $\mathbf{R}_{\mathbf{h}_c}$ and the estimate in (19). However, the fact that the RIS multipaths are confined to only the half space associated with the active side of the RIS means that, even with isotropic scattering in that half space, an identity covariance for $\mathbf{R}_{GR}$ and $\mathbf{R}_{HR}$ only holds for certain special RIS geometries, none of which correspond to the standard uniform planar array (UPA) architecture assumed for RIS [37]. Furthermore, assumptions of uncorrelated fading are hard to justify in RIS-aided wireless systems, which are often 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 $\mathbf{X}_t^F = \begin{bmatrix} \mathbf{x}_{t,1}^F & \cdots & \mathbf{x}_{t,N_c}^F \end{bmatrix}$ 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:

$$\mathbf{y}_{t,n}^F = \sqrt{P}\left(\mathbf{H}_{d,n}^F + \mathbf{H}_n^F \mathbf{\Phi}_{t,n} \mathbf{G}_n^{FH}\right) \mathbf{x}_{t,n}^F + \mathbf{n}_{t,n}^F \ , \tag{23}$$

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 is used to construct channel estimates for others [4]. An alternative approach proposed in [13] 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.,* $\mathbf{\Phi}_{t,n} = \mathbf{\Phi}_t$, but this is generally true only for relatively narrow bandwidths [38], [39]. 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

$$\mathbf{y}_{t,s} = \sqrt{P} \sum_{k=0}^{L-1} \Big(\mathbf{H}_d(k) + \mathbf{H}(k) \mathbf{\Phi}_{t,s-k} \mathbf{G}^H(k) \Big) \mathbf{x}_{t,s-k} + \mathbf{n}_{t,s} \ , \tag{24}$$

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

$$\mathbf{y}_{t,s} = \sqrt{P}\sum_{k=0}^{L-1}\left[\tilde{\boldsymbol{\phi}}_{t,s-k}^T \otimes \mathbf{x}_{t,s-k}^T \otimes \mathbf{I}_M\right]\mathbf{h}_c(k) + \mathbf{n}_{t,s} \quad (25a)$$

$$= \sqrt{P}\sum_{k=0}^{L-1}\mathbf{Z}_{t,s-k}\mathbf{h}_c(k) + \mathbf{n}_{t,s} \quad (25b)$$

$$= \sqrt{P}\Big[\mathbf{Z}_{t,s} \;\; \mathbf{Z}_{t,s-1} \;\; \cdots \;\; \mathbf{Z}_{t,s-L+1}\Big]\mathfrak{h}_c + \mathbf{n}_{t,s} \quad (25c)$$

$$\mathfrak{y}_t = \begin{bmatrix}\mathbf{y}_{t,1}\\ \vdots \\ \mathbf{y}_{t,N_c}\end{bmatrix} = \sqrt{P}\mathfrak{Z}_t\mathfrak{h}_c + \mathfrak{n}_t\ , \quad (25d)$$

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

$$\mathfrak{y} = \sqrt{P}\begin{bmatrix}\mathfrak{Z}_1\\ \vdots \\ \mathfrak{Z}_{T_o}\end{bmatrix}\mathfrak{h}_c + \mathfrak{n} = \sqrt{P}\mathfrak{Z}\mathfrak{h}_c + \mathfrak{n}\ . \quad (26)$$

The time-domain approach assumes only pilot data is transmitted first, followed by payload data. The total number of pilot symbols required is $T = T_oN_c \geq KL(N+1)$. 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

$$\mathbf{y}_t = \sqrt{P}\left(\mathbf{h}_d + \mathbf{H}\,\text{diag}(\mathbf{g}^*)\boldsymbol{\phi}_t\right)x_t + \mathbf{n}_t\ , \quad (27)$$

where the composite channel is now $\mathbf{H}_c = \mathbf{g}^H \diamond \mathbf{H} = \mathbf{H}\,\text{diag}(\mathbf{g}^*)$. 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

$$y_t = \sqrt{P}\mathbf{h}_c^T\tilde{\boldsymbol{\phi}}_t x_t + n_t \quad (28a)$$

$$\mathbf{h}_c^T = \begin{bmatrix}h_d & \mathbf{g}^H \odot \mathbf{h}^T\end{bmatrix} = \begin{bmatrix}h_d & \bar{\mathbf{h}}_c^T\end{bmatrix}\ , \quad (28b)$$

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 [4], [6]. High channel correlation occurs when adjacent RIS elements are closely spaced; retaining the flexibility of arbitrary phase shifts for such elements does not provide a significant increase in degrees of freedom for beamforming design, since the designed phases would likely be nearly identical. 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

$$\mathbf{H}_c\boldsymbol{\phi}_t = \mathbf{H}_c(\boldsymbol{\phi}_t' \otimes \mathbf{1}_J) = \mathbf{H}_c(\mathbf{I}_{N'} \otimes \mathbf{1}_J)\boldsymbol{\phi}_t' = \mathbf{H}_c'\boldsymbol{\phi}_t'\,, \quad (29)$$

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 [6] 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

$$\mathbf{Z}^T = \mathbf{U}\mathbf{V} \tag{30}$$

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 (30) is generally not possible if $T < K(N+1)$. It may, however, be possible to approximately solve (30) 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 amount of data required to estimate $\mathbf{R}_{\mathbf{h}_c}$ as well as 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:

$$\hat{\mathbf{h}}_c = \frac{\sqrt{P}}{\sigma^2}\mathbf{U}\left[\mathbf{I}_r - \mathbf{W}\left(\mathbf{W} + \frac{\sigma^2}{P}\mathbf{I}_r\right)^{-1}\right]\mathbf{U}^H\mathbf{Z}^H\boldsymbol{\mathfrak{y}}, \tag{31}$$

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 $\mathbf{H}$ [10]. 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 [9], [12]. 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 $\boldsymbol{\Lambda}$ in (2), so step 1 provides us with an estimate of $\tilde{\mathbf{H}}$, and we can set the first row of $\tilde{\mathbf{G}}^H$ to $\mathbf{1}_K$. With the estimate $\hat{\tilde{\mathbf{H}}}$, during step 2 the training data model is approximately given by

$$\begin{aligned}
\mathbf{y}_t &\simeq \sqrt{P}\left(\tilde{\mathbf{H}}_d + \hat{\tilde{\mathbf{H}}}\boldsymbol{\Phi}_t\tilde{\mathbf{G}}^H\right)\mathbf{x}_t + \mathbf{n}_t && \text{(32a)}\\
&\simeq \sqrt{P}\underbrace{\mathbf{x}_t^T \otimes \left[\mathbf{I}_M \quad \hat{\tilde{\mathbf{H}}}\boldsymbol{\Phi}_t\right]}_{M\times(K-1)(M+N)}\begin{bmatrix}\tilde{\mathbf{h}}_d\\ \tilde{\mathbf{g}}^*\end{bmatrix} + \mathbf{n}_t && \text{(32b)}\\
&\simeq \sqrt{P}\tilde{\mathbf{Z}}_t\tilde{\mathbf{h}}_c + \mathbf{n}_t\ , && \text{(32c)}
\end{aligned}$$

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 $\hat{\tilde{\mathbf{h}}}_c = \mathbf{Z}^\dagger\boldsymbol{\mathfrak{y}}/\sqrt{P}$ 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

$$T_{min} = (N+1) + \left(\frac{N}{M}+1\right)(K-1)\ , \tag{33}$$

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

$$\mathbf{a}_x(\omega_x) = [1 \;\; e^{j\omega_x} \;\; e^{j2\omega_x} \;\; \cdots \;\; e^{j(M_x-1)\omega_x}]^T\ , \tag{34}$$

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 $M_x \times M_y$ UPA with antenna separations of Δ_x and Δ_y in the x and y directions, the array response vector can be written as

$$\mathbf{a}(\boldsymbol{\omega}) = \mathbf{a}_x(\omega_x) \otimes \mathbf{a}_y(\omega_y)\ , \tag{35}$$

where the vertical array response component is similar to (34),

$$\mathbf{a}_y(\omega_y) = [1 \;\; e^{j\omega_2} \;\; e^{j2\omega_2} \;\; \cdots \;\; e^{j(M_y'-1)\omega_2}]^T\ , \tag{36}$$

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 UPA, there is a one-to-one correspondence between the angles and spatial frequencies as long as both Δ_x and Δ_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.

[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 [26].

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 often made for scenarios with low-rank near-specular propagation at high frequencies, where blockages are common. However, such blockages may not be permanent, and given the lower path loss that a direct path signal would likely undergo, the direct channel $\mathbf{H}_d$ can make a significant contribution to the received signal. In such cases, $\mathbf{H}_d$ could be estimated using standard algorithms and relatively little pilot data with the RIS switched off. Consequently, we will not explicitly consider estimation of $\mathbf{H}_d$ in this section. 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 UPA, 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:

$$\mathbf{H} = \sum_{k=1}^{d_H} \gamma_{H,k} \mathbf{a}_B(\boldsymbol{\omega}_{BH,k}) \mathbf{a}_U^H(\boldsymbol{\omega}_{RH,k}) \tag{37a}$$

$$= \mathbf{A}_B(\boldsymbol{\omega}_{BH}) \boldsymbol{\Gamma}_H \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \,, \tag{37b}$$

where the columns of

$$\mathbf{A}_B(\boldsymbol{\omega}_{BH}) = [\mathbf{a}_B(\omega_{BH,1}) \;\; \cdots \;\; \mathbf{a}_B(\omega_{BH,d_H})] \tag{37c}$$

$$\mathbf{A}_R(\boldsymbol{\omega}_{RH}) = [\mathbf{a}_R(\boldsymbol{\omega}_{RH,1}) \;\; \cdots \;\; \mathbf{a}_R(\boldsymbol{\omega}_{RH,d_H})] \tag{37d}$$

respectively represent the steering vectors for the propagation paths with AoA spatial frequencies $\boldsymbol{\omega}_{BH} = [\omega_{BH,1} \; \cdots \; \omega_{BH,d_H}]^T$ at the BS and AoD spatial frequencies $\boldsymbol{\omega}_{RH} = [\boldsymbol{\omega}_{RH,1}^T \; \cdots \; \boldsymbol{\omega}_{RH,d_H}^T]^T$ from the RIS. The diagonal matrix $\boldsymbol{\Gamma}_H = \mathrm{diag}\{\boldsymbol{\gamma}_H\} = \mathrm{diag}\{\gamma_{H,1} \; \cdots \; \gamma_{H,d_H}\}$ 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:

$$\boldsymbol{\omega}_{RH,k} = \begin{bmatrix} \omega_{RH,k,x} \\ \omega_{RH,k,y} \end{bmatrix} . \tag{38}$$

Parametric models like (37) 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:

$$\mathbf{Y}' = \mathbf{A}(\boldsymbol{\omega}')\mathbf{S}' + \mathbf{N}' \,, \tag{39}$$

where $\mathbf{Y}'$ is $M' \times n$, $\mathbf{S}'$ is $d \times n$, $\mathbf{N}'$ is noise, $\mathbf{A} = [\mathbf{a}(\boldsymbol{\omega}'_1) \;\; \cdots \;\; \mathbf{a}(\boldsymbol{\omega}'_d)]$ is the $M' \times d$ array response matrix. 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

$$p_B(\boldsymbol{\omega}) = \mathbf{a}^H(\boldsymbol{\omega}) \mathbf{R}_{Y'} \mathbf{a}(\boldsymbol{\omega}) \,, \tag{40}$$

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

$$\mathbf{R}_{Y'} = \frac{1}{n} \mathbf{Y}' \mathbf{Y}'^H \,. \tag{41}$$

Alternatively, one can employ higher resolution algorithms such as MUSIC [40] or ESPRIT [41], which require computation of the eigendecompositon of $\mathbf{R}_{Y'}$. If $\mathbf{N}'$ is spatially and temporally white, the (deterministic) maximimum likelihood (ML) method [42] finds the AoA estimates from the d-dimensional (or $2d$-dimensional for azimuth/elevation angles) problem

$$\hat{\boldsymbol{\omega}}'_{ML} = \arg\min_{\boldsymbol{\omega}} \mathrm{trace}\left(\mathbf{P}_A^{\perp}(\boldsymbol{\omega}) \mathbf{R}_{Y'}\right) \,, \tag{42}$$

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 $\mathrm{rank}(\mathbf{S}') = d$ and thus that $n \geq d$, but the beamforming and ML methods are theoretically viable for smaller values of n. The theoretical limit for identifiability is $d < (M' + n)/2$, so for large enough M', n can be as small as 1 [43]. For a massive antenna array where M' is large, all of the above methods provide asymptotically efficient AoA estimates [44], 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 [45–48]. For example, we can represent the array response vectors for the BS side as

$$\mathbf{A}_B(\boldsymbol{\omega}_{BH}) = \mathbf{A}_{BD} \mathbf{Q}_{BH} \,, \tag{43}$$

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 (43) 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:

$$\mathbf{A}_R(\boldsymbol{\omega}_{RH}) = \mathbf{A}_{RD} \mathbf{Q}_{RH} \,, \tag{44}$$

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. Substituting (43) and (44) into (37), we can write

$$\mathbf{H} = \mathbf{A}_{BD} \mathbf{Q}_{BH} \boldsymbol{\Gamma}_H \mathbf{Q}_{RH}^T \mathbf{A}_{RD}^H \tag{45a}$$

$$\mathfrak{h} = \mathrm{vec}\,(\mathbf{H}) = (\mathbf{A}_{RD}^* \otimes \mathbf{A}_{BD})\, \boldsymbol{\gamma}_{HD} \,, \tag{45b}$$

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 (39):

$$\mathbf{Y}' = \mathbf{A}_D\mathbf{Q}'\mathbf{S}' + \mathbf{N}' = \mathbf{A}_D\mathbf{C}' + \mathbf{N}' \,, \tag{46}$$

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 $\mathbf{S}'$. Provided that $M' \geq O(d\log(N_D/d))$, the problem of finding the sparse entries in $\mathbf{C}'$ can be solved by a number of different CS approaches, such as Iterative Hard Thresholding (IHT) [49], Orthogonal Matching Pursuit (OMP) [50], Alternating Direction Method of Multipliers (ADMM) [51], or using AMP-based algorithms [52], [53]. As mentioned above, the accuracy of dictionary-based approaches 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 grid-based AoA estimates is to apply a small angular rotation to each selected column of $\mathbf{A}_D$ that maximizes the correlation with the received data [21]:

$$\boldsymbol{\omega}'_k = \arg\max_{\boldsymbol{\omega}} \left\| \text{diag}\left(\mathbf{a}(\boldsymbol{\omega})\right) \mathbf{a}_{D,k} \mathbf{Y}'^H \right\|_2 \tag{47}$$

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 solve the basis mismatch problem for one-dimensional spatial frequencies is to recast the sparse recovery problem using the atomic norm [22], [54].

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:

$$\mathbf{G} = \mathbf{A}_U(\boldsymbol{\omega}_{UG})\boldsymbol{\Gamma}_G\mathbf{A}_R^H(\boldsymbol{\omega}_{RG}) \tag{48a}$$

$$= \mathbf{A}_{UD}\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G\mathbf{Q}_{RG}^T\mathbf{A}_{RD}^H \tag{48b}$$

$$\mathfrak{g} = \text{vec}\left(\mathbf{G}\right) = \left(\mathbf{A}_{RD}^* \otimes \mathbf{A}_{UD}\right)\boldsymbol{\gamma}_{GD} \,, \tag{48c}$$

where we assume d_G paths with gains $\boldsymbol{\gamma}_G = [\gamma_{G,1} \;\; \cdots \;\; \gamma_{G,d_G}]^T$ and $\boldsymbol{\Gamma}_G = \text{diag}\{\boldsymbol{\gamma}_G\}$. In this case, $\boldsymbol{\gamma}_{GD} = \text{vec}(\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G\mathbf{Q}_{RG}^T)$ 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:

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{h}_c + \mathfrak{n} \,, \tag{49}$$

where

$$\mathbf{Z} = \begin{bmatrix} \boldsymbol{\phi}_1^T \otimes \mathbf{x}_1^T \otimes \mathbf{I}_M \\ \vdots \\ \boldsymbol{\phi}_T^T \otimes \mathbf{x}_T^T \otimes \mathbf{I}_M \end{bmatrix} \tag{50}$$

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 $\mathbf{h}_c = \text{vec}(\mathbf{H}_c) = \text{vec}\left(\mathbf{G}^* \diamond \mathbf{H}\right)$. 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:

$$\mathbf{H}_c = \underbrace{\left(\mathbf{A}_U^*(\boldsymbol{\omega}_{UG}) \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH})\right)}_{MK \times d_H d_G} \underbrace{\boldsymbol{\Gamma}_{GH}}_{d_H d_G \times d_H d_G} \times \underbrace{\left(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\right)}_{d_H d_G \times N} \tag{51a}$$

$$= \underbrace{\left(\mathbf{A}_{UD}^* \otimes \mathbf{A}_{BD}\right)}_{MK \times N_{UD}N_{BD}} \underbrace{\mathbf{Q}_{GH}}_{N_{UD}N_{BD} \times N_{RD}^2} \underbrace{\left(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H\right)}_{N_{RD}^2 \times N} \,, \tag{51b}$$

where

$$\boldsymbol{\Gamma}_{GH} = \boldsymbol{\Gamma}_G^* \otimes \boldsymbol{\Gamma}_H \tag{52a}$$

$$\mathbf{Q}_{GH} = \underbrace{\left(\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G^*\mathbf{Q}_{RG}^T\right)}_{N_{UD} \times N_{RD}} \otimes \underbrace{\left(\mathbf{Q}_{BH}\boldsymbol{\Gamma}_H\mathbf{Q}_{RH}^T\right)}_{N_{BD} \times N_{RD}} \tag{52b}$$

$$= \underbrace{\left(\mathbf{Q}_{UG} \otimes \mathbf{Q}_{BH}\right)}_{N_{UD}N_{BD} \times d_H d_G} \boldsymbol{\Gamma}_{GH} \underbrace{\left(\mathbf{Q}_{RG} \otimes \mathbf{Q}_{RH}\right)^T}_{d_H d_G \times N_{RD}^2} \,. \tag{52c}$$

As in the non-parametric case, not all of the parameters or decompositions shown in (51) are identifiable. To see this, let

$$\boldsymbol{\Lambda} = \text{diag}\left(\mathbf{a}_R(\boldsymbol{\omega}')\right) \tag{53}$$

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

$$\begin{aligned} \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) &= \mathbf{A}_R^T(\boldsymbol{\omega}_{RG})\boldsymbol{\Lambda} \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\boldsymbol{\Lambda}^* \\ &= \mathbf{A}_R^T(\boldsymbol{\omega}'_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}'_{RH}) \end{aligned} \tag{54}$$

where the elements of $\boldsymbol{\omega}'_{RH}$ and $\boldsymbol{\omega}'_{RG}$ are defined as $\boldsymbol{\omega}'_{RH,k} = \boldsymbol{\omega}_{RH,k} + \boldsymbol{\omega}'$ and $\boldsymbol{\omega}'_{RG,k} = \boldsymbol{\omega}_{RG,k} + \boldsymbol{\omega}'$. 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 (51), 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 $\boldsymbol{\gamma}_{H,1} = 1$. 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.

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.

1) Channel Estimation for the General Case

Using (51), the composite channel is given by

$$\mathbf{h}_c = \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\gamma}_{GH} \tag{55a}$$
$$= \mathbf{A}_D \mathbf{q}_{GH} \ , \tag{55b}$$

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

$$\mathbf{A}(\boldsymbol{\omega}) = \Big(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\Big)^T \diamond (\mathbf{A}_U^*(\boldsymbol{\omega}_{UG}) \otimes \mathbf{A}_B(\boldsymbol{\omega}_{BH})) \tag{56a}$$

$$\mathbf{A}_D = \big(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H\big)^T \otimes (\mathbf{A}_{UD}^* \otimes \mathbf{A}_{BD}) \ . \tag{56b}$$

and $\boldsymbol{\omega}^T = \big[\boldsymbol{\omega}_{BH}^T \ \ \boldsymbol{\omega}_{RH}^T \ \ \boldsymbol{\omega}_{RG}^T \ \ \boldsymbol{\omega}_{UG}^T\big]$. Similarly, there are two forms of the general data model in (49):

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{A}(\boldsymbol{\omega})\boldsymbol{\gamma}_{GH} + \mathfrak{n} \tag{57a}$$
$$= \sqrt{P}\mathbf{Z}\mathbf{A}_D\mathbf{q}_{GH} + \mathfrak{n} \ . \tag{57b}$$

At first glance, Eq. (57a) 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

$$[\mathbf{A}(\boldsymbol{\omega})]_{:k} = \big[\mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})\big]_{:k} \otimes \mathbf{a}_U^*(\omega_{UG,\ell}) \otimes \mathbf{a}_B(\omega_{BH,p}), \tag{58}$$

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

$$\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH}) = \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \ . \tag{59}$$

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 $\boldsymbol{\omega}_{RH}$ [21]:

$$\big[\mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})\big]_{:k} = \big[\big(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\big)_{k:}\big]^T \tag{60a}$$
$$= \big[\mathbf{a}_R^T(\boldsymbol{\omega}_{RG,\ell}) \odot \mathbf{a}_R^H(\boldsymbol{\omega}_{RH,p})\big]^T \tag{60b}$$
$$= \mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) \ . \tag{60c}$$

Thus, from (58), we see that the k-th composite steering vector in (57a) is parameterized by different pairs of entries from $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\omega}_{RH}$, and from $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}$:

$$[\mathbf{A}(\boldsymbol{\omega})]_{:k} = \mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) \otimes \mathbf{a}_U^*(\omega_{UG,\ell}) \otimes \mathbf{a}_B(\omega_{BH,p}) \ . \tag{61}$$

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:

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \ \mathfrak{y}^H \mathbf{P}^{\perp}_{\mathbf{ZA}(\boldsymbol{\omega})} \mathfrak{y} \ , \tag{62}$$

where $\mathbf{P}^{\perp}_{\mathbf{ZA}(\boldsymbol{\omega})}$ is the projection orthogonal to the columns of the effective array response $\mathbf{ZA}(\boldsymbol{\omega})$. 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, (57a) simplifies to $\mathbf{h}_c = \gamma \mathbf{a}(\boldsymbol{\omega})$, where

$$\mathbf{a}(\boldsymbol{\omega}) = \mathbf{a}_R(\boldsymbol{\omega}_{RG}) \otimes \mathbf{a}_U^*(\omega_{UG}) \otimes \mathbf{a}_B(\omega_{BH}) \ , \tag{63}$$

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 $\mathfrak{y}^H \mathbf{P}^{\perp}_{\mathbf{Za}(\boldsymbol{\omega})} \mathfrak{y}$, and in either case setting $\hat{\gamma} = (\mathbf{Z}\mathbf{a}(\hat{\boldsymbol{\omega}}))^{\dagger} \mathfrak{y}/\sqrt{P}$. 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 (57a) and (61) as

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{A}(\boldsymbol{\omega}')\boldsymbol{\gamma} + \mathfrak{n} \tag{64a}$$
$$\mathbf{A}(\boldsymbol{\omega}') = \big[\mathbf{a}(\boldsymbol{\omega}'_1) \ \ \cdots \ \ \mathbf{a}(\boldsymbol{\omega}'_{d_H d_G})\big] \tag{64b}$$
$$\mathbf{a}(\boldsymbol{\omega}'_k) = \mathbf{a}_R(\boldsymbol{\omega}'_{k,1}) \otimes \mathbf{a}_U^*(\boldsymbol{\omega}'_{k,2}) \otimes \mathbf{a}_B(\boldsymbol{\omega}'_{k,3}) \ , \tag{64c}$$

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, $\hat{\boldsymbol{\gamma}}' = \big(\mathbf{Z}\mathbf{A}(\hat{\boldsymbol{\omega}}')\big)^{\dagger} \mathfrak{y}/\sqrt{P}$.

CSI estimation for the dictionary-based model in (57b) also requires an unconventional approach, due to the sparsity pattern in $\mathbf{q}_{GH}$ [15], [21], [23]. 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 (56b). 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 via Eq. (52b). The matrix $\mathbf{Q}_1 = \big(\mathbf{Q}_{UG}\boldsymbol{\Gamma}_G^*\mathbf{Q}_{RG}^T\big)$ has d_G non-zero elements at

[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 UPA). 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.

the row/column coordinates corresponding to the AoA/AoD pairs of $\mathbf{G}$, while $\mathbf{Q}_2 = \left(\mathbf{Q}_{BH}\mathbf{\Gamma}_H\mathbf{Q}_{RH}^T\right)$ 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 depends only on $d_H + d_G - 1$ unique complex values. While cumbersome, imposing the required structured sparsity constraint on $\mathbf{q}_{GH}$ is possible [15], [23].

A bigger issue with the CS approach in the general case is the size of the dictionary in (57b), which has $N_{BD}N_{UD}N_{RD}^2$ elements. A smaller dimensional problem can be formulated based on the parameterization in (64), such that

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{A}'_D\mathbf{q}' + \mathfrak{n}\ , \tag{65}$$

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

$$\mathbf{Z}\mathbf{A}'_D = \mathbf{Z}\left(\mathbf{A}_{RD} \otimes \mathbf{A}^*_{UD} \otimes \mathbf{A}_{BD}\right)\ . \tag{66}$$

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 [22], [25]. 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 $\boldsymbol{\phi}$, which results in the following received signal at the BS:

$$\begin{aligned}\mathbf{Y}_1 &= \sqrt{P}\mathbf{H}\bar{\mathbf{\Phi}}\mathbf{G}^H\mathbf{X}_1 + \mathbf{N}_1 && \text{(67a)}\\ &= \sqrt{P}\mathbf{A}_B(\boldsymbol{\omega}_{BH})\mathbf{\Xi}\mathbf{A}_U^H(\boldsymbol{\omega}_{UG})\mathbf{X}_1 + \mathbf{N}_1 && \text{(67b)}\\ &= \sqrt{P}\mathbf{A}_{BD}\mathbf{Q}_{BH}\mathbf{\Xi}\mathbf{Q}_{UG}^T\mathbf{A}_{UD}^H\mathbf{X}_1 + \mathbf{N}_1\ , && \text{(67c)}\end{aligned}$$

where

$$\mathbf{\Xi} = \mathbf{\Gamma}_H\mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\mathbf{\Phi}\mathbf{A}_R(\boldsymbol{\omega}_{RG})\mathbf{\Gamma}_G^*\ . \tag{68}$$

We see immediately that (67b)-(67c) are in the form of a standard AoA estimation problem as in (39), and as long as $M \geq \max\{2d_H - T_1, d_H + 1\}$ [43], $\boldsymbol{\omega}_{BH}$ could be estimated from $\mathbf{Y}_1$ using an AoA estimation algorithm[5]. Similarly, estimation of $\boldsymbol{\omega}_{UG}$ can also be performed separately based on the following equation, provided that $K \geq d_G + 1$ since presumably $M > d_G$:

$$\begin{aligned}\frac{\mathbf{X}_1\mathbf{Y}_1^H}{T_1\sqrt{P}} &= \mathbf{A}_U(\boldsymbol{\omega}_{UG})\mathbf{\Xi}^H\mathbf{A}_B^H(\boldsymbol{\omega}_{BH}) + \frac{\mathbf{X}_1\mathbf{N}_1^H}{T_1\sqrt{P}} && \text{(69a)}\\ &= \mathbf{A}_U(\boldsymbol{\omega}_{UG})\mathbf{S} + \frac{1}{T_1\sqrt{P}}\mathbf{X}_1\mathbf{N}_1^H\ . && \text{(69b)}\end{aligned}$$

For the second stage, the UE repeatedly sends T_2/d_G blocks of the pilot signal matrix $\mathbf{X}_2 \in \mathbb{C}^{K\times d_G}$ during T_2 time slots, while the reflection shift $\mathbf{\Phi}_b$ varies from block to block. The observed signal in the b-th pilot block ($b = 1, ..., T_2/d_G$) is given by

$$\mathbf{Y}_b = \sqrt{P}\mathbf{H}\mathbf{\Phi}_b\mathbf{G}^H\mathbf{X}_2 + \mathbf{N}_{2,b}. \tag{70}$$

With sufficiently accurate estimates of $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}$ from Stage 1, we have

$$\begin{aligned}\mathbf{Y}_{\mathrm{B},b} &= \frac{1}{\sqrt{P}}\mathbf{A}_B^\dagger(\hat{\boldsymbol{\omega}}_{BH})\mathbf{Y}_b\mathbf{B}^\dagger\\ &\simeq \mathbf{\Gamma}_H\mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\mathbf{\Phi}_b\mathbf{A}_R(\boldsymbol{\omega}_{RG})\mathbf{\Gamma}_G^* + \mathbf{N}_{\mathrm{B},b}, && \text{(71)}\end{aligned}$$

where $\mathbf{N}_{\mathrm{B},b} = \frac{1}{\sqrt{P}}\mathbf{A}_B^\dagger(\hat{\boldsymbol{\omega}}_{BH})\mathbf{N}_{2,b}\left(\mathbf{A}_U^H(\hat{\boldsymbol{\omega}}_{UG})\mathbf{X}_2\right)$ and $\mathbf{B} = \mathbf{A}_U^H(\boldsymbol{\omega}_{UG})\mathbf{X}_2$. The $b = 1, \cdots, T_2/d_G$ blocks defined in (71) are gathered together to form

$$\begin{aligned}\mathbf{Y}_{\mathrm{B}} =&[\mathrm{vec}(\mathbf{Y}_{\mathrm{B},1}), ..., \mathrm{vec}(\mathbf{Y}_{\mathrm{B},\frac{T_2}{d_G}})]^{\mathrm{T}} && \text{(72a)}\\ =&\mathbf{\Psi}^*\underbrace{\left(\mathbf{A}_R^T(\boldsymbol{\omega}_{RG}) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\right)^T}_{\mathbf{C}^T(\boldsymbol{\omega}_{RG},\boldsymbol{\omega}_{RH})}\mathbf{\Gamma}_{GH} + \mathbf{N}_{\mathrm{B}} && \text{(72b)}\\ =&\mathbf{\Psi}^*\left(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H\right)^T(\mathbf{Q}_{RG} \otimes \mathbf{Q}_{RH})\mathbf{\Gamma}_{GH} + \mathbf{N}_{\mathrm{B}}, && \text{(72c)}\end{aligned}$$

where $\mathbf{\Psi}^H = [\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_{\frac{T_2}{d_G}}]$, $\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$ is defined in (59). and $\mathbf{N}_{\mathrm{B}}$ is defined similarly to $\mathbf{Y}_{\mathrm{B}}$.

A sparse estimation problem could be set up for $\boldsymbol{\omega}_{RG}$ and $\boldsymbol{\omega}_{RH}$ based on vectorizing (72c), and while the resulting dictionary is significantly smaller than in (57b), it still has N_{RD}^2 elements. Instead, a much simpler solution can be found by noting the structure of $\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$ described in (60), which indicates that each of the $d_G d_H$ columns of (72b) 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,

$$\mathbf{y}_{B,k} \simeq \gamma_k\mathbf{\Psi}^*\mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) + \mathbf{n}_{B,k}\ , \tag{73}$$

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 (70) 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 $\mathbf{c}_k^T$ represents the k-th row of $\mathbf{\Gamma}_{GH}\mathbf{C}(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})$, then $\mathbf{c}_k^T$ 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 $\mathbf{Y}_B$:

$$\hat{\mathbf{c}}_k = \hat{\gamma}_k\mathbf{a}_R(\hat{\boldsymbol{\omega}}_k)\ . \tag{74}$$

[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 [5], [8], [17], [55].

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 approaches can be employed. We will focus on the time-domain approach here as it is a bit more straightforward (see [17] for methods that focus on estimating the individual subchannel responses). Our starting point is (26), which differs from (57) 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 $\mathfrak{h}_c$:

$$\mathfrak{Z} = [\mathfrak{Z}(0) \;\; \mathfrak{Z}(1) \;\; \cdots \;\; \mathfrak{Z}(L-1)] \; . \tag{75}$$

Using (55), we can write

$$\mathfrak{y} = \sqrt{P}\,\mathfrak{A}(\mathbf{\Omega}) \begin{bmatrix} \boldsymbol{\gamma}_{GH}(0) \\ \vdots \\ \boldsymbol{\gamma}_{GH}(L-1) \end{bmatrix} + \mathfrak{n} \tag{76a}$$

$$= \sqrt{P}\,\mathfrak{A}_D \begin{bmatrix} \mathbf{q}_{GH}(0) \\ \vdots \\ \mathbf{q}_{GH}(L-1) \end{bmatrix} + \mathfrak{n} \, , \tag{76b}$$

where

$$\mathfrak{A}(\mathbf{\Omega}) = \Big[\mathfrak{Z}(0)\mathbf{A}(\boldsymbol{\omega}(0)) \;\; \cdots \;\; \mathfrak{Z}(L-1)\mathbf{A}(\boldsymbol{\omega}(L-1))\Big] \tag{77a}$$

$$\mathfrak{A}_D = \Big[\mathfrak{Z}(0)\mathbf{A}_D \;\; \cdots \;\; \mathfrak{Z}(L-1)\mathbf{A}_D\Big] \, , \tag{77b}$$

$\mathbf{\Omega} = \big[\boldsymbol{\omega}(0)^T \;\; \cdots \;\; \boldsymbol{\omega}^T(L-1)\big]^T$, 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 (76) that the full-scale parameterization of the problem is similar to that in (57), 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 $\boldsymbol{\phi}$. Using (24), 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:

$$\mathbf{Y}_{1,t} = \sqrt{P}\Big[\mathbf{H}(0)\mathbf{\Phi}\mathbf{G}^H(0) \;\; \cdots \;\; \mathbf{H}(L-1)\mathbf{\Phi}\mathbf{G}^H(L-1)\Big] \times \mathbf{X}_{1,t} + \mathbf{N}_{1,t} \tag{78a}$$

$$= \sqrt{P}\mathbf{A}_B(\mathbf{\Omega}_{BH})\mathbf{\Pi}\mathbf{X}_{1,t} + \mathbf{N}_{1,t} \tag{78b}$$

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

$$\mathbf{A}_B(\mathbf{\Omega}_{BH}) = \Big[\mathbf{A}_B(\boldsymbol{\omega}_{BH}(0) \;\; \cdots \;\; \mathbf{A}_B(\boldsymbol{\omega}_{BH}(L-1)\Big] \tag{79a}$$

$$\mathbf{\Pi} = \text{blkdiag}\Big(\big\{\mathbf{\Xi}(k)\mathbf{A}_U^H(\boldsymbol{\omega}_{UG}(k))\big\}_{k=0}^{L-1}\Big) \, , \tag{79b}$$

where $\mathbf{\Xi}(k)$ is the matrix corresponding to (68) for tap k. Concatenating data from the T_{o1} OFDM symbols yields

$$\mathbf{Y}_1 = \sqrt{P}\,[\mathbf{Y}_{1,1} \;\; \cdots \;\; \mathbf{Y}_{1,T_{o1}}] \tag{80a}$$

$$= \sqrt{P}\,\underbrace{\mathbf{A}_B(\mathbf{\Omega}_{BH})}_{M \times Ld_H} \underbrace{\mathbf{\Pi}\mathbf{X}_1}_{Ld_H \times T_1} + \mathbf{N}_1 \, , \tag{80b}$$

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

$$\frac{(\mathbf{Y}_1\mathbf{X}_1^\dagger)^H}{\sqrt{P}T_1} = \begin{bmatrix} \mathbf{D}(0) \\ \vdots \\ \mathbf{D}(L-1) \end{bmatrix} + \frac{(\mathbf{N}_1\mathbf{X}_1^\dagger)^H}{\sqrt{P}T_1} \, , \tag{81}$$

where $\mathbf{D}(k) = \mathbf{A}_U(\boldsymbol{\omega}_{UG}(k))\mathbf{\Xi}(k)\mathbf{A}_B^H(\boldsymbol{\omega}_{BH}(k))$. Assuming $K > d_G$, the UE angles are found by solving AoA estimation problems on the L blocks in (81).

In Stage 2, T_{o2} additional OFDM training symbols are transmitted, for a total of $T_2 = T_{o2}N_c$ samples. As in (70), the received signal in the b-th symbol block ($b = 1, ..., T_{o2}/Ld_G$) is given by

$$\mathbf{Y}_{b,s} = \sqrt{P}\sum_{k=0}^{L-1} \mathbf{H}(k)\mathbf{\Phi}_{b,s}\mathbf{G}^H(k)\mathbf{X}_{s-k} + \mathbf{N}_{t,s}$$

$$= \mathbf{A}_B(\mathbf{\Omega}_{BH})\text{blkdiag}\left(\{\mathbf{V}_k\}_{k=0}^{L-1}\right)\mathbf{B}_s^H + \mathbf{N}_{t,s} \tag{82a}$$

$$\mathbf{V}_k = \mathbf{\Gamma}_H(k)\mathbf{A}_R^H(\boldsymbol{\omega}_{RH}(k))\mathbf{\Phi}_{b,s}\mathbf{A}_R(\boldsymbol{\omega}_{RG}(k))\mathbf{\Gamma}_G^*(k) \tag{82b}$$

$$\mathbf{B}_s = \sqrt{P}[\mathbf{X}_{2,s}^H\mathbf{A}_U(\boldsymbol{\omega}_{UG}(0)) \;\; \cdots \;\; \mathbf{X}_{2,s-L+1}^H\mathbf{A}_U(\boldsymbol{\omega}_{UG}(L-1))] \, , , \tag{82c}$$

where $\mathbf{X}_{s-k}$ is $K \times Ld_G$. Replacing $\boldsymbol{\omega}_{UG}(k)$ and $\boldsymbol{\omega}_{BH}(k)$ with their estimates from Stage 1, and assuming $M \geq Ld_H$ and $K \geq d_G$, we multiply each $\mathbf{Y}_{b,s}$ on the left by the estimate of the pseudo-inverse of $\mathbf{A}_B(\mathbf{\Omega}_{BH})$ and on the right by the estimate of the pseudo-inverse of $\mathbf{B}_s^H$, and we stack the vectorized $\mathbf{V}_k$ together as

$$\mathbf{y}_{B,b,s} = \begin{bmatrix} \text{vec}(\mathbf{V}_0) \\ \vdots \\ \text{vec}(\mathbf{V}_{L-1}) \end{bmatrix} + \mathbf{n}_{B,b,s}$$

$$= \begin{bmatrix} \mathbf{\Gamma}_{GH}(0)\mathbf{C}(0)\boldsymbol{\phi}_{b,s} \\ \vdots \\ \mathbf{\Gamma}_{GH}(L-1)\mathbf{C}(L-1)\boldsymbol{\phi}_{b,s-L+1} \end{bmatrix} + \mathbf{n}_{B,b,s}, \tag{83}$$

where $\mathbf{C}(k) = \mathbf{C}(\boldsymbol{\omega}_{RG}(k), \boldsymbol{\omega}_{RH}(k))$. We stack the N_c vectors defined in (83) together and take the transpose as in (72):

$$\underbrace{\mathbf{Y}_{B,b}}_{N_c \times Ld_Hd_G} = [\underbrace{\mathbf{Y}_{B,b,0}}_{N_c \times d_Hd_G} \;\; \cdots \;\; \mathbf{Y}_{B,b,L-1}]$$

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

$$\mathbf{\Psi}_b^* = \big[\mathbf{\Psi}_{b,0}^* \;\; \cdots \;\; \mathbf{\Psi}_{b,L-1}^*\big] \, ,$$

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

$$\bar{\mathbf{Y}}_{B,k} \simeq \begin{bmatrix} \mathbf{\Psi}_{1,k}^* \\ \vdots \\ \mathbf{\Psi}_{N_c,k}^* \end{bmatrix} \mathbf{C}^T(k)\mathbf{\Gamma}_{GH}(k) + \mathbf{N}_{B,k} \ , \tag{85}$$

where $\bar{\mathbf{Y}}_{B,k}$ holds columns $k+1$ through $k+L$ of $\mathbf{Y}_B$. This equation is equivalent in form to (72b), 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 (57a) still has $d_H d_G$ columns, now given by

$$[\mathbf{A}(\boldsymbol{\omega})]_{:k} = \mathbf{a}_R(\boldsymbol{\omega}_{RG,\ell} - \boldsymbol{\omega}_{RH,p}) \otimes \mathbf{a}_B(\omega_{BH,p}) \ , \tag{86}$$

where $\ell = \lfloor k/d_H \rfloor$ and $p = \text{mod}_{d_H}(k)$. We assume without loss of generality that $x_t = 1, \forall\, t$, so (50) simplifies to

$$\mathbf{Z} = \begin{bmatrix} \boldsymbol{\phi}_1^T \otimes \mathbf{I}_M \\ \vdots \\ \boldsymbol{\phi}_T^T \otimes \mathbf{I}_M \end{bmatrix} \ . \tag{87}$$

Ignoring the structure of $\boldsymbol{\gamma}_{GH}$ in (57a), the DML criterion in (62) 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 (64), 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 (57b) for single-antenna UEs can also be approached in two ways. The first retains the full geometric structure of the channel, with

$$\mathbf{A}_D = \left(\mathbf{A}_{RD}^T \diamond \mathbf{A}_{RD}^H\right)^T \otimes \mathbf{A}_{BD} \tag{88a}$$

$$\mathbf{Q}_{GH} = \underbrace{\left(\boldsymbol{\gamma}_G^H \mathbf{Q}_{RG}^T\right)}_{1 \times N_{RD}} \otimes \underbrace{\left(\mathbf{Q}_{BH}\mathbf{\Gamma}_H \mathbf{Q}_{RH}^T\right)}_{N_{BD} \times N_{RD}} \ . \tag{88b}$$

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 (65), 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

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

$$\mathbf{Y}_B = \frac{1}{\sqrt{P}} \left[\mathbf{A}_B^\dagger(\hat{\boldsymbol{\omega}}_{BH})\mathbf{Y}_1\right]^T \tag{89a}$$

$$\simeq \mathbf{\Psi}^* \mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH}) \left(\boldsymbol{\gamma}_G^* \otimes \mathbf{\Gamma}_H\right) + \mathbf{N}_B \ , \tag{89b}$$

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

$$\mathbf{y}_{B,k} \simeq \gamma_{H,k}\mathbf{\Psi}^* \sum_{n=1}^{d_G} \boldsymbol{\gamma}_{G,n}^* \mathbf{a}_R(\boldsymbol{\omega}_{RG,n} - \boldsymbol{\omega}_{RH,k}) + \mathbf{n}_{B,k} \tag{90a}$$

$$\simeq \gamma_{H,K}\mathbf{\Psi}^* \text{diag}\Big(\mathbf{A}_R(\boldsymbol{\omega}_{RG})\boldsymbol{\gamma}_G^*\Big)\mathbf{a}_R(-\boldsymbol{\omega}_{RH,k}) + \mathbf{n}_{B,k}, \tag{90b}$$

This special structure allows for a simpler solution than that required in the multi-antenna UE case considered in Section IV-C2 [21]. 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

$$\mathbf{y}_{B,1} \simeq \mathbf{\Psi}^* \mathbf{A}_R(\boldsymbol{\omega}_{RG})\boldsymbol{\gamma}_G^* + \mathbf{n}_{B,1} \ . \tag{91}$$

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 algorithm such as beamforming or DML, or using a d_G-sparse CS approach. Once estimated, these parameters can be substituted into (90), 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, for Stage 2, 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

$$\mathbf{y} = \mathbf{\Psi}^* \mathbf{C}^T(\boldsymbol{\omega}_{RG}, \boldsymbol{\omega}_{RH})(\boldsymbol{\gamma}_G^* \otimes \boldsymbol{\gamma}_H) + \mathbf{n} \tag{92a}$$

$$= \mathbf{\Psi}^* \sum_{k=1}^{d_G} \sum_{n=1}^{d_H} \gamma_{G,k}^* \gamma_{H,n} \mathbf{a}_R\left(\boldsymbol{\omega}_{RG,k} - \boldsymbol{\omega}_{RH,n}\right) + \mathbf{n} \ . \tag{92b}$$

This leads to an equivalent single-snapshot AoA estimation problem, although with a nonlinear dependence on the gains via the Kronecker product. 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:

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_U(\boldsymbol{\omega}_{UG}^1)\mathbf{\Gamma}_{G^1}\mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^1) \\ \vdots \\ \mathbf{A}_U(\boldsymbol{\omega}_{UG}^U)\mathbf{\Gamma}_{G^U}\mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^U) \end{bmatrix} \equiv \begin{bmatrix} \mathbf{G}^1 \\ \vdots \\ \mathbf{G}^U \end{bmatrix} \ , \tag{93}$$

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

$$\mathbf{H}_c = \begin{bmatrix} \mathbf{G}^1 \diamond \mathbf{H} \\ \vdots \\ \mathbf{G}^U \diamond \mathbf{H} \end{bmatrix} \tag{94a}$$

$$\mathbf{h}_c = \begin{bmatrix} \mathbf{h}_c^1 \\ \vdots \\ \mathbf{h}_c^U \end{bmatrix} = \mathbf{J}\text{vec}(\mathbf{H}_c) \ , \tag{94b}$$

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 (57):

$$\mathfrak{y} = \sqrt{P}\mathbf{Z}\mathbf{J} \begin{bmatrix} \mathbf{A}(\boldsymbol{\omega}^1)\boldsymbol{\gamma}_{G^1H} \\ \vdots \\ \mathbf{A}(\boldsymbol{\omega}^U)\boldsymbol{\gamma}_{G^UH} \end{bmatrix} + \mathfrak{n} \tag{95a}$$

$$= \sqrt{P}\underbrace{\mathbf{Z}\left[\mathbf{J}_1\mathbf{A}(\boldsymbol{\omega}^1) \ \cdots \ \mathbf{J}_U\mathbf{A}(\boldsymbol{\omega}^U)\right]}_{\mathbf{A}(\boldsymbol{\omega})} \underbrace{\begin{bmatrix} \boldsymbol{\gamma}_{G^1H} \\ \vdots \\ \boldsymbol{\gamma}_{G^UH} \end{bmatrix}}_{\boldsymbol{\gamma}_{GH}} + \mathfrak{n} \tag{95b}$$

$$= \sqrt{P}\underbrace{\mathbf{Z}\left[\mathbf{J}_1\mathbf{A}_D \ \cdots \ \mathbf{J}_U\mathbf{A}_D\right]}_{\text{effective dictionary}} \underbrace{\begin{bmatrix} \mathbf{q}_{G^1H} \\ \vdots \\ \mathbf{q}_{G^UH} \end{bmatrix}}_{\mathbf{q}_{GH}} + \mathfrak{n}, \tag{95c}$$

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 (55). 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 (69):

$$\frac{\mathbf{X}_1\mathbf{Y}_1^H}{T_1\sqrt{P}} = \begin{bmatrix} \mathbf{A}_U(\boldsymbol{\omega}_{UG}^1) & & \\ & \ddots & \\ & & \mathbf{A}_U(\boldsymbol{\omega}_{UG}^U) \end{bmatrix} \mathbf{S} + \frac{\mathbf{X}_1\mathbf{N}_1^H}{T_1\sqrt{P}} \tag{96}$$

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

$$\boldsymbol{\Xi}^H = \boldsymbol{\Gamma}_G \begin{bmatrix} \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^1) \\ \vdots \\ \mathbf{A}_R^H(\boldsymbol{\omega}_{RG}^U) \end{bmatrix} \boldsymbol{\Phi}^*\mathbf{A}_R(\boldsymbol{\omega}_{RH})\boldsymbol{\Gamma}_H^* \ . \tag{97}$$

Assuming $K_u > d_{G^u}$, UE AoD estimates can be found separately for each user by processing different block rows of (96). Once estimates of $\boldsymbol{\omega}_{BH}$ and $\boldsymbol{\omega}_{UG}^u$ are determined, we collect additional training data and proceed as in (71) where

$$\mathbf{B} = \text{blkdiag}\left(\mathbf{A}_U^H(\boldsymbol{\omega}_{UG}^u)_{u=1}^U\right)\mathbf{X}_2. \tag{98}$$

Stacking the vectorized block columns of (71) together, we have

$$\begin{aligned}\mathbf{y}_{B,b} =& \begin{bmatrix} \text{vec}(\boldsymbol{\Gamma}_H\mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\boldsymbol{\Phi}_b\mathbf{A}_R(\boldsymbol{\omega}_{RG}^1)\boldsymbol{\Gamma}_G^{1,*}) \\ \vdots \\ \text{vec}(\boldsymbol{\Gamma}_H\mathbf{A}_R^H(\boldsymbol{\omega}_{RH})\boldsymbol{\Phi}_b\mathbf{A}_R(\boldsymbol{\omega}_{RG}^U)\boldsymbol{\Gamma}_G^{U,*}) \end{bmatrix} + \mathbf{n}_{B,b} \\ =& \boldsymbol{\Gamma}_{GH} \underbrace{\begin{bmatrix} \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}^1) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \\ \vdots \\ \mathbf{A}_R^T(\boldsymbol{\omega}_{RG}^U) \diamond \mathbf{A}_R^H(\boldsymbol{\omega}_{RH}) \end{bmatrix}}_{\mathbf{C}\left(\boldsymbol{\omega}_{RG}^1, \cdots, \boldsymbol{\omega}_{RG}^U, \boldsymbol{\omega}_{RH}\right)} \boldsymbol{\phi}_b + \mathbf{n}_{B,b} \ . \end{aligned} \tag{99}$$

Collecting the vectors from (99) together as columns of a matrix as in (72) leads to

$$\mathbf{Y}_B \simeq \boldsymbol{\Psi}^*\mathbf{C}\left(\boldsymbol{\omega}_{RG}^1, \cdots, \boldsymbol{\omega}_{RG}^U, \boldsymbol{\omega}_{RH}\right)\boldsymbol{\Gamma}_{GH} + \mathbf{N}_B \ . \tag{100}$$

As before, each column of (100) 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 [21]:

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 (69) if there is only one user with multiple antennas, or as in (96) 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

$$\mathbf{Y} = \mathbf{H}\bar{\boldsymbol{\Phi}}\mathbf{A}_R(\boldsymbol{\omega}_{RG})\boldsymbol{\Gamma}_G\mathbf{A}_U^H(\boldsymbol{\omega}_{UG})\mathbf{X} + \mathbf{N} \ . \tag{101}$$

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 [21]. 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 (57), (76), or (95) can be estimated using least-squares; *e.g.,* for (57a) we have

$$\hat{\boldsymbol{\gamma}}_{GH} = \frac{1}{\sqrt{P}}\Big(\mathbf{Z}\mathbf{A}(\hat{\boldsymbol{\omega}})\Big)^{\dagger}\mathfrak{y}\,. \tag{102}$$

V. Unstructured vs. Geometric Models: The Trade-Offs

The choice of channel model and estimation algorithm for RIS-based systems depends on many factors including estimation accuracy, training overhead, algorithm complexity, and robustness to the modeling assumptions. We will examine these factors in this section, and then focus on estimation performance for some specific scenarios in Section VI. The discussion below in this section will assume the case of either a single K-antenna UE or K single-antenna UEs.

A. Estimation Accuracy

Rather than study the performance of individual algorithms, of which there are too many examples to consider comprehensively in this paper, we can get a sense for what is achievable by examining performance bounds based just on the models themselves. The most common example is the CRB of the composite channel $\breve{\mathbf{h}}_c^T = \big[\mathrm{Re}(\mathbf{h}_c)^T \;\; \mathrm{Im}(\mathbf{h}_c)^T\big]$, where $\mathrm{Re}(\mathbf{h}_c)$ and $\mathrm{Im}(\mathbf{h}_c)$ respectively represent the real and imaginary parts of $\mathbf{h}_c$.

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

$$f_u(\breve{\mathbf{h}}_c) = -MT\ln(\pi\sigma^2) - \frac{1}{\sigma^2}\left\|\mathfrak{y} - \sqrt{P}\mathbf{Z}\mathbf{h}_c\right\|^2\,, \tag{103}$$

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

$$\mathrm{CRB}_u(\breve{\mathbf{h}}_c) = \mathrm{FIM}_u^{-1}(\breve{\mathbf{h}}_c) \tag{104}$$

$$\mathrm{FIM}_u(\breve{\mathbf{h}}_c) = \mathbb{E}\left\{\frac{\partial f_u(\breve{\mathbf{h}}_c)}{\partial \breve{\mathbf{h}}_c}\left(\frac{\partial f_u(\breve{\mathbf{h}}_c)}{\partial \breve{\mathbf{h}}_c}\right)^T\right\}\,, \tag{105}$$

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

$$\mathrm{CRB}_u(\breve{\mathbf{h}}_c) = \frac{\sigma^2}{2P}\left(\breve{\mathbf{Z}}^T\breve{\mathbf{Z}}\right)^{-1}\,, \tag{106}$$

where

$$\breve{\mathbf{Z}} = \begin{bmatrix} \mathrm{Re}(\mathbf{Z}) & -\mathrm{Im}(\mathbf{Z}) \\ \mathrm{Im}(\mathbf{Z}) & \mathrm{Re}(\mathbf{Z}) \end{bmatrix} \tag{107}$$

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

$$\mathrm{CRB}_u(\breve{\mathbf{h}}_c) = \frac{\sigma^2}{2PT}\mathbf{I}_{2MK(N+1)}\,, \tag{108}$$

signifying that the lower bound is identical for every element of $\breve{\mathbf{h}}_c$ (note that we have assumed the RIS element gains are $\beta = 1$). An important observation is 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, and is of much smaller dimension than for the unstructured model when the channels are dominated by a small number of propagation paths. The CRB for the structured $\boldsymbol{\eta}$ is more difficult to compute, requiring a large number of tedious derivative calculations which we do not include here (for an example of such derivations, see [19] for a detailed derivation for the special case of a single-antenna BS and UE). When one forms an estimate of the composite channel $\mathbf{h}_c$ from the estimate of the parameters in $\boldsymbol{\eta}$, the resulting CRB for the channel is given by

$$\mathrm{CRB}_s(\breve{\mathbf{h}}_c) = \frac{\partial \breve{\mathbf{h}}_c}{\partial \boldsymbol{\eta}}\mathrm{CRB}(\boldsymbol{\eta})\left(\frac{\partial \breve{\mathbf{h}}_c}{\partial \boldsymbol{\eta}}\right)^T\,, \tag{109}$$

where here the subscript s is for "structured." Simplification of (109) for the general RIS-based model in (57a) is still an open problem, but some general observations based on studies from traditional AoA estimation can be made (*e.g.,* see [56]). In particular, like the unstructured case, $\mathrm{CRB}(\boldsymbol{\eta})$ and hence $\mathrm{CRB}_s(\breve{\mathbf{h}}_c)$ are inversely proportional to T and the SNR P/σ^2. However, unlike the unstructured case, whose performance cannot be improved by increasing the number of antennas (except in the sense that T must increase with N to maintain identifiability), the geometric CRB will improve as either M or N increases. For example, the CRB for estimation of the AoAs at the BS decreases proportional to M^3. In Section VI an example will be given to show the decrease with N. The examples in Section VI all show that the geometric model provides a much smaller channel estimation error for all of the scenarios considered. However, as discussed below in Section V-C, some of this performance advantage will inevitably be lost in practice due to modeling errors.

For large arrays such as those considered in RIS-based systems, the CRB is achievable by standard AoA estimation algorithms under the ideal model, even for small (non-asymptotic) values for T [44]. Such algorithms estimate the spatial frequencies as continuous variables, unlike CS-based methods that employ a dictionary on a finite grid. Since the density of the grid cannot be made arbitrarily fine due to the resulting computational complexity and ill-conditioning, for CS-based AoA estimation algorithms the grid spacing becomes the limiting factor rather than the noise, and thus the CRB is an optimistic performance indicator. The examples in Section VI will demonstrate that even with the loss due to the use of

an AoA grid, the geometric model still provides a sizable performance advantage.

B. Training Overhead

When considering the amount of training required for a given model, there are two aspects to consider: the minimum number of samples required for unique identification of the channel model, and the number of samples needed to obtain some desired level of performance. In many cases, these two quantities may be quite different. Some performance-related issues were discussed above and will be illustrated in Section VI, so here we focus on the uniqueness of the models. A more comprehensive analysis of the impact of the training would account for its impact on the system throughput, both in terms of time lost for data transmission and the impact of the resulting channel estimation errors on the achievable rate (see [57] for a recent study that addresses this question).

We know from our discussion in Section III-A1 that identifiability of the unstructured model requires $T \geq K(N+1)$ pilot symbols. Establishing a similar condition for the general geometric model in (57) is difficult since it amounts to a single-snapshot multi-dimensional frequency estimation problem, with a complicated inter-relationship between the frequency dimensions. A simple dimensionality analysis gives the necessary condition that the number of observations MT must exceed the number of parameters to be estimated, which is $5d_H + 5d_G - 4$. For typical values of d_H and d_G assumed for sparse channels and a large BS array, this condition could be satisfied for T as small as 1 or 2. However, the complexity of directly using (57) for estimating the channel parameters is prohibitive due to the non-convex optimization required for $\boldsymbol{\omega}$ based on (57a) using classical AoA estimation methods, or the huge dictionary needed by CS methods for the model in (57b).

Instead, consider the simpler decoupled approach discussed in Section IV-C2, which assumes the more stringent case for identifiability with a single K-antenna UE where the channel depends on the d_G AoDs at the UE. This approach assumes $K \geq d_G + 1$ and a training interval divided into two parts: $T = T_1 + T_2$. For large enough M, the first stage requires at least $T_1 \geq K$ pilot symbols to estimate the BS AoAs and the UE AoDs, while the second stage requires that $T_2/d_G \geq 4$ in order to uniquely estimate the 2D angle and gain for each of the $d_H d_G$ columns defined in (73) [58]. Together, this implies that a minimum of $T = 5K = 5d_G + 1$ pilots are required to identify the composite channel. Note that for the case where there are K single-antenna UEs rather than one K-antenna UE, the condition $K \geq d_G + 1$ can be eliminated since estimation of $\boldsymbol{\omega}_{UG}$ is no longer required. In this case, the minimum number of training samples is $T = K + 4d_G$. In either case, the training overhead is considerably less than for the unstructured model.

In addition to the identifiability conditions that must be satisfied for unique AoA/AoD estimation, CS-based methods must also satisfy conditions that will guarantee with high probability the existence of the desired sparse factorization. Unfortunately, such conditions are not possible to specify precisely. For the decoupled approach with a single K-antenna UE, sparse estimation of $\mathbf{Q}_{BH}$ in (67c) requires $M \geq O(d_H \log(N_{BD}/d_H))$, while estimation of $\mathbf{Q}_{UG}$ in (67c) and (69) requires $K \geq O(d_G \log(N_{UD}/d_G))$, neither with a specific requirement for T_1. Stage 2, on the other hand, requires $T_2 \geq d_G O(\log(N_{RD}))$ to find a sparse factorization for the 1-sparse problems in (69), which is still likely much smaller than for the unstructured model. As before, the condition $K \geq O(d_G \log(N_{UD}/d_G))$ can be dropped when the scenario involves K single-antenna users.

C. Robustness

While geometric methods typically enjoy an advantage in terms of training overhead, their reliance on a more detailed parameterized model makes them subject to performance degradation when the actual system deviates from the assumed model. In particular, the geometric model requires that one determine the number of propagation paths in each link, and relies on prior knowledge of the array response or calibration in order to determine the AoAs and AoDs. Errors in determining the model order or inevitable deviations of the array response from the nominal model will lead to a corresponding performance loss. To get a sense for the magnitude of this loss compared to that due to background noise, consider the following simple model involving a single source with spatial frequency ω that transmits T pilot symbols:

$$\mathbf{Y} = \sqrt{P}\,(\mathbf{a}(\omega) + \tilde{\mathbf{a}})\,\mathbf{x} + \mathbf{N}\,, \tag{110}$$

where $\mathbf{Y}$ and $\mathbf{N}$ are $M \times T$ matrices containing the received signal and noise, respectively, $\mathbf{x}$ is a $1 \times T$ vector of pilot data, and $\tilde{\mathbf{a}}$ is a vector that represents the deviation of the $M \times 1$ array response from the nominal model $\mathbf{a}(\omega)$. A conventional matched filter beamformer steered to angle ω produces the following statistic

$$\frac{1}{MT}\mathbf{a}^H(\omega)\mathbf{Y}\mathbf{x}^H = \underbrace{\sqrt{P}}_{\text{desired}} + \underbrace{\frac{\sqrt{P}}{M}\mathbf{a}^H(\omega)\tilde{\mathbf{a}}}_{\zeta_a} + \underbrace{\frac{1}{MT}\mathbf{a}^H(\omega)\mathbf{N}\mathbf{x}^H}_{\zeta_n} \tag{111}$$

assuming unit amplitude symbols and antenna gains. For simplicity, assume that the elements of $\tilde{\mathbf{a}}$ are uncorrelated random perturbations with variance σ_a^2. Then the variance of the calibration error term ζ_a is $\sigma_a^2 P/M$, while that of the noise term ζ_n is σ^2/MT. Thus, the impact of the array calibration error boils down to the relative size of σ_a^2 compared with σ^2/PT. For example, if we assume that the antennas are calibrated to within a 5% accuracy ($\sigma_a = 0.05$), then the array calibration errors would dominate the effects of the noise when $T > 400$ for an SNR of 0dB, and when $T > 125$ for an SNR of 5dB. This simple analysis suggests that array calibration errors, while not negligible, are only an important source of error at higher SNRs not typical of scenarios envisioned for RIS-based systems (especially those at millimeter wave and higher frequencies). But further study is needed to confirm this result for the more complicated channel conditions when RIS are involved.

D. Algorithm Complexity

Another important factor to consider in choosing the type of channel model is the complexity required to compute the channel estimate. For the LS approach described in (13)-(15) and implemented with orthogonal $\mathbf{X}$ and $\boldsymbol{\Psi}$, the required

number of operations is $O(M(K+N)T) \approx O(MNT)$, where the approximation is due to the fact that typically $N \gg K$. Since $T = O(KN)$ is needed for identifiability, the complexity order is $O(MKN^2)$, which is quadratic in the number of RIS elements. If non-orthogonal RIS phase sequences are employed, computation of the pseudo-inverse of $\mathbf{\Psi}^H$ requires $O(N^3)$ operations in addition to the $O(MKN^2)$ needed to calculate the channel estimate, and hence the relative values of MK and N determine the overall computational load.

For the geometric case, tackling the estimation problem implied by the general data model in (57a) via classical AoA estimation methods is only practical for very small values of d_H and d_G. The sparsity-based model in (57b) is difficult to implement even for small d_H and d_G since the size of dictionary is $N_{BD}N_{UD}N_{RD}^2$, which is usually of order $O(MKN^2)$ if we assume the resolution of the grid points is inversely proportional to the number of antennas on each side of the link. Since the overall complexity order would involve the product of the dictionary size with the sparsity level $d_H d_G$ and the number of measurements MT, it is clear that this direct approach is not feasible.

Instead, we again consider the decoupled approach of Section IV-C2 as a practical alternative. We will focus on the complexity of CS-based approaches since they have been most frequently considered for RIS-based channel estimation. For a single K-antenna user, the steps involved in the decoupled approach are a d_H-sparse estimation with an $M \times N_{BD}$ dictionary, followed by a d_G-sparse estimation with a $K \times N_{UD}$ dictionary, and concluding with $d_H d_G$ 1-sparse problems with an $N \times N_{RD}$ dictionary. The total complexity is thus of order

$$\begin{aligned} &O(d_H M N_{BD}) + O(d_G K N_{UD}) + O(d_H d_G N N_{RD}) \\ &\qquad \approx O(d_H M^2) + O(d_G K^2) + O(d_H d_G N^2)\,, \end{aligned} \tag{112}$$

where our approximation again assumes grid resolution on the order of the reciprocal of the number of antennas. We see that the decoupled CS-based approach will have a lower complexity provided that $d_H \ll KN^2/M$, $d_G \ll MN^2/K$, and $d_H d_G \ll MK$, all of which would be considered reasonable for large RIS and BS arrays. The computational load is lower for the case where there are K single-antenna users, based on the approach described in Section IV-E1.

VI. Numerical Examples

The examples below compare $\mathrm{CRB}_u(\breve{\mathbf{h}}_c)$ and $\mathrm{CRB}_s(\breve{\mathbf{h}}_c)$ for several different scenarios. The first set of examples also includes the performance of the LS algorithm and an implementation of the decoupled estimator for the geometric channel model described in Section IV-C2, referred to as "CS-Geom" in the plots and discussion below. The CS optimizations were implemented using OMP on a grid with M points for the BS and N points for the RIS. 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 $K = 1$ antenna, 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 $[0°, 90°]$ 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 $\mathbf{\Phi}_t$ are not chosen to maximize the coherent gain offered by the RIS, the RIS provides sufficient combining gain such that $\|\mathbf{H}\mathbf{\Phi}_t\mathbf{G}\| \gg \|\mathbf{H}_d\|$. The SNR in the examples below is defined as P/σ^2, but the effective SNR will be a function of the channel gain $\|\mathbf{H}_d + \mathbf{H}\mathbf{\Phi}_t\mathbf{G}\|$, which will change as a function of the number of antennas, RIS elements, and path gains.

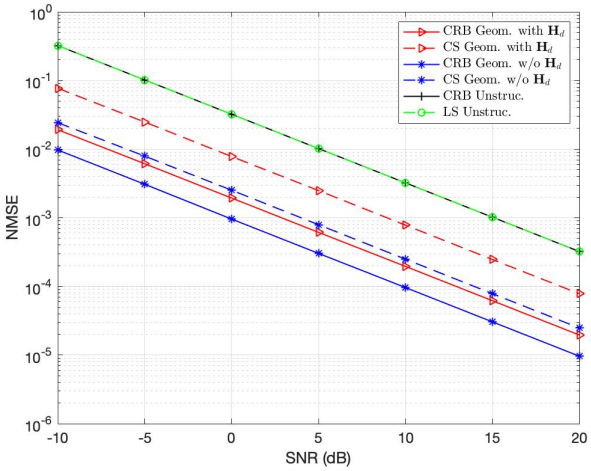

Fig. 2: Channel estimate NMSE and CRB for various channel models vs. SNR. Channel parameters are $M = 30, N = 30, T = 31, K = 1, d_F = 5, d_G = 5, d_H = 2$.

Figs. 2-4 assume a BS with $M = 30$ antennas and an RIS with $N = 30$ elements. Fig. 2 plots the performance 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 $K = 1$ antenna UE. The CRB for the unstructured model and the geometric model with and without the direct BS-UE channel $\mathbf{H}_d$ are included, along with the performance of the LS and CS-Geom algorithms. The CRB curves are plotted with solid lines, while the performance of the algorithms is plotted with dashed lines. We see that the performance of the LS algorithm exactly matches the unstructured CRB, while for CS-Geom there is a significant gap between the bound and the accuracy of the algorithm. This is due to a number of factors, the most important being the rather coarse grid that was used to estimate the AoAs and AoDs. Despite this gap, the CS-Geom approach still provides a 10dB improvement in channel estimation accuracy across all SNRs. It is important to note that this gap can be closed considerably using a technique that estimates the AoAs/AoDs as continuous variables.

The advantage of the geometric model in terms of training

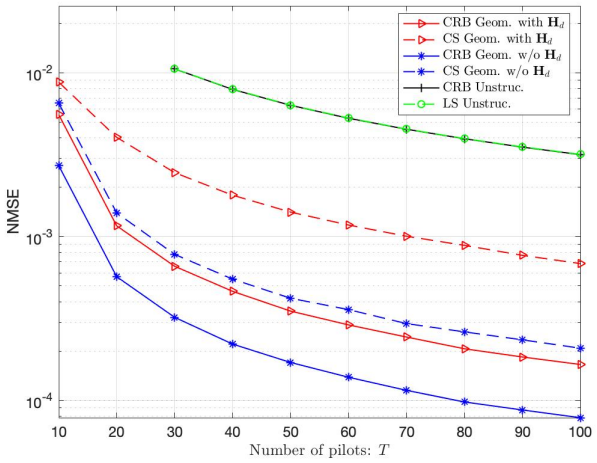

Fig. 3: Channel estimate NMSE and CRB for various channel models vs. number of training samples T. Channel parameters are $M = 30, N = 30$, SNR $= 5$dB, $K = 1, d_F = 5, d_G = 5, d_H = 2$.

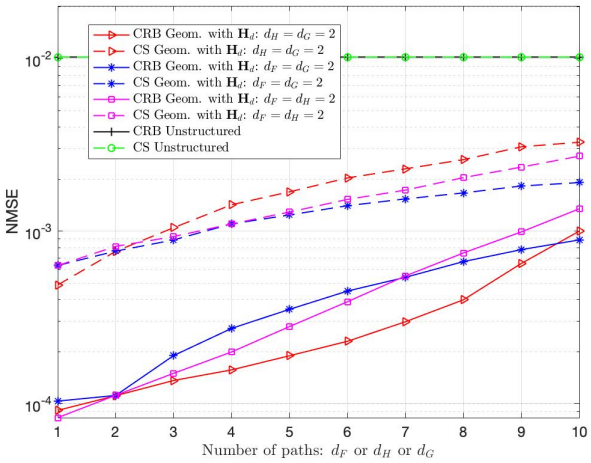

Fig. 4: Channel estimate NMSE and CRB for various models vs. number of propagation paths. Channel parameters are $M = 30, N = 30, T = 31, K = 1$, SNR $= 5$dB.

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 $\|\mathbf{H}\mathbf{\Phi}_t\mathbf{G}\| \gg \|\mathbf{H}_d\|$. 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 performance as a function of the number of paths present in the geometric channel model, again assuming $M = N = 30$ and the minimum number of $T = K(N + 1) = 31$ training samples. The unstructured model is unaffected by these changes, and both the CRB and LS algorithm show constant accuracy versus the number of paths. 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 and NMSE of CS-Geom increase with the number of paths. It is clear from this example 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 channel estimation performance as a function of N assuming that $M + N = 70$, with the other parameters set as before ($K = 2, d_F = 5, d_G = 5, d_H = 2$, SNR $=$ 5dB). The curve for the unstructured model assumes that T is increasing with N according to $T = KN$, which explains why the performance improves as the size of the RIS increases. The blue curves for the geometric model assume the same increasing values for T, while the red lines assume a fixed value of $T = 30$. 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, so for these plots we include only the CRB results. Unlike the previous case, there are $U = 2$ users with 2 antennas each ($K = 4$),

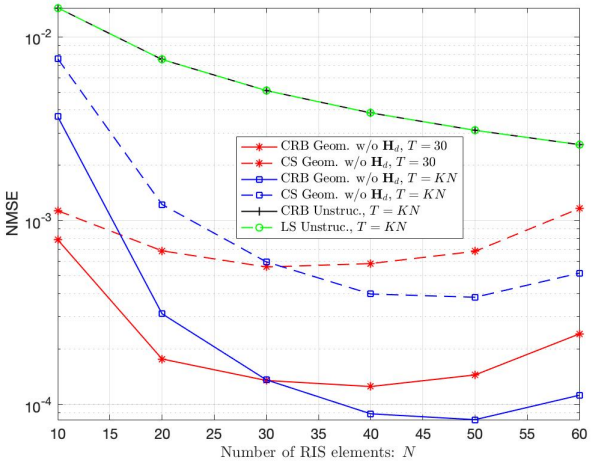

Fig. 5: Channel estimate NMSE and CRB for various channel models vs. number of RIS elements N, where $M + N = 70$. Channel parameters are SNR $= 5$dB, $K = 2, d_F = 5, d_G = 5, d_H = 2$. For the unstructured model and the blue curves (geometric model), we have $T = KN$. For the red curves (geometric model), $T = 30$.

the RIS training sequences are not random but rather chosen such that the diagonal CRB in (108) 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 $\mathbf{H}_d$" - bound for estimate of $\mathbf{H}_d$ assuming geometric model and with the RIS channel present, (2) "Geometric $\mathbf{H}_c$" - bound for estimate of $\mathbf{H}_c$ using geometric model with $\mathbf{H}_d$ included, (3) "Unstructured" - bound for unstructured model (independent of $\mathbf{H}_d$), (4) "Geom. $\mathbf{H}_d$ only" - bound for $\mathbf{H}_d$ assuming geometric model when no RIS is present, and (5)

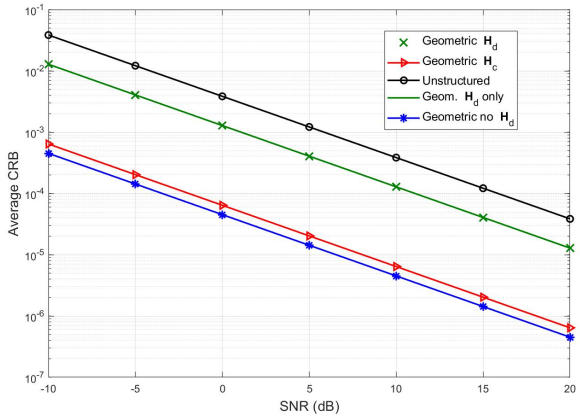

Fig. 6: CRB for various channel models vs. SNR. Channel parameters are $M = 6, N = 64, T = 260, U = 2, K = 4, d_F = 2, d_G = 3, d_H = 2$.

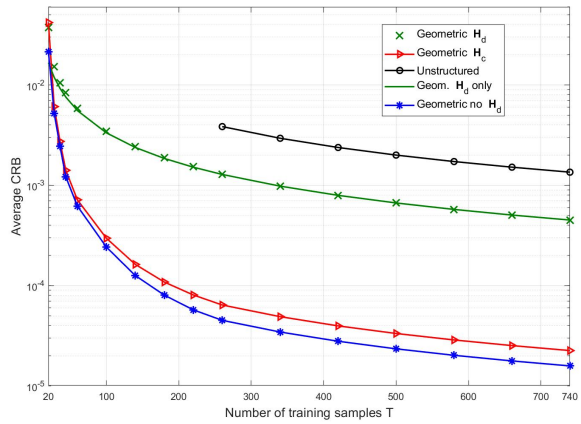

Fig. 7: CRB for various channel models vs. number of training samples T. Channel parameters are $M = 6, N = 64, \text{SNR} = 0\text{dB}, U = 2, K = 4, d_F = 2, d_G = 3, d_H = 2$.

"Geometric no $\mathbf{H}_d$" - bound for $\mathbf{H}_c$ assuming geometric model 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 to be identifiable for the case of the largest RIS.

VII. Emerging Topics

A. Alternatives to Passive Reflection-Only RIS

The methods discussed above have all assumed a passive reflecting RIS with elements that only apply a tunable phase

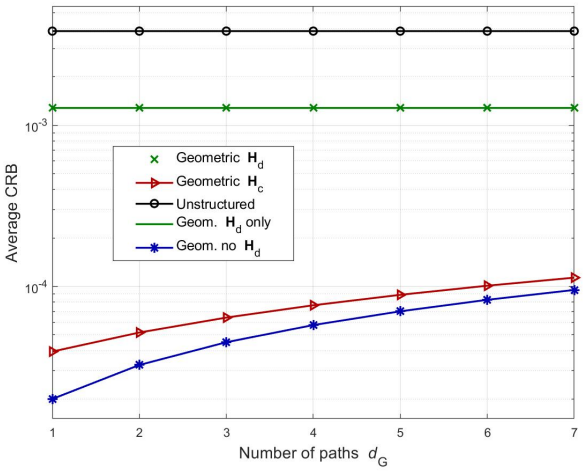

Fig. 8: CRB for various models vs. number of propagation paths d_G. Channel parameters are $M = 6, N = 64, T = 260, U = 2, K = 4, \text{SNR} = 0\text{dB}, d_F = 2, d_H = 2$.

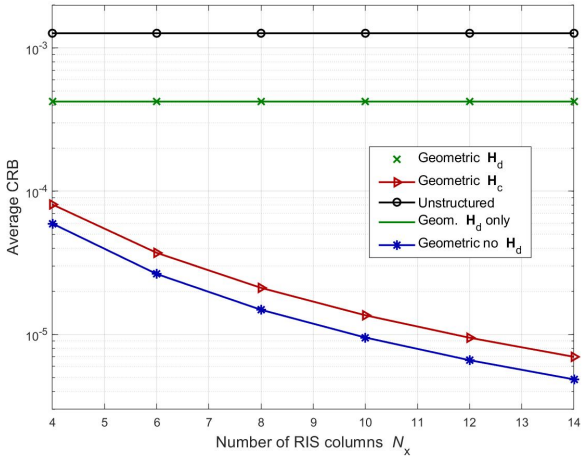

Fig. 9: CRB for various channel models vs. number of columns N_x in a square RIS with $N = N_x^2$ elements. Channel parameters are $M = 6, T = 788, \text{SNR} = 0\text{dB}, U = 2, K = 4, d_F = 2, d_G = 3, d_H = 2$.

shift to the impinging signal, together with a phase- and frequency-dependent attenuation factor. Recently, several new RIS devices and architectures have been introduced that include active receivers, local baseband processing, or the ability to perform passive "transmission" through the RIS (akin to refraction) rather than just reflection. Proponents argue that the resulting benefits of such implementations outweigh the resulting increase in hardware complexity and power consumption at the RIS. The work focusing on active RIS elements has assumed only active receivers and not active transmission, 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 and is reflected by the RIS, data collected by an active RIS elements will have higher SNR due to lower path loss and could lead to higher fidelity channel estimates. In addition, data from the active receivers would allow the channels $\mathbf{H}$ and $\mathbf{G}$ to 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 most common channel estimation scenario considered thus far involves an RIS with a few dedicated active receivers whose collected data are used to estimate the gains and angles of the geometric channel model. These estimates are then used in turn to infer the full $\mathbf{H}$ and $\mathbf{G}$ channel matrices using the known RIS geometry [59–61] or a deep neural network (DNN) [62]. This matrix completion approach is similar to methods proposed for channel estimation in hybrid digital/analog systems. A tensor completion approach is proposed in [63], 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 [64].

A more general hybrid architecture is proposed in [65], [66], where each RIS element is designed to act as both a passive reflector and an active receiver. In such designs, not only is the reflected phase a tunable parameter, but also the fraction of the power absorbed at each element and applied to the receiver. The development of channel estimation methods that exploit this additional flexibility is an interesting open problem. Simultaneous Transmit and Receive (STAR)-RIS are another recently proposed RIS architecture in which each element can split the received energy into reflected or refracted ("transmitted") signals in a controllable way. LS-based channel estimators for a STAR-RIS system are presented in [67] for two cases, one where the elements time-switch between reflection and refraction (a simple extension of existing methods), and one where the received energy is split between reflection and refraction. We also mention here the related work of [68] 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 has been studied for certain special cases. In [69], 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 [70], 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 stage, 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:

$$T = O\Big(\frac{K N_1 N_2}{M} + K(N_1 + N_2)\Big)\,, \tag{113}$$

which can be large when the number of BS antennas M is small. An earlier approach to the double-RIS problem in [71] assumes 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 [72] 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 [73]. However, [73] 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 [35], [74] or compressive sensing in the geometric model [75]. Multi-stage DNN denoising networks are proposed in [76], [77], 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 [62] that uses a DNN to map the channel estimated with a few active RIS elements to the full RIS array. The approach in [60] 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. Finally, we mention here the approach of [78], which uses deep unfolding of the LS fit of the channel to the data with a nuclear norm regularization to promote a low-rank estimate.

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.

Reference	Geometric	Wideband	Multi UE	MIMO BS	MIMO UE	Active Rcvr	Double RIS	Learning
3				✓				
4, 39		✓						
5, 8, 9, 11, 12, 14, 32, 36, 68			✓	✓				
6								
7			✓					
10		✓	✓					
13		✓		✓				
15, 21, 23	✓		✓	✓				
16	✓		✓	✓	✓			
17	✓	✓	✓	✓				
18, 24	✓			✓				
19	✓				✓			
20, 22, 25, 27, 29, 30, 34	✓			✓	✓			
26	✓	✓						
28	✓	✓		✓				
33			✓	✓	✓			
35				✓				✓
59	✓		✓	✓		✓		
60	✓					✓		✓
61	✓		✓	✓	✓	✓		
62	✓		✓	✓	✓	✓		✓
63	✓	✓	✓	✓		✓		
64	✓					✓		
69							✓	
70			✓	✓			✓	
71				✓			✓	
72, 74			✓	✓				✓
73		✓	✓	✓	✓			✓
75				✓	✓	✓		✓
76								✓
77, 78				✓				✓

TABLE II: Summary of topics covered by references related to channel estimation.

VIII. Conclusions and Future Work

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, they have a higher computational complexity, and their achievable accuracy is limited due to the large number of channel coefficients that must be estimated. On the other hand, geometric channel models lead to estimation of many fewer parameters and hence have a much smaller training burden, lower overall complexity, and can achieve dramatically better performance. However, this improvement comes at the cost of determining the model order (*i.e.,* number of propagation paths), and some of the performance gain will be lost in practice due to the use of grid-based estimators and 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 Table II, we summarize the references related to channel estimation that we have discussed in this work, and to aid those seeking more details, we indicate if these papers consider channel estimation for geometric channel models, wideband channel models, multiple UEs, multi-antenna BS, multi-antenna UEs, active receivers, double RIS, and learning-based algorithms.

In addition to these emerging topics discussed in Section VII, many open research problems related to RIS remain. For example, while we have considered only relatively simple scenarios involving a single RIS, a challenging question to

answer is how to efficiently coordinate the estimation of channels for multiple RIS when they are visible to multiple basestations and their reflections interact. We have considered only simple models for RIS element behavior, but future work should investigate the implications of more realistic models that account for the coupled dependence of the RIS element gains and phases, their variation in frequency for wideband scenarios, mutual coupling, nonlinear properties, etc. More work is also needed to further reduce the training overhead and computational complexity associated with RIS channel estimation. Possible topics include reducing the complexity of compressive-sensing based approaches that normally require large dictionaries for RIS-based CSI estimation, exploiting 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, and advancing the application of learning-based methods. Finally, a systematic study of the impact of modeling errors on RIS-based systems is needed to more rigorously establish the potential benefits of geometric models.

ACKNOWLEDGEMENT

This work was supported by the U.S. National Science Foundation under Grants ECCS-2030029 and CNS-2107182, and by the National Natural Science Foundation of China under Grant 61971088.

REFERENCES

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

[2] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[3] 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*, Barcelona, Spain, May 2020, pp. 5000–5004.

[4] 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.

[5] 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.

[6] 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.

[7] 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. Commun. Soc.*, vol. 1, pp. 661–680, May 2020.

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

[9] 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.

[10] 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.

[11] 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.

[12] Y. Wei, M.-M. Zhao, M.-J. Zhao, and Y. Cai, "Channel estimation for IRS-aided multiuser communications with reduced error propagation," *IEEE Trans. Wireless Commun.*, 2021.

[13] 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.

[14] L. Wei, C. Huang, G. C. Alexandropoulos, Z. Yang, C. Yuen, and Z. Zhang, "Joint channel estimation and signal recovery in RIS-assisted multi-user MISO communications," in *Proc. IEEE Wireless Commun. & Networking Conf. (WCNC)*, Nanjing, China, Mar. 2021.

[15] 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

[16] 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," *IEEE J. Sel. Topics Signal Proc.*, vol. 15, no. 3, pp. 789–802, Feb. 2021.

[17] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmwave massive MIMO systems," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, Dublin, Ireland, Jul. 2020.

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

[19] S. Noh, H. Yu, and Y. Sung, "Training signal design for sparse channel estimation in intelligent reflecting surface-assisted millimeter-wave communication," *IEEE Trans. Wireless Commun.*, 2022 (to appear).

[20] 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, May 2020.

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

[22] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmwave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5786–5797, Apr. 2021.

[23] 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.

[24] W. Zhang, J. Xu, W. Xu, D. W. K. Kwan, and H. Sun, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593–597, Feb. 2021.

[25] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Proc. Lett.*, vol. 28, pp. 513–517, Feb. 2021.

[26] S. Ma, W. Shen, J. An, and L. Hanzo, "Wideband channel estimation for IRS-aided systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6240–6253, Oct. 2021.

[27] 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.

[28] Y. Liu, S. Zhang, F. Gao, J. Tang, and O. A. Dobre, "Cascaded channel estimation for RIS assisted mmWave MIMO transmissions," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 2065–2069, Sept. 2021.

[29] H. Liu, J. Zhang, Q. Wu, H. Xiao, and B. Ai, "ADMM based channel estimation for RISs aided millimeter wave communications," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2894–2898, Sept. 2021.

[30] 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," *IEEE Trans. Wireless Commun.*, 2022.

[31] E. Björnson, "Optimizing a binary intelligent reflecting surface for OFDM communications under mutual coupling," in *Proc. 25th Int'l ITG Workshop on Smart Antennas (WSA)*, Nov. 2021.

[32] 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

[33] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[34] 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, Jun. 2021.

[35] 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. Commun. Soc.*, vol. 2, pp. 471–487, Mar. 2021.

[36] 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, Aug. 2021.

[37] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Letters*, vol. 10, no. 4, pp. 830–834, Apr. 2021.

[38] 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, Jul. 2020.

[39] W. Yang, H. Li, M. Li, Y. Liu, and Q. Liu, "Channel estimation for practical IRS-assisted OFDM systems," in *Proc. IEEE Wireless Commun. & Networking Conf. (WCNC) Workshops*, Nanjing, China, Mar. 2021.

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

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

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

[43] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. 37, no. 7, pp. 996–1000, Jul. 1989.

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

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

[46] 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, Jul. 2011.

[47] 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, Apr. 2012.

[48] 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, Nov. 2016.

[49] R. Carrillo and K. Barner, "Lorentzian iterative hard thresholding: Robust compressed sensing with prior information," *IEEE Trans. Signal Proc.*, vol. 61, no. 19, pp. 4822–4833, Oct. 2013.

[50] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, p. 4655–4666, Dec. 2007.

[51] K. Yu, M. Shen, R. Wang, and Y. He, "Joint nuclear norm and l1–2-regularization sparse channel estimation for mmwave massive MIMO systems," *IEEE Access*, vol. 8, pp. 155 409–155 416, Aug. 2020.

[52] J. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Wireless Commun.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[53] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing — Part I: Derivation," *IEEE Trans. Signal Proc.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

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

[55] 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, Nov. 2020.

[56] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. 37, no. 5, pp. 720–741, May 1989.

[57] A. Zappone, M. Di Renzo, F. Shams, X. Qian, and M. Debbah, "Overhead-aware design of reconfigurable intelligent surfaces in smart radio environments," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 126–141, Jan. 2021.

[58] X. Liu and N. Sidiropoulos, "Almost sure identifiability of constant modulus multidimensional harmonic retrieval," *IEEE Trans. Signal Proc.*, vol. 50, no. 9, pp. 2366–2368, Sept. 2002.

[59] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1132–1142, Feb. 2022.

[60] 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, Mar. 2021.

[61] X. Chen, J. Shi, Z. Yang, and L. Wu, "Low-complexity channel estimation for intelligent reflecting surface-enhanced massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 996–1000, May 2021.

[62] Y. Jin, J. Zhang, X. Zhang, H. Xiao, B. Ai, and D. W. K. Ng, "Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11 083–11 088, Oct. 2021.

[63] 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, Jun. 2021.

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

[65] G. Alexandropoulos, N. Shlezinger, I. Alamzadeh, M. Imani, H. Zhang, and Y. Eldar, "Hybrid reconfigurable intelligent metasurfaces: Enabling simultaneous tunable reflections and sensing for 6G wireless communications," April 2021. [Online]. Available: https://arxiv.org/abs/2104.04690

[66] I. Alamzadeh, G. Alexandropoulos, N. Shlezinger, and M. Imani, "A reconfigurable intelligent surface with integrated sensing capability," *Scientific Reports*, vol. 11, 2021, Article Number 20737.

[67] C. Wu, C. You, Y. Liu, X. Gu, and Y. Cai, "Channel estimation for STAR-RIS-aided wireless communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 652–656, Mar. 2022.

[68] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," *IEEE Trans. Wireless Commun.*, 2021.

[69] 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.

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

[71] 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, Aug. 2020.

[72] 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, Sept. 2020.

[73] 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

[74] 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.*, vol. 21, no. 2, pp. 898–912, Feb. 2022.

[75] 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. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.

[76] 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, Aug. 2021.

[77] 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, Jun. 2021.

[78] J. He, H. Wymeersch, M. Di Renzo, and M. Juntti, "Learning to estimate RIS-aided mmwave channels," *IEEE Wireless Commun. Letters*, 2022.

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 375 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, 2010 and 2022 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.

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 surfaces and signal processing.

Rang Liu (Graduate Student Member, IEEE) received the B.S. degree in electronics information engineering from Dalian University of Technology, Dalian, China, in 2018. She is now studying toward the Ph.D. degree with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China. Her current research interests are focused on signal processing, massive MIMO systems, reconfigurable intelligent surfaces, and integrated sensing and communication. She was the recipient of the National Scholarship in 2020.

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. From 2019 to 2021, he was a Lecturer in the same university. Since 2021, he is a full professor in Southeast University. His research interests mainly include reconfigurable intelligent surfaces (RIS), intelligent reflection surface (IRS), ultra-reliable low latency communication, machine learning, unmanned aerial vehicles, the Internet of Things, and mobile edge computing, and he has co-authored over 120 IEEE journal papers. He currently serves as an Editor for IEEE Wireless Communication Letters, IEEE Communications Letters and IEEE ACCESS. He has served as a Guest Editor or Lead Guest Editor for special issues in the IEEE Journal on Selected Areas in Communications, IEEE Journal of Selected Topics in Signal Processing, IEEE Vehicular Technology Magazine, IEEE Open Journal of Vehicular Technology, and IEEE ACCESS. He has organized Workshops on RIS in IEEE ICCC 2021 and IEEE Globecom 2021, and he is currently the Workshops and Symposia officer for the Reconfigurable Intelligent Surfaces Emerging Technology Initiative. He is workshop chair for IEEE WCNC 2024, and TPC co-chair for IEEE ICCT 2022. He has served as a TPC member for numerous conferences, such as ICC and GLOBECOM, and was the Student Travel Grant Chair for ICC 2019. He received the IEEE ComSoc Leonard G. Abraham Prize in 2022.

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 Jan. 2011 to Aug. 2013, he was a Post Doctoral Research Associate with the Department of Electrical Engineering, SUNY-Buffalo. From Aug. 2013 to June 2014, Dr. Li 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 presently a Professor. His current research interests are in the general areas of communication theory and signal processing with applications to integrated sensing and communication, reconfigurable intelligent surfaces, millimeter-wave communications, massive MIMO systems, and secure wireless communications. He was the co-chair for the IEEE INFOCOM 2019 workshop on Wireless Communications and Networking in Extreme Environments.