Low-Overhead Channel Estimation for RIS-Aided Multi-Cell Networks in the Presence of Phase Quantization Errors

Qingchao Li, *Graduate Student Member, IEEE*, Mohammed El-Hajjar, *Senior Member, IEEE*, Ibrahim Hemadeh, *Senior Member, IEEE*, Arman Shojaeifard, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Deploying reconfigurable intelligent surfaces (RIS) is promising for enhancing the transmission reliability of wireless communications by controlling the wireless environment, in which the active beamforming at the base station and the passive beamforming at the RIS are jointly designed based on the acquisition of channel state information. Hence, channel estimation is crucial for RIS-aided systems. Due to the lack of active radio frequency chains at the RIS to process and transmit pilot sequences, only the cascaded twin-hop transmitter-RIS-receiver channel can be estimated, which results in extremely high pilot overhead, when a large number of RIS reflecting elements is used. As a remedy, we propose a channel estimation method relying on low pilot overhead, namely the Karhunen-Loève transformation based linear minimal mean square error (KL-LMMSE) estimator. This exploits the spatial correlation of the RIS-cascaded channels, for our multi-cell multiple-input and multiple-output RIS-aided systems. Furthermore, we extend our investigations to the effects of realistic phase quantization errors. Additionally, we derive the theoretical mean square error (MSE) of our proposed channel estimators verified by numerical simulations, and compare the results to various benchmark schemes. We show that the MSE performance of our proposed KL-LMMSE estimator is better than that of the state-of-the-art low-overhead channel estimators.**

***Index Terms*—Reconfigurable intelligent surfaces (RIS), channel estimation, linear minimal mean square error (LMMSE), Karhunen-Loève (KL) transformation, phase quantization error.**

I. Introduction

In the fifth-generation (5G) systems, various advanced technologies, such as millimeter wave (mmWave) [1] and massive multiple-input and multiple-output (MIMO) schemes [2], are employed for increasing the performance of ultra-dense networks. Recently, reconfigurable intelligent surfaces (RIS) have been widely hailed as a promising technique of improving the spectral efficiency, energy efficiency and coverage range of the next-generation systems [3]. Briefly, the RIS is an artificial metamaterial surface comprised of a large number of passive reflecting elements employed between the base station (BS) and the user equipment (UE) [4]–[7]. By adjusting the RIS coefficients, i.e. the phase and amplitude of the reflecting elements, the wireless channel environment can be beneficially ameliorated. The RIS has a wide range of promising applications, such as increasing the channel gains via passive beamforming at the RIS and the active beamforming at the BS [8], [9], constructing physical layer security networks [10], reducing the inter-cell interference for cell-edge users [11], providing simultaneous wireless information and power transfer [12], and aiding mobile edge computation [13] and over-the-air computation [14], etc. However, to support these applications and realize performance enhancements for wireless systems, high-precision channel state information (CSI) acquisition is a crucial issue. Since the RIS operating passive, without active signal processing units, the BS-RIS channel links and RIS-UE channel links cannot be directly estimated. Therefore, channel estimation is more challenging in RIS-aided systems than in conventional wireless networks.

Firstly, in the case of a single UE, the least square (LS) method is widely employed [15]–[18] for the channel estimation of RIS-aided wireless communications. In [15], Mishra *et al.* proposed simple ON/OFF RIS training patterns for estimating the cascaded BS-RIS-UE links. However, in the context of ON/OFF patterns, only a single reflecting element is activated in each symbol slot for the estimation of the cascaded link. Hence, a large number of elements are not fully utilized, which degrades the channel estimation accuracy. To deal with this issue, Jensen *et al.* [16] conceived a discrete Fourier transformation (DFT) matrix based method for designing RIS training patterns, demonstrating that the DFT-based RIS training patterns succeed in achieving the minimal mean square error (MSE) of channel estimation. However, in practice the RIS elements has finite phase resolution, hence only a set of discrete phase shifts can be used by the RIS elements. In this practical context, Zhou *et al.* [17] employed the Hadamard matrix for the design of RIS patterns, where only two phase shifts, namely 0 and π, are used for each RIS reflecting element. They demonstrated that only a slight MSE performance erosions imposed compared to the DFT-based RIS training patterns [16]. In [18], An *et al.* proposed an optimal pilot power allocation strategy for the LS estimator to

(Corresponding author: Lajos Hanzo.)

Qingchao Li, Mohammed El-Hajjar and Lajos Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: Qingchao.Li@soton.ac.uk; meh@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Ibrahim Hemadeh and Arman Shojaeifard are with InterDigital, London EC2A 3QR, U.K. (e-mail: Ibrahim.Hemadeh@InterDigital.com; Arman.Shojaeifard@InterDigital.com).

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1, EP/X01228X/1 and EP/Y026721/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

improve the throughput of RIS-aided communication systems. In this LS method, the receiver estimates the channels purely based on its own observations without any prior knowledge of the channels. When the statistical channel information is available, the linear minimum mean square error (LMMSE) method can be employed for enhancing the channel estimation performance [19]. In [20], Zhang *et al.* proposed an optimized channel estimator in a closed-form for RIS-assisted systems in mmWave channels based on the MSE minimization criterion.

However, the above methods have extremely high pilot overhead, when a large number of reflecting elements are employed, since at least $N+1$ symbol slots are required in each coherence intervals at the stage of channel estimation for an N-element RIS system. To deal with this, an element grouping based idea was adopted in [21]–[24] for the LS estimator to cut down pilot overhead. Specifically, the N reflecting elements of the RIS are partitioned into $N_{\mathcal{G}}$ compact groups, and the cascaded channel links corresponding to the elements in the same group are assumed to have identical CSI by exploiting the high spatial correlation among adjacent RIS-cascaded channels. In this case, only a minimum of $N_{\mathcal{G}}+1$ symbol slots are required in each coherence interval for channel estimation, which can significantly reduce the pilot overhead upon reducing the number of groups $N_{\mathcal{G}}$, albeit at the cost of degrading the estimation accuracy. Furthermore, in [25], Huang *et al.* proposed a novel semi-blind channel estimator which exploits the data symbols for channel estimation enhancement and only a fraction of full-pilot signaling overhead is required.

On the other hand, upon considering multiple UEs, Alwazani *et al.* [26] proposed to extend the single-UE channel method. Specifically, when there are K UEs in the RIS-aided wireless system, each RIS training pattern lasts K symbol slots, in which the K UEs send orthogonal pilot sequences to eliminate the inter-user interference. At the BS, the CSI of each UE can be estimated similarly to the single-UE case by using a Bayesian minimum mean squared error (MMSE) estimator. In this case, at least $K(N+1)$ symbol slots are required in each coherence interval at the stage of channel estimation for an N-element RIS system. In [27], [28], the authors cut down the pilot overhead by exploiting the property that all users share the common BS-RIS link. Specifically, first the cascaded BS-RIS-UE channel of a selected reference UE is estimated via the same methods in the single-UE case. Then the RIS-UE links of other $K-1$ UEs are estimated based on the estimation of the selected reference UE, which requires the minimal pilot overhead of $K+N+\max(K-1,(K-1)\lceil\frac{N}{M}\rceil)$ instead of $K(N+1)$, where M is the number of BS antennas. However, the minimum number of symbol slots required for channel estimation in each coherence interval in [27], [28] is still larger than the number of RIS elements, resulting in a high pilot overhead when a large number of RIS elements are used. Besides, in [27], [28], the authors focused on single-cell wireless communication. While the RIS is promising in terms of reducing the inter-cell interference in multi-cell systems [29], it relies on the accurate of CSI acquisition, including both the channels from the UEs in the local cell to the BS and the channels from the UEs in the adjacent cells to the BS. Furthermore, the above channel estimation methods are based on the assumption that the phase shift of each RIS element can be adjusted perfectly. However, due to the limited phase shift resolution, the RIS having phase quantization errors is inevitable in the reflecting elements. Thus, it is meaningful to take the RIS phase quantization error into account in designing the channel estimation algorithm.

In this paper, we propose a low-overhead instantaneous CSI acquisition method for RIS-aided multi-cell MIMO communications, namely the Karhunen-Loève transformation based LMMSE (KL-LMMSE) channel estimator, by exploiting the spatial correlation of the cascaded channels. Furthermore, our channel estimators takes into account the phase quantization error of RIS. Additionally, we verify the accuracy of our theoretical analysis by simulation results. Furthermore, we compare them to various benchmark schemes. Against this background, the novelty of the proposed low pilot overhead channel estimator is compared to the existing solutions in the literature [15]–[28] of the RIS-aided systems in Table I. The contributions of our paper are further derived as follows:

- By considering the high pilot overhead of the state-of-the-art (SoA) LMMSE estimator, we propose the novel KL-LMMSE estimator, where we compress the cascaded-RIS channels via the Karhunen-Loève (KL) transform applied to the cascaded-RIS channel covariance matrix for reducing the pilot overhead. Since the KL transform is the optimal transform for linear approximations, it leads to lower MSE for the estimated channels compared to the estimators based on grouping ideas proposed in [21]–[24].
- We further extend the SoA LMMSE estimator and the proposed KL-LMMSE estimator by considering practical limited phase shift resolution of the RIS, including the RIS phase quantization error having both von Mises distribution and uniform distribution, and we also derive the theoretical MSE.
- Finally, the theoretical analysis and simulation results demonstrate that the LMMSE estimators have lower MSE than the LS estimators due to the utilization of the first moment and second moment of signal links. Furthermore, it is shown that the MSE performance of our proposed KL-LMMSE estimator tends to that of the SoA LMMSE estimator upon increasing the RIS spatial correlation, despite having a reduced pilot overhead.

The rest of this paper is organized as follows. In Section II, we present the system model. The low-overhead LMMSE channel estimator is derived in Section III. Section IV presents the theoretical analysis of the proposed channel estimator, while our simulation results are presented in Section V. Finally we conclude in Section VI.

Notations: Vectors and matrices are denoted by boldface lower and upper case letters, respectively, $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, and $(\cdot)^{\mathrm{H}}$ represent the transpose, conjugate and Hermitian transpose operation, respectively, $\lceil a \rceil$ is the minimal integer larger than a, $\mathbb{C}^{m\times n}$ denotes the space of $m\times n$ complex-valued matrix, $\odot$ and $\otimes$ denotes the Hadamard product and Kronecker product, respectively, a_i is the ith element in vector $\mathbf{a}$, $\mathbf{0}_n$ and $\mathbf{1}_n$ represent the $n\times 1$ zero vector and $n\times 1$ one vector, respectively,

TABLE I: The novelty comparison of the proposed low pilot overhead channel estimation for the RIS-aided wireless communications to the existing solutions in the literatures [15]–[28].

	Our paper	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]
Multi-user	✔												✔	✔	✔
Spatial correlated RIS	✔							✔	✔	✔	✔		✔	✔	✔
Estimation method	KL-LMMSE	LS	LS	LS	LS	LMMSE	MMSE	LS	LS	LS	LS	Semi-blind	MMSE	LMMSE	LS
Reduced pilot overhead	✔							✔	✔	✔	✔	✔		✔	✔
RIS phase quantization error	✔														
Multi-cell	✔														

$\mathbf{O}_{n_1\times n_2}$ represents the $n_1 \times n_2$ zero matrix, $\mathbf{I}_n$ represents the $n \times n$ identity matrix, $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal elements being the elements in $\mathbf{a}$ in order. A circularly symmetric complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbb{E}[\mathbf{x}]$ represents the mean of the random vector $\mathbf{x}$, $\mathbf{C}_{\mathbf{xy}}$ is the covariance matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$.

II. System Model

The system model of the RIS-aided multi-cell wireless MIMO systems [29], [30] is shown in Fig. 1, where L cells are supported by a RIS to increase the transmission reliability and diminish the inter-cell interference. We denote these cells as cell-1, cell-2, $\cdots$, cell-L and the corresponding BSs as BS-1, BS-2, $\cdots$, BS-L, each of which is equipped with $M_1, M_2, \cdots, M_L$ antennas, respectively, to communicate with several multi-antenna users. We denote the number of users in cell-l as U_l, where each of them is equipped with V_l antennas. Thus, the total number of user antennas (UANs) in cell-l is given by $K_l = U_l V_l$ with each of them denoted as UAN-k_l ($k_l = 1, 2, \cdots, K_l$). Thus, cell-l is essentially an $M_l \times K_l$ MIMO system supported by a RIS. The parameters of the RIS elements are configured by the RIS controller, which can deliver the control data with the aid of the BSs via a Central Processing Unit (CPU). We consider the uplink in a time division duplex (TDD) channel estimation scenario by exploiting the channel's reciprocity [31]. Specifically, the phase of each RIS element is designed based on the pre-defined RIS patterns, and the UANs transmit user pilot sequences (orthogonal or non-orthogonal) to the BSs. The BSs estimate the CSI, including the RIS-related links and the direct UAN-BS links, based on its observation.

Although employing RIS is beneficial for enhancing the channel capacity, it brings about many challenges for the channel estimators. Specifically, as shown in Fig. 1, the observation at BS-l includes the pilot sequences transmitted not only from the UANs in cell-l but also from the UANs in the other $L-1$ cells, although the CSI in cell-l can be accurately estimated at BS-l, provided that all the UANs in all cells employ orthogonal pilot sequences. However, this results in an extremely high pilot overhead, when a large number of RIS elements is used due to the limitation of each coherence interval length. In a practical system, usually non-orthogonal pilot assignment is adopted, where the pilot sequences transmitted within the same cell are orthogonal, but they are reused in all the cells [31]. In this case, the intra-cell interference can be eliminated, but some inter-cell interference will still exist. Furthermore, compared to the family of conventional wireless communication systems operating without RIS, the RIS may in fact precipitate extra inter-cell interference, which degrades the channel estimation performance. For example, in Fig. 1, the inter-cell interference at BS-l results from both the direct UAN-BS links plus also from the RIS-related links arriving from the user antennas in the other $L-1$ cells to BS-l. Furthermore, the inter-cell interference is exacerbated with the increase of the number of RIS elements. As shown in Fig. 1, the inter-cell interference at UAN-RIS links can be strengthened upon increasing the number of deployed RIS elements. Harnessing the RIS for enhancing the channel capacity also requires accurate channel estimation, including both the desired links and inter-cell interference links. Therefore, we propose a channel estimation method for RIS-aided multi-cell communications, which can effectively deal with the issue of inter-cell interference, despite requiring a low pilot overhead.

A. RIS Architecture

The RIS is a uniform rectangular planar array (URPA) containing $N = N_x \times N_y$ reflecting elements, where N_x and N_y represent the numbers of reflecting elements in the horizontal and vertical directions. The distance between the adjacent RIS elements is $\delta^{(x)}$ and $\delta^{(y)}$ in the horizontal and vertical directions, respectively. The phase shift of each RIS element is configured by the RIS controller, as shown in Fig. 1. We denote the RIS phase shift vector as $\boldsymbol{\theta} = [\mathrm{e}^{j\theta_1}, \mathrm{e}^{j\theta_2}, \cdots, \mathrm{e}^{j\theta_N}]$, where θ_n is the phase shift of the n-th RIS element and $\theta_n \in [-\pi, \pi)$. Due to the limited RIS phase shift resolution, the phase shift of each RIS element cannot be precisely designed. Therefore, in a practical RIS, the phase shift of each reflecting element can be modelled as $\theta'_n = \theta_n + \tilde{\theta}_n$, where θ_n represents the expected phase shift of the n-th reflecting element and $\tilde{\theta}_n$ represents the phase quantization error of the n-th reflecting element [32].

Referring to [32]–[34], the phase shift of each reflecting element is subject to RIS phase error due to the finite-resolution quantization, which results from hardware impairments (HWI), since the high-resolution phase shift of each reflecting element is tuned by the bias voltage of the varactor diode on the RIS metasurface [35]. Due to the uncertainty of the hardware circuit response, the RIS phase error is time-varying. Therefore,

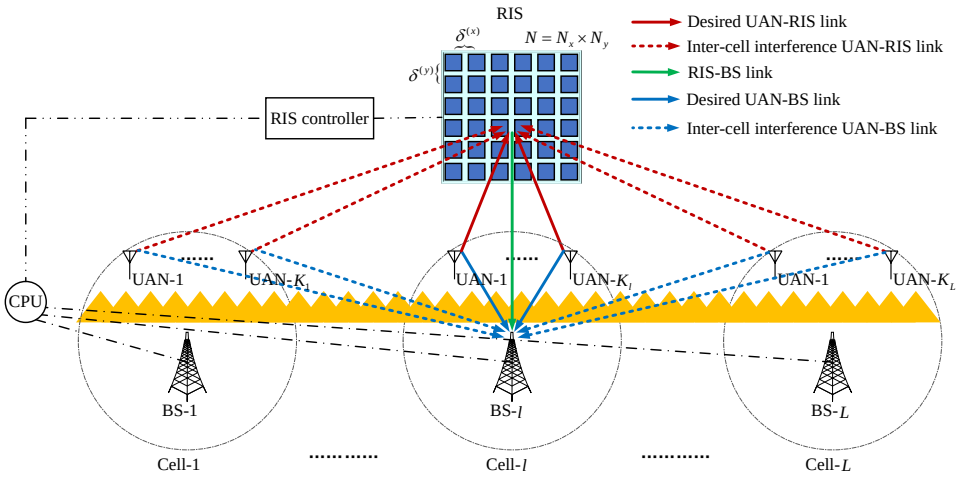


Fig. 1: System model of the RIS-aided multi-cell MIMO wireless communications, where only the links to BS-l is presented and the links to other BSs can be obtained similarly.

the phase quantization errors can be modelled as independent and identically distributed (i.i.d.) random variables having the mean of 0 [32]–[34]. The most widely used probability density functions to describe the phase quantization error are the von Mises distribution and the uniform distribution, denoted as $\tilde{\theta}_n \sim \mathcal{VM}(0, \kappa_{\rm p})$ and $\tilde{\theta}_n \sim \mathcal{U}(-\iota_{\rm p}, \iota_{\rm p})$, respectively, where $\kappa_{\rm p}$ is the concentration parameter in the von Mises distribution and $(-\iota_{\rm p}, \iota_{\rm p})$ is the support in the uniform distribution. If we denote the desired RIS phase shift vector by $\boldsymbol{\theta}$, and the phase quantization error vector by $\tilde{\boldsymbol{\theta}} = [{\rm e}^{j\tilde{\theta}_1}, {\rm e}^{j\tilde{\theta}_2}, \cdots, {\rm e}^{j\tilde{\theta}_N}]$, then the practical phase shift vector $\boldsymbol{\theta}' = \boldsymbol{\theta} \odot \tilde{\boldsymbol{\theta}}$.

B. Channel Model

For simplicity, we combine the UAN index sets $\{k_1|k_1 = 1, 2, \cdots, K_1\}$, $\{k_2|k_2 = 1, 2, \cdots, K_2\}$, $\cdots$, $\{k_L|k_L = 1, 2, \cdots, K_L\}$ into the set $\{k|k = 1, 2, \cdots, K\}$, where $K = \sum_{l=1}^{L} K_l$. We denote the large scale fading between UAN-k and BS-l as $\varrho_{\mathbf{b}^{(k,l)}}$, from the UAN-$k$ to the RIS as $\varrho_{\mathbf{g}^{(k)}}$, and from the RIS to the BS-l as $\varrho_{\mathbf{A}^{(l)}}$, where $\varrho_{\mathbf{b}^{(k,l)}} = {\rm C}_0 d_{\mathbf{b}^{(k,l)}}^{-\alpha_{\mathbf{b}}}$, $\varrho_{\mathbf{g}^{(k)}} = {\rm C}_0 d_{\mathbf{g}^{(k)}}^{-\alpha_{\mathbf{g}}}$, and $\varrho_{\mathbf{A}^{(l)}} = {\rm C}_0 d_{\mathbf{A}^{(l)}}^{-\alpha_{\mathbf{A}}}$ [36]. Here ${\rm C}_0$ is the path loss at the reference distance of 1 meter. Furthermore, $d_{\mathbf{b}^{(k,l)}}$ represents the distance between UAN-k and the BS-l, $d_{\mathbf{g}^{(k)}}$ is the distance between the UAN-k and the RIS, and $d_{\mathbf{A}^{(l)}}$ is the distance between the RIS and the BS-l. Finally, $\alpha_{\mathbf{b}}$, $\alpha_{\mathbf{g}}$ $\alpha_{\mathbf{A}}$ represent the path loss exponent of the UAN-BS path, UAN-RIS path, and RIS-BS path, respectively.

As for the small-scale fading, the direct links impinging from UAN-k to the BS-l is denoted as $\mathbf{b}^{(k,l)} \in \mathbb{C}^{M_l \times 1}$. The links impinging from UAN-k to the RIS is denoted as $\mathbf{g}^{(k)} \in \mathbb{C}^{N \times 1}$. The link spanning from the RIS to the BS-l is given by $\mathbf{A}^{(l)} = [\mathbf{a}_1^{(l)\rm T}, \mathbf{a}_2^{(l)\rm T}, \cdots, \mathbf{a}_{M_l}^{(l)\rm T}]^{\rm T}$, where $\mathbf{a}_{m_l}^{(l)} = [a_{m_l,1}^{(l)}, a_{m_l,2}^{(l)}, \cdots, a_{m_l,N}^{(l)}]$ represents the channel vector from the RIS to the m_lth BS-l antenna. Therefore, the equivalent channel spanning from UAN-k to BS-l can be represented as

$$\mathbf{h}^{(k,l)} = \sqrt{\varrho_{\mathbf{g}^{(k)}} \varrho_{\mathbf{A}^{(l)}}} \mathbf{A}^{(l)} \boldsymbol{\Theta} \mathbf{g}^{(k)} + \sqrt{\varrho_{\mathbf{b}^{(k,l)}}} \mathbf{b}^{(k,l)}, \quad (1)$$

where $\boldsymbol{\Theta} = {\rm diag}\{\boldsymbol{\theta}\}$.

We assume that the direct links from the UANs to BS-l are blocked, so the RIS creates reflected paths between the UANs and BS-l [37]. Therefore, it is reasonable to assume that the UAN-BS links follow Rayleigh fading, while the UAN-RIS links and the RIS-BS link obey Rician fading. The Rayleigh channel spanning from UAN-k to BS-l is given by

$$\mathbf{b}^{(k,l)} \sim \mathcal{CN}(\mathbf{0}_{M_l}, \mathbf{I}_{M_l}), \quad (2)$$

and the Rician channel impinging from UAN-k to the RIS is

$$\mathbf{g}^{(k)} = \sqrt{\frac{\kappa_{\mathbf{g}}}{1+\kappa_{\mathbf{g}}}} \overline{\mathbf{g}}^{(k)} + \sqrt{\frac{1}{1+\kappa_{\mathbf{g}}}} \widetilde{\mathbf{g}}^{(k)}, \quad (3)$$

where $\kappa_{\mathbf{g}}$ is the Rician factor of the UAN-RIS links, while $\overline{\mathbf{g}}^{(k)} \in \mathbb{C}^{N \times 1}$ and $\widetilde{\mathbf{g}}^{(k)} \in \mathbb{C}^{N \times 1}$ represent the line-of-sight (LoS) and non-line-of-sight (NLoS) components, respectively. The LoS component $\overline{\mathbf{g}}^{(k)}$ is expressed as [38]

$$\overline{\mathbf{g}}^{(k)} = \big[1, {\rm e}^{-j\frac{2\pi}{\lambda}(\delta^{(x)} \sin\phi_k^{(\rm A)} \cos\varphi_k^{(\rm A)} + \delta^{(y)} \cos\phi_k^{(\rm A)})}, \cdots, {\rm e}^{-j\frac{2\pi}{\lambda}(\delta^{(x)}(N_x-1) \sin\phi_k^{(\rm A)} \cos\varphi_k^{(\rm A)} + \delta^{(y)}(N_y-1) \cos\phi_k^{(\rm A)})}\big]^{\rm T}, \quad (4)$$

where λ is the carrier wavelength, $\phi_k^{(\rm A)}$ and $\varphi_k^{(\rm A)}$ are the elevation and azimuth angle of arrival (AoA) of signals from UAN-k to the RIS, respectively. The NLoS component $\widetilde{\mathbf{g}}^{(k)}$ is represented as $\widetilde{\mathbf{g}}^{(k)} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{R})$, with $\mathbf{R} \in \mathbb{C}^{N \times N}$ being the covariance matrix of the channel vector $\widetilde{\mathbf{g}}^{(k)}$. In most literature, it was assumed that the NLoS links corresponding to all RIS elements are uncorrelated, so $\mathbf{R} = \mathbf{I}_N$. However, since a large number of passive RIS elements are densely employed in a practical system, it is reasonable to model the NLoS links corresponding to all RIS elements as being spatially correlated [39]. Here, we employed the widely used exponential correlation channel model to describe the spatial correlation between the RIS-related links [39]–[42], since it can characterize the impact of the reflecting element distance on the channel correlation. It is worth noting that our proposed channel estimation method is also suitable for other spatial correlation models, such as the Bessel correlation model [43]. According to the exponential correlation channel model, $\mathbf{R}$ is

determined by the distance between the RIS elements, where $[\mathbf{R}]_{n_1,n_2} = \mathrm{e}^{-\frac{\delta_{n_1,n_2}}{\delta_0}}$, in which δ_{n_1,n_2} is the distance between the n_1th and n_2th RIS element, and δ_0 is a constant that controls the level of correlation. For a user equipment, δ_0 would be around half a wavelength, whereas for a stationary equipment it could be as high as tens of wavelengths [44]. Therefore, $\mathbf{g}^{(k)}$ is distributed as

$$\mathbf{g}^{(k)} \sim \mathcal{CN}\Big(\sqrt{\frac{\kappa_{\mathbf{g}}}{1+\kappa_{\mathbf{g}}}}\overline{\mathbf{g}}^{(k)}, \frac{1}{1+\kappa_{\mathbf{g}}}\mathbf{R}\Big). \tag{5}$$

The Rician channel from the RIS to BS-l is given by

$$\mathbf{A}^{(l)} = \sqrt{\frac{\kappa_{\mathbf{A}}}{1+\kappa_{\mathbf{A}}}}\overline{\mathbf{A}}^{(l)} + \sqrt{\frac{1}{1+\kappa_{\mathbf{A}}}}\widetilde{\mathbf{A}}^{(l)}, \tag{6}$$

where $\kappa_{\mathbf{A}}$ is the Rician factor of the RIS-BS links, $\overline{\mathbf{A}}^{(l)} \in \mathbb{C}^{M_l \times N}$ and $\widetilde{\mathbf{A}}^{(l)} \in \mathbb{C}^{M_l \times N}$ represent the LoS and N-LoS components, respectively. The LoS component $\overline{\mathbf{A}}^{(l)} = \mathbf{f}_{\mathrm{AoA}}^{(\mathbf{A}^{(l)})}\mathbf{f}_{\mathrm{AoD}}^{(\mathbf{A}^{(l)})}$, where $\mathbf{f}_{\mathrm{AoD}}^{(\mathbf{A}^{(l)})}$ is the response of an $N = N_x \times N_y$ URPA at the RIS, given by [38]

$$\mathbf{f}_{\mathrm{AoD}}^{(\mathbf{A}^{(l)})} = \big[1, \mathrm{e}^{-j\frac{2\pi}{\lambda}(\delta^{(x)}\sin\phi_l^{(\mathrm{D})}\cos\varphi_l^{(\mathrm{D})} + \delta^{(y)}\cos\phi_l^{(\mathrm{D})})}, \cdots, \mathrm{e}^{-j\frac{2\pi}{\lambda}(\delta^{(x)}(N_x-1)\sin\phi_l^{(\mathrm{D})}\cos\varphi_l^{(\mathrm{D})} + \delta^{(y)}(N_y-1)\cos\phi_l^{(\mathrm{D})})}\big], \tag{7}$$

where $\phi_l^{(\mathrm{D})}$ and $\varphi_l^{(\mathrm{D})}$ are the elevation and azimuth angle of departure (AoD) of signals from the RIS to BS-l, respectively. Furthermore, $\mathbf{f}_{\mathrm{AoA}}^{(\mathbf{A}^{(l)})}$ is the response of the M_l-antenna uniform linear array (ULA) at BS-l, based on [38]

$$\mathbf{f}_{\mathrm{AoA}}^{(\mathbf{A}^{(l)})} = \big[1, \mathrm{e}^{-j\frac{2\pi}{\lambda}d_0\sin\psi_l^{(\mathrm{A})}}, \cdots, \mathrm{e}^{-j\frac{2\pi}{\lambda}d_0(M_l-1)\sin\psi_l^{(\mathrm{A})}}\big]^{\mathrm{T}}, \tag{8}$$

where d_0 is the distance between adjacent BS antennas, and $\psi^{(\mathrm{A})}$ is the AoA of signals from the RIS to BS-l. We introduce $\overline{\mathbf{A}}^{(l)} = [\overline{\mathbf{a}}_1^{\mathrm{T}}, \overline{\mathbf{a}}_2^{\mathrm{T}}, \cdots, \overline{\mathbf{a}}_{M_l}^{\mathrm{T}}]^{\mathrm{T}}$, where $\overline{\mathbf{a}}_{m_l}$ represents the LoS channel from the RIS to the m_lth BS-l antenna. The NLoS component is $\widetilde{\mathbf{A}}^{(l)} = [\widetilde{\mathbf{a}}_1^{\mathrm{T}}, \widetilde{\mathbf{a}}_2^{\mathrm{T}}, \cdots, \widetilde{\mathbf{a}}_{M_l}^{\mathrm{T}}]^{\mathrm{T}}$, where $\widetilde{\mathbf{a}}_{m_l}$ represents the NLoS channel from the RIS to the m_lth BS-l antenna, given by $\widetilde{\mathbf{a}}_{m_l}^{\mathrm{T}} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{R})$, with $\mathbf{R} \in \mathbb{C}^{N\times N}$ being the covariance matrix of the channel vector $\widetilde{\mathbf{a}}_{m_l}$. Since the UAN-RIS links and the RIS-BS links share the same reflecting surfaces, the covariance matrix of the channel vector $\widetilde{\mathbf{a}}_{m_l}$ is the same as that of UAN-RIS links. Therefore, $\mathbf{a}_{m_l}$ is distributed as

$$\mathbf{a}_{m_l}^{\mathrm{T}} \sim \mathcal{CN}\Big(\sqrt{\frac{\kappa_{\mathbf{A}}}{1+\kappa_{\mathbf{A}}}}\overline{\mathbf{a}}_{m_l}^{\mathrm{T}}, \frac{1}{1+\kappa_{\mathbf{A}}}\mathbf{R}\Big). \tag{9}$$

III. Low-Overhead LMMSE Channel Estimation

The channel estimation regime of RIS-aided wireless systems is shown in Fig. 2. The statistical CSI, which mainly depends on the AoA and AoD of each link and on the location of mobile equipments, and the RIS covariance matrix $\mathbf{R}$, remain unchanged in a stationary interval. The statistical CSI is estimated at the beginning of each stationary block, and each stationary block can be divided into Q coherence intervals, within each the instantaneous CSI are considered constant [45]. Each coherence interval is comprised of 3 stages, including the instantaneous CSI estimation, joint RIS and BS beamforming based on the estimated channels, and data transmission. In this paper, we assume that the statistical CSI, which can be estimated based on [46], is known and our objective is that of estimating the instantaneous CSI in each coherence interval.

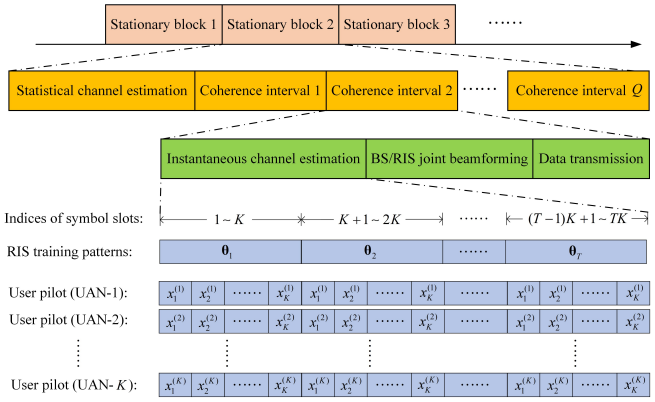


Fig. 2: Diagram of channel estimation for RIS-aided wireless communication systems.

A total of TK symbol slots are utilized for instantaneous CSI estimation at the beginning of each coherence interval, and T pre-defined RIS training patterns $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_T$ are activated in order. Specifically, from the 1st to the Kth symbol slot, the RIS employs $\boldsymbol{\theta}_1$. Then from the $(K+1)$st to the $2K$th symbol slot, the RIS employs $\boldsymbol{\theta}_2$ and so on. In the duration of each RIS training pattern, orthogonal user pilots are transmitted by the K UANs in the K symbol slots to prevent pilot contamination. Specifically, the $K \times K$ DFT matrix can be employed as orthogonal user pilots for K UANs. We denote the $K \times K$ DFT matrix as $\boldsymbol{\Omega}$, given by $\boldsymbol{\Omega} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(K)}]$, where $\mathbf{x}^{(k)} \in \mathbb{C}^{K\times 1}$ represents the kth column in $\boldsymbol{\Omega}$ with $x_i^{(k)}$ representing the pilot transmitted for UAN-k during the ith symbol slot in each RIS training pattern.

Without loss of generality, we focus on the estimation of the links to cell-l based on the observation at BS-l. The channel estimation of links to the other $L-1$ cells can be carried out similarly. Thus, the subscript or superscript l is omitted in the following analysis.

A. Channel Estimations for a Perfect RIS Architecture

Firstly, we focus our attention on RIS-aided communications relying on a perfect RIS architecture, i.e. without RIS phase quantization error. The equivalent channel $\mathbf{h}^{(k)}$ is

$$\begin{aligned}\mathbf{h}^{(k)} =& \sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{A}\boldsymbol{\Theta}\mathbf{g}^{(k)} + \sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{b}^{(k)} \\ =& [\sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M, \sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M \otimes \boldsymbol{\theta}]\mathbf{s}^{(k)},\end{aligned} \tag{10}$$

where $\mathbf{s}^{(k)} = [\mathbf{b}^{(k)\mathrm{T}}, \mathbf{s}_1^{(k)\mathrm{T}}, \cdots, \mathbf{s}_M^{(k)\mathrm{T}}]^{\mathrm{T}}$, in which $\mathbf{s}_m^{(k)} \in \mathbb{C}^{N\times 1}$ is the cascaded channel given by $\mathbf{s}_m^{(k)} = \mathbf{a}_m^{\mathrm{T}} \odot \mathbf{g}^{(k)} = [a_{m,1}g_1^{(k)}, a_{m,2}g_2^{(k)}, \cdots, a_{m,N}g_N^{(k)}]^{\mathrm{T}}$. The UAN-RIS channel $\mathbf{g}^{(k)}$ and the RIS-BS channel $\mathbf{A}$ cannot be estimated separately, since the RIS elements are passive. Fortunately, estimating the cascaded channel $\mathbf{s}^{(k)}$ is sufficient for designing the RIS phase shift matrix for data transmission without loss of

optimality. Therefore, our objective is to estimate $\mathbf{s}^{(k)}$. Firstly, we present the SoA LMMSE method, followed by our new low-overhead method, namely the KL transformation based LMMSE estimator.

1) SoA LMMSE method: In the state-of-the-art LMMSE method, the observation at the BS in the ith symbol slot of the tth ($t = 1, 2, \cdots, T$) RIS training pattern from the kth UAN, denoted as $\mathbf{y}_{i,t}^{(k)} \in \mathbb{C}^{M\times 1}$, is given by

$$\mathbf{y}_{i,t}^{(k)} = \sqrt{P}[\sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M, \sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M \otimes \boldsymbol{\theta}_t]\mathbf{s}^{(k)}x_i^{(k)} + \mathbf{n}_{i,t}^{(k)}, \tag{11}$$

where P is the transmit power at UANs, $\mathbf{n}_{i,t}^{(k)} \in \mathbb{C}^{M\times 1}$ is the noise at the BS antennas, and $\mathbf{n}_{i,t}^{(k)} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_{\mathrm{n}}^2\mathbf{I}_M)$ with σ_{n}^2 being the noise power at the BS antennas, while $\boldsymbol{\theta}_t = [\mathrm{e}^{j\theta_{t,1}}, \mathrm{e}^{j\theta_{t,2}}, \cdots, \mathrm{e}^{j\theta_{t,N}}]$. By stacking the observation $\mathbf{y}_{i,1}^{(k)}, \mathbf{y}_{i,2}^{(k)}, \cdots, \mathbf{y}_{i,T}^{(k)}$, we can get

$$\mathbf{y}_i^{(k)} = [\mathbf{y}_{i,1}^{(k)\mathrm{T}}, \mathbf{y}_{i,2}^{(k)\mathrm{T}}, \cdots, \mathbf{y}_{i,T}^{(k)\mathrm{T}}]^{\mathrm{T}} = \sqrt{P}\mathbf{Z}^{(k)}\mathbf{s}^{(k)}x_i^{(k)} + \mathbf{n}_i^{(k)}, \tag{12}$$

where $\mathbf{n}_i^{(k)} = [\mathbf{n}_{i,1}^{(k)\mathrm{T}}, \mathbf{n}_{i,2}^{(k)\mathrm{T}}, \cdots, \mathbf{n}_{i,T}^{(k)\mathrm{T}}]^{\mathrm{T}}$ and

$$\mathbf{Z}^{(k)} = \begin{bmatrix} \sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M, \sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M \otimes \boldsymbol{\theta}_1 \\ \sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M, \sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M \otimes \boldsymbol{\theta}_2 \\ \vdots \\ \sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M, \sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M \otimes \boldsymbol{\theta}_T \end{bmatrix}. \tag{13}$$

Since $\mathbf{s}^{(k)}$ is an $M(N+1)\times 1$ vector and $\mathbf{Z}^{(k)}$ is an $MT \times M(N+1)$ matrix, to ensure that $\mathbf{s}^{(k)}$ can be uniquely estimated, the number of RIS training patterns must satisfy $T \geq N+1$. As for the design of RIS training patterns, we have the following two scenarios.

- *DFT matrix:* Referring to [16], a $2^{\lceil \log_2 T\rceil}\times 2^{\lceil \log_2 T\rceil}$ DFT matrix can be employed for the design of the RIS training patterns $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_T$. Specifically, $[1, \boldsymbol{\theta}_t]$ is used for the first $N+1$ elements in the tth row of the $2^{\lceil \log_2 T\rceil}\times 2^{\lceil \log_2 T\rceil}$ DFT matrix.
- *Hadamard matrix:* Referring to [17], since the RIS training patterns based on the DFT matrix require accurate phase shift resolution, we can employ a $2^{\lceil \log_2 T\rceil}\times 2^{\lceil \log_2 T\rceil}$ Hadamard matrix to replace the DFT matrix. In this scenario the Hadamard matrix representing the RIS pattern assists us in avoiding the RIS phase quantization error, since it is based on two discrete phase shifts, i.e. 0 and π.

In each RIS training pattern, the conjugate transpose of the user pilot sequences, i.e. $[x_1^{(k)*}, x_2^{(k)*}, \cdots, x_K^{(k)*}]$, is designed as the receive combining vector for the BS observation $\mathbf{y}_1^{(k)}, \mathbf{y}_2^{(k)}, \cdots, \mathbf{y}_K^{(k)}$ to estimate $\mathbf{s}^{(k)}$. Thus, we can get the equivalent BS observation for the estimation of $\mathbf{s}^{(k)}$, denoted by $\mathbf{y}^{(k)}$, as

$$\begin{aligned}\mathbf{y}^{(k)} =& \sum_{i=1}^{K}\mathbf{y}_i^{(k)}x_i^{(k)*} \\ =& \sqrt{P}\sum_{i=1}^{K}\Big(\sum_{k=1}^{K}\mathbf{Z}^{(k)}\mathbf{s}^{(k)}x_i^{(k)}x_i^{(k)*}\Big) + \sum_{i=1}^{K}\mathbf{n}_i^{(k)}x_i^{(k)*} \\ =& K\sqrt{P}\mathbf{Z}^{(k)}\mathbf{s}^{(k)} + \mathbf{n}'^{(k)},\end{aligned} \tag{14}$$

where $\mathbf{n}'^{(k)} \sim \mathcal{CN}(\mathbf{0}_{MT}, K\sigma_{\mathrm{n}}^2\mathbf{I}_{MT})$. Therefore, (14) is equivalent to

$$\mathbf{y}^{(k)} = \sqrt{KP}\mathbf{Z}^{(k)}\mathbf{s}^{(k)} + \mathbf{n}^{(k)}, \tag{15}$$

with $\mathbf{n}^{(k)} \sim \mathcal{CN}(\mathbf{0}_{MT}, \sigma_{\mathrm{n}}^2\mathbf{I}_{MT})$.

To design the LMMSE algorithm, the mean and the covariance matrix of $\mathbf{s}^{(k)}$ should be obtained. Firstly, the mean of $\mathbf{s}^{(k)}$ is given by

$$\begin{aligned}\mathbb{E}[\mathbf{s}^{(k)}] =& \mathbb{E}[\mathbf{b}^{(k)\mathrm{T}}, \mathbf{s}_1^{(k)\mathrm{T}}, \cdots, \mathbf{s}_M^{(k)\mathrm{T}}]^{\mathrm{T}} \\ =& \sqrt{\frac{\kappa_{\mathbf{A}}\kappa_{\mathbf{g}}}{(1+\kappa_{\mathbf{A}})(1+\kappa_{\mathbf{g}})}}\begin{bmatrix}\mathbf{0}_M \\ \overline{\mathbf{a}}_1^{\mathrm{T}}\odot\overline{\mathbf{g}}^{(k)} \\ \vdots \\ \overline{\mathbf{a}}_M^{\mathrm{T}}\odot\overline{\mathbf{g}}^{(k)}\end{bmatrix}.\end{aligned} \tag{16}$$

Then, the covariance matrix of $\mathbf{s}^{(k)}$ is formulated as $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} = \mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}] - \mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)\mathrm{H}}]$, where $\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)\mathrm{H}}]$ can be calculated based on (16) as

$$\begin{aligned}&\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)\mathrm{H}}] \\ =& \begin{bmatrix}\mathbf{O}_{M\times M} & \mathbf{O}_{M\times N} & \cdots & \mathbf{O}_{M\times N} \\ \mathbf{O}_{N\times M} & \overline{\mathbf{A}}_{1,1}\odot\overline{\mathbf{G}}^{(k)} & \cdots & \overline{\mathbf{A}}_{1,M}\odot\overline{\mathbf{G}}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{N\times M} & \overline{\mathbf{A}}_{M,1}\odot\overline{\mathbf{G}}^{(k)} & \cdots & \overline{\mathbf{A}}_{M,M}\odot\overline{\mathbf{G}}^{(k)}\end{bmatrix},\end{aligned} \tag{17}$$

and $\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]$ can be expressed as

$$\begin{aligned}&\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}] \\ =& \begin{bmatrix}\mathbf{O}_{M\times M} & \mathbf{O}_{M\times N} & \cdots & \mathbf{O}_{M\times N} \\ \mathbf{O}_{N\times M} & \overline{\mathbf{A}}_{1,1}\odot\mathbf{V}^{(k)} & \cdots & \overline{\mathbf{A}}_{1,M}\odot\mathbf{V}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{N\times M} & \overline{\mathbf{A}}_{M,1}\odot\mathbf{V}^{(k)} & \cdots & \overline{\mathbf{A}}_{M,M}\odot\mathbf{V}^{(k)}\end{bmatrix} \\ &+ \begin{bmatrix}\mathbf{I}_M & \mathbf{O}_{M\times MN} \\ \mathbf{O}_{MN\times M} & \mathbf{I}_M\otimes(\mathbf{R}_{\mathbf{A}}\odot\mathbf{V}^{(k)})\end{bmatrix},\end{aligned} \tag{18}$$

where $\mathbf{V}^{(k)} = \overline{\mathbf{G}}^{(k)} + \mathbf{R}_{\mathbf{g}}$. In (17) and (18), $\mathbf{R}_{\mathbf{A}} = \frac{1}{1+\kappa_{\mathbf{A}}}\mathbf{R}$, $\mathbf{R}_{\mathbf{g}} = \frac{1}{1+\kappa_{\mathbf{g}}}\mathbf{R}$, $\overline{\mathbf{G}}^{(k)} = \frac{\kappa_{\mathbf{g}}}{1+\kappa_{\mathbf{g}}}\overline{\mathbf{g}}^{(k)}\overline{\mathbf{g}}^{(k)\mathrm{H}}$ and $\overline{\mathbf{A}}_{m_1,m_2} = \frac{\kappa_{\mathbf{A}}}{1+\kappa_{\mathbf{A}}}\overline{\mathbf{a}}_{m_1}^{\mathrm{T}}\overline{\mathbf{a}}_{m_2}^{*}$. According to (17) and (18), we arrive at

$$\begin{aligned}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} =& \begin{bmatrix}\mathbf{O}_{M\times M} & \mathbf{O}_{M\times N} & \cdots & \mathbf{O}_{M\times N} \\ \mathbf{O}_{N\times M} & \overline{\mathbf{A}}_{1,1}\odot\mathbf{R}_{\mathbf{g}} & \cdots & \overline{\mathbf{A}}_{1,M}\odot\mathbf{R}_{\mathbf{g}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{N\times M} & \overline{\mathbf{A}}_{M,1}\odot\mathbf{R}_{\mathbf{g}} & \cdots & \overline{\mathbf{A}}_{M,M}\odot\mathbf{R}_{\mathbf{g}}\end{bmatrix} \\ &+ \begin{bmatrix}\mathbf{I}_M & \mathbf{O}_{M\times MN} \\ \mathbf{O}_{MN\times M} & \mathbf{I}_M\otimes(\mathbf{R}_{\mathbf{A}}\odot\mathbf{V}^{(k)})\end{bmatrix}.\end{aligned} \tag{19}$$

Based on (16) and (19), the SoA LMMSE estimator of $\mathbf{s}^{(k)}$, denoted as $\hat{\mathbf{s}}_{\mathcal{S}}^{(k)}$, is given by

$$\hat{\mathbf{s}}_{\mathcal{S}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}^{-1}(\mathbf{y}^{(k)} - \mathbb{E}[\mathbf{y}^{(k)}]), \tag{20}$$

where the detailed derivation can be found in [31]. Since $\mathbb{E}[\mathbf{y}^{(k)}] = \sqrt{KP}\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]$,

$\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}} = \sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}}$ and $\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}} = KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT}$, and we can get the SoA LMMSE estimator of $\mathbf{s}^{(k)}$ as

$$\hat{\mathbf{s}}_{\mathcal{S}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}}(KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT})^{-1}(\mathbf{y}^{(k)} - \sqrt{KP}\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]). \quad (21)$$

In the SoA LMMSE estimator, the number of RIS training patterns T must satisfy $T \geq N+1$. Since a large number of reflecting elements are used in a practical RIS-aided communication systems, the SoA LMMSE cannot work when the coherence interval length is limited. Therefore, it is necessary to conceive low-overhead channel estimation methods. In [21]–[24], the authors partitioned the N RIS elements into $N_{\mathcal{G}}$ groups, each of which contains $\frac{N}{N_{\mathcal{G}}}$ elements, where it was assumed that the channel links corresponding to the reflecting elements in the same group have identical CSI. Therefore, we can combine the grouping idea and the LMMSE method to derive a grouping-based LMMSE estimator to reduce the pilot overhead. However, the grouping idea has the drawback that it assumes the NLoS CSI corresponding to the RIS elements in the same group is identical, while the NLoS CSI corresponding to the RIS elements in different groups is uncorrelated. In practice, this assumption is difficult to satisfy, because this grouping idea is based on the assumption that the spatial correlation matrix is $\mathbf{R} = \mathbf{I}_{N_{\mathcal{G}}} \otimes \left(\mathbf{1}_{\frac{N}{N_{\mathcal{G}}}} \cdot \mathbf{1}_{\frac{N}{N_{\mathcal{G}}}}^{\mathrm{T}}\right)$. To overcome these drawbacks, we propose the following KL-LMMSE method.

2) KL transformation based LMMSE method: As described in [47], an $N \times 1$ correlated random vector can be approximately represented by an $N_{\mathcal{E}}$ ($N_{\mathcal{E}} \leq N$) parameters after the Karhunen-Loève transform. Specifically, if an $N \times 1$ correlated random vector $\mathbf{x} \in \mathbb{C}^{N\times 1}$ has a mean of $\mathbf{0}_N$ and a covariance matrix of $\mathbf{C}_{\mathbf{xx}}$, the $N_{\mathcal{E}}$-order ($1 \leq N_{\mathcal{E}} \leq N$) Karhunen-Loève transform of $\mathbf{x}$ is given by $\mathbf{x}_{\mathcal{E}} = \sum_{i=1}^{N_{\mathcal{E}}}\sqrt{\lambda_i}w_i\mathbf{u}_i$, where $\lambda_1, \lambda_2, \cdots, \lambda_N$ are the eigenvalues of $\mathbf{C}_{\mathbf{xx}}$ in the descending order, and $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N$ are the corresponding eigen-vectors. This means that the $N \times 1$ vector $\mathbf{x}$ can be approximately represented by $N_{\mathcal{E}}$ parameters $w_1, w_2, \cdots .w_{N_{\mathcal{E}}}$. The corresponding Karhunen-Loève transform error is $\mathbf{x}_{\mathcal{E},\mathrm{e}} = \mathbf{x} - \mathbf{x}_{\mathcal{E}} = \sum_{i=N_{\mathcal{E}}+1}^{N}\sqrt{\lambda_i}w_i\mathbf{u}_i$. The random variables $w_1, w_2, \cdots, w_N$ are i.i.d. with zero mean and a variance of 1. The Karhunen-Loève transform is optimal for linear approximations, since it results in the least mean square error of $\mathbf{x}_{\mathcal{E},\mathrm{e}}$ [47], where the value of $N_{\mathcal{E}}$ determines the approximation accuracy. Therefore, we can employ the Karhunen-Loève transform for estimating the vectors $\mathbf{s}^{(k)}$. Here, we reserve the direct link $\mathbf{b}^{(k)}$ in $\mathbf{s}^{(k)}$, and only employ the Karhunen-Loève transform for the vectors $\mathbf{s}_1^{(k)}, \mathbf{s}_2^{(k)}, \cdots, \mathbf{s}_M^{(k)}$. Specifically, we select $\mathbf{s}_1^{(k)}, \mathbf{s}_2^{(k)}, \cdots, \mathbf{s}_M^{(k)}$ from $\mathbf{s}^{(k)}$ and introduce $\mathbf{s}'^{(k)} = [\mathbf{s}_1^{(k)\mathrm{T}}, \mathbf{s}_2^{(k)\mathrm{T}}, \cdots, \mathbf{s}_M^{(k)\mathrm{T}}]^{\mathrm{T}}$. The covariance of $\mathbf{s}'^{(k)}$ can be readily obtained by selecting the corresponding elements in $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}$. The Karhunen-Loève transform of $\mathbf{s}^{(k)}$, denoted as $\mathbf{s}_{\mathcal{E}}^{(k)}$, is given by

$$\mathbf{s}_{\mathcal{E}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \mathbf{T}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}, \quad (22)$$

while the Karhunen-Loève transform error of $\mathbf{s}^{(k)}$ is

$$\mathbf{s}_{\mathcal{E},\mathrm{e}}^{(k)} = \mathbf{T}_{\mathcal{E},\mathrm{e}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}, \quad (23)$$

where $\mathbf{w}_{\mathcal{E}}^{(k)} = [\mathbf{b}^{(k)\mathrm{T}}, w_1^{(k)}, w_2^{(k)}, \cdots, w_{MN_{\mathcal{E}}}^{(k)}]^{\mathrm{T}}$, $\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)} = [w_{MN_{\mathcal{E}}+1}^{(k)}, w_{MN_{\mathcal{E}}+2}^{(k)}, \cdots, w_{MN}^{(k)}]^{\mathrm{T}}$ and

$$\mathbf{T}_{\mathcal{E}} = \begin{bmatrix} \mathbf{I}_M & \mathbf{O}_{M\times MN_{\mathcal{E}}} \\ \mathbf{O}_{MN\times M} & \left[\sqrt{\lambda_1^{(k)}}\mathbf{u}_1^{(k)}, \cdots, \sqrt{\lambda_{MN_{\mathcal{E}}}^{(k)}}\mathbf{u}_{MN_{\mathcal{E}}}^{(k)}\right] \end{bmatrix}, \quad (24)$$

$$\mathbf{T}_{\mathcal{E},\mathrm{e}} = \left[\sqrt{\lambda_{MN_{\mathcal{E}}+1}^{(k)}}\mathbf{u}_{MN_{\mathcal{E}}+1}^{(k)}, \cdots, \sqrt{\lambda_{MN}^{(k)}}\mathbf{u}_{MN}^{(k)}\right], \quad (25)$$

in which $\lambda_1^{(k)}, \lambda_2^{(k)}, \cdots, \lambda_{MN}^{(k)}$ are the eigen-values of $\mathbf{C}_{\mathbf{s}'^{(k)}\mathbf{s}'^{(k)}}$ in the descending order, and $\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \cdots, \mathbf{u}_{MN}^{(k)}$ are the corresponding eigen-vectors. Then, the KL-LMMSE method has the following two stages.

In the first stage, the LMMSE associated with the RIS training patterns satisfying $T \geq N_{\mathcal{E}}+1$ is employed for estimating the vector $\mathbf{w}_{\mathcal{E}}^{(k)}$ based on the BS observation $\mathbf{y}^{(k)}$. Here, the BS observation $\mathbf{y}^{(k)}$ can be represented as

$$\begin{aligned}\mathbf{y}^{(k)} &= \sqrt{KP}\mathbf{Z}^{(k)}\mathbf{s}^{(k)} + \mathbf{n}^{(k)} \\ &= \mathbb{E}[\mathbf{y}^{(k)}] + \sqrt{KP}\mathbf{Z}^{(k)}(\mathbf{s}^{(k)} - \mathbb{E}[\mathbf{s}^{(k)}]) + \mathbf{n}^{(k)} \\ &= \mathbb{E}[\mathbf{y}^{(k)}] + \sqrt{KP}\mathbf{Z}^{(k)}(\mathbf{T}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)} + \mathbf{T}_{\mathcal{E},\mathrm{e}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}) + \mathbf{n}^{(k)},\end{aligned} \quad (26)$$

where $\mathbf{Z}^{(k)}$ is given in (13). Based on (26) and on the definition of the LMMSE algorithm, the LMMSE estimator of $\mathbf{w}_{\mathcal{E}}^{(k)}$ is formulated as

$$\hat{\mathbf{w}}_{\mathcal{E}}^{(k)} = \mathbb{E}[\mathbf{w}_{\mathcal{E}}^{(k)}] + \mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}^{-1}(\mathbf{y}^{(k)} - \mathbb{E}[\mathbf{y}^{(k)}]), \quad (27)$$

where $\mathbb{E}[\mathbf{y}^{(k)}] = \sqrt{KP}\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]$, and $\mathbb{E}[\mathbf{w}_{\mathcal{E}}^{(k)}] = \mathbf{0}_{MT}$. $\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}$ and $\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}$ are given by

$$\begin{aligned}\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}} =& \sqrt{KP}(\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}} \\ &+ \mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}}\mathbf{T}_{\mathcal{E},\mathrm{e}}^{\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}),\end{aligned} \quad (28)$$

$$\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}} = KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT}, \quad (29)$$

in which $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}$ is given in (19), and

$$\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}} = \mathbf{I}_{M(N_{\mathcal{E}}+1)}, \quad (30)$$

$$\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}} = \mathbf{O}_{M(N_{\mathcal{E}}+1)\times M(N-N_{\mathcal{E}})}. \quad (31)$$

Thus, upon substituting (28) and (29) into (27), we can formulate the LMMSE estimator of $\mathbf{w}_{\mathcal{E}}^{(k)}$ as

$$\begin{aligned}\hat{\mathbf{w}}_{\mathcal{E}}^{(k)} =& \sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}(KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} \\ &+ \sigma_{\mathrm{n}}^2\mathbf{I}_{MT})^{-1}(\mathbf{y}^{(k)} - \sqrt{KP}\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]).\end{aligned} \quad (32)$$

In the second stage, when the vector $\mathbf{w}_{\mathcal{E}}^{(k)}$ is estimated, the recovered vector $\mathbf{s}$ is $\hat{\mathbf{s}}_{\mathcal{E}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \mathbf{T}_{\mathcal{E}}^{(k)}\hat{\mathbf{w}}_{\mathcal{E}}^{(k)}$. According to (32), we can get the KL-LMMSE estimator of $\mathbf{s}^{(k)}$ as

$$\hat{\mathbf{s}}_{\mathcal{E}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}(KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\cdot$$

$$\mathbf{Z}^{(k)\mathrm{H}}+\sigma_{\mathrm{n}}^2\mathbf{I}_{MT})^{-1}(\mathbf{y}^{(k)}-\sqrt{KP}\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]). \quad (33)$$

It is notable that when the direct channel is unobstructed, the performance gain attained by deploying a RIS would become less obvious due to the twin-hop path-loss characterize of the RIS, i.e. $\varrho_{\mathbf{b}^{(k)}} \gg \varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}$ [36]. However, the channel estimation method proposed in this paper would still work well, since it is not affected by the presence or absence of the direct channel link.

B. Channel Estimation in the Face of RIS Phase Quantization Error

As we mentioned in Section II-A, the RIS phase quantization error is inevitable due to its limited phase shift resolution [33]. In this section, we derive the SoA LMMSE algorithm and the KL-LMMSE algorithm for the channel estimator of multi-cell MIMO systems while considering the RIS phase quantization error.

When considering the RIS phase quantization error, the perfect RIS training patterns $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\cdots,\boldsymbol{\theta}_T$ having no phase quantization error will be replaced by the more realistic imperfect RIS training patterns $\boldsymbol{\theta}'_1,\boldsymbol{\theta}'_2,\cdots,\boldsymbol{\theta}'_T$ with phase quantization error, where the matrix $\mathbf{Z}^{(k)}$ presented in (13) will become $\mathbf{Z}'^{(k)}$ given by

$$\mathbf{Z}'^{(k)}=\begin{bmatrix}\sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M,\sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M\otimes\boldsymbol{\theta}'_1\\ \sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M,\sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M\otimes\boldsymbol{\theta}'_2\\ \vdots\\ \sqrt{\varrho_{\mathbf{b}^{(k)}}}\mathbf{I}_M,\sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M\otimes\boldsymbol{\theta}'_T\end{bmatrix}, \quad (34)$$

in which $\boldsymbol{\theta}'_t$ is the tth RIS training pattern with RIS phase quantization error, given by $\boldsymbol{\theta}'_t=\boldsymbol{\theta}_t\odot\tilde{\boldsymbol{\theta}}_t=[\mathrm{e}^{j(\theta_{t,1}+\tilde{\theta}_{t,1})},\mathrm{e}^{j(\theta_{t,2}+\tilde{\theta}_{t,2})},\cdots,\mathrm{e}^{j(\theta_{t,N}+\tilde{\theta}_{t,N})}]$, with $\boldsymbol{\theta}_t$ being the desired RIS phase shift vector and $\tilde{\boldsymbol{\theta}}_t$ being the RIS phase quantization error vector. We define $\Delta\boldsymbol{\theta}_t$ and $\Delta\mathbf{Z}$ as

$$\Delta\boldsymbol{\theta}_t=\boldsymbol{\theta}'_t-\boldsymbol{\theta}_t=[\mathrm{e}^{j(\theta_{t,1}+\tilde{\theta}_{t,1})}-\mathrm{e}^{j\theta_{t,1}},\\ \mathrm{e}^{j(\theta_{t,2}+\tilde{\theta}_{t,2})}-\mathrm{e}^{j\theta_{t,2}},\cdots,\mathrm{e}^{j(\theta_{t,N}+\tilde{\theta}_{t,N})}-\mathrm{e}^{j\theta_{t,N}}], \quad (35)$$

$$\Delta\mathbf{Z}^{(k)}=\mathbf{Z}'^{(k)}-\mathbf{Z}^{(k)}=\begin{bmatrix}\mathbf{O}_{M\times M},\sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M\otimes\Delta\boldsymbol{\theta}_1\\ \mathbf{O}_{M\times M},\sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M\otimes\Delta\boldsymbol{\theta}_2\\ \vdots\\ \mathbf{O}_{M\times M},\sqrt{\varrho_{\mathbf{g}^{(k)}}\varrho_{\mathbf{A}}}\mathbf{I}_M\otimes\Delta\boldsymbol{\theta}_T\end{bmatrix}. \quad (36)$$

In the following, we present three corollaries.

Corollary 1. When considering the RIS phase quantization error, the mean of $\Delta\mathbf{Z}^{(k)}$ is given by

$$\mathbb{E}[\Delta\mathbf{Z}^{(k)}]=(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)}, \quad (37)$$

where $\xi=\frac{I_1(\kappa_p)}{I_0(\kappa_p)}$ when the RIS phase quantization error follows the von Mises distribution of $\mathcal{VM}(0,\kappa_p)$, and $\xi=\mathrm{sinc}(\iota_{\mathrm{p}})$ when the RIS phase quantization error follows the uniform distribution of $\mathcal{U}(-\iota_{\mathrm{p}},\iota_{\mathrm{p}})$, where $I_i(\cdot)$ represents the modified Bessel functions of the first kind of order i, and $\mathrm{sinc}(\iota_{\mathrm{p}})$ represent the sinc function given by $\mathrm{sinc}(\iota_{\mathrm{p}})=\frac{\sin(\iota_{\mathrm{p}})}{\iota_{\mathrm{p}}}$. The RIS phase quantization error power is $\sigma_{\mathrm{p}}^2=\mathbb{E}[\tilde{\theta}^2]=\frac{1}{\kappa_p}$ and $\sigma_{\mathrm{p}}^2=\mathbb{E}[\tilde{\theta}^2]=\frac{1}{3}\iota_{\mathrm{p}}^2$, when it follows the von Mises distribution and the uniform distribution, respectively. Furthermore, $\mathbf{Z}_{\mathrm{p}}^{(k)}=[\mathbf{O}_{MT\times M},\mathbf{1}_{MT}\mathbf{1}_{MN}^{\mathrm{T}}]\odot\mathbf{Z}^{(k)}$, which retains the RIS-related link component and turns the direct link component to 0 in matrix $\mathbf{Z}^{(k)}$.

Proof: See Appendix A. ■

Corollary 2. When considering the RIS phase quantization error, the means of $\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}$ and $\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}$ are identical, having the value of $(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}$.

Proof: See Appendix B. ■

Corollary 3. When considering the RIS phase quantization error, the mean of $\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}$ is given by

$$\begin{aligned}&\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]\\ =&(\xi-1)^2\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}+(1-\xi^2)\mathbf{C}_{\mathrm{p}},\end{aligned} \quad (38)$$

where $\mathbf{C}_{\mathrm{p}}$ is formulated as:

$$\begin{aligned}\mathbf{C}_{\mathrm{p}}=&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\mathbf{I}_T\otimes\begin{bmatrix}\mathbf{D}_{1,1}^{(k)}&\cdots&\mathbf{D}_{1,M}^{(k)}\\ \vdots&\ddots&\vdots\\ \mathbf{D}_{M,1}^{(k)}&\cdots&\mathbf{D}_{M,M}^{(k)}\end{bmatrix}\\ &+\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\mathbf{I}_{MT}\otimes(\mathbf{1}_N^{\mathrm{T}}(\mathbf{R}_{\mathbf{A}}\odot\mathbf{V}^{(k)})\mathbf{1}_N),\end{aligned} \quad (39)$$

with $\mathbf{D}_{m_1,m_2}^{(k)}=\mathbf{1}_N^{\mathrm{T}}(\overline{\mathbf{A}}_{m_1,m_2}\odot\mathbf{V}^{(k)})\mathbf{1}_N$.

Proof: See Appendix C. ■

When considering the RIS phase quantization error, the mean of the observation $\mathbf{y}^{(k)}$ is given by

$$\begin{aligned}\mathbb{E}[\mathbf{y}^{(k)}]&=\sqrt{KP}\mathbb{E}[\mathbf{Z}'^{(k)}\mathbf{s}^{(k)}]\\ &\overset{(a)}{=}\sqrt{KP}\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]+\sqrt{KP}\mathbb{E}[\Delta\mathbf{Z}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]\\ &\overset{(b)}{=}\sqrt{KP}(\mathbf{Z}^{(k)}+(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})\mathbb{E}[\mathbf{s}^{(k)}],\end{aligned} \quad (40)$$

where (a) is based on the fact that the random matrix $\Delta\mathbf{Z}^{(k)}$ is independent of $\mathbf{s}^{(k)}$, and (b) is based on *Corollary* 1. Furthermore, $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}$ is given by

$$\begin{aligned}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}&=\mathbb{E}[(\mathbf{s}^{(k)}-\mathbb{E}[\mathbf{s}^{(k)}])(\mathbf{y}^{(k)}-\mathbb{E}[\mathbf{y}^{(k)}])^{\mathrm{H}}]\\ &=\sqrt{KP}\mathbb{E}[(\mathbf{s}^{(k)}-\mathbb{E}[\mathbf{s}^{(k)}])(\mathbf{s}^{(k)}-\mathbb{E}[\mathbf{s}^{(k)}])^{\mathrm{H}}\mathbf{Z}'^{(k)\mathrm{H}}]\\ &\overset{(a)}{=}\sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}}+\sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbb{E}[\Delta\mathbf{Z}^{(k)}]^{\mathrm{H}}\\ &\overset{(b)}{=}\sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}}+\sqrt{KP}(\xi-1)\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\\ &=\sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}(\mathbf{Z}^{(k)}+(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})^{\mathrm{H}},\end{aligned} \quad (41)$$

where (a) is based on the fact that $\Delta\mathbf{Z}^{(k)}$ is independent of $\mathbf{s}^{(k)}$, and (b) is based on *Corollary* 1.

When considering the RIS phase quantization error, the covariance matrix $\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}$ is formulated as:

$$\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}=\mathbb{E}[\mathbf{y}^{(k)}\mathbf{y}^{(k)\mathrm{H}}]-\mathbb{E}[\mathbf{y}^{(k)}]\mathbb{E}[\mathbf{y}^{(k)}]^{\mathrm{H}}. \quad (42)$$

Based on (40), $\mathbb{E}[\mathbf{y}^{(k)}]\mathbb{E}[\mathbf{y}^{(k)}]^{\mathrm{H}}$ is given by

$$\begin{aligned}\mathbb{E}[\mathbf{y}^{(k)}]\mathbb{E}[\mathbf{y}^{(k)}]^{\mathrm{H}}&=KP\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}\\ &+KP(\xi-1)\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\\ &+KP(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}\end{aligned}$$

$$+ KP(\xi-1)^2\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}$$
$$\overset{(a)}{=} KP\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}$$
$$+ KP(\xi^2-1)\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}, \quad (43)$$

where (a) is based on $\mathbf{Z}_{\mathrm{p}}^{(k)} = [\mathbf{O}_{MT\times M}, \mathbf{1}_{MT}\mathbf{1}_{MN}^{\mathrm{T}}] \odot \mathbf{Z}^{(k)}$, and $\mathbb{E}[\mathbf{y}^{(k)}\mathbf{y}^{(k)\mathrm{H}}]$ is given by

$$\mathbb{E}[\mathbf{y}^{(k)}\mathbf{y}^{(k)\mathrm{H}}] = KP\mathbb{E}[\mathbf{Z}'^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\mathbf{Z}'^{(k)\mathrm{H}}] + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT}$$
$$= KP\mathbb{E}[(\mathbf{Z}^{(k)} + \Delta\mathbf{Z}^{(k)})\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}(\mathbf{Z}^{(k)} + \Delta\mathbf{Z}^{(k)})^{\mathrm{H}}] + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT}$$
$$= KP(\mathbf{Z}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}^{(k)\mathrm{H}} + (\xi^2-1)$$
$$(\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}} - \mathbf{C}_{\mathrm{p}})) + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT}. \quad (44)$$

Based on $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} = \mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}] - \mathbb{E}[\mathbf{s}^{(k)}]\mathbb{E}[\mathbf{s}^{(k)}]^{\mathrm{H}}$, and substituting (43) and (44) into (42), we can get

$$\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}} = KP(\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} + (\xi^2-1)$$
$$(\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}} - \mathbf{C}_{\mathrm{p}})) + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT}. \quad (45)$$

Firstly, we focus our attention on the SoA LMMSE method, where we substitute (40), (41) and (45) into (20), and then we can get the LMMSE estimator of $\mathbf{s}^{(k)}$ as

$$\hat{\mathbf{s}}_{\mathcal{S}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \sqrt{KP}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}(\mathbf{Z}^{(k)} + (\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})^{\mathrm{H}}$$
$$(KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} + KP(\xi^2-1)\cdot$$
$$(\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}} - \mathbf{C}_{\mathrm{p}}) + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT})^{-1}\cdot$$
$$(\mathbf{y}^{(k)} - \sqrt{KP}(\mathbf{Z}^{(k)} + (\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})\mathbb{E}[\mathbf{s}^{(k)}]). \quad (46)$$

Then, we focus on the KL-LMMSE method, where $\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}$ is given by

$$\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}} = \mathbb{E}[(\mathbf{w}_{\mathcal{E}}^{(k)} - \mathbb{E}[\mathbf{w}_{\mathcal{E}}^{(k)}])(\mathbf{y}^{(k)} - \mathbb{E}[\mathbf{y}^{(k)}])^{\mathrm{H}}]$$
$$= \sqrt{KP}\mathbb{E}[\mathbf{T}_{\mathcal{E}}^{(k)}((\mathbf{w}_{\mathcal{E}}^{(k)} - \mathbb{E}[\mathbf{w}_{\mathcal{E}}^{(k)}])(\mathbf{w}_{\mathcal{E}}^{(k)} - \mathbb{E}[\mathbf{w}_{\mathcal{E}}^{(k)}])^{\mathrm{H}}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}}$$
$$+ (\mathbf{w}_{\mathcal{E}}^{(k)} - \mathbb{E}[\mathbf{w}_{\mathcal{E}}^{(k)}])(\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)} - \mathbb{E}[\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}])^{\mathrm{H}}\mathbf{T}_{\mathcal{E},\mathrm{e}}^{(k)\mathrm{H}})\mathbf{Z}'^{(k)\mathrm{H}}]$$
$$\overset{(a)}{=} \sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)}(\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}} + \mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}}\mathbf{T}_{\mathcal{E},\mathrm{e}}^{(k)\mathrm{H}})\mathbf{Z}^{(k)\mathrm{H}} +$$
$$\sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)}(\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}} + \mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}}\mathbf{T}_{\mathcal{E},\mathrm{e}}^{\mathrm{H}})\mathbb{E}[\Delta\mathbf{Z}^{(k)}]^{\mathrm{H}}$$
$$\overset{(b)}{=} \sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)}(\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}} + \mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}}\mathbf{T}_{\mathcal{E},\mathrm{e}}^{(k)\mathrm{H}})\mathbf{Z}^{(k)\mathrm{H}} +$$
$$\sqrt{KP}(\xi-1)\mathbf{T}_{\mathcal{E}}^{(k)}(\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E}}^{(k)}}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}} + \mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{w}_{\mathcal{E},\mathrm{e}}^{(k)}}\mathbf{T}_{\mathcal{E},\mathrm{e}}^{(k)\mathrm{H}})\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}$$
$$\overset{(c)}{=} \sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}}(\mathbf{Z}^{(k)} + (\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})^{\mathrm{H}}, \quad (47)$$

where (a) is based on the fact that the random matrix $\Delta\mathbf{Z}^{(k)}$ is independent of $\mathbf{w}_{\mathcal{E}}^{(k)}$, (b) is based on *Corollary* 1, and (c) is based on (30) and (31). Then, upon substituting (40), (45) and (47) into (33), we can get the KL-LMMSE estimator of $\mathbf{s}^{(k)}$ as

$$\hat{\mathbf{s}}_{\mathcal{E}}^{(k)} = \mathbb{E}[\mathbf{s}^{(k)}] + \sqrt{KP}\mathbf{T}_{\mathcal{E}}^{(k)}\mathbf{T}_{\mathcal{E}}^{(k)\mathrm{H}}(\mathbf{Z}^{(k)} + (\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})^{\mathrm{H}}\cdot$$
$$(KP\mathbf{Z}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}^{(k)\mathrm{H}} + KP(\xi^2-1)\cdot$$
$$(\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}} - \mathbf{C}_{\mathrm{p}}) + \sigma_{\mathrm{n}}^2\mathbf{I}_{MT})^{-1}\cdot$$
$$(\mathbf{y}^{(k)} - \sqrt{KP}(\mathbf{Z}^{(k)} + (\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)})\mathbb{E}[\mathbf{s}^{(k)}]). \quad (48)$$

TABLE II: Pilot overhead of various estimators in cell-l.

Channel estimation methods	**Pilot overhead τ_p**
SoA LS/LMMSE (OUP)	$(\sum_{l'=1}^{L} K_{l'})T_{\mathcal{S}}$
SoA LS/LMMSE (NOUP)	$K_l T_{\mathcal{S}}$
G-LS/G-LMMSE	$(\sum_{l'=1}^{L} K_{l'})T_{\mathcal{G}}$
KL-LMMSE	$(\sum_{l'=1}^{L} K_{l'})T_{\mathcal{E}}$

IV. Performance Analysis

In this section, firstly we analyze the theoretical normalized MSE performance of the SoA LMMSE estimator, and of the proposed KL-LMMSE methods. Then, we compare the pilot overhead and computational complexity of these channel estimation methods.

A. Normalized Mean Square Error

In the SoA LMMSE method, the estimation error of $\mathbf{s}^{(k)}$, denoted as $\check{\mathbf{s}}_{\mathcal{S}}^{(k)}$, is defined as $\check{\mathbf{s}}_{\mathcal{S}}^{(k)} = \mathbf{s}^{(k)} - \hat{\mathbf{s}}_{\mathcal{S}}^{(k)}$. We can get the estimation error covariance matrix of $\mathbf{s}^{(k)}$ as

$$\mathbf{C}_{\check{\mathbf{s}}_{\mathcal{S}}^{(k)}\check{\mathbf{s}}_{\mathcal{S}}^{(k)}} = \mathbb{E}[(\mathbf{s}^{(k)} - \hat{\mathbf{s}}_{\mathcal{S}}^{(k)})(\mathbf{s}^{(k)} - \hat{\mathbf{s}}_{\mathcal{S}}^{(k)})^{\mathrm{H}}]$$
$$\overset{(a)}{=} \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} - \mathbf{C}_{\hat{\mathbf{s}}_{\mathcal{S}}^{(k)}\mathbf{s}^{(k)}} - \mathbf{C}_{\mathbf{s}^{(k)}\hat{\mathbf{s}}_{\mathcal{S}}^{(k)}} + \mathbf{C}_{\hat{\mathbf{s}}_{\mathcal{S}}^{(k)}\hat{\mathbf{s}}_{\mathcal{S}}^{(k)}}$$
$$\overset{(b)}{=} \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} - \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}^{-1}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{s}^{(k)}}^{\mathrm{H}}, \quad (49)$$

where (a) is based on the same mean of $\mathbf{s}^{(k)}$ and $\hat{\mathbf{s}}_{\mathcal{S}}^{(k)}$, (b) is based on (21), while $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}$, $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}$ and $\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}$ are given in (19), (41) and (45), respectively. Therefore, the normalized MSE for UAN-k of the SoA LMMSE method is given by $\mathrm{NMSE}_{\mathcal{S}}^{(k)} = \frac{1}{M(N+1)}\mathrm{Tr}(\mathbf{C}_{\check{\mathbf{s}}_{\mathcal{S}}^{(k)}\check{\mathbf{s}}_{\mathcal{S}}^{(k)}})$.

In the KL-LMMSE method, the estimation error of $\mathbf{s}^{(k)}$, denoted as $\check{\mathbf{s}}_{\mathcal{E}}^{(k)}$, is defined as $\check{\mathbf{s}}_{\mathcal{E}}^{(k)} = \mathbf{s}^{(k)} - \hat{\mathbf{s}}_{\mathcal{E}}^{(k)}$. We can express the estimation error covariance matrix of $\mathbf{s}^{(k)}$ as

$$\mathbf{C}_{\check{\mathbf{s}}_{\mathcal{E}}^{(k)}\check{\mathbf{s}}_{\mathcal{E}}^{(k)}} = \mathbb{E}[(\mathbf{s}^{(k)} - \hat{\mathbf{s}}_{\mathcal{E}}^{(k)})(\mathbf{s}^{(k)} - \hat{\mathbf{s}}_{\mathcal{E}}^{(k)})^{\mathrm{H}}]$$
$$\overset{(a)}{=} \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} - \mathbf{C}_{\hat{\mathbf{s}}_{\mathcal{E}}^{(k)}\mathbf{s}^{(k)}} - \mathbf{C}_{\mathbf{s}^{(k)}\hat{\mathbf{s}}_{\mathcal{E}}^{(k)}} + \mathbf{C}_{\hat{\mathbf{s}}_{\mathcal{E}}^{(k)}\hat{\mathbf{s}}_{\mathcal{E}}^{(k)}}$$
$$\overset{(b)}{=} \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}} - \mathbf{T}_{\mathcal{E}}\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}^{-1}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}^{\mathrm{H}}$$
$$- \mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}^{-1}\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}^{\mathrm{H}}\mathbf{T}_{\mathcal{E}}^{\mathrm{H}}$$
$$+ \mathbf{T}_{\mathcal{E}}\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}^{-1}\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}^{\mathrm{H}}\mathbf{T}_{\mathcal{E}}^{\mathrm{H}}, \quad (50)$$

where (a) is based on the same mean of $\mathbf{s}^{(k)}$ and $\hat{\mathbf{s}}_{\mathcal{E}}^{(k)}$, (b) is based on (48), while $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}$, $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}$, $\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}$ and $\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}$ are given in (19), (41), (45) and (47), respectively. Therefore, the normalized MSE of the UAN-k in the KL-LMMSE method is given by $\mathrm{NMSE}_{\mathcal{E}}^{(k)} = \frac{1}{M(N+1)}\mathrm{Tr}(\mathbf{C}_{\check{\mathbf{s}}_{\mathcal{E}}^{(k)}\check{\mathbf{s}}_{\mathcal{E}}^{(k)}})$.

B. Pilot Overhead

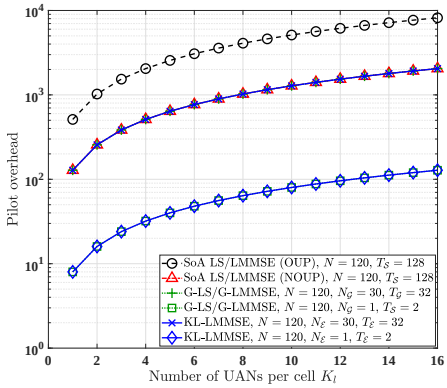
We define pilot overhead τ_p as the number of symbol slots required for channel estimators in each coherence interval. Table. II compares the pilot overhead of the SoA LS

method [15]–[18], of the SoA LMMSE method [19], including orthogonal user pilot (OUP) and non-orthogonal user pilot (NOUP) sequences, the grouping based LS (G-LS) method [21]–[24], the grouping based LMMSE (G-LMMSE) method and the KL-LMMSE method of cell-l, where the variables $T_{\mathcal{S}}$, $T_{\mathcal{G}}$ and $T_{\mathcal{E}}$ denote the number of RIS training patterns in the SoA methods, grouping methods and the proposed KL-LMMSE method, respectively, satisfying $T_{\mathcal{S}} \geq N+1$, $T_{\mathcal{G}} \geq N_{\mathcal{G}}+1$ and $T_{\mathcal{E}} \geq N_{\mathcal{E}}+1$. By exploiting the spatial correlation of RIS-related channels, the pilot overhead of the G-LS/G-LMMSE method and the KL-LMMSE method are approximately proportional to $N_{\mathcal{G}}$ and $N_{\mathcal{E}}$ respectively, instead of the number of RIS elements N. When a large number of passive RIS elements are used at a high spatial density, the G-LS/G-LMMSE method and the KL-LMMSE method can promise significantly lower pilot overhead than the SoA LS/LMMSE method.

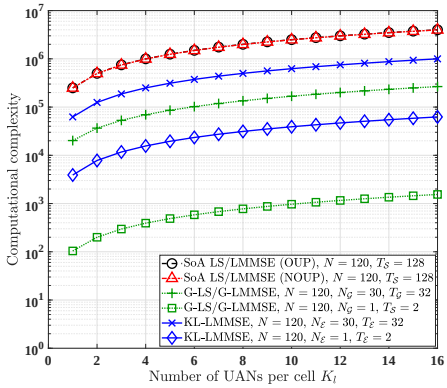
The channel capacity of UAN-k can be shown to be $R_k = \frac{\tau_c-\tau_p}{\tau_c}\log 2(1+\gamma_k)$, where τ_c represents the total number of samples in each coherence interval and γ_k represents the signal-to-interference-plus-noise-ratio (SINR) at UAN-k [31]. Hence, the channel capacity of UAN-k is determined by both the pre-log factor $\frac{\tau_c-\tau_p}{\tau_c}$ and the SINR γ_k. Compared to the SoA LS/LMMSE (OUP) estimator, the G-LS/G-LMMSE estimator and our proposed KL-LMMSE estimator can ensure higher pre-log factor $\frac{\tau_c-\tau_p}{\tau_c}$ due to their reduced pilot overhead τ_p, but this is achieved at the cost of degradation of γ_k. It means that in fast fading channels both the low-overhead G-LS/G-LMMSE estimator and the KL-LMMSE estimator achieve higher channel capacity than the SoA LS/LMMSE (OUP) estimator. Furthermore, the value of $T_{\mathcal{G}}$ in the G-LS/G-LMMSE estimator and $T_{\mathcal{E}}$ in the KL-LMMSE estimator can be adaptively adjusted to increase the channel capacity based on the total number of samples in each coherence interval.

C. Computational Complexity

The complexity of various channel estimation methods can be quantified in terms of the number of complex multiplications in each coherence interval. Since the statistical information, e.g. the values of $\mathbb{E}[\mathbf{s}^{(k)}]$, $\mathbb{E}[\mathbf{y}^{(k)}]$, $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}$, $\mathbf{C}_{\mathbf{y}^{(k)}\mathbf{y}^{(k)}}$, $\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{y}^{(k)}}$ and $\mathbf{C}_{\mathbf{w}_{\mathcal{E}}^{(k)}\mathbf{y}^{(k)}}$, are updated only at the beginning of each stationary block, the complexity of these values can be ignored. In each coherence interval, firstly the received signal vector $\mathbf{y}^{(k)}$ is combined with the conjugate transpose of the user pilot sequences, in which the calculation complexity is equal to the pilot overhead times the number of antennas at BS-l. Then, the CSIs are estimated for each UAN by the various channel estimators. Table. III compares the calculation complexity of the SoA LS method, the SoA LMMSE method, including the OUP sequences and NOUP sequences, the G-LS method, the G-LMMSE method, and the KL-LMMSE method for estimating the CSI in cell-l. Since the number of RIS elements N is significantly higher than the number of UANs in all L cells in a practical RIS-aided wireless system, the G-LS/G-LMMSE method and the KL-LMMSE method have the complexity order of $\mathcal{O}(M_l^2 N_{\mathcal{G}}^2)$ and $\mathcal{O}(M_l^2 N_{\mathcal{E}} N)$, respectively, which are significantly lower than the SoA LS/LMMSE complexity order of $\mathcal{O}(M_l^2 N^2)$.



(a) Pilot overhead



(b) Computational complexity

Fig. 3: Comparison of the pilot overhead and computational complexity versus the number of UANs per cell K_l in SoA LS/LMMSE (OUP), SoA LS/LMMSE (NOUP), G-LS, G-LMMSE, and KL-LMMSE method.

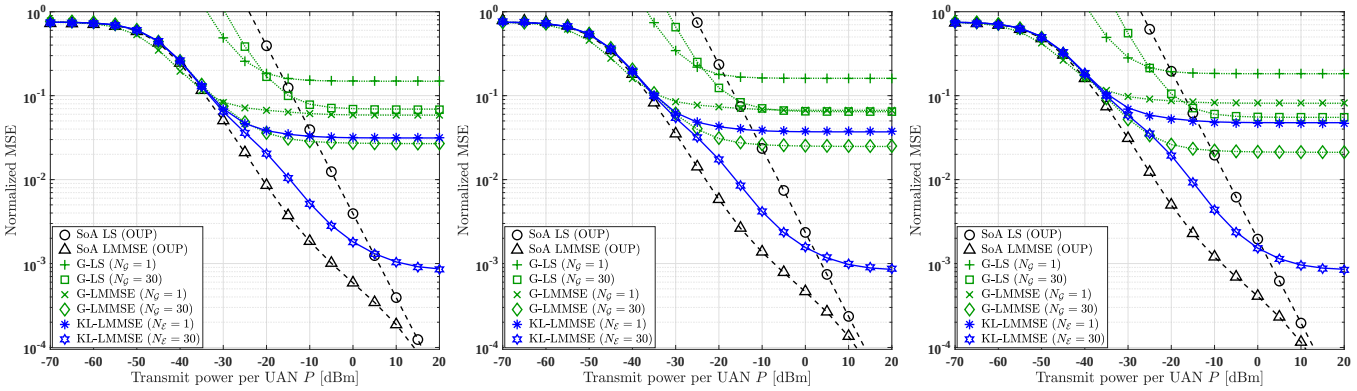
V. Simulation Results

In this section, we qualify both the pilot overhead and the computational complexity, as well as the theoretical and simulation based normalized MSE of different channel estimators, including the SoA LS method, the SoA LMMSE method, the G-LS method in [21]–[24], the G-LMMSE method, and our proposed KL-LMMSE method, all in multi-cell RIS-aided wireless communications.

We use the system model of Fig. 1, where a RIS is supporting four cells. The 3-dimensional Cartesian coordinate of the RIS is (0m, 30m, 8m), and the four BSs are located at (0m, 0m, 20m), ($-D_c$, 0m, 20m), (0m, $-D_c$, 20m) and (D_c, 0m, 20m), respectively, where D_c is the distance between adjacent BSs. We focused our attention on the channel estimation performance of the users in cell-1 to BS-1, i.e. the one having coordinates of (0m, 0m, 20m). Therefore, the signals transmitted from the users in others cells surrounding cell-1 can be viewed as interference. The users in each cell are at 25m in front of the corresponding BS, at the height of 3m and the adjacent users are 3m apart. Unless otherwise specified, the simulation parameters employed in this section are as follows. The number of antennas at BS-1 is $M_1 = 4$. The number of RIS elements is $N = 12 \times 10 = 120$. The number of users in each cell is 4 with each user equipped

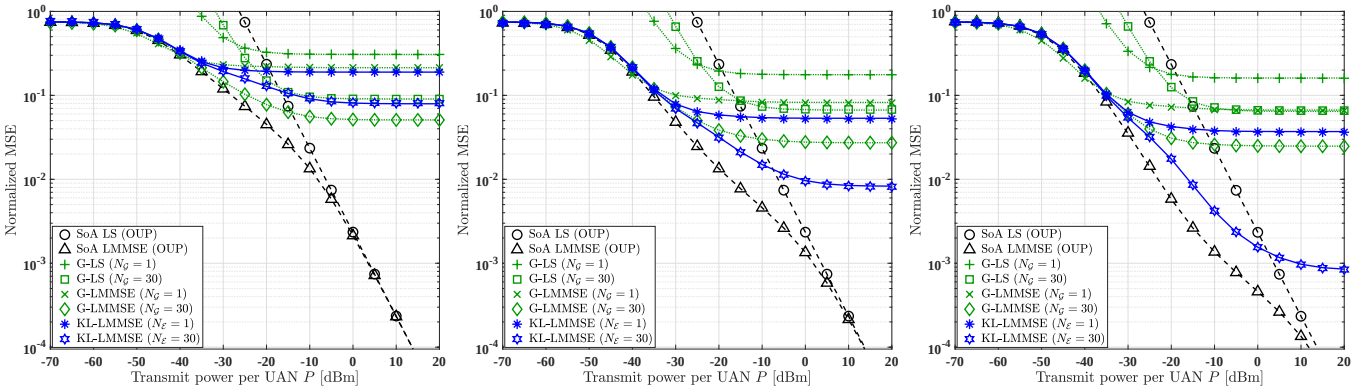
TABLE III: Calculation complexity of various channel estimators in cell-l.

Channel estimation methods	**Combined with user pilots**	**Estimated per user**	**Total complexity**
SoA LS/LMMSE (OUP)	$M_l(\sum_{l'=1}^{L} K_{l'})T_{\mathcal{S}}$	$M_l^2(N+1)T_{\mathcal{S}}$	$T_{\mathcal{S}}(M_l^2K_l(N+1)+M_l(\sum_{l'=1}^{L} K_{l'}))$
SoA LS/LMMSE (NOUP)	$M_lK_lT_{\mathcal{S}}$	$M_l^2(N+1)T_{\mathcal{S}}$	$T_{\mathcal{S}}(M_l^2K_l(N+1)+M_lK_l)$
G-LS/G-LMMSE	$M_l(\sum_{l'=1}^{L} K_{l'})T_{\mathcal{G}}$	$M_l^2(N_{\mathcal{G}}+1)T_{\mathcal{G}}+M_lN_{\mathcal{G}}$	$T_{\mathcal{G}}(M_l^2K_l(N_{\mathcal{G}}+1)+M_lN_{\mathcal{G}}+M_l(\sum_{l'=1}^{L} K_{l'}))$
KL-LMMSE	$M_l(\sum_{l'=1}^{L} K_{l'})T_{\mathcal{E}}$	$M_l^2(N+1)T_{\mathcal{E}}$	$T_{\mathcal{E}}(M_l^2K_l(N+1)+M_l(\sum_{l'=1}^{L} K_{l'}))$



(a) $K_l = 2$ (b) $K_l = 4$ (c) $K_l = 8$

Fig. 4: Comparison of the normalized MSE versus transmit power per UAN P in SoA LS (OUP), SoA LMMSE (OUP), G-LS, G-LMMSE, and KL-LMMSE method, with different number of total UANs in each cell.



(a) $\delta_0 = 1\lambda$ (b) $\delta_0 = 10\lambda$ (c) $\delta_0 = 100\lambda$

Fig. 5: Comparison of the normalized MSE versus transmit power per UAN P in SoA LS (OUP), SoA LMMSE (OUP), G-LS, G-LMMSE, and KL-LMMSE method, with different correlation reference distance.

with 2 antennas. The Rician factor are $\kappa_{\mathbf{g}} = \kappa_{\mathbf{A}} = 0$dB. Since the direct UE-RIS links are blocked, referring to [18], the path loss exponents $\alpha_{\mathbf{b}}$ would be much larger than $\alpha_{\mathbf{A}}$ and $\alpha_{\mathbf{g}}$. Here we set the path loss exponents to $\alpha_{\mathbf{A}} = 2$, $\alpha_{\mathbf{g}} = 2.2$ and $\alpha_{\mathbf{b}} = 4.8$, along with $\varrho_0 = -30$dB. The additive noise power is $\sigma_{\mathrm{n}}^2 = -100$dBm. The distance between the adjacent RIS elements is $\delta^{(x)} = \delta^{(y)} = \frac{1}{30}\lambda$. The correlation reference distance is $\delta_0 = 100\lambda$. The distance between adjacent BS antennas $d_0 = \frac{1}{2}\lambda$. AoA at BS-1 is $\psi_1^{(\mathrm{A})} = \frac{\pi}{3}$. The distance between adjacent cells is $D_{\mathrm{c}} = 60$m. The values of $\phi_1^{(\mathrm{A})}, \phi_2^{(\mathrm{A})}, \cdots, \phi_K^{(\mathrm{A})}$, $\varphi_1^{(\mathrm{A})}, \varphi_2^{(\mathrm{A})}, \cdots, \varphi_K^{(\mathrm{A})}$, $\phi_1^{(\mathrm{D})}$ and $\varphi_1^{(\mathrm{D})}$ are calculated according to the relative position of the BSs, the RIS and the UANs.

We compare the SoA LS, SoA LMMSE, G-LS, G-LMMSE and our proposed KL-LMMSE methods. We set the number of RIS training patterns to $T_{\mathcal{S}} = 128$ for the SoA LS/LMMSE method, set $T_{\mathcal{G}} = 2$ and 32 for the G-LS/G-LMMSE method with $N_{\mathcal{G}} = 1$ and 30, respectively, and $T_{\mathcal{E}} = 2$ and 32 for the KL-LMMSE method with $N_{\mathcal{E}} = 1$ and 30, respectively. In all channel estimation methods, the generation of the correlated channel links of all channel estimation methods in the simulation are identical. In the G-LMMSE estimator, it just considers the correlation among different groups while it assumes that the links corresponding to the elements in

the same group have identical CSI, which is an impractical assumption. This is the flaw of the grouping based estimation method, and our proposed KL-LMMSE estimator avoids this assumption.

Fig. 3 shows both the pilot overhead and computational complexity versus the number of UANs per cell K_l in the SoA LS/LMMSE (OUP), SoA LS/LMMSE (NOUP), G-LS, G-LMMSE, and the proposed KL-LMMSE method, based on Table II and Table III. Explicitly, both the estimators based on grouping ideas and our proposed KL-LMMSE method have significantly lower pilot overhead than the SoA LS/LMMSE (OUP) method, while the G-LS/G-LMMSE ($N_{\mathcal{G}} = 30$), and KL-LMMSE ($N_{\mathcal{E}} = 30$) have the same pilot overhead as the SoA LS/LMMSE (NOUP) method. Additionally, both the estimators based on grouping ideas and our proposed KL-LMMSE method have a significantly lower computational complexity than the SoA LS/LMMSE method, as shown in Fig. 3.

From Fig. 4 to Fig. 7, we utilize lines, e.g. '$--$' and '$-\cdot-\cdot$', to represent the theoretical analysis of the normalized MSE, and utilize markers, e.g. '$\square$' and '$\times$', to represent the simulation results. In order to investigate the normalized MSE performance of different channel estimators, in Fig. 4, Fig. 5 and Fig. 6, the RIS patterns are designed based on the Hadamard matrix to avoid phase quantization error, i.e. $\sigma_{\mathrm{p}}^2 = 0$. It is worth noting that the normalized MSE performance of the SoA LMMSE (OUP) method is better than that of the G-LMMSE and of our proposed KL-LMMSE methods, albeit at the cost of a higher pilot overhead and computational complexity, as shown in Fig. 3. In Fig. 4, we compare the normalized MSE versus the transmit power per UAN P in the SoA LS (OUP), SoA LMMSE (OUP), G-LS, G-LMMSE, and KL-LMMSE methods, with the number of users in each cell set to $1, 2, 4$, respectively. Each user is equipped with 2 antennas, i.e. the total number of user antennas per cell is $K_l = 2, 4, 8$, respectively. Fig. 4 shows that the MSE performance of the SoA LMMSE method is better than that of the SoA LS method, and the G-LMMSE method performs better than the G-LS method, especially in the low transmit power region. This is because the statistical information of the first-order moment and the second-order moment of the channels are utilized in the LMMSE methods. Additionally, in the high transmit power region of Fig. 4, the normalized MSE of the G-LMMSE method and the KL-LMMSE method performs worse than the SoA LMMSE (OUP) method. Explicitly, a residence MSE arises, since non-negligible channel estimation error is inflicted owing to the discarded channel information in the G-LMMSE and KL-LMMSE, which dominates the gradually diminishing noise at the BS. However, in the low transmit power region, the normalized MSE of the G-LMMSE method and the KL-LMMSE method is comparable to that of the SoA LMMSE (OUP) method, since the BS's noise is the main factor governing the channel estimation error. Furthermore, the KL-LMMSE outperforms the G-LMMSE, since the KL-LMMSE relies on the Karhunen-Loève transform, which results in the least mean square error during information compression. Furthermore, Fig. 4 shows that with the increase of $N_{\mathcal{E}}$, the MSE performance of KL-LMMSE method tends towards that of the SoA LMMSE (OUP) method.

In Fig. 5, we compare the normalized MSE versus the transmit power per UAN P in SoA LS (OUP), SoA LMMSE (OUP), G-LS ($N_{\mathcal{G}} = 1, 30$), G-LMMSE ($N_{\mathcal{G}} = 1, 30$), and KL-LMMSE ($N_{\mathcal{E}} = 1, 30$) methods, at the correlation reference distance of $\delta_0 = 1\lambda, 10\lambda, 100\lambda$. Fig. 5 shows that the normalized MSE performance of the KL-LMMSE method tends to that of the SoA LMMSE (OUP) method upon increasing of the correlation reference distance. However, the G-LMMSE has a consistently high normalized MSE because the G-LMMSE method assumes that the NLoS CSI corresponding to the RIS elements in the same group is identical. However, when the LoS links exist, i.e. $\kappa_{\mathbf{A}} > 0$ or $\kappa_{\mathbf{g}} > 0$, according to (19) the covariance of channels corresponding to the RIS element in the same group is not necessarily identical. Therefore, the low pilot overhead KL-LMMSE method is suitable for the RIS-aided systems with a large number of compact RIS elements employed.

In Fig. 6, we compare the SoA LS (NOUP) and SoA LMMSE (NOUP), G-LS ($N_{\mathcal{G}} = 30$) and G-LMMSE ($N_{\mathcal{G}} = 30$), and KL-LMMSE ($N_{\mathcal{E}} = 30$) methods, while using the same pilot overhead for fairness of comparison and also employing the adjacent BSs distances of $D_{\mathrm{c}} = 60\mathrm{m}, 120\mathrm{m}, 180\mathrm{m}$. Since the user pilot sequences in the G-LS, G-LMMSE and KL-LMMSE methods are orthogonal, varying the distance between the adjacent cells has no effect on the channel estimation performance. However, in the SoA LS (NOUP) and SoA LMMSE (NOUP) methods, the normalized MSE performance degrades with the reduction of the distance between adjacent cells, since the neighbouring cells cause higher interference levels. As shown in Fig. 6, the performance of the proposed KL-LMMSE method is significantly better than that of the SoA LS (NOUP) method and SoA LMMSE (NOUP) method, even when the distance between adjacent cells is large.

Next, we investigate the normalized MSE performance of various LMMSE channel estimation methods in the face of RIS phase quantization error, when the RIS patterns are designed based on the DFT matrix. To ensure the same pilot overhead for the SoA LMMSE (NOUP) and the low pilot overhead estimator of G-LMMSE and KL-LMMSE, we set $T_{\mathcal{G}} = 32$ along with $N_{\mathcal{G}} = 30$ and $T_{\mathcal{E}} = 32$ in conjunction with $N_{\mathcal{E}} = 30$. In Fig. 7, we compare the normalized MSE versus the transmit power per UAN P in the SoA LMMSE (OUP), SoA LMMSE (NOUP), G-LMMSE ($N_{\mathcal{G}} = 30$), and KL-LMMSE ($N_{\mathcal{E}} = 30$) methods, where the RIS phase quantization error has a power of $\sigma_{\mathrm{p}}^2 = 0.05$ and follows the von Mises distribution and uniform distribution. In each figure, we compare the normalized MSE performance both with and without RIS phase quantization error, where 'consider' in legend means considering the RIS phase quantization error in the design of the estimator, while 'ignore' in legend means ignoring the RIS phase quantization error in the design of the estimator. Fig. 7 shows that our proposed KL-LMMSE method performs significantly better than the SoA LMMSE (NOUP) method and the G-LMMSE method. Furthermore, observe in Fig. 7 that as expected, when the power of RIS phase quantization error is $\sigma_{\mathrm{p}}^2 = 0.05$, the normalized MSE performance

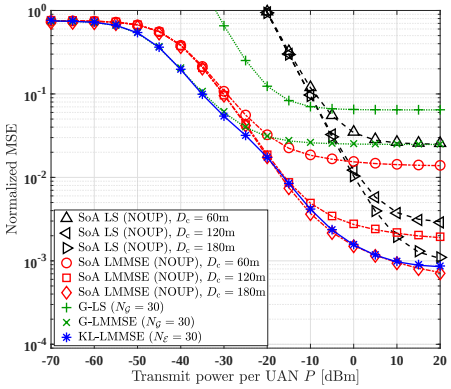


Fig. 6: Comparison of the normalized MSE versus transmit power per UAN P in SoA LS (NOUP), SoA LMMSE (NOUP), G-LS, G-LMMSE, and KL-LMMSE method, with the adjacent BSs distance $D_{\mathrm{c}} = 60\mathrm{m}, 120\mathrm{m}, 180\mathrm{m}$.

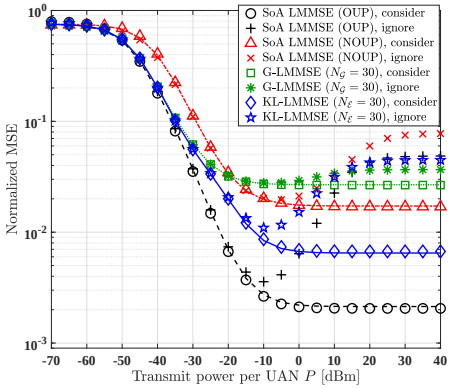
is better upon considering the RIS phase quantization error during the design of channel estimation methods than upon ignoring it. Explicitly, Fig. 7 shows that when the RIS phase quantization error is ignored, the normalized MSE escalates in the region of high transmit power values, since the ignored statistical information concerning the RIS phase quantization error in (40), (41), (45) and (47) increases upon increasing the transmit power.

VI. Conclusions

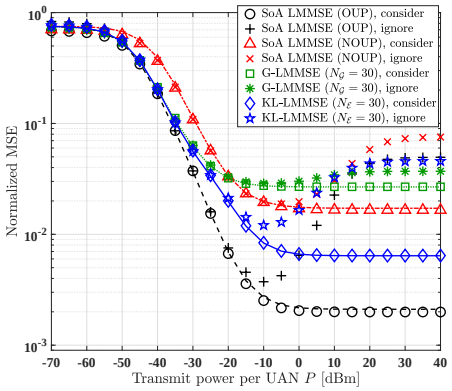
In this paper, a low-overhead channel estimation method, termed as the KL transformation based LMMSE estimator, was proposed for RIS-aided multi-cell MIMO systems by exploiting the spatial correlation of the channels. The numerical results show that the MSE performance of our proposed method tends to that of the SoA LMMSE estimator relying on orthogonal pilot sequences upon increasing the RIS spatial correlation, despite its reduced pilot overhead. Furthermore, by exploiting the statistical spatial correlation information of RIS-cascaded channel links, our proposed KL transformation based LMMSE estimator has better MSE performance than that of the grouping based LMMSE estimator. In general, we provided a methodology for designing low-overhead channel estimation methods by exploiting the spatial correlation of cascaded RIS-aided channel links and by theoretically analyzing the corresponding MSE performance. Furthermore, we evaluated the impact of the RIS phase quantization error on the channel estimation accuracy and showed the significance of considering the RIS phase quantization error effects on the channel estimator's design. Finally, this work opens new research directions for the low-overhead channel estimation of RIS-aided multi-cell wireless systems.

Appendix A
Proof of *Corollary* 1

Since the RIS phase quantization error $\tilde{\theta}_{t,n}$ are i.i.d. for $t = 1, 2, \cdots, T$ and $n = 1, 2, \cdots, N$, we can omit the subscripts in $\tilde{\theta}_{t,n}$ and we denote it as $\tilde{\theta}$ for simplicity. Since the distribution of RIS phase quantization error is also independent of the



(a) Von Mises distribution



(b) Uniform distribution

Fig. 7: Comparison of the normalized MSE versus transmit power per UAN P in SoA LMMSE (OUP), SoA LMMSE (NOUP), G-LMMSE, and KL-LMMSE method, with the RIS phase quantization error having the power $\sigma_{\mathrm{p}}^2 = 0.05$ following von Mises distribution and uniform distribution, respectively.

matrix $\mathbf{Z}^{(k)}$, based on (35) and (36), we can get the mean of $\Delta\mathbf{Z}^{(k)}$ as $\mathbb{E}[\Delta\mathbf{Z}^{(k)}] = \mathbb{E}[\mathrm{e}^{j\tilde{\theta}} - 1]\mathbf{Z}_{\mathrm{p}}^{(k)}$. Firstly, we focus on the RIS phase quantization error $\tilde{\theta}$ following the von Mises distribution, i.e. $\tilde{\theta} \sim \mathcal{VM}(0, \kappa_{\mathrm{p}})$. Since $\mathbb{E}[\mathrm{e}^{j\tilde{\theta}}] = \frac{I_1(\kappa_{\mathrm{p}})}{I_0(\kappa_{\mathrm{p}})}$ when $\tilde{\theta} \sim \mathcal{VM}(0, \kappa_{\mathrm{p}})$ [48], we can get $\mathbb{E}[\mathrm{e}^{j\tilde{\theta}} - 1] = \frac{I_1(\kappa_{\mathrm{p}})}{I_0(\kappa_{\mathrm{p}})} - 1$, and the RIS phase quantization error power $\sigma_{\mathrm{p}}^2 = \mathbb{E}[\tilde{\theta}^2] = \frac{1}{\kappa_p}$. Then, we focus on the RIS phase quantization error $\tilde{\theta}$ following the uniform distribution, i.e. $\tilde{\theta} \sim \mathcal{U}(-\iota_{\mathrm{p}}, \iota_{\mathrm{p}})$. When $\tilde{\theta} \sim \mathcal{U}(-\iota_{\mathrm{p}}, \iota_{\mathrm{p}})$, the ith-order moment of $\tilde{\theta}$, denoted as $\mathbb{E}[\tilde{\theta}^i]$, equal to 0 when i is odd and equal to $\frac{1}{i+1}\iota_{\mathrm{p}}^i$ when i is even. Thus, we can arrive at $\mathbb{E}[\mathrm{e}^{j\tilde{\theta}} - 1] = \sum_{i=1}^{\infty} \frac{(-1)^i}{(2i)!}\mathbb{E}[\tilde{\theta}^{2i}] = \sum_{i=1}^{\infty} \frac{(-1)^i \iota_{\mathrm{p}}^{2i}}{(2i+1)!} = \frac{\sin(\iota_{\mathrm{p}}) - \iota_{\mathrm{p}}}{\iota_{\mathrm{p}}} = \mathrm{sinc}(\iota_{\mathrm{p}}) - 1$. Besides, the RIS phase quantization error power $\sigma_{\mathrm{p}}^2 = \mathbb{E}[\tilde{\theta}^2] = \frac{1}{3}\iota_{\mathrm{p}}^2$.

Appendix B
Proof of *Corollary* 2

In $\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}$, the element with the row corresponding to the m_1th BS antenna in the t_1th RIS training pattern and the column corresponding to the m_2th BS antenna in the t_2th RIS training pattern, denoted as

$\{\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\}^{(t_1,t_2),(m_1,m_2)}$, is given by

$$\begin{aligned}&\{\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\}^{(t_1,t_2),(m_1,m_2)}\\ =&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbb{E}[s_{m_1,n_1}s_{m_2,n_2}^{*}]\mathrm{e}^{\theta_{t_1,n_1}}\mathrm{e}^{-\theta_{t_2,n_2}},\end{aligned}\tag{51}$$

where $s_{m_1,n}$ represents the element corresponding to the nth reflecting surface in $\mathbf{s}^{(k)}$ when $n>0$, and represents the element corresponding to the direct link in $\mathbf{s}^{(k)}$ when $n=0$.

In $\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}]$ the element with the row corresponding to the m_1th BS antenna in the t_1th RIS training pattern and the column corresponding to the m_2th BS antenna in the t_2th RIS training pattern, denoted as $\{\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}]\}^{(t_1,t_2),(m_1,m_2)}$, is given by

$$\begin{aligned}&\{\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}]\}^{(t_1,t_2),(m_1,m_2)}\\ =&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbb{E}[s_{m_1,n_1}s_{m_2,n_2}^{*}]\mathrm{e}^{j\theta_{t_1,n_1}}\mathrm{e}^{-j\theta_{t_2,n_2}}\cdot\\ &\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)]+\sqrt{\varrho_{\mathbf{b}^{(k)}}}\sqrt{\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}}\cdot\\ &\sum_{n_1=1}^{N}\mathbb{E}[s_{m_1,n_1}s_{m_2,0}^{*}]\mathrm{e}^{j\theta_{t_1,n_1}}\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)]\\ \overset{(a)}{=}&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbb{E}[s_{m_1,n_1}s_{m_2,n_2}^{*}]\mathrm{e}^{j\theta_{t_1,n_1}}\mathrm{e}^{-j\theta_{t_2,n_2}}\cdot\\ &\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)]\\ \overset{(b)}{=}&(\xi-1)\{\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\}^{(t_1,t_2),(m_1,m_2)}.\end{aligned}\tag{52}$$

where (a) is based on that $\mathbb{E}[s_{m_1,n_1}s_{m_2,0}^{*}]=0$, and (b) is based on (51). Thus, $\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\mathbf{Z}^{(k)\mathrm{H}}]=(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}$. Similarly, we can get that $\mathbb{E}[\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]=(\xi-1)\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}$.

Appendix C
Proof of *Corollary* 3

In $\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]$ the element with the row corresponding to the m_1th BS antenna in the t_1th RIS training pattern and the column corresponding to the m_2th BS antenna in the t_2th RIS training pattern, denoted as $\{\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]\}^{(t_1,t_2),(m_1,m_2)}$, is given by

$$\begin{aligned}&\{\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]\}^{(t_1,t_2),(m_1,m_2)}\\ =&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbb{E}[s_{m_1,n_1}^{(k)}s_{m_2,n_2}^{(k)*}]\mathrm{e}^{j\theta_{t_1,n_1}}\mathrm{e}^{-j\theta_{t_2,n_2}}\cdot\\ &\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)(\mathrm{e}^{-j\tilde{\theta}_{t_2,n_2}}-1)].\end{aligned}\tag{53}$$

In (53), when $t_1\neq t_2$ or $n_1\neq n_2$, $\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)(\mathrm{e}^{-j\tilde{\theta}_{t_2,n_2}}-1)]$ is given by

$$\begin{aligned}&\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)(\mathrm{e}^{-j\tilde{\theta}_{t_2,n_2}}-1)]\\ =&\mathbb{E}[\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1]\mathbb{E}[\mathrm{e}^{-j\tilde{\theta}_{t_2,n_2}}-1]=(\xi-1)^2;\end{aligned}\tag{54}$$

otherwise, when $t_1=t_2$ and $n_1=n_2$,

$$\begin{aligned}\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t_1,n_1}}-1)(\mathrm{e}^{-j\tilde{\theta}_{t_2,n_2}}-1)]=&\mathbb{E}[(\mathrm{e}^{j\tilde{\theta}_{t,n}}-1)(\mathrm{e}^{-j\tilde{\theta}_{t,n}}-1)]\\ =&2(1-\xi).\end{aligned}\tag{55}$$

Then, substituting (54) and (55) into (53), we can get when $t_1\neq t_2$,

$$\begin{aligned}&\{\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]\}^{(t_1,t_2),(m_1,m_2)}\\ =&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbb{E}[s_{m_1,n_1}^{(k)}s_{m_2,n_2}^{(k)*}]\mathrm{e}^{j\theta_{t_1,n_1}}\mathrm{e}^{-j\theta_{t_2,n_2}}(\xi-1)^2\\ =&(\xi-1)^2\{\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\}^{(t_1,t_2),(m_1,m_2)};\end{aligned}\tag{56}$$

and when $t_1=t_2$, we have

$$\begin{aligned}&\{\mathbb{E}[\Delta\mathbf{Z}^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]\}^{(t_1,t_2),(m_1,m_2)}\\ =&\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbb{E}[s_{m_1,n_1}^{(k)}s_{m_2,n_2}^{(k)*}]\mathrm{e}^{j\theta_{t_1,n_1}}\mathrm{e}^{-j\theta_{t_2,n_2}}(\xi-1)^2\\ &+(1-\xi^2)\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n=1}^{N}\mathbb{E}[s_{m_1,n}^{(k)}s_{m_2,n}^{(k)*}]\\ =&(\xi-1)^2\{\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbb{E}[\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{H}}]\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}\}^{(t_1,t_2),(m_1,m_2)}\\ &+(1-\xi^2)\varrho_{\mathbf{A}}\varrho_{\mathbf{g}^{(k)}}\sum_{n=1}^{N}\mathbb{E}[s_{m_1,n}^{(k)}s_{m_2,n}^{(k)*}].\end{aligned}\tag{57}$$

In (57), according to (16) and (19), $\sum_{n=1}^{N}\mathbb{E}[s_{m_1,n}^{(k)}s_{m_2,n}^{(k)*}]$ is

$$\sum_{n=1}^{N}\mathbb{E}[s_{m_1,n}^{(k)}s_{m_2,n}^{(k)*}]=\begin{bmatrix}\mathbf{D}_{1,1}^{(k)}&\cdots&\mathbf{D}_{1,M}^{(k)}\\ \vdots&\ddots&\vdots\\ \mathbf{D}_{M,1}^{(k)}&\cdots&\mathbf{D}_{M,M}^{(k)}\end{bmatrix}.\tag{58}$$

Thus, according to (39), (56), (57) and (58), we can arrive at $\mathbb{E}[\Delta\mathbf{Z}^{(k)}\tilde{\mathbf{s}}^{(k)}\tilde{\mathbf{s}}^{(k)\mathrm{H}}\Delta\mathbf{Z}^{(k)\mathrm{H}}]=(\xi-1)^2\mathbf{Z}_{\mathrm{p}}^{(k)}\mathbf{C}_{\mathbf{s}^{(k)}\mathbf{s}^{(k)}}\mathbf{Z}_{\mathrm{p}}^{(k)\mathrm{H}}+(1-\xi^2)\mathbf{C}_{\mathrm{p}}$.

References

[1] H. Que, J. Yang, C.-K. Wen, S. Xia, X. Li, and S. Jin, "Joint beam management and SLAM for mmWave communication systems," *IEEE Trans. Commun.*, 2023.

[2] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi, "Deep learning-based channel estimation for wideband hybrid mmWave massive MIMO," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3679–3693, 2023.

[3] X. Cao, B. Yang, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Han, H. V. Poor, and L. Hanzo, "Massive access of static and mobile users via reconfigurable intelligent surfaces: protocol design and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1253–1269, 2022.

[4] P. Zeng, D. Qiao, H. Qian, and Q. Wu, "Joint beamforming design for IRS aided multiuser MIMO with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 10 729–10 743, 2022.

[5] P. Zeng, D. Qiao, Q. Wu, and Y. Wu, "Throughput maximization for active intelligent reflecting surface-aided wireless powered communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 992–996, 2022.

[6] Q. Li, M. El-Hajjar, I. A. Hemadeh, A. Shojaeifard, A. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6367–6383, 2022.

[7] Q. Li, M. El-Hajjar, Y. Sun, I. Hemadeh, A. Shojaeifard, Y. Liu, and L. Hanzo, "Achievable rate analysis of the STAR-RIS aided NOMA uplink in the face of imperfect CSI and hardware impairments," *IEEE Trans. Commun.*, 2023.

[8] J. Yao, J. Xu, W. Xu, D. W. K. Ng, C. Yuen, and X. You, "Robust beamforming design for RIS-aided cell-free systems with CSI uncertainties and capacity-limited backhaul," *IEEE Trans. Commun.*, 2023.

[9] Q. Li, M. El-Hajjar, I. Hemadeh, D. Jagyasi, A. Shojaeifard, and L. Hanzo, "Performance analysis of active RIS-aided systems in the face of imperfect CSI and phase shift noise," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8140–8145, 2023.

[10] Q. Zhang, J. Liu, Z. Gao, Z. Li, Z. Peng, Z. Dong, and H. Xu, "Robust beamforming design for RIS-aided NOMA secure networks with transceiver hardware impairments," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3637–3649, 2023.

[11] M. Zhou, Y. Li, Y. Sun, and Z. Ding, "Outage performance of RIS-assisted V2I communications with inter-cell interference," *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 962–966, 2023.

[12] Y. Fang, Y. Tao, H. Ma, Y. Li, and M. Guizani, "Design of a reconfigurable intelligent surface-assisted FM-DCSK-SWIPT scheme with nonlinear energy harvesting model," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 1863–1877, 2023.

[13] Z. Liu, Z. Li, M. Wen, Y. Gong, and Y.-C. Wu, "STAR-RIS-aided mobile edge computing: Computation rate maximization with binary amplitude coefficients," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 4313–4327, 2023.

[14] W. Zhang, J. Xu, W. Xu, X. You, and W. Fu, "Worst-case design for RIS-aided over-the-air computation with imperfect CSI," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2136–2140, 2022.

[15] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE, 2019, pp. 4659–4663.

[16] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE, 2020, pp. 5000–5004.

[17] 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, 2020.

[18] 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. 16 202–16 206, 2020.

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

[20] W. Zhang, J. Xu, W. Xu, D. W. K. Ng, and H. Sun, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593–597, 2020.

[21] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, 2019.

[22] B. Zheng, C. You, and R. Zhang, "Fast channel estimation for IRS-assisted OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 580–584, 2020.

[23] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, 2020.

[24] 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, 2020.

[25] C. Huang, J. Xu, W. Zhang, W. Xu, and D. W. K. Ng, "Semi-blind channel estimation for RIS-assisted MISO systems using expectation maximization," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10 173–10 178, 2022.

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

[27] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, 2020.

[28] Y. Wei, M.-M. Zhao, M.-J. Zhao, and Y. Cai, "Channel estimation for IRS-aided multiuser communications with reduced error propagation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2725–2741, 2021.

[29] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.

[30] M. Hua, Q. Wu, D. W. K. Ng, J. Zhao, and L. Yang, "Intelligent reflecting surface-aided joint processing coordinated multipoint transmission," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1650–1665, 2020.

[31] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[32] X. Qian, M. Di Renzo, J. Liu, A. Kammoun, and M.-S. Alouini, "Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 77–81, 2020.

[33] M.-A. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 184–188, 2019.

[34] A. Papazafeiropoulos, C. Pan, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation and beamforming design," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2077–2092, 2021.

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

[36] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.

[37] 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, 2022.

[38] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Reconfigurable intelligent surface assisted MIMO symbiotic radio networks," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4832–4846, 2021.

[39] J. Wang, H. Wang, Y. Han, S. Jin, and X. Li, "Joint transmit beamforming and phase shift design for reconfigurable intelligent surface assisted MIMO systems," *IEEE Trans. Cogn. Commun. and Netw.*, vol. 7, no. 2, pp. 354–368, 2021.

[40] Z. Shi, H. Wang, Y. Fu, G. Yang, S. Ma, and F. Gao, "Outage analysis of reconfigurable intelligent surface aided MIMO communications with statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 823–839, 2021.

[41] Z. Abdullah, A. Papazafeiropoulos, S. Kisseleff, S. Chatzinotas, and B. Ottersten, "Impact of phase-noise and spatial correlation on double-RIS-assisted multiuser MISO networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1473–1477, 2022.

[42] Q. Li, M. El-Hajjar, I. Hemadeh, D. Jagyasi, A. Shojaeifard, E. Basar, and L. Hanzo, "The reconfigurable intelligent surface-aided multi-node IoT downlink: Beamforming design and performance analysis," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 6400–6414, 2022.

[43] L. Huikan, "Spatial correlation functions of fields in a reverberation chamber based on expansion of spherical bessel functions," *IEEE trans. electromagn. compat.*, vol. 48, no. 2, pp. 427–428, 2006.

[44] X. Chen, "Antenna correlation and its impact on multi-antenna system," *Progress In Electromagnetics Research B*, vol. 62, pp. 241–253, 2015.

[45] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. Di Renzo, A. L. Swindlehurst, R. Zhang, and A. Y. Zhang, "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, 2022.

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

[47] X. Zhang, *Matrix analysis and applications*. Cambridge University Press, 2017.

[48] T. Hillen, K. J. Painter, A. C. Swan, and A. D. Murtha, "Moments of von Mises and Fisher distributions and applications," *Mathematical Biosciences & Engineering*, vol. 14, no. 3, p. 673, 2017.