Optimal Time and Power Allocation for Phase-Shift Configuration and Downlink Channel Estimation in RIS-Aided Systems

Suji Naduvilpattu, *Student Member, IEEE*, and Neelesh B. Mehta, *Fellow, IEEE*

***Abstract*— A fundamental trade-off exists between the time and energy allocated to pilots and data, accuracy of channel estimates, and data rate in a reconfigurable intelligent surface (RIS)-aided system. We optimize the above trade-off for a two-phase training scheme. In the first phase, the base station (BS) estimates the channel from the uplink pilots and configures the RIS. In the second phase, the user equipment estimates the channel from the downlink pilots and coherently demodulates the data. We derive an expression for the achievable rate that accounts for the impact of the channel estimation errors on the RIS phase-shift configuration and data demodulation. Our analysis uses a novel tractable approximation for the effective downlink channel gain and a novel proof that it is asymptotically Gaussian even in the presence of spatial correlation. Our analysis applies to the scenario where enough pilots are sent to estimate the cascaded channels and the cascaded channel grouping scenario that uses fewer pilots. We also study two channel models that depend on the location of the RIS relative to the BS. We derive insightful, closed-form expressions for the optimal powers and training durations. The optimal solution highlights the importance of boosting the pilot powers.**

***Index Terms*— Reconfigurable intelligent surface, channel estimation, power allocation, pilots, achievable rate, grouping.**

I. Introduction

RECONFIGURABLE intelligent surfaces (RISs) consist of thin reflecting elements of metamaterials or patch-arrays that can change the electromagnetic properties of the impinging waves, such as phase, in a controllable manner [2]. RISs are easily deployable as no dedicated power supply is required by them to transmit in their passive configuration [3]. An appropriately configured RIS can enhance coverage in scenarios where the user equipment (UE) is blocked from its base station (BS), enhance data rate by improving the signal-to-noise ratio (SNR), or null an interferer [4], [5]. This makes RIS an appealing technology for 6G [5].

Manuscript received 9 June 2023; revised 3 December 2023; accepted 27 January 2024. Date of publication 13 February 2024; date of current version 14 August 2024. This work was supported in part by the "Next Generation Wireless Research and Standardization on 5G and Beyond" Project funded by the Ministry of Electronics and Information Technology (MeitY), Government of India. A part of this paper has been accepted for presentation at the IEEE Intl. Conf. on Communications (ICC) 2024 [1]. The associate editor coordinating the review of this article and approving it for publication was T. Tsiftsis. *(Corresponding author: Suji Naduvilpattu.)*

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science (IISc), Bengaluru 560012, India (e-mail: sujivalsala@gmail.com; nbmehta@iisc.ac.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2024.3362383.

Digital Object Identifier 10.1109/TWC.2024.3362383

In order to benefit from the RIS, the reflection coefficients of its elements need to be configured appropriately. Broadly, three methods have been studied for configuring the RIS in the literature. In the instantaneous channel state information (CSI)-based methods, the instantaneous CSI of the direct channel between the BS and the UE and the cascaded channels is used to configure the RIS [6], [7], [8]. In codebook-based methods, the best RIS phase-shift configuration is selected from a set of predefined codebook of configurations [9], [10]. In statistical CSI-based methods, the RIS configuration is based on channel statistics such as covariance [11], [12], [13]. The above methods vary in the training overhead they require and the passive RIS beamforming gains they can generate.

For the instantaneous CSI-based method, the channel training is based on the on/off method, in which the RIS elements are turned on sequentially one at a time [7], [8], [14], [15], or the discrete Fourier transform (DFT)/Hadamard matrix-based method, in which the RIS configuration is chosen from the columns of a DFT/Hadamard matrix [6], [7], [16]. In [6], the least squares (LS) estimator and linear minimum mean squared error (LMMSE) estimator-based cascaded channel estimation is studied. The RIS phase-shifts and power allocation are optimized to maximize the rate. In [7], pilot and data power allocations are optimized to maximize the achievable rate. In [8], an achievable rate-maximizing pilot power allocation is investigated. Cascaded channel estimation when the transmitter and receiver have multiple antennas is studied in [16]. While the beamforming gain of the instantaneous-CSI based methods is large, the training overhead is proportional to the number of the RIS elements.

To reduce the training overhead, the cascaded channel grouping technique, in which only the sum of cascaded channels per group needs to be estimated, is studied in [14], [15]. In [15], the focus is on optimizing the RIS configuration and the transmit power to maximize the achievable rate. The optimal grouping size that maximizes the achievable rate is investigated in [14]. The channel estimation scheme proposed in [17] makes use of the eigenvectors of the spatial correlation matrix of the cascaded channel gain vector. As a result, the pilot overhead is proportional to the rank of the correlation matrix, which is smaller. Machine learning-based approaches are studied for cascaded channel estimation in [18] and [19]. A convolutional neural network-based deep learning scheme is proposed in [18], while a federated learning-based approach is studied in [19]. A compressed sensing-based two-stage cascaded channel estimation method, which exploits the low-rank nature of the sparse channel to reduce the training overhead, is studied in [20].

In the codebook-based methods, the training overhead and the beamforming gain depend on the size of the codebook [21]. Two codebook designs, one with random phase-shift entries and the other in which the Euclidean distance between the codes is maximized, are studied in [9], [10]. In [9], rate-maximizing active and passive beamforming are studied. The focus in [10] is on minimizing the transmit power.

In the statistical CSI-based methods, the training overhead is lower when amortized over time since the channel statistics vary slowly. However, the beamforming gain is also lower due to the reduced degrees of freedom in configuring the RIS. To configure the RIS reflection coefficients, a genetic algorithm-based approach is studied in [12], a gradient ascent algorithm is developed in [11], and a majorization-minimization-based approach is employed in [13]. The above three works focus on maximizing the uplink rate.

A. Contributions

We focus on the instantaneous CSI-based method because it was among the first methods proposed for determining the RIS phase-shift configuration and has attracted considerable attention in the literature. In this method, noise during channel estimation leads to channel estimation errors that affect the RIS-aided system in multiple ways. First, the channel estimation errors result in a sub-optimal configuration of the RIS reflection coefficients. Furthermore, the estimation errors have a non-linear impact on the phase-shifts of the RIS elements. This alters the effective gain of the BS-to-RIS-to-UE channel, as it is a function of the cascaded channel gains and reflection coefficients of the RIS elements. Second, the errors lead to an imperfect estimate of the above effective channel gain itself. Increasing the pilot energy or training duration improves the accuracy of the channel estimates that are used for configuring the RIS phase-shifts and for demodulating the data symbols. However, doing so reduces the number of data symbols that can be transmitted in a coherence interval and the energy available at the BS for transmitting them. We comprehensively address the above fundamental trade-off and characterize the cumulative impact of the channel estimation errors on the downlink achievable rate for the data transmitted by the BS to the UE via the RIS. We make the following contributions:

1) Comprehensive Channel Models and Training Scenarios: We consider two channel models that depend on the location of the RIS relative to the BS. In the first model, the BS-RIS link is a deterministic line-of-sight (LoS) link. This arises when the RIS is placed at a height close to the BS [6]. In the second model, the BS-RIS link is a non-line-of-sight (NLoS) link [22]. This is suitable for an urban scenario where the LoS between the BS and the RIS can be blocked. In both models, the RIS-UE link is a NLoS link.

We study a two-phase training scheme that exploits channel reciprocity of time-division duplex (TDD) systems. In the first uplink training phase, the BS estimates the cascaded channel gains from the uplink pilots and then configures the phase-shifts of the RIS. In the second downlink training phase, the BS sends pilots followed by data. The UE estimates the effective BS-UE channel gain and uses it to coherently demodulate the data. Thus, no feedback from the BS to the UE is required.

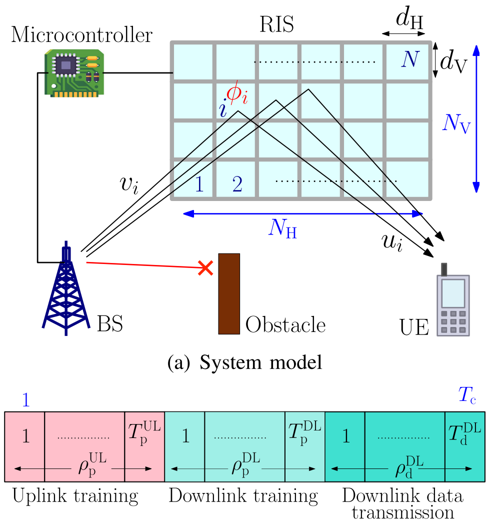
Another comprehensive aspect of our study is that it covers two training scenarios. In the first scenario, enough pilots are sent to estimate each of the individual cascaded channel gains. In the second scenario, which arises in cascaded channel grouping, the number of pilots is less than the number of reflecting elements.

2) Lower Bound on Achievable Data Rate: We derive a novel lower bound on the achievable ergodic data rate for the above training scheme that accounts for the training overheads and the combined impact of the imperfect phase-shift configuration of the RIS, due to the uplink channel estimation errors, and the downlink channel estimation errors on data demodulation. To make the derivation tractable, we propose a novel approximation for the effective BS-UE channel gain with channel estimation errors.

3) Gaussianity of BS-UE Channel Gain for a Large RIS: The above bound on the achievable rate assumes that the effective BS-UE channel gain in the presence of channel estimation errors is a Gaussian random variable (RV). This is motivated by the central limit theorem. When a large number of independent and identically distributed RVs are summed, it is sufficient for the RVs to have a finite variance for the central limit theorem to apply. However, this is no longer sufficient when the RVs are correlated. Instead, our proof employs more sophisticated ρ-mixing arguments [23]. We show numerically that this result also applies to planar RISs with just a few elements. We note that the Gaussian assumption has been employed in the RIS literature without any such theoretical backing [7].

4) Numerical Results: Our numerical results show that the proposed optimal scheme outperforms the allocation of equal power for training and data transmission and the on/off-based training scheme. We find that grouping of the RIS elements significantly improves the achievable rate. We determine the optimal group size that maximizes the achievable data rate. We note that a comprehensive comparison with the codebook-based methods [9], [10] and statistical CSI-based methods [11], [12], [13] is beyond the scope of this paper. Such a comparison depends on several system variables, the codebook size, and how often the channel gains and their statistics change.

Comparison with Literature: There are fundamental differences between our system model, training model, and analysis of the achievable rate and those considered in the literature. In [7], the focus is on the uplink. As a result, the BS can employ the same channel estimates to configure the RIS and demodulate the data. In [8], the impact of noise on only the phase-shift configuration is studied. However, the effective channel gain is implicitly assumed to be perfectly known at the UE when it demodulates the data. In [6], [15], [16], the actual channel gain is replaced with the estimated gain in the rate expression. The rate expression so obtained is neither the channel capacity nor an achievable rate. In [14], [17],



(a) System model

(b) Training scheme

Fig. 1. System model showing an RIS-aided BS transmitting data to a UE. Also shown are the uplink training, downlink training, and data transmission phases, with the corresponding durations and transmit powers.

[18], [19], [20], the focus is on minimizing the mean squared error (MSE) or the training overhead. The impact of channel estimation error on the rate is not studied. Lastly, the cascaded channel gains are assumed to be uncorrelated in [7], [8], [14].

B. Outline and Notations

Section II describes the system model, the training scheme, and the achievable data rate-maximization problem. In Section III, we prove the asymptotic Gaussianity of the downlink channel gain, numerically verify it for smaller RIS systems, and derive the optimal training durations and transmit powers. We analyze the impact of grouping of RIS elements on the data rate in Section IV. Numerical results and our conclusions follow in Sections V and VI, respectively.

Notation: We show scalar variables in normal font and matrices in bold font. $\mathbb{E}[X]$, $\mathbb{E}[X|Y]$, $\mathrm{Var}[X]$, and $F_X(.)$ denote expectation, expectation conditioned on Y, variance, and cumulative distribution function (CDF) of X, respectively. $\mathrm{Cov}[X,Y]$ denotes the covariance between X and Y. The operators $|.|$, $\angle.$, $(.)^*$, $\Re[.]$, and $\Im[.]$ represent the absolute value, angle, complex conjugate, real part, and imaginary part, respectively. $\mathbf{X}^H$ and $\mathbf{X}^T$ denote Hermitian and transpose of $\mathbf{X}$, respectively. $\mathbf{I}_n$ denotes the identity matrix of order n. The notation $X \sim \mathcal{CN}(\mu, \tau^2)$ means that X has a complex normal distribution with mean μ and variance τ^2. The $(i,k)^{\text{th}}$ element of $\mathbf{X}$ is denoted as $\mathbf{X}(i,k)$.

II. System Model and Problem Formulation

An RIS with N passive reflecting elements is deployed, as shown in Figure 1(a). Let d_{H} and d_{V} be the width and height of an RIS element normalized by the wavelength, and N_{H} and N_{V} be the number of RIS elements in the horizontal and vertical directions. Let $N = N_{\mathrm{H}} N_{\mathrm{V}}$. The reflection coefficient of the i^{th} reflecting element is $\phi_i = e^{j\theta_i}$, where $\theta_i \in [0, 2\pi)$ [6]. The BS configures the phases $\theta_1, \ldots, \theta_N$ by sending control signals over a wired link.

The channel coefficients from the BS to the i^{th} element of the RIS and from the i^{th} element of the RIS to the UE are denoted as v_i and u_i, respectively. The cascaded channel gain h_i between the BS and the UE through the i^{th} element of the RIS is $h_i = v_i u_i$. Let $\mathbf{v} = [v_1, \ldots, v_N]^T$, $\mathbf{u} = [u_1, \ldots, u_N]^T$, and $\mathbf{h} = [h_1, \ldots, h_N]^T$. The direct link between the BS and UE is assumed to be blocked by obstacles. We consider a single-antenna BS and UE because the cumulative impact of channel estimation errors on the statistics of the effective channel gain and the achievable rate, and the joint optimization of the training durations, pilot power, and data power to address this impact are not fully understood even for this model in the literature. In multiple antenna systems, the training overhead is larger and the impact of the channel estimation errors on the precoding vector at the BS also needs to be considered.

The RIS-UE link is an NLoS Rayleigh channel. Then, $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \beta_{\mathrm{u}} \mathbf{R})$, where β_{u} is the pathloss from the RIS to the UE. Here, $\mathbf{R} = \mathbb{E}[\mathbf{u}\mathbf{u}^H]/\beta_{\mathrm{u}}$ denotes the spatial correlation matrix of the RIS. We focus on the correlation model developed in [22] for an isotropic scattering environment. In it, the $(i,k)^{\text{th}}$ entry of $\mathbf{R}$ is given by $\mathbf{R}(i,k) = \mathrm{sinc}(2d_{i,k})$, where $d_{i,k}$ is the wavelength-normalized distance between the left-bottom corners of the i^{th} and the k^{th} elements of the RIS. Based on the nature of the BS-RIS link, we consider the following two channel models.

1) LoS-NLoS Model: In this model, the BS-RIS link is a deterministic LoS link. Let φ_{az} and φ_{el} be the azimuth and elevation angles of arrival at the RIS. Then, $v_i = \sqrt{\beta_{\mathrm{v}}} e^{j2\pi[\mathrm{mod}(i-1, N_{\mathrm{H}}) d_{\mathrm{H}} \cos\varphi_{\mathrm{el}} \sin\varphi_{\mathrm{az}} + \lfloor (i-1)/N_{\mathrm{H}} \rfloor d_{\mathrm{V}} \sin\varphi_{\mathrm{el}}]}$ for $i \in \{1, \ldots, N\}$ [22]. Here, β_{v} is the pathloss in the BS-RIS link. Thus, $h_i \sim \mathcal{CN}(0, \beta_{\mathrm{u}}\beta_{\mathrm{v}})$. We note that β_{u} and β_{v} are proportional to the area of the reflecting element [22].

2) NLoS-NLoS Model: In this model, the BS-RIS channels undergo Rayleigh fading. Then, $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \beta_{\mathrm{v}}\mathbf{R})$. Furthermore, v_i and u_i are independent since the BS and the UE are far apart [22]. Thus, $\mathbb{E}[h_i] = 0$ and $\mathrm{Var}[h_i] = \beta_{\mathrm{u}}\beta_{\mathrm{v}}$ for all $i \in \{1, \ldots, N\}$.

A. Training Scheme

The proposed training scheme is shown in Figure 1(b). A coherence block of length T_{c} symbols consists of the following two phases.

1) Uplink Training Phase: The UE transmits $T_{\mathrm{p}}^{\mathrm{UL}}$ pilots in the uplink with transmit power $\rho_{\mathrm{p}}^{\mathrm{UL}}$. Let $x_{\mathrm{p}}^{\mathrm{ul}}(k)$ be the pilot symbol in the k^{th} transmission. Without loss of generality, $x_{\mathrm{p}}^{\mathrm{ul}}(k) = 1$ for all $k \in \{1, \ldots, T_{\mathrm{p}}^{\mathrm{UL}}\}$. Then, the received signal vector $\mathbf{y}_{\mathrm{p}}^{\mathrm{UL}}$ at the BS is given by

$$\mathbf{y}_{\mathrm{p}}^{\mathrm{UL}} = \sqrt{\rho_{\mathrm{p}}^{\mathrm{UL}}} \boldsymbol{\Phi}^{\mathrm{UL}} \mathbf{h} + \mathbf{n}^{\mathrm{UL}}. \tag{1}$$

Here, $\boldsymbol{\Phi}^{\mathrm{UL}} \in \mathcal{C}^{T_{\mathrm{p}}^{\mathrm{UL}} \times N}$ is the reflection coefficient matrix of the RIS during uplink training and $\mathbf{n}^{\mathrm{UL}} \sim \mathcal{CN}\left(0, \sigma^2 \mathbf{I}_{T_{\mathrm{p}}^{\mathrm{UL}}}\right)$ is the additive white Gaussian noise (AWGN) at the BS. We choose $\boldsymbol{\Phi}^{\mathrm{UL}}$ such that $\left(\boldsymbol{\Phi}^{\mathrm{UL}}\right)^H \boldsymbol{\Phi}^{\mathrm{UL}} = T_{\mathrm{p}}^{\mathrm{UL}} \mathbf{I}_N$ and $|\boldsymbol{\Phi}^{\mathrm{UL}}(i,j)|^2 = 1$ for all $i \in \{1, \ldots, T_{\mathrm{p}}^{\mathrm{UL}}\}$ and $j \in \{1, \ldots, N\}$. These requirements are fulfilled when $\boldsymbol{\Phi}^{\mathrm{UL}}$ is constructed by taking, for example, N columns of a DFT matrix or a Hadamard matrix of size $T_{\rm p}^{\rm UL} \times T_{\rm p}^{\rm UL}$. This decorrelates the noise components in the observations corresponding to each of the cascaded channels [6], [24].

From $\mathbf{y}_{\rm p}^{\rm UL}$, the BS estimates $\mathbf{h}$. Let the estimate of h_i be $\hat{h}_i$ and let $\widehat{\mathbf{h}} = \left[\hat{h}_1, \ldots, \hat{h}_N\right]^T$. The BS then configures the reflection coefficients of the RIS elements as a function of $\widehat{\mathbf{h}}$ such that the effective channel strength is maximized. Thus, it sets the reflection coefficient $\hat{\phi}_i$ of the $i^{\rm th}$ reflecting element as [8]

$$\hat{\phi}_i = e^{j\angle \hat{h}_i^*}. \tag{2}$$

2) Downlink Training and Data Transmission Phase: After configuring the RIS in the uplink training phase, the BS transmits $T_{\rm p}^{\rm DL}$ pilots followed by $T_{\rm d}^{\rm DL}$ data symbols in the downlink with powers $\rho_{\rm p}^{\rm DL}$ and $\rho_{\rm d}^{\rm DL}$, respectively. The received pilot signal vector $\mathbf{y}_{\rm p}^{\rm DL}$ at the UE is

$$\mathbf{y}_{\rm p}^{\rm DL} = \sqrt{\rho_{\rm p}^{\rm DL}} g(\widehat{\mathbf{h}}) \mathbf{x}_{\rm p}^{\rm DL} + \mathbf{n}_{\rm p}^{\rm DL}, \tag{3}$$

where the pilot vector $\mathbf{x}_{\rm p}^{\rm DL} \in \mathcal{C}^{T_{\rm p}^{\rm DL} \times 1}$ satisfies $\left(\mathbf{x}_{\rm p}^{\rm DL}\right)^H \mathbf{x}_{\rm p}^{\rm DL} = T_{\rm p}^{\rm DL}$, $\mathbf{n}_{\rm p}^{\rm DL} \sim \mathcal{CN}\left(0, \sigma^2 \mathbf{I}_{T_{\rm p}^{\rm DL}}\right)$ denotes the AWGN vector at the UE during pilot reception, and $g(\widehat{\mathbf{h}})$ is the effective channel gain between the BS and UE. It is given by

$$g(\widehat{\mathbf{h}}) = \sum_{i=1}^{N} \hat{\phi}_i h_i. \tag{4}$$

Notice that $g(\widehat{\mathbf{h}})$ is affected by the noise in the uplink training phase.

The received data signal $y_{\rm d}^{\rm DL}(k)$ at the UE when the $k^{\rm th}$ data symbol $x_{\rm d}^{\rm DL}(k)$ is transmitted is

$$y_{\rm d}^{\rm DL}(k) = \sqrt{\rho_{\rm d}^{\rm DL}} g(\widehat{\mathbf{h}}) x_{\rm d}^{\rm DL}(k) + n_{\rm d}^{\rm DL}(k). \tag{5}$$

Here, $n_{\rm d}^{\rm DL}(k) \sim \mathcal{CN}\left(0, \sigma^2\right)$ is AWGN and $\mathbb{E}\left[|x_{\rm d}^{\rm DL}(k)|^2\right] = 1$ for all k.

From $\mathbf{y}_{\rm p}^{\rm DL}$, the UE estimates $g(\widehat{\mathbf{h}})$. Let its estimate be denoted by $\widehat{g}(\widehat{\mathbf{h}})$. Then, the UE employs $\widehat{g}(\widehat{\mathbf{h}})$ to coherently demodulate $x_{\rm d}^{\rm DL}(k)$ from $y_{\rm d}^{\rm DL}(k)$. Thus, $\widehat{g}(\widehat{\mathbf{h}})$ is affected by the noise in the uplink and downlink training phases.

B. Problem Formulation

Our aim is to optimize the number of training symbols and the transmit powers during training and data transmission to maximize the ergodic rate, subject to the following constraints.

1) UE Energy Budget Constraint: The energy consumed by the UE during the uplink training in a coherence block is limited to $E_{\rm UE}$. Thus, $\rho_{\rm p}^{\rm UL} T_{\rm p}^{\rm UL} = E_{\rm UE}$.

2) BS Energy Budget Constraint: The total energy consumed by the BS during the downlink training and data transmission phases in a coherence block is $E_{\rm BS}$. It can be written as $\rho_{\rm d}^{\rm DL} T_{\rm d}^{\rm DL} = \alpha E_{\rm BS}$ and $\rho_{\rm p}^{\rm DL} T_{\rm p}^{\rm DL} = (1-\alpha) E_{\rm BS}$. Here, the *energy allocation factor* $\alpha \in [0,1]$ is fraction of energy spent on downlink data transmission and is a parameter that we shall optimize.

Note: The above formulation does not combine the energy budgets of the UE and the BS. This acknowledges the fact that the energy expended on the uplink is by an energy-constrained UE, while that on the downlink is by the BS, which is not as constrained because it is often connected to the power grid. A single total energy budget is a special case of this formulation.

Additionally, the uplink and downlink training and the data transmission must occur within the coherence interval. Let $C\left(\alpha, T_{\rm p}^{\rm UL}, T_{\rm p}^{\rm DL}\right)$ denote the ergodic rate. Then, the ergodic rate-maximization problem can be formulated as follows:

$$\mathcal{P}_0: \max_{\alpha \in [0,1], T_{\rm p}^{\rm UL} \geq N, T_{\rm p}^{\rm DL} \geq 1} \left\{C\left(\alpha, T_{\rm p}^{\rm UL}, T_{\rm p}^{\rm DL}\right)\right\}, \tag{6}$$

$$\text{s.t.} \quad \rho_{\rm p}^{\rm UL} T_{\rm p}^{\rm UL} = E_{\rm UE}, \tag{7}$$

$$\rho_{\rm d}^{\rm DL} T_{\rm d}^{\rm DL} = \alpha E_{\rm BS}, \tag{8}$$

$$\rho_{\rm p}^{\rm DL} T_{\rm p}^{\rm DL} = (1-\alpha) E_{\rm BS}, \tag{9}$$

$$T_{\rm p}^{\rm UL} + T_{\rm d}^{\rm DL} + T_{\rm p}^{\rm DL} = T_{\rm c}. \tag{10}$$

III. Optimal Solution

We first derive in Section III-A a tractable lower bound for $C\left(\alpha, T_{\rm p}^{\rm UL}, T_{\rm p}^{\rm DL}\right)$, which accounts for the channel estimation errors in both uplink and downlink training phases, assuming that $g(\widehat{\mathbf{h}})$ is a Gaussian RV. Next, in Section III-B, we derive closed-form expressions for the mean and the variance of $g(\widehat{\mathbf{h}})$. In Section III-C, we prove that $g(\widehat{\mathbf{h}})$ is asymptotically Gaussian as N increases. In Section III-D, we solve $\mathcal{P}_0$.

A. Tractable Lower Bound on $C\left(\alpha, T_p^{UL}, T_p^{DL}\right)$

Let $g_{\rm e}(\widehat{\mathbf{h}})$ be the estimation error in estimating $g(\widehat{\mathbf{h}})$. Thus, $g_{\rm e}(\widehat{\mathbf{h}}) = g(\widehat{\mathbf{h}}) - \widehat{g}(\widehat{\mathbf{h}})$. Then, we can rewrite the received downlink data signal in (5) as

$$y_{\rm d}^{\rm DL}(k) = \sqrt{\rho_{\rm d}^{\rm DL}} \widehat{g}(\widehat{\mathbf{h}}) x_{\rm d}^{\rm DL}(k) + \sqrt{\rho_{\rm d}^{\rm DL}} g_{\rm e}(\widehat{\mathbf{h}}) x_{\rm d}^{\rm DL}(k) + n_{\rm d}^{\rm DL}(k). \tag{11}$$

Consider the downlink pilot signal in (3). Let $\tilde{y}_{\rm p}^{\rm DL} = \left(\mathbf{x}_{\rm p}^{\rm DL}\right)^H \mathbf{y}_{\rm p}^{\rm DL} / \left(\sqrt{\rho_{\rm p}^{\rm DL}} T_{\rm p}^{\rm DL}\right)$. Assuming that $g(\widehat{\mathbf{h}})$ is a Gaussian RV, the minimum mean square error (MMSE) estimate $\widehat{g}(\widehat{\mathbf{h}})$ of $g(\widehat{\mathbf{h}})$ is given by [25, Ch. 12.3]

$$\begin{aligned}\widehat{g}(\widehat{\mathbf{h}}) = &\frac{\mathrm{Cov}\left[g(\widehat{\mathbf{h}}), \tilde{y}_{\rm p}^{\rm DL}\right]}{\mathrm{Var}\left[\tilde{y}_{\rm p}^{\rm DL}\right]} \tilde{y}_{\rm p}^{\rm DL} + \mathbb{E}\left[g(\widehat{\mathbf{h}})\right] \\ &- \frac{\mathrm{Cov}\left[g(\widehat{\mathbf{h}}), \tilde{y}_{\rm p}^{\rm DL}\right]}{\mathrm{Var}\left[\tilde{y}_{\rm p}^{\rm DL}\right]} \mathbb{E}\left[\tilde{y}_{\rm p}^{\rm DL}\right].\end{aligned} \tag{12}$$

Let $\mathbb{E}\left[g(\widehat{\mathbf{h}})\right] = \mu_g$ and $\mathrm{Var}\left[g(\widehat{\mathbf{h}})\right] = \sigma_g^2$. We can show that $\mathrm{Cov}\left[g(\widehat{\mathbf{h}}), \tilde{y}_{\rm p}^{\rm DL}\right] = \sigma_g^2$, $\mathrm{Var}\left[\tilde{y}_{\rm p}^{\rm DL}\right] = \sigma_g^2 + \sigma^2 / \left(\rho_{\rm p}^{\rm DL} T_{\rm p}^{\rm DL}\right)$, and $\mathbb{E}\left[\tilde{y}_{\rm p}^{\rm DL}\right] = \mu_g$. Substituting these expressions in (12) yields

$$\widehat{g}(\widehat{\mathbf{h}}) = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma^2}{\rho_{\rm p}^{\rm DL} T_{\rm p}^{\rm DL}}} \tilde{y}_{\rm p}^{\rm DL} + \left(1 - \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma^2}{\rho_{\rm p}^{\rm DL} T_{\rm p}^{\rm DL}}}\right) \mu_g. \tag{13}$$

Since $\widehat{g}(\widehat{\mathbf{h}})$ is an MMSE estimate of $g(\widehat{\mathbf{h}})$, we know that $\mathbb{E}\left[g_{\text{e}}(\widehat{\mathbf{h}})|\widehat{g}(\widehat{\mathbf{h}})\right] = 0$ [26]. Furthermore, $\mathbb{E}\left[x_{\text{d}}^{\text{DL}}(k)\, g_{\text{e}}^*(\widehat{\mathbf{h}})|\widehat{g}(\widehat{\mathbf{h}})\right] = 0$ since $x_{\text{d}}^{\text{DL}}(k)$ is independent of $g_{\text{e}}(\widehat{\mathbf{h}})$. Let $x_{\text{d}}^{\text{DL}}(k) \sim \mathcal{CN}(0,1)$. Then, $C\left(\alpha, T_{\text{p}}^{\text{UL}}, T_{\text{p}}^{\text{DL}}\right)$ can be lower bounded as [26], [27]:

$$\begin{aligned} &C\left(\alpha, T_{\text{p}}^{\text{UL}}, T_{\text{p}}^{\text{DL}}\right) \\ &\geq C_{\text{LB}}\left(\alpha, T_{\text{p}}^{\text{UL}}, T_{\text{p}}^{\text{DL}}\right) \\ &= \frac{T_{\text{d}}^{\text{DL}}}{T_{\text{c}}} \mathbb{E}_{\widehat{g}(\widehat{\mathbf{h}})}\left[\log_2\left(1 + \frac{\rho_{\text{d}}^{\text{DL}}|\widehat{g}(\widehat{\mathbf{h}})|^2}{\rho_{\text{d}}^{\text{DL}}\mathbb{E}\left[|g_{\text{e}}(\widehat{\mathbf{h}})|^2|\widehat{g}(\widehat{\mathbf{h}})\right] + \sigma^2}\right)\right], \end{aligned} \tag{14}$$

where the expectation is over $\widehat{g}(\widehat{\mathbf{h}})$. The above expression arises when the term $\sqrt{\rho_{\text{d}}^{\text{DL}}} g_{\text{e}}(\widehat{\mathbf{h}}) x_{\text{d}}^{\text{DL}}(k) + n_{\text{d}}^{\text{DL}}(k)$ in (11) is treated as the effective noise, which is uncorrelated with the first term given $\widehat{g}(\widehat{\mathbf{h}})$.

Henceforth, we focus on the case where $T_{\text{d}}^{\text{DL}} > 1$ since the data rate in (14) scales linearly with T_{d}^{DL}.

To further simplify (14), we compute $\mathbb{E}\left[|g_{\text{e}}(\widehat{\mathbf{h}})|^2|\widehat{g}(\widehat{\mathbf{h}})\right]$ and $\mathbb{E}\left[|\widehat{g}(\widehat{\mathbf{h}})|^2\right]$ in closed-form below.

Lemma 1: $\mathbb{E}\left[|g_{\text{e}}(\widehat{\mathbf{h}})|^2|\widehat{g}(\widehat{\mathbf{h}})\right] = \sigma_g^2\sigma^2/\left(\rho_{\text{p}}^{\text{DL}} T_{\text{p}}^{\text{DL}} \sigma_g^2 + \sigma^2\right)$. And, $\mathbb{E}\left[|\widehat{g}(\widehat{\mathbf{h}})|^2\right] = \sigma_g^2 - \frac{\sigma_g^2\sigma^2}{\rho_{\text{p}}^{\text{DL}} T_{\text{p}}^{\text{DL}}\sigma_g^2 + \sigma^2} + |\mu_g|^2$.

Proof: The proof is given in Appendix A. ■

Applying Lemma 1 in (14) and simplifying yields

$$C_{\text{LB}}\left(\alpha, T_{\text{p}}^{\text{UL}}, T_{\text{p}}^{\text{DL}}\right) = \frac{T_{\text{d}}^{\text{DL}}}{T_{\text{c}}} \mathbb{E}\left[\log_2\left(1 + \text{SNR}_{\text{eff}}|\widehat{g}_{\text{norm}}(\widehat{\mathbf{h}})|^2\right)\right], \tag{15}$$

where $|\widehat{g}_{\text{norm}}(\widehat{\mathbf{h}})|^2 = |\widehat{g}(\widehat{\mathbf{h}})|^2 / \left(\sigma_g^2 - \frac{\sigma_g^2\sigma^2}{\rho_{\text{p}}^{\text{DL}} T_{\text{p}}^{\text{DL}}\sigma_g^2 + \sigma^2} + |\mu_g|^2\right)$ and

$$\text{SNR}_{\text{eff}} \triangleq \frac{\rho_{\text{d}}^{\text{DL}}\rho_{\text{p}}^{\text{DL}} T_{\text{p}}^{\text{DL}}\sigma_g^2\left(\sigma_g^2 + |\mu_g|^2\right) + \rho_{\text{d}}^{\text{DL}}|\mu_g|^2\sigma^2}{\rho_{\text{d}}^{\text{DL}}\sigma_g^2\sigma^2 + \rho_{\text{p}}^{\text{DL}} T_{\text{p}}^{\text{DL}}\sigma_g^2\sigma^2 + \sigma^4}. \tag{16}$$

Since $\mathbb{E}\left[|\widehat{g}_{\text{norm}}(\widehat{\mathbf{h}})|^2\right] = 1$, which follows from Lemma 1, SNR_{eff} can be interpreted as the *effective SNR* in the downlink. It is a function of T_{p}^{UL}, T_{p}^{DL}, and α.

B. Expressions for Mean μ_g and Variance σ_g^2 of $g(\widehat{\mathbf{h}})$

Let $\tilde{y}_{\text{p},i}^{\text{UL}} = \left[\mathbf{\Phi}^{\text{UL}}(:,i)\right]^H \mathbf{y}_{\text{p}}^{\text{UL}} / \left(\sqrt{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}\right)$. Let $\hat{h}_i$ be the LMMSE estimate of h_i. Then, as in (12), $\hat{h}_i$ is given by

$$\hat{h}_i = \frac{\beta_{\text{u}}\beta_{\text{v}}}{\beta_{\text{u}}\beta_{\text{v}} + \frac{\sigma^2}{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}} \tilde{y}_{\text{p},i}^{\text{UL}}. \tag{17}$$

Substituting the simplified expression for $\tilde{y}_{\text{p},i}^{\text{UL}}$ from (1) and rearranging, we get

$$\hat{h}_i = \gamma h_i + \epsilon_i, \tag{18}$$

where $\gamma = \beta_{\text{u}}\beta_{\text{v}} / \left(\beta_{\text{u}}\beta_{\text{v}} + \frac{\sigma^2}{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}\right)$ and $\epsilon_i = \gamma\left[\mathbf{\Phi}^{\text{UL}}(:,i)\right]^H \mathbf{n}^{\text{UL}} / \left(\sqrt{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}\right)$. Note that ϵ_i and ϵ_k are uncorrelated for all $i \neq k$. This is because $\mathbb{E}\left[\epsilon_i\epsilon_k^*\right] = \gamma^2\sigma^2\left[\mathbf{\Phi}^{\text{UL}}(:,i)\right]^H \mathbf{\Phi}^{\text{UL}}(:,k) / \left(\rho_{\text{p}}^{\text{UL}}\left(T_{\text{p}}^{\text{UL}}\right)^2\right)$ and the orthogonality of the columns of $\mathbf{\Phi}^{\text{UL}}$ implies that $\left[\mathbf{\Phi}^{\text{UL}}(:,i)\right]^H \mathbf{\Phi}^{\text{UL}}(:,k) = 0$ for $i \neq k$.

From the definition of $\hat{\phi}_i$ in (2), we get

$$\hat{\phi}_i = \frac{\hat{h}_i^*}{|\hat{h}_i|} = \frac{\gamma h_i^* + \epsilon_i^*}{|\gamma h_i + \epsilon_i|}. \tag{19}$$

We note that noise appears in both the numerator and the denominator of (19), which makes the derivation of μ_g and σ_g^2 intractable. To get a tractable expression for $\hat{\phi}_i$, we use the following approximation, which is motivated by considering the high SNR regime during the uplink training phase. In this regime, $\gamma \to 1$ and $\epsilon_i \to 0$. Then,

$$\hat{\phi}_i \approx \frac{\gamma h_i^* + \epsilon_i^*}{|h_i|}. \tag{20}$$

We shall assess its accuracy in the low SNR regime below and also in Section V. Substituting (20) in the expression for $g(\widehat{\mathbf{h}})$ in (4) and rearranging terms yield

$$g(\widehat{\mathbf{h}}) \approx \sum_{i=1}^{N}\left(\gamma|h_i| + \epsilon_i^* e^{j\angle h_i}\right). \tag{21}$$

An alternate approximation can be obtained by substituting $\gamma = 1$ in the numerator as well as the denominator of (19). However, it is less accurate.

1) Accuracy of Approximation: In Figure 2, we plot the MSE $\mathbb{E}\left[\left|\sum_{i=1}^{N}\frac{\gamma h_i^* + \epsilon_i^*}{|\gamma h_i + \epsilon_i h_i|}h_i - \sum_{i=1}^{N}\left(\gamma|h_i| + \epsilon_i^* e^{j\angle h_i}\right)\right|^2\right]$ between the unapproximated effective channel gain $g(\widehat{\mathbf{h}}) = \sum_{i=1}^{N}\frac{\gamma h_i^* + \epsilon_i^*}{|\gamma h_i + \epsilon_i h_i|}h_i$, obtained by substituting (19) in (4), and its approximation in (21) as a function of $E_{\text{UE}}\beta_{\text{u}}\beta_{\text{v}}/\sigma^2$ for the two channel models. It is normalized by $\mathbb{E}\left[\left|\sum_{i=1}^{N}\frac{\gamma h_i^* + \epsilon_i^*}{|\gamma h_i + \epsilon_i h_i|}h_i\right|^2\right]$, which is the mean square value of the unapproximated $g(\widehat{\mathbf{h}})$. The simulation is done for $N = 512$ and $(\beta_{\text{u}}\beta_{\text{v}}/\sigma^2) = 0$ dB. The expectations are computed numerically by Monte Carlo simulations for 5×10^4 uplink noise samples and fading states. The normalized MSE (NMSE) decreases as E_{UE} increases for both channel models. It is less than 10% when $E_{\text{UE}}\beta_{\text{u}}\beta_{\text{v}}/\sigma^2$ is as small as 5 dB for both models.

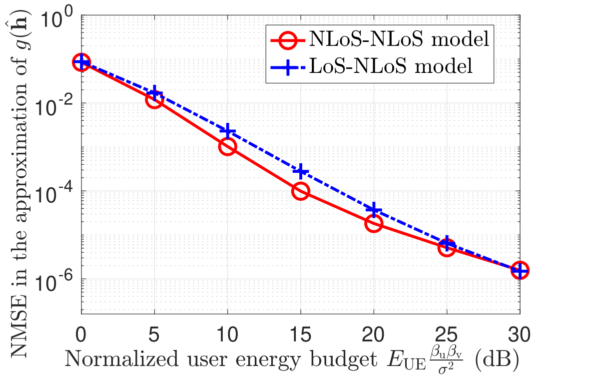


Fig. 2. Normalized MSE of approximation of $g(\widehat{\mathbf{h}})$ as a function of $E_{\text{UE}}\beta_{\text{u}}\beta_{\text{v}}/\sigma^2$ $\left(N = 64 \times 8,\ d_{\text{V}} = d_{\text{H}} = \frac{1}{4},\ \varphi_{\text{az}} = 36°, \text{ and } \varphi_{\text{el}} = 60°\right)$.

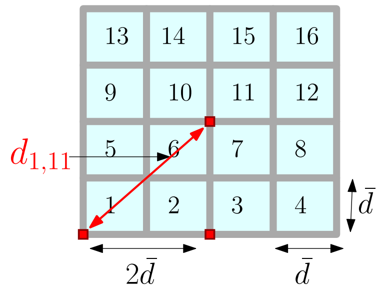


Fig. 3. Illustration of an RIS plane that shows indexing of reflecting elements and their inter-element distance $(N = 4 \times 4)$.

2) Expressions: We now derive μ_g and σ_g^2 for the two channel models.

Lemma 2: For the LoS-NLoS model, μ_g and σ_g^2 are given by

$$\mu_g = \frac{\sqrt{\pi}}{2} N\gamma\sqrt{\beta_{\mathrm{u}}\beta_{\mathrm{v}}}, \tag{22a}$$

$$\sigma_g^2 = \frac{\pi}{4}\gamma^2 \sum_{i=1}^{N}\sum_{k=1}^{N} \beta_{\mathrm{u}}\beta_{\mathrm{v}} \mathcal{F}\left(-\frac{1}{2}, -\frac{1}{2}; 1; [\mathbf{R}(i,k)]^2\right) + N\frac{\gamma^2\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}} T_{\mathrm{p}}^{\mathrm{UL}}} - \frac{\pi}{4} N^2\gamma^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}, \tag{22b}$$

where $\mathcal{F}(.;.;.)$ is the hypergeometric function [28, Ch. 15]. For the NLoS-NLoS model, the corresponding expressions are

$$\mu_g = \frac{\pi}{4} N\gamma\sqrt{\beta_{\mathrm{u}}\beta_{\mathrm{v}}}, \tag{23a}$$

$$\sigma_g^2 = \frac{\pi^2}{16}\gamma^2 \sum_{i=1}^{N}\sum_{k=1}^{N} \beta_{\mathrm{u}}\beta_{\mathrm{v}} \left[\mathcal{F}\left(-\frac{1}{2}, -\frac{1}{2}; 1; [\mathbf{R}(i,k)]^2\right)\right]^2 + N\frac{\gamma^2\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}} T_{\mathrm{p}}^{\mathrm{UL}}} - \frac{\pi^2}{16} N^2\gamma^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}. \tag{23b}$$

Proof: The proof is given in Appendix B. ■

C. Asymptotic Gaussianity of $g(\widehat{\mathbf{h}})$

For ease of exposition, we focus on $d_{\mathrm{H}} = d_{\mathrm{V}} = \bar{d}$. A similar, but more involved, proof also holds for $d_{\mathrm{H}} \neq d_{\mathrm{V}}$. Let the i^{th} summand of $g(\widehat{\mathbf{h}})$ in (21) be denoted by χ_i. Then,

$$\chi_i = \gamma|h_i| + \epsilon_i^* e^{j\angle h_i}. \tag{24}$$

Let $d_{i,k}$ be the distance between the points $\left(\mathrm{mod}\,(i-1, N_{\mathrm{H}})\,\bar{d}, \lfloor (i-1)/N_{\mathrm{H}} \rfloor \bar{d}\right)$ and $(\mathrm{mod}(k-1, N_{\mathrm{H}})\,\bar{d}, \lfloor (k-1)/N_{\mathrm{H}} \rfloor \bar{d})$ in the RIS plane. These points correspond to the left-bottom corners of the i^{th} and the k^{th} RIS elements. For example, $d_{1,11}$ is shown in Figure 3.

The normalized correlation coefficient between χ_i and χ_k is defined as

$$\rho(\chi_i, \chi_k) = \frac{\mathbb{E}[\chi_i, \chi_k^*] - \mathbb{E}[\chi_i]\,\mathbb{E}[\chi_k^*]}{\sqrt{\mathrm{Var}[\chi_i]\,\mathrm{Var}[\chi_k]}}. \tag{25}$$

Let

$$\rho_{\text{sup}}(n) \triangleq \sup_{k \in \mathcal{S}_n} \{|\rho(\chi_1, \chi_k)|\}, \tag{26}$$

where $\sup\{.\}$ denotes the supremum and $\mathcal{S}_n$ is the set of indices of the RIS elements that satisfy $d_{1,k} \geq n\bar{d}$. For example, in Figure 3, $\mathcal{S}_2 = \{3, 4, 7, 8, \ldots, 16\}$. Thus, when $\rho_{\text{sup}}(n) \to 0$, $\rho(\chi_1, \chi_k) \to 0$ for all $k \in \mathcal{S}_n$.

Definition [23]: A sequence $\{\chi_i\}_{i=1}^{N}$ is said to be ρ-mixing when $\chi_1, \ldots, \chi_N$ are asymptotically independent, i.e., $\rho_{\text{sup}}(n) \to 0$ as $n \to \infty$.

We now derive the expressions for $\rho_{\text{sup}}(n)$.

Lemma 3: Let i_{sup} be the index of an RIS element such that $\rho_{\text{sup}}(n) = |\rho(\chi_1, \chi_{i_{\text{sup}}})|$. Then, $\rho_{\text{sup}}(n)$ for the LoS-NLoS and NLoS-NLoS models is given by

$$\rho_{\text{sup}}(n) = \begin{cases} \dfrac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}\left[\mathcal{F}\left(-\frac{1}{2}, -\frac{1}{2}; 1; [\mathbf{R}(1, i_{\text{sup}})]^2\right) - 1\right]}{4\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}} T_{\mathrm{p}}^{\mathrm{UL}}}\right)}, & \text{for LoS-NLoS model,} \quad (27a) \\[2ex] \dfrac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}\left[\left[\mathcal{F}\left(-\frac{1}{2}, -\frac{1}{2}; 1; [\mathbf{R}(1, i_{\text{sup}})]^2\right)\right]^2 - 1\right]}{16\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{16} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}} T_{\mathrm{p}}^{\mathrm{UL}}}\right)}, & \text{for NLoS-NLoS model.} \quad (27b) \end{cases}$$

Proof: The proof is given in Appendix C. ■

Using Lemma 3, we now prove that the sequence $\{\chi_i\}_{i=1}^{N}$ is ρ-mixing.

Lemma 4: The sequence $\{\chi_i\}_{i=1}^{N}$ is ρ-mixing when $N \to \infty$.

Proof: The proof is given in Appendix D. ■

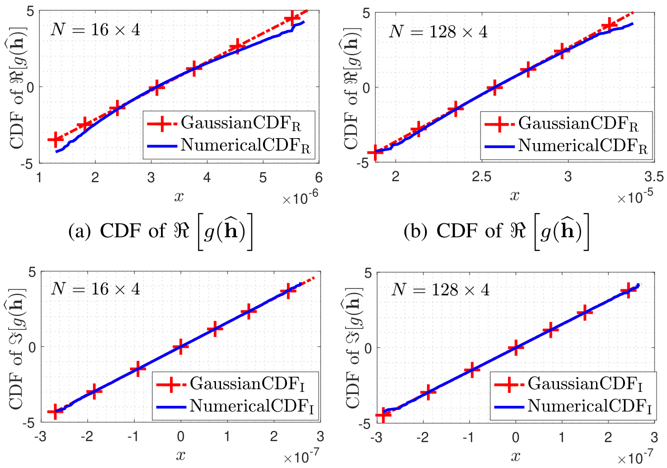
From [23, Th. 0], the sum of a ρ-mixing sequence $\sum_{i=1}^{n}\chi_i$ becomes a Gaussian RV as $n \to \infty$ when the following three conditions hold: (a) $\mathrm{Var}[\chi_i] < \infty$, (b) $\lim_{n\to\infty} \mathrm{Var}[\sum_{i=1}^{n}\chi_i] \to \infty$, and (c) $\sum_{i=1}^{\infty}\rho_{\text{sup}}(2^i) < \infty$. Lemma 4 leads to the following key theorem.

Theorem 1: $g(\widehat{\mathbf{h}}) = \sum_{i=1}^{N}\chi_i$ becomes a complex Gaussian RV as $N \to \infty$.

Proof: The proof is given in Appendix E. ■

Note that the real and imaginary parts of $g(\widehat{\mathbf{h}}) = \sum_1^N \chi_i$ are independent. This is because the real and imaginary parts of χ_i are independent since h_i and ϵ_i are independent of each other and both are circularly symmetric.

1) Numerical Assessment: We first consider the NLoS-NLoS model. Let $\mathbb{E}\left[\Re\left[g(\widehat{\mathbf{h}})\right]\right] = \mu_{\mathrm{R}}$ and $\mathrm{Var}\left[\Re\left[g(\widehat{\mathbf{h}})\right]\right] = \sigma_{\mathrm{R}}^2$. When $\Re\left[g(\widehat{\mathbf{h}})\right]$ is Gaussian, we know that $F_{\Re[g(\widehat{\mathbf{h}})]}(x) = 1 - Q((x-\mu_{\mathrm{R}})/\sigma_{\mathrm{R}})$, where $Q(.)$ is the Gaussian-Q function. Figures 4(a) and 4(b) compare $Q^{-1}\left(1 - F_{\Re[g(\widehat{\mathbf{h}})]}(x)\right)$ (denoted by NumericalCDF$_{\mathrm{R}}$) with $(x-\mu_{\mathrm{R}})/\sigma_{\mathrm{R}}$ (denoted by GaussianCDF$_{\mathrm{R}}$) for two values of N. Here, $F_{\Re[g(\widehat{\mathbf{h}})]}(x)$, μ_{R}, and σ_{R} are determined numerically from Monte Carlo simulations. The closer $Q^{-1}\left(1 - F_{\Re[g(\widehat{\mathbf{h}})]}(x)\right)$ is to the straight line $(x-\mu_{\mathrm{R}})/\sigma_{\mathrm{R}}$, the more Gaussian the distribution is. This method is referred to as the Gaussian probability paper test [29, Ch. 6]. We see



(a) CDF of $\Re\left[g(\widehat{\mathbf{h}})\right]$ (b) CDF of $\Re\left[g(\widehat{\mathbf{h}})\right]$

(c) CDF of $\Im\left[g(\widehat{\mathbf{h}})\right]$ (d) CDF of $\Im\left[g(\widehat{\mathbf{h}})\right]$

Fig. 4. CDF of real and imaginary parts of $g(\widehat{\mathbf{h}})$ on Gaussian probability paper for two values of N ($\bar{d} = 1/4$).

that as N increases, the two curves are closer to each other over a wider range.

The corresponding results for $\Im\left[g(\widehat{\mathbf{h}})\right]$ are shown in Figures 4(c) and 4(d). In them, we compare $Q^{-1}\left(1 - F_{\Im[g(\widehat{\mathbf{h}})]}(x)\right)$ (denoted by NumericalCDF$_{\mathrm{I}}$) with $(x-\mu_{\mathrm{I}})/\sigma_{\mathrm{I}}$ (denoted by GaussianCDF$_{\mathrm{I}}$). Here, $\mathbb{E}\left[\Im[g(\widehat{\mathbf{h}})]\right] = \mu_{\mathrm{I}}$ and $\mathrm{Var}\left[\Im[g(\widehat{\mathbf{h}})]\right] = \sigma_{\mathrm{I}}^2$. We see that NumericalCDF$_{\mathrm{I}}$ converges to GaussianCDF$_{\mathrm{I}}$ even for small values of N. This follows from (21) because $\Im[g(\widehat{\mathbf{h}})] = \sum_{i=1}^{N} \Im\left[\epsilon_i^* e^{j\angle h_i}\right]$, and h_i and ϵ_i are independent of each other and both are circularly symmetric. Thus, Figure 4 shows that Theorem 1 can be applied even for small N. Similar results arise for the LoS-NLoS model. We skip them to conserve space.

D. Optimal Training Durations and Transmit Powers

We first find the optimal power allocation factor as a function of $T_{\mathrm{d}}^{\mathrm{DL}}$.

Lemma 5: Given $T_{\mathrm{d}}^{\mathrm{DL}}$, the optimal value $\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)$ of α is given by

$$\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) = \zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) - \sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)\left[\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) - \kappa\right]}, \tag{28}$$

where

$$\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) = \frac{E_{\mathrm{BS}}\sigma_g^2 + \sigma^2}{E_{\mathrm{BS}}\sigma_g^2}\frac{T_{\mathrm{d}}^{\mathrm{DL}}}{T_{\mathrm{d}}^{\mathrm{DL}} - 1}, \tag{29}$$

$$\kappa = \frac{E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2 + |\mu_g|^2\right) + |\mu_g|^2\sigma^2}{E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2 + |\mu_g|^2\right)}. \tag{30}$$

Proof: The proof is given in Appendix F. ■

Substituting $\rho_{\mathrm{d}}^{\mathrm{DL}} = \alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)E_{\mathrm{BS}}/T_{\mathrm{d}}^{\mathrm{DL}}$, $\rho_{\mathrm{p}}^{\mathrm{DL}}T_{\mathrm{p}}^{\mathrm{DL}} = \left[1 - \alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)\right]E_{\mathrm{BS}}$, and the expression of $\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)$ from (28) in (15) yields

$$\begin{aligned} &C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right), T_{\mathrm{p}}^{\mathrm{UL}}, T_{\mathrm{p}}^{\mathrm{DL}}\right) \\ &= \frac{T_{\mathrm{d}}^{\mathrm{DL}}}{T_{\mathrm{c}}}\mathbb{E}\left[\log_2\left(1 + |\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2\frac{E_{\mathrm{BS}}\left(\sigma_g^2 + |\mu_g|^2\right)}{\sigma^2\left(T_{\mathrm{d}}^{\mathrm{DL}} - 1\right)}\right.\right. \\ &\quad \left.\left.\times\left[\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)} - \sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) - \kappa}\right]^2\right)\right]. \end{aligned} \tag{31}$$

We prove the following monotonicity property of $C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right), T_{\mathrm{p}}^{\mathrm{UL}}, T_{\mathrm{p}}^{\mathrm{DL}}\right)$.

Theorem 2: $C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right), T_{\mathrm{p}}^{\mathrm{UL}}, T_{\mathrm{p}}^{\mathrm{DL}}\right)$ is a monotonically increasing function of $T_{\mathrm{d}}^{\mathrm{DL}}$.

Proof: The proof is given in Appendix G. ■

This can be explained as follows. When $T_{\mathrm{p}}^{\mathrm{UL}}$ decreases, the uplink channel estimates become less accurate, due to which the effective channel strength and SNR$_{\mathrm{eff}}$ decrease. However, this decrease is within a logarithmic term. On the other hand, the increase in $T_{\mathrm{d}}^{\mathrm{DL}}$ is in a pre-log factor in (15), which dominates. The same logic applies to $T_{\mathrm{p}}^{\mathrm{DL}}$.

It follows from Theorem 2 that the optimal uplink training duration $T_{\mathrm{p,opt}}^{\mathrm{UL}}$ and the optimal downlink training duration $T_{\mathrm{p,opt}}^{\mathrm{DL}}$ are given by

$$T_{\mathrm{p,opt}}^{\mathrm{UL}} = N \text{ and } T_{\mathrm{p,opt}}^{\mathrm{DL}} = 1, \tag{32}$$

as these are the minimum number of pilots required to estimate N cascaded channel gains in the uplink and a single effective channel gain in the downlink, respectively. Then, $T_{\mathrm{d}}^{\mathrm{DL}} = T_{\mathrm{c}} - N - 1$.

Corollary 1: The optimal value α_{opt} of the energy allocation factor is given by

$$\alpha_{\mathrm{opt}} = \frac{E_{\mathrm{BS}}\sigma_g^2 + \sigma^2}{E_{\mathrm{BS}}\sigma_g^2}\omega - \psi, \tag{33}$$

where $\omega = (T_{\mathrm{c}} - N - 1)/(T_{\mathrm{c}} - N - 2)$ and $\psi = \sqrt{\frac{\left[E_{\mathrm{BS}}\sigma_g^2+\sigma^2\right]\omega}{E_{\mathrm{BS}}\sigma_g^2}\left(\frac{\left[E_{\mathrm{BS}}\sigma_g^2+\sigma^2\right]\omega}{E_{\mathrm{BS}}\sigma_g^2} - \frac{E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)+|\mu_g|^2\sigma^2}{E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)}\right)}$. Furthermore, the optimal uplink pilot power $\rho_{\mathrm{p,opt}}^{\mathrm{UL}}$, the optimal downlink pilot power $\rho_{\mathrm{p,opt}}^{\mathrm{DL}}$, and the optimal downlink data power $\rho_{\mathrm{d,opt}}^{\mathrm{DL}}$ are given by

$$\rho_{\mathrm{p,opt}}^{\mathrm{UL}} = \frac{E_{\mathrm{UE}}}{N}, \tag{34}$$

$$\rho_{\mathrm{d,opt}}^{\mathrm{DL}} = \frac{E_{\mathrm{BS}}}{T_{\mathrm{c}} - N - 1}\left(\frac{E_{\mathrm{BS}}\sigma_g^2 + \sigma^2}{E_{\mathrm{BS}}\sigma_g^2}\omega - \psi\right), \tag{35}$$

$$\rho_{\mathrm{p,opt}}^{\mathrm{DL}} = E_{\mathrm{BS}}\left(1 - \frac{E_{\mathrm{BS}}\sigma_g^2 + \sigma^2}{E_{\mathrm{BS}}\sigma_g^2}\omega + \psi\right). \tag{36}$$

Proof: Equation (33) follows from (28). Similarly, (34), (35), and (36) follow from the constraints in (7), (8), and (9). ■

IV. Impact of Cascaded Channel Grouping

In grouping, the N reflecting elements are divided into K groups. Each group contains N/K elements. We assume N/K to be an integer. All the elements in a group use the same reflection coefficient [15]. The training and data transmission phases are now as follows.

A. Uplink Channel Estimation and RIS Configuration

The UE sends $T_{\mathrm{p}}^{\mathrm{UL}} \geq K$ pilots in the uplink with power $\rho_{\mathrm{p}}^{\mathrm{UL}}$. Then, as in (1), the received pilot vector at the BS is given by

$$\mathbf{y}_{\mathrm{G}} = \sqrt{\rho_{\mathrm{p}}^{\mathrm{UL}}}\boldsymbol{\Phi}_{\mathrm{G}}\left[G_1, \ldots, G_K\right]^T + \mathbf{n}_{\mathrm{G}}. \tag{37}$$

Here, $\mathbf{\Phi}_{\text{G}}$ is a $T_{\text{p}}^{\text{UL}} \times K$ matrix, $G_k = \sum_{i \in \mathcal{C}_k} h_i$ is the effective cascaded channel gain of the k^{th} group, $\mathcal{C}_k$ is the set of indices of the reflecting elements that belong to the k^{th} group, and $\mathbf{n}_{\text{G}} \sim \mathcal{CN}\left(0, \sigma^2 \mathbf{I}_K\right)$ is AWGN. As before, $\left(\mathbf{\Phi}_{\text{G}}\right)^H \mathbf{\Phi}_{\text{G}} = T_{\text{p}}^{\text{UL}} \mathbf{I}_K$ and $|\mathbf{\Phi}_{\text{G}}(i,k)| = 1$. It can be constructed by choosing K columns of a DFT matrix or a Hadamard matrix of size $T_{\text{p}}^{\text{UL}} \times T_{\text{p}}^{\text{UL}}$.

We assume that $G_1, \ldots, G_K$ are statistically identical. This is reasonable for large N. Then, $\mathbb{E}[G_k] = 0$. Let $\sigma_{\text{grp}}^2 = \text{Var}[G_k] = \sum_{i \in \mathcal{C}_k} \sum_{i' \in \mathcal{C}_k} \mathbb{E}[h_i h_{i'}^*]$. For any $1 \leq k \leq K$, we can show that

$$\sigma_{\text{grp}}^2 = \begin{cases} \sum_{i \in \mathcal{C}_k} \sum_{i' \in \mathcal{C}_k} \beta_{\text{u}} v_i v_{i'}^* \mathbf{R}(i,i'), & \text{for LoS-NLoS model}, \quad (38) \\ \sum_{i \in \mathcal{C}_k} \sum_{i' \in \mathcal{C}_k} \beta_{\text{u}} \beta_{\text{v}} \left[\mathbf{R}(i,i')\right]^2, & \text{for NLoS-NLoS model}. \quad (39) \end{cases}$$

As in (12), we can find the LMMSE estimate $\widehat{G}_k$ of G_k from $\tilde{y}_{\text{G},k} = \left[\mathbf{\Phi}_{\text{G}}(:,k)\right]^H \mathbf{y}_{\text{G}} / \left(\sqrt{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}\right)$ as $\widehat{G}_k = \gamma_{\text{G}} \tilde{y}_{\text{G},k}$, where $\gamma_{\text{G}} = \sigma_{\text{grp}}^2 / \left(\sigma_{\text{grp}}^2 + \frac{\sigma^2}{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}\right)$. Substituting the expression for $\tilde{y}_{\text{G},k}$ and simplifying yields

$$\widehat{G}_k = \gamma_{\text{G}} G_k + \epsilon_{\text{G},k}, \tag{40}$$

where $\epsilon_{\text{G},k} = \gamma_{\text{G}} \left[\mathbf{\Phi}_{\text{G}}(:,k)\right]^H \mathbf{n}_{\text{G}} / \left(\sqrt{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}}\right)$. The BS now sets the reflection coefficients of all the elements in group k as $\hat{\phi}_k = e^{j\angle \widehat{G}_k^*}$, and the effective channel gain is given as

$$g(\widehat{\mathbf{G}}) = \sum_{k=1}^{K} \hat{\phi}_k G_k = \sum_{k=1}^{K} e^{j\angle \widehat{G}_k^*} G_k, \tag{41}$$

where $\widehat{\mathbf{G}} = \left[\widehat{G}_1, \ldots, \widehat{G}_K\right]^T$. It is a function of the noise during the uplink training phase.

B. Optimal Training Durations and α

We focus on the LoS-NLoS model. The derivation of the mean and the variance of $g(\widehat{\mathbf{G}})$ is intractable for the NLoS-NLoS model because the correlation coefficient for the envelopes among the effective channels of the groups are not known to the best of our knowledge. For the LoS-NLoS model, we have $G_k \sim \mathcal{CN}\left(0, \sigma_{\text{grp}}^2\right)$ for all $k \in \{1, \ldots, K\}$. As in Section III-B, $\hat{\phi}_k \approx \left(\gamma_{\text{G}} G_k + \epsilon_{\text{G},k}\right) / |G_k|$. Substituting this in (41) and simplifying yields the following expression of $g(\widehat{\mathbf{G}})$:

$$g(\widehat{\mathbf{G}}) \approx \sum_{k=1}^{K} \left(\gamma_{\text{G}} |G_k| + \epsilon_{\text{G},k}^* e^{j\angle G_k}\right). \tag{42}$$

This leads to the following result about the mean and the variance of $g(\widehat{\mathbf{G}})$.

Lemma 6: The mean and the variance of $g(\widehat{\mathbf{G}})$ with grouping are given by

$$\mathbb{E}\left[g(\widehat{\mathbf{G}})\right] = \sqrt{\pi} K \gamma_{\text{G}} \frac{\sqrt{\sigma_{\text{grp}}^2}}{2}, \tag{43a}$$

$$\begin{aligned} \text{Var}\left[g(\widehat{\mathbf{G}})\right] &= \frac{\pi}{4} \sum_{k=1}^{K} \sum_{k'=1}^{K} \gamma_{\text{G}}^2 \sigma_{\text{grp}}^2 \\ &\times \mathcal{F}\left(-\frac{1}{2}, -\frac{1}{2}; 1; \left[\frac{\sum_{i \in \mathcal{C}_k} \sum_{i' \in \mathcal{C}_{k'}} \beta_{\text{u}} v_i v_{i'}^* \mathbf{R}(i,i')}{\sigma_{\text{grp}}^2}\right]^2\right) \\ &+ K \frac{\gamma_{\text{G}}^2 \sigma^2}{\rho_{\text{p}}^{\text{UL}} T_{\text{p}}^{\text{UL}}} - \frac{\pi}{4} K^2 \gamma_{\text{G}}^2 \sigma_{\text{grp}}^2. \end{aligned} \tag{43b}$$

Proof: The proof is given in Appendix H. ■

Substituting (43a) and (43b) in (12) gives the linear estimator for $g(\widehat{\mathbf{G}})$. We can prove the asymptotic Gaussianity of $g(\widehat{\mathbf{G}})$ for sufficiently large K in a manner similar to Section III-C. We skip the details to conserve space. With grouping, the same expression for the achievable rate in (31) applies except that μ_g and σ_g^2 are as given in (43a) and (43b), respectively. Similar to Theorem 2, we can prove that the achievable rate is an increasing function of T_{d}^{DL}. Therefore, the optimal uplink and downlink training durations are

$$T_{\text{p}}^{\text{UL}} = K \text{ and } T_{\text{p}}^{\text{DL}} = 1. \tag{44}$$

Substituting $T_{\text{d}}^{\text{DL}} = T_{\text{c}} - K - 1$ in (28) yields the expression for the optimal α.

Since $K \leq N$, the rate-maximizing uplink training duration is smaller than that without grouping. Thus, more time is available for data transmission. However, as the number of elements per group N/K increases, the SNR in the downlink decreases since $|g(\widehat{\mathbf{G}})|^2$ decreases. This is because of a decrease in the degrees of freedom available for combining the multiple reflected paths constructively.

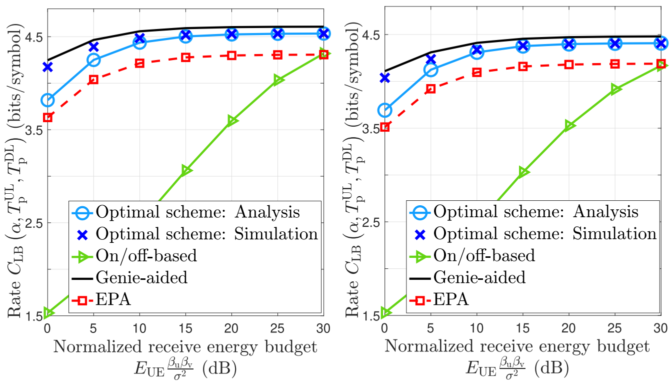
V. Numerical Results

We quantify the dependence of the rate $C_{\text{LB}}\left(\alpha, T_{\text{p}}^{\text{UL}}, T_{\text{p}}^{\text{DL}}\right)$, the optimal downlink pilot and data powers, and the extent of grouping on various system parameters and the training overhead. We consider the uniform planar array configuration for the RIS. We set $d_{\text{H}} = d_{\text{V}} = \bar{d} = 1/4$, $\varphi_{\text{az}} = 36°$, and $\varphi_{\text{el}} = 60°$. For $\bar{d} = 1/4$, we set $\beta_{\text{u}} \beta_{\text{v}} / \sigma^2 = 0$ dB.

The analytical curves for the achievable rate are obtained from (31), with the outer expectation being computed numerically. They are compared with the value measured from Monte Carlo simulations. In the simulations, $\hat{\phi}_i$ is computed as per (2) (without any high SNR approximation), μ_g and σ_g^2 are measured numerically from (4), and then the rate is computed as per (15) for a fixed α. Then, the rate-maximizing α is determined numerically by a bisection search method.

We benchmark the optimal power allocation in Section III-D with the following schemes:

- *Equal Power Allocation (EPA):* The BS employs equal power for the downlink data and the pilot transmissions, i.e., $\rho_{\text{p}}^{\text{DL}} = \rho_{\text{d}}^{\text{DL}} = E_{\text{BS}}/(T_{\text{c}} - N)$ [26]. Here, T_{d}^{DL} and α are determined numerically to maximize $C_{\text{LB}}\left(\alpha, T_{\text{p}}^{\text{UL}}, T_{\text{p}}^{\text{DL}}\right)$.
- *On/off-Based Scheme:* The RIS elements are turned on sequentially one by one during the uplink training. This scheme is employed in [7] and [8]. Then, we get



(a) LoS-NLoS model (b) NLoS-NLoS model

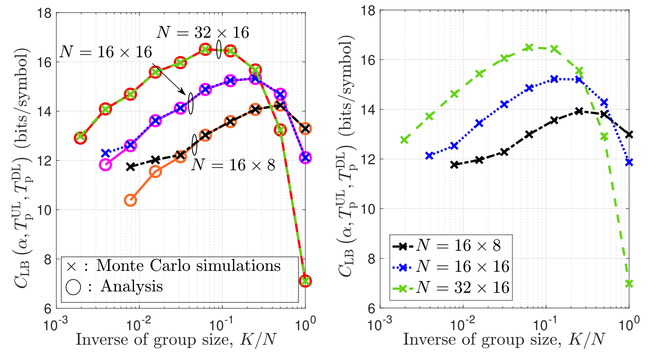
Fig. 5. Rate as a function of $E_{\rm UE}\beta_{\rm u}\beta_{\rm v}/\sigma^2$ for the two channel models $\left(N = 64 \times 8,\ E_{\rm BS}\beta_{\rm u}\beta_{\rm v}/\sigma^2 = 10 \text{ dB},\ T_{\rm c} = 800, \text{ and no grouping}\right)$.

$\gamma = \beta_{\rm u}\beta_{\rm v}/\left(\beta_{\rm u}\beta_{\rm v} + \frac{\sigma^2}{\rho_{\rm p}^{\rm UL}}\right)$ and $\epsilon_i = \gamma \mathbf{n}^{\rm UL}(i)/\sqrt{\rho_{\rm p}^{\rm UL}}$. We substitute these two expressions in (18). The rate is computed numerically from (15).

- *Genie-Aided Scheme:* $g(\widehat{\mathbf{h}})$ is assumed to be perfectly known to the UE and the entire BS energy is used for the downlink data transmission. This assumption is made in [8]. Then, from (5), the rate can be shown to be equal to $T_{\rm d}^{\rm DL}\mathbb{E}_{g(\widehat{\mathbf{h}})}\left[\log_2\left(1 + \rho_{\rm d}^{\rm DL}|g(\widehat{\mathbf{h}})|^2/\sigma^2\right)\right]/T_{\rm c}$, which we compute numerically.

Figure 5 compares the rates of the optimal scheme, EPA, on/off-based scheme, and genie-aided scheme for the LoS-NLoS model in Figure 5(a) and for the NLoS-NLoS model in Figure 5(b). For a given $E_{\rm UE}$, the accuracy of the uplink estimates in the on/off-based scheme is lower compared to DFT-based training. And, EPA requires a larger training duration in the downlink than the optimal scheme, which reduces the available time for data transmission. Hence, the optimal scheme outperforms the on/off-based scheme and EPA. The rate of the genie-aided scheme serves as an upper bound because all the BS energy is spent on data transmission and the channel estimate in the downlink is assumed to be noise-free. As $E_{\rm UE}$ increases, the rates of all schemes increase because the channel estimates at the BS become more accurate. The gap between the analysis and simulation curves of the optimal scheme decreases as $E_{\rm UE}$ increases and becomes negligible for $E_{\rm UE}\beta_{\rm u}\beta_{\rm v}/\sigma^2 \geq 10$ dB. For small $E_{\rm UE}$, the analytical expression lower bounds the numerical value. Similarly, as $E_{\rm BS}$ increases, the rate increases since the accuracy of the downlink channel gain estimate increases and so does the effective SNR. We do not show the dependence of the rate on $E_{\rm BS}$ to avoid clutter. The trends are qualitatively similar. Due to the deterministic nature of the BS-RIS link, the LoS-NLoS model achieves a higher rate than the NLoS-NLoS model.

The impact of grouping of the RIS elements on the rate is shown for the LoS-NLoS model in Figure 6(a) and for the NLoS-NLoS model in Figure 6(b). As K/N increases, the rate initially increases because fewer elements per group leads to an increase in the SNR. Then, the rate peaks and decreases thereafter. This is because the increase in the training overhead counteracts the increase in the SNR. The rate that the optimum grouping size achieves is 230% of that without



(a) LoS-NLoS model (b) NLoS-NLoS model

Fig. 6. Rate as a function of the inverse of the group size for different RIS planar sizes for the two channel models $\left(E_{\rm BS}\beta_{\rm u}\beta_{\rm v}/\left[(T_{\rm c} - K)\,\sigma^2\right] = \rho_{\rm p}^{\rm UL}\beta_{\rm u}\beta_{\rm v}/\sigma^2 = 7 \text{ dB and } T_{\rm c} = 800\right)$.

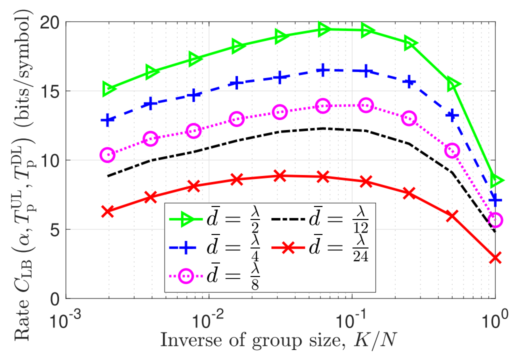


Fig. 7. Rate as a function of group size for different inter-element spacings $\left(N = 32 \times 16,\ E_{\rm BS}\beta_{\rm u}\beta_{\rm v}/\left[(T_{\rm c} - K)\,\sigma^2\right] = \rho_{\rm p}^{\rm UL}\beta_{\rm u}\beta_{\rm v}/\sigma^2 = 7 \text{ dB when } \bar{d} = 1/4,\ T_{\rm c} = 800, \text{ and LoS-NLoS model}\right)$.

grouping ($K/N = 1$). The trends are qualitatively similar for the two models. Figure 6(a) also plots the analytical results. They match match well with the simulation results for $K/N \geq 0.03$. For smaller K/N, there is a gap between the two because $g(\widehat{\mathbf{h}})$ does not contain sufficiently many terms for the Gaussian approximation to be accurate.

Figure 7 plots the rate as a function of K/N for different distances between the RIS elements. We see that as $\bar{d}$ increases, the rate increases for all values of K/N. This is because the pathloss $\beta_{\rm u}\beta_{\rm v}$ is proportional to $\bar{d}^4$ [22]. This dominates the decrease in the rate due to the decrease in the spatial correlation. The optimal group size turns out to be insensitive to $\bar{d}$ except when $\bar{d}$ is very small.

Figure 8 plots the normalized downlink pilot and data powers $\rho_{\rm p,opt}^{\rm DL}\beta_{\rm u}\beta_{\rm v}/\sigma^2$ and $\rho_{\rm d,opt}^{\rm DL}\beta_{\rm u}\beta_{\rm v}/\sigma^2$ as a function of $T_{\rm c}$ for two values of N. For both N, we see that substantially more power is allocated to the downlink pilots than the data. The data power is insensitive to both $T_{\rm c}$ and N because much more time is available for data transmission compared to downlink training, which needs only 1 symbol duration. Furthermore, as N increases, both $T_{\rm d}^{\rm DL}$ and $E_{\rm BS}$ decrease. Since $T_{\rm p}^{\rm DL}$ is fixed, a smaller power is allocated during training because of which $\rho_{\rm p,opt}^{\rm DL}$ decreases as N increases.

In Figure 9, we study the impact of the training overhead on the rate. We plot the rate as a function of the number of downlink training symbols for different numbers of the uplink training symbols. For a given $T_{\rm p}^{\rm DL}$, as $T_{\rm p}^{\rm UL}$ increases from its optimal value of 256, the rate decreases. When $T_{\rm p}^{\rm UL}$ increases, the uplink channel estimates become more accurate and the

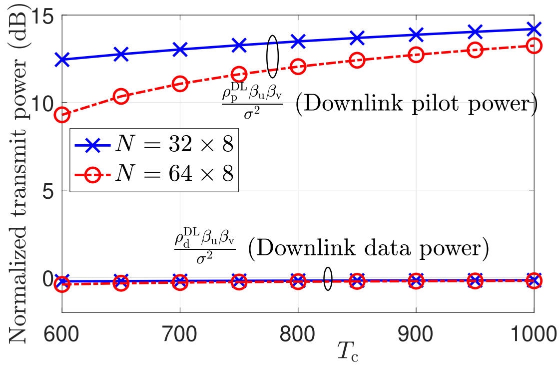


Fig. 8. Comparison of optimal downlink pilot power and data power as a function of T_c for different values of N $\left(E_{BS}\beta_u\beta_v/\left[(T_c - N)\,\sigma^2\right] = 0 \text{ dB}, E_{UE}\beta_u\beta_v/\sigma^2 = 25 \text{ dB, no grouping, and NLoS-NLoS model}\right)$.

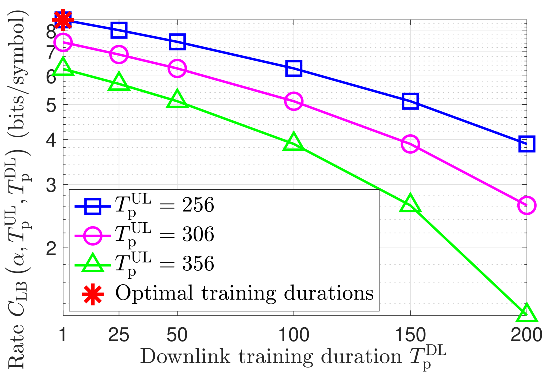


Fig. 9. Rate as a function of the downlink training duration for different values of the uplink training duration ($E_{BS}\beta_u\beta_v/\sigma^2 = 25$ dB, $E_{UE}\beta_u\beta_v/\sigma^2 = 25$ dB, $T_c = 600$, $N = 64 \times 4$, NLoS-NLoS model, and no grouping).

effective SNR SNR_{eff} in (15) increases. However, this increase is within a logarithmic term, while the decrease in the data duration due to the increased training overhead is in a pre-log factor. The latter dominates. Similarly, for a given T_p^{UL}, as T_p^{DL} increases from its optimal value 1, the rate decreases. This validates the optimality of the result in (32).

VI. Conclusion

We studied a training scheme for an RIS-aided TDD system, in which both BS and UE estimated the channel gains. We did so for two scenarios. While sufficient number of uplink pilots were employed to estimate the cascaded channel gains in the first scenario, fewer pilots were sent in the second scenario that used cascaded channel grouping. We characterized the effective downlink channel gain through the RIS in the presence of the channel estimation errors. For both LoS-NLoS and NLoS-NLoS channel models, we saw that it had a Gaussian distribution even in the presence of spatial correlation (due to closely-spaced RIS elements) and channel estimation errors. Following this, we derived a novel lower bound on the achievable rate that accounted for the cumulative impact of the uplink and downlink channel estimation errors on the RIS phase-shift configuration and data demodulation. The closed-form expressions for the optimal pilot durations and transmit powers that maximized the achievable rate under energy constraints at the BS and UE brought out the importance of boosting the pilot powers relative to the data power.

The proposed scheme achieved a larger rate compared to equal power allocation and the on/off-based scheme. Grouping led to a higher rate as it incurred a lower optimal training overhead and exploited spatial correlation. Extending this study to other spatial correlation models and fading models, such as Rician fading, is an interesting avenue for future work. Another avenue is analyzing the impact of channel estimation errors on multiple antenna, multi-user RIS systems and systems with active RISs.

Appendix

A. Brief Proof of Lemma 1

Since $\widehat{g}(\widehat{\mathbf{h}})$ and $g_e(\widehat{\mathbf{h}})$ are independent, $\mathbb{E}\left[|g_e(\widehat{\mathbf{h}})|^2|\widehat{g}(\widehat{\mathbf{h}})\right] = \mathbb{E}\left[|g_e(\widehat{\mathbf{h}})|^2\right]$. From [25, Ch. 12.3], we have $\mathbb{E}\left[|g_e(\widehat{\mathbf{h}})|^2\right] = \text{Var}\left[g(\widehat{\mathbf{h}})\right] - \left|\text{Cov}\left[g(\widehat{\mathbf{h}}), \tilde{y}_p^{DL}\right]\right|^2/\text{Var}\left[\tilde{y}_p^{DL}\right]$. Substituting the expressions for each term from Section III-A and simplifying yields $\mathbb{E}\left[|g_e(\widehat{\mathbf{h}})|^2\right] = \sigma_g^2\sigma^2/\left(\rho_p^{DL}T_p^{DL}\sigma_g^2 + \sigma^2\right)$.

From $g(\widehat{\mathbf{h}}) = \widehat{g}(\widehat{\mathbf{h}}) + g_e(\widehat{\mathbf{h}})$, we get $\text{Var}\left[\widehat{g}(\widehat{\mathbf{h}})\right] = \sigma_g^2 - \mathbb{E}\left[|g_e(\widehat{\mathbf{h}})|^2\right]$. Since $\widehat{g}(\widehat{\mathbf{h}})$ is unbiased, $\mathbb{E}\left[\widehat{g}(\widehat{\mathbf{h}})\right] = \mu_g$. Then, $\mathbb{E}\left[|\widehat{g}(\widehat{\mathbf{h}})|^2\right] = \text{Var}\left[\widehat{g}(\widehat{\mathbf{h}})\right] + \left|\mathbb{E}\left[\widehat{g}(\widehat{\mathbf{h}})\right]\right|^2 = \sigma_g^2 - \frac{\sigma_g^2\sigma^2}{\rho_p^{DL}T_p^{DL}\sigma_g^2+\sigma^2} + |\mu_g|^2$.

B. Proof of Lemma 2

Taking expectation on both sides of (21) yields

$$\mu_g = \sum_{i=1}^{N}\left(\gamma\mathbb{E}\left[|h_i|\right] + \mathbb{E}\left[\epsilon_i^* e^{j\angle h_i}\right]\right) = \gamma N\mathbb{E}\left[|h_1|\right]. \quad (45)$$

The second equality follows because $h_1, \ldots, h_N$ are identically distributed, and, from Section III-B, ϵ_i is independent of h_i and is zero mean. Similarly, we can show that

$$\begin{aligned}\sigma_g^2 &= \sum_{i=1}^{N}\sum_{k=1}^{N}\left(\gamma^2\mathbb{E}\left[|h_i||h_k|\right] + \mathbb{E}\left[\epsilon_i^*\epsilon_k\right]\mathbb{E}\left[e^{j(\angle h_i - \angle h_k)}\right]\right) \\ &\quad - |\mu_g|^2, \\ &= \gamma^2\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbb{E}\left[|v_i||v_k|\right]\mathbb{E}\left[|u_i||u_k|\right] + N\frac{\gamma^2\sigma^2}{\rho_p^{UL}T_p^{UL}} - |\mu_g|^2.\end{aligned} \quad (46)$$

Here, the second equality follows because $|h_i| = |v_i||u_i|$ and v_i and u_i are independent. Furthermore, $\mathbb{E}\left[\epsilon_i^*\epsilon_k\right] = 0$ for $i \neq k$ and $\mathbb{E}\left[|\epsilon_i|^2\right] = \gamma^2\sigma^2/\left(\rho_p^{UL}T_p^{UL}\right)$.

We now derive $\mathbb{E}\left[|h_1|\right]$ and $\mathbb{E}\left[|h_i||h_k|\right]$ for the LoS-NLoS and NLoS-NLoS models.

a) LoS-NLoS Model: In it, $|h_1|$ is a Rayleigh RV with mean $\sqrt{\pi\beta_u\beta_v}/2$ [30, Ch. 1]. Substituting this in (45) yields (22a).

From Section II, we get $\mathbb{E}\left[|v_i||v_k|\right] = \beta_v$ and $\mathbb{E}\left[u_i u_k^*\right] = \beta_u\mathbf{R}(i,k)$. Then, from the expression for the correlation coefficient for Rayleigh envelopes in [30, Sec. 1.3], we get

$$\mathbb{E}\left[|u_i||u_k|\right] = \frac{\pi}{4}\beta_u\mathcal{F}\left(-\frac{1}{2}, -\frac{1}{2}; 1; \left[\mathbf{R}\left(i,k\right)\right]^2\right). \quad (47)$$

Hence, $\mathbb{E}\left[|h_i||h_k|\right] = \frac{\pi}{4}\beta_{\mathrm{v}}\beta_{\mathrm{u}}\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(i,k\right)\right]^2\right)$. Substituting this along with the expression for μ_g from (22a) in (46) yields (22b).

b) NLoS-NLoS Model: In it, $|v_i|$ and $|u_i|$ are Rayleigh RVs with means $\sqrt{\pi\beta_{\mathrm{v}}}/2$ and $\sqrt{\pi\beta_{\mathrm{u}}}/2$, respectively. Hence, $\mathbb{E}\left[|h_1|\right] = \pi\sqrt{\beta_{\mathrm{v}}\beta_{\mathrm{u}}}/4$. Substituting this in (45) yields (23a).

As above, we can show that $\mathbb{E}\left[|v_i||v_k|\right] = \frac{\pi}{4}\beta_{\mathrm{v}}\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(i,k\right)\right]^2\right)$ and $\mathbb{E}\left[|u_i||u_k|\right] = \frac{\pi}{4}\beta_{\mathrm{u}}\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(i,k\right)\right]^2\right)$. Thus,

$$\mathbb{E}\left[|h_i||h_k|\right] = \frac{\pi^2}{16}\beta_{\mathrm{v}}\beta_{\mathrm{u}}\left[\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(i,k\right)\right]^2\right)\right]^2.$$

Substituting this and μ_g from (23a) in (46) yields (23b).

C. Proof of Lemma 3

From Section III-B, we have $\mathbb{E}\left[\epsilon_{i_{\mathrm{sup}}}\right] = \mathbb{E}\left[\epsilon_1\right] = 0$, $\mathbb{E}\left[\epsilon_1^*\epsilon_{i_{\mathrm{sup}}}\right] = 0$, and ϵ_1 and $\epsilon_{i_{\mathrm{sup}}}$ are independent of h_1 and $h_{i_{\mathrm{sup}}}$. This combined with the expression for χ_i in (24) imply $\mathbb{E}\left[\chi_1, \chi_{i_{\mathrm{sup}}}^*\right] = \gamma^2\mathbb{E}\left[|h_1||h_{i_{\mathrm{sup}}}|\right]$ and $\mathbb{E}\left[\chi_1\right] = \mathbb{E}\left[\chi_{i_{\mathrm{sup}}}\right] = \gamma\mathbb{E}\left[|h_1|\right]$. Furthermore, $\mathrm{Var}\left[\chi_1\right] = \gamma^2\mathbb{E}\left[|h_1|^2\right] + \mathbb{E}\left[|\epsilon_1|^2\right] - \gamma^2|\mathbb{E}\left[|h_1|\right]|^2 = \mathrm{Var}\left[\chi_{i_{\mathrm{sup}}}\right]$. Substituting these in (25), we get

$$\rho\left(\chi_1, \chi_{i_{\mathrm{sup}}}\right) = \frac{\gamma^2\mathbb{E}\left[|h_1||h_{i_{\mathrm{sup}}}|\right] - \gamma^2|\mathbb{E}\left[|h_1|\right]|^2}{\gamma^2\mathbb{E}\left[|h_1|^2\right] + \mathbb{E}\left[|\epsilon_1|^2\right] - \gamma^2|\mathbb{E}\left[|h_1|\right]|^2}. \quad (48)$$

We now compute the terms in (48) for the LoS-NLoS and NLoS-NLoS models.

a) LoS-NLoS Model: The expression for $\mathbb{E}\left[|h_1||h_{i_{\mathrm{sup}}}|\right]$ is given in Appendix B. Substituting this along with $\mathbb{E}\left[h_1\right] = \sqrt{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}/2$, $\mathbb{E}\left[|h_1|^2\right] = \beta_{\mathrm{u}}\beta_{\mathrm{v}}$, and $\mathbb{E}\left[|\epsilon_1|^2\right] = \gamma^2\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right)$ in (48) and simplifying yields (27a).

b) NLoS-NLoS Model: Substituting the expression for $\mathbb{E}\left[|h_1||h_{i_{\mathrm{sup}}}|\right]$ from Appendix B, $\mathbb{E}\left[|h_1|\right] = \pi\sqrt{\beta_{\mathrm{u}}\beta_{\mathrm{v}}}/4$, $\mathbb{E}\left[|h_1|^2\right] = \beta_{\mathrm{u}}\beta_{\mathrm{v}}$, and $\mathbb{E}\left[|\epsilon_1|^2\right] = \gamma^2\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right)$ in (48) and simplifying yields (27b).

D. Proof of Lemma 4

From the definition of i_{sup}, we have, $\rho_{\mathrm{sup}}\left(n\right) = |\rho\left(\chi_1, \chi_{i_{\mathrm{sup}}}\right)|$. From the definition of $\rho_{\mathrm{sup}}\left(n\right)$ in (26), $d_{1,i_{\mathrm{sup}}} \geq n\bar{d}$. Hence, as $n \to \infty$, $d_{1,i_{\mathrm{sup}}} \to \infty$. Therefore, $\mathbf{R}(1, i_{\mathrm{sup}}) = \mathrm{sinc}\left(2d_{1,i_{\mathrm{sup}}}\right) \to 0$ and $\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2\right) \to 1$. Substituting this in (27a) and (27b), we get $\rho_{\mathrm{sup}}\left(n\right) \to 0$ as $n \to \infty$ for both models.

E. Proof of Theorem 1

1) LoS-NLoS Model: The three parts of the proof are as follows.

a) Proof of Var$\left[\chi_i\right] < \infty$: From (24), $\mathrm{Var}\left[\chi_i\right] = \gamma^2\mathbb{E}\left[|h_i|^2\right] + \mathbb{E}\left[|\epsilon_i|^2\right] - \gamma^2|\mathbb{E}\left[|h_i|\right]|^2$. Substituting $\mathbb{E}\left[|h_i|\right] = \sqrt{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}/2$, $\mathbb{E}\left[|h_i|^2\right] = \beta_{\mathrm{u}}\beta_{\mathrm{v}}$, and $\mathbb{E}\left[|\epsilon_i|^2\right] = \gamma^2\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right)$ in the above expression, we get

$$\mathrm{Var}\left[\chi_i\right] = \left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}} - \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4}\right)\gamma^2 < \infty. \quad (49)$$

b) Proof of $\lim_{N\to\infty}$ Var$\left[\sum_{i=1}^N \chi_i\right] \to \infty$: From (24), this is equivalent to proving that $\lim_{N\to\infty}\sigma_g^2 \to \infty$. Since $\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;x\right) \geq 1$ for $x \in [0,1]$, it follows from (22b) that $\sigma_g^2 \geq N\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right)$. From this, it is easy to see that $N\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right) \to \infty$ as $N \to \infty$. Hence, $\lim_{N\to\infty}\sigma_g^2 \to \infty$.

c) Proof of $\sum_{i=1}^{\infty}\rho_{sup}\left(2^i\right) < \infty$: $\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;x\right)$ is a convex function that monotonically increases from 1 to $4/\pi$ for $x \in [0,1]$ [28, Ch. 15]. Thus,

$$\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;x\right) \leq 1 + \frac{(4-\pi)\,x}{\pi} \text{ for } x \in [0,1]. \quad (50)$$

This implies that $\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;x\right) \leq 1 + x,$ for $x \in [0,1]$.

Thus, applying $\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2\right) \leq 1 + \left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2$ in (27a) and simplifying yields

$$\begin{aligned}\rho_{\mathrm{sup}}\left(2^i\right) = |\rho\left(\chi_1, \chi_{i_{\mathrm{sup}}}\right)| &\leq \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}\left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2}{4\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}\right)} \\ &= \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}\mathrm{sinc}^2\left(2d_{1,i_{\mathrm{sup}}}\right)}{4\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}\right)}. \end{aligned} \quad (51)$$

We know that $\mathrm{sinc}^2\left(2d_{1,i_{\mathrm{sup}}}\right) \leq \frac{1}{4d_{1,i_{\mathrm{sup}}}^2} \leq \frac{1}{4\left(2^i\bar{d}\right)^2}$ since $d_{1,i_{\mathrm{sup}}} \geq 2^i\bar{d}$. Hence,

$$\rho_{\mathrm{sup}}\left(2^i\right) \leq \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}\right)}\frac{1}{4^{i+1}\bar{d}^2}. \quad (52)$$

Since $\{1/4^i\}_{i=1}^{\infty}$ converges, we get

$$\sum_{i=1}^{\infty}\rho_{\mathrm{sup}}\left(2^i\right) \leq \frac{\frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4}}{\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{4} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}}\left(\frac{1}{4\bar{d}^2}\right)\sum_{i=1}^{\infty}\frac{1}{4^i} < \infty. \quad (53)$$

2) NLoS-NLoS Model: As above, we can show that $\mathrm{Var}\left[\chi_i\right] = \left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}} - \frac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{16}\right)\gamma^2 < \infty$ and $\mathrm{Var}\left[\sum_{i=1}^N \chi_i\right] = \sigma_g^2 \geq N\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right) \to \infty$ as $N \to \infty$.

Applying $\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2\right) \leq 1 + \left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2$ in (27b) and simplifying yields

$$\rho_{\mathrm{sup}}\left(2^i\right) \leq \frac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}\left[\left(\left[\mathbf{R}\left(1,i_{\mathrm{sup}}\right)\right]^2 + 1\right)^2 - 1\right]}{16\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{16} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}\right)}. \quad (54)$$

Simplifying (54) and upper bounding along lines similar to (52), we can show that

$$\begin{aligned}\sum_{i=1}^{\infty}\rho_{\mathrm{sup}}\left(2^i\right) \leq &\frac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{16\left(\beta_{\mathrm{u}}\beta_{\mathrm{v}} - \frac{\pi^2\beta_{\mathrm{u}}\beta_{\mathrm{v}}}{16} + \frac{\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}\right)} \\ &\times\left(\frac{1}{16\bar{d}^4}\sum_{i=1}^{\infty}\frac{1}{16^i} + \frac{1}{2\bar{d}^2}\sum_{i=1}^{\infty}\frac{1}{4^i}\right). \end{aligned} \quad (55)$$

Since both $\{1/16^i\}_{i=1}^{\infty}$ and $\{1/4^i\}_{i=1}^{\infty}$ converge, we get $\sum_{i=1}^{\infty}\rho_{\mathrm{sup}}\left(2^i\right) < \infty$.

F. Proof of Lemma 5

Substituting $\rho_{\mathrm{d}}^{\mathrm{DL}} = \alpha E_{\mathrm{BS}}/T_{\mathrm{d}}^{\mathrm{DL}}$ from (8) and $\rho_{\mathrm{p}}^{\mathrm{DL}} T_{\mathrm{p}}^{\mathrm{DL}} = (1-\alpha) E_{\mathrm{BS}}$ from (9) in (16) yields

$$\mathrm{SNR}_{\mathrm{eff}} = \frac{E_{\mathrm{BS}}\left(\sigma_g^2+|\mu_g|^2\right)}{\sigma^2\left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)}\left(\frac{\alpha\kappa-\alpha^2}{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\alpha}\right), \tag{56}$$

where $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) = \frac{E_{\mathrm{BS}}\sigma_g^2+\sigma^2}{E_{\mathrm{BS}}\sigma_g^2}\frac{T_{\mathrm{d}}^{\mathrm{DL}}}{T_{\mathrm{d}}^{\mathrm{DL}}-1}$ and $\kappa = \left[E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)+|\mu_g|^2\sigma^2\right]/E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)$.

By first order conditions, the optimal α satisfies $\frac{\partial \mathrm{SNR}_{\mathrm{eff}}}{\partial\alpha} = 0$. Solving this equation yields $\alpha = \zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) \pm \sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)\left[\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa\right]}$. We now prove that $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) \geq \kappa$, which implies that α is real. Since $\left(E_{\mathrm{BS}}\sigma_g^2+\sigma^2\right)\left(\sigma_g^2+|\mu_g|^2\right) \geq E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)+\sigma^2|\mu_g|^2$ and $T_{\mathrm{d}}^{\mathrm{DL}}E_{\mathrm{BS}}\sigma_g^2 > \left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)E_{\mathrm{BS}}\sigma_g^2$, it follows that

$$\begin{aligned}\left[T_{\mathrm{d}}^{\mathrm{DL}}E_{\mathrm{BS}}\sigma_g^2\right]\left(E_{\mathrm{BS}}\sigma_g^2+\sigma^2\right)\left(\sigma_g^2+|\mu_g|^2\right) &\geq \left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)\\ \times E_{\mathrm{BS}}\sigma_g^2\left(E_{\mathrm{BS}}\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)+\sigma^2|\mu_g|^2\right).\end{aligned} \tag{57}$$

Rearranging the terms in (57) and substituting the definitions of $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)$ and κ, we can show that $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) \geq \kappa$.

Furthermore, $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) - \sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)\left[\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa\right]}$ is the optimal solution since $\alpha \in [0,1]$.

G. Proof of Theorem 2

We shall prove that $\frac{\partial C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right),T_{\mathrm{p}}^{\mathrm{UL}},T_{\mathrm{p}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}} \geq 0$ for $T_{\mathrm{d}}^{\mathrm{DL}} > 1$. From (31), we get

$$\begin{aligned}&\frac{\partial C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right),T_{\mathrm{p}}^{\mathrm{UL}},T_{\mathrm{p}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\\ &= \frac{\mathbb{E}\left[\frac{T_{\mathrm{d}}^{\mathrm{DL}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2}{1+\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2}\frac{\partial\mathrm{SNR}_{\mathrm{eff}}}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\right]}{\log(2)T_{\mathrm{c}}}\\ &\quad+\frac{\mathbb{E}\left[\log\left(1+\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2\right)\right]}{\log(2)T_{\mathrm{c}}}.\end{aligned} \tag{58}$$

From the expression for $\mathrm{SNR}_{\mathrm{eff}}$ in (31), we get

$$\begin{aligned}&\frac{\partial\mathrm{SNR}_{\mathrm{eff}}}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\\ &= -\frac{E_{\mathrm{BS}}^2\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)}{E_{\mathrm{BS}}\sigma_g^2\sigma^2\left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)}\\ &\quad\times\frac{\left[\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}-\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa}\right]^2}{\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)\left[\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa\right]}}\frac{\partial\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\\ &\quad+\left[\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}-\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa}\right]^2E_{\mathrm{BS}}^2\sigma_g^2\left(\sigma_g^2+|\mu_g|^2\right)\\ &\quad\times\frac{\partial}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\left(\frac{1}{E_{\mathrm{BS}}\sigma_g^2\sigma^2\left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)}\right).\end{aligned} \tag{59}$$

Substituting $\frac{\partial\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}} = -\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)/\left[T_{\mathrm{d}}^{\mathrm{DL}}\left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)\right]$, which follows from the definition of $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)$ in (30), and $\frac{\partial}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\left(\frac{1}{E_{\mathrm{BS}}\sigma_g^2\sigma^2\left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)}\right) = -1/\left(E_{\mathrm{BS}}\sigma_g^2\sigma^2\left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)^2\right)$ in (59) and rearranging terms yields

$$\frac{\partial\mathrm{SNR}_{\mathrm{eff}}}{\partial T_{\mathrm{d}}^{\mathrm{DL}}} = \frac{\mathrm{SNR}_{\mathrm{eff}}}{T_{\mathrm{d}}^{\mathrm{DL}}-1}\left[\frac{1}{T_{\mathrm{d}}^{\mathrm{DL}}}\sqrt{\frac{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa}}-1\right]. \tag{60}$$

Let $\Delta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) = \frac{T_{\mathrm{d}}^{\mathrm{DL}}}{T_{\mathrm{d}}^{\mathrm{DL}}-1}\left[1-\frac{1}{T_{\mathrm{d}}^{\mathrm{DL}}}\sqrt{\frac{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa}}\right]$. Substituting (60) in (58) and rearranging yields

$$\begin{aligned}&\frac{\partial C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right),T_{\mathrm{p}}^{\mathrm{UL}},T_{\mathrm{p}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\\ &= \frac{\mathbb{E}\left[\log\left(1+\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2\right)\right]}{\log(2)T_{\mathrm{c}}}\\ &\quad-\frac{\mathbb{E}\left[\frac{\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2}{1+\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2}\Delta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)\right]}{\log(2)T_{\mathrm{c}}}.\end{aligned} \tag{61}$$

From Appendix F, we know that $\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa > 0$. Hence, $\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)/\left(\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa\right)} > 1$, from which we get $1-\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)}\Big/\left[T_{\mathrm{d}}^{\mathrm{DL}}\sqrt{\zeta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right)-\kappa}\right] < \left(T_{\mathrm{d}}^{\mathrm{DL}}-1\right)/T_{\mathrm{d}}^{\mathrm{DL}}$. Rearranging this inequality yields $\Delta\left(T_{\mathrm{d}}^{\mathrm{DL}}\right) < 1$. Hence,

$$\begin{aligned}&\frac{\partial C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right),T_{\mathrm{p}}^{\mathrm{UL}},T_{\mathrm{p}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}}\\ &\geq \frac{\mathbb{E}\left[\log\left(1+\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2\right)\right]}{\log(2)T_{\mathrm{c}}}\\ &\quad-\frac{\mathbb{E}\left[\frac{\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2}{1+\mathrm{SNR}_{\mathrm{eff}}|\widehat{g}_{\mathrm{norm}}(\widehat{\mathbf{h}})|^2}\right]}{\log(2)T_{\mathrm{c}}}.\end{aligned}$$

Applying $\log(1+x) \geq x/(1+x)$, for $x \geq 0$, in the above expression yields $\frac{\partial C_{\mathrm{LB}}\left(\alpha^\star\left(T_{\mathrm{d}}^{\mathrm{DL}}\right),T_{\mathrm{p}}^{\mathrm{UL}},T_{\mathrm{p}}^{\mathrm{DL}}\right)}{\partial T_{\mathrm{d}}^{\mathrm{DL}}} \geq 0$.

H. Proof of Lemma 6

The net channel gains of all groups are statistically identical, $\mathbb{E}\left[\epsilon_{\mathrm{G},k}\right] = 0$, and $\epsilon_{\mathrm{G},k}$ is independent of G_k. Therefore, from (42), we get

$$\mathbb{E}\left[g(\widehat{\mathbf{G}})\right] = \gamma_{\mathrm{G}}\sum_{k=1}^{K}\mathbb{E}\left[|G_k|\right] = K\gamma_{\mathrm{G}}\mathbb{E}\left[|G_1|\right]. \tag{62}$$

Since $G_k \sim \mathcal{CN}\left(0,\sigma_{\mathrm{grp}}^2\right)$, substituting $\mathbb{E}\left[|G_k|\right] = \sqrt{\pi\sigma_{\mathrm{grp}}^2}/2$ in the above expression yields (43a).

From (25), the normalized correlation coefficient between G_k and $G_{k'}$ is given by

$$\rho\left(G_k,G_{k'}\right) = \frac{\sum_{i\in\mathcal{C}_k}\sum_{i'\in\mathcal{C}_{k'}}v_i v_{i'}^*\mathbb{E}\left[u_i u_{i'}^*\right]}{\sigma_{\mathrm{grp}}^2}, \tag{63}$$

where $\mathbb{E}\left[u_i u_{i'}^*\right] = \beta_{\mathrm{u}}\mathbf{R}(i,i')$. As in Appendix B, we get

$$\mathbb{E}\left[|G_k|,|G_{k'}|\right] = \frac{\pi}{4}\sigma_{\mathrm{grp}}^2\mathcal{F}\left(-\frac{1}{2},-\frac{1}{2};1;\rho^2\left(G_k,G_{k'}\right)\right). \tag{64}$$

Also, since $\mathbb{E}\left[\epsilon_{\mathrm{G},k}\epsilon_{\mathrm{G},k'}^*\right] = 0$ and $\mathbb{E}\left[|\epsilon_{\mathrm{G},k'}|^2\right] = \gamma_{\mathrm{G}}^2\sigma^2/\left(\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}\right)$, we get

$$\begin{aligned}\mathrm{Var}\left[g(\widehat{\mathbf{G}})\right] &= \gamma_{\mathrm{G}}^2\sum_{k=1}^{K}\sum_{k'=1}^{K}\mathbb{E}\left[|G_k|,|G_{k'}|\right]+K\frac{\gamma_{\mathrm{G}}^2\sigma^2}{\rho_{\mathrm{p}}^{\mathrm{UL}}T_{\mathrm{p}}^{\mathrm{UL}}}\\ &\quad-\left|\mathbb{E}\left[g(\widehat{\mathbf{G}})\right]\right|^2.\end{aligned} \tag{65}$$

Substituting the expression for $\rho\left(G_k,G_{k'}\right)$ in (64) and then in (65) along with (43a) yields (43b).

ACKNOWLEDGMENT

The authors would like to thank inputs from Sriram Ganesan, which helped in improving the quality of their work.

REFERENCES

[1] S. Naduvilpattu and N. B. Mehta, "Insights into cumulative impact of channel estimation errors on RIS phase-shift configuration and data demodulation," in *Proc. IEEE ICC*, Jun. 2024.

[2] Y. Liu et al., "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1546–1577, 3rd Quart., 2021.

[3] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[4] M. Di Renzo et al., "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[5] Q. Wu et al., "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[6] Q. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, 2020.

[7] B. Shamasundar, N. Daryanavardan, and A. Nosratinia, "Channel training & estimation for reconfigurable intelligent surfaces: Exposition of principles, approaches, and open problems," *IEEE Access*, vol. 11, pp. 6717–6734, 2023.

[8] J. An, L. Wang, C. Xu, L. Gan, and L. Hanzo, "Optimal pilot power based channel estimation improves the throughput of intelligent reflective surface assisted systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16202–16206, Dec. 2020.

[9] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1245–1260, Feb. 2022.

[10] J. An, C. Xu, L. Wang, Y. Liu, L. Gan, and L. Hanzo, "Joint training of the superimposed direct and reflected links in reconfigurable intelligent surface assisted multiuser communications," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 739–754, Jun. 2022.

[11] K. Zhi et al., "Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI," *IEEE Trans. Inf. Theory*, vol. 69, no. 5, pp. 3001–3033, May 2023.

[12] K. Zhi, C. Pan, H. Ren, and K. Wang, "Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3558–3574, May 2022.

[13] K. Zhi, C. Pan, G. Zhou, H. Ren, M. Elkashlan, and R. Schober, "Is RIS-aided massive MIMO promising with ZF detectors and imperfect CSI?" *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 3010–3026, Oct. 2022.

[14] N. K. Kundu, Z. Li, J. Rao, S. Shen, M. R. McKay, and R. Murch, "Optimal grouping strategy for reconfigurable intelligent surface assisted wireless communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 1082–1086, May 2022.

[15] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

[16] Z. Zhou, N. Ge, Z. Wang, and L. Hanzo, "Joint transmit precoding and reconfigurable intelligent surface phase adjustment: A decomposition-aided channel estimation approach," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1228–1243, Feb. 2021.

[17] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Exploiting array geometry for reduced-subspace channel estimation in RIS-aided communications," in *Proc. IEEE 12th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2022, pp. 455–459.

[18] M. Xu, S. Zhang, C. Zhong, J. Ma, and O. A. Dobre, "Ordinary differential equation-based CNN for channel extrapolation over RIS-assisted communication," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 1921–1925, Jun. 2021.

[19] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and RIS-assisted massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4255–4268, Jun. 2022.

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

[21] J. An et al., "Codebook-based solutions for reconfigurable intelligent surfaces and their open challenges," *IEEE Wireless Commun.*, early access, Nov. 15, 2022, doi: 10.1109/MWC.010.2200312.

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

[23] M. Peligrad, "On the central limit theorem for ρ-mixing sequences of random variables," *Ann. Probab.*, vol. 15, no. 4, pp. 1387–1394, Oct. 1987.

[24] C. Pan et al., "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[25] S. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. London, U.K.: Pearson, 1993.

[26] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[27] E. Bjornson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. Boston, MA, USA: Now Foundations and Trends, 2017.

[28] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1965.

[29] J. Chambers, *Graphical Methods for Data Analysis*. Boca Raton, FL, USA: CRC Press, 1983.

[30] W. C. Jakes, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1974.

Suji Naduvilpattu (Student Member, IEEE) received the Bachelor of Technology degree in electronics and communication engineering from the Government College of Engineering, Kannur, Kerala, India, in 2009, and the Master of Technology degree in telecommunication from the National Institute of Technology, Calicut, Kerala, in 2012. She is currently pursuing the Ph.D. degree with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru. Her research interests include wireless communication, spectrum sharing, and reconfigurable intelligent surfaces.

Neelesh B. Mehta (Fellow, IEEE) received the B.Tech. degree in electronics and communications engineering from the Indian Institute of Technology (IIT) Madras in 1996 and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, USA, in 1997 and 2001, respectively. He is currently a Professor with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru. He is a fellow of the Indian National Science Academy, Indian Academy of Sciences, Indian National Academy of Engineering, and National Academy of Sciences India. He was a recipient of the J. C. Bose Fellowship, Shanti Swarup Bhatnagar Award, Khosla Award, Vikram Sarabhai Research Award, and the Swarnjayanti Fellowship. He served on the Board of Governors for the IEEE ComSoc from 2012 to 2015. He served on the executive editorial committee for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2014 to 2017 and served as the Chair from 2017 to 2018. He also serves as the chair of its steering committee. He has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE WIRELESS COMMUNICATION LETTERS in the past.