Combating Inter-Operator Pilot Contamination in Reconfigurable Intelligent Surfaces Assisted Multi-Operator Networks

Doğa Gürgünoğlu, *Student Member, IEEE*, Emil Björnson, *Fellow, IEEE*, Gábor Fodor, *Senior Member, IEEE*

Abstract—In this paper, we study a new kind of pilot contamination appearing in multi-operator reconfigurable intelligent surfaces (RIS) assisted networks, where multiple operators provide services to their respective served users. The operators use dedicated frequency bands, but each RIS inadvertently reflects the transmitted uplink signals of the user equipment devices in multiple bands. Consequently, the concurrent reflection of pilot signals during the channel estimation phase introduces a new inter-operator pilot contamination effect. We investigate the implications of this effect in systems with either deterministic or correlated Rayleigh fading channels, specifically focusing on its impact on channel estimation quality, signal equalization, and channel capacity. The numerical results demonstrate the substantial degradation in system performance caused by this phenomenon and highlight the pressing need to address inter-operator pilot contamination in multi-operator RIS deployments. To combat the negative effect of this new type of pilot contamination, we propose to use orthogonal RIS configurations during uplink pilot transmission, which can mitigate or eliminate the negative effect of inter-operator pilot contamination at the expense of some inter-operator information exchange and orchestration.

Index Terms—Reconfigurable intelligent surface, channel estimation, pilot contamination.

I. INTRODUCTION

Pilot contamination is a key problem that frequently arises in wireless communication systems [2]. When multiple users use the same pilot sequences simultaneously in the same band, due to the limited channel coherence time, the base station (BS) cannot distinguish their channels. This typically results in poor channel estimates and extra beamformed interference from or towards the user equipments (UEs) that reuse the same pilot sequence. Therefore, pilot contamination adversely affects the coherent reception of data, and methods to mitigate pilot contamination—including adaptive pilot reuse, power control, user grouping, multi-cell coordination, and coded random access techniques—have been widely studied in the communication literature [2]–[6].

In recent years, reconfigurable intelligent surfaces (RISs) have arisen as a new technology component for 6G [7]. An RIS is a surface consisting of multiple reflecting elements that have sub-wavelength spacing and controllable reflection properties [8]. This feature provides partial control of the propagation environment that can lead to better services for users, especially when their serving BS is not in their line of sight (LOS). By adjusting the impedances of the individual elements via a RIS controller, the elements are capable of adding desired phase shifts to the reflected signals, thereby forming reflected beams in desired directions that can significantly boost the signal-to-interference-plus-noise ratio (SINR) and reduce the symbol estimation error at the receiver [7], [9].

On the other hand, the addition of RISs to existing systems introduces new design and operational challenges [10]. For example, the length of the pilot signal required by a single UE is proportional to the number of RIS elements (e.g., tens or hundreds), because the RIS must change its configuration to explore all channel dimensions [11], [12]. In addition, the path loss of the reflected path through a passive RIS is proportional to the multiplication of the path losses to and from the RIS [13], so a larger surface is needed to achieve a decent signal-to-noise ratio (SNR) improvement. Active RISs, on the other hand, use amplifiers to overcome the large path loss but have the traditional issues of relays: increased power consumption, higher cost, and additional noise [14], [15]. While the aforementioned problems caused by the RIS have been recognized [16], pilot contamination caused by the presence of multiple RISs has not been studied in the literature.

Wireless communication systems use standardized protocols, interfaces, and well-defined pilot sequences and codebooks to ensure inter-operability [17]. While employing RISs in cellular networks have not been studied by the relevant standards organizations yet, it may be expected that the configuration sequences that facilitate the deployment of RISs while maintaining interoperability will be specified. Consequently, when multiple cellular networks are deployed by different network operators in overlapping geographical areas, the RISs may adopt identical or overlapping pilot sequences and cause pilot contamination. The number of orthogonal pilot sequences is limited by the length of the pilot sequence, and increasing the pilot sequence length not only creates more channel estimation overhead but also is infeasible due to the limited coherence budget of the channel. As a consequence, the need for repeating pilot sequences comes up very often.

In this paper, we argue that when multiple RISs are

deployed for the purpose of shaping the propagation characteristics of the environment, the propagation characteristics might change in unintended ways. For example, an RIS belonging to another operator might modify the propagation of a UE's own pilot signal, leading to pilot contamination even in the absence of any interfering signals or intra-band pilot reuse. The underlying reason is that an RIS element—although designed for a particular frequency—does not act as a bandpass filter, but reflects all frequencies with varying amplitude and phase. Indeed, as pointed out in [18], due to the lack of baseband signal processing, the RIS reflects the impinging broadband signal with frequency-flat reflection coefficients. Therefore, in realistic system models of, for example, passive RIS assisted 5G New Radio systems, we need to take into account that the RIS inadvertently reflects the transmitted uplink signals of the user equipment devices in multiple bands as in [19], [20].

Specifically, in this paper, we identify this new pilot contamination phenomenon as a major practical challenge when multiple RIS assisted operator networks are deployed over a geographical area, including the important practical scenario of inter-operator site sharing [21], [22], [23]. In such an environment, a UE that transmits pilots to its serving BS via multiple RIS, which may change their configurations simultaneously, is exposed to new pilot-related ambiguities that have not been studied before. Since this phenomenon exacerbates the pilot contamination problem, it is clear that pilot contamination due to the presence of multiple RISs must be dealt with.

To the best of the authors' knowledge, the problem of inter-operator pilot contamination in RIS-aided wireless communication systems has not been addressed before, except in the preliminary version of this manuscript [1], which assumed deterministic rather than stochastically fading channels. In this paper, we derive the capacity lower bound of the system under pilot contamination and imperfect channel state information (CSI) assuming Rayleigh fading. Our major contributions can be summarized as follows:

- For the case when inter-operator pilot contamination is neglected, we provide a misspecified maximum likelihood (ML) estimator under the assumption that all the channels in the system setup are deterministic. We also derive the resulting channel estimation mean squared error (MSE) for different choices of the RIS configurations.
- Based on the obtained results for the channel estimation error under inter-operator pilot contamination, we provide the data signal estimation MSE for a misspecified minimum mean squared error (MMSE) estimator.
- In addition to deterministic channels, we also consider the case where the channels are Rayleigh fading with spatial correlation. For generic channel spatial covariances, we derive the misspecified MMSE estimator and the resulting MSE.
- Based on the channel estimation error model, obtained for spatially correlated Rayleigh fading channels, we derive a capacity lower bound given the channel estimates. Our numerical results show that the choice of RIS configurations during channel estimation makes a

significant impact on the capacity lower bound.

The rest of the manuscript is organized as follows: in Section II, we provide the received signal model, in Section III, we provide the misspecified ML estimator where the inter-operator pilot contamination is neglected and the channels are assumed to be deterministic. Section IV builds on top of section III by providing the data estimation MSE as a result of inter-operator pilot contamination. In Section V, we derive the impact of inter-operator pilot contamination in closed-form for the case of spatially correlated Rayleigh-fading channels by considering the channel estimation MSE as our performance metric. Since the data estimation MSE has a dependence on individual channel realizations, an alternative performance metric for data transmission is needed to capture the behavior of fading channels. To this end, we derive the capacity lower bound under channel side information in Section VI. We provide the numerical results in Section VII, and conclude the manuscript in Section VIII.

II. SYSTEM MODEL

In this paper, we study the pilot contamination caused by the presence of multiple RISs by considering the uplink of a system consisting of two wide-band RISs, two single-antenna UEs, and two co-located single-antenna BSs. The UEs are subscribed to different operators and use non-overlapping frequency bands. Each RIS is dedicated to and controlled by a single operator, but both UE signals are reflected from both RISs. In this scenario, although there is no interference between the two UEs, both RISs affect both frequency bands.

Although the two BSs are closely located, their coupling effects are ignored due to their casings providing enough electromagnetic shielding. For co-located BSs, the coupling can be avoided by separating the BSs by 1.5 meters [24]. Inter-RIS coupling is also ignored since the RISs are far apart. While the coupling among the elements in a dense RIS affects the system behavior, it can generally be treated as part of the spatial channel correlation. We consider arbitrary spatial correlation in this paper and are not explicitly modeling the coupling to avoid convoluting the notation but focus on the inter-operator pilot contamination effect. However, explicit modeling of the coupling effect among the RIS elements can be done as in [25], [26] and is recommended for future work.

In Fig. 1, we graphically describe the system that we consider. The components associated with the two different operators are depicted in two different colors: blue BS, RIS, UE and the channels belong to operator 1, while the red ones belong to operator 2. The operators can potentially use site-sharing (as in the figure) to reduce deployment costs but transmit over two disjoint narrow frequency bands to their respective serving BSs. Each RIS has N reflecting elements, and is dedicated to and controlled by a single operator but affects both bands. To focus on the fundamentals of pilot contamination, we consider an environment where the direct UE-BS paths are blocked, while the UE-RIS and RIS-BS paths are operational. Since the BSs and RISs have fixed deployment locations, we assume the RIS-BS channels $\mathbf{h}_k \in \mathbb{C}^N$ are known, while the UE-RIS channels $\mathbf{g}_k \in \mathbb{C}^N$ are unknown and to be estimated, for $k = 1, 2$.
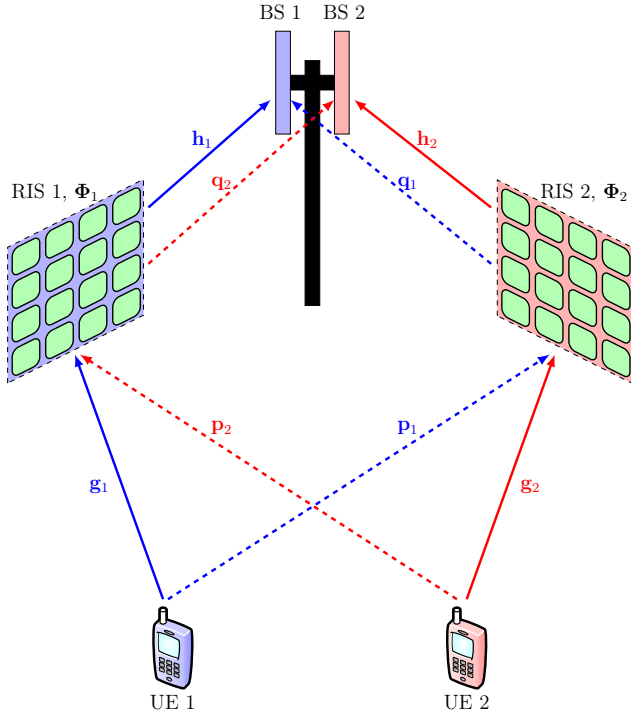
Fig. 1: The considered setup with two UEs, two RISs, and two co-located single-antenna BSs. The blue channels correspond to frequency band 1, and the red channels correspond to frequency band 2, subscribed by UEs 1 and 2, respectively. The desired channels are denoted by solid lines, while the undesired channels (whose existences might be unknown to the BSs) are denoted by dashed lines. Each channel vector is N-dimensional because each RIS has N elements.

The signal transmitted by UE k reaches its serving BS through the channels $\mathbf{h}_k$ and $\mathbf{g}_k$, for $k = 1, 2$. Importantly, each UE's transmitted signal is also reflected by the non-serving operator's RIS and parts of the reflected signal will reach the serving BS. This effect contaminates the pilot signal reflected by the serving RIS and we will study the implications. The phenomenon is illustrated in Fig. 1, where the resulting UE-RIS and RIS-BS channels are denoted by $\mathbf{p}_k \in \mathbb{C}^N$ and $\mathbf{q}_k \in \mathbb{C}^N$, respectively, for $k = 1, 2$. Defining the pilot signal of UE k as $s_k \in \mathbb{C}$, the received signals on bands 1 and 2 at the BSs can be expressed as

$$y_{p1} = \sqrt{P_p}\mathbf{h}_1^T\mathbf{\Phi}_1\mathbf{g}_1 s_1 + \sqrt{P_p}\mathbf{q}_1^T\mathbf{\Phi}_2\mathbf{p}_1 s_1 + w_{p1}, \quad (1a)$$
$$y_{p2} = \sqrt{P_p}\mathbf{h}_2^T\mathbf{\Phi}_2\mathbf{g}_2 s_2 + \sqrt{P_p}\mathbf{q}_2^T\mathbf{\Phi}_1\mathbf{p}_2 s_2 + w_{p2}, \quad (1b)$$

where $y_{pk} \in \mathbb{C}$ denotes the received signal, $w_{pk} \sim \mathcal{CN}(0, 1)$ denotes the receiver noise for band k, and $\mathbf{\Phi}_k = \mathrm{diag}(e^{-j\phi_{k1}}, \ldots, e^{-j\phi_{kN}})$ denotes the kth RIS's reflection matrix, and P_p denotes the pilot signal's transmission power. We assume $s_1 = s_2 = 1$ without loss of generality. When analyzing channel estimation, it is more convenient to rewrite (1) as

$$y_{p1} = \sqrt{P_p}\boldsymbol{\phi}_1^T\mathbf{D}_{\mathbf{h}_1}\mathbf{g}_1 + \sqrt{P_p}\boldsymbol{\phi}_2^T\mathbf{D}_{\mathbf{q}_1}\mathbf{p}_1 + w_{p1}, \quad (2a)$$
$$y_{p2} = \sqrt{P_p}\boldsymbol{\phi}_2^T\mathbf{D}_{\mathbf{h}_2}\mathbf{g}_2 + \sqrt{P_p}\boldsymbol{\phi}_1^T\mathbf{D}_{\mathbf{q}_2}\mathbf{p}_2 + w_{p2}, \quad (2b)$$

where $\mathbf{D}_{\mathbf{h}_k}$ and $\mathbf{D}_{\mathbf{q}_k}$ represent the $N \times N$ diagonal matrices containing the elements of $\mathbf{h}_k$ and $\mathbf{q}_k$, and $\boldsymbol{\phi}_k \in \mathbb{C}^N$ denotes the column vectors containing the diagonal entries of $\mathbf{\Phi}_k$ for $k = 1, 2$.

As there are N parameters in $\mathbf{g}_1$ and $\mathbf{g}_2$, at least N linearly independent observations are needed to estimate them uniquely. To this end, we perform $L \geq N$ pilot transmissions over time, and we vertically stack the received signals to obtain

$$\mathbf{y}_{p1} = \sqrt{P_p}\mathbf{B}_1\mathbf{D}_{\mathbf{h}_1}\mathbf{g}_1 + \sqrt{P_p}\mathbf{B}_2\mathbf{D}_{\mathbf{q}_1}\mathbf{p}_1 + \mathbf{w}_{p1}, \quad (3a)$$
$$\mathbf{y}_{p2} = \sqrt{P_p}\mathbf{B}_2\mathbf{D}_{\mathbf{h}_2}\mathbf{g}_2 + \sqrt{P_p}\mathbf{B}_1\mathbf{D}_{\mathbf{q}_2}\mathbf{p}_2 + \mathbf{w}_{p2}, \quad (3b)$$

where $\mathbf{y}_{pk} = [y_{pk}[1], \ldots, y_{pk}[L]]^T \in \mathbb{C}^L$ denotes the sequence of received signals from the kth UE over L time instances, and the matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ represent the sequence of RIS configurations over L time instances; that is, $\mathbf{B}_k \triangleq \begin{bmatrix} \boldsymbol{\phi}_k[1] & \ldots & \boldsymbol{\phi}_k[L] \end{bmatrix}^T \in \mathbb{C}^{L \times N}$ for $k = 1, 2$.

A. Construction of $\mathbf{B}_1$ and $\mathbf{B}_2$

We consider $\mathbf{B}_1$ and $\mathbf{B}_2$ to be orthogonal matrices, that is, $\mathbf{B}_k^H\mathbf{B}_k = L\mathbf{I}_N$. Each entry of the matrices must have a unit modulus, and there are multiple designs that perform equally well. One way to design such matrices is to start from an L-dimensional Discrete Fourier Transform (DFT) matrix $\tilde{\mathbf{B}} \in \mathbb{C}^{L \times L}$. We need $L \geq N$ to estimate all channel components. After constructing $\tilde{\mathbf{B}}$, we perform the following operation:

- For $\mathbf{B}_1 = \mathbf{B}_2$, we assign the sub-matrix consisting of the first N column vectors of $\tilde{\mathbf{B}}$ to both $\mathbf{B}_1$ and $\mathbf{B}_2$.
- To obtain $\mathbf{B}_1^H\mathbf{B}_2 = \mathbf{0}$, we need at least $2N$ vectors that are orthogonal to each other, therefore, we need at least $L \geq 2N$. With this in mind, we assign the sub-matrix consisting of the first N columns of $\tilde{\mathbf{B}}$ and assign it to $\mathbf{B}_1$. Then we assign the sub-matrix consisting of the second N column vectors of $\tilde{\mathbf{B}}$ to $\mathbf{B}_2$. Consequently, for $L = 2N$, for example, we obtain

$$\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix}. \quad (4)$$

B. Extension to Multiple-Antenna BSs

While we considered single-antenna base stations to focus on the multi-RIS pilot contamination phenomenon more, it is indeed possible to consider multi-antenna base stations. For example, suppose both base stations have M antennas. In this case, the pilot signal received at a single time instant at the k'th base station can be expressed as

$$\mathbf{y}_p = \sqrt{P_p}(\mathbf{H}_k\mathbf{\Phi}_k\mathbf{g}_k + \mathbf{Q}_k\mathbf{\Phi}_j\mathbf{p}_k)s_k + \mathbf{w}_k \in \mathbb{C}^{M \times 1}, \quad (5)$$

where $\mathbf{H}_k, \mathbf{Q}_k \in \mathbb{C}^{M \times N}$ denote the channel between RIS k and BS k and the channel between RIS j and BS k, respectively. Since $\mathbf{H}_k$ and $\mathbf{Q}_k$ are matrices (as opposed to vectors used in the manuscript), it is non-trivial to make the transition from (1) to (2). To this end, we define $\bar{\boldsymbol{\phi}}_n \triangleq \phi_n\mathbf{I}_{M \times M}$, $\boldsymbol{\phi}_k \triangleq [\bar{\boldsymbol{\phi}}_1, \ldots, \bar{\boldsymbol{\phi}}_N]$ and $\mathbf{D}_{\mathbf{H}_k} \triangleq \mathrm{diag}(\mathbf{h}_k^{(1)}, \ldots, \mathbf{h}_k^{(N)})$ with $\mathbf{h}_k^{(n)}$ denoting the n'th $M \times 1$ column of

$$\mathbf{y}_p = \sqrt{P_p}(\boldsymbol{\phi}_k\mathbf{D}_{\mathbf{H}_k}\mathbf{g}_k + \boldsymbol{\phi}_j\mathbf{D}_{\mathbf{Q}_k}\mathbf{p}_k)s_k + \mathbf{w}_k \in \mathbb{C}^{M \times 1}, \quad (6)$$

which corresponds to eq. (2) in the multi-antenna BS setting. Without loss of generality, we assume that $s_k = 1$. After this step, we stack the $M \times 1$ received pilot observations over time vertically to obtain a $LM \times 1$ collection of pilot observations:

$$\mathbf{Y}_p = \sqrt{P_p}(\mathbf{B}_k\mathbf{D}_{\mathbf{H}_k}\mathbf{g}_k + \mathbf{B}_j\mathbf{D}_{\mathbf{Q}_k}\mathbf{p}_k) + \mathbf{W}_k \in \mathbb{C}^{LM \times 1}. \quad (7)$$

This equation corresponds to (3) in our manuscript. Note that the "cascaded" channels going over each RIS in this case correspond to $\mathbf{D}_{\mathbf{H}_k}\mathbf{g}_k \in \mathbb{C}^{MN \times 1}$, that is, the true dimensions of the channels are represented. It also has to be noted that $\mathbf{B}_k \in \mathbb{C}^{LM \times MN}$ in this case. Compared to the original $\mathbf{B}_k$ matrices in our manuscript, these $\mathbf{B}_k$ matrices are their Kronecker-multiplied versions, therefore, using the same dynamics with $L \geq 2N$ and DFT matrices, it is possible to obtain both $\mathbf{B}_1 = \mathbf{B}_2$ and $\mathbf{B}_1^H\mathbf{B}_2 = \mathbf{0}$. Since we assume that the static channels $\mathbf{H}_k$ and $\mathbf{Q}_k$ are known, we still have N unknown channel dimensions, hence we do not have additional complexity in configuring the RISs to estimate the unknown channels coming from having multiple antennas at the BSs.

In the remainder of this paper, we will analyze channel estimation and the resulting communication performance for deterministic and fading channels, respectively.

III. MAXIMUM LIKELIHOOD ESTIMATION OF DETERMINISTIC CHANNELS

In this section, we will consider channel estimation for deterministic channels. The same assumptions and results will then be considered in Section IV for data transmission.

We assume that $\mathbf{g}_k$ is a deterministic and unknown channel without any known structure. That is, $\mathbf{g}_k$ is an $N \times 1$ vector of complex deterministic parameters to be estimated. In addition, we assume that $\mathbf{h}_k$ is perfectly known. On the other hand, the BSs do not know the existence of $\mathbf{q}_k$ and $\mathbf{p}_k$. We assume that the BSs are unable to obtain enough fading observations to construct a statistical model, therefore, the BSs employ the classical estimation framework. In contrast, we assume that statistics of the fading channels of interest $\mathbf{g}_1$ and $\mathbf{g}_2$ are available in Section V, and we consider $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{q}_1$, and $\mathbf{q}_2$ to be deterministic since those links are between fixed deployments. On the other hand, we consider the statistics of the fading channels $\mathbf{p}_1$ and $\mathbf{p}_2$ only for MSE analysis, and assume that the BSs are unaware of these channels. We can denote the received signal models assumed by the BSs as

$$\hat{\mathbf{y}}_{p1} = \sqrt{P_p}\mathbf{B}_1\mathbf{D}_{\mathbf{h}_1}\mathbf{g}_1 + \mathbf{w}_{p1}, \quad (8a)$$

$$\hat{\mathbf{y}}_{p2} = \sqrt{P_p}\mathbf{B}_2\mathbf{D}_{\mathbf{h}_2}\mathbf{g}_2 + \mathbf{w}_{p2}. \quad (8b)$$

Since $\mathbf{g}_k$ does not have a known structure and hence consists of N scalars, BS k requires at least N independent observations to estimate it. To this end, both $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{C}^{L \times N}$ must have full column rank. Furthermore, we require that the RIS configurations used at different time instances are mutually orthogonal and contain entries on the unit circle that can be realized using a reflecting element. These assumptions result in $\mathbf{B}_k^H\mathbf{B}_k = L\mathbf{I}_N$. In classical non-Bayesian parameter estimation, the ML estimator is widely used, which maximizes the likelihood function of the received observation over the

unknown parameter. Since the BSs have misspecified received pilot signal models, they will instead maximize the likelihood functions obtained from the misspecified model, leading to misspecified maximum likelihood (MML) estimators. For (8), the MML estimator can be expressed as

$$\hat{\mathbf{g}}_k = \arg \max_{\mathbf{g}_k} f(\mathbf{y}_{pk}; \mathbf{g}_k)$$

$$= \arg \max_{\mathbf{g}_k} \frac{1}{(\pi\sigma_w^2)^L} \exp\left(-\frac{\|\mathbf{y}_{pk} - \sqrt{P_p}\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\mathbf{g}_k\|^2}{\sigma_w^2}\right)$$

$$= \arg \min_{\mathbf{g}_k} \left\|\mathbf{y}_{pk} - \sqrt{P_p}\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\mathbf{g}_k\right\|^2$$

$$= \frac{1}{\sqrt{P_p}}\mathbf{D}_{\mathbf{h}_k}^{-1}(\mathbf{B}_k^H\mathbf{B}_k)^{-1}\mathbf{B}_k^H\mathbf{y}_{pk}$$

$$= \frac{1}{L\sqrt{P_p}}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{B}_k^H\mathbf{y}_{pk}. \quad (9)$$

In the following subsections, we describe the behavior of this estimator for two different choices of the $\mathbf{B}_k$ matrices.

A. Case 1: The RISs Adopt the Same Configuration Sequence

We discussed in the introduction that in the absence of inter-operator cooperation, it is highly likely that the RISs will use the same standardized sequence of configurations during the channel estimation phase, which corresponds to $\mathbf{B}_1 = \mathbf{B}_2 = \mathbf{B}$.[1] In this case, (9) becomes

$$\hat{\mathbf{g}}_k = \mathbf{g}_k + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{D}_{\mathbf{q}_k}\mathbf{p}_k + \frac{1}{L\sqrt{P_p}}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{B}^H\mathbf{w}_{pk}. \quad (10)$$

Since we consider the channels as deterministic parameters, we obtain the probability distribution

$$\hat{\mathbf{g}}_k \sim \mathcal{CN}\left(\mathbf{g}_k + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{D}_{\mathbf{q}_k}\mathbf{p}_k, \frac{\sigma_w^2}{LP_p}(\mathbf{D}_{\mathbf{h}_k}^H\mathbf{D}_{\mathbf{h}_k})^{-1}\right). \quad (11)$$

We notice that $\hat{\mathbf{g}}_k$ is biased; that is, $\mathbf{b}_k \triangleq \mathbb{E}[\hat{\mathbf{g}}_k - \mathbf{g}_k] = \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{D}_{\mathbf{Q}_k}\mathbf{p}_k \neq \mathbf{0}$. The estimator bias does not vanish when increasing P_p or L, or decreasing σ_w^2. Hence, this estimator is not asymptotically unbiased. This is a new instance of an extensively studied phenomenon in the massive multiple-input multiple-output (MIMO) literature, namely pilot contamination [2], [3]. Interestingly, the RISs cause pilot contamination even between two non-overlapping frequency bands, which has not been recognized in the previous literature.

B. Case 2: The RISs Adopt Different Configuration Sequences

In this section, we consider the generic case of $\mathbf{B}_1 \neq \mathbf{B}_2$. To motivate the proposed method for configuring $\mathbf{B}_1$ and $\mathbf{B}_2$, we first consider the case where the BSs are aware of the true signal model in (3), and therefore can estimate both $\mathbf{g}_k$ and $\mathbf{r}_k \triangleq \mathbf{D}_{\mathbf{q}_k}\mathbf{p}_{\mathbf{k}}$. The resulting system model can be expressed as

$$\mathbf{y}_{p1} = \sqrt{P_p}\begin{bmatrix}\mathbf{B}_1\mathbf{D}_{\mathbf{h}_1} & \mathbf{B}_2\end{bmatrix}\begin{bmatrix}\mathbf{g}_1 \\ \mathbf{r}_1\end{bmatrix} + \mathbf{w}_{p1}, \quad (12a)$$

$$\mathbf{y}_{p2} = \sqrt{P_p}\begin{bmatrix}\mathbf{B}_2\mathbf{D}_{\mathbf{h}_2} & \mathbf{B}_1\end{bmatrix}\begin{bmatrix}\mathbf{g}_2 \\ \mathbf{r}_2\end{bmatrix} + \mathbf{w}_{p2}. \quad (12b)$$

---

[1]The analysis in this paper can be easily extended to the case when $\mathbf{B}_1 = \mathbf{U}\mathbf{B}_2$ for some unitary matrix $\mathbf{U}$, so that the configuration sequences have identical spans. It is the overlap of the spans that can cause issues.

In (12), a known linear transformation is applied to the parameter vector of interest in the presence of additive noise. Consequently, the ML estimates of UE 1's channels become

$$\begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{r}}_1 \end{bmatrix} = \frac{1}{\sqrt{P_p}} \begin{bmatrix} L\mathbf{D}_{\mathbf{h}_1}^H \mathbf{D}_{\mathbf{h}_1} & \mathbf{D}_{\mathbf{h}_1}^H \mathbf{B}_1^H \mathbf{B}_2 \\ \mathbf{B}_2^H \mathbf{B}_1 \mathbf{D}_{\mathbf{h}_1} & L\mathbf{I}_N \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_{\mathbf{h}_1}^H \mathbf{B}_1^H \\ \mathbf{B}_2^H \end{bmatrix} \mathbf{y}_{p1}.$$

(13)

Note that in this case, the total dimension of the unknown parameter vector is $2N$, hence, at least $2N$ independent observations are required for the matrix inverse to exist.[2] The structure in (13) applies to UE 2 with alternated indices, and it gives the ML estimator, which is both unbiased and efficient, since (12) is a linear observation model with additive Gaussian noise [27, Th. 7.3]. Hence, (13) is unbiased irrespective of other parameters such as σ_w^2, L, and P_p, and it achieves the Cramér-Rao Lower Bound (CRLB), which provides a lower bound on the MSE of any unbiased estimator [28]. It has to be noted that when $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$, (13) becomes

$$\begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{r}}_1 \end{bmatrix} = \frac{1}{L\sqrt{P_p}} \begin{bmatrix} \mathbf{D}_{\mathbf{h}_1}^H \mathbf{D}_{\mathbf{h}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_{\mathbf{h}_1}^H \mathbf{B}_1^H \\ \mathbf{B}_2^H \end{bmatrix} \mathbf{y}_{p1}$$

$$= \frac{1}{L\sqrt{P_p}} \begin{bmatrix} \mathbf{D}_{\mathbf{h}_1}^{-1} \mathbf{B}_1^H \\ \mathbf{B}_2^H \end{bmatrix} \mathbf{y}_{p1}.$$

(14)

Note that the expression for $\hat{\mathbf{g}}_1$ in (14) is the same as in (9). This shows that when $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$, the MML in (9) coincides with the ML estimator; that is, the misspecified model is sufficient when the configuration sequences are designed to alleviate pilot interference because the missing terms anyway vanish in the receiver processing. The probability distribution of $\hat{\mathbf{g}}_k$ in this case can be expressed as

$$\hat{\mathbf{g}}_k \sim \mathcal{CN} \left(\mathbf{g}_k, \frac{\sigma_w^2}{LP_p} (\mathbf{D}_{\mathbf{h}_k}^H \mathbf{D}_{\mathbf{h}_k})^{-1} \right),$$

(15)

which shows that choosing the RIS configuration sequences such that $\mathbf{B}_1$ and $\mathbf{B}_2$ remove the bias from the MML estimator. However, the major setback of this approach is that the minimum number of observations required for this channel estimation procedure is $2N$ instead of N, due to the fact that the $2N$-many L-dimensional columns must all be mutually orthogonal, for which $L \geq 2N$ is required. Considering that the estimator bias in (11) does not vanish with increasing L, this is a necessary sacrifice. Hence, it has to be noted that the number of pilot transmissions increases linearly with the number of RISs deployed in proximity.

C. MSE During Channel Estimation

The estimation error can be quantified through the MSE, which is the trace of the error covariance matrix. We will derive the MSE in this section. In this derivation, we do not assume a particular choice of $\mathbf{B}_1$ and $\mathbf{B}_2$, but we utilize the basic assumption $\mathbf{B}_k \mathbf{B}_k^H = L\mathbf{I}_N$. Consequently, we use $\mathbf{b}_k$

[2]The pseudo-inverse could be used when there are fewer observations, but it will not provide a useful estimate.

to denote the potential estimator bias. We can then compute the error covariance matrix as

$$\boldsymbol{\Sigma}_{e,k} = \mathbb{E} \left[(\hat{\mathbf{g}}_k - \mathbf{g}_k)(\hat{\mathbf{g}}_k - \mathbf{g}_k)^H \right]$$

$$= \mathbf{b}_k \mathbf{b}_k^H + \frac{1}{LP_p} \mathbb{E} \left[\mathbf{D}_{\mathbf{h}_k}^{-1} \mathbf{w}_{pk} \mathbf{w}_{pk}^H \mathbf{D}_{\mathbf{h}_k}^{-H} \right]$$

$$= \mathbf{b}_k \mathbf{b}_k^H + \frac{\sigma_w^2}{LP_p} \left(\mathbf{D}_{\mathbf{h}_k} \mathbf{D}_{\mathbf{h}_k}^H \right)^{-1}.$$

(16)

Consequently, the trace of the error covariance matrix becomes

$$\text{tr}(\boldsymbol{\Sigma}_{e,k}) = \|\mathbf{b}_k\|^2 + \frac{\sigma_w^2}{LP_p} \sum_{n=1}^{N} \frac{1}{|h_{kn}|^2}.$$

(17)

Note that for high P_p, L, and/or low σ_w^2, the second term in (17) vanishes, and the trace of the error covariance converges to $\|\mathbf{b}_k\|^2$ which depends on the configurations of $\mathbf{B}_1, \mathbf{B}_2$. For the two previously considered cases, we have

$$\|\mathbf{b}_k\|^2 = \begin{cases} \sum_{n=1}^{N} \frac{|r_{kn}|^2}{|h_{kn}|^2} & \mathbf{B}_1 = \mathbf{B}_2, \\ 0 & \mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}. \end{cases}$$

(18)

This result shows that configuring the RISs such that $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ removes the asymptotic floor on the average MSE, which comes from the energy of the estimator bias. On the other hand, when the intended RIS-BS links $\mathbf{h}_1, \mathbf{h}_2$ are strong relative to the unintended and unknown overall link $\mathbf{r}_k$, the estimator bias will be weaker and the performance loss associated with choosing $\mathbf{B}_1 = \mathbf{B}_2$ will be lower. Nevertheless, pilot contamination results in a fundamental error floor, even if the RISs are utilized in different bands. In the next section, we consider the estimation of data based on the channel estimation performed in this section and analyze the consequence of pilot contamination in this phase.

IV. DATA SIGNAL ESTIMATION WITH DETERMINISTIC CHANNELS

The channel estimation is followed by data transmission over the same deterministic channel as in Section III. The receiver can use the channel estimate derived in the last section when determining the transmitted signal. Practical channels are never fully deterministic but might have a long coherence time. Moreover, the impact of estimation errors is only relevant when the data packet has a modest size so we cannot afford to spend much resources on pilots. For this reason, we cannot consider the channel capacity as performance metric but will instead consider the MSE.

Defining the data signal transmitted by the kth UE as $x_k \sim \mathcal{CN}(0, 1)$, we can express the received data as

$$y_1 = \sqrt{P_d}(\mathbf{h}_1^T \hat{\boldsymbol{\Phi}}_1 \mathbf{g}_1 + \mathbf{q}_1^T \hat{\boldsymbol{\Phi}}_2 \mathbf{p}_1)x_1 + w_1,$$

(19a)

$$y_2 = \sqrt{P_d}(\mathbf{h}_2^T \hat{\boldsymbol{\Phi}}_2 \mathbf{g}_2 + \mathbf{q}_2^T \hat{\boldsymbol{\Phi}}_1 \mathbf{p}_2)x_2 + w_2,$$

(19b)

where $w_k \sim \mathcal{CN}(0, \sigma_w^2)$ denotes the receiver noise, P_d denotes the data transmission power, and the RIS configuration matrices $\hat{\boldsymbol{\Phi}}_k$ are selected based on the estimated channels to

maximize the average gain of the desired cascaded channel as shown in [12, Sec. II]:

$$\hat{\phi}_{kn} = \arg(h_{kn}) + \arg(\hat{g}_{kn}),$$
$$\hat{\boldsymbol{\Phi}}_k = \text{diag}\left(e^{-j\hat{\phi}_{k1}}, \ldots, e^{-j\hat{\phi}_{kN}}\right). \quad (20)$$

Note that we use two different notations for the noise acting upon the pilot and data transmissions to emphasize that they are two distinct i.i.d. random variables which becomes an important fact when one considers data signal estimation based on finite-SNR channel estimates. However, since the BSs are unaware of the unintended reflections and base their data reception on the previously obtained channel estimates, they assume the following misspecified received data signal models:

$$\hat{y}_1 = \sqrt{P_d}\mathbf{h}_1^T\hat{\boldsymbol{\Phi}}_1\hat{\mathbf{g}}_1 x_1 + w_1, \quad (21a)$$
$$\hat{y}_2 = \sqrt{P_d}\mathbf{h}_2^T\hat{\boldsymbol{\Phi}}_2\hat{\mathbf{g}}_2 x_2 + w_2. \quad (21b)$$

Introducing the notation $m_k \triangleq \sqrt{P_d}(\mathbf{h}_k^T\hat{\boldsymbol{\Phi}}_k\mathbf{g}_k + \mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_j\mathbf{p}_k)$ for $j, k \in \{1, 2\}, j \neq k$, and $\hat{m}_k \triangleq \sqrt{P_d}\mathbf{h}_k^T\hat{\boldsymbol{\Phi}}_k\hat{\mathbf{g}}_k$, (19) and (21) can be expressed as

$$y_k = m_k x_k + w_k, \quad k = 1, 2, \quad (22a)$$
$$\hat{y}_k = \hat{m}_k x_k + w_k, \quad k = 1, 2. \quad (22b)$$

Based on the misspecified observation model in (22b), the BSs estimate x_k by using the misspecified MMSE estimator

$$\hat{x}_k = \frac{\hat{m}_k^*}{|\hat{m}_k|^2 + \sigma_w^2} y_k, \quad k = 1, 2. \quad (23)$$

In this section, we consider the MSE between x_k and $\hat{x}_k$ as the performance metric for the data transmission. We derive the data estimation MSE for UE k as

$$\mathbb{E}\left[|x_k - \hat{x}_k|^2\right] = 1 + \mathbb{E}\left[|\hat{x}_k|^2\right] - 2\text{Re}(\mathbb{E}[x_k\hat{x}_k^*])$$
$$= 1 + \mathbb{E}\left[\frac{|\hat{m}_k|^2(|m_k|^2 + \sigma_w^2)}{(|\hat{m}_k|^2 + \sigma_w^2)^2}\right]$$
$$- 2\text{Re}\left(\mathbb{E}\left[\frac{\hat{m}_k m_k^*}{|\hat{m}_k|^2 + \sigma_w^2}\right]\right)$$
$$= 1 + \mathbb{E}\left[\frac{|\hat{m}_k|^2(|m_k|^2 + \sigma_w^2) - 2\text{Re}(\hat{m}_1 m_1^*)(|\hat{m}_k|^2 + \sigma_w^2)}{(|\hat{m}_k|^2 + \sigma_w^2)^2}\right]$$
$$= 1 + \mathbb{E}\left[\frac{|\hat{m}_k|^2(|m_k|^2 + \sigma_w^2) - 2\text{Re}(\hat{m}_1 m_1^*)(|\hat{m}_k|^2 + \sigma_w^2)}{(|\hat{m}_k|^2 + \sigma_w^2)^2}\right]$$
$$+ \mathbb{E}\left[\frac{\sigma_w^2(|m_k|^2 + \sigma_w^2) - \sigma_w^2(|m_k|^2 + \sigma_w^2)}{(|\hat{m}_k|^2 + \sigma_w^2)^2}\right]$$
$$= 1 + \mathbb{E}\left[\frac{(|\hat{m}_k|^2 + \sigma_w^2)(|m_k|^2 + \sigma_w^2 - 2\text{Re}(\hat{m}_k m_k^*))}{(\hat{m}_k^2 + \sigma_w^2)^2}\right]$$
$$- \mathbb{E}\left[\frac{\sigma_w^2(|m_k|^2 + \sigma_w^2)}{(\hat{m}_k^2 + \sigma_w^2)^2}\right]$$
$$= \mathbb{E}\left[\frac{|m_k - \hat{m}_k|^2 + 2\sigma_w^2}{|\hat{m}_k|^2 + \sigma_w^2}\right] - \mathbb{E}\left[\frac{\sigma_w^2(|m_k|^2 + \sigma_w^2)}{(|\hat{m}_k|^2 + \sigma_w^2)^2}\right]. \quad (24)$$

Defining $\epsilon_k \triangleq m_k - \hat{m}_k$, (24) can be rewritten as

$$\mathbb{E}[|x_k - \hat{x}_k|^2] = \mathbb{E}\left[\frac{|\epsilon_k|^2 + 2\sigma_w^2}{|m_k - \epsilon_k|^2 + \sigma_w^2} - \frac{\sigma_w^2(|m_k|^2 + \sigma_w^2)}{(|m_k - \epsilon_k|^2 + \sigma_w^2)^2}\right]. \quad (25)$$

To examine the impact of pilot contamination on the data estimation performance more clearly, we now consider channel estimation at high SNRs, so that the estimation error only comes from the estimator bias, i.e., pilot contamination. This happens when L or P_p is high and/or σ_w^2 is low, which results in the estimator covariances in (11) and (15) becoming zero. For notational convenience, we consider the case where P_p is arbitrarily large so that $\lim_{P_p \to \infty} \hat{\mathbf{g}}_k = \mathbf{g}_k + \mathbf{b}_k$, where

$$\mathbf{b}_k = \begin{cases} \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{D}_{\mathbf{q}_k}\mathbf{p}_k & \mathbf{B}_1 = \mathbf{B}_2, \\ \mathbf{0} & \mathbf{B}_1^H\mathbf{B}_2 = \mathbf{0}. \end{cases} \quad (26)$$

A. Data MSE with Channel Estimation at High SNR

In (25), ϵ_k and m_k are functions of $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$, therefore as $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ converge to their means, ϵ_k and m_k become

$$\overline{m}_k = \sqrt{P_d}(\mathbf{h}_k^T\bar{\boldsymbol{\Phi}}_k\mathbf{g}_k + \mathbf{q}_k^T\bar{\boldsymbol{\Phi}}_j\mathbf{p}_k), \quad (27a)$$
$$\bar{\epsilon}_k = \sqrt{P_d}(\mathbf{q}_k^T\bar{\boldsymbol{\Phi}}_j\mathbf{p}_k - \mathbf{h}_k^T\bar{\boldsymbol{\Phi}}_k\mathbf{b}_k), \quad (27b)$$

for $j, k \in \{1, 2\}$ and $j \neq k$. $\bar{\boldsymbol{\Phi}}_k$ denotes the RIS configuration computed according to (20) with $\hat{\mathbf{g}}_k = \mathbf{g}_k + \mathbf{b}_k$. At high SNR, the MSE in (25) can be rewritten as

$$\text{MSE} = \frac{|\bar{\epsilon}_k|^2 + 2\sigma_w^2}{|\overline{m}_k - \bar{\epsilon}_k|^2 + \sigma_w^2} - \frac{\sigma_w^2(|\overline{m}_k|^2 + \sigma_w^2)}{(|\overline{m}_k - \bar{\epsilon}_k|^2 + \sigma_w^2)^2}. \quad (28)$$

This is a practically achievable limit since RIS-aided systems require long pilot sequences over a narrow bandwidth, thus, the effective SNR (proportional to $P_p L$) during pilot transmission can be much larger than in the data transmission phase.

B. Data MSE with Transmission at High SNR

In the previous section, we obtained the expression for the data MSE when the channels are estimated with a high pilot SNR, while the data transmission is done at an arbitrary SNR. To study the case when also the data transmission is conducted at a high SNR, we let $\sigma_w^2 \to 0$, which results in the limit

$$\lim_{\sigma_w^2 \to 0} \text{MSE} = \frac{|\bar{\epsilon}_k|^2}{|\overline{m}_k - \bar{\epsilon}_k|^2}. \quad (29)$$

Note that the resulting expression denotes the ratio between the estimated overall single input single output (SISO) channel $\hat{m}_k$'s power and the mismatch parameter ϵ_k's power. Recall that $\mathbf{b}_k$ depends on which sequence of RIS configurations is utilized. For $\mathbf{B}_1 = \mathbf{B}_2$, we can obtain ϵ_k as

$$\epsilon_k = \sqrt{P_d}\mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_j\mathbf{p}_k - \sqrt{P_d}\mathbf{h}_k^T\hat{\boldsymbol{\Phi}}_k\mathbf{D}_{\mathbf{h}_1}^{-1}\mathbf{D}_{\mathbf{q}_1}\mathbf{p}_1$$
$$= \sqrt{P_d}\mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_j\mathbf{p}_k - \sqrt{P_d}\hat{\phi}_k\mathbf{D}_{\mathbf{h}_k}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{D}_{\mathbf{q}_1}\mathbf{p}_1$$
$$= \sqrt{P_d}\mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_j\mathbf{p}_k - \sqrt{P_d}\hat{\phi}_k\mathbf{D}_{\mathbf{q}_1}\mathbf{p}_1$$
$$= \sqrt{P_d}\mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_j\mathbf{p}_k - \sqrt{P_d}\mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_k\mathbf{p}_k$$
$$= \sqrt{P_d}\mathbf{q}_k^T(\hat{\boldsymbol{\Phi}}_j - \hat{\boldsymbol{\Phi}}_k)\mathbf{p}_k. \quad (30)$$

On the other hand, $\mathbf{B}_1^H\mathbf{B}_2 = \mathbf{0}$ removes $\mathbf{b}_k$ for $k \in \{1, 2\}$. Consequently, we have

$$\epsilon_k = \begin{cases} \sqrt{P_d}\mathbf{q}_k^T(\hat{\boldsymbol{\Phi}}_j - \hat{\boldsymbol{\Phi}}_k)\mathbf{p}_k & \mathbf{B}_1 = \mathbf{B}_2, \\ \sqrt{P_d}\mathbf{q}_k^T\hat{\boldsymbol{\Phi}}_j\mathbf{p}_k & \mathbf{B}_1^H\mathbf{B}_2 = \mathbf{0}. \end{cases} \quad (31)$$

We notice that ϵ_k corresponds to only the unintended reflection path when $\mathbf{B}_1^H\mathbf{B}_2 = \mathbf{0}$. On the other hand, $\mathbf{B}_1 = \mathbf{B}_2$ yields an expression depending on the difference between the two RISs' configurations during data transmission. Since each RIS is configured based on the (estimated) channels of their respective users, it is highly unlikely that the configurations will be close. Moreover, it has to be noted that the RIS configuration of the non-serving RIS is different between the two cases since the channel estimates are also different.

V. CHANNEL ESTIMATION BASED ON CORRELATED RAYLEIGH FADING PRIORS

We now switch focus to consider fading channels that can be modeled using the Bayesian framework. In this section, we consider channel estimation and assume that all the UE-RIS channels exhibit spatially correlated Rayleigh fading: $\mathbf{g}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{g}_k})$ and $\mathbf{p}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{p}_k})$. The covariance matrices $\boldsymbol{\Sigma}_{\mathbf{g}_k}, \boldsymbol{\Sigma}_{\mathbf{p}_k} \in \mathbb{C}^{N \times N}$ are generic positive semi-definite matrices. In addition, the BSs know $\mathbf{h}_k$ perfectly while they consider $\mathbf{q}_k$ as deterministic and unknown channels for $k = 1, 2$. The pilot transmission model assumed by the BSs for $k = 1, 2$ can be expressed as

$$\hat{\mathbf{y}}_{p1} = \sqrt{P_p}\mathbf{B}_1\mathbf{D}_{\mathbf{h}_1}\mathbf{g}_1 + \mathbf{w}_{p1} \in \mathbb{C}^L, \quad (32a)$$

$$\hat{\mathbf{y}}_{p2} = \sqrt{P_p}\mathbf{B}_2\mathbf{D}_{\mathbf{h}_2}\mathbf{g}_2 + \mathbf{w}_{p2} \in \mathbb{C}^L. \quad (32b)$$

Based on (32), the BSs can estimate $\mathbf{g}_k$ via a misspecified MMSE estimator, which can be expressed as

$$\hat{\mathbf{g}}_1 = \frac{1}{\sqrt{P_p}}\boldsymbol{\Sigma}_{\mathbf{g}_1}\mathbf{D}_{\mathbf{h}_1}^H\mathbf{B}_1^H$$
$$\times \left(\mathbf{B}_1\mathbf{D}_{\mathbf{h}_1}\boldsymbol{\Sigma}_{\mathbf{g}_1}\mathbf{D}_{\mathbf{h}_1}^H\mathbf{B}_1^H + \frac{\sigma_w^2}{P_p}\mathbf{I}_L\right)^{-1}\mathbf{y}_{p1}, \quad (33a)$$

$$\hat{\mathbf{g}}_2 = \frac{1}{\sqrt{P_p}}\boldsymbol{\Sigma}_{\mathbf{g}_2}\mathbf{D}_{\mathbf{h}_2}^H\mathbf{B}_2^H$$
$$\times \left(\mathbf{B}_2\mathbf{D}_{\mathbf{h}_2}\boldsymbol{\Sigma}_{\mathbf{g}_2}\mathbf{D}_{\mathbf{h}_2}^H\mathbf{B}_2^H + \frac{\sigma_w^2}{P_p}\mathbf{I}_L\right)^{-1}\mathbf{y}_{p2}. \quad (33b)$$

The diagonal entries of the error covariance matrix represent the MSEs of the corresponding channel entry. We first define $\mathbf{r}_k \triangleq \mathbf{D}_{\mathbf{q}_k}\mathbf{p}_k$, which results in $\mathbf{r}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{r}_k})$ where $\boldsymbol{\Sigma}_{\mathbf{r}_k} \triangleq \mathbf{D}_{\mathbf{q}_k}\boldsymbol{\Sigma}_{\mathbf{p}_k}\mathbf{D}_{\mathbf{q}_k}^H$. To simplify the representation of the channel estimation error covariance matrix, we introduce the following notation:

$$\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}} \triangleq \mathbb{E}[\mathbf{g}_k\hat{\mathbf{y}}_{pk}^H] = \mathbb{E}[\mathbf{g}_k\mathbf{y}_{pk}^H] = \sqrt{P_p}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H, \quad (34a)$$

$$\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} \triangleq \mathbb{E}[\hat{\mathbf{y}}_{pk}\hat{\mathbf{y}}_{pk}^H] = P_p\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H + \sigma_w^2\mathbf{I}_L, \quad (34b)$$

$$\mathbf{C}_{\mathbf{y}\mathbf{y}} \triangleq \mathbb{E}[\mathbf{y}_{pk}\mathbf{y}_{pk}^H] = \mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} + P_p\mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H. \quad (34c)$$

With this notation, the error covariance matrix can be represented as

$$\mathbb{E}[(\mathbf{g}_k - \hat{\mathbf{g}}_k)(\mathbf{g}_k - \hat{\mathbf{g}}_k)^H] =$$
$$= \boldsymbol{\Sigma}_{\mathbf{g}_k} + \mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{y}\mathbf{y}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H - 2\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H$$
$$= \boldsymbol{\Sigma}_{\mathbf{g}_k} + \mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}(\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} + P_p\mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H)\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H$$
$$- 2\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H$$
$$= \underbrace{\boldsymbol{\Sigma}_{\mathbf{g}_k} - \mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H}_{\text{error covariance for } \mathbf{y}=\hat{\mathbf{y}}} + \underbrace{P_p\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H}_{\text{term coming from pilot contamination}}. \quad (35)$$

Note that the additive term in (35) coming from pilot contamination depends on $\boldsymbol{\Sigma}_{\mathbf{r}_k}$. If we consider the case where the channel $\mathbf{r}_k$ does not exist, that is, $\boldsymbol{\Sigma}_{\mathbf{r}_k} = \mathbf{0}$, then we would obtain the error covariance in the form of a typical MMSE estimation error covariance matrix. It is also important to address the dependence of the term coming from pilot contamination on P_p: while the terms $\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}$ scale with $\sqrt{P_p}$ each, the terms $\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}$ scale with $1/P_p$ each, and along with the leading P_p multiplier, we can see that P_p-dependent terms cancel each other out, hence leaving a non-vanishing pilot contamination term.

A. High-SNR Channel Estimation

Now, we investigate the behavior of the error covariance matrix in (35) when σ_w^2 is low or P_p is high. Note that the first term in (35) is already the error covariance matrix for an MMSE estimator without misspecification that estimates $\mathbf{g}_k$ based on the observation $\hat{\mathbf{y}}_{pk}$. Therefore, we know that this term vanishes at high SNR. Consequently, the asymptotic error covariance matrix is governed by the high-SNR behavior of the term coming from pilot contamination, that is

$$\lim_{\sigma_w^2 \to 0} \mathbb{E}[(\mathbf{g}_k - \hat{\mathbf{g}}_k)(\mathbf{g}_k - \hat{\mathbf{g}}_k)^H]$$
$$= \lim_{\sigma_w^2 \to 0} P_p\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{C}_{\mathbf{g}\hat{\mathbf{y}}}^H$$
$$= P_p^2\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H\left(P_p\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H + \sigma_w^2\mathbf{I}_L\right)^{-1}$$
$$\times \mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H\left(P_p\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H + \sigma_w^2\mathbf{I}_L\right)^{-1}\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}$$
$$= \lim_{\sigma_w^2 \to 0} \boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H\left(\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H + \frac{\sigma_w^2}{P_p}\mathbf{I}_L\right)^{-1}$$
$$\times \mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H\left(\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H + \frac{\sigma_w^2}{P_p}\mathbf{I}_L\right)^{-1}\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}. \quad (36)$$

Note that since $\mathbf{B}_k^H\mathbf{B}_k = L\mathbf{I}_N$, the pseudoinverse corresponds to $\mathbf{B}_k^\dagger = \frac{1}{L}\mathbf{B}_k^H$. Consequently, we have

$$\lim_{\sigma_w^2 \to 0} \mathbb{E}[(\mathbf{g}_k - \hat{\mathbf{g}}_k)(\mathbf{g}_k - \hat{\mathbf{g}}_k)^H]$$
$$= \frac{1}{L^4}\boldsymbol{\Sigma}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{h}_k}^H\mathbf{B}_k^H\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}^{-H}\boldsymbol{\Sigma}_{\mathbf{g}_k}^{-1}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{B}_k^H$$
$$\times \mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}^{-H}\boldsymbol{\Sigma}_{\mathbf{g}_k}^{-1}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{B}_k^H\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}\boldsymbol{\Sigma}_{\mathbf{g}_k}$$
$$= \frac{1}{L^2}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{B}_k^H\mathbf{B}_j\boldsymbol{\Sigma}_{\mathbf{r}_k}\mathbf{B}_j^H\mathbf{B}_k\mathbf{D}_{\mathbf{h}_k}^{-H}. \quad (37)$$

From (37), one can observe that the asymptotic behavior of the pilot contamination depends on the choice of $\mathbf{B}_1$ and $\mathbf{B}_2$. To analyze this, we consider two cases.

1) Case 1: $\mathbf{B}_1 = \mathbf{B}_2$: In this case, we have that $\mathbf{B}_k^H \mathbf{B}_j = L\mathbf{I}_N$ for $j, k = 1, 2$ and $j \neq k$. Consequently, (37) becomes

$$\frac{1}{L^2}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{B}_k^H \mathbf{B}_j \mathbf{\Sigma}_{\mathbf{r}_k} \mathbf{B}_j^H \mathbf{B}_k \mathbf{D}_{\mathbf{h}_k}^{-H} = \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H}. \quad (38)$$

Note that this result does not depend on L, which implies that this channel estimation error caused by pilot contamination cannot be eliminated by increasing the number of pilots when $\mathbf{B}_1 = \mathbf{B}_2$.

2) The RISs are configured such that $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$: In this case, $\mathbf{B}_k^H \mathbf{B}_j = \mathbf{0}$ for $k \neq j$ and $k, j \in \{1, 2\}$. This implies that (37) is zero; thus, BS k can estimate $\mathbf{g}_k$ even without being aware of $\mathbf{r}_k$. Also note that (37) implies that any choice of $\mathbf{B}_1$ and $\mathbf{B}_2$ that does not satisfy $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ will result in pilot contamination while estimating correlated Rayleigh fading channels. We can summarize the high-SNR behavior of the channel estimation error covariance matrix as

$$\lim_{\sigma_w^2 \to 0} \mathbb{E}[(\mathbf{g}_k - \hat{\mathbf{g}}_k)(\mathbf{g}_k - \hat{\mathbf{g}}_k)^H]$$
$$= \begin{cases} \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H} & \mathbf{B}_1 = \mathbf{B}_2, \\ \mathbf{0} & \mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}. \end{cases} \quad (39)$$

This result shows that in order to estimate $\mathbf{g}_k$ reliably, it is necessary to configure the RISs such that $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$. Note that (39) shows us that the cascaded channel over the foreign operator's RIS acts as channel estimation noise that does not vanish with high transmission power. Additionally, the link between the operator's own RIS and BS counteracts the noise component. This result is also in line with the result obtained in Section III, in (16), that is, when the channels get a prior distribution, $\mathbf{b}_k$ becomes a complex Gaussian random vector with the covariance matrix provided in (39).

VI. CAPACITY LOWER BOUND FOR RELIABLE COMMUNICATION UNDER IMPERFECT CSI

In this section, we compute a lower bound on the ergodic capacity based on the imperfect CSI obtained in the previous section via channel estimation. In particular, we consider the impact of pilot contamination and the effect of the signal model misspecification on the channel capacity. We derive the channel capacity lower bound for the two cases considered in Section V.

A. Capacity Lower Bound of a SISO Channel with Channel Side Information

Consider a generic SISO system with the following received signal model:

$$y = hx + w \quad (40)$$

with $w \sim \mathcal{CN}(0, \sigma_w^2)$ and $x \sim \mathcal{CN}(0, 1)$. Suppose that the receiver has partial information on h, denoted by Ω. Then the capacity lower bound is [29, Eq. 2.46]

$$C \geq \mathbb{E}_\Omega \left[\log_2 \left(1 + \frac{|\mathbb{E}[h|\Omega]|^2}{\text{Var}(h|\Omega) + \text{Var}(w|\Omega)} \right) \right]. \quad (41)$$

This bound is valid under certain conditions, which can be listed as follows [29, Section 2.3.5]:

- The noise w has zero-mean conditioned on Ω, that is, $\mathbb{E}[w|\Omega] = 0$.
- The transmitted signal x and the noise w are uncorrelated conditioned on Ω, that is, $\mathbb{E}[xw^*|\Omega] = \mathbb{E}[x|\Omega]\mathbb{E}[w^*|\Omega]$.
- The received signal hx and the noise w are uncorrelated conditioned on Ω, that is, $\mathbb{E}[hxw^*|\Omega] = \mathbb{E}[hx|\Omega]\mathbb{E}[w^*|\Omega]$.

In our setup, the data signal model for the two users can be expressed as

$$y_1 = \sqrt{P_d}(\mathbf{h}_1^T \mathbf{\Phi}_1 \mathbf{g}_1 + \mathbf{q}_1^T \mathbf{\Phi}_2 \mathbf{p}_1)x_1 + w_1, \quad (42a)$$
$$y_2 = \sqrt{P_d}(\mathbf{h}_2^T \mathbf{\Phi}_2 \mathbf{g}_2 + \mathbf{q}_2^T \mathbf{\Phi}_1 \mathbf{p}_2)x_2 + w_2, \quad (42b)$$

where $x_k \sim \mathcal{CN}(0, 1)$ denotes the transmitted data for $k = 1, 2$ and P_d denotes the data transmission power. During the data transmission phase, both RISs are configured to phase-align the cascaded channel, that is:

$$\phi_{kn} = \arg(h_{kn}) + \arg(\hat{g}_{kn}). \quad (43)$$

For both BSs, we consider the side information Ω as the knowledge of $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$, $\mathbf{h}_1$, $\mathbf{h}_2$, $\hat{\mathbf{g}}_1$, and $\hat{\mathbf{g}}_2$. Consequently, the outer expectation in (41) refers to the expectation with respect to the marginal distributions of $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$.

Lemma 1. *The system setup described by the signal model in* (42) *that uses the RIS configurations described in* (43) *satisfies the three regularity conditions required by the capacity bound in* (41).

Proof. First, let us identify the h and w that we had defined in our system setup in (42):

$$h = \sqrt{P_d}(\mathbf{h}_k^T \mathbf{\Phi}_k \mathbf{g}_k + \mathbf{q}_k^T \mathbf{\Phi}_j \mathbf{p}_k) \quad (44a)$$
$$w = w_k \quad (44b)$$

We can now prove that the conditional mean of the noise conditioned on the channel side information, which we can describe as $\Omega = \hat{\mathbf{g}}_k$, is zero. Note that $\hat{\mathbf{g}}_k$ and w_k are independent: for $\mathbf{B}_1 = \mathbf{B}_2$, the less trivial case, we have that $\hat{\mathbf{g}}_k = \mathbf{g}_k + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{r}_k$, that is, no dependence on the noise. At lower SNRs, the vanishing components contain the realizations of the noise received during channel estimation, and considering the fact that the receiver noise is white over time, that does not affect the independence between $\hat{\mathbf{g}}_k$ and w_k either. Therefore, $\mathbb{E}[w_k|\hat{\mathbf{g}}_k] = \mathbb{E}[w_k] = 0$.

It is also straightforward to prove that the transmitted signal is uncorrelated with the noise conditioned on Ω due to the fact that x_k and w_k are independent of Ω individually. So the expression $\mathbb{E}[x_k w_k^*|\Omega]$ does not have anything that depends on Ω, i.e., $\mathbb{E}[x_k w_k^*|\Omega] = \mathbb{E}[x_k w_k^*] = \mathbb{E}[x_k]\mathbb{E}[w_k^*] = \mathbb{E}[x_k|\Omega]\mathbb{E}[w_k^*|\Omega]$ can be obtained, proving that x and w are uncorrelated with each other conditioned on Ω.

The last point is also quite straightforward since Ω does not contain any relations between the overall channel, the receiver noise, and the transmitted signal. Therefore it is easy to claim that this regularity condition also holds, hence the capacity bound provided in (41) is applicable to our system setup. $\square$

B. Capacity Lower Bound with High-SNR Channel Estimates Available at the BSs

If we assume that the channel estimation is performed at a high SNR, we can model the channel estimation error according to (39). Consequently, we can express the channel estimates in terms of the true channels and the channel estimation error as

$$\hat{\mathbf{g}}_k = \mathbf{g}_k + \mathbf{e}_k, \tag{45}$$

where the channel estimation error is

$$\mathbf{e}_k = \begin{cases} \mathbf{0} & \mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}, \\ \mathbf{D}_{\mathbf{h}_k}^{-1} \mathbf{r}_k & \mathbf{B}_1 = \mathbf{B}_2. \end{cases} \tag{46}$$

Hence, BS k knows $\mathbf{g}_k$ perfectly if the RISs are configured such that $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ during channel estimation, and $\mathbf{e}_k = \mathbf{D}_{\mathbf{h}_k}^{-1} \mathbf{r}_k$ when $\mathbf{B}_1 = \mathbf{B}_2$. Also note that even when we have $\mathbf{B}_1 = \mathbf{B}_2$, the channel and the channel estimation error are independent. Consequently, we can rewrite the overall SISO channel as

$$v_k \triangleq \sqrt{P_d}(\phi_k^T \mathbf{D}_{\mathbf{h}_k} \mathbf{g}_k + \phi_j^T \mathbf{r}_k). \tag{47}$$

The mean of the overall SISO channel conditioned on the side information can be expressed as

$$\mathbb{E}[v_k|\Omega] = \sqrt{P_d}(\phi_k^T \mathbf{D}_{\mathbf{h}_k} \mathbb{E}[\mathbf{g}_k|\hat{\mathbf{g}}_k] + \phi_j^T \mathbb{E}[\mathbf{r}_k|\hat{\mathbf{g}}_k]) \tag{48}$$

Here, we can utilize the channel estimate structure provided by (45) and (46). Note that when we consider $\mathbb{E}[\mathbf{g}_k|\hat{\mathbf{g}}_k]$, it can be thought as estimating $\mathbf{g}_k$ based on observing $\hat{\mathbf{g}}_k$ since both $\mathbf{g}_k$ and $\mathbf{e}_k$ and they are independent from each other. The same goes for computing $\mathbb{E}[\mathbf{r}_k|\hat{\mathbf{g}}_k]$. Since the $\mathbf{g}_k - \hat{\mathbf{g}}_k$ and $\mathbf{r}_k - \hat{\mathbf{g}}_k$ are jointly Gaussian, the MMSE estimate, also known as the conditional mean estimate coincides with the linear minimum mean squared error (LMMSE) estimate, therefore, we can use the LMMSE formulation here:

$$\mathbb{E}[\mathbf{g}_k|\hat{\mathbf{g}}_k] = \mathbb{E}[\mathbf{g}_k \hat{\mathbf{g}}_k^H](\mathbb{E}[\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H])^{-1}\hat{\mathbf{g}}_k \tag{49a}$$

$$\mathbb{E}[\mathbf{r}_k|\hat{\mathbf{g}}_k] = \mathbb{E}[\mathbf{r}_k \hat{\mathbf{g}}_k^H](\mathbb{E}[\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H])^{-1}\hat{\mathbf{g}}_k \tag{49b}$$

which can be expressed more explicitly as

$$\mathbb{E}[\mathbf{g}_k|\hat{\mathbf{g}}_k] = \mathbf{\Sigma}_{\mathbf{g}_k}(\mathbf{\Sigma}_{\mathbf{g}_k} + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H})^{-1}\hat{\mathbf{g}}_k, \tag{50a}$$

$$\mathbb{E}[\mathbf{r}_k|\hat{\mathbf{g}}_k] = \begin{cases} \mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H}(\mathbf{\Sigma}_{\mathbf{g}_k} + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H})^{-1}\hat{\mathbf{g}}_k & \mathbf{B}_1 = \mathbf{B}_2, \\ \mathbf{0} & \mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}. \end{cases} \tag{50b}$$

On the other hand, the variance of v_k conditioned on $\hat{\mathbf{g}}_k$ can be expressed as

$$\begin{aligned} \mathrm{Var}(v_k|\hat{\mathbf{g}}_k) &= P_d \phi_k^T \mathbf{D}_{\mathbf{h}_k} \mathrm{Var}(\mathbf{g}_k|\hat{\mathbf{g}}_k) \mathbf{D}_{\mathbf{h}_k}^H \phi_k^* \\ &+ P_d \phi_j^T \mathrm{Var}(\mathbf{r}_k|\hat{\mathbf{g}}_k)\phi_j^* + 2\mathrm{Re}(\phi_k^T \mathbf{D}_{\mathbf{h}_k} \mathbb{E}[\mathbf{g}_k \mathbf{r}_k^H|\hat{\mathbf{g}}_k]\phi_j^*) \\ &- 2\mathrm{Re}(\phi_k^T \mathbf{D}_{\mathbf{h}_k} \mathbb{E}[\mathbf{g}_k|\hat{\mathbf{g}}_k]\mathbb{E}[\mathbf{r}_k^H|\hat{\mathbf{g}}_k]\phi_j^*) \end{aligned} \tag{51}$$

where $\mathbb{E}[\mathbf{g}_k|\hat{\mathbf{g}}_k]$ and $\mathbb{E}[\mathbf{r}_k|\hat{\mathbf{g}}_k]$ are provided by (50). In addition, we can use LMMSE formulation results for $\mathrm{Var}(\mathbf{g}_k|\hat{\mathbf{g}}_k)$ and $\mathrm{Var}(\mathbf{r}_k|\hat{\mathbf{g}}_k)$ which correspond to the error covariance matrices as a result of estimating $\mathbf{g}_k$ and $\mathbf{r}_k$ with an LMMSE

estimator based on the observation $\hat{\mathbf{g}}_k$. Consequently, these two terms can be expressed as

$$\mathrm{Var}(\mathbf{g}_k|\hat{\mathbf{g}}_k) = \mathbb{E}[\mathbf{g}_k \mathbf{g}_k^H] - \mathbb{E}[\mathbf{g}_k \hat{\mathbf{g}}_k^H](\mathbb{E}[\mathbf{g}_k \hat{\mathbf{g}}_k^H])^{-1}\mathbb{E}[\hat{\mathbf{g}}_k \mathbf{g}_k^H], \tag{52a}$$

$$\mathrm{Var}(\mathbf{r}_k|\hat{\mathbf{g}}_k) = \mathbb{E}[\mathbf{r}_k \mathbf{r}_k^H] - \mathbb{E}[\mathbf{r}_k \hat{\mathbf{g}}_k^H](\mathbb{E}[\mathbf{r}_k \hat{\mathbf{g}}_k^H])^{-1}\mathbb{E}[\hat{\mathbf{g}}_k \mathbf{r}_k^H]. \tag{52b}$$

Computing the expectations above and also the cross-term $\mathbb{E}[\mathbf{g}_k \mathbf{r}_k^H|\hat{\mathbf{g}}_k]$, we can obtain the implicit expressions in (51) as

$$\mathrm{Var}(\mathbf{g}_k|\hat{\mathbf{g}}_k) = \mathbf{\Sigma}_{\mathbf{g}_k} - \mathbf{\Sigma}_{\mathbf{g}_k}(\mathbf{\Sigma}_{\mathbf{g}_k} + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H})^{-1}\mathbf{\Sigma}_{\mathbf{g}_k} \tag{53a}$$

$$\begin{aligned} &\mathrm{Var}(\mathbf{r}_k|\hat{\mathbf{g}}_k) \\ &= \mathbf{\Sigma}_{\mathbf{r}_k} - \mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H}(\mathbf{\Sigma}_{\mathbf{g}_k} + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H})^{-1}\mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k} \end{aligned} \tag{53b}$$

$$\begin{aligned} &\mathbb{E}[\mathbf{g}_k \mathbf{r}_k^H|\hat{\mathbf{g}}_k] \\ &= \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H(\mathbf{\Sigma}_{\mathbf{g}_k} + \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H})^{-1}\mathbf{\Sigma}_{\mathbf{r}_k}\mathbf{D}_{\mathbf{h}_k}^{-H} - \mathbf{D}_{\mathbf{h}_k}^{-1}\mathbf{\Sigma}_{\mathbf{r}_k} \end{aligned} \tag{53c}$$

for $\mathbf{B}_1 = \mathbf{B}_2$ and

$$\mathrm{Var}(\mathbf{g}_k|\hat{\mathbf{g}}_k) = \mathbf{0} \tag{54a}$$

$$\mathrm{Var}(\mathbf{r}_k|\hat{\mathbf{g}}_k) = \mathbf{\Sigma}_{\mathbf{r}_k} \tag{54b}$$

$$\mathbb{E}[\mathbf{g}_k \mathbf{r}_k^H|\hat{\mathbf{g}}_k] = \mathbf{0} \tag{54c}$$

for $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$. Obtaining $\mathrm{Var}(w_k|\Omega) = \sigma_w^2$ is straightforward since w_k is independent from Ω. Also note that $\hat{\mathbf{g}}_k$ depends on the choice of $\mathbf{B}_1$ and $\mathbf{B}_2$. As a result, the capacity lower bound can be expressed as

$$C_k \geq \mathbb{E}_\Omega \left[\log_2 \left(1 + \frac{|\mathbb{E}[v_k|\Omega]|^2}{\mathrm{Var}(v_k|\Omega) + \sigma_w^2} \right) \right] \tag{55}$$

with $\mathbb{E}[v_k|\Omega]$ and $\mathrm{Var}(v_k|\Omega)$ taking values according to the choice of $\mathbf{B}_1$ and $\mathbf{B}_2$.

VII. NUMERICAL RESULTS

In this section, we provide numerical examples to demonstrate the implications of the analytical results obtained in Sections III-VI. First, we provide the numerical results for the channel and data estimation MSEs when deterministic channels are considered. For correlated Rayleigh fading, we demonstrate the impact of pilot contamination on channel estimation MSE and the resulting capacity lower bound.

A. Estimation Performance With Deterministic Channels

For deterministic channel estimation, we consider the normalized mean squared error (NMSE) as our performance metric. Moreover, we consider the results for a single UE, since the results for different UEs only differ by the channel realizations. For the deterministic channel $\mathbf{g}_k$, we obtain the NMSE as

$$\mathrm{NMSE} = \frac{\mathrm{MSE}}{\|\mathbf{g}_k\|^2}. \tag{56}$$

In Fig. 2, we plot (17) for different values of P_p, and we also

TABLE I: Parameters used in Figures 2 and 3.

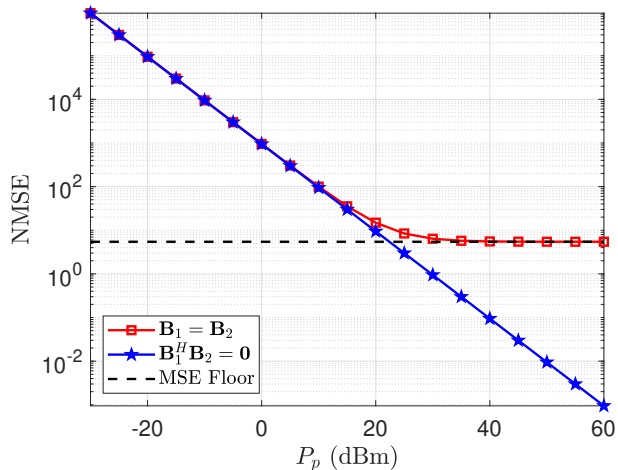| Parameter	Value
P_p or P_d	$-30, -25, \ldots, 40$ dBm[3]
UE-RIS path loss	-80 dB
RIS-BS path loss	-60 dB
σ_w^2	-90 dBm
N	256
L	513

Fig. 2: Pilot transmission power versus the channel estimation NMSE for deterministic channels. Since the prior distribution of the parameter vector is not considered in the non-random parameter estimation framework, it is highly likely to obtain NMSEs greater than 1.

provide the high-SNR floor for the case where $\mathbf{B}_1 = \mathbf{B}_2$. The set of parameters used to generate Fig. 2 and 3 are provided in Table I. On the other hand, the range of transmission power in Fig. 6 is between -10 and 60 dBm. In addition, for Figs. 4, 5 and 6, we consider a 8×8 uniform rectangular array (URA) geometry with $\lambda/2$ spacing in both vertical and horizontal axes. Therefore, the parameter values $N = 64$ and $L = 128$ apply to those figures. Note that at lower transmission powers, the covariance matrix of the estimator acts dominantly, hence, both RIS configurations perform nearly the same. However, after $P_p = 20$ dBm, the power of the estimator bias starts to dominate, and the average MSE for $\mathbf{B}_1 = \mathbf{B}_2$ converges to the floor denoted by the black dashed line, which is given by (18). On the other hand, the average MSE for $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ does not stop there but keeps decreasing towards zero. As mentioned before, the MML estimators used by the BSs coincide with the true ML estimators when the RISs are configured such that $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$.

B. Data Estimation with Deterministic Channels

In Fig. 3, the data estimation MSE performance with the two RIS pilot configurations is analyzed when the channel estimation SNR is high as in Section IV-A. That is, (28) is plotted for $\mathbf{B}_1 = \mathbf{B}_2$ and $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$. In addition, the case where all of the channels are perfectly known is

[3]The results for $P_p = 45, 50, 55,$ and 60 dBm are also demonstrated in Fig. 2 to display the high SNR floor more clearly.
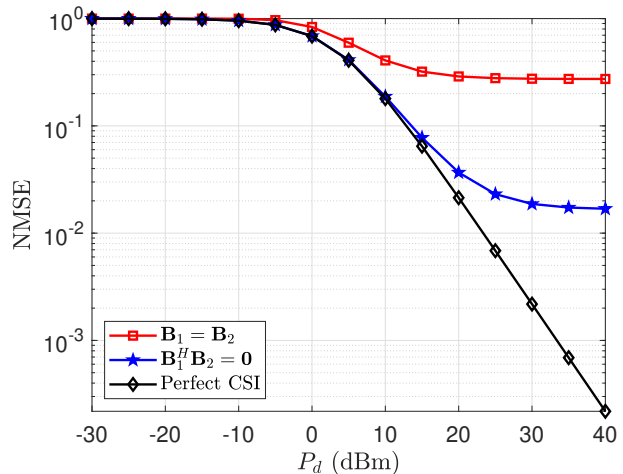

Fig. 3: Data transmission power versus the data estimation NMSE for deterministic channels with high channel estimation SNR.

plotted to serve as the golden standard, labeled as *Perfect CSI*. However, even when all the channels are perfectly known, each RIS is assumed to be optimized independently according to the subscribed UE's CSI. Note that although the channel estimation SNR is high, $\mathbf{B}_1 = \mathbf{B}_2$ yields biased estimates of $\mathbf{g}_1$ due to pilot contamination caused by self-interference. On the other hand, $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ yields the true $\mathbf{g}_1$ as the estimate, however, since BS 1 is unaware of the path through the second RIS, the data estimate is biased, hence, there is still a high data transmission SNR floor. At around $P_d = 5$ dBm, $\mathbf{B}_1 = \mathbf{B}_2$ starts to approach the high-SNR floor. On the other hand, $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ does not suffer from the lack of awareness of the second RIS path until around $P_d = 20$ dBm. Hence, Fig. 3 clearly shows the benefit of configuring the RIS pilot configurations sequences orthogonally.

C. Channel Estimation Based on Correlated Rayleigh Fading Priors

In Fig. 4, the channel estimation performance with the two RIS pilot configurations are analyzed for correlated Rayleigh fading channels. Fig. 4 is generated by computing the trace of the channel estimation error covariance matrix provided in (35) and normalizing it by the factor of $\text{tr}(\mathbf{\Sigma}_{\mathbf{g}_k})$ for different P_p values and different spatial channel correlation matrices. The spatial channel correlation matrices are computed for isotropic scattering based on the different RIS element geometries according to [30, Prop. 1]. Note that configuring the RISs such that $\mathbf{B}_1 = \mathbf{B}_2$ causes severe problems in channel estimation, that is, the MSE increases as pilot transmission power increases for all geometries while $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ completely eliminates pilot contamination. In addition, one can note that as the spatial correlation increases, channel estimation performance also increases since different channel parameters contain more information from one another.

Furthermore, in Fig. 5, we demonstrate the two different components of the channel estimation error in the presence of inter-operator pilot contamination. While the red curve corresponds to the total NMSE and the blue curve corresponds
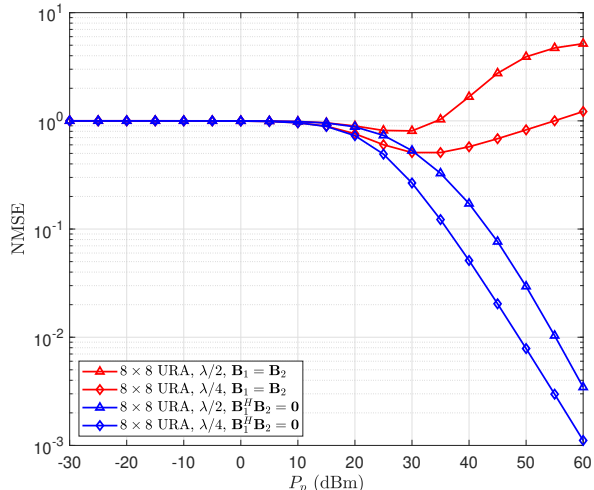
Fig. 4: Pilot transmission power versus channel estimation MSE for different RIS geometries.
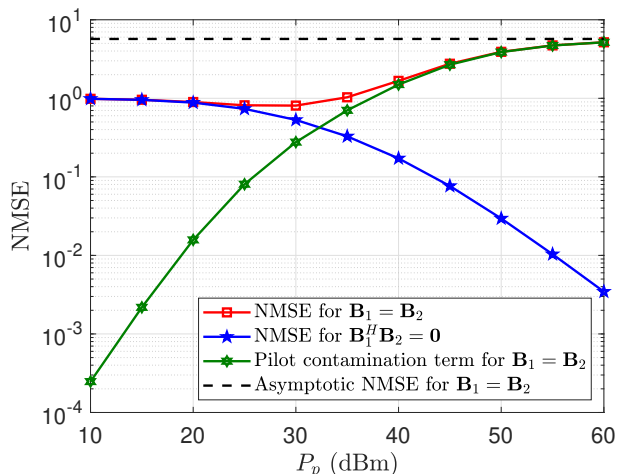

Fig. 5: Different components of the channel estimation NMSE

to the NMSE in the absence of pilot contamination as usual, the green curve demonstrates the term coming from pilot contamination, as in (35). On the other hand, the black dashed line represents the asymptote of the channel estimation NMSE in the presence of pilot contamination, which is provided by (39). Note that as the transmission power increases, the NMSE coming from pilot contamination also increases up to a certain point, and converges to the trace of (39) for $\mathbf{B}_1 = \mathbf{B}_2$ due to the fact that at very high transmission powers, the increase in pilot contamination cancels out with the increasing ability to estimating the channel.

D. Capacity Lower Bound for Reliable Communication Under Imperfect CSI

In Fig. 6, the capacity lower bound derived in Section VI is plotted against the data transmission power. This is performed by generating several channel realizations and computing (41). Note that when $\mathbf{B}_1 = \mathbf{B}_2$, the capacity lower bound stops increasing after $P_d = 30$ dBm while this happens at
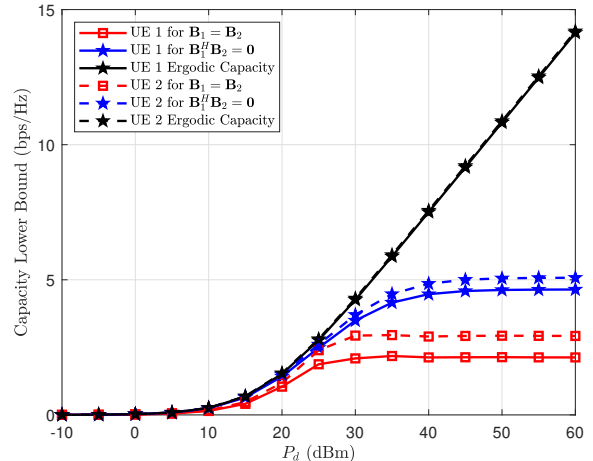

Fig. 6: Data transmission power versus capacity lower bound and the ergodic capacity for the two users. Note that after $P_d = 30$ dBm, the gap between the ergodic capacity of the channel and the capacity lower bound based on the channel estimates grows significantly.

around $P_d = 40$ dBm for $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ when the effect of the misspecified channel during data transmission starts to appear. In any case, it is clear that configuring the RISs such that $\mathbf{B}_1^H \mathbf{B}_2 = \mathbf{0}$ almost doubles the capacity lower bound. In addition, we plotted the ergodic capacity to show the gap between the capacity bound and the achievable gold standard. We assume that each operator configures its RIS to maximize its user's capacity, and they ignore the inter-operator pilot contamination, that is:

$$\phi_1 = \exp(-j(\arg(\mathbf{h}_1) + \arg(\mathbf{g}_1))), \tag{57a}$$

$$\phi_2 = \exp(-j(\arg(\mathbf{h}_2) + \arg(\mathbf{g}_2))). \tag{57b}$$

With this in mind, we consider the overall SISO channels to be perfectly known, which we define as

$$v_1 \triangleq \mathbf{h}_1^T \boldsymbol{\Phi}_1 \mathbf{g}_1 + \mathbf{q}_1 \boldsymbol{\Phi}_2 \mathbf{p}_1, \tag{58a}$$

$$v_2 \triangleq \mathbf{h}_2^T \boldsymbol{\Phi}_2 \mathbf{g}_2 + \mathbf{q}_2 \boldsymbol{\Phi}_1 \mathbf{p}_2. \tag{58b}$$

Recall that while we consider $\mathbf{h}_k$ and $\mathbf{q}_k$ channels to be static, we consider $\mathbf{g}_k$ and $\mathbf{p}_k$ channels to be stochastic, therefore, the ergodic capacity expression contains an expectation with respect to these channels [31]:

$$C_k = \mathbb{E}_{\mathbf{g}_k, \mathbf{g}_j, \mathbf{p}_k} \left[\log_2 \left(1 + P_d \frac{|v_k|^2}{\sigma_w^2} \right) \right]. \tag{59}$$

Note that after $P_d = 30$ dBm, the gap in Fig. 6 between the ergodic capacity of the channel and the capacity lower bound based on the channel estimates grows significantly. At low transmission power, the dominant impeding factor is the noise, while as we transmit at higher powers, pilot contamination, and the signal model misspecification take over. This shows that inter-operator interference significantly degrades the system performance.

VIII. Conclusions

In this paper, we have studied the impact of pilot contamination in a system consisting of two wide-band RISs,

two single-antenna UEs, and two co-located single-antenna BSs. We have demonstrated that the presence of multiple RISs in the same area causes pilot contamination, although the UEs are subscribed to different operators and transmit over disjoint narrow frequency bands. To combat this new type of pilot contamination, we have proposed the use of orthogonal RIS configurations during pilot transmission. For two different sets of assumptions, that is, deterministic and correlated Rayleigh-fading channel models, we have derived the channel and data estimation MSEs and the capacity lower bound in closed-form. In the numerical results, we have clearly shown that the proposed approach eliminates pilot contamination completely, and decreases data estimation MSE significantly for deterministic channels. On the other hand, we have also shown that the capacity lower bound almost doubles when the RISs are configured orthogonally during the pilot transmission step. While one might argue that this doubling comes at the expense of doubling the number of pilots, the estimates can be used for many data transmissions if the channel is static enough, resulting in a higher overall data rate. While this study covers the channel estimation performance in multi-operator RIS-based pilot contamination scenarios for both deterministic and stochastic channels, further analysis is needed for parametric channel models, which opens a new set of possibilities.

REFERENCES

[1] D. Gürgünoğlu, E. Björnson, and G. Fodor, "Impact of pilot contamination between operators with interfering reconfigurable intelligent surfaces," in *2023 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2023, pp. 27–32.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[3] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward Massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, 2020.

[4] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[5] V. Saxena, G. Fodor, and E. Karipidis, "Mitigating pilot contamination by pilot reuse and power control schemes for massive MIMO systems," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015, pp. 1–6.

[6] G. Fodor, N. Rajatheva, W. Zirwas, L. Thiele, M. Kurras, K. Guo, A. Tolli, J. H. Sorensen, and E. De Carvalho, "An overview of massive MIMO technology components in METIS," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 155–161, 2017.

[7] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 14–20, 2021.

[8] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three myths and two critical questions," *IEEE Commun. Mag.*, no. 12, pp. 90–96, 2020.

[9] G. T. de Araújo, A. L. F. de Almeida, R. Boyer, and G. Fodor, "Semiblind joint channel and symbol estimation for IRS-assisted MIMO systems," *IEEE Transactions on Signal Processing*, vol. 71, pp. 1184–1199, 2023.

[10] R. Liu, M. Li, H. Luo, Q. Liu, and A. L. Swindlehurst, "Integrated sensing and communication with reconfigurable intelligent surfaces: Opportunities, applications, and future directions," *IEEE Wireless Communications*, vol. 30, no. 1, pp. 50–57, 2023.

[11] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4144–4157, 2021.

[12] E. Björnson and P. Ramezani, "Maximum likelihood channel estimation for RIS-aided communications with LOS channels," in *Asilomar Conference on Signals, Systems and Computers*, 2022.

[13] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[14] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6g?" *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1707–1725, 2023.

[15] M. Rihan, A. Zappone, S. Buzzi, G. Fodor, and M. Debbah, "Passive vs. active reconfigurable intelligent surfaces for integrated sensing and communication: Challenges and opportunities," *IEEE Network*, pp. 1–1, 2023.

[16] N. Garg, H. Ge, and T. Ratnarajah, "Generalized superimposed training scheme in IRS-assisted cell-free massive mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 1157–1171, 2022.

[17] 3GPP, "NR; Physical channels and modulation," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.211, 09 2022, version 17.4.0.

[18] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1823–1838, 2020.

[19] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Bjrnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627–8640, 2021.

[20] V. Tapio, A. Shojaeifard, I. Hemadeh, A. Mourad, and M. Juntti, "Reconfigurable intelligent surface for 5G NR uplink coverage enhancement," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–5.

[21] Y. Zhao and X. Lv, "Network coexistence analysis of RIS-assisted wireless communications," *IEEE Access*, vol. 10, pp. 63 442–63 454, 2022.

[22] M. Vincenzi, A. Antonopoulos, E. Kartsakli, J. Vardakas, L. Alonso, and C. Verikoukis, "Cooperation incentives for multi-operator C-RAN energy efficient sharing," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[23] H.-T. Chien, Y.-D. Lin, H.-W. Chang, and C.-L. Lai, "Multi-operator fairness in transparent RAN sharing by soft-partition with blocking and dropping mechanism," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 597–11 605, 2018.

[24] H.-S. Guan, Y.-L. Kuang, G.-Z. Huang, and X.-B. Li, "Analysis of farfield radiation and multi antenna coupling characteristics of antenna on shared tower," in *2020 6th Global Electromagnetic Compatibility Conference (GEMCCON)*, 2020, pp. 1–4.

[25] P. Zheng, X. Ma, and T. Y. Al-Naffouri, "On the impact of mutual coupling on RIS-assisted channel estimation," *arXiv:2309.04990*, 2023.

[26] Z. Yu and D. Yuan, "Resource optimization with interference coupling in multi-RIS-assisted multi-cell systems," *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 98–110, 2022.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[28] H. V. Poor, *An Introduction to Signal Detection and Estimation (2nd Ed.)*. Berlin, Heidelberg: Springer-Verlag, 1994.

[29] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

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

[31] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.