Two-Timescale Design for Reconfigurable Intelligent Surface-Aided Massive MIMO Systems with Imperfect CSI

Kangda Zhi, Cunhua Pan, Hong Ren, Kezhi Wang, Maged Elkashlan, Marco Di Renzo, *Fellow, IEEE*, Robert Schober, *Fellow, IEEE*, H. Vincent Poor, *Fellow, IEEE*, Jiangzhou Wang, *Fellow, IEEE* and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—This paper investigates the two-timescale transmission scheme for reconfigurable intelligent surface (RIS)-aided massive multiple-input multiple-output (MIMO) systems, where the beamforming at the base station (BS) is adapted to the rapidly-changing instantaneous channel state information (CSI), while the nearly-passive beamforming at the RIS is adapted to the slowly-changing statistical CSI. Specifically, we first consider a system model with spatially independent Rician fading channels, which leads to tractable expressions and offers analytical insights on the power scaling laws and on the impact of various system parameters. Then, we analyze a more general system model with spatially correlated Rician fading channels and consider the impact of electromagnetic interference (EMI) caused by any uncontrollable sources present in the considered environment. For both case studies, we apply the linear minimum mean square error (LMMSE) estimator to estimate the aggregated channel from the users to the BS, utilize the low-complexity maximal ratio combining (MRC) detector, and derive a closed-form expression for a lower bound of the achievable rate. Besides, an accelerated gradient ascent-based algorithm is proposed for solving the minimum user rate maximization problem. Numerical results show that, in the considered setup, the spatially independent model without EMI is sufficiently accurate when the inter-distance of the RIS elements is sufficiently large and the EMI is mild. In the presence of spatial correlation, we show that an RIS can better tailor the wireless environment. Furthermore, it is shown that deploying an RIS in a massive MIMO network brings significant gains when the RIS is deployed close to the cell-edge users. On the other hand, the gains obtained by the users distributed over a large area are shown to be modest.**

***Index Terms*—Reconfigurable intelligent surface (RIS), massive MIMO, two-timescale transmission scheme, channel estimation, spatial correlation, electromagnetic interference (EMI).**

Part of this work has been presented in the IEEE SPAWC, 2021 [1]. This work was supported in part by the National Natural Science Foundation of China (Grant No. 62101128, No. 62201137), Basic Research Project of Jiangsu Provincial Department of Science and Technology (Grant No. BK20210205), and the U.S. National Science Foundation under Grant CCF1908308. The work of M. Di Renzo was supported in part by the European Commission through the H2020 ARIADNE project under grant agreement number 871464 and through the H2020 RISE-6G project under grant agreement number 101017011. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). R. Schober's work was partly supported by the Federal Ministry of Education and Research of Germany under the programme of "Souveran. Digital. Vernetzt." joint project 6G-RIC (project identification number: PIN 16KISK023) and the German Science Foundation under project SCHO 831/15-1. The work of K. Zhi was supported by China Scholarship Council. *(Corresponding author: Cunhua Pan).*

K. Zhi, M Elkashlan are with the School of Electronic Engineering and Computer Science at Queen Mary University of London, London E1 4NS, U.K. (e-mail: k.zhi, maged.elkashlan@qmul.ac.uk).

C. Pan, H. Ren are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (e-mail: c.pan, hren@seu.edu.cn).

K. Wang is with Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH (e-mail: kezhi.wang@brunel.ac.uk).

M. Di Renzo is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 3 Rue Joliot-Curie, 91192 Gif-sur-Yvette, France. (marco.di-renzo@universite-paris-saclay.fr)

R. Schober is with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany (e-mail: robert.schober@fau.de).

H. V. Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

J. Wang is with the School of Engineering and Digital Arts, University of Kent, UK. (e-mail: J.Z.Wang@kent.ac.uk).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk.).

I. Introduction

As an emerging candidate for next-generation communication systems, reconfigurable intelligent surfaces (RISs), also termed intelligent reflecting surfaces (IRSs), have attracted significant interest from academia and industry [2], [3]. An RIS is a reconfigurable engineered surface that does not require active radio frequency (RF) chains, power amplifiers, and digital signal processing units, and is usually made of a large number of low cost and passive scattering elements that are coupled with simple low power electronic circuits. By intelligently tuning the phase shifts of the impinging waves with the aid of a controller, an RIS can constructively strengthen the desired signal or can deconstructively weaken the interference signals, which results in an appealing nearly-passive beamforming gain.

Compared with existing multi-antenna systems [4]–[9], it has been demonstrated that RIS-aided systems have the potential to achieve better performance in terms of cost and energy consumption [10]–[17]. Recently, RISs have been considered for being integrated into various communication scenarios, such as terahertz, sub-terahertz, and millimeter-wave systems [18], [19], simultaneous wireless information and power transfer (SWIPT) [20], unmanned aerial vehicle (UAV) communications [21], cell-free systems [22], physical-layer security [23]–[25], mobile edge computing (MEC) [26]–[28], device-to-device (D2D) communications [29], [30]. Furthermore, the effectiveness of RIS-aided systems in the pres-

ence of practical imperfections has been demonstrated in [31]–[34]. Specifically, relying on imperfect instantaneous channel state information (CSI), the robust transmission design of RISs was studied in [31], [32]. The authors of [33] studied the RIS beamforming design by considering transceiver hardware impairments. With the consideration of RF impairments and phase noises, the authors of [34] conducted a theoretical study on the fundamental tradeoffs between the spectral and energy efficiency of an RIS communication network. In addition, a valuable experimental investigation of RIS-assisted channels was carried out in [35].

While several benefits of RISs have been demonstrated in the above-mentioned contributions, most of them considered the design of the nearly-passive beamforming at the RIS under the assumption that the instantaneous CSI is estimated in each channel coherence interval. In practice, however, instantaneous CSI-based schemes face two challenges. The first one is the overhead for the acquisition of the instantaneous CSI. Due to the absence of power amplifiers, digital signal processing units, and radio frequency chains at the RISs, many authors proposed to estimate the cascaded user-RIS-BS channels instead of the separated user-RIS and RIS-BS channels [36], [37]. The pilot overhead of these channel estimation schemes is proportional to the number of RIS elements. However, an RIS generally consists of a large number of reflecting elements to ensure the desired coverage enhancement [38], which incurs in a prohibitively high pilot overhead. Secondly, in each channel coherence time interval, the BS needs to calculate the optimal beamforming coefficients for the RIS, and needs to send them back to the RIS controller via dedicated feedback links. For instantaneous CSI-based schemes, therefore, the beamforming calculation and information feedback need to be executed frequently in each channel coherence interval, which results in a high computational complexity, feedback overhead, and energy consumption.

To address these two practical challenges, recently, Han *et al.* [39] proposed a novel two-timescale based RIS scheme, which facilitates the deployment and operation of RIS-aided systems. This promising two-timescale scheme was further analyzed in recent research works [40]–[51]. In the two-timescale scheme, the BS beamforming is designed based on the instantaneous aggregated CSI, which includes the direct and RIS-reflected links. The dimension of this aggregated channel is the same as for conventional RIS-free systems, which is independent of the number of RIS elements. Hence, in the two-timescale scheme, the number of pilot signals needs to be only larger than the number of users, which significantly reduces the channel estimation overhead. More importantly, the two-timescale scheme aims to optimize the RISs only based on long-term statistical CSI, such as the locations and the angles of arrival and departure of the users with respect to the BS and the RIS, which vary much slower than the instantaneous CSI, for typical applications in the sub-6 GHz bands. The phase shifts of the RIS elements need to be updated only when the large-scale channel information changes. Compared with instantaneous CSI-based designs that need to update the phase shifts of the RIS elements in each channel coherence interval, therefore, RIS-aided designs based on statistical CSI can significantly reduce the computational complexity, feedback overhead and energy consumption.

In addition, massive MIMO technology has been identified as the cornerstone of the fifth generation (5G) and future communication systems [52], [53]. Massive MIMO exploits tens or hundreds of BS antennas to serve multiple users simultaneously. Due to the complexity of wireless propagation environments, e.g., the presence of large blocking objects, however, the signal power received at the end-users may be still too weak, and it may be insufficient to support emerging applications that entail high date rate requirements, such as virtual reality (VR) or augmented reality (AR). Inspired by the capability of RISs to customize the wireless propagation environment, a natural idea is to integrate them into massive MIMO systems. By constructing alternative transmission paths, it is envisioned that RIS-aided massive MIMO systems can achieve significant performance gains, especially when the direct links between the BS and the users are blocked by obstacles. In RIS-aided massive MIMO systems, the transmission scheme needs to be carefully designed, and the channel estimation overhead needs to be taken into account considering the large channel dimension. The application of instantaneous CSI-assisted schemes, in particular, may lead to a prohibitive complexity and overhead. Instead, due to the reduced channel estimation and feedback overhead, the two-timescale scheme is deemed more suitable for RIS-aided massive MIMO systems.

Even though RIS-aided massive MIMO systems have been investigated in some recent works [49], [50], [54], [55], three key issues are still not well understood. Firstly, it is crucial to identify the ultimate performance limits of RIS-aided massive MIMO systems based on the two-timescale scheme under imperfect CSI. In the presence of channel estimation errors, the impact of key system parameters, the achievable rate scaling law, and the power scaling law are unknown. To tackle these open problems, it is necessary to derive explicit information-theoretic analytical frameworks that provide guidelines for system design. Secondly, it is essential to adopt realistic channel models that account for line-of-sight (LoS) and non-LoS (NLoS) components, so that the impact of the LoS and the scattered power can be appropriately modeled and analyzed. This enables one to provide guidelines for the deployment of RISs. Thirdly, some unique and realistic characteristics need to be considered when analyzing RIS-aided systems, including the spatial correlation among the RIS elements and the electromagnetic interference (EMI). To date, the impact of spatial correlation and EMI have not been examined in RIS-aided massive MIMO systems based on the two-timescale scheme and in the presence of imperfect CSI. To be specific, due to the planar structure of the RIS, the channel spatial correlation among the RIS elements cannot be ignored [56]. To model the LoS and NLoS channel components and the spatial correlation among the RIS elements, the correlated Rician fading model is considered an appropriate choice. Also, due to the large aperture, an RIS may be subject to a large amount of EMI, which is generated by any uncontrollable external sources (e.g., the signals from adjacent cells and the natural background radiation) [57], [58]. Therefore, the EMI

re-radiated by a large RIS towards the intended receiver might deteriorate the channel estimation quality and reduce the end-to-end SINR, especially when the RIS is large and the useful signal power is weak. These three open research problems motivate the present research work.

In this paper, we analyze the uplink (UL) two-timescale transmission of an RIS-aided massive MIMO system that is subject to imperfect aggregated CSI. The Rician channel model is adopted to evaluate the impact of the LoS and NLoS channel components. To gain some initial design insights, we first analyze a channel model with spatially independent Rician fading, which admits tractable expressions of the achievable rate, and enables us to develop a comprehensive theoretical framework to evaluate the impact of critical system parameters and power scaling laws. Then, we generalize our analysis to a channel model with spatially correlated Rician fading and EMI. In this context, we focus our attention on the impact of spatial correlation and EMI on the achievable rate and the power scaling laws. Finally, we propose a gradient ascent method to solve the minimum user rate maximization problem based only on statistical CSI. The specific contributions of this paper are summarized as follows.

- To begin with, we consider the spatially independent Rician fading model. The aggregated channel is estimated by relying on the linear minimum mean square error (LMMSE) method and its performance in terms of mean square error (MSE) and normalized MSE (NMSE) is analyzed. Under the assumption of MRC detectors, we derive closed-form expressions for the use-and-then-forget (UatF) bound of the achievable rate. The derived results hold for an arbitrary number of BS antennas and RIS elements. Then, we analyze the impact of important system parameters, the asymptotic behavior of the rate, and the power scaling laws. We specialize our findings to the single-user case in order to obtain further engineering insights.
- Next, we consider a more general system model that includes spatial correlation at the RIS and the EMI captured by the RIS. Also in this case, we compute the LMMSE channel estimates and formulate the UatF bound of the achievable rate in a closed-form expression. Our analysis shows that the presence of spatial correlation provides the RIS with an enhanced capability of customizing the wireless environment. On the other hand, the presence of severe EMI may result in different power scaling laws.
- For both the spatially independent and spatially correlated channel models, we propose an accelerated gradient ascent-based algorithm to solve the minimum user rate maximization problem. We first apply a log-sum-exp approximation to obtain a smooth objective function. Then, we compute the gradient vectors with respect to the angle vectors. The performance loss in the projection is avoided since the objective function is periodic with the angles and the unit modulus constraint holds for all the angles. Besides, closed-form solutions are obtained in the special case of a single user.
- Numerical results validate the accuracy of analytical insights derived by neglecting the spatial correlation and EMI. In the presence of spatial correlation and EMI, the obtained numerical results show that similar trends hold when the spatial correlation and the EMI are moderate. Specifically, our numerical study reveals that (i) an RIS with a large number of elements may benefit from the presence of spatial correlation; (ii) in the presence of severe EMI, an RIS-aided system may not offer better performance than a conventional massive MIMO system; (iii) the integration of RISs in massive MIMO systems is especially beneficial when the RISs are deployed near the cell edge users.

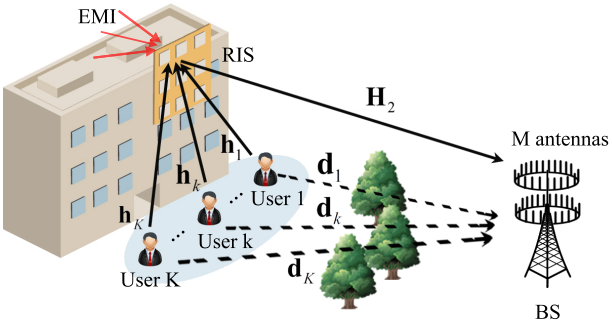

Fig. 1: An RIS-aided massive MIMO system.

The remainder of this paper is organized as follows. The performance analysis based on spatially independent channels without EMI is carried out in Sections II, III, and IV. Specifically, the system model is introduced in Section II, the LMMSE channel estimator is derived and analyzed in Section III, and a closed-form lower bound expression of the achievable rate is obtained in Section IV. The extension to spatially correlated channels in the presence of EMI is discussed in Section V. In Section VI, a gradient ascent-based algorithm for solving the minimum user rate maximization problem is introduced. Extensive numerical results are illustrated in Section VII and the conclusions are drawn in SectionVIII.

Notations: Vectors and matrices are denoted by boldface lower case and upper case letters, respectively. The transpose, conjugate, conjugate transpose, and inverse of matrix $\mathbf{X}$ are denoted by $\mathbf{X}^T$, $\mathbf{X}^*$, $\mathbf{X}^H$ and $\mathbf{X}^{-1}$, respectively. $[\mathbf{X}]_{m,n}$ denotes the (m,n)th entry of matrix $\mathbf{X}$. The real, imaginary, trace, expectation, and covariance operators are denoted by $\mathrm{Re}\{\cdot\}$, $\mathrm{Im}\{\cdot\}$, $\mathrm{Tr}\{\cdot\}$, $\mathbb{E}\{\cdot\}$, and $\mathrm{Cov}\{\cdot\}$, respectively. The l_2 norm of a vector and the absolute value of a complex number are denoted by $\|\cdot\|$ and $|\cdot|$, respectively. $\mathbb{C}^{M\times N}$ denotes the space of $M \times N$ complex matrices. $\mathbf{I}_M$ and $\mathbf{0}$ denote the $M \times M$ identity matrix and all-zero matrix with appropriate dimension, respectively. The operator mod returns the remainder after division, and $\lfloor x \rfloor$ denotes the nearest integer smaller than x. $\mathbf{x} \sim \mathcal{CN}(\bar{\mathbf{x}}, \mathbf{C})$ is a complex Gaussian distributed vector with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{C}$. $\mathcal{O}$ denotes the standard big-O notation. Besides, for ease of reference, the main symbols used in this work are listed in Table I.

TABLE I: List of Main Symbols

Symbol	Definition	Symbol	Definition
$M/N/K$	Number of BS antennas/RIS elements/users	p	Transmit power for each user
θ_n	Phase shift of the n-th RIS element	$\boldsymbol{\theta}$	Phase shift vector equal to $[\theta_1, \theta_2, \ldots, \theta_N]^T$
$\boldsymbol{c}$	Vector equal to $e^{j\boldsymbol{\theta}}$	$\boldsymbol{\Phi}$	RIS phase shifts matrix, $\boldsymbol{\Phi} = \text{diag}(\mathbf{c})$
σ^2/σ_e^2, ρ	Power of thermal noise/EMI, $\rho = \frac{\sigma_e^2}{\sigma^2}$	$\mathbf{x}/\mathbf{n}/\boldsymbol{v}$	Signal/noise/EMI vector
d_{ris}/d_{bs}	Element spacing of RIS/BS	λ	Wavelength
τ/τ_c	Lengths of pilot signal/coherence interval	$\mathbf{s}_k$, $\mathbf{S}$	User k's pilot sequence, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K]$
$\mathbf{N}/\mathbf{V}$	Noise/EMI vectors over τ time slots	γ_k	Pathloss of user k's direct link
α_k	Pathloss of user k-RIS link	β	Pathloss of RIS-BS link
δ	Rician factor of RIS-BS link	ε_k	Rician factor of user k-RIS link
$\mathbf{d}_k$, $\tilde{\mathbf{d}}_k$	User k-BS direct link, $\mathbf{d}_k = \sqrt{\gamma_k}\tilde{\mathbf{d}}_k$	$\mathbf{h}_k$, $\overline{\mathbf{h}}_k$, $\tilde{\mathbf{h}}_k$	User k-RIS link, comprised of $\overline{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$
$\mathbf{H}_2/\mathbf{H}_{c,2}$	RIS-BS link without/with correlation	$\tilde{\mathbf{H}}_2/\tilde{\mathbf{H}}_{c,2}$	NLoS part of $\mathbf{H}_2/\mathbf{H}_{c,2}$
$\mathbf{q}_k/\mathbf{q}_{c,k}$	Aggregated link without/with correlation	$\mathbf{Q}/\mathbf{Q}_c$	Matrix with the k-th column of $\mathbf{q}_k/\mathbf{q}_{c,k}$
$\hat{\mathbf{q}}_k/\hat{\mathbf{q}}_{c,k}$	Channel estimate of $\mathbf{q}_k/\mathbf{q}_{c,k}$	$\hat{\mathbf{Q}}/\hat{\mathbf{Q}}_c$	Matrix with the k-th column of $\hat{\mathbf{q}}_k/\hat{\mathbf{q}}_{c,k}$
$\mathbf{q}_k^1$-$\mathbf{q}_k^4$, $\underline{\mathbf{q}}_k$	Notations defined in (10)	$\hat{\mathbf{q}}_k^1$-$\hat{\mathbf{q}}_k^4$, $\underline{\hat{\mathbf{q}}}_k$	Notations defined in (17)
$\mathbf{y}/\mathbf{y}_c$	BS received signal without/with correlation	$\mathbf{r}/\mathbf{r}_c$	Decoded symbols from $\mathbf{y}/\mathbf{y}_c$
$\mathbf{Y}_P/\mathbf{Y}_{c,P}$	Received pilot signals at the BS	$\mathbf{y}_p^k/\mathbf{y}_{c,p}^k$	Observation vector without/with correlation
$\mathbf{a}_M(.)/\mathbf{a}_N(.)$	Array response vector for BS/RIS	E_u	A constant used in the power scaling laws
c_k, $\widehat{c}_k$	$c_k = \frac{\beta\alpha_k}{(\delta+1)(\varepsilon_k+1)}$, $\widehat{c}_k = \frac{\beta\alpha_k}{\delta+1}$	a_{k1} - a_{k4}	Notations defined in Lemma 1 and Theorem 1
e_{k1} - e_{k3}	Notations defined in Lemma 2	$f_k(\boldsymbol{\Phi})$	Scalar equal to $\mathbf{a}_N^H\boldsymbol{\Phi}\overline{\mathbf{h}}_k$
$\mathbf{A}_k$, $\mathbf{B}_k$	Matrices defined in Theorem 1	$\boldsymbol{\Upsilon}_k$	Matrix defined in Theorem 3
$\underline{R}_k/\underline{R}_{c,k}$	Rate of use k without/with correlation	$f(\boldsymbol{\theta})/f_c(\boldsymbol{\theta})$	Approximated minimum user rate
$\mathbf{R}_{ris}$,$\mathbf{R}_{emi}$	Spatial correlation matrices	$\boldsymbol{f}_d(.)$, $\mathbf{z}_k(.)$	Function defined in Lemma 4, 5
$f_{c,1}(\boldsymbol{\Phi})$, $f_{c,k,2}(\boldsymbol{\Phi})$ - $f_{c,k,7}(\boldsymbol{\Phi})$, $f_{c,ki,8}(\boldsymbol{\Phi})$ - $f_{c,ki,9}(\boldsymbol{\Phi})$		Scalar functions defined in (79)	
$\boldsymbol{f}'_{c,1}(\boldsymbol{\theta})$, $\boldsymbol{f}'_{c,k,2}(\boldsymbol{\theta})$ - $\boldsymbol{f}'_{c,k,7}(\boldsymbol{\theta})$, $\boldsymbol{f}'_{c,ki,8}(\boldsymbol{\theta})$ - $\boldsymbol{f}'_{c,ki,9}(\boldsymbol{\theta})$		Gradient vectors defined in Lemma 6	
E_k^{signal}, I_{ki}, E_k^{leak}, E_k^{noise}		Signal, interference, leakage, and noise in Theorem 2	
$E_{c,k}^{\text{signal}}$, $I_{c,ki}$, $E_{c,k}^{\text{leak}}$, $E_{c,k}^{\text{emi}}$, $E_{c,k}^{\text{noise}}$		Signal, interference, leakage, EMI and noise in Theorem 4	

II. System Model

To begin with, we consider an RIS-aided massive MIMO system under spatially uncorrelated channels and in the absence of EMI. These two aspects will be analyzed in Section V. Specifically, as illustrated in Fig. 1, we consider the UL transmission of an RIS-aided massive MIMO system, where an RIS is deployed in the proximity of K users to assist their UL transmissions to the BS. For convenience, we denote the set of users as $\mathcal{K} = \{1, 2, \ldots, K\}$. The BS is equipped with M active antennas, the RIS comprises N nearly-passive reflecting elements, and the K users are equipped with a single transmit antenna. The channels from user k, $k \in \mathcal{K}$ to the BS, from user k, $k \in \mathcal{K}$ to the RIS, and from the RIS to the BS are denoted by $\mathbf{d}_k \in \mathbb{C}^{M\times 1}$, $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$, and $\mathbf{H}_2 \in \mathbb{C}^{M\times N}$, respectively. Additionally, we define $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$ and $\mathbf{H}_1 = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$.

The RIS shapes the propagation environment by phase-shifting the impinging signals. Its phase shift matrix is denoted by $\boldsymbol{\Phi} = \text{diag}\left\{e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}\right\}$, where $\theta_n \in [0, 2\pi)$ represents the phase shift of the nth reflecting element. Based on these definitions, the cascaded user k-RIS-BS channel can be written as $\mathbf{g}_k = \mathbf{H}_2\boldsymbol{\Phi}\mathbf{h}_k$, and the cascaded channels of the K users are collected in the matrix $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K] = \mathbf{H}_2\boldsymbol{\Phi}\mathbf{H}_1 \in \mathbb{C}^{M\times K}$.

The K users transmit their data in the same UL time-frequency resource. For ease of exposition, let $\mathbf{Q} = \mathbf{G} + \mathbf{D} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K] \in \mathbb{C}^{M\times K}$ denote the aggregated instantaneous channel matrix from the users to the BS. Thereby, the signal vector received at the BS is given by

$$\mathbf{y} = \sqrt{p}\mathbf{Q}\mathbf{x} + \mathbf{n} = \sqrt{p}\sum\nolimits_{k=1}^{K}\mathbf{q}_k x_k + \mathbf{n}, \tag{1}$$

where p is the average transmit power of each user, $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T$ are the transmit symbols of the K users, and $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_M\right)$ denotes the noise vector.

The BS applies a low-complexity MRC receiver to detect the transmitted symbols. Before designing the MRC matrix, the channel $\mathbf{Q}$ has to be estimated at the BS. A standard LMMSE estimator is employed to obtain the estimated channel $\hat{\mathbf{Q}}$, as explained in the next section[1]. Relying on the channel estimate, the BS performs MRC by multiplying the received signal $\mathbf{y}$ with $\hat{\mathbf{Q}}^H$, as follows

$$\mathbf{r} = \hat{\mathbf{Q}}^H\mathbf{y} = \sqrt{p}\hat{\mathbf{Q}}^H\mathbf{Q}\mathbf{x} + \hat{\mathbf{Q}}^H\mathbf{n}. \tag{2}$$

Then, the kth element of the vector $\mathbf{r}$ can be expressed as

$$r_k = \sqrt{p}\hat{\mathbf{q}}_k^H\mathbf{q}_k x_k + \sqrt{p}\sum\nolimits_{i=1, i\neq k}^{K}\hat{\mathbf{q}}_k^H\mathbf{q}_i x_i + \hat{\mathbf{q}}_k^H\mathbf{n}, \quad k \in \mathcal{K}, \tag{3}$$

where $\hat{\mathbf{q}}_k$ is the kth column of $\hat{\mathbf{Q}}$.

[1]Given the LMMSE channel estimator, this work is focused on the possible benefits of deploying RISs in massive MIMO systems. It is meaningful to investigate other channel estimators (such as the least-squares and element-wise MMSE [53], [59], [60]) and to evaluate the trade-off between estimation quality and implementation complexity. The comparison between different channel estimation strategies is postponed to a future work.

A. Channel Model

Since the users may be located far away from the BS and a large number of environmental blocking objects (i.e., blockages such as trees, vehicles, buildings) may exist in the area of interest, the LoS path between the users and the BS could be blocked. As in [39], [47], [48], we adopt the Rayleigh fading model to describe the NLoS channel between the user k and the BS, as follows

$$\mathbf{d}_k = \sqrt{\gamma_k}\tilde{\mathbf{d}}_k, \quad k \in \mathcal{K}, \tag{4}$$

where γ_k denotes the distance-dependent path-loss, and $\tilde{\mathbf{d}}_k$ denotes the fast fading NLoS channel. The entries of $\tilde{\mathbf{d}}_k$ are independent and identically distributed (i.i.d.) complex Gaussian random variables, i.e., $\tilde{\mathbf{d}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$.

Considering that the RIS is often installed on the facades of high-rise buildings and it could be placed near the users, the channels between the users and the RIS have a high LoS probability. In addition, the RIS and the BS are usually deployed at some heights above the ground, which implies that LoS paths are likely to exist between the RIS and the BS. Therefore, as in [39], [47]–[50], we adopt the Rician fading model for the user-RIS and RIS-BS channels, as follows

$$\mathbf{h}_k = \sqrt{\frac{\alpha_k}{\varepsilon_k+1}}\left(\sqrt{\varepsilon_k}\,\overline{\mathbf{h}}_k + \tilde{\mathbf{h}}_k\right), k \in \mathcal{K}, \tag{5}$$

$$\mathbf{H}_2 = \sqrt{\frac{\beta}{\delta+1}}\left(\sqrt{\delta}\,\overline{\mathbf{H}}_2 + \tilde{\mathbf{H}}_2\right), \tag{6}$$

where α_k and β represent the path-loss coefficients, ε_k and δ are the Rician factors that account for the ratio of the LoS power to the NLoS power of the corresponding propagation paths. Furthermore, $\overline{\mathbf{h}}_k$ and $\overline{\mathbf{H}}_2$ denote the LoS components, whereas $\tilde{\mathbf{h}}_k$ and $\tilde{\mathbf{H}}_2$ represent the NLoS components. For the NLoS paths, the components of $\tilde{\mathbf{h}}_k$ and $\tilde{\mathbf{H}}_2$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. For the LoS paths, the uniform linear array (ULA) and uniform squared planar array (USPA) models are adopted for the BS and the RIS, respectively. Hence, $\overline{\mathbf{h}}_k$ and $\overline{\mathbf{H}}_2$ are, respectively, modelled as follows

$$\overline{\mathbf{h}}_k = \mathbf{a}_N\left(\varphi_{kr}^a, \varphi_{kr}^e\right), k \in \mathcal{K}, \tag{7}$$

$$\overline{\mathbf{H}}_2 = \mathbf{a}_M\left(\phi_r^a, \phi_r^e\right)\mathbf{a}_N^H\left(\varphi_t^a, \varphi_t^e\right), \tag{8}$$

where φ_{kr}^a (φ_{kr}^e) is the azimuth (elevation) angle of arrival (AoA) of the incident signal at the RIS from the user k, φ_t^a (φ_t^e) is the azimuth (elevation) angle of departure (AoD) reflected by the RIS towards the BS, and ϕ_r^a (ϕ_r^e) is the azimuth (elevation) AoA of the signal received at the BS from the RIS, respectively. Furthermore, $\mathbf{a}_X\left(\vartheta^a, \vartheta^e\right) \in \mathbb{C}^{X\times 1}$ denotes the array response vector, whose x-th entry is

$$\begin{aligned}
\left[\mathbf{a}_M\left(\vartheta^a, \vartheta^e\right)\right]_x &= \exp\left\{j2\pi\frac{d_{bs}}{\lambda}(x-1)\sin\vartheta^e\sin\vartheta^a\right\}, \\
\left[\mathbf{a}_N\left(\vartheta^a, \vartheta^e\right)\right]_x &= \exp\Bigg\{j2\pi\frac{d_{ris}}{\lambda}\left(\left\lfloor(x-1)/\sqrt{N}\right\rfloor\sin\vartheta^e\sin\vartheta^a\right. \\
&\quad \left.+\left((x-1) \bmod \sqrt{N}\right)\cos\vartheta^e\right)\Bigg\},
\end{aligned} \tag{9}$$

where d_{bs}, d_{ris}, and λ denote the BS antenna spacing, the RIS element spacing, and the wavelength, respectively.

To simplify the notation, in the sequel, we denote $\mathbf{a}_M\left(\phi_r^a, \phi_r^e\right)$ and $\mathbf{a}_N\left(\varphi_t^a, \varphi_t^e\right)$ simply by $\mathbf{a}_M$ and $\mathbf{a}_N$, respectively. Then, the aggregated channel from the user k to the BS can be expressed as

$$\begin{aligned}
\mathbf{q}_k &= \mathbf{g}_k + \mathbf{d}_k = \mathbf{H}_2\mathbf{\Phi}\mathbf{h}_k + \mathbf{d}_k \\
&= \underbrace{\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k}_{\mathbf{q}_k^1} + \underbrace{\sqrt{c_k\delta}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k}_{\mathbf{q}_k^2} \\
&+ \underbrace{\sqrt{c_k\varepsilon_k}\,\tilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k}_{\mathbf{q}_k^3} + \underbrace{\sqrt{c_k}\,\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k}_{\mathbf{q}_k^4} + \sqrt{\gamma_k}\,\tilde{\mathbf{d}}_k \\
&\triangleq \underline{\mathbf{q}}_k + \mathbf{d}_k,
\end{aligned} \tag{10}$$

where $c_k \triangleq \frac{\beta\alpha_k}{(\delta+1)(\varepsilon_k+1)}$, and $\underline{\mathbf{q}}_k = \sum_{\omega=1}^4 \mathbf{q}_k^\omega$. Note that $\underline{\mathbf{q}}_k$ and $\mathbf{d}_k$ are mutually independent.

III. Channel Estimation

In this section, we use the LMMSE method to obtain the estimated aggregated instantaneous channel $\hat{\mathbf{Q}}$. Specifically, the BS estimates the aggregated channel matrix $\mathbf{Q}$ based on some predefined pilot signals. Let τ_c and τ denote the length of the channel coherence interval and the number of time slots used for channel estimation, respectively, where τ is no smaller than K, i.e., $\tau \geq K$. In each channel coherence interval, the K users simultaneously transmit mutually orthogonal pilot sequences to the BS. The pilot sequence of user k is denoted by $\mathbf{s}_k \in \mathbb{C}^{\tau\times 1}$. By defining $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K]$, we have $\mathbf{S}^H\mathbf{S} = \mathbf{I}_K$. Then, the $M \times \tau$ pilot signals received at the BS can be written as

$$\mathbf{Y}_p = \sqrt{\tau p}\mathbf{Q}\mathbf{S}^H + \mathbf{N}, \tag{11}$$

where τp is the transmit pilot power, and $\mathbf{N}$ denotes the $M\times\tau$ noise matrix whose entries are i.i.d. complex Gaussian random variables with zero mean and variance σ^2. Multiplying (11) by $\frac{\mathbf{s}_k}{\sqrt{\tau p}}$ and exploiting the orthogonality of the pilot signals, the BS obtains the following observation vector for user k

$$\mathbf{y}_p^k = \frac{1}{\sqrt{\tau p}}\mathbf{Y}_p\mathbf{s}_k = \mathbf{q}_k + \frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k. \tag{12}$$

The optimal estimate of the k-th user's channel based on the observation vector $\mathbf{y}_p^k$ can be determined based on the MMSE criterion, which has been widely utilized in conventional massive MIMO systems [61]–[63]. In RIS-aided massive MIMO systems where Rician fading is considered for all RIS-aided channels, however, it is challenging to obtain the MMSE estimator. This is because the cascaded user-RIS-BS channel $\mathbf{G}$ in RIS-aided systems is not Gaussian distributed, but double Gaussian distributed [64]. To obtain closed-form channel estimates, as is needed to obtain useful design insights, we adopt the sub-optimal but tractable LMMSE estimator. This is because the LMMSE estimator only requires the knowledge of the first and second order statistics, and therefore it does not need to know the exact channel distributions. In the following lemma, we present the required statistics for the channel vector $\mathbf{q}_k$ and the observation vector $\mathbf{y}_p^k$.

Lemma 1 *For $k \in \mathcal{K}$, the mean vectors and covariance matrices that are needed to compute the LMMSE estimator are given by*

$$\mathbb{E}\left\{\mathbf{q}_k\right\} = \mathbb{E}\left\{\mathbf{y}_p^k\right\} = \sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k, \tag{13}$$

$$\begin{aligned}\mathrm{Cov}\left\{\mathbf{q}_k, \mathbf{y}_p^k\right\} &= \mathrm{Cov}\left\{\mathbf{y}_p^k, \mathbf{q}_k\right\} = \mathrm{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\} \\ &= a_{k1}\mathbf{a}_M\mathbf{a}_M^H + a_{k2}\mathbf{I}_M,\end{aligned} \tag{14}$$

$$\begin{aligned}\mathrm{Cov}\left\{\mathbf{y}_p^k, \mathbf{y}_p^k\right\} &= \mathrm{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\} + \frac{\sigma^2}{\tau p}\mathbf{I}_M \\ &= a_{k1}\mathbf{a}_M\mathbf{a}_M^H + \left(a_{k2} + \frac{\sigma^2}{\tau p}\right)\mathbf{I}_M,\end{aligned} \tag{15}$$

where $a_{k1} = Nc_k\delta$ and $a_{k2} = Nc_k\left(\varepsilon_k + 1\right) + \gamma_k$ are two auxiliary variables.

Proof: See Appendix B. ■

Theorem 1 *Using the observation vector $\mathbf{y}_p^k$, the LMMSE estimate $\hat{\mathbf{q}}_k$ of the channel vector $\mathbf{q}_k$ is given by (17) at the bottom of this page,*[2] *where*

$$\mathbf{A}_k = \mathbf{A}_k^H = a_{k3}\mathbf{a}_M\mathbf{a}_M^H + a_{k4}\mathbf{I}_M, \tag{18}$$

$$\mathbf{B}_k = \left(\mathbf{I}_M - \mathbf{A}_k\right)\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k, \tag{19}$$

$$a_{k3} = \frac{a_{k1}\frac{\sigma^2}{\tau p}}{\left(a_{k2} + \frac{\sigma^2}{\tau p}\right)\left(a_{k2} + \frac{\sigma^2}{\tau p} + Ma_{k1}\right)}, \tag{20}$$

$$a_{k4} = \frac{a_{k2}}{a_{k2} + \frac{\sigma^2}{\tau p}}, \tag{21}$$

and the NMSE of the estimate of $\mathbf{q}_k$ is

$$\begin{aligned}\mathrm{NMSE}_k &= \frac{\mathrm{Tr}\left\{\mathrm{Cov}\left\{\mathbf{q}_k - \hat{\mathbf{q}}_k, \mathbf{q}_k - \hat{\mathbf{q}}_k\right\}\right\}}{\mathrm{Tr}\left\{\mathrm{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\}\right\}} \\ &= \frac{\frac{\sigma^2}{\tau p}\left(Ma_{k1}a_{k2} + a_{k2}^2 + \left(a_{k1} + a_{k2}\right)\frac{\sigma^2}{\tau p}\right)}{\left(a_{k2} + \frac{\sigma^2}{\tau p}\right)\left(a_{k2} + \frac{\sigma^2}{\tau p} + Ma_{k1}\right)\left(a_{k1} + a_{k2}\right)}.\end{aligned} \tag{22}$$

Proof: See Appendix C. ■

As evident from Theorem 1, we only estimate the aggregated channel matrix $\mathbf{Q} \in \mathbb{C}^{M\times K}$ including the reflected and direct channels, which has the same dimension as the user-BS channel matrix in conventional massive MIMO systems. Therefore, we only require that the length of the pilot sequences is no smaller than the number of users, i.e., $\tau \geq K$.

[2]Note that $\hat{\mathbf{q}}_k^1$ and $\mathbf{q}_k^1$ are identical due to the unbiased estimation. However, we define two symbols in order to simplify the analytical formulation and make the derivations easier to understand (see (141) and (146) for example).

Compared to methods that estimate the MN individual channels in RIS-aided communications [36], [37], the proposed method has a lower overhead and computational complexity.

Remark 1 *When $c_k = 0, \forall k$, i.e., the RIS-assisted channels are absent, we have $a_{k1} = 0$, $a_{k2} = \gamma_k$, $a_{k3} = 0$, $a_{k4} = \frac{\gamma_k}{\gamma_k + \frac{\sigma^2}{\tau p}}$ and $\mathbf{B}_k = \mathbf{0}$. In this case, the estimate in (16) reduces to $\hat{\mathbf{q}}_k = \frac{\gamma_k}{\gamma_k + \frac{\sigma^2}{\tau p}}\mathbf{y}_p^k$ and the MSE matrix in (133) reduces to $\mathbf{MSE}_k = \frac{\gamma_k\frac{\sigma^2}{\tau p}}{\gamma_k + \frac{\sigma^2}{\tau p}}\mathbf{I}_M$, which, as expected, is the same as the MSE in conventional massive MIMO systems [62]. If the RIS channels only have the LoS components, i.e., $\delta, \varepsilon_k \to \infty, \forall k$, we also obtain $a_{k1} \to 0$ and $a_{k2} \to \gamma_k$. In this case, the MSE matrix in (133) is again the same as that in conventional massive MIMO systems. This is because the LoS channels are deterministic and known, and, thus, they do not introduce additional estimation errors.*

Corollary 1 *In the low pilot power-to-noise ratio regime, high pilot power-to-noise ratio regime, and large N regime, the asymptotic NMSE is, respectively, given by*

$$\lim\nolimits_{\frac{\sigma^2}{\tau p}\to\infty} \mathrm{NMSE}_k \to 1, \tag{23}$$

$$\lim\nolimits_{\frac{\sigma^2}{\tau p}\to 0} \mathrm{NMSE}_k \to 0, \tag{24}$$

$$\lim\nolimits_{N\to\infty} \mathrm{NMSE}_k \to 0. \tag{25}$$

Besides, assume that the power p is scaled proportionally to $p = E_u/N$, where E_u denotes a constant. As $N \to \infty$, we have

$$\lim\nolimits_{p=\frac{E_u}{N}, N\to\infty} \mathrm{NMSE}_k < 1. \tag{26}$$

Proof: When $\frac{\sigma^2}{\tau p} \to \infty$ or $N \to \infty$, by selecting the dominant terms in (22), which scale with $\left(\frac{\sigma^2}{\tau p}\right)^2$ or N^3, we arrive at (23) and (25), respectively. Substituting $\frac{\sigma^2}{\tau p} = 0$ into (22), its numerator reduces to zero, which leads to (24). Replacing the power p in (22) with $p = E_u/N$, as $N \to \infty$, we can readily find that all the dominant terms in the numerator are present in the denominator as well, which results in (26). We omit the specific limit of (26) since it is a complex expression but is simple to compute. ■

It is worth noting that NMSE values between 0 (i.e., perfect estimation) and 1 (i.e., using the mean value of the variable as the estimate) quantify the relative estimation error [53]. In conventional massive MIMO systems, a common method for reducing the NMSE is to increase the length of the pilot sequence τ. In RIS-aided massive MIMO systems, Corollary 1 indicates that increasing the number of RIS elements N can

$$\hat{\mathbf{q}}_k = \mathbf{A}_k\mathbf{y}_p^k + \mathbf{B}_k \tag{16}$$

$$= \underbrace{\underbrace{\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k}_{\hat{\mathbf{q}}_k^1} + \underbrace{\left(Ma_{k3} + a_{k4}\right)\sqrt{c_k\delta}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k}_{\hat{\mathbf{q}}_k^2} + \underbrace{\sqrt{c_k\varepsilon_k}\mathbf{A}_k\tilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k}_{\hat{\mathbf{q}}_k^3} + \underbrace{\sqrt{c_k}\mathbf{A}_k\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k}_{\hat{\mathbf{q}}_k^4}}_{\underline{\hat{\mathbf{q}}}_k} + \sqrt{\gamma_k}\mathbf{A}_k\tilde{\mathbf{d}}_k + \frac{1}{\sqrt{\tau p}}\mathbf{A}_k\mathbf{N}\mathbf{s}_k, \tag{17}$$

play a similar role as increasing τ. Therefore, increasing the number of RIS elements not only helps improve the system rate, but it also helps reduce the NMSE. Additionally, (26) reveals that an RIS equipped with a large number of reflecting elements N can help the NMSE converge to a limit lower than one, even for low pilot powers.

To better understand the impact of increasing N for channel estimation, we present the following asymptotic results.

Corollary 2 *When $\tau \to \infty$, we have $\hat{\mathbf{q}}_k \to \mathbf{q}_k$, which implies $\mathbf{e}_k \to \mathbf{0}$ and therefore the channel estimation is perfect. When $N \to \infty$, by contrast, we have*

$$\hat{\mathbf{q}}_k \to \mathbf{q}_k + \frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k, \tag{27}$$

$$\mathbf{e}_k = \mathbf{q}_k - \hat{\mathbf{q}}_k \to \frac{-1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k, \tag{28}$$

$$\mathbf{MSE}_k = \mathbb{E}\left\{\mathbf{e}_k\mathbf{e}_k^H\right\} \to \frac{\sigma^2}{\tau p}\mathbf{I}_M. \tag{29}$$

Proof: When $\tau \to \infty$ or $N \to \infty$, based on Theorem 1, we have $a_{k3} \to 0$, $a_{k4} \to 1$, and $Ma_{k3} + a_{k4} \to 1$, which yields $\mathbf{A}_k \to \mathbf{I}_M$. If $\tau \to \infty$, we further get $\frac{1}{\sqrt{\tau p}} \to 0$, which completes the proof. ■

Although the NMSE converges to zero as $N \to \infty$ (see (25)), Corollary 2 shows that, in contrast to increasing τ, the MSE of the LMMSE estimator converges to a non-zero constant as $N \to \infty$. If we estimate the channel $\mathbf{q}_k$ based on the least-squares (LS) estimator [53, (3.35)], it is interesting to note that we obtain the same results as in (27) and (29). In general, the LS estimator, which does not exploit any prior channel statistics, has worse estimation performance (higher MSE) than the LMMSE estimator [37], [53], [59]. Therefore, Corollary 2 indicates that the MSE performance of the LMMSE estimation converges towards an upper bound, which is the MSE performance of the LS estimation, as $N \to \infty$. This result will be validated in Section VII.

Corollary 3 *When the RIS-BS channel reduces to the Rayleigh channel (i.e., $\delta = 0$), the estimated channel vector, MSE, and NMSE, respectively, simplify to*

$$\hat{\mathbf{q}}_k = a_{k4}\mathbf{I}_M\mathbf{y}_p^k = \frac{N\beta\alpha_k + \gamma_k}{N\beta\alpha_k + \gamma_k + \frac{\sigma^2}{\tau p}}\mathbf{y}_p^k, \tag{30}$$

$$\mathbf{MSE}_k = \frac{\left(N\beta\alpha_k + \gamma_k\right)\frac{\sigma^2}{\tau p}}{N\beta\alpha_k + \gamma_k + \frac{\sigma^2}{\tau p}}\mathbf{I}_M, \tag{31}$$

$$\mathrm{NMSE}_k = \frac{\frac{\sigma^2}{\tau p}}{N\beta\alpha_k + \gamma_k + \frac{\sigma^2}{\tau p}}. \tag{32}$$

Proof: When $\delta = 0$, we have $a_{k1} = 0$, $a_{k2} = N\beta\alpha_k + \gamma_k$, $a_{k3} = 0$, $a_{k4} = \frac{N\beta\alpha_k + \gamma_k}{N\beta\alpha_k + \gamma_k + \frac{\sigma^2}{\tau p}}$ and $\mathbf{B}_k = \mathbf{0}$. The proof follows by inserting these results in Theorem 1 and (133). ■

Corollary 3 corresponds to a scenario where a large number of scatterers exist nearby the RIS and the BS, and the LoS path between the RIS and the BS is negligible. Therefore, the RIS-BS channel is dominated by the NLoS paths. In this case, both the MSE and NMSE have simple analytical expressions, which help us better understand the conclusions drawn in Corollary 1 and Corollary 2. It is apparent that the MSE (represented by the trace of $\mathbf{MSE}_k$ in (31)) and the NMSE (represented by NMSE_k in (32)) are decreasing functions of the pilot power τp. As a function of N, on the other hand, the MSE is an increasing function, while the NMSE is a decreasing function. When $N \to \infty$, we have $\mathbf{MSE}_k \to \frac{\sigma^2}{\tau p}\mathbf{I}_M$ but $\mathrm{NMSE}_k \to 0$. Note that we can obtain the MSE and NMSE for conventional massive MIMO systems by setting $N = 0$ in (31) and (32). Therefore, the obtained result implies that the MSE of RIS-aided massive MIMO systems is worse than the MSE of massive MIMO systems without RISs, while the NMSE of RIS-aided massive MIMO systems is better than the NMSE of massive MIMO systems without RISs. The reason is that an RIS introduces N additional paths to the system, but the pilot length τ does not increase correspondingly, which increases the estimation error. However, the presence of an RIS results in better channel gains, which help decrease the normalized error.

Furthermore, if we reduce the power as $p = E_u/N$, as $N \to \infty$, the NMSE in (32) converges to a limit less than one, as follows

$$\lim_{\delta=0,\, p=\frac{E_u}{N},\, N\to\infty} \mathrm{NMSE}_k \to \frac{\sigma^2}{\tau E_u \beta\alpha_k + \sigma^2} < 1. \tag{33}$$

IV. Analysis of the Achievable Rate

Based on the channel estimates provided in Theorem 1, closed-form expressions for a lower bound of the achievable rate are derived and analyzed in this section[3]. In Section VI, the obtained analytical expressions are utilized for optimizing the phase shifts of the RIS based on statistical CSI.

A. Derivation of the Rate

As in [59], [65]–[67], we utilize the so called UatF bound, which is a tractable lower bound, to characterize the ergodic rate of RIS-aided massive MIMO systems. First, we rewrite r_k in (3) as

$$r_k = \underbrace{\sqrt{p}\,\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}x_k}_{\text{Desired signal}} + \underbrace{\sqrt{p}\left(\hat{\mathbf{q}}_k^H\mathbf{q}_k - \mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}\right)x_k}_{\text{Signal leakage}} + \underbrace{\sqrt{p}\sum\nolimits_{i=1,i\neq k}^{K}\hat{\mathbf{q}}_k^H\mathbf{q}_i x_i}_{\text{Multi-user interference}} + \underbrace{\hat{\mathbf{q}}_k^H\mathbf{n}}_{\text{Noise}}. \tag{34}$$

Then, we formulate the lower bound of the k-th user's ergodic rate as $\underline{R}_k = \frac{\tau_c - \tau}{\tau_c}\log_2\left(1 + \mathrm{SINR}_k\right)$, where the pre-log factor $\frac{\tau_c - \tau}{\tau_c}$ represents the rate loss that originates from the pilot overhead, and the SINR is given in (35) at the bottom of the next page.

To simplify the expression of $\underline{R}_k$, we define three auxiliary variables e_{k1}, e_{k2}, and e_{k3}. These variables capture the performance degradation due to the imperfect knowledge of the CSI.

[3]To avoid verbose expressions, "lower bound of the achievable rate" is replaced with "achievable rate" in the rest of this paper. It is, however, implied that we compute a lower bound.

Lemma 2 *For $k \in \mathcal{K}$, we have* $\mathrm{Tr}\{\mathbf{A}_k\} = Me_{k1}$, $\mathbf{A}_k\overline{\mathbf{H}}_2 = e_{k2}\overline{\mathbf{H}}_2$ *and* $\mathrm{Tr}\{\mathbf{A}_k\mathbf{A}_k\} = Me_{k3}$, *where*

$$e_{k1} \triangleq a_{k3} + a_{k4}, \tag{36}$$

$$e_{k2} \triangleq Ma_{k3} + a_{k4}, \tag{37}$$

$$e_{k3} \triangleq Ma_{k3}^2 + 2a_{k3}a_{k4} + a_{k4}^2. \tag{38}$$

Furthermore, e_{k1}, e_{k2} and e_{k3} are bounded in $[0, 1]$. When $\tau p \to \infty$ or $N \to \infty$, we have $e_{k1}, e_{k2}, e_{k3} \to 1$. When $\tau p \to 0$, by contrast, we have $e_{k1}, e_{k2}, e_{k3} \to 0$.

Proof: See Appendix D. ■

In the following theorem, we derive a closed-form expression for the achievable rate.

Theorem 2 *A lower bound for the ergodic rate of the k-th user is given by*[4]

$$\underline{R}_k = \tau^o \log_2\left(1 + \mathrm{SINR}_k\right),$$

$$\mathrm{SINR}_k = \frac{pE_k^{\mathrm{signal}}\left(\mathbf{\Phi}\right)}{pE_k^{\mathrm{leak}}\left(\mathbf{\Phi}\right) + p\sum\limits_{i=1, i\neq k}^{K} I_{ki}\left(\mathbf{\Phi}\right) + \sigma^2 E_k^{\mathrm{noise}}\left(\mathbf{\Phi}\right)}, \tag{39}$$

where $\tau^o = \frac{\tau_c - \tau}{\tau_c}$, $E_k^{\mathrm{signal}}\left(\mathbf{\Phi}\right) = \left\{E_k^{\mathrm{noise}}\left(\mathbf{\Phi}\right)\right\}^2$. The expressions of $E_k^{\mathrm{noise}}\left(\mathbf{\Phi}\right)$, $E_k^{\mathrm{leak}}\left(\mathbf{\Phi}\right)$, and $I_{ki}\left(\mathbf{\Phi}\right)$ are given in (40), (41), and (42) at the bottom of this and the next pages, respectively, with

$$f_k(\mathbf{\Phi}) \triangleq \mathbf{a}_N^H \mathbf{\Phi} \overline{\mathbf{h}}_k = \sum\nolimits_{n=1}^{N} e^{j\left(\zeta_n^k + \theta_n\right)}, \tag{43}$$

$$\zeta_n^k = 2\pi\frac{d}{\lambda}\Big(\lfloor (n-1)/\sqrt{N}\rfloor\left(\sin\varphi_{kr}^e \sin\varphi_{kr}^a - \sin\varphi_t^e \sin\varphi_t^a\right) + ((n-1) \bmod \sqrt{N})\left(\cos\varphi_{kr}^e - \cos\varphi_t^e\right)\Big). \tag{44}$$

Proof: See Appendix E. ■

The closed-form expression in Theorem 2 does not involve the calculation of inverse matrices and the numerical computation of integrals. In contrast to time-consuming Monte Carlo simulations, the evaluation of the rate based on Theorem 2 has a low computational complexity even if M and N are large numbers, as usually is in RIS-aided massive MIMO systems. Besides, Theorem 2 only relies on statistical CSI. Therefore, by using the analytical expression of the rate in (39) as an objective function for system design, we are able to optimize the phase shifts of the RIS only based on long-term statistical CSI. For clarity and analytical tractability, the statistical CSI is assumed to be perfectly known [40], [53], [54]. In practice, due to the user mobility, there may exist location and angular estimation errors based on, e.g., GPS (Global Positioning System) information, which could result in some performance loss for the design of receiver at the BS and passive beamforming at the RIS. The impact of imperfect statistical CSI can be analyzed by averaging the angular estimation error in the expression of the achievable rate similar to [68]. This analysis is interesting and is left to a future research work.

By comparing the formulation in Theorem 2 with that given in [49, Theorem 1], it can be seen that the impact of imperfect CSI is completely characterized by the parameters e_{k1}, e_{k2}, e_{k3} and $\frac{\sigma^2}{\tau p}$. In the perfect CSI scenario, we have $\tau \to \infty$, which leads to $e_{k1} = e_{k2} = e_{k3} = 1$ and $\frac{\sigma^2}{\tau p} = 0$. Based on Theorem 2, we can analyze the performance of RIS-aided massive MIMO systems for arbitrary system parameters. Even though the obtained analytical expressions may look cumbersome at the first sight, they provide clear insights in terms of the key system parameters M, N, and $f_k(\mathbf{\Phi})$, $\forall k$. For example, since the interference term I_{ki} scales as $\mathcal{O}(M^2)$, we infer that RIS-aided massive MIMO systems suffer from stronger multi-user interference than conventional massive MIMO systems. In the following, we provide a comprehensive analysis of RIS-aided massive MIMO systems, including the asymptotic behavior of the rate for large values of M and N, the power scaling laws, and the impact of the Rician factors. To this end, we begin with a useful lemma.

Lemma 3
- *If $N = 1$, for arbitrary $\mathbf{\Phi}$, we have $|f_k(\mathbf{\Phi})| = 1$ in (43).*
- *If $N > 1$, by optimizing $\mathbf{\Phi}$, the range of values $0 \leq |f_k(\mathbf{\Phi})| \leq N$ is achievable in (43).*
- *If we configure the phase shifts of the RIS to achieve $|f_k(\mathbf{\Phi})| = N$, unless the user $i, i \neq k$, has the same azimuth and elevation AoA as the user k, the function $|f_i(\mathbf{\Phi})|$ in (43) is bounded when $N \to \infty$.*
- *Unless the user $i, i \neq k$, has the same azimuth and elevation AoA as the user k, the term $\left|\overline{\mathbf{h}}_k^H \overline{\mathbf{h}}_i\right|^2$ is bounded when $N \to \infty$.*

Proof: See Appendix F. ■

B. Multi-user Case

In this section, we consider the general multi-user scenario, i.e., $K > 1$. Since any two users are unlikely to be in the same location, we assume that the azimuth and elevation AoA

[4]The phase shift matrix $\mathbf{\Phi}$ is assumed to be fixed when deriving the achievable rate. After obtaining the achievable rate, we will design $\mathbf{\Phi}$ so that the derived rate is optimized.

$$\mathrm{SINR}_k = \frac{p\left|\mathbb{E}\left\{\hat{\mathbf{q}}_k^H \mathbf{q}_k\right\}\right|^2}{p\left(\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H \mathbf{q}_k\right|^2\right\} - \left|\mathbb{E}\left\{\hat{\mathbf{q}}_k^H \mathbf{q}_k\right\}\right|^2\right) + p\sum\limits_{i=1, i\neq k}^{K}\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H \mathbf{q}_i\right|^2\right\} + \sigma^2\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_k\right\|^2\right\}}. \tag{35}$$

$$E_k^{\mathrm{noise}}\left(\mathbf{\Phi}\right) = M\Big\{\left|f_k(\mathbf{\Phi})\right|^2 c_k\delta\varepsilon_k + Nc_k\delta e_{k2} + \left(Nc_k\left(\varepsilon_k + 1\right) + \gamma_k\right)e_{k1}\Big\}, \tag{40}$$

of any two users are different, i.e., $(\varphi_{kr}^a, \varphi_{kr}^e) \neq (\varphi_{ir}^a, \varphi_{ir}^e)$. To begin with, we investigate the asymptotic behavior of the rate in (39) for large values of M and N.

Remark 2 *From Theorem 2, we observe that, as a function of M, $E_k^{\text{signal}}(\mathbf{\Phi})$, $E_k^{\text{leak}}(\mathbf{\Phi})$ and $I_{ki}(\mathbf{\Phi})$ behave asymptotically as $\mathcal{O}(M^2)$. Therefore, the rate $\underline{R}_k$ converges to a finite limit when $M \to \infty$. If, on the other hand, we align the phase shifts of the RIS for maximizing the intended signal for the user k, i.e., we set $|f_k(\mathbf{\Phi})| = N$, then we have $\underline{R}_k \to \infty$ for user k, and $\underline{R}_i \to 0$ for the other users $i \neq k$ as $N \to \infty$, based on Lemma 3. In a multi-user scenario, this implies that it is necessary to enforce some fairness requirements among the users when designing the phase shifts of the RIS.*

Next, we study the power scaling laws of RIS-aided massive MIMO systems with different Rician factors. Specifically, the Rician factor characterizes the fading severity of the environment and the richness of scatterers in the environment. The smaller the Rician factor, the larger the number of scatterers in the environment. If the Rician factor is zero, we retrieve the Rayleigh fading channel as a special case in which only the NLoS components exist. If the Rician factor tends to infinity, the channel is deterministic and is characterized only by the LoS component. It is worth mentioning that, under the assumption of imperfect CSI, decreasing the transmit power p results in a reduction of the power used for both the data and pilot signals.

We analyze several scenarios for the RIS-BS and user-RIS channels. For ease of exposition, we summarize the obtained power scaling laws as a function of M and N in Table II. Specifically, the following notations are used. "Imperfect CSI" and "Perfect CSI" are referred to the power scaling laws obtained for imperfect and perfect CSI, respectively. By setting $e_{k1} = e_{k2} = e_{k3} = 1$ and $\frac{\sigma^2}{\tau p} = 0$, which are obtained when $\tau \to \infty$, the imperfect CSI setup reduces to the perfect CSI setup. The notation "(Rician, Rician)" means that the RIS-BS channel and all the user-RIS channels are Rician distributed, i.e., $\delta > 0$ and $\varepsilon_k > 0, \forall k$. Similarly, the notation "(Rician, Rayleigh)" means that the RIS-BS channel is Rician distributed and all the user-RIS channels are Rayleigh distributed, i.e., $\delta > 0$ and $\varepsilon_k = 0, \forall k$. The notations "$1/M$", "$1/\sqrt{M}$" and "$1/N$" imply that the rate tends to a non-zero value if the transmit power scales proportionally to $1/M$, $1/\sqrt{M}$ and $1/N$, respectively. We mention, for completeness, that the readers interested in the power scaling laws as a function of M in conventional massive MIMO systems without RISs may refer to [61] and [62]. Besides, we note that the rate does not depend on the RIS phase shift matrix $\mathbf{\Phi}$ if $\delta = 0$ or $\varepsilon_k = 0, \forall k$, which will be proved in Corollary 5. In the following, we mainly consider the proof for the imperfect CSI case, since the perfect CSI setup can be obtained in a similar manner, by setting $e_{k1} = e_{k2} = e_{k3} = 1$ and $\frac{\sigma^2}{\tau p} = 0$.

Corollary 4 *("$1/M$" for "(Rician, Rician)" and "(Rician, Rayleigh)") Assume that the transmit power p is scaled as $p = E_u/M$. For $M \to \infty$, the rate of user k, $k \in \mathcal{K}$, is lower bounded by (45) shown at the bottom of the next page, where*

$$E_k^{\text{leak}}(\mathbf{\Phi}) = N|f_k(\mathbf{\Phi})|^2 c_k^2 \delta^2 \varepsilon_k \left(e_{k2}^2 + 1\right) + \frac{\sigma^2}{\tau E_u}|f_k(\mathbf{\Phi})|^2 c_k \delta \varepsilon_k e_{k2}^2 + N^2 c_k^2 \delta^2 e_{k2}^2 + \frac{\sigma^2}{\tau E_u} N c_k \delta e_{k2}^2, \tag{46}$$

$$I_{ki}(\mathbf{\Phi}) = |f_k(\mathbf{\Phi})|^2 |f_i(\mathbf{\Phi})|^2 c_k c_i \delta^2 \varepsilon_k \varepsilon_i + N|f_k(\mathbf{\Phi})|^2 c_k c_i \delta^2 \varepsilon_k + |f_i(\mathbf{\Phi})|^2 c_i \delta \varepsilon_i e_{k2}^2 \left(N c_k \delta + \frac{\sigma^2}{\tau E_u}\right) + N^2 c_k c_i \delta^2 e_{k2}^2 + N \frac{\sigma^2}{\tau E_u} c_i \delta e_{k2}^2, \tag{47}$$

$$e_{k2} = \frac{N c_k \delta}{\frac{\sigma^2}{\tau E_u} + N c_k \delta}. \tag{48}$$

Proof: If $p = E_u/M$ and $M \to \infty$, we have $e_{k1} \to 0$, $e_{k3} \to 0$, and e_{k2} tends to (48). The proof is completed by

$$\begin{aligned} E_k^{\text{leak}}(\mathbf{\Phi}) = {} & M|f_k(\mathbf{\Phi})|^2 c_k^2 \delta \varepsilon_k \Big\{ N\left(M\delta + \varepsilon_k + 1\right)\left(e_{k2}^2 + 1\right) + 2\left(M e_{k1} + e_{k2}\right)\left(e_{k2} + 1\right) \Big\} \\ & + M|f_k(\mathbf{\Phi})|^2 c_k \delta \varepsilon_k \left\{ \gamma_k + \left(\gamma_k + \frac{\sigma^2}{\tau p}\right) e_{k2}^2 \right\} \\ & + M^2 N^2 c_k^2 \delta^2 e_{k2}^2 + M N^2 c_k^2 \left\{ 2\delta\left(\varepsilon_k + 1\right) e_{k2}^2 + \left(\varepsilon_k + 1\right)^2 e_{k3} \right\} + M^2 N c_k^2 \left\{ \left(2\varepsilon_k + 1\right) e_{k1}^2 + 2\delta e_{k1} e_{k2} \right\} \\ & + MNc_k \left\{ c_k \left(2\delta e_{k2}^2 + \left(2\varepsilon_k + 1\right) e_{k3}\right) + \left(2\gamma_k + \frac{\sigma^2}{\tau p}\right)\left(\delta e_{k2}^2 + \left(\varepsilon_k + 1\right) e_{k3}\right) \right\} + M\gamma_k \left(\gamma_k + \frac{\sigma^2}{\tau p}\right) e_{k3}, \end{aligned} \tag{41}$$

$$\begin{aligned} I_{ki}(\mathbf{\Phi}) = {} & M^2 |f_k(\mathbf{\Phi})|^2 |f_i(\mathbf{\Phi})|^2 c_k c_i \delta^2 \varepsilon_k \varepsilon_i + M|f_k(\mathbf{\Phi})|^2 c_k \delta \varepsilon_k \left\{ c_i \left(MN\delta + N\varepsilon_i + N + 2Me_{k1}\right) + \gamma_i \right\} \\ & + M|f_i(\mathbf{\Phi})|^2 c_i \delta \varepsilon_i \left\{ c_k e_{k2} \left(MN\delta e_{k2} + N\varepsilon_k e_{k2} + N e_{k2} + 2M e_{k1}\right) + \left(\gamma_k + \tfrac{\sigma^2}{\tau p}\right) e_{k2}^2 \right\} \\ & + M^2 N^2 c_k c_i \delta^2 e_{k2}^2 + MN^2 c_k c_i \left\{ \delta\left(\varepsilon_k + \varepsilon_i + 2\right) e_{k2}^2 + \left(\varepsilon_k + 1\right)\left(\varepsilon_i + 1\right) e_{k3} \right\} \\ & + M^2 N c_k c_i e_{k1} \left\{ \left(\varepsilon_k + \varepsilon_i + 1\right) e_{k1} + 2\delta e_{k2} \right\} + M^2 c_k c_i \varepsilon_k \varepsilon_i e_{k1} \left(\left| \overline{\mathbf{h}}_k^H \overline{\mathbf{h}}_i \right|^2 e_{k1} + 2\delta \operatorname{Re}\left\{ f_k^H(\mathbf{\Phi}) f_i(\mathbf{\Phi}) \overline{\mathbf{h}}_i^H \overline{\mathbf{h}}_k \right\} \right) \\ & + MN \left\{ \left(\gamma_k + \tfrac{\sigma^2}{\tau p}\right) c_i \left(\delta e_{k2}^2 + \left(\varepsilon_i + 1\right) e_{k3}\right) + \gamma_i c_k \left(\delta e_{k2}^2 + \left(\varepsilon_k + 1\right) e_{k3}\right) \right\} + M\gamma_i \left(\gamma_k + \tfrac{\sigma^2}{\tau p}\right) e_{k3}, \end{aligned} \tag{42}$$

TABLE II: Power scaling laws in the multi-user case.

		(RIS-BS channel, user-RIS channels)			
		(Rician, Rician)	(Rician, Rayleigh)	(Rayleigh, Rician)	(Rayleigh, Rayleigh)
Imperfect CSI	M	$1/M$	$1/M$	$1/\sqrt{M}$	$1/\sqrt{M}$
	N	\	$1/N$		
Perfect CSI	M	$1/M$			
	N	\	$1/N$		

substituting $p = E_u/M$ into Theorem 2 and retaining the non-zero terms whose asymptotic behavior is $\mathcal{O}(M)$. ■

For a massive number of antennas, Corollary 4 shows that the rate of all the users tends to a non-zero value when the transmit power scales as $p = E_u/M$. From (45), we evince that the rate $\underline{R}_k$ is still non-zero if $\varepsilon_k = 0, \forall k$, i.e., all the user-RIS channels are Rayleigh distributed. This proves the power scaling law "$1/M$" for the "(Rician, Rayleigh)" setup in Table II. However, the rate $\underline{R}_k$ in (45) reduces to zero if $c_k = 0$ or $\delta = 0$, i.e., the RIS-aided channels are absent or the RIS-BS channel is Rayleigh distributed. This indicates that the power scaling law "$1/M$" does not hold for these two case studies. Specifically, the considered system degenerates to an RIS-free massive MIMO system with Rayleigh fading if $c_k = 0, \forall k$. In this case, it has been proven that the rate can maintain a non-zero value when the power scales as $p = E_u/\sqrt{M}$ [62, (37)]. As for the power scaling law for $\delta = 0$, we first provide an analytical expression of the rate when $\delta = 0$.

Corollary 5 *If the RIS-BS channel is Rayleigh distributed ($\delta = 0$), the rate of user k, $k \in \mathcal{K}$, is lower bounded by*

$$\underline{R}_k^{(\mathrm{NL}_1)} = \tau^o \log_2\left(1 + \frac{pE_k^{\mathrm{signal}}}{pE_k^{\mathrm{leak}} + p\sum_{i=1,i\neq k}^{K} I_{ki} + \sigma^2 E_k^{\mathrm{noise}}}\right), \tag{49}$$

where

$$E_k^{\mathrm{signal}} = M\left(Nc_k\left(\varepsilon_k+1\right)+\gamma_k\right)^2 e_{k1}, \tag{50}$$

$$E_k^{\mathrm{noise}} = Nc_k\left(\varepsilon_k+1\right)+\gamma_k, \tag{51}$$

$$\begin{aligned} E_k^{\mathrm{leak}} &= N^2 c_k^2\left(\varepsilon_k+1\right)^2 e_{k1} + MNc_k^2\left(2\varepsilon_k+1\right)e_{k1} \\ &\quad + Nc_k\left\{c_k\left(2\varepsilon_k+1\right)+\left(2\gamma_k+\frac{\sigma^2}{\tau p}\right)\left(\varepsilon_k+1\right)\right\}e_{k1} \\ &\quad + \gamma_k\left(\gamma_k+\frac{\sigma^2}{\tau p}\right)e_{k1}, \end{aligned} \tag{52}$$

$$\begin{aligned} I_{ki} &= N^2 c_k c_i\left(\varepsilon_k+1\right)\left(\varepsilon_i+1\right)e_{k1} \\ &\quad + MNc_k c_i(\varepsilon_k+\varepsilon_i+1)\,e_{k1} + Mc_k c_i\varepsilon_k\varepsilon_i\left|\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_i\right|^2 e_{k1} \\ &\quad + N\left\{\left(\gamma_k+\frac{\sigma^2}{\tau p}\right)c_i\left(\varepsilon_i+1\right)+\gamma_i c_k\left(\varepsilon_k+1\right)\right\}e_{k1} \\ &\quad + \gamma_i\left(\gamma_k+\frac{\sigma^2}{\tau p}\right)e_{k1}, \end{aligned} \tag{53}$$

and

$$e_{k1} = \frac{N\beta\alpha_k+\gamma_k}{N\beta\alpha_k+\gamma_k+\frac{\sigma^2}{\tau p}}. \tag{54}$$

Proof: When $\delta = 0$, we have $a_{k1} = 0$, $a_{k2} = N\beta\alpha_k + \gamma_k$, and $a_{k3} = 0$. Thus, we obtain $e_{k3} = e_{k1}^2$, where $e_{k1} = a_{k4}$ is given in (54). Substituting $\delta = 0$ into Theorem 2 and using $e_{k3} = e_{k1}^2$, the proof follows with the aid of some algebraic simplifications. ■

It is observed that the rate in Corollary 5 does not depend on $\mathbf{\Phi}$. Therefore, in a fully NLoS RIS-BS channel, any RIS phase shift matrix results in the same ergodic rate. This is because the RIS phase shift matrix $\mathbf{\Phi}$ is a unitary matrix and the entries of the NLoS channel $\tilde{\mathbf{H}}_2$ are Gaussian distributed. Therefore, $\tilde{\mathbf{H}}_2\mathbf{\Phi}$ has the same statistical properties as $\tilde{\mathbf{H}}_2$. Likewise, there is no need to design the RIS phase shifts if all the user-RIS links are fully NLoS. This conclusion is apparent from (39) by setting $\varepsilon_k = 0, \forall k$.

By analyzing the dominant terms of (49) when $M, N \to \infty$, we evince that the rate increases without bound for all the users. This implies that fairness requirements among the users are implicitly guaranteed in this special case. As $N \to \infty$, specifically, the dominant terms in (49) scale asymptotically as $\mathcal{O}\left(N^2\right)$, and the rate converges to

$$\underline{R}_k^{(\mathrm{NL}_1)} \to \tau^o \log_2\left(1+\frac{M\alpha_k}{\sum_{i=1}^K \alpha_i}\right), \text{ as } N \to \infty, \tag{55}$$

$$= \tau^o \log_2\left(1+M/K\right), \text{ if } \alpha_1 = \ldots = \alpha_K. \tag{56}$$

From (55), we evince that the SINR, $\frac{M\alpha_k}{\sum_{i=1}^K \alpha_i}$, does not depend on the pilot power τp and it increases linearly with M. Therefore, good performance can be obtained if $\delta = 0$ and $N \to \infty$.

$$\underline{R}_k \to \tau^o \log_2\left(1+\frac{E_u c_k^2\delta^2\left(\left|f_k(\mathbf{\Phi})\right|^2\varepsilon_k + Ne_{k2}\right)^2}{E_u E_k^{\mathrm{leak}}\left(\mathbf{\Phi}\right) + E_u\sum\limits_{i=1,i\neq k}^{K} I_{ki}\left(\mathbf{\Phi}\right) + \sigma^2 c_k\delta\left(\left|f_k(\mathbf{\Phi})\right|^2\varepsilon_k + Ne_{k2}\right)}\right), \tag{45}$$

With the aid of Corollary 5, we investigate, in the following corollaries, the power scaling laws as a function of M and N when $\delta = 0$.

Corollary 6 *("$1/\sqrt{M}$" for "(Rayleigh, Rician)" and "(Rayleigh, Rayleigh)") If the RIS-BS channel is Rayleigh distributed ($\delta = 0$), and the power is scaled as $p = E_u/\sqrt{M}$ with $M \to \infty$, the rate of user k, $k \in \mathcal{K}$ tends to $\underline{R}_k^{(\mathrm{NL}_1)} \to \tau^o \log_2(1 + \mathrm{SINR}_k)$, where the effective SINR is given by (57) at the bottom of this page.*

Proof: First, we substitute $p = E_u/\sqrt{M}$ into Corollary 5 and ignore the terms that tend to zero as $M \to \infty$. Then, we divide the numerator and denominator of the SINR by $\frac{N\beta\alpha_k+\gamma_k}{\sigma^2}$. This yields (57) and the proof is completed. ■

From (57), we evince that the numerator of the SINR scales with $\mathcal{O}(N^2)$, but the denominator of the SINR only scales with $\mathcal{O}(N)$. Therefore, Corollary 6 indicates that the rate scales logarithmically with N if $p = E_u/\sqrt{M}$ and $M \to \infty$, which is a promising result for RIS-aided massive MIMO systems. Besides, it is worth noting that (57) reduces to the same expression as in [62, Eq. (37)] when $c_k = 0, \forall k$.

Corollary 7 *("$1/N$" for "(Rayleigh, Rician)" and "(Rayleigh, Rayleigh)") If the RIS-BS channel is Rayleigh distributed ($\delta = 0$) and the power is scaled as $p = E_u/N$ with $N \to \infty$, the rate of user k, $k \in \mathcal{K}$, is lower bounded by (58) at the bottom of this page.*

Proof: First, we substitute $p = E_u/N$ into Corollary 5. When $N \to \infty$, we have $e_{k1} \to \frac{\beta\alpha_k}{\beta\alpha_k + \frac{\sigma^2}{\tau E_u}}$. Then, we remove the non-dominant terms that do not scale as $\mathcal{O}(N)$. By noting that $c_k(\varepsilon_k + 1) = \beta\alpha_k, \forall k$, and dividing the numerator and denominator of the SINR by $\beta\alpha_k$, we obtain (58). This completes the proof. ■

Corollary 7 sheds some interesting insights. Firstly, we note that Corollary 6 has unveiled that the transmit power p can only be reduced proportionally to $1/\sqrt{M}$, while maintaining a non-zero rate, when $\delta = 0$. Corollary 7, on the other hand, proves that the transmit power can be reduced proportionally to $1/N$, while maintaining a non-zero rate, when $\delta = 0$. This reveals the positive role of deploying RISs in massive MIMO systems. Secondly, the obtained power scaling law does not depend on the Rician factors of the user-RIS links, i.e., $\varepsilon_k, \forall k$. This implies that the rate in (58) is the same for LoS-only and NLoS-only user-RIS channels. Thirdly, in (58), the desired signal term in (58) scales as $\mathcal{O}(M)$ and the interference term scales as $\mathcal{O}(1)$. As a result, the rate scales logarithmically with the number of BS antennas. When the number of antennas is large, the power of the interference is relatively small compared with the power of the desired signal, and then a good rate can be guaranteed with the setup stated in Corollary 7. Therefore, a rich-scattering environment between the RIS and the BS ($\delta = 0$) is beneficial in RIS-aided massive MIMO systems, since it can provide sufficient spatial multiplexing gains and help mitigate the multi-user interference. Finally, (58) unveils that, if the users are all located at the same distance from the RIS, i.e., $\alpha_1 = \ldots = \alpha_K$, they all achieve the same rate. Therefore, fairness requirements can be guaranteed in this special case.

Corollary 7 sheds light on the achievable rate when the RIS-BS channel is Rayleigh distributed ($\delta = 0$). In the next corollary, we analyze the opposite scenario in which the user-RIS channels are Rayleigh distributed ($\varepsilon_k = 0, \forall k$).

Corollary 8 *("$1/N$" for "(Rician, Rayleigh)") Assume $\delta > 0$. If the user-RIS channels are Rayleigh distributed ($\varepsilon_k = 0, \forall k$) and the power is scaled as $p = E_u/N$ with $N \to \infty$, the rate of user k, $k \in \mathcal{K}$, is lower bounded by (59) at the bottom of the next page, where*

$$E_k^{\mathrm{leak}} + \sum_{i=1, i\neq k}^{K} I_{ki} = \sum_{i=1}^{K} c_i \left\{ M c_k \delta^2 e_{k2}^2 + c_k \left(2\delta e_{k2}^2 + e_{k3}\right) + \frac{\sigma^2}{\tau E_u}\left(\delta e_{k2}^2 + e_{k3}\right)\right\}, \tag{60}$$

$$a_{k3} = \frac{c_k \delta \frac{\sigma^2}{\tau E_u}}{\left(c_k + \frac{\sigma^2}{\tau E_u}\right)\left(c_k + \frac{\sigma^2}{\tau E_u} + M c_k \delta\right)}, \tag{61}$$

$$a_{k4} = \frac{c_k}{c_k + \frac{\sigma^2}{\tau E_u}}. \tag{62}$$

Proof: It follows from Theorem 2 by setting $\varepsilon_k = 0, \forall k$ and $p = E_u/N$, and by keeping only the dominant terms for $N \to \infty$. ■

Corollary 8 characterizes the achievable rate when the user-RIS channels are characterized by rich scattering. The obtained performance trends are different from those unveiled in Corollary 7 (i.e., the RIS-BS channel characterized by rich scattering). In contrast to Corollary 7, in particular, both the desired signal and the interference in (59) scale as $\mathcal{O}(M)$. As a result, if the user-RIS channels are Rayleigh distributed, the rate in (59) is still bounded from above even if the number of BS antennas is very large. Besides, it is not hard to prove that

$$\mathrm{SINR}_k = \frac{\tau E_u^2 \left(N c_k (\varepsilon_k + 1) + \gamma_k\right)^2}{\tau E_u^2 N c_k^2 (2\varepsilon_k + 1) + \sum_{i=1, i\neq k}^{K} \tau E_u^2 c_k c_i \left\{ N(\varepsilon_k + \varepsilon_i + 1) + \varepsilon_k \varepsilon_i \left| \overline{\mathbf{h}}_k^H \overline{\mathbf{h}}_i \right|^2 \right\} + \sigma^4}. \tag{57}$$

$$\underline{R}_k^{(\mathrm{NL}_1)} \to \tau^o \log_2 \left(1 + \frac{E_u M \beta \alpha_k}{\sum_{i=1}^{K} \left(E_u \beta \alpha_i + \frac{\alpha_i}{\alpha_k} \frac{\sigma^2}{\tau} \right) + \sigma^2 \left(1 + \frac{\sigma^2}{\tau E_u \beta \alpha_k} \right)} \right). \tag{58}$$

the rate in (59) reduces to the same expression as (58) if we set $\delta = 0$. This result confirms the conclusion in Corollary 7 that the scaling law unrelated to the Rician factor ε_k if $\delta = 0$.

From Corollary 7 and Corollary 8, we conclude that a small value of δ is beneficial in terms of power scaling laws. This is because a small δ corresponds to a high-rank RIS-BS channel, which provides sufficient spatial diversity for multi-user communications. It is known that, due to the product pathloss law that characterizes RIS-aided links in the far-field region, it is better to deploy an RIS either close to the BS or close to the users [69], [70]. Our analysis reveals that the best deployment for an RIS depends on the spatial diversity provided by the RIS-BS channel. When the RIS is deployed close to the users, δ could be small since the Rician factor commonly decreases with the communication distance [71]. Therefore, placing the RIS close to the users is still a good choice since this results in a high rank RIS-BS channel. If the RIS is deployed near the BS, δ could be large and the RIS-BS channel could become rank-deficient. In this context, other methods are needed to improve the rank of the channel such as introducing some artificial scatterers between the BS and the RIS or placing the RIS very close to the BS [46].

C. Single-user Case

In this subsection, we analyze the power scaling laws in the special case with only one user, i.e., $K = 1$. Without loss of generality, the user is referred to as user k. Since no other user exists, the rate can be obtained from Theorem 2 by ignoring the multi-user interference term, i.e., by setting $I_{ki}(\mathbf{\Phi}) = 0$. For analytical tractability, we further assume that the number of RIS elements is large. In this scenario (single-user and large N), it can be proved that the optimal phase shift matrix that maximizes the rate corresponds to the condition $|f_k(\mathbf{\Phi})| = N$. This statement is formally proved in the next section (Theorem 5).

Therefore, by setting $I_{ki} = 0$ and $|f_k(\mathbf{\Phi})| = N$ in Theorem 2, we obtain that the power of the desired signal scales as $\mathcal{O}\left(M^2N^4\right)$, the power of the signal leakage scales as $\mathcal{O}\left(M^2N^3\right)$, and the power of the noise term scales as $\mathcal{O}\left(MN^2\right)$. Therefore, the rate is bounded for $M \to \infty$, but it can grow without bound for $N \to \infty$. For ease of exposition, similar to the multi-user case, we summarize the obtained power scaling laws in Table III. In the following, we report the proofs only for some (those that lead to insightful design guidelines) system setups that are summarized in Table III. The proof of each case study can, in fact, be obtained by using analytical steps similar to the multi-user case. Finally, we mention that the power scaling laws in the single-user case with perfect CSI can be derived readily based on [39, Eq. (17)].

Corollary 9 *Consider a single-user system with $|f_k(\mathbf{\Phi})| = N$. If the transmit power is scaled as $p = E_u/\left(MN^2\right)$ with $M, N \to \infty$, the rate is lower bounded by*

$$\underline{R}_k \to \tau^o \log_2\left(1 + \frac{E_u}{\sigma^2}\frac{\beta\alpha_k\delta\varepsilon_k}{(\delta+1)(\varepsilon_k+1)}\right). \tag{63}$$

If the transmit power is scaled as $p = E_u/N^2$ with $N \to \infty$, the rate is lower bounded by

$$\underline{R}_k \to \tau^o \log_2\left(1 + \frac{E_u}{\sigma^2}Mc_k\delta\varepsilon_k\right). \tag{64}$$

Proof: Let us set $p = E_u/\left(MN^2\right)$, $|f_k(\mathbf{\Phi})| = N$ and $I_{ki} = 0$ in Theorem 2. The rate in (63) follows because $e_{k1}, e_{k2}, e_{k3} \to 0$ and by retaining the dominant terms that scale as $\mathcal{O}\left(MN^2\right)$ for $M, N \to \infty$. Similarly, let us set $p = E_u/N^2$, $|f_k(\mathbf{\Phi})| = N$ and $I_{ki} = 0$ in Theorem 2. The rate in (64) follows by retaining the dominant terms that scale as $\mathcal{O}\left(N^2\right)$ for $N \to \infty$. ■

The SNRs in (63) and (64) do not depend on τ, and except for a pre-log scaling factor, the same SNR as for perfect CSI-based systems can be obtained from [39, Eq. (17)]. We evince, therefore, that $\tau = K = 1$ is the optimal pilot length based on (63) and (64). Therefore, the overhead for channel estimation is relatively low. Furthermore, the rates in (63) and (64) are increasing functions with the Rician factors δ and ε_k, which unveils that LoS-dominated environments are favorable for RIS-aided single-user systems. If both $\delta \to \infty$ and $\varepsilon_k \to \infty$, (63) and (64) are maximized. On the contrary, if $\delta = 0$ or $\varepsilon_k = 0$, we observe that (63) and (64) tend to zero. This implies that the power scaling law $1/N^2$ does not hold anymore. In these two cases, the transmit power can be scaled only proportionally to $1/N$ to maintain a non-zero rate when $N \to \infty$. Mathematically, the corresponding power scaling laws can be proved from Corollary 7 and Corollary 8 by setting the multi-user interference to zero. As an example, the case study for $\delta = 0$ is analyzed in the following corollary.

Corollary 10 *Consider a single-user system with $\delta = 0$. If the transmit power is scaled as $p = E_u/N$ with $N \to \infty$, the rate is lower bounded by*

$$\underline{R}_k^{(\mathrm{NL}_1)} \to \tau^o \log_2\left(1 + \frac{E_u M\beta\alpha_k}{E_u\beta\alpha_k + \frac{\sigma^2}{\tau} + \sigma^2\left(1 + \frac{\sigma^2}{\tau E_u\beta\alpha_k}\right)}\right). \tag{65}$$

As τ increases, the denominator of the SNR of (65) decreases. Therefore, the SNR of (65) is an increasing function of τ. Therefore, $\tau = 1$ is not guaranteed to be optimal in a rich-scattering environment ($\delta = 0$), and a relatively large number of pilot signals may be needed. Thus, Corollary 10 also unveils that LoS environments are favorable for RIS-aided single-user systems.

$$\underline{R}_k^{(\mathrm{NL}_2)} \to \tau^o \log_2\left(1 + \frac{E_u M c_k^2\left(\delta e_{k2} + e_{k1}\right)^2}{E_u\left(E_k^{\mathrm{leak}} + \sum_{i=1, i\neq k}^{K} I_{ki}\right) + \sigma^2 c_k\left(\delta e_{k2} + e_{k1}\right)}\right), \tag{59}$$

TABLE III: Power scaling laws in the single-user case.

		(RIS-BS channel, user-RIS channel)			
		(Rician, Rician)	(Rician, Rayleigh)	(Rayleigh, Rician)	(Rayleigh, Rayleigh)
Imperfect CSI	M	$1/M$	$1/M$	$1/\sqrt{M}$	$1/\sqrt{M}$
	N	$1/N^2$	$1/N$		
Perfect CSI	M	$1/M$			
	N	$1/N^2$	$1/N$		

V. Extension to Correlated Channels with EMI

In this section, we generalize the analysis in Section IV by considering the impact of spatial correlation at the RIS and the presence of EMI. We ignore the spatial correlation at the BS, since a ULA with half-wavelength antenna spacing is assumed at the BS. On the other hand, the RIS is usually modeled as a UPA and the spatial correlation cannot be ignored in general [56]. Specifically, this section has two objectives: (1) to analyze the impact of spatial correlation and EMI in RIS-aided massive MIMO systems; and (2) to study to what extent the findings obtained in Section IV hold in the presence of spatial correlation and EMI.

A. Channel Model with Spatial Correlation

The evaluation conducted in Section IV indicates that it is appropriate to place the RIS near the users. In this scenario, the LoS components dominate the user-RIS channels, and therefore the Rician factor ε_k is relatively large. For ease of analysis and brevity, this section is focused on the scenario where the user-RIS channels are characterized only by the LoS component (i.e., $\varepsilon_k \to \infty$, $\forall k$).[5] In the following, we present the generalized system model in the presence of spatial correlation and EMI. For the avoidance of doubt, the subscript c is utilized to indicate the existence of spatial correlation.

In the presence of spatial correlation and EMI, the received signal at the BS is

$$\mathbf{y}_c = \sqrt{p}\mathbf{Q}_c\mathbf{x} + \mathbf{H}_{c,2}\mathbf{\Phi}\boldsymbol{v} + \mathbf{n}, \tag{66}$$

where $\boldsymbol{v} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_e^2\mathbf{R}_{emi}\right)$ denotes the EMI received at the RIS whose spatial correlation matrix is $\mathbf{R}_{emi}$. Specifically, the EMI is reflected by the RIS and reaches the BS through the RIS-BS channel $\mathbf{H}_{c,2}$ resulting in the term $\mathbf{H}_{c,2}\mathbf{\Phi}\boldsymbol{v}$ in (66). The matrix $\mathbf{Q}_c = [\mathbf{q}_{c,1}, \mathbf{q}_{c,2}, \ldots, \mathbf{q}_{c,K}] \in \mathbb{C}^{M\times K}$ denotes the spatially correlated aggregated channel from the K users to the BS, where $\mathbf{q}_{c,k} = \mathbf{H}_{c,2}\mathbf{\Phi}\mathbf{h}_k + \mathbf{d}_k$ is the aggregated channel of user k. The user k-RIS channel $\mathbf{h}_k$ and the RIS-BS channel $\mathbf{H}_{c,2}$ are, respectively, given by

$$\mathbf{h}_k = \sqrt{\alpha_k}\overline{\mathbf{h}}_k, \tag{67}$$

$$\mathbf{H}_{c,2} = \sqrt{\frac{\beta}{\delta+1}}\left(\sqrt{\delta}\overline{\mathbf{H}}_2 + \tilde{\mathbf{H}}_{c,2}\right), \tag{68}$$

where $\tilde{\mathbf{H}}_{c,2} = \tilde{\mathbf{H}}_2\mathbf{R}_{ris}^{1/2}$ and $\mathbf{R}_{ris}$ denotes the spatial correlation matrix of the NLoS channel components. Assuming an isotropic scattering environment for $\boldsymbol{v}$ and $\tilde{\mathbf{H}}_{c,2}$, the spatial correlation matrices $\mathbf{R}_{emi}$ and $\mathbf{R}_{ris}$ at the RIS can be formulated as $\mathbf{R}_{emi} = \mathbf{R}_{ris} = \mathbf{R}$ with [56], [57]

$$[\mathbf{R}]_{a,b} = \operatorname{sinc}\left(\frac{2\left\|\mathbf{u}_a - \mathbf{u}_b\right\|}{\lambda}\right), 1 \le a, b \le N, \tag{69}$$

where $\|\mathbf{u}_a - \mathbf{u}_b\|$ denotes the distance between the a-th and b-th elements of the RIS, which depends on the RIS element spacing d_{ris}. Since $\operatorname{sinc}(\cdot)$ is an even function, we have $\mathbf{R} = \mathbf{R}^H$. For ease of writing, we define $\widehat{c}_k = \frac{\alpha_k\beta}{\delta+1}$. Therefore, based on (67) and (68), the spatially correlated aggregated channel of user k can be expressed as

$$\begin{aligned}\mathbf{q}_{c,k} &= \mathbf{H}_{c,2}\mathbf{\Phi}\mathbf{h}_k + \mathbf{d}_k \\ &= \sqrt{\widehat{c}_k\delta}\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k + \sqrt{\widehat{c}_k}\tilde{\mathbf{H}}_{c,2}\mathbf{\Phi}\overline{\mathbf{h}}_k + \sqrt{\gamma_k}\tilde{\mathbf{d}}_k.\end{aligned} \tag{70}$$

B. Channel Estimation

In this section, we derive the LMMSE channel estimate $\hat{\mathbf{q}}_{c,k}$ for the aggregated channel of the k-th user. During the channel estimation phase, the BS receives the $M \times \tau$ pilot signal as follows

$$\mathbf{Y}_{c,P} = \sqrt{\tau p}\mathbf{Q}_c\mathbf{S}^H + \mathbf{H}_{c,2}\mathbf{\Phi}\mathbf{V} + \mathbf{N}, \tag{71}$$

where $\mathbf{V} = \mathbf{R}_{emi}^{1/2}\tilde{\mathbf{V}} \in \mathbb{C}^{N\times\tau}$ and each element of $\tilde{\mathbf{V}} \in \mathbb{C}^{N\times\tau}$ is independently distributed as $\mathcal{CN}\left(0, \sigma_e^2\right)$. After correlating $\mathbf{Y}_{c,P}$ with $\mathbf{s}_k$, the observation vector for the channel of the k-th user $\mathbf{q}_{c,k}$ is given by

$$\mathbf{y}_{c,p}^k = \frac{1}{\sqrt{\tau p}}\mathbf{Y}_{c,P}\mathbf{s}_k = \mathbf{q}_{c,k} + \frac{\left(\mathbf{H}_{c,2}\mathbf{\Phi}\mathbf{V} + \mathbf{N}\right)\mathbf{s}_k}{\sqrt{\tau p}}. \tag{72}$$

Theorem 3 *Based on $\mathbf{y}_{c,p}^k$, the LMMSE channel estimate for $\mathbf{q}_{c,k}$ is given by*

$$\begin{aligned}\hat{\mathbf{q}}_{c,k} = &\sqrt{\widehat{c}_k\delta}\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k + \sqrt{\widehat{c}_k}\mathbf{\Upsilon}_k\tilde{\mathbf{H}}_{c,2}\mathbf{\Phi}\overline{\mathbf{h}}_k + \sqrt{\gamma_k}\mathbf{\Upsilon}_k\tilde{\mathbf{d}}_k \\ &+ \frac{\mathbf{\Upsilon}_k\mathbf{H}_{c,2}\mathbf{\Phi}\mathbf{V}\mathbf{s}_k}{\sqrt{\tau p}} + \frac{\mathbf{\Upsilon}_k\mathbf{N}\mathbf{s}_k}{\sqrt{\tau p}},\end{aligned} \tag{73}$$

where $\mathbf{\Upsilon}_k$ is shown at the bottom of the next page.

Proof: See Appendix G. ■

Besides, applying [72, Eq. (12.21)], the MSE matrix is given by

$$\mathbf{MSE}_{c,k} = \left(\widehat{c}_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k + \gamma_k\right)\left(\mathbf{I}_M - \mathbf{\Upsilon}_k\right). \tag{75}$$

Equation (75) embodies the impact of spatial correlation and EMI on channel estimation. By the direct inspection of (75), we can make the following observations. On the one hand, the MSE may be degraded by the EMI power σ_e^2 through the

[5] Many research works have revealed that the rate is marginally affected by the Rician factor when it is greater than 10 [39], [40]. Thus, the considered scenario serves as a tractable approximation when ε_k can be assumed to be relatively large. The analysis of arbitrary values for the Rician fading factors $\varepsilon_k, \forall k$, is postponed to a future research work.

term $\boldsymbol{\Upsilon}_k$. On the other hand, the unitary matrices $\boldsymbol{\Phi}^H$ and $\boldsymbol{\Phi}$ do not cancel out in the presence of spatial correlation, i.e., the matrices $\mathbf{R}_{ris}$ and $\mathbf{R}_{emi}$ are not identity matrices. This implies that an RIS can be utilized for improving the channel estimation accuracy for transmission over spatially correlated channels. This is a benefit that spatial correlation brings in RIS-aided systems. If the spatial correlation is negligible, by contrast, we obtain $\mathbf{R}_{ris} = \mathbf{R}_{emi} = \mathbf{I}_N$ and the MSE matrix in (75) no longer depends on $\boldsymbol{\Phi}$, and therefore we cannot optimize the phase shifts of the RIS to improve the quality of channel estimation.

C. Achievable Rate

Based on the estimated channel $\hat{\mathbf{q}}_{c,k}$, the MRC detector can be obtained and the corresponding UatF bound of the achievable rate can be computed in the presence of spatial correlation and EMI as well. Specifically, by pre-multiplying the MRC decoding matrix $\hat{\mathbf{Q}}_c^H = [\hat{\mathbf{q}}_{c,1}, \ldots, \hat{\mathbf{q}}_{c,K}]^H$ with the received signal $\mathbf{y}_c$ in (66), the decoded symbols at the BS are given by

$$\mathbf{r}_c = \hat{\mathbf{Q}}_c^H \mathbf{y}_c = \sqrt{p}\hat{\mathbf{Q}}_c^H \mathbf{Q}_c \mathbf{x} + \hat{\mathbf{Q}}_c^H \mathbf{H}_{c,2} \boldsymbol{\Phi} \boldsymbol{v} + \hat{\mathbf{Q}}_c^H \mathbf{n}. \quad (76)$$

Then, the k-th entry of $\mathbf{r}_c$ can be expressed as follows

$$\begin{aligned} r_{c,k} =& \sqrt{p}\mathbb{E}\{\hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,k}\} x_k + \sqrt{p}\big(\hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,k} - \mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,k}\right\}\big) x_k \\ &+ \sqrt{p} \sum_{i=1, i\neq k}^{K} \hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,i} x_i + \hat{\mathbf{q}}_{c,k}^H \mathbf{H}_{c,2} \boldsymbol{\Phi} \boldsymbol{v} + \hat{\mathbf{q}}_{c,k}^H \mathbf{n}. \end{aligned} \quad (77)$$

Accordingly, the SINR of user k can be written as

$$\mathrm{SINR}_{c,k} = \frac{p E_{c,k}^{\mathrm{signal}}}{p E_{c,k}^{\mathrm{leak}} + p \sum_{i=1, i\neq k}^{K} I_{c,ki} + \sigma_e^2 E_{c,k}^{\mathrm{emi}} + \sigma^2 E_{c,k}^{\mathrm{noise}}}, \quad (78)$$

where the desired signal is $E_{c,k}^{\mathrm{signal}} = \left| \mathbb{E}\left\{ \hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,k} \right\} \right|^2$, the signal leakage is $E_{c,k}^{\mathrm{leak}} = \mathbb{E}\left\{ \left| \hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,k} \right|^2 \right\} - \left| \mathbb{E}\left\{ \hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,k} \right\} \right|^2$, the interference is $I_{c,ki} = \mathbb{E}\left\{ \left| \hat{\mathbf{q}}_{c,k}^H \mathbf{q}_{c,i} \right|^2 \right\}$, the EMI is $E_{c,k}^{\mathrm{emi}} = \mathbb{E}\left\{ \hat{\mathbf{q}}_{c,k}^H \mathbf{H}_{c,2} \boldsymbol{\Phi} \mathbf{R}_{emi} \boldsymbol{\Phi}^H \mathbf{H}_{c,2}^H \hat{\mathbf{q}}_{c,k} \right\}$, and the noise is $E_{c,k}^{\mathrm{noise}} = \mathbb{E}\left\{ \|\hat{\mathbf{q}}_{c,k}\|^2 \right\}$.

In order to obtain a compact expression for the UatF bound of the achievable rate, we introduce the following shorthand functions, for $1 \leq k, i \leq K$

$$\begin{aligned} &f_{c,1}(\boldsymbol{\Phi}) = \mathrm{Tr}\left\{ \mathbf{R}_{ris} \boldsymbol{\Phi} \mathbf{R}_{emi} \boldsymbol{\Phi}^H \right\}, \\ &f_{c,k,2}(\boldsymbol{\Phi}) = \overline{\mathbf{h}}_k^H \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \overline{\mathbf{h}}_k, \\ &f_{c,k,3}(\boldsymbol{\Phi}) = \mathrm{Tr}\left\{ \boldsymbol{\Upsilon}_k^2 \overline{\mathbf{H}}_2 \boldsymbol{\Phi} \mathbf{R}_{emi} \boldsymbol{\Phi}^H \overline{\mathbf{H}}_2^H \right\}, \\ &f_{c,k,4}(\boldsymbol{\Phi}) = \mathrm{Tr}\left\{ \boldsymbol{\Upsilon}_k^2 \right\}, \quad f_{c,k,5}(\boldsymbol{\Phi}) = \left| \mathrm{Tr}\left\{ \boldsymbol{\Upsilon}_k \right\} \right|^2, \\ &f_{c,k,6}(\boldsymbol{\Phi}) = \overline{\mathbf{h}}_k^H \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \mathbf{R}_{emi} \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \overline{\mathbf{h}}_k, \\ &f_{c,k,7}(\boldsymbol{\Phi}) = \left| f_k(\boldsymbol{\Phi}) \right|^2, \; f_{c,ki,8}(\boldsymbol{\Phi}) = \overline{\mathbf{h}}_i^H \boldsymbol{\Phi}^H \overline{\mathbf{H}}_2^H \boldsymbol{\Upsilon}_k^2 \overline{\mathbf{H}}_2 \boldsymbol{\Phi} \overline{\mathbf{h}}_i, \\ &f_{c,ki,9}(\boldsymbol{\Phi}) = \overline{\mathbf{h}}_i^H \boldsymbol{\Phi}^H \overline{\mathbf{H}}_2^H \boldsymbol{\Upsilon}_k \overline{\mathbf{H}}_2 \boldsymbol{\Phi} \mathbf{R}_{emi} \boldsymbol{\Phi}^H \overline{\mathbf{H}}_2^H \boldsymbol{\Upsilon}_k^H \overline{\mathbf{H}}_2 \boldsymbol{\Phi} \overline{\mathbf{h}}_i. \end{aligned} \quad (79)$$

Theorem 4 *In the presence of spatial correlation and EMI, the UatF bound for the achievable rate of the k-th user is given by*

$$\underline{R}_{c,k} = \tau^o \log_2 \left(1 + \mathrm{SINR}_{c,k} \right), \quad (80)$$

$$\mathrm{SINR}_{c,k} = \frac{p E_{c,k}^{\mathrm{signal}}}{p E_{c,k}^{\mathrm{leak}} + p \sum_{i=1, i\neq k}^{K} I_{c,ki} + \sigma_e^2 E_{c,k}^{\mathrm{emi}} + \sigma^2 E_{c,k}^{\mathrm{noise}}}, \quad (81)$$

where the signal term is $E_{c,k}^{\mathrm{signal}} = \left(E_{c,k}^{\mathrm{noise}} \right)^2$ and the noise term is

$$\begin{aligned} E_{c,k}^{\mathrm{noise}} =& M \widehat{c}_k \delta \left| f_k(\boldsymbol{\Phi}) \right|^2 + \widehat{c}_k \, \mathrm{Tr}\left\{ \boldsymbol{\Upsilon}_k \right\} \overline{\mathbf{h}}_k^H \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \overline{\mathbf{h}}_k \\ &+ \gamma_k \, \mathrm{Tr}\left\{ \boldsymbol{\Upsilon}_k \right\}. \end{aligned} \quad (82)$$

$$\begin{aligned} \boldsymbol{\Upsilon}_k = \boldsymbol{\Upsilon}_k^H =& \left(\widehat{c}_k \overline{\mathbf{h}}_k^H \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \overline{\mathbf{h}}_k + \gamma_k \right) \times \\ & \left\{ \left(\widehat{c}_k \overline{\mathbf{h}}_k^H \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \overline{\mathbf{h}}_k + \gamma_k + \frac{\sigma^2}{\tau p} + \frac{\sigma_e^2 \beta \, \mathrm{Tr}\left\{ \mathbf{R}_{emi} \boldsymbol{\Phi}^H \mathbf{R}_{ris} \boldsymbol{\Phi} \right\}}{\tau p (\delta + 1)} \right) \mathbf{I}_M + \frac{\sigma_e^2 \beta \delta \overline{\mathbf{H}}_2 \boldsymbol{\Phi} \mathbf{R}_{emi} \boldsymbol{\Phi}^H \overline{\mathbf{H}}_2^H}{\tau p (\delta + 1)} \right\}^{-1}. \end{aligned} \quad (74)$$

The EMI term is given by $E_{c,k}^{\mathrm{emi}} = \frac{\beta}{\delta+1}\sum_{\omega=1}^{8} E_{c,k}^{\omega,\mathrm{emi}}$ *where*

$$
\begin{aligned}
E_{c,k}^{1,\mathrm{emi}} &= M^2\widehat{c}_k\delta^2 f_{c,k,7}(\mathbf{\Phi})\mathbf{a}_N^H\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\mathbf{a}_N,\\
E_{c,k}^{2,\mathrm{emi}} &= \Bigg(\widehat{c}_k\delta f_{c,k,2}(\mathbf{\Phi}) + \frac{2\beta\delta\sigma_e^2}{\tau p(\delta+1)} f_{c,1}(\mathbf{\Phi})\\
&\quad + \delta\left(\gamma_k + \frac{\sigma^2}{\tau p}\right)\Bigg) f_{c,k,3}(\mathbf{\Phi}),\\
E_{c,k}^{3,\mathrm{emi}} &= \Bigg(M\widehat{c}_k\delta f_{c,k,7}(\mathbf{\Phi}) + \Bigg(\frac{\sigma^2}{\tau p} + \gamma_k + \widehat{c}_k f_{c,k,2}(\mathbf{\Phi})\\
&\quad + \frac{\beta\sigma_e^2}{\tau p(\delta+1)} f_{c,1}(\mathbf{\Phi})\Bigg) f_{c,k,4}(\mathbf{\Phi})\Bigg) f_{c,1}(\mathbf{\Phi}),\\
E_{c,k}^{4,\mathrm{emi}} &= \frac{\beta\delta^2\sigma_e^2}{\tau p(\delta+1)}\,\mathrm{Tr}\left\{\left(\mathbf{R}_{emi}\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbf{\Upsilon}_k\overline{\mathbf{H}}_2\mathbf{\Phi}\right)^2\right\},\\
E_{c,k}^{5,\mathrm{emi}} &= 2\widehat{c}_k\delta\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k\right\}\\
&\quad \times \mathrm{Re}\left\{\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}_{emi}\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k\right\},\\
E_{c,k}^{6,\mathrm{emi}} &= \frac{2\beta\delta\sigma_e^2}{\tau p(\delta+1)}\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k\right\}\\
&\quad \times \mathrm{Tr}\left\{\mathbf{R}_{emi}\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbf{\Upsilon}_k^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\right\},\\
E_{c,k}^{7,\mathrm{emi}} &= \widehat{c}_k f_{c,k,5}(\mathbf{\Phi}) f_{c,k,6}(\mathbf{\Phi}),\\
E_{c,k}^{8,\mathrm{emi}} &= \frac{\beta\sigma_e^2}{\tau p(\delta+1)} f_{c,k,5}(\mathbf{\Phi})\,\mathrm{Tr}\left\{\left(\mathbf{R}_{ris}\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\right)^2\right\}.
\end{aligned}
\tag{83}
$$

The interference term is $I_{c,ki} = \sum_{\omega=1}^{8} I_{c,ki}^{\omega}$, *where*

$$
\begin{aligned}
I_{c,ki}^{1} &= \gamma_i E_{c,k}^{\mathrm{noise}} + M^2\widehat{c}_k\widehat{c}_i\delta^2 f_{c,k,7}(\mathbf{\Phi}) f_{c,i,7}(\mathbf{\Phi}),\\
I_{c,ki}^{2} &= \Bigg\{M\widehat{c}_k\widehat{c}_i\delta f_{c,k,7}(\mathbf{\Phi})\\
&\quad + \left(\widehat{c}_i\left(\gamma_k + \frac{\sigma^2}{\tau p}\right) + \frac{\widehat{c}_i\beta\sigma_e^2}{\tau p(\delta+1)} f_{c,1}(\mathbf{\Phi})\right) f_{c,k,4}(\mathbf{\Phi})\\
&\quad + \frac{\widehat{c}_i\beta\delta\sigma_e^2}{\tau p(\delta+1)} f_{c,k,3}(\mathbf{\Phi})\Bigg\} f_{c,i,2}(\mathbf{\Phi}),\\
I_{c,ki}^{3} &= \left\{\widehat{c}_k\widehat{c}_i\delta f_{c,ki,8}(\mathbf{\Phi}) + \widehat{c}_k\widehat{c}_i f_{c,k,4}(\mathbf{\Phi}) f_{c,i,2}(\mathbf{\Phi})\right\} f_{c,k,2}(\mathbf{\Phi}),\\
I_{c,ki}^{4} &= \left\{\frac{\widehat{c}_i\beta\delta\sigma_e^2}{\tau p(\delta+1)} f_{c,1}(\mathbf{\Phi}) + \widehat{c}_i\delta\left(\gamma_k + \frac{\sigma^2}{\tau p}\right)\right\} f_{c,ki,8}(\mathbf{\Phi}),\\
I_{c,ki}^{5} &= \left\{\widehat{c}_k\widehat{c}_i\left|\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_i\right|^2 + \frac{\widehat{c}_i\beta\sigma_e^2}{\tau p(\delta+1)} f_{c,i,6}(\mathbf{\Phi})\right\}\\
&\quad \times f_{c,k,5}(\mathbf{\Phi}),\\
I_{c,ki}^{6} &= 2\widehat{c}_k\widehat{c}_i\delta\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k\right\}\\
&\quad \times \mathrm{Re}\left\{\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k\right\},\\
I_{c,ki}^{7} &= \frac{\widehat{c}_i\beta\delta^2\sigma_e^2}{\tau p(\delta+1)} f_{c,ki,9}(\mathbf{\Phi}),\\
I_{c,ki}^{8} &= \frac{2\widehat{c}_i\beta\delta\sigma_e^2}{\tau p(\delta+1)}\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k\right\}\\
&\quad \times \mathrm{Re}\left\{\overline{\mathbf{h}}_i^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbf{\Upsilon}_k^H\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_i\right\}.
\end{aligned}
\tag{84}
$$

The signal leakage term is $E_{c,k}^{\mathrm{leak}} = \sum_{\omega=1}^{8} E_{c,k}^{\omega,\mathrm{leak}}$, *where*

$$
\begin{aligned}
E_{c,k}^{1,\mathrm{leak}} &= M\widehat{c}_k\delta\gamma_k f_{c,k,7}(\mathbf{\Phi}),\\
E_{c,k}^{2,\mathrm{leak}} &= \Bigg\{M\widehat{c}_k^2\delta f_{c,k,7}(\mathbf{\Phi}) + \widehat{c}_k^2\delta f_{c,kk,8}(\mathbf{\Phi})\\
&\quad + \left(\widehat{c}_k^2 f_{c,k,2}(\mathbf{\Phi}) + 2\widehat{c}_k\gamma_k + \frac{\widehat{c}_k\sigma^2}{\tau p}\right) f_{c,k,4}(\mathbf{\Phi})\Bigg\}\\
&\quad \times f_{c,k,2}(\mathbf{\Phi}),\\
E_{c,k}^{3,\mathrm{leak}} &= \left\{\widehat{c}_k\delta\gamma_k + \frac{\widehat{c}_k\beta\delta\sigma_e^2}{\tau p(\delta+1)} f_{c,1}(\mathbf{\Phi}) + \frac{\widehat{c}_k\delta\sigma^2}{\tau p}\right\} f_{c,kk,8}(\mathbf{\Phi}),\\
E_{c,k}^{4,\mathrm{leak}} &= \Bigg\{\gamma_k^2 + \frac{\gamma_k\sigma^2}{\tau p}\\
&\quad + \frac{\beta\sigma_e^2}{\tau p(\delta+1)}\left(\gamma_k + \widehat{c}_k f_{c,k,2}(\mathbf{\Phi})\right) f_{c,1}(\mathbf{\Phi})\Bigg\} f_{c,k,4}(\mathbf{\Phi}),\\
E_{c,k}^{5,\mathrm{leak}} &= \frac{\widehat{c}_k\beta\delta^2\sigma_e^2}{\tau p(\delta+1)} f_{c,kk,9}(\mathbf{\Phi}),\\
E_{c,k}^{6,\mathrm{leak}} &= \frac{2\widehat{c}_k\beta\delta\sigma_e^2}{\tau p(\delta+1)}\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k^H\right\}\\
&\quad \times \mathrm{Re}\left\{\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbf{\Upsilon}_k\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k\right\},\\
E_{c,k}^{7,\mathrm{leak}} &= \frac{\beta\delta\sigma_e^2}{\tau p(\delta+1)}\left\{\gamma_k + \widehat{c}_k f_{c,k,2}(\mathbf{\Phi})\right\} f_{c,k,3}(\mathbf{\Phi}),\\
E_{c,k}^{8,\mathrm{leak}} &= \frac{\widehat{c}_k\beta\sigma_e^2}{\tau p(\delta+1)} f_{c,k,5}(\mathbf{\Phi}) f_{c,k,6}(\mathbf{\Phi}).
\end{aligned}
\tag{85}
$$

Proof: See Appendix H. ■

By comparing the rate $\underline{R}_{c,k}$ in Theorem 4 with the rate $\underline{R}_k$ in Theorem 2, we can unveil the impact of spatial correlation and EMI. The impact of spatial correlation on the achievable rate is discussed in the following remark.

Remark 3 *As briefly mentioned for the MSE in (75), the presence of spatial correlation could enhance the capabilities of an RIS to tailor a wireless channel. This is apparent by the direct inspection of the rate in Theorem 4 as well. To be specific, consider the term* $\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k$ *as an example. If the spatial correlation is negligible, this term is fixed and equal to* N *without any possibility to be adjusted by the RIS, since the matrix* $\mathbf{\Phi}$ *is a unitary matrix and* $\mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}_N$. *However, the same term can be shaped by an RIS in the presence of spatial correlation. For simplicity, let us assume the most severe setup in terms of spatial correlation, i.e.,* $\mathbf{R}_{ris} = \mathbf{1}_{N\times N}$ *so that* $\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k = \left|\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{1}_{N\times 1}\right|^2$. *Based on the proof of Lemma 3, we have* $0 \le \left|\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{1}_{N\times 1}\right|^2 \le N^2$, *which demonstrates the enhanced adjustment ability of an RIS to shape the channel in the presence of spatial correlation.*

Next, we discuss the impact of the EMI on the power scaling laws. Due to the complex expressions in (83) and the fact that the optimal design of the RIS phase shifts matrix $\mathbf{\Phi}$ cannot be obtained in a closed-form expression, general conclusions cannot be drawn. However, some special cases are discussed in the following corollary based on the proof by contradiction method.

Corollary 11 *The power scaling laws summarized in Table II are not guaranteed to hold in the presence of EMI.*

Proof: We first give a counterexample for the power scaling laws as a function of M. Specifically, we note that the desired signal $E_{c,k}^{\text{signal}}$ and the EMI term $E_{c,k}^{1,\text{emi}}$ in (83) scale as $\mathcal{O}\left(M^2\right)$. If the power is scaled proportionally to $p = 1/M$, therefore, the SINR in (81) tends to zero when $M \to \infty$. Let us now give a counterexample for the power scaling laws as a function of N. Consider the case study in which only the NLoS components of the channels are present, i.e., $\delta = 0$, and no spatial correlation is present, i.e., $\mathbf{R}_{ris} = \mathbf{R}_{emi} = \mathbf{I}_N$. Accordingly, $\boldsymbol{\Upsilon}_k$ simplifies as follows

$$\text{Tr}\{\boldsymbol{\Upsilon}_k\} = \frac{M(N\widehat{c}_k + \gamma_k)}{N\widehat{c}_k + \gamma_k + \frac{\sigma^2}{\tau p} + \frac{N\sigma_e^2\beta}{\tau p}}. \tag{86}$$

Then, we have $E_{c,k}^{\text{signal}} = \left(E_{c,k}^{\text{noise}}\right)^2$ where

$$E_{c,k}^{\text{noise}} = \frac{M(N\widehat{c}_k + \gamma_k)^2}{N\widehat{c}_k + \gamma_k + \frac{\sigma^2}{\tau p} + \frac{N\sigma_e^2\beta}{\tau p}}. \tag{87}$$

If the power is scaled proportionally to $p = 1/N$ when $N \to \infty$, (87) implies that $E_{c,k}^{\text{signal}} \to (\frac{\tau M \widehat{c}_k^2}{\sigma_e^2 \beta})^2$, which implies $pE_{c,k}^{\text{signal}} = E_{c,k}^{\text{signal}}/N \to 0$. Therefore, the SINR would tend to zero. This special case demonstrates that the power scaling laws with respect to N are not guaranteed to hold in the presence of EMI. ■

A simple explanation for Corollary 11 is the following. If the users' transmit power p is scaled proportionally to $1/M$ or $1/N$, as M or N increases, the intended signal power received by the RIS becomes weaker and weaker while the power of the EMI received by the RIS is unaffected. Thus, the EMI becomes stronger and stronger as compared to the intended signal. In other words, as $M, N \to \infty$, the useful power becomes extremely weak and the EMI power dominates the received signal at the RIS.

Nevertheless, we note that the importance of the power scaling laws does not lie in the performance limits in the asymptotic regime for $M, N \to \infty$. In practice, neither the number of BS antennas nor the number of RIS elements can be infinite. The analysis of the power scaling laws is insightful to understand whether the transmit power of the users can be reduced by increasing M or N while not significantly sacrificing the rate. Therefore, we are usually interested in the power scaling laws when M or N is large but finite. The considered channel model can, in addition, be applied in the far-field region of the BS and RIS, and hence it is not possible to consider an infinite number of BS antennas or RIS elements. Besides, the users share the same RIS-BS channel in RIS-aided systems, which results in strong multi-user interference when applying MRC, as noted in Remark 2. Even though the EMI re-radiated by an RIS may be stronger than the thermal noise, it may not necessarily be stronger than the multi-user interference when M or N is not very large. Specifically, some numerical examples about the impact of the EMI on the achievable rate and power scaling laws are reported in Section VII.

VI. Design of the RIS Phase Shifts

In this section, we optimize the phase shifts of the RIS to maximize the achievable rate derived in Theorem 2 and Theorem 4. Since the derived ergodic rate depends only on statistical CSI, we need to update the phase shifts of the RIS according to the time variations of the long-term CSI. This results in less frequent updates of the RIS phase shifts especially in the sub-6 GHz frequency range, which, in turn, reduces the channel acquisition overhead and the computational complexity.

A. Single-user Case

Before tackling the general optimization problem, we first justify the statement made in Section IV-C that the optimal phase shift matrix that maximizes the rate in the single-user case fulfills the condition $|f_k(\boldsymbol{\Phi})| = N$. To this end, this subsection aims to solve the phase shifts optimization problem in the single-user case and in the absence of spatial correlation and EMI.

In the single-user case, only the user k is present. We aim to find the phase shifts matrix $\boldsymbol{\Phi}$ that maximizes the lower bound of the ergodic rate $\underline{R}_k$ in Theorem 2 by setting $I_{ki}(\boldsymbol{\Phi}) = 0$. Only the scenarios with $N > 1$, $\delta > 0$ and $\varepsilon_k > 0, \forall k$ are considered, since $\boldsymbol{\Phi}$ can be set arbitrarily otherwise. It can be observed that the phase shifts matrix $\boldsymbol{\Phi}$ appears only in the term $|f_k(\boldsymbol{\Phi})|^2$. For clarity, we denote $x = |f_k(\boldsymbol{\Phi})|^2$ as the optimization variable. Then, the rate $\underline{R}_k$ in Theorem 2 can be rewritten as given in (88), which depends on some constants s_1, s_2, t_1 and t_2, as follows

$$\begin{aligned}\underline{R}_k &= \tau^o \log_2\left(1 + \text{SNR}_k(x)\right) \\ &= \tau^o \log_2\left(1 + \frac{E_k^{\text{signal}}(x)}{E_k^{\text{leak}}(x) + \frac{\sigma^2}{p} E_k^{\text{noise}}(x)}\right) \\ &= \tau^o \log_2\left(1 + \frac{(s_1 x + s_2)^2}{t_1 x + t_2}\right).\end{aligned} \tag{88}$$

The expressions of s_1, s_2, t_1 and t_2 can be derived by direct inspection of Theorem 2 and therefore are omitted for brevity. Besides, it is not difficult to prove that $s_1, s_2, t_1, t_2 > 0$. From Lemma 3, we know that the domain of the variable x is $0 \leq x \leq N^2$. Based on (88), therefore, the optimization problem can be formulated as follows

$$\max_x \quad \text{SNR}_k(x) = \frac{(s_1 x + s_2)^2}{t_1 x + t_2}, \tag{89a}$$

$$\text{s.t.} \quad 0 \leq x \leq N^2. \tag{89b}$$

To solve the problem in (89), we compute the first-order derivative of $\text{SNR}_k(x)$ with respect to x, as follows

$$\frac{\partial \text{SNR}_k(x)}{\partial x} = \frac{(s_1 x + s_2)(s_1 t_1 x + 2 s_1 t_2 - s_2 t_1)}{(t_1 x + t_2)^2}. \tag{90}$$

The first-order derivative of $\text{SNR}_k(x)$ is positive or negative depending on the numerator in (90), which is a quadratic function of x, i.e., a parabola opening upward, with two roots.

The two roots can be obtained by setting (90) equal to zero, which yields

$$x_0^L = \frac{-s_2}{s_1}, \quad x_0^R = \frac{s_2 t_1 - 2 s_1 t_2}{s_1 t_1}, \tag{91}$$

where $x_0^L < 0$ while x_0^R can be positive.

We can design the optimal configuration of $\mathbf{\Phi}$ by analyzing the derivative $\frac{\partial \mathrm{SNR}_k(x)}{\partial x}$ in the domain of x, i.e., (89b), which depends on x_0^R. For example, if $x_0^R \leq 0$, for a parabola opening upward, we obtain $\frac{\partial \mathrm{SNR}_k(x)}{\partial x} \geq 0$ in the domain $0 \leq x \leq N^2$. The complete optimal design criterion is summarized in the following theorem.

Theorem 5 *For RIS-aided single-user systems subject to imperfect CSI, the optimal phase shift matrix $\mathbf{\Phi}$ obtained by maximizing the UatF bound of the achievable rate can be summarized as follows.*

- *It is optimal to set $|f_k(\mathbf{\Phi})| = N$ if (1) $x_0^R \leq 0$; or (2) $0 < x_0^R < N^2$ and $\mathrm{SNR}_k(0) \leq \mathrm{SNR}_k(N^2)$; or (3) $N \to \infty$.*
- *It is optimal to set $|f_k(\mathbf{\Phi})| = 0$ if (4) $0 < x_0^R < N^2$ and $\mathrm{SNR}_k(0) > \mathrm{SNR}_k(N^2)$; or (5) $x_0^R \geq N^2$.*

Proof: It follows by direct inspection of x_0^R. If $x_0^R \leq 0$, we obtain $\frac{\partial \mathrm{SNR}_k(x)}{\partial x} \geq 0$ in the domain $0 \leq x \leq N^2$. Thus, the SNR is an increasing function of x in its domain, which implies that the maximum SNR is reached at the endpoint $x = N^2$. Therefore, it is optimal to set $|f_k(\mathbf{\Phi})| = N$. If $x_0^R \geq N^2$, we obtain $\frac{\partial \mathrm{SNR}_k(x)}{\partial x} \leq 0$ in the domain of x. Thus, the SNR is a decreasing function of x, which implies that the maximum SNR is reached at the endpoint $x = 0$. Therefore, it is optimal to set $|f_k(\mathbf{\Phi})| = 0$. If $0 < x_0^R < N^2$, the SNR first decreases for $x < x_0^R$, and then increases for $x > x_0^R$. Therefore, the maximum SNR is obtained either at $x = 0$ or at $x = N^2$. By comparing $\mathrm{SNR}_k(0)$ with $\mathrm{SNR}_k(N^2)$, we can identify the optimal design. Finally, we focus on a special case of $N \to \infty$. In this context, we have $\mathrm{SNR}_k(0) < \mathrm{SNR}_k(N^2)$, since $\mathrm{SNR}_k(0)$ is bounded while $\mathrm{SNR}_k(N^2) \to \infty$. Therefore, it is optimal to set $|f_k(\mathbf{\Phi})| = N$ if $N \to \infty$. ■

Finally, we note that the optimal design obtained for $N \to \infty$ substantiates the analysis reported in Section IV-C for large N.

B. Multi-user Case

In this subsection, we consider the design of the RIS phase shifts in the general multi-user scenario with $K > 1$. In the multi-user case, as mentioned in Remark 2, it is necessary to guarantee some fairness requirements among the different users. To this end, we aim to maximize the minimum rate of the users. As a result, the optimization problem can be formulated as follows

$$\max_{\mathbf{\Phi}} \min_{k \in \mathcal{K}} \ \underline{R}_k(\mathbf{\Phi}) \text{ or } \underline{R}_{c,k}(\mathbf{\Phi}), \tag{92a}$$

$$\text{s.t.} \quad \left| [\mathbf{\Phi}]_{n,n} \right| = 1, \forall n, \tag{92b}$$

where $\underline{R}_k(\mathbf{\Phi})$ is given by (39) in Theorem 2 and $\underline{R}_{c,k}(\mathbf{\Phi})$ is given by (80) in Theorem 4. Constraint (92b) is the unit modulus constraint for the RIS phase shifts matrix.

For tractability, we introduce the vectors $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_N]^T$ and $\boldsymbol{c} = [e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}]^T$ so that $\boldsymbol{c} = e^{j\boldsymbol{\theta}}$ and $\mathbf{\Phi} = \mathrm{diag}(\boldsymbol{c})$. Then, the problem in (92) can be solved effectively based on the gradient ascent method with respect to the real variable $\boldsymbol{\theta}$. It is worth noting that our proposed method is different from existing works which adopted the projected gradient ascent method with respect to complex variable $\boldsymbol{c}$ [46]. To be specific, after updating $\boldsymbol{c}$, the projected gradient ascent method needs a projection operation to ensure that the updated variable $\boldsymbol{c}_{new}$ fulfills the unit modulus constraint $|\boldsymbol{c}_{new}| = \mathbf{1}$. By contrast, the proposed gradient ascent method avoids the suboptimality caused by the projection operation since the complex exponential functions are periodic with $\boldsymbol{\theta}$ and the unit modulus constraint holds for every phase shifts vector $\boldsymbol{\theta}$. Besides, the performance of the gradient ascent method highly depends on the step size, and working with real variables makes the algorithm more robust to the choice of this tuning parameter [73].

The gradient with respect to $\boldsymbol{\theta}$ is given as follows. Since the objective function in (92) includes the min function, which is not differentiable, we first approximate the objective function in (92) as

$$\min_k \underline{R}_k(\boldsymbol{\theta}) \approx -\frac{1}{\mu} \ln \left\{ \sum_{k=1}^K \exp\left\{ -\mu \underline{R}_k(\boldsymbol{\theta}) \right\} \right\} \triangleq f(\boldsymbol{\theta}), \tag{93}$$

$$\min_k \underline{R}_{c,k}(\boldsymbol{\theta}) \approx -\frac{1}{\mu} \ln \left\{ \sum_{k=1}^K \exp\left\{ -\mu \underline{R}_{c,k}(\boldsymbol{\theta}) \right\} \right\} \triangleq f_c(\boldsymbol{\theta}), \tag{94}$$

where μ is a constant value for controlling the accuracy of the approximation. It can be proved that the approximation error is smaller than $\frac{\ln K}{\mu}$ based on the method in [74]. Thus, the problem in (92) can be recast as

$$\max_{\boldsymbol{\theta}} \ f(\boldsymbol{\theta}) \text{ or } f_c(\boldsymbol{\theta}), \tag{95a}$$

$$\text{s.t.} \quad 0 \leq \theta_n < 2\pi, \forall n. \tag{95b}$$

As mentioned, the constraint (95b) can be neglected thanks to the periodicity of the objective functions $f(\boldsymbol{\theta})$ and $f_c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Therefore, there is no need to perform any projection operation after updating variable $\boldsymbol{\theta}$. Then, we need to calculate the gradient of $f(\boldsymbol{\theta})$ and $f_c(\boldsymbol{\theta})$. Since these two gradients can be calculated in a similar way, we only provide the detailed process for $\frac{\partial f_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. Based on the chain rule, we have

$$\frac{\partial f_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\tau^o \sum_{k=1}^K \left\{ \frac{\exp\left\{ -\mu \underline{R}_{c,k}(\boldsymbol{\theta}) \right\}}{1 + \mathrm{SINR}_{c,k}(\boldsymbol{\theta})} \frac{\partial \mathrm{SINR}_{c,k}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}}{(\ln 2) \left(\sum_{k=1}^K \exp\left\{ -\mu \underline{R}_{c,k}(\boldsymbol{\theta}) \right\} \right)}, \tag{96}$$

and $\frac{\partial \mathrm{SINR}_{c,k}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ shown at the bottom of the next page.

Therefore, the gradient of $f_c(\boldsymbol{\theta})$ can be obtained after calculating $\frac{\partial E_{c,k}^{\mathrm{signal}}}{\partial \boldsymbol{\theta}}$, $\frac{\partial E_{c,k}^{\mathrm{leak}}}{\partial \boldsymbol{\theta}}$, $\frac{\partial I_{c,ki}}{\partial \boldsymbol{\theta}}$, $\frac{\partial E_{c,k}^{\mathrm{emi}}}{\partial \boldsymbol{\theta}}$, and $\frac{\partial E_{c,k}^{\mathrm{noise}}}{\partial \boldsymbol{\theta}}$ in (97). Based on Theorem 4, we note that $E_{c,k}^{\mathrm{signal}}$, $I_{c,ki}$, $E_{c,k}^{\mathrm{leak}}$, $E_{c,k}^{\mathrm{emi}}$ and $E_{c,k}^{\mathrm{noise}}$ can be computed from the functions in (79). For ease of description, we first provide two useful lemmas and then use them to calculate the gradient of the terms in (79).

Lemma 4 *Given the deterministic matrices* $\mathbf{A}$ *and* $\mathbf{B}$*, the gradient of* $\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}$ *with respect to* $\boldsymbol{\theta}$ *is given by*

$$\begin{aligned}\frac{\partial \mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial \boldsymbol{\theta}} &= j\boldsymbol{\Phi}^T\left(\mathbf{A}^T\odot\mathbf{B}\right)\boldsymbol{c}^* - j\boldsymbol{\Phi}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\boldsymbol{c}\\ &\triangleq \boldsymbol{f}_d(\mathbf{A},\mathbf{B}).\end{aligned} \tag{98}$$

If $\mathbf{A}=\mathbf{A}^H, \mathbf{B}=\mathbf{B}^H$*, we further have*

$$\frac{\partial \mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial \boldsymbol{\theta}} = 2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\boldsymbol{c}\right\}. \tag{99}$$

Proof: See Appendix I. ■

Lemma 5 *Define* $\psi_k^1 = \widehat{c}_k\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\mathbf{R}_{ris}\boldsymbol{\Phi}\overline{\mathbf{h}}_k + \gamma_k$ *and* $\boldsymbol{\Upsilon}_k = \psi_k^1\boldsymbol{\Upsilon}_k^1$*. Then, given the deterministic matrix* $\mathbf{T}$*, the gradient of* $\mathrm{Tr}\left\{\mathbf{T}\boldsymbol{\Upsilon}_k\right\}$ *with respect to* $\boldsymbol{\theta}$ *is given by*

$$\begin{aligned}&\frac{\partial \mathrm{Tr}\left\{\mathbf{T}\boldsymbol{\Upsilon}_k\right\}}{\partial \boldsymbol{\theta}}\\ &= 2\widehat{c}_k\left\{\mathrm{Tr}\left\{\mathbf{T}\boldsymbol{\Upsilon}_k^1\right\} - \psi_k^1\,\mathrm{Tr}\left\{\mathbf{T}\left(\boldsymbol{\Upsilon}_k^1\right)^2\right\}\right\}\\ &\quad\times\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{R}_{ris}\odot\left(\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\right)^T\right)\boldsymbol{c}\right\}\\ &\quad-\frac{2\sigma_e^2\beta}{\tau p(\delta+1)}\psi_k^1\,\mathrm{Tr}\left\{\mathbf{T}\left(\boldsymbol{\Upsilon}_k^1\right)^2\right\}\mathrm{Im}\left\{\boldsymbol{\Phi}^H(\mathbf{R}_{ris}\odot\mathbf{R}_{emi})\,\boldsymbol{c}\right\}\\ &\quad-\frac{\sigma_e^2\beta\delta}{\tau p(\delta+1)}\psi_k^1\boldsymbol{f}_d\left(\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^1\mathbf{T}\boldsymbol{\Upsilon}_k^1\overline{\mathbf{H}}_2,\mathbf{R}_{emi}\right)\\ &\triangleq \mathbf{z}_k(\mathbf{T})\end{aligned} \tag{100}$$

Proof: The proof is similar to the proof of Lemma 4 after applying the chain rule to the inverse matrix $\partial\left(\mathbf{X}^{-1}\right) = -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1}$. ■

With the aid of Lemma 4 and 5, we obtain the following lemma.

Lemma 6 *The gradients of the functions defined in (79) are given by*

$$\begin{aligned}\boldsymbol{f}_{c,1}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,1}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}} = 2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{R}_{ris}\odot\mathbf{R}_{emi}\right)\boldsymbol{c}\right\},\\ \boldsymbol{f}_{c,k,2}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,k,2}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}} = 2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{R}_{ris}\odot\left(\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\right)^T\right)\boldsymbol{c}\right\},\\ \boldsymbol{f}_{c,k,3}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,k,3}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}} = \mathbf{z}_k\left(\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k\right)\\ &\quad+2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^2\overline{\mathbf{H}}_2\odot\mathbf{R}_{emi}\right)\boldsymbol{c}\right\}\\ &\quad+\mathbf{z}_k\left(\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\right),\end{aligned} \tag{101}$$

$$\begin{aligned}\boldsymbol{f}_{c,k,4}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,k,4}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}} = 2\mathbf{z}_k\left(\boldsymbol{\Upsilon}_k\right),\\ \boldsymbol{f}_{c,k,5}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,k,5}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}} = 2\,\mathrm{Tr}\left\{\boldsymbol{\Upsilon}_k\right\}\mathbf{z}_k\left(\mathbf{I}_M\right),\\ \boldsymbol{f}_{c,k,6}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,k,6}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}}\\ &= 2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{R}_{ris}\boldsymbol{\Phi}\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\mathbf{R}_{ris}\odot\mathbf{R}_{emi}\right)\boldsymbol{c}\right\}\\ &\quad+2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{R}_{ris}\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\mathbf{R}_{ris}\odot\left(\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\right)^T\right)\boldsymbol{c}\right\},\end{aligned} \tag{102}$$

$$\begin{aligned}\boldsymbol{f}_{c,k,7}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,k,7}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}}\\ &= 2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\mathbf{a}_N\mathbf{a}_N^H\odot\left(\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\right)^T\right)\boldsymbol{c}\right\},\\ \boldsymbol{f}_{c,ki,8}'(\boldsymbol{\theta}) &= \frac{\partial f_{c,ki,8}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}} = \mathbf{z}_k\left(\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\right)\\ &\quad+\mathbf{z}_k\left(\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k\right)\\ &\quad+2\,\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^2\overline{\mathbf{H}}_2\odot\left(\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\right)^T\right)\boldsymbol{c}\right\},\end{aligned} \tag{103}$$

$$\begin{aligned}&\boldsymbol{f}_{c,ki,9}'(\boldsymbol{\theta}) = \frac{\partial f_{c,ki,9}(\boldsymbol{\Phi})}{\partial\boldsymbol{\theta}}\\ &= \mathbf{z}_k\left(\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^H\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\right)\\ &+2\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^H\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\odot\mathbf{R}_{emi}\right)\boldsymbol{c}\right\}\\ &+\mathbf{z}_k\left(\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\right)\\ &+2\mathrm{Im}\left\{\boldsymbol{\Phi}^H\left(\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^H\overline{\mathbf{H}}_2\odot\left(\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\right)^T\right)\boldsymbol{c}\right\}.\end{aligned} \tag{104}$$

Proof: It follows by applying the chain rule to compute the derivatives and using Lemma 4 and 5. Consider $\boldsymbol{f}_{c,ki,3}'(\boldsymbol{\theta})$ as an example. By applying the chain rule, we have

$$\begin{aligned}\boldsymbol{f}_{c,k,3}'(\boldsymbol{\theta}) &= \frac{\partial \mathrm{Tr}\left\{\boldsymbol{\Upsilon}_k\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\right\}}{\partial\boldsymbol{\theta}}\\ &= \left.\frac{\partial \mathrm{Tr}\left\{\mathbf{T}\boldsymbol{\Upsilon}_k\right\}}{\partial\boldsymbol{\theta}}\right|_{\mathbf{T}=\boldsymbol{\Upsilon}_k\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H}\\ &+\left.\frac{\partial \mathrm{Tr}\left\{\mathbf{T}\boldsymbol{\Upsilon}_k\right\}}{\partial\boldsymbol{\theta}}\right|_{\mathbf{T}=\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k}\\ &+\left.\frac{\partial \mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial\boldsymbol{\theta}}\right|_{\mathbf{A}=\overline{\mathbf{H}}_2^H\boldsymbol{\Upsilon}_k^2\overline{\mathbf{H}}_2,\mathbf{B}=\mathbf{R}_{emi}}.\end{aligned} \tag{105}$$

The proof follows by applying Lemma 4 and 5. The other terms can be obtained similarly. ■

$$\begin{aligned}&\frac{\partial \mathrm{SINR}_{c,k}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\\ &= \frac{p\frac{\partial E_{c,k}^{\mathrm{signal}}}{\partial\boldsymbol{\theta}}}{pE_{c,k}^{\mathrm{leak}}+p\sum_{i=1,i\neq k}^K I_{c,ki}+\sigma_e^2E_{c,k}^{\mathrm{emi}}+\sigma^2E_{c,k}^{\mathrm{noise}}} - pE_{c,k}^{\mathrm{signal}}\frac{p\frac{\partial E_{c,k}^{\mathrm{leak}}}{\partial\boldsymbol{\theta}}+p\sum_{i=1,i\neq k}^K\frac{\partial I_{c,ki}}{\partial\boldsymbol{\theta}}+\sigma_e^2\frac{\partial E_{c,k}^{\mathrm{emi}}}{\partial\boldsymbol{\theta}}+\sigma^2\frac{\partial E_{c,k}^{\mathrm{noise}}}{\partial\boldsymbol{\theta}}}{\left(pE_{c,k}^{\mathrm{leak}}+p\sum_{i=1,i\neq k}^K I_{c,ki}+\sigma_e^2E_{c,k}^{\mathrm{emi}}+\sigma^2E_{c,k}^{\mathrm{noise}}\right)^2}.\end{aligned} \tag{97}$$

Algorithm 1 Accelerated Gradient Ascent Algorithm

1: Initialize $\boldsymbol{\theta}_0$ randomly, $i=0$, $a_0=1$, $\boldsymbol{x}_{-1}=\boldsymbol{\theta}_0$;
2: **while** 1 **do**
3: Calculate the gradient vector $\boldsymbol{f}_c'(\boldsymbol{\theta}_i)=\frac{\partial f_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_i}$;
4: Obtain the step size κ_i based on the backtracking line search;
5: $\boldsymbol{x}_i=\boldsymbol{\theta}_i+\kappa_i \boldsymbol{f}_c'(\boldsymbol{\theta}_i)$;
6: $a_{i+1}=(1+\sqrt{4a_i^2+1})/2$;
7: $\boldsymbol{\theta}_{i+1}=\boldsymbol{x}_i+(a_i-1)(\boldsymbol{x}_i-\boldsymbol{x}_{i-1})/a_{i+1}$;
8: **if** $f_c(\boldsymbol{\theta}_{i+1})-f_c(\boldsymbol{\theta}_i)<10^{-4}$ **then**
9: $\boldsymbol{\theta}^*=\boldsymbol{\theta}_{i+1}$, break;
10: **end if**
11: $i=i+1$;
12: **end while**

Therefore, the gradient of $\frac{\partial f_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ in (96) follows from (97), Lemmas 4, 5, 6 and by applying the chain rule. For example, we have

$$\frac{\partial E_{c,k}^{\text{signal}}}{\partial \boldsymbol{\theta}}=\frac{\partial\left\{\left(E_{c,k}^{\text{noise}}\right)^2\right\}}{\partial \boldsymbol{\theta}}=2E_{c,k}^{\text{noise}}\frac{\partial E_{c,k}^{\text{noise}}}{\partial \boldsymbol{\theta}}, \tag{106}$$

and

$$\frac{\partial E_{c,k}^{\text{noise}}}{\partial \boldsymbol{\theta}}=M\widehat{c}_k\delta \boldsymbol{f}_{c,k,7}'(\boldsymbol{\theta})+\{\widehat{c}_k f_{c,k,2}(\boldsymbol{\Phi})+\gamma_k\}\,\mathbf{z}_k\left(\mathbf{I}_M\right) + \widehat{c}_k\,\text{Tr}\left\{\boldsymbol{\Upsilon}_k\right\}\boldsymbol{f}_{c,k,2}'(\boldsymbol{\theta}). \tag{107}$$

All the other terms in $\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\frac{\partial f_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ can be obtained similarly to (106). For brevity, the final analytical expressions of $\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\frac{\partial f_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ are given in the extended version of the present paper [75, Appendix K]. It is known that gradient-based methods may have a slow convergence rate. To tackle this issue, we apply Nesterov's accelerated gradient method, which effectively increases the convergence speed of the gradient method [76]. For completeness, the algorithm for optimizing $f_c(\boldsymbol{\theta})$ is presented in Algorithm 1 where steps 6-7 correspond to Nesterov's acceleration method.

VII. Numerical Results

In this section, we evaluate the performance of RIS-aided massive MIMO systems and validate the impact of key system parameters unveiled in the previous sections. We first consider a typical RIS-aided scenario where an RIS is deployed in close proximity to some cell-edge users. In this case, the direct links are relatively weak, and therefore an RIS may improve the end-to-end system performance. Accordingly, we assume that $K=8$ users are evenly distributed on a semicircle centered at the RIS and of radius $d_{UI}=20$ m. The distance between the RIS and the BS is $d_{IB}=700$ m. The distance between the user k and the BS is obtained from the network topology, i.e., $\left(d_k^{\text{UB}}\right)^2=\left(d_{\text{IB}}-d_{\text{UI}}\cos\left(\frac{\pi}{9}k\right)\right)^2+\left(d_{\text{UI}}\sin\left(\frac{\pi}{9}k\right)\right)^2$. The path-loss exponent of the direct links is larger than the path-loss exponent of the RIS-assisted links in order to characterize the more severe signal attenuation due to the presence of blocking objects on the ground. Specifically, we set the distance-dependent path-loss factors equal to $\alpha_k=10^{-3}d_{\text{UI}}^{-2}$, $\beta=10^{-3}d_{\text{IB}}^{-2.5}$ and $\gamma_k=10^{-3}\left(d_k^{\text{UB}}\right)^{-4}, \forall k$. The number of symbols in each channel coherence time interval is $\tau_c=196$ [61], [62], and $\tau=K=8$ symbols are utilized for channel estimation. The noise power is $\sigma^2=-104$ dBm (corresponding to a noise spectral density equal to -174 dBm/Hz over a bandwidth of 10 MHz). The other simulation parameters (unless stated otherwise) are listed in Table IV.

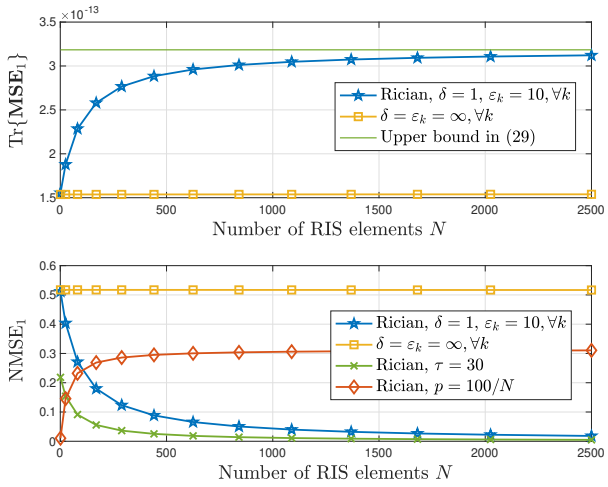

Fig. 2: MSE and NMSE of user 1 versus the number of RIS elements.

A. Spatially independent Channels in the Absence of EMI

We first validate the obtained analytical results by assuming that the channels are spatially independent and the EMI is not present. This help us obtain initial but useful insights on the performance offered by RIS-aided systems thanks to the simpler analytical expressions of the rate and the explicit analytical insights obtained in Section IV. Specifically, the analytical results are obtained by using Theorem 2 and related corollaries. The Monte Carlo simulations, which are referred to as "Simulation" in the legends of the figures, are obtained from (35) by averaging over 10^5 random channel realizations. The phase shifts are obtained by solving Problem (95) with respect to $f(\boldsymbol{\theta})$.

1) Quality of the LMMSE Channel Estimation: To begin with, we investigate the MSE and NMSE of the proposed channel estimation scheme. The MSE and NMSE of the channel estimation algorithm of the k-th user are characterized through the functions $\text{Tr}\left\{\mathbf{MSE}_k\right\}$ and NMSE_k, respectively. Without loss of generality, Fig. 2 illustrates the MSE and NMSE of user 1 versus the number of RIS elements N. In general Rician channels, we observe that the MSE is an increasing function of N while the NMSE is a decreasing function of N, which is consistent with Corollaries 1, 2 and 3. This is because the number of communication paths increases with N, but the pilot length τ does not increase correspondingly, which increases the estimation error. However, the intensity of the channel gains increases with N, which, in turn, decreases the normalized errors. In purely LoS

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3227538

TABLE IV: Simulation parameters.

$(\varphi_t^a, \varphi_t^e)$	$(4.17, 0.09)$	(ϕ_r^a, ϕ_r^e)	$(6.28, 4.21)$
$(\varphi_{1r}^a, \varphi_{1r}^e)$	$(5.20, 4.32)$	$(\varphi_{2r}^a, \varphi_{2r}^e)$	$(0.41, 2.52)$
$(\varphi_{3r}^a, \varphi_{3r}^e)$	$(3.84, 1.78)$	$(\varphi_{4r}^a, \varphi_{4r}^e)$	$(1.35, 4.15)$
$(\varphi_{5r}^a, \varphi_{5r}^e)$	$(5.08, 5.76)$	$(\varphi_{6r}^a, \varphi_{6r}^e)$	$(4.75, 1.56)$
$(\varphi_{7r}^a, \varphi_{7r}^e)$	$(4.74, 5.36)$	$(\varphi_{8r}^a, \varphi_{8r}^e)$	$(0.09, 1.40)$
BS antennas	$M = 64$	RIS elements	$N = 64$
Transmit power	$p = 30$ dBm	Antenna spacing	$d_{bs} = \lambda/2$
Rician factors	$\delta = 1$, $\varepsilon_k = 10, \forall k$	Approximation factor	$\mu = 100$

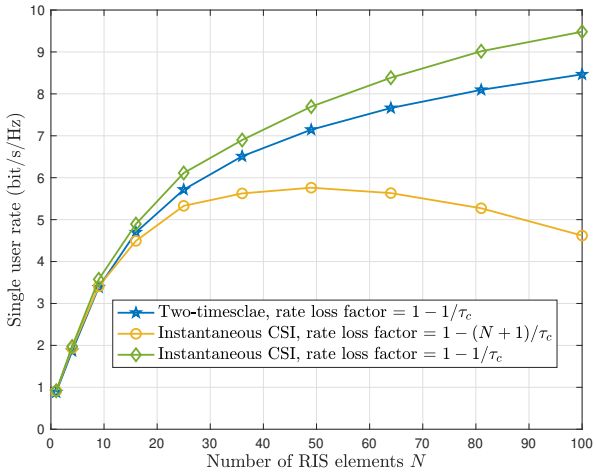

Fig. 3: Comparison of the two-timescale design and instantaneous CSI-based design.

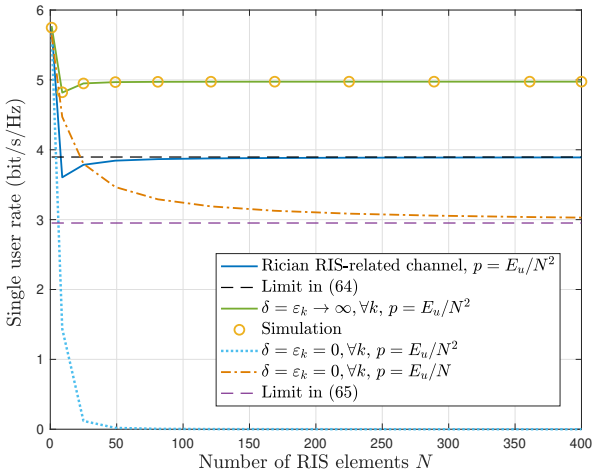

Fig. 4: Rate versus N in a single-user system. The transmit power is scaled as $p = E_u/N^2$ or $p = E_u/N$, where $E_u = 20$ dB.

RIS-assisted channels ($\delta = \varepsilon_k \to \infty$), the MSE and NMSE are, on the other hand, independent of N. This is because LoS channels are deterministic, and therefore do not introduce additional estimation errors. Also, we see that the MSE tends to an upper bound but the NMSE tends to zero when $N \to \infty$, which validates Corollary 1 and 2. By increasing the length of the pilot signals from 8 to 30, we see that the NMSE decreases. However, the NMSE that is obtained for $\tau = 30$ can also be obtained for $\tau = 8$ but by using a larger value for N. This validates our remark that increasing the RIS elements can play a similar role as increasing τ. Finally, we see that the NMSE tends to a limit less than 1 when the transmit power is scaled proportionally to $p = 100/N$, as $N \to \infty$. This validates the correctness of (26).

2) Single-user Case: Next, we evaluate the ergodic achievable rate in the single-user scenario, where only user 1 is present.

In Fig. 3, we compare the proposed two-timescale scheme with the conventional instantaneous CSI-based scheme. For brevity, the detailed implementation of the instantaneous CSI-based scheme is presented in the companion extended version of the present paper [75, Appendix J]. By assuming the same rate loss factor (ideal but not achievable), it is seen that the instantaneous CSI-based scheme outperforms the proposed two-timescale scheme, especially when N is large. This is because the LoS and NLoS channel components are both exploited in the instantaneous CSI-based RIS design. By contrast, the fast-fading NLoS channel information is averaged out in the proposed statistical CSI-based RIS design. When considering the actual channel estimation overhead, however, the proposed scheme outperforms the instantaneous CSI-based scheme. This is because the instantaneous CSI-based scheme requires a longer pilot length, which is proportional to N, even though it results in a higher SNR. When N is large, the instantaneous CSI-based scheme needs a large number of time slots to transmit the pilot sequence, and then only a few symbols are left for data transmission. As a result of the high estimation overhead, the instantaneous CSI-based scheme incurs in a rate loss, which leads to a severe decrease of the rate in the large N regime. Therefore, Fig. 3 validates the effectiveness of the proposed two-timescale scheme.

In Fig. 4, we illustrate the power scaling law as a function of N in a single-user scenario. In agreement with Corollary 9, the rate converges to a limit if we reduce the power proportionally to $1/N^2$ in Rician fading channels. Also, the limit is maximized in LoS-only RIS-assisted channels ($\delta = \varepsilon_k \to \infty$). In NLoS-only RIS-assisted channels ($\delta = \varepsilon_k = 0$), scaling the power proportionally to $1/N^2$ reduces the rate to zero. As proved in Corollary 10, in NLoS-only RIS-assisted channels, the power can only be scaled proportionally to $1/N$ for maintaining a non-zero rate. These observations highlight that LoS environments are preferable for the deployment of RIS-

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3227538

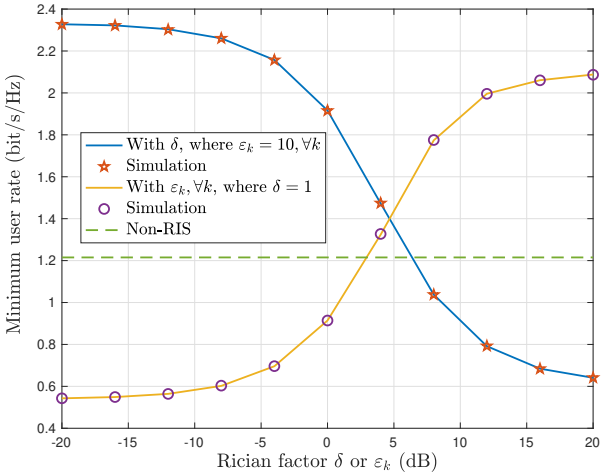

Fig. 5: Minimum user rate versus the Rician factor δ or $\varepsilon_k, \forall k$.

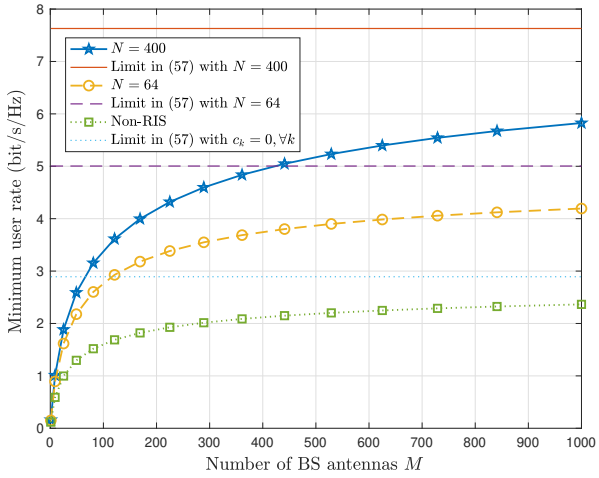

Fig. 7: Minimum user rate versus M when $\delta = 0$. The transmit power is scaled as $p = E_u/\sqrt{M}$, where $E_u = 10$ dB.

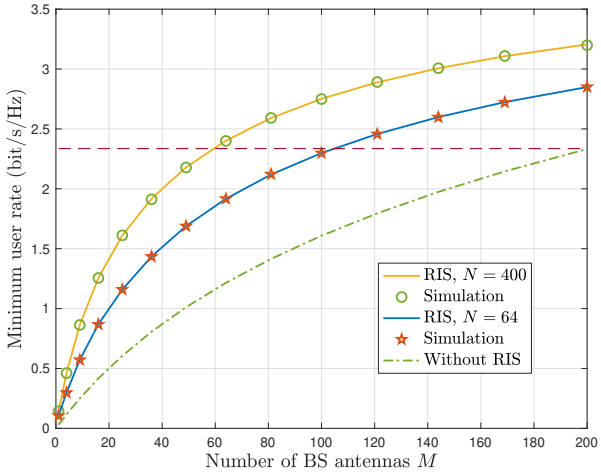

Fig. 6: Minimum user rate versus M.

aided single-user systems.

3) Multi-user Case: In Figs. 5-8, we evaluate the performance of RIS-aided systems in the general multi-user scenario.

Fig. 5 shows the impact of the Rician factors. It can be observed that the achievable rate is a decreasing function of δ but an increasing function of $\varepsilon_k, \forall k$. This is because the rank of the LoS component $\overline{\mathbf{H}}_2$ between the RIS and the BS is 1, while the rank of the LoS component $\overline{\mathbf{H}}_1$ between the users and the RIS is not. When $\delta \to \infty$, the rank of the RIS-BS channel tends to 1, which leads to a rank-1 cascaded user-RIS-BS channel. As a result, the RIS-assisted channel becomes rank-deficient, which cannot effectively sustain the transmission of multiple users simultaneously. It is known that the RIS should be deployed either near the BS or near the users so that the product pathloss effect is mitigated [69]. In addition, Fig. 5 provides some suggestions with respect to the spatial diversity gain provided by the deployment of an RIS. To increase ε_k, it is beneficial to install the RIS at a certain height with respect to the ground, which results in increasing the strength of the LoS components of the RIS-user channels. Besides, it is necessary to guarantee a high-rank RIS-BS channel. This condition holds for small values of δ under the considered Rician fading model. Since small values of δ are typically obtained when the RIS is deployed far away from the BS, it is still a good choice to place the RIS near the users after taking into consideration the impact of spatial diversity. On the contrary, if the RIS is deployed near the BS, δ could be large and the BS-RIS channel could be rank-deficiency under the considered Rician fading model. In this case, possible options for increasing the rank of the channel may be the deployment of artificial scatterers between the BS and the RIS or placing the RIS very close to the BS so that the spherical wave model is valid [46].

In Fig. 6, we evaluate the rate as a function of the number of BS antennas. The figure illustrates the impact of deploying an RIS in conventional massive MIMO systems. It is observed that the deployment of an RIS effectively improves the rate, and the improvement increases with the number of RIS elements. It is worth nothing that this performance gain is obtained by using a simple MRC receiver at the BS, and that the LMMSE channel estimator requires the same amount of overhead as conventional massive MIMO systems. With the help of an RIS, we can achieve the same rate as conventional massive MIMO systems, but with a much smaller number of BS antennas. In particular, the rate obtained by a 200-antenna BS in conventional massive MIMO systems can be obtained by a 100-antenna BS in RIS-aided massive MIMO systems with $N = 64$ RIS elements. The number of BS antennas can be further decreased to $M = 64$ if the number of RIS elements is increased to $N = 400$. Since the cost and energy consumption of one RIS element is much lower than that of one BS antenna, we conclude that the integration of RISs in conventional massive MIMO systems is a promising and cost-effective solution for future wireless communication systems.

In Fig. 7 and Fig. 8, finally, we investigate the power scaling law over a purely NLoS RIS-BS channel ($\delta = 0$) and a purely

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3227538

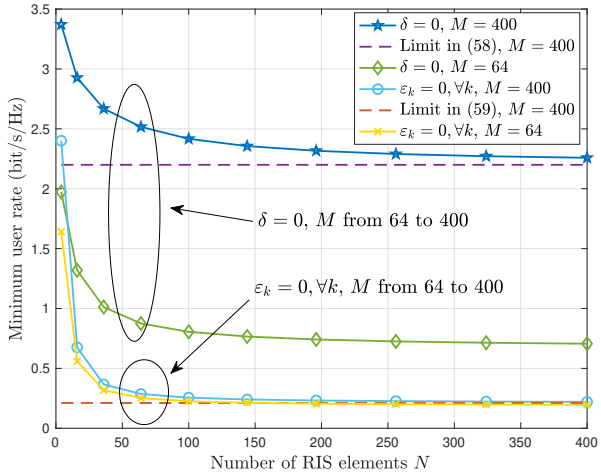

Fig. 8: Minimum user rate versus N when $\delta = 0$ or $\varepsilon_k = 0$. The transmit power is scaled as $p = E_u/N$, where $E_u = 10$ dB.

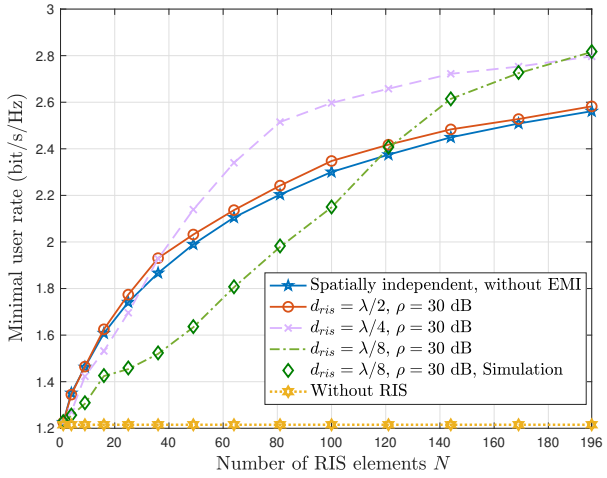

Fig. 9: Achievable rate versus N for different values of the RIS element spacing d_{ris}.

NLoS user-RIS channels ($\varepsilon_k = 0, \forall k$). In Fig. 7, the transmit power is scaled proportionally to $1/\sqrt{M}$ for the NLoS RIS-BS channel ($\delta = 0$). In agreement with Corollary 6, if $\delta = 0$, the rate can be maintained to a non-zero value when the power is scaled proportionally to $1/\sqrt{M}$ as $M \to \infty$. Compared with conventional massive MIMO systems, the deployment of an RIS effectively improves the asymptotic limit when $M \to \infty$, and the rate gain could be further improved by increasing N.

In Fig. 8, the transmit power is scaled proportionally to $1/N$ over a purely NLoS RIS-BS channel ($\delta = 0$) or purely NLoS user-RIS channels ($\varepsilon_k = 0, \forall k$). For $N \to \infty$, the rate maintains a non-zero value, which is consistent with Corollaries 7 and 8. Besides, in agreement with Corollary 7, the asymptotic limit for $\delta = 0$ when $N \to \infty$ can be significantly improved by increasing the number of BS antennas from $M = 64$ to $M = 400$. This is because the RIS-BS channel has a high rank if $\delta = 0$, which decreases the spatial correlation among the users and mitigates the multi-user interference. Furthermore, in agreement with Corollary 8, the asymptotic limit for $\varepsilon_k = 0, \forall k$ when $N \to \infty$ only marginally increases when increasing M from 64 to 400. This observation confirms once again that guaranteeing a high spatial diversity between the RIS and the BS could offer a good rate in RIS-aided massive MIMO systems.

B. Spatially correlated Channels in the Presence of EMI

The results illustrated in Figs. 2-8 showcase the gain of RIS over spatially independent channels and in the absence of EMI. In this section, some numerical examples are presented to explore the impact of spatial correlation and EMI and study under what conditions the spatial correlation and the EMI can be ignored as a function of the inter-distance between the RIS elements and the strength of the EMI. Specifically, the strength of EMI with respect to the thermal noise at the BS is characterized by the following ratio [58]

$$\rho = \frac{\sigma_e^2}{\sigma^2}. \tag{108}$$

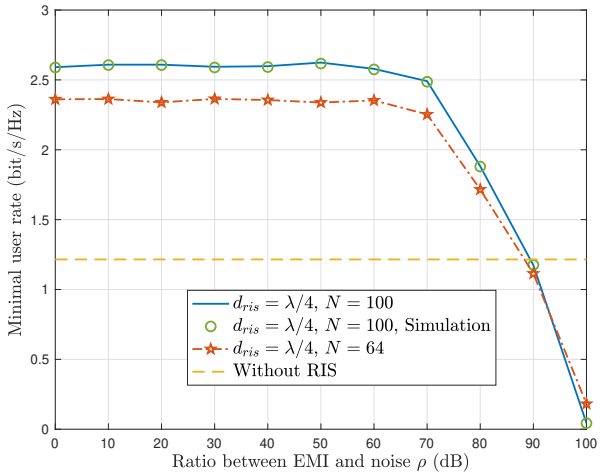

Fig. 10: Impact of the EMI.

Fig. 9 illustrates the impact of channel spatial correlation, which is due to the sub-wavelength spacing between the RIS elements, on the achievable rate. In this context, the objective function of the optimization problem (92) is $\underline{R}_{c,k}(\mathbf{\Phi})$ where the impact of spatial correlation is taken into account in the design of the RIS phase shifts. First, as expected, we see that the impact of spatial correlation can be safely ignored when the inter-distance between the RIS elements is half of the wavelength ($d_{ris} = \lambda/2$) and the EMI is light ($\rho = 30$ dB). This confirms that the analytical insights drawn in Section IV over spatially independent channels and in the absence of EMI are meaningful to understand the fundamental performance limits of RIS-aided systems in practically relevant scenarios. As the spacing between the RIS elements decreases ($d_{ris} = \lambda/4$, $\lambda/8$), however, the spatial correlation cannot be ignored and it has a non-negligible impact on the rate. Specifically, we identify two operating regions: (i) small values of RIS elements N and (ii) large values of RIS elements N.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3227538

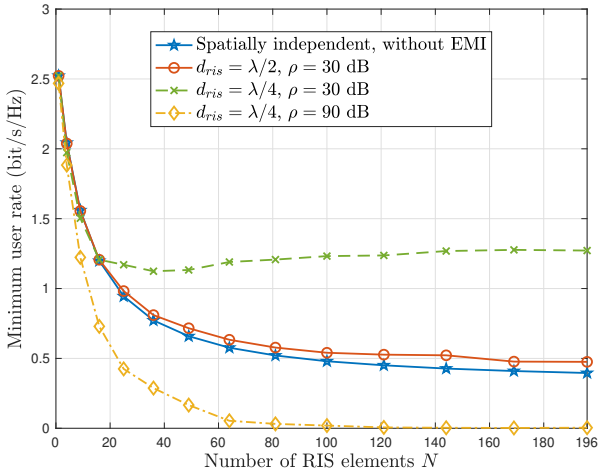

Fig. 11: Achievable rate when the power is scaled proportionally to $p = 10/N$.

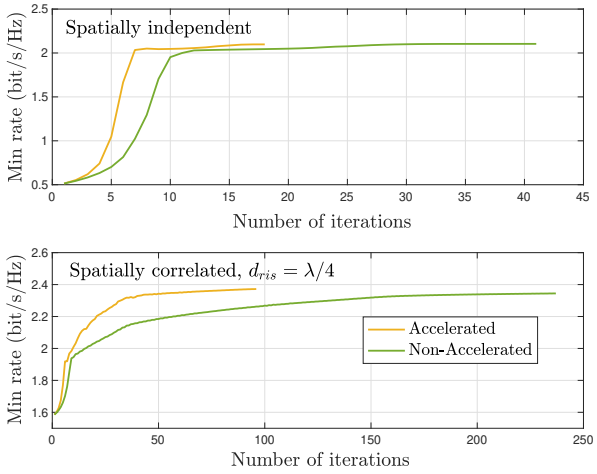

Fig. 12: Convergence behavior for spatially independent and spatially correlated cases, $M = N = 64$.

For small values of N, the rate decreases as the inter-distance decreases. This is attributed to the decrease of the channel rank. For large values of N, the channel rank still decreases but we can leverage the large number of RIS elements and the greater ability of an RIS to customize the wireless channels in the presence of channel correlation, as discussed in Theorem 3 and Remark 3. For large values of N, the beamforming gains provided by optimizing RIS outweigh the negative impact of spatial correlation, which in turn results in a better achievable rate.

The impact of EMI is studied in Fig. 10. When the power of the EMI is sufficiently small with respect to the noise ($\rho < 60$ dB), the impact of the EMI on the achievable rate is negligible. This is attributed to the strong multi-user interference when using MRC. As a result, when the EMI is mild, its impact is negligible as compared with the multi-user interference. As ρ increases, the EMI becomes more severe, and it eventually becomes the dominant contribution. For large values of the EMI, RIS-aided systems may even perform worse than conventional massive MIMO systems.

Fig. 11 illustrates the power scaling laws as a function of the channel spatial correlation and EMI. Specifically, Fig. 11 shows the achievable rate when the power is scaled as $p = 10/N$. The figures validate Corollary 11: if the EMI is mild, the power scaling law as a function of the transmit power is confirmed. On the other hand, it does not hold anymore in the presence of strong EMI. As a function of the inter-distance d_{ris}, Fig. 11 is in agreement with Fig. 9.

In Fig. 12, we study the convergence behavior of the proposed accelerated gradient method compared with its non-accelerated counterpart. By applying the proposed acceleration method, it can be observed that the speed of convergence is effectively improved. In the presence of spatially independent channels, the algorithm converges very quickly due to the simple expression of the achievable rate. By contrast, when considering spatial correlation of $d_{ris} = \lambda/4$, the expression of the rate becomes more complex and the optimization variable

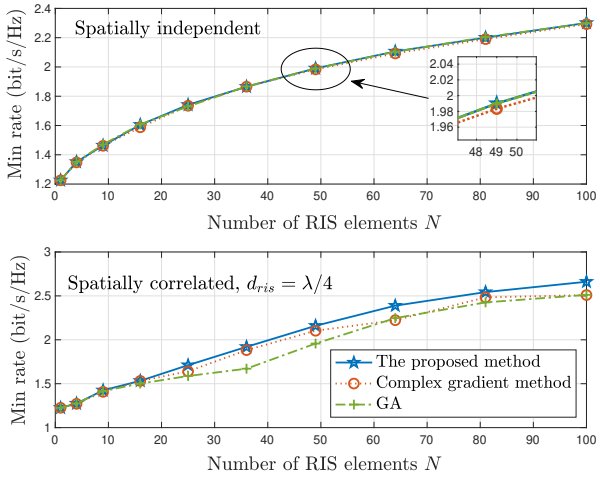

Fig. 13: Performance comparison between different optimization algorithms.

$\boldsymbol{\Phi}$ appears more frequently, as discussed in Remark 3. As a result, the number of iterations needed for convergence increases. Nevertheless, it can be observed that the accelerated gradient algorithm converges within 100 iterations even though the number of optimization variables is 64.

Fig. 13 compares the performance of the proposed method with two benchmark algorithms, i.e., the genetic algorithm (GA) [49] and the gradient ascent method formulated in terms of the complex variables $\boldsymbol{c} = e^{j\boldsymbol{\theta}}$ [46]. In the absence of spatial correlation, it can be observed that the three algorithms provide almost the same performance. This is because the objective function possesses a simple and tractable form. Nevertheless, the proposed algorithm performs slightly better than the gradient ascent method applied to complex-valued variables. This is because the proposed method treats the angles as optimization variables and therefore avoids the performance loss due to the projection operation. In the presence of spatial correlation, the

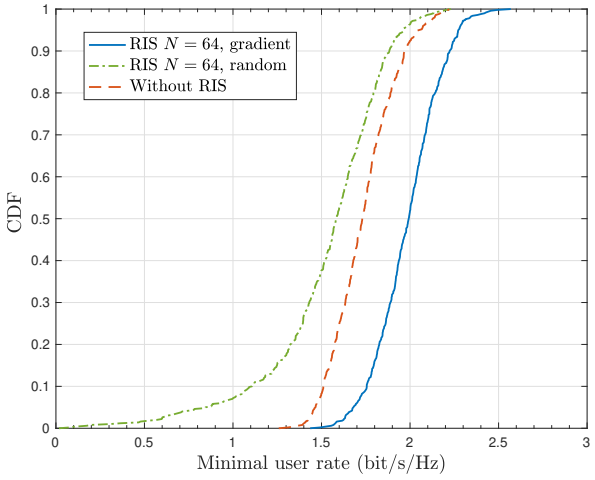

Fig. 14: CDF of the minimal user rate for RIS-aided and RIS-free systems, $d_{ris} = \lambda/4$, $\rho = 30$ dB, $p = 20$ dBm.

objective function of the optimization problem becomes more complex. In this case, it can be seen that the proposed method outperforms the other two methods especially when N is large.

C. Randomly Distributed Users in a Large Area

The numerical results illustrated in the previous figures are obtained by assuming that the RIS is deployed near the cell-edge users. In this subsection we examine the case study in which the users are distributed over a large area and the transmit power may not be very high due to the deployment of many BS antennas. We set the transmit power equal to $p = 20$ dBm and assume that the users are randomly distributed in a 100 m $\times$ 100 m area identified by the coordinates $(200\text{ m}, 0)$ to $(300\text{ m}, 100)$ [54]. The BS and the RIS are deployed in $(0, 0)$ and $(200\text{ m}, 0)$, respectively. Also, we assume $d_{ris} = \lambda/4$ and $\rho = 30$ dB for the spacing between the RIS elements and the EMI, respectively.

In Fig. 14, we illustrate the average rate of RIS-aided systems as a function of 500 random locations of the users and compare it against the rate provided by conventional massive MIMO systems. We observe that the deployment of an RIS still provides some performance gains, but these are reduced as compared to the optimized deployment of the RIS near the cell-edge users. As expected, in addition, the achievable rate is low if the RIS phase shifts matrix is not optimized with the gradient ascent algorithm, but the phase shifts are randomly set.

VIII. Conclusion

This paper investigated the two-timescale design for RIS-aided massive MIMO systems by taking into account the impact of channel estimation errors. We first considered a spatially independent channel model in the absence of EMI, and we then extended the study to a spatially correlated channel model in the presence of EMI. In both cases, we obtained the LMMSE channel estimator for the user-BS aggregated channels, employed the MRC detector, derived the UatF bound of the achievable rate, and optimized the phase shifts of the RIS based on a gradient ascent method. To gain a better understanding of the performance offered by RIS-aided systems, we unveiled fundamental scaling laws over spatially independent channel models. We demonstrated that the transmit power can be reduced proportionally to $1/M$, while maintaining a non-zero rate, as $M \to \infty$, over RIS-BS Rician channels. If the RIS-BS channel is Rayleigh distributed, on the other hand, a non-zero rate can be maintained when the power is reduced proportionally to $1/\sqrt{M}$ as $M \to \infty$ or proportionally to $1/N$ as $N \to \infty$. Over spatially correlated channels and in the presence of EMI, we demonstrated that the presence of spatial correlation is beneficial in terms of shaping the wireless channels. We also found that it is beneficial to place the RIS close to the cell-edge users to compensate for the product path-loss law in the far-field region. Finally, we proved that the scaling laws in the absence of EMI may not be preserved in the presence of EMI, especially if the EMI is strong enough.[6]

Appendix A
Some Useful Results

Lemma 7 *Consider a matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$, $m, n \geq 1$, whose entries are i.i.d. random variables with zero mean and v_x variance. Consider a deterministic matrix $\mathbf{W} \in \mathbb{C}^{n\times n}$. Then, we have*

$$\mathbb{E}\left\{\mathbf{X}\mathbf{W}\mathbf{X}^H\right\} = v_x \operatorname{Tr}\{\mathbf{W}\}\mathbf{I}_m. \tag{109}$$

Proof: See [75]. ■

Lemma 8 *Consider the deterministic matrices $\mathbf{W} \in \mathbb{C}^{N\times N}$ and vectors $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{C}^{N\times 1}$, and $\mathbf{w}_3, \mathbf{w}_4 \in \mathbb{C}^{M\times 1}$. Then, we have*

$$\mathbb{E}\left\{\tilde{\mathbf{H}}_2\mathbf{W}\tilde{\mathbf{H}}_2\right\} = \mathbb{E}\left\{\operatorname{Re}\left\{\tilde{\mathbf{H}}_2\mathbf{W}\tilde{\mathbf{H}}_2\right\}\right\} = \mathbf{0}, \tag{110}$$

$$\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{w}_1\tilde{\mathbf{h}}_k^H\mathbf{w}_2\right\} = \mathbb{E}\left\{\operatorname{Re}\left\{\tilde{\mathbf{h}}_k^H\mathbf{w}_1\tilde{\mathbf{h}}_k^H\mathbf{w}_2\right\}\right\} = 0, \tag{111}$$

$$\mathbb{E}\left\{\mathbf{w}_3^H\tilde{\mathbf{d}}_k\mathbf{w}_4^H\tilde{\mathbf{d}}_k\right\} = \mathbb{E}\left\{\operatorname{Re}\left\{\mathbf{w}_3^H\tilde{\mathbf{d}}_k\mathbf{w}_4^H\tilde{\mathbf{d}}_k\right\}\right\} = 0. \tag{112}$$

Proof: See [75]. ■

Lemma 9 *For deterministic matrices $\mathbf{C} \in \mathbb{C}^{M\times M}$ and $\mathbf{W} \in \mathbb{C}^{N\times N}$, if $\mathbf{C} = \mathbf{C}^H$, we obtain*

$$\begin{aligned}&\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H\mathbf{C}\tilde{\mathbf{H}}_2\mathbf{W}\tilde{\mathbf{H}}_2^H\mathbf{C}\tilde{\mathbf{H}}_2\right\}\\ &= \operatorname{Tr}\{\mathbf{W}\}\operatorname{Tr}\left\{\mathbf{C}^2\right\}\mathbf{I}_N + |\operatorname{Tr}\{\mathbf{C}\}|^2\mathbf{W},\end{aligned} \tag{113}$$

$$\begin{aligned}&\mathbb{E}\left\{\tilde{\mathbf{H}}_{c,2}^H\mathbf{C}\tilde{\mathbf{H}}_{c,2}\mathbf{W}\tilde{\mathbf{H}}_{c,2}^H\mathbf{C}\tilde{\mathbf{H}}_{c,2}\right\}\\ &= \operatorname{Tr}\left\{\mathbf{R}_{ris}\mathbf{W}\right\}\operatorname{Tr}\left\{\mathbf{C}^2\right\}\mathbf{R}_{ris} + |\operatorname{Tr}\{\mathbf{C}\}|^2\mathbf{R}_{ris}\mathbf{W}\mathbf{R}_{ris}.\end{aligned} \tag{114}$$

[6]For brevity, the appendices provide a sketch of the proofs of the main results of the present paper. The interested readers may find the detailed proof in the companion extended version of the present paper [75].

Proof: Define $\tilde{\mathbf{H}}_2 = [\mathbf{J}_1, \ldots, \mathbf{J}_N]$, where $\mathbf{J}_n \in \mathbb{C}^{M\times 1}$, $1 \le n \le N$, are independent of each other, and $\mathbf{J}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Denoting $[\mathbf{W}]_{m,n} = w_{mn}$, then we have

$$\left[\tilde{\mathbf{H}}_2^H \mathbf{C}\tilde{\mathbf{H}}_2 \mathbf{W}\tilde{\mathbf{H}}_2^H \mathbf{C}\tilde{\mathbf{H}}_2\right]_{i,j} = \sum_{h=1}^{N}\sum_{m=1}^{N} \mathbf{J}_i^H \mathbf{C}\mathbf{J}_m w_{mh} \mathbf{J}_h^H \mathbf{C}\mathbf{J}_j. \tag{115}$$

Note that $\mathbb{E}\left\{\left|\mathbf{J}_i^H \mathbf{C}\mathbf{J}_i\right|^2\right\} = |\operatorname{Tr}(\mathbf{C})|^2 + \operatorname{Tr}\left(\mathbf{C}^2\right)$ [77, (35)]. The expectation of the i-th diagonal term can be calculated as

$$\begin{aligned}
&\mathbb{E}\left\{\left[\tilde{\mathbf{H}}_2^H \mathbf{C}\tilde{\mathbf{H}}_2 \mathbf{W}\tilde{\mathbf{H}}_2^H \mathbf{C}\tilde{\mathbf{H}}_2\right]_{i,i}\right\} \\
&= \mathbb{E}\left\{\mathbf{J}_i^H \mathbf{C}\mathbf{J}_i w_{ii} \mathbf{J}_i^H \mathbf{C}\mathbf{J}_i\right\} \\
&\quad + \mathbb{E}\left\{\sum_{m=1, m\neq i}^{N} \mathbf{J}_i^H \mathbf{C}\mathbf{J}_m w_{mm} \mathbf{J}_m^H \mathbf{C}\mathbf{J}_i\right\} \\
&= w_{ii}\mathbb{E}\left\{\left|\mathbf{J}_i^H \mathbf{C}\mathbf{J}_i\right|^2\right\} \\
&\quad + \mathbb{E}\left\{\sum_{m=1, m\neq i}^{N} w_{mm} \mathbf{J}_i^H \mathbf{C}\mathbb{E}\left\{\mathbf{J}_m \mathbf{J}_m^H\right\} \mathbf{C}\mathbf{J}_i\right\} \\
&= w_{ii}|\operatorname{Tr}(\mathbf{C})|^2 + \operatorname{Tr}\{\mathbf{W}\}\operatorname{Tr}\left\{\mathbf{C}^2\right\}.
\end{aligned} \tag{116}$$

The expectation of the (i,j)-th non-diagonal term is given by

$$\begin{aligned}
&\mathbb{E}\left\{\left[\tilde{\mathbf{H}}_2^H \mathbf{C}\tilde{\mathbf{H}}_2 \mathbf{W}\tilde{\mathbf{H}}_2^H \mathbf{C}\tilde{\mathbf{H}}_2\right]_{i,j}\right\} = \mathbb{E}\left\{\mathbf{J}_i^H \mathbf{C}\mathbf{J}_i w_{ij} \mathbf{J}_j^H \mathbf{C}\mathbf{J}_j\right\} \\
&= w_{ij}\mathbb{E}\left\{\mathbf{J}_i^H \mathbf{C}\mathbf{J}_i\right\}\mathbb{E}\left\{\mathbf{J}_j^H \mathbf{C}\mathbf{J}_j\right\} = w_{ij}|\operatorname{Tr}\{\mathbf{C}\}|^2.
\end{aligned} \tag{117}$$

Combining (116) and (117) completes the proof of (113). Then, we prove (114) by using $\tilde{\mathbf{H}}_{c,2} = \tilde{\mathbf{H}}_2 \mathbf{R}_{ris}^{1/2}$. ■

Applying Lemma 9, we obtain some useful results summarized in the following Lemma.

Lemma 10 *For a deterministic matrix $\mathbf{W} \in \mathbb{C}^{N\times N}$, we have*

$$\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H \mathbf{A}_k \tilde{\mathbf{H}}_2 \mathbf{W}\tilde{\mathbf{H}}_2^H \mathbf{A}_k \tilde{\mathbf{H}}_2\right\} = e_{k1}^2 M^2 \mathbf{W} + e_{k3} M \operatorname{Tr}\{\mathbf{W}\}\mathbf{I}_N, \tag{118}$$

$$\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H \tilde{\mathbf{H}}_2 \mathbf{W}\tilde{\mathbf{H}}_2^H \tilde{\mathbf{H}}_2\right\} = M^2 \mathbf{W} + M \operatorname{Tr}\{\mathbf{W}\}\mathbf{I}_N, \tag{119}$$

$$\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H \mathbf{a}_M \mathbf{a}_M^H \tilde{\mathbf{H}}_2 \mathbf{W}\tilde{\mathbf{H}}_2^H \mathbf{a}_M \mathbf{a}_M^H \tilde{\mathbf{H}}_2\right\} = M^2 \mathbf{W} + M^2 \operatorname{Tr}\{\mathbf{W}\}\mathbf{I}_N. \tag{120}$$

where $\mathbf{A}_k$ is defined in (18), and e_{k1} and e_{k3} are defined in Lemma 2. For a deterministic matrix $\mathbf{C} \in \mathbb{C}^{M\times M}$ and a random vector $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, we have

$$\mathbb{E}\left\{\mathbf{u}\mathbf{u}^H \mathbf{C}\mathbf{u}\mathbf{u}^H\right\} = \mathbf{C} + \operatorname{Tr}\{\mathbf{C}\}\mathbf{I}_M, \tag{121}$$

$$\mathbb{E}\left\{\|\mathbf{u}\|^4\right\} = \operatorname{Tr}\left\{\mathbb{E}\left\{\mathbf{u}\mathbf{u}^H \mathbf{u}\mathbf{u}^H\right\}\right\} = M^2 + M. \tag{122}$$

APPENDIX B

Recalling the definition of $\mathbf{q}_k$ in (10), where $\tilde{\mathbf{H}}_2$, $\tilde{\mathbf{h}}_k$, $\tilde{\mathbf{d}}_k$, and $\mathbf{N}$ are independent of each other and composed of zero-mean entries, we have

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{y}_p^k\right\} &= \mathbb{E}\left\{\mathbf{q}_k\right\} + \frac{1}{\sqrt{\tau p}}\mathbb{E}\left\{\mathbf{N}\right\}\mathbf{s}_k = \mathbb{E}\left\{\mathbf{q}_k\right\} \\
&= \sqrt{c_k \delta \varepsilon_k}\,\overline{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k.
\end{aligned} \tag{123}$$

The covariance matrix between the unknown channel $\mathbf{q}_k$ and the observation vector $\mathbf{y}_p^k$ can be written as

$$\begin{aligned}
&\operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{y}_p^k\right\} \\
&= \mathbb{E}\left\{\left(\mathbf{q}_k - \mathbb{E}\left\{\mathbf{q}_k\right\}\right)\left(\mathbf{q}_k + \tfrac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k - \mathbb{E}\left\{\mathbf{q}_k\right\}\right)^H\right\} \\
&= \operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\},
\end{aligned} \tag{124}$$

and

$$\begin{aligned}
\operatorname{Cov}\left\{\mathbf{y}_p^k, \mathbf{q}_k\right\} &= \left(\operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{y}_p^k\right\}\right)^H \\
&= \left(\operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\}\right)^H = \operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\}.
\end{aligned} \tag{125}$$

Invoking the definition of $\mathbf{q}_k$, we obtain

$$\begin{aligned}
&\operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\} = \mathbb{E}\left\{\left(\mathbf{q}_k - \mathbb{E}\left\{\mathbf{q}_k\right\}\right)\left(\mathbf{q}_k - \mathbb{E}\left\{\mathbf{q}_k\right\}\right)^H\right\} \\
&= \mathbb{E}\Big\{c_k \delta \overline{\mathbf{H}}_2 \mathbf{\Phi} \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H \mathbf{\Phi}^H \overline{\mathbf{H}}_2^H + c_k \varepsilon_k \tilde{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k \overline{\mathbf{h}}_k^H \mathbf{\Phi}^H \tilde{\mathbf{H}}_2^H \\
&\qquad + c_k \tilde{\mathbf{H}}_2 \mathbf{\Phi} \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H \mathbf{\Phi}^H \tilde{\mathbf{H}}_2^H + \gamma_k \tilde{\mathbf{d}}_k \tilde{\mathbf{d}}_k^H\Big\} \\
&\overset{(b)}{=} N c_k \delta \mathbf{a}_M \mathbf{a}_M^H + \left(N c_k\left(\varepsilon_k + 1\right) + \gamma_k\right)\mathbf{I}_M,
\end{aligned} \tag{126}$$

where (b) exploits Lemma 7 and the mutual independence of $\tilde{\mathbf{H}}_2$ and $\tilde{\mathbf{h}}_k$.

Similarly, we have

$$\begin{aligned}
\operatorname{Cov}\left\{\mathbf{y}_p^k, \mathbf{y}_p^k\right\} &= \mathbb{E}\left\{\left(\mathbf{q}_k - \mathbb{E}\left\{\mathbf{q}_k\right\}\right)\left(\mathbf{q}_k - \mathbb{E}\left\{\mathbf{q}_k\right\}\right)^H\right\} \\
&\quad + \frac{1}{\tau p}\mathbb{E}\left\{\mathbf{N}\mathbf{s}_k \mathbf{s}_k^H \mathbf{N}^H\right\} \\
&= \operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{q}_k\right\} + \frac{\sigma^2}{\tau p}\mathbf{I}_M.
\end{aligned} \tag{127}$$

Finally, by introducing the auxiliary variables $a_{k1} = N c_k \delta$ and $a_{k2} = N c_k\left(\varepsilon_k + 1\right) + \gamma_k$, the proof is completed.

APPENDIX C

The LMMSE estimate of the channel $\mathbf{q}_k$ based on the observation vector $\mathbf{y}_p^k$ can be written as [72, Chapter 12.5]

$$\begin{aligned}
\hat{\mathbf{q}}_k &= \mathbb{E}\left\{\mathbf{q}_k\right\} \\
&\quad + \operatorname{Cov}\left\{\mathbf{q}_k, \mathbf{y}_p^k\right\}\operatorname{Cov}^{-1}\left\{\mathbf{y}_p^k, \mathbf{y}_p^k\right\}\left(\mathbf{y}_p^k - \mathbb{E}\left\{\mathbf{y}_p^k\right\}\right),
\end{aligned} \tag{128}$$

where the mean and covariance matrices are given in Lemma 1.

Let us compute $\operatorname{Cov}^{-1}\left\{\mathbf{y}_p^k, \mathbf{y}_p^k\right\}$. Using the Woodbury matrix identity [72, Page 571], we have

$$\begin{aligned}
\operatorname{Cov}^{-1}\left\{\mathbf{y}_p^k, \mathbf{y}_p^k\right\} &= \left(a_{k2} + \frac{\sigma^2}{\tau p}\right)^{-1}\mathbf{I}_M \\
&\quad - \frac{a_{k1}\left(a_{k2} + \frac{\sigma^2}{\tau p}\right)^{-2}}{1 + M a_{k1}\left(a_{k2} + \frac{\sigma^2}{\tau p}\right)^{-1}}\mathbf{a}_M \mathbf{a}_M^H.
\end{aligned} \tag{129}$$

As a result, we have

$$\begin{aligned}&\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{y}_p^k\right\}\mathrm{Cov}^{-1}\left\{\mathbf{y}_p^k,\mathbf{y}_p^k\right\}\\&=\frac{a_{k1}\frac{\sigma^2}{\tau p}}{\left(a_{k2}+\frac{\sigma^2}{\tau p}\right)\left\{\left(a_{k2}+\frac{\sigma^2}{\tau p}\right)+Ma_{k1}\right\}}\mathbf{a}_M\mathbf{a}_M^H\\&\quad+\frac{a_{k2}}{a_{k2}+\frac{\sigma^2}{\tau p}}\mathbf{I}_M\\&\triangleq a_{k3}\mathbf{a}_M\mathbf{a}_M^H+a_{k4}\mathbf{I}_M\triangleq\mathbf{A}_k=\mathbf{A}_k^H.\end{aligned}\tag{130}$$

Since we have $\mathbb{E}\{\mathbf{q}_k\}=\mathbb{E}\left\{\mathbf{y}_p^k\right\}=\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_k$, the LMMSE channel estimate in (128) is calculated as

$$\begin{aligned}\hat{\mathbf{q}}_k&=\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_k+\mathbf{A}_k\left(\mathbf{y}_p^k-\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_k\right)\\&=\mathbf{A}_k\mathbf{y}_p^k+\left(\mathbf{I}_M-\mathbf{A}_k\right)\sqrt{c_k\delta\varepsilon_k}\,\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_k\triangleq\mathbf{A}_k\mathbf{y}_p^k+\mathbf{B}_k.\end{aligned}\tag{131}$$

Additionally, we can expand the above linear expression and rewrite it as

$$\begin{aligned}\hat{\mathbf{q}}_k&=\mathbf{A}_k\left(\mathbf{q}_k+\frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k\right)+\mathbf{B}_k\\&=\mathbf{q}_k^1+\sum\nolimits_{\omega=2}^{4}\mathbf{A}_k\mathbf{q}_k^\omega+\sqrt{\gamma_k}\mathbf{A}_k\tilde{\mathbf{d}}_k+\frac{1}{\sqrt{\tau p}}\mathbf{A}_k\mathbf{N}\mathbf{s}_k.\end{aligned}\tag{132}$$

Then, by exploiting the property $\mathbf{A}_k\overline{\mathbf{H}}_2=\left(a_{k3}\mathbf{a}_M\mathbf{a}_M^H+a_{k4}\mathbf{I}_M\right)\mathbf{a}_M\mathbf{a}_N^H=\left(Ma_{k3}+a_{k4}\right)\overline{\mathbf{H}}_2$, we arrive at (17).

Based on the estimate $\hat{\mathbf{q}}_k$, we can obtain the estimation error $\mathbf{e}_k=\mathbf{q}_k-\hat{\mathbf{q}}_k$. By direct inspection, the mean of $\mathbf{e}_k$ is zero. Exploiting [72, Eq. (12.21)], Lemma 1 and (130), the MSE matrix of the estimation error is calculated as

$$\begin{aligned}\mathbf{MSE}_k&=\mathbb{E}\left\{\mathbf{e}_k\mathbf{e}_k^H\right\}=\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{q}_k\right\}\\&\quad-\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{y}_p^k\right\}\mathrm{Cov}^{-1}\left\{\mathbf{y}_p^k,\mathbf{y}_p^k\right\}\mathrm{Cov}\left\{\mathbf{y}_p^k,\mathbf{q}_k\right\}\\&=\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{q}_k\right\}-\mathbf{A}_k\,\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{q}_k\right\}\\&=\left(\mathbf{I}_M-\mathbf{A}_k\right)\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{q}_k\right\}\\&=\left(a_{k1}\left(1-a_{k4}\right)-Ma_{k1}a_{k3}-a_{k2}a_{k3}\right)\mathbf{a}_M\mathbf{a}_M^H\\&\quad+a_{k2}\left(1-a_{k4}\right)\mathbf{I}_M\\&\triangleq a_{k5}\mathbf{a}_M\mathbf{a}_M^H+a_{k6}\mathbf{I}_M,\end{aligned}\tag{133}$$

where $a_{k5}=\frac{a_{k1}\left(\frac{\sigma^2}{\tau p}\right)^2}{\left(a_{k2}+\frac{\sigma^2}{\tau p}\right)\left(a_{k2}+\frac{\sigma^2}{\tau p}+Ma_{k1}\right)}$, and $a_{k6}=\frac{a_{k2}\frac{\sigma^2}{\tau p}}{a_{k2}+\frac{\sigma^2}{\tau p}}$. Based on the MSE matrix, the NMSE of the estimation error is expressed as [53, Eq. (3.20)]

$$\begin{aligned}\mathrm{NMSE}_k&=\frac{\mathrm{Tr}\left\{\mathbf{MSE}_k\right\}}{\mathrm{Tr}\left\{\mathrm{Cov}\left\{\mathbf{q}_k,\mathbf{q}_k\right\}\right\}}\\&=\frac{\frac{\sigma^2}{\tau p}\left(Ma_{k1}a_{k2}+a_{k2}^2+\left(a_{k1}+a_{k2}\right)\frac{\sigma^2}{\tau p}\right)}{\left(a_{k2}+\frac{\sigma^2}{\tau p}\right)\left(a_{k2}+\frac{\sigma^2}{\tau p}+Ma_{k1}\right)\left(a_{k1}+a_{k2}\right)}.\end{aligned}\tag{134}$$

Hence, the proof is completed.

APPENDIX D

Recalling that $\mathbf{A}_k=a_{k3}\mathbf{a}_M\mathbf{a}_M^H+a_{k4}\mathbf{I}_M$, and $\overline{\mathbf{H}}_2=\mathbf{a}_M\mathbf{a}_N^H$, we obtain

$$\begin{aligned}\mathrm{Tr}\left\{\mathbf{A}_k\right\}&=M\left(a_{k3}+a_{k4}\right)\triangleq Me_{k1},\\\mathbf{A}_k\overline{\mathbf{H}}_2&=a_{k3}\mathbf{a}_M\mathbf{a}_M^H\mathbf{a}_M\mathbf{a}_N^H+a_{k4}\mathbf{I}_M\mathbf{a}_M\mathbf{a}_N^H\\&=\left(Ma_{k3}+a_{k4}\right)\overline{\mathbf{H}}_2\triangleq e_{k2}\overline{\mathbf{H}}_2,\\\mathrm{Tr}\left\{\mathbf{A}_k\mathbf{A}_k\right\}&=\mathrm{Tr}\left\{Ma_{k3}^2\mathbf{a}_M\mathbf{a}_M^H+2a_{k3}a_{k4}\mathbf{a}_M\mathbf{a}_M^H+a_{k4}^2\mathbf{I}_M\right\}\\&=M\left(Ma_{k3}^2+2a_{k3}a_{k4}+a_{k4}^2\right)\triangleq Me_{k3}.\end{aligned}\tag{135}$$

By direct inspection of e_{k1},e_{k2},e_{k3}, we evince that they are composed of non-negative terms. Therefore, we have $e_{k1},e_{k2},e_{k3}\geq 0$. Then, we aim to prove $e_{k1},e_{k2},e_{k3}\leq 1$. We first focus on the parameter e_{k2}. Using the expressions of a_{k3} and a_{k4} in (20) and (21), we can expand e_{k2} as

$$\begin{aligned}e_{k2}&=Ma_{k3}+a_{k4}\\&=\frac{a_{k2}\left(a_{k2}+\frac{\sigma^2}{\tau p}+Ma_{k1}\right)+Ma_{k1}\frac{\sigma^2}{\tau p}}{a_{k2}\left(a_{k2}+\frac{\sigma^2}{\tau p}+Ma_{k1}\right)+Ma_{k1}\frac{\sigma^2}{\tau p}+\frac{\sigma^2}{\tau p}\left(a_{k2}+\frac{\sigma^2}{\tau p}\right)}.\end{aligned}\tag{136}$$

It is clear that the numerator in (136) is smaller than the denominator. Therefore, we proved that $e_{k2}\leq 1$. Then, we can directly obtain $e_{k1}\leq e_{k2}\leq 1$ and $e_{k3}\leq e_{k2}^2\leq e_{k2}\leq 1$. Finally, when $\tau p\to\infty$ or $N\to\infty$, we have $a_{k3}\to 0$ and $a_{k4}\to 1$, which implies that $e_{k1}=e_{k2}=e_{k3}\to 1$. When $\tau p\to 0$, we have $a_{k3},a_{k4}\to 0$, which gives $e_{k1}=e_{k2}=e_{k3}\to 0$. This completes the proof.

APPENDIX E

A. Signal Term and Noise Term

According to the orthogonality property of the LMMSE estimator, we have $\mathbb{E}\left\{\mathbf{e}_k\left(\mathbf{y}_p^k\right)^H\right\}=\mathbf{0}$. Besides, since $\mathbf{e}_k$ has zero mean, we obtain $\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{e}_k\right\}=\mathbb{E}\left\{\left(\mathbf{A}_k\mathbf{y}_p^k+\mathbf{B}_k\right)^H\mathbf{e}_k\right\}=0$. Therefore, we have

$$\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}=\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\hat{\mathbf{q}}_k\right\}+\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{e}_k\right\}=\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_k\right\|^2\right\}.\tag{137}$$

Denote the intended signal in (35) as $\left|\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}\right|^2\triangleq E_k^{\mathrm{signal}}(\boldsymbol{\Phi})$, and the noise in (35) as $\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_k\right\|^2\right\}\triangleq E_k^{\mathrm{noise}}(\boldsymbol{\Phi})$. Clearly, $\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_k\right\|^2\right\}$ is a real variable. Then, from (137), we obtain

$$\begin{aligned}E_k^{\mathrm{signal}}(\boldsymbol{\Phi})&=\left|\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}\right|^2\\&=\left(\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_k\right\|^2\right\}\right)^2=\left(E_k^{\mathrm{noise}}(\boldsymbol{\Phi})\right)^2.\end{aligned}\tag{138}$$

Let us now derive $E_k^{\mathrm{noise}}(\boldsymbol{\Phi})$. Recall the expressions in (10) and (17). Since $\tilde{\mathbf{H}}_2$, $\tilde{\mathbf{h}}_k$, $\tilde{\mathbf{d}}_k$ and $\mathbf{N}$ are independent of each other and they all have zero mean, we can derive the term

$\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}$ by selecting the non-zero terms in the expansion as

$$\begin{aligned}
&E_k^{\text{noise}}\left(\mathbf{\Phi}\right)=\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_k\right\|^2\right\}=\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}\\
&=\sum_{\omega=1}^{4}\mathbb{E}\left\{\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_k^{\omega}\right\}+\gamma_k\mathbb{E}\left\{\tilde{\mathbf{d}}_k^H\mathbf{A}_k^H\tilde{\mathbf{d}}_k\right\}\\
&=c_k\delta\varepsilon_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k+e_{k2}c_k\delta\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\right\}\\
&\quad+c_k\varepsilon_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H\mathbf{A}_k^H\tilde{\mathbf{H}}_2\right\}\mathbf{\Phi}\overline{\mathbf{h}}_k\\
&\quad+c_k\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H\mathbf{A}_k^H\tilde{\mathbf{H}}_2\right\}\mathbf{\Phi}\tilde{\mathbf{h}}_k\right\}+\gamma_k\mathrm{Tr}\left\{\mathbf{A}_k^H\right\}\\
&\overset{(c)}{=}M\left\{\left|f_k(\mathbf{\Phi})\right|^2c_k\delta\varepsilon_k+Nc_k\delta e_{k2}+\left(Nc_k\left(\varepsilon_k+1\right)+\gamma_k\right)e_{k1}\right\},
\end{aligned}\tag{139}$$

where (c) applies Lemma 7 and exploits the identities $\mathrm{Tr}\{\mathbf{A}_k\}=Me_{k1}$, $\mathbf{\Phi}^H\mathbf{\Phi}=\mathbf{I}_N$, and $\mathrm{Tr}\left\{\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\right\}=MN$. Substituting (139) into (138), we complete the calculation of $E_k^{\text{signal}}\left(\mathbf{\Phi}\right)$ and $E_k^{\text{noise}}\left(\mathbf{\Phi}\right)$. Besides, by using a procedure similar to that used for obtaining (139), we obtain $\mathbb{E}\left\{\mathbf{q}_k^H\mathbf{q}_k\right\}$, $\mathbb{E}\left\{\underline{\hat{\mathbf{q}}}_k^H\underline{\hat{\mathbf{q}}}_k\right\}$, $\mathbb{E}\left\{\underline{\mathbf{q}}_k^H\mathbf{A}_k\mathbf{A}_k^H\underline{\mathbf{q}}_k\right\}$, $\mathbb{E}\left\{\underline{\mathbf{q}}_i^H\mathbf{A}_k\mathbf{A}_k^H\underline{\mathbf{q}}_i\right\}$, and $\mathbb{E}\left\{\underline{\hat{\mathbf{q}}}_k^H\underline{\mathbf{q}}_k\right\}$. The details are omitted for brevity.

B. Interference Term

In this subsection, we derive the interference term in (35). The interference term is denoted by $\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H\mathbf{q}_i\right|^2\right\}\triangleq I_{ki}\left(\mathbf{\Phi}\right)$. First, it is worth noting that the derivation of the interference term in the presence of imperfect CSI and double-Rician channels in RIS-aided massive MIMO systems has two main differences compared to conventional massive MIMO systems. Firstly, the channel $\mathbf{q}_k$ and $\mathbf{q}_i$ are not independent, since different users experience the same RIS-BS channel. This can be readily validated by examining that $\mathbb{E}\left\{\mathbf{q}_k^H\mathbf{q}_i\right\}\neq\mathbb{E}\left\{\mathbf{q}_k^H\right\}\mathbb{E}\left\{\mathbf{q}_i\right\}$. Secondly, the LMMSE error $\mathbf{e}_k$ is uncorrelated with but dependent on the estimate $\hat{\mathbf{q}}_k$, since the cascaded channel is not Gaussian distributed. To tackle these two challenges, we derive the interference term by decomposing it as

$$\begin{aligned}
&I_{ki}\left(\mathbf{\Phi}\right)=\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H\mathbf{q}_i\right|^2\right\}\\
&=\mathbb{E}\left\{\left|\left(\underline{\hat{\mathbf{q}}}_k+\mathbf{A}_k\mathbf{d}_k+\frac{1}{\sqrt{\tau p}}\mathbf{A}_k\mathbf{N}\mathbf{s}_k\right)^H\left(\underline{\mathbf{q}}_i+\mathbf{d}_i\right)\right|^2\right\}\\
&=\mathbb{E}\Big\{\Big|\underline{\hat{\mathbf{q}}}_k^H\underline{\mathbf{q}}_i+\underline{\hat{\mathbf{q}}}_k^H\mathbf{d}_i+\mathbf{d}_k^H\mathbf{A}_k^H\underline{\mathbf{q}}_i+\mathbf{d}_k^H\mathbf{A}_k^H\mathbf{d}_i\\
&\quad+\frac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\underline{\mathbf{q}}_i+\frac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\mathbf{d}_i\Big|^2\Big\}\\
&=\mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_k^H\underline{\mathbf{q}}_i\right|^2\right\}+\mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_k^H\mathbf{d}_i\right|^2\right\}+\mathbb{E}\left\{\left|\mathbf{d}_k^H\mathbf{A}_k^H\underline{\mathbf{q}}_i\right|^2\right\}\\
&\quad+\mathbb{E}\left\{\left|\mathbf{d}_k^H\mathbf{A}_k^H\mathbf{d}_i\right|^2\right\}+\frac{1}{\tau p}\mathbb{E}\left\{\left|\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\underline{\mathbf{q}}_i\right|^2\right\}\\
&\quad+\frac{1}{\tau p}\mathbb{E}\left\{\left|\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\mathbf{d}_i\right|^2\right\}.
\end{aligned}\tag{140}$$

The derivation of the interference term follows by calculating the six expectations in (140) one by one. Based on the independence between the noise and the channels, it is not difficult to complete the derivation of the 2nd - 6th expectation terms. The details are presented in [75]. To derive the first term $\mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_k^H\underline{\mathbf{q}}_i\right|^2\right\}$, we first expand it as

$$\begin{aligned}
&\mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_k^H\underline{\mathbf{q}}_i\right|^2\right\}=\mathbb{E}\left\{\left|\sum_{\omega=1}^{4}\sum_{\psi=1}^{4}\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_i^{\psi}\right|^2\right\}\\
&=\sum_{\omega=1}^{4}\sum_{\psi=1}^{4}\mathbb{E}\left\{\left|\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_i^{\psi}\right|^2\right\}\\
&\quad+\sum_{\substack{\omega1,\psi1,\omega2,\psi2,\\(\omega1,\psi1)\neq(\omega2,\psi2)}}^{4}\mathbb{E}\left\{\left(\left(\hat{\mathbf{q}}_k^{\omega1}\right)^H\mathbf{q}_i^{\psi1}\right)\left(\left(\hat{\mathbf{q}}_k^{\omega2}\right)^H\mathbf{q}_i^{\psi2}\right)^H\right\},
\end{aligned}\tag{141}$$

where $\mathbf{q}_k^1$ - $\mathbf{q}_k^4$ are defined in (10), and $\hat{\mathbf{q}}_k^1$ - $\hat{\mathbf{q}}_k^4$ are defined in (17).

Equation (141) can be derived by calculating the expectations of the 16 modulus-square terms and the expectations of the other cross-terms. The derivations utilize Lemma 7, Lemma 10, the independence between $\tilde{\mathbf{H}}_2$, $\tilde{\mathbf{h}}_k$, and $\tilde{\mathbf{h}}_i$, and some algebraic manipulations. The derivation is lengthy. Therefore, we provide the details only for some terms, as an example. Firstly, we need to derive the 16 modulus-square terms in (141), i.e., $\mathbb{E}\left\{\left|\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_i^{\psi}\right|^2\right\}$, $1\leq\omega\leq4$, $1\leq\psi\leq4$. When $\omega=2$ and $\psi=4$, based on the independence and using Lemma 7, we have

$$\begin{aligned}
&\mathbb{E}\left\{\left|e_{k2}\sqrt{c_k\delta}\sqrt{c_i}\,\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_i\right|^2\right\}\\
&=e_{k2}^2c_kc_i\delta\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbb{E}\left\{\tilde{\mathbf{H}}_2\mathbf{\Phi}\mathbb{E}\left\{\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^H\right\}\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\right\}\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\right\}\\
&=e_{k2}^2c_kc_i\delta\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbb{E}\left\{\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^H\right\}\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\right\}\\
&=e_{k2}^2c_kc_i\delta N\,\mathrm{Tr}\left\{\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\right\}=e_{k2}^2c_kc_i\delta MN^2.
\end{aligned}\tag{142}$$

When $\omega=3$ and $\psi=3$, using Lemma 10, we have

$$\begin{aligned}
&\mathbb{E}\left\{\left|\sqrt{c_k\varepsilon_k}\sqrt{c_i\varepsilon_i}\,\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\mathbf{A}_k^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_i\right|^2\right\}\\
&=c_kc_i\varepsilon_k\varepsilon_i\\
&\times\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H\mathbf{A}_k^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\mathbf{A}_k\tilde{\mathbf{H}}_2\right\}\mathbf{\Phi}\overline{\mathbf{h}}_k\\
&=c_kc_i\varepsilon_k\varepsilon_i\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\\
&\times\left(e_{k1}^2M^2\mathbf{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\mathbf{\Phi}^H+e_{k3}M\,\mathrm{Tr}\left\{\mathbf{\Phi}\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\mathbf{\Phi}^H\right\}\mathbf{I}_N\right)\mathbf{\Phi}\overline{\mathbf{h}}_k\\
&=c_kc_i\varepsilon_k\varepsilon_i\left(e_{k1}^2M^2\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\overline{\mathbf{h}}_k+e_{k3}M\,\mathrm{Tr}\left\{\overline{\mathbf{h}}_i\overline{\mathbf{h}}_i^H\right\}\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_k\right)\\
&=c_kc_i\varepsilon_k\varepsilon_i\left(e_{k1}^2M^2\left|\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_i\right|^2+e_{k3}MN^2\right).
\end{aligned}\tag{143}$$

For the remaining cross-terms in (141), even though they are 16×15 in total, only a few terms are non-zero. The non-zero terms can be identified by exploiting Lemma 8, the independence, and the zero-mean of $\tilde{\mathbf{H}}_2$, $\tilde{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_i$. Then, based on methods similar to (142) and (143), we can complete the derivation. The details can be found in [75].

C. Signal Leakage

In this subsection, we derive the signal leakage term in (35) as

$$E_k^{\text{leak}}\left(\mathbf{\Phi}\right)=\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H\mathbf{q}_k\right|^2\right\}-\left|\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}\right|^2,\tag{144}$$

where $\mathbb{E}\left\{\hat{\mathbf{q}}_k^H\mathbf{q}_k\right\}$ is given in (139). Therefore, we only need to derive the expectation $\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H\mathbf{q}_k\right|^2\right\}$. By exploiting the zero-mean properties of $\mathbf{d}_k$ and $\mathbf{N}$, and exploiting the independence between the cascaded channel, the direct channel, and the noise, we can expand this term and remove the terms with zero expectation as

$$\begin{aligned}
&\mathbb{E}\left\{\left|\hat{\mathbf{q}}_k^H\mathbf{q}_k\right|^2\right\}\\
&=\mathbb{E}\left\{\left|\left(\hat{\underline{\mathbf{q}}}_k^H+\mathbf{d}_k^H\mathbf{A}_k^H+\tfrac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\right)\left(\underline{\mathbf{q}}_k+\mathbf{d}_k\right)\right|^2\right\}\\
&=\mathbb{E}\Big\{\Big|\hat{\underline{\mathbf{q}}}_k^H\underline{\mathbf{q}}_k+\hat{\underline{\mathbf{q}}}_k^H\mathbf{d}_k+\mathbf{d}_k^H\mathbf{A}_k^H\underline{\mathbf{q}}_k+\mathbf{d}_k^H\mathbf{A}_k^H\mathbf{d}_k\\
&\quad+\tfrac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\underline{\mathbf{q}}_k+\tfrac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\mathbf{d}_k\Big|^2\Big\}\\
&\overset{(d)}{=}\mathbb{E}\left\{\left|\hat{\underline{\mathbf{q}}}_k^H\underline{\mathbf{q}}_k\right|^2\right\}+\mathbb{E}\left\{\left|\hat{\underline{\mathbf{q}}}_k^H\mathbf{d}_k\right|^2\right\}+\mathbb{E}\left\{\left|\mathbf{d}_k^H\mathbf{A}_k^H\underline{\mathbf{q}}_k\right|^2\right\}\\
&+\mathbb{E}\left\{\left|\mathbf{d}_k^H\mathbf{A}_k^H\mathbf{d}_k\right|^2\right\}+\mathbb{E}\left\{\left|\tfrac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\underline{\mathbf{q}}_k\right|^2\right\}\\
&+\mathbb{E}\left\{\left|\tfrac{1}{\sqrt{\tau p}}\mathbf{s}_k^H\mathbf{N}^H\mathbf{A}_k^H\mathbf{d}_k\right|^2\right\}\\
&+2\,\mathrm{Re}\left\{\mathbb{E}\left\{\hat{\underline{\mathbf{q}}}_k^H\underline{\mathbf{q}}_k\left(\mathbf{d}_k^H\mathbf{A}_k^H\mathbf{d}_k\right)^H\right\}\right\},
\end{aligned}\tag{145}$$

where in (d) the cross-term $\mathbb{E}\left\{\hat{\underline{\mathbf{q}}}_k^H\mathbf{d}_k\left(\mathbf{d}_k^H\mathbf{A}_k^H\underline{\mathbf{q}}_k\right)^H\right\}$ is zero due to Lemma 8, and the cross-term $\mathbb{E}\left\{\hat{\underline{\mathbf{q}}}_k^H\mathbf{d}_k\left(\mathbf{d}_k^H\mathbf{A}_k^H\mathbf{d}_k\right)^H\right\}$ is zero because the odd-order central moments of a zero-mean Gaussian variable are zero [78, Eq. (12)].

The derivation of the signal leakage follows by calculating seven expectations in (145). The derivations of the 2nd - 7th terms are not difficult and the details are omitted for brevity. On the other hand, the first term is expanded as

$$\begin{aligned}
&\mathbb{E}\left\{\left|\hat{\underline{\mathbf{q}}}_k^H\underline{\mathbf{q}}_k\right|^2\right\}=\mathbb{E}\left\{\left|\textstyle\sum_{\omega=1}^4\sum_{\psi=1}^4\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_k^{\psi}\right|^2\right\}\\
&=\textstyle\sum_{\omega=1}^4\sum_{\psi=1}^4\mathbb{E}\left\{\left|\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_k^{\psi}\right|^2\right\}\\
&+\textstyle\sum_{\substack{\omega1,\psi1,\omega2,\psi2\\(\omega1,\psi1)\neq(\omega2,\psi2)}}^4\mathbb{E}\left\{\left(\left(\hat{\mathbf{q}}_k^{\omega1}\right)^H\mathbf{q}_k^{\psi1}\right)\left(\left(\hat{\mathbf{q}}_k^{\omega2}\right)^H\mathbf{q}_k^{\psi2}\right)^H\right\}.
\end{aligned}\tag{146}$$

To derive (146), we need to calculate the 16 modulus-square terms $\mathbb{E}\left\{\left|\left(\hat{\mathbf{q}}_k^{\omega}\right)^H\mathbf{q}_k^{\psi}\right|^2\right\}$, $1\le\omega\le4$, $1\le\psi\le4$, and the remaining cross-terms. Similar to the derivation of the interference term, we only provide some examples. The details are presented in [75]. When $\omega=2$ and $\psi=4$, using (121), we arrive at

$$\begin{aligned}
&\mathbb{E}\left\{\left|e_{k2}\sqrt{c_k\delta}\sqrt{c_k}\,\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\right|^2\right\}\\
&=c_k^2\delta e_{k2}^2\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\right\}\\
&\overset{(e)}{=}c_k^2\delta e_{k2}^2\\
&\times\mathbb{E}_{\tilde{\mathbf{h}}_k}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\mathbb{E}_{\tilde{\mathbf{H}}_2}\left\{\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\right\}\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\,|\,\tilde{\mathbf{h}}_k\right\}\\
&=c_k^2\delta e_{k2}^2\mathbb{E}_{\tilde{\mathbf{h}}_k}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\,\mathrm{Tr}\left\{\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\right\}\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\mid\tilde{\mathbf{h}}_k\right\}\\
&\overset{(f)}{=}c_k^2\delta e_{k2}^2\mathbb{E}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\left(\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k\right)\right\}\\
&=c_k^2\delta e_{k2}^2\,\mathrm{Tr}\left\{\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbb{E}\left\{\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\right\}\right\}\\
&\overset{(g)}{=}c_k^2\delta e_{k2}^2\,\mathrm{Tr}\left\{\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}(N+1)\mathbf{I}_N\right\}\\
&=c_k^2\delta e_{k2}^2MN(N+1),
\end{aligned}\tag{147}$$

where (e) utilizes the law of total expectation, which calculates the conditional expectation of $\tilde{\mathbf{H}}_2$ given $\tilde{\mathbf{h}}_k$, and then calculates the expectation of $\tilde{\mathbf{h}}_k$. Since $\tilde{\mathbf{H}}_2$ is independent of $\tilde{\mathbf{h}}_k$, the conditional expectation of $\tilde{\mathbf{H}}_2$ given $\tilde{\mathbf{h}}_k$ is the same as its unconditional expectation; (f) comes from $\mathrm{Tr}\left\{\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\right\}=\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k$ which is a scalar number and its place can be arbitrarily moved; and (g) applies a special case of (121).

When $\omega=4$ and $\psi=4$, we have (148) shown at the bottom of the next page, where (i) uses (122).

Similar to the interference term, after the calculation of the 16 modulus-square terms, we identity the non-zero cross-terms and calculate their expectations. Interested readers can find the details in [75].

Appendix F

Recall the definition of $f_k(\mathbf{\Phi})$ in (43). If $N=1$, we have $\zeta_1^k=0$. Then, for any θ_1, we have $|f_k(\mathbf{\Phi})|=\left|e^{j\theta_1}\right|=1$. If $N>1$, we aim to prove the inequalities $0\le|f_k(\mathbf{\Phi})|\le N$. Firstly, by invoking the triangle inequality, we have

$$|f_k(\mathbf{\Phi})|=\left|\sum\nolimits_{n=1}^N e^{j\left(\zeta_n^k+\theta_n\right)}\right|\le\sum\nolimits_{n=1}^N\left|e^{j\left(\zeta_n^k+\theta_n\right)}\right|=N.\tag{149}$$

The equality holds if the phase shifts of all the RIS elements are aligned as $\theta_n=-\zeta_n^k+C_0,\forall n$, where C_0 is an arbitrary constant. Next, we aim to prove that the minimum value of $|f_k(\mathbf{\Phi})|$ is zero. Firstly, if N is even, the minimum value 0 is obtained when $\theta_{2i-1}+\zeta_{2i-1}^k=\left(\theta_{2i}+\zeta_{2i}^k\right)+\pi$, $1\le i\le\frac{N}{2}$. Otherwise, if N is odd, the minimum value 0 is still achievable for

$$\begin{aligned}
&\theta_{2i-1}+\zeta_{2i-1}^k=\left(\theta_{2i}+\zeta_{2i}^k\right)+\pi,1\le i\le\tfrac{N-1}{2}-1,\\
&\theta_{N-2}+\zeta_{N-2}^k=\tfrac{\pi}{3},\ \ \theta_{N-1}+\zeta_{N-1}^k=-\tfrac{\pi}{3},\ \ \theta_N+\zeta_N^k=\pi.
\end{aligned}\tag{150}$$

Next, we aim to prove that the term $|f_i(\mathbf{\Phi})|$ for user i, $i\neq k$, is bounded as $N\to\infty$ if the phase shifts of the RIS are designed to maximize $|f_k(\mathbf{\Phi})|$. This result can be proved rigorously for the one-dimensional ULA model.

By ignoring the elevation direction in (7) and (8) of the USPA model, we obtain a one-dimensional ULA model for

$\overline{\mathbf{h}}_k$ and $\mathbf{a}_N$ with φ_{kr}^a and φ_t^a being the AoA and the AoD, respectively. Then, we rewrite $f_k(\mathbf{\Phi})$ as

$$\begin{aligned}
&\overline{\mathbf{h}}_k \triangleq \mathbf{a}_N\left(\varphi_{kr}^a\right) = \left[1, e^{j2\pi\frac{d}{\lambda}\sin\varphi_{kr}^a}, \ldots, e^{j2\pi\frac{d}{\lambda}(N-1)\sin\varphi_{kr}^a}\right]^T,\\
&\mathbf{a}_N \triangleq \mathbf{a}_N\left(\varphi_t^a\right) = \left[1, e^{j2\pi\frac{d}{\lambda}\sin\varphi_t^a}, \ldots, e^{j2\pi\frac{d}{\lambda}(N-1)\sin\varphi_t^a}\right]^T,\\
&f_k(\mathbf{\Phi}) = \mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k = \textstyle\sum_{n=1}^N e^{j2\pi\frac{d}{\lambda}(n-1)(\sin\varphi_{kr}^a - \sin\varphi_t^a)+j\theta_n}.
\end{aligned} \tag{151}$$

If $\theta_n = 2\pi\frac{d}{\lambda}(n-1)\left(\sin\varphi_t^a - \sin\varphi_{kr}^a\right)$, we have $|f_k(\mathbf{\Phi})| = N$. Also, for user i, we have

$$\begin{aligned}
f_i(\mathbf{\Phi}) &= \mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_i = \sum\nolimits_{n=1}^N e^{j2\pi\frac{d}{\lambda}(n-1)(\sin\varphi_{ir}^a - \sin\varphi_t^a)+j\theta_n}\\
&= \sum\nolimits_{n=1}^N e^{j2\pi\frac{d}{\lambda}(n-1)(\sin\varphi_{ir}^a - \sin\varphi_t^a + \sin\varphi_t^a - \sin\varphi_{kr}^a)}\\
&= \sum\nolimits_{n=1}^N e^{j2\pi\frac{d}{\lambda}(n-1)(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a)}.
\end{aligned} \tag{152}$$

Then, by using properties of geometric progressions, we obtain

$$\begin{aligned}
|f_i(\mathbf{\Phi})| &= \left|\sum\nolimits_{n=1}^N \left(e^{j2\pi\frac{d}{\lambda}(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a)}\right)^{(n-1)}\right|\\
&= \left|\frac{1 - e^{j2\pi\frac{d}{\lambda}N(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a)}}{1 - e^{j2\pi\frac{d}{\lambda}\left(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a\right)}}\right|\\
&= \left|\frac{e^{-j\pi\frac{d}{\lambda}N(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a)} - e^{j\pi\frac{d}{\lambda}N(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a)}}{e^{-j\pi\frac{d}{\lambda}\left(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a\right)} - e^{j\pi\frac{d}{\lambda}\left(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a\right)}}\right.\\
&\quad \left.\times\frac{e^{j\pi\frac{d}{\lambda}N(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a)}}{e^{j\pi\frac{d}{\lambda}\left(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a\right)}}\right|\\
&= \frac{\sin\left(\pi\frac{d}{\lambda}N\left(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a\right)\right)}{\sin\left(\pi\frac{d}{\lambda}\left(\sin\varphi_{ir}^a - \sin\varphi_{kr}^a\right)\right)}.
\end{aligned} \tag{153}$$

Therefore, if the user i does not have the same AoA as user k, the term $|f_i(\mathbf{\Phi})|$ is bounded when $N \to \infty$. Using a similar process, we can prove that the term $\left|\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_i\right|^2$ is bounded when $N \to \infty$.

APPENDIX G

To begin with, we need to derive the first and second order statistical properties of the aggregated channel and the observation vector. The expectation is $\mathbb{E}\left\{\mathbf{y}_{c,p}^k\right\} = \sqrt{\widehat{c}_k\delta}\,\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k$. Aided by Lemma 7, the covariances between $\mathbf{q}_{c,k}$ and $\mathbf{y}_{c,p}^k$ is given by

$$\begin{aligned}
&\mathrm{Cov}\left\{\mathbf{q}_{c,k}, \mathbf{y}_{c,p}^k\right\} = \mathbb{E}\left\{\left(\mathbf{q}_{c,k} - \mathbb{E}\left\{\mathbf{q}_{c,k}\right\}\right)\left(\mathbf{y}_{c,p}^k - \mathbb{E}\left\{\mathbf{y}_{c,p}^k\right\}\right)^H\right\}\\
&= \mathbb{E}\left\{\left(\sqrt{\widehat{c}_k}\tilde{\mathbf{H}}_{c,2}\mathbf{\Phi}\overline{\mathbf{h}}_k + \sqrt{\gamma_k}\tilde{\mathbf{d}}_k\right)\left(\sqrt{\widehat{c}_k}\tilde{\mathbf{H}}_{c,2}\mathbf{\Phi}\overline{\mathbf{h}}_k + \sqrt{\gamma_k}\tilde{\mathbf{d}}_k\right)^H\right\}\\
&= \mathbb{E}\left\{\widehat{c}_k\tilde{\mathbf{H}}_2\mathbf{R}_{ris}^{1/2}\mathbf{\Phi}\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}^{1/2}\tilde{\mathbf{H}}_2^H + \gamma_k\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k^H\right\}\\
&= \left(\widehat{c}_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k + \gamma_k\right)\mathbf{I}_M.
\end{aligned} \tag{154}$$

Using Lemma 7, the definition of $\mathbf{H}_{c,2}$ in (68), the fact $\mathbf{V} = \mathbf{R}_{emi}^{1/2}\tilde{\mathbf{V}}$, and the independence between the channels, noise, and EMI, the covariance of $\mathbf{y}_{c,p}^k$ is calculated as (155) shown at the bottom of the next page.

Then, the LMMSE channel estimate for channel $\mathbf{q}_{c,k}$ is given by

$$\begin{aligned}
&\hat{\mathbf{q}}_{c,k} = \mathbb{E}\left\{\mathbf{q}_{c,k}\right\}\\
&+ \mathrm{Cov}\left\{\mathbf{q}_{c,k}, \mathbf{y}_{c,p}^k\right\}\mathrm{Cov}^{-1}\left\{\mathbf{y}_{c,p}^k, \mathbf{y}_{c,p}^k\right\}\left(\mathbf{y}_{c,p}^k - \mathbb{E}\left\{\mathbf{y}_{c,p}^k\right\}\right).
\end{aligned} \tag{156}$$

Combining (156) with (154) and (155) completes the proof.

APPENDIX H

With the aid of Lemmas 7 and 9, the proof follows as the proof of Appendix E. Using the orthogonal property, the noise is given by

$$\begin{aligned}
E_{c,k}^{\text{noise}} &= \mathbb{E}\left\{\left\|\hat{\mathbf{q}}_{c,k}\right\|^2\right\} = \mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,k}\right\}\\
&= \widehat{c}_k\delta\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k + \gamma_k\mathbb{E}\left\{\tilde{\mathbf{d}}_k^H\mathbf{\Upsilon}_k^H\tilde{\mathbf{d}}_k\right\}\\
&\quad + \mathbb{E}\left\{\widehat{c}_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H\mathbf{\Upsilon}_k^H\tilde{\mathbf{H}}_{c,2}\mathbf{\Phi}\overline{\mathbf{h}}_k\right\}\\
&= M\widehat{c}_k\delta\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k + \gamma_k\mathbb{E}\left\{\tilde{\mathbf{d}}_k^H\mathbf{\Upsilon}_k^H\tilde{\mathbf{d}}_k\right\}\\
&\quad + \widehat{c}_k\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}^{1/2}\mathbb{E}\left\{\tilde{\mathbf{H}}_2^H\mathbf{\Upsilon}_k^H\tilde{\mathbf{H}}_2\right\}\mathbf{R}_{ris}^{1/2}\mathbf{\Phi}\overline{\mathbf{h}}_k\\
&= M\widehat{c}_k\delta\left|f_k(\mathbf{\Phi})\right|^2 + \widehat{c}_k\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k^H\right\}\overline{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbf{R}_{ris}\mathbf{\Phi}\overline{\mathbf{h}}_k\\
&\quad + \gamma_k\,\mathrm{Tr}\left\{\mathbf{\Upsilon}_k^H\right\}.
\end{aligned} \tag{157}$$

By substituting $\mathbf{H}_{c,2} = \sqrt{\frac{\beta}{\delta+1}}\left(\sqrt{\delta}\overline{\mathbf{H}}_2 + \tilde{\mathbf{H}}_{c,2}\right)$, the EMI is calculated as

$$\begin{aligned}
&\mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\mathbf{H}_{c,2}\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\mathbf{H}_{c,2}^H\hat{\mathbf{q}}_{c,k}\right\}\\
&= \frac{\beta}{\delta+1}\left(\delta\mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\hat{\mathbf{q}}_{c,k}\right\}\right.\\
&\quad + 2\sqrt{\delta}\mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H\hat{\mathbf{q}}_{c,k}\right\}\\
&\quad \left.+ \mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\tilde{\mathbf{H}}_{c,2}\mathbf{\Phi}\mathbf{R}_{emi}\mathbf{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H\hat{\mathbf{q}}_{c,k}\right\}\right).
\end{aligned} \tag{158}$$

$$\begin{aligned}
&\mathbb{E}\left\{\left|\sqrt{c_k}\sqrt{c_k}\,\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\mathbf{A}_k^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\right|^2\right\} = c_k^2\mathbb{E}_{\tilde{\mathbf{h}}_k}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\mathbb{E}_{\tilde{\mathbf{H}}_2}\left\{\tilde{\mathbf{H}}_2^H\mathbf{A}_k^H\tilde{\mathbf{H}}_2\mathbf{\Phi}\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\tilde{\mathbf{H}}_2^H\mathbf{A}_k\tilde{\mathbf{H}}_2\right\}\mathbf{\Phi}\tilde{\mathbf{h}}_k \mid \tilde{\mathbf{h}}_k\right\}\\
&= c_k^2\mathbb{E}_{\tilde{\mathbf{h}}_k}\left\{\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\left(e_{k1}^2M^2\mathbf{\Phi}\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H + e_{k3}M\,\mathrm{Tr}\left\{\mathbf{\Phi}\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\mathbf{\Phi}^H\right\}\mathbf{I}_N\right)\mathbf{\Phi}\tilde{\mathbf{h}}_k\right\}\\
&= c_k^2\mathbb{E}\left\{e_{k1}^2M^2\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k + e_{k3}M\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H\tilde{\mathbf{h}}_k\right\} \overset{(i)}{=} c_k^2\left\{e_{k1}^2M^2N(N+1) + e_{k3}MN(N+1)\right\},
\end{aligned} \tag{148}$$

Equation (158) is derived by inserting the definition of $\hat{\mathbf{q}}_{c,k}$ from (73), using Lemmas 7 and 9, and utilizing the independence between $\tilde{\mathbf{H}}_{c,2}^H$, $\mathbf{V}$, and $\mathbf{N}$. Details of the proof are omitted for brevity.

Next, we discuss the derivation of the multi-user interference. For notational simplicity, define

$$\begin{aligned}\underline{\hat{\mathbf{q}}}_{c,k}^H &= \sqrt{\widehat{c}_k}\delta\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H + \sqrt{\widehat{c}_k}\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H\boldsymbol{\Upsilon}_k^H,\\ \underline{\mathbf{q}}_{c,i} &= \sqrt{\widehat{c}_i}\delta\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_i + \sqrt{\widehat{c}_i}\tilde{\mathbf{H}}_{c,2}\boldsymbol{\Phi}\overline{\mathbf{h}}_i.\end{aligned} \tag{159}$$

Then, based on the independence, the multi-user interference term can be decomposed as

$$\mathbb{E}\left\{\left|\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,i}\right|^2\right\} = \mathbb{E}\left\{\left|\hat{\mathbf{q}}_{c,k}^H\underline{\mathbf{q}}_{c,i}\right|^2\right\} + \mathbb{E}\left\{\left|\sqrt{\gamma_i}\hat{\mathbf{q}}_{c,k}^H\tilde{\mathbf{d}}_i\right|^2\right\}, \tag{160}$$

where

$$\begin{aligned}&\mathbb{E}\left\{\left|\sqrt{\gamma_i}\hat{\mathbf{q}}_{c,k}^H\tilde{\mathbf{d}}_i\right|^2\right\} = \mathbb{E}\left\{\gamma_i\hat{\mathbf{q}}_{c,k}^H\tilde{\mathbf{d}}_i\tilde{\mathbf{d}}_i^H\hat{\mathbf{q}}_{c,k}\right\}\\ &= \gamma_i\mathbb{E}\left\{\left\|\hat{\mathbf{q}}_{c,k}^H\right\|^2\right\} = \gamma_i E_{c,k}^{\text{noise}},\end{aligned} \tag{161}$$

and

$$\begin{aligned}&\mathbb{E}\left\{\left|\hat{\mathbf{q}}_{c,k}^H\underline{\mathbf{q}}_{c,i}\right|^2\right\} = \mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_{c,k}^H\underline{\mathbf{q}}_{c,i}\right|^2\right\}\\ &+\mathbb{E}\left\{\left|\left(\sqrt{\gamma_k}\tilde{\mathbf{d}}_k^H\boldsymbol{\Upsilon}_k^H + \frac{\mathbf{s}_k^H\mathbf{V}^H\boldsymbol{\Phi}^H\mathbf{H}_{c,2}^H\boldsymbol{\Upsilon}_k^H}{\sqrt{\tau p}} + \frac{\mathbf{s}_k^H\mathbf{N}^H\boldsymbol{\Upsilon}_k^H}{\sqrt{\tau p}}\right)\right.\right.\\ &\left.\left.\times\,\underline{\mathbf{q}}_{c,i}\right|^2\right\}\\ &= \mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_{c,k}^H\underline{\mathbf{q}}_{c,i}\right|^2\right\} + \mathbb{E}\left\{\left(\gamma_k + \frac{\sigma^2}{\tau p}\right)\underline{\mathbf{q}}_{c,i}^H\boldsymbol{\Upsilon}_k^2\underline{\mathbf{q}}_{c,i}\right\}\\ &+\mathbb{E}\left\{\frac{\sigma_e^2}{\tau p}\underline{\mathbf{q}}_{c,i}^H\boldsymbol{\Upsilon}_k\mathbf{H}_{c,2}\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\mathbf{H}_{c,2}^H\boldsymbol{\Upsilon}_k^H\underline{\mathbf{q}}_{c,i}\right\}.\end{aligned} \tag{162}$$

By utilizing Lemma 7 and Lemma 9 and utilizing steps similar to Appendix E, the computation of the multi-user interference can be completed by calculating the three expectations in (162).

Finally, we derive the signal leakage. Recall that $E_{c,k}^{\text{leak}} = \mathbb{E}\left\{\left|\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,k}\right|^2\right\} - \left|\mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,k}\right\}\right|^2$ and $\mathbb{E}\left\{\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,k}\right\}$ are derived in (157). Therefore, we only need to derive $\mathbb{E}\left\{\left|\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,k}\right|^2\right\}$, which can be decomposed as

$$\begin{aligned}&\mathbb{E}\left\{\left|\hat{\mathbf{q}}_{c,k}^H\mathbf{q}_{c,k}\right|^2\right\} = \mathbb{E}\left\{\left|\left(\underline{\hat{\mathbf{q}}}_{c,k}^H + \sqrt{\gamma_k}\tilde{\mathbf{d}}_k^H\boldsymbol{\Upsilon}_k^H\right)\mathbf{q}_{c,k}\right|^2\right\}\\ &+\mathbb{E}\left\{\left|\frac{\mathbf{s}_k^H\mathbf{V}^H\boldsymbol{\Phi}^H\mathbf{H}_{c,2}^H\boldsymbol{\Upsilon}_k^H}{\sqrt{\tau p}}\mathbf{q}_{c,k}\right|^2\right\} + \mathbb{E}\left\{\left|\frac{\mathbf{s}_k^H\mathbf{N}^H\boldsymbol{\Upsilon}_k^H}{\sqrt{\tau p}}\mathbf{q}_{c,k}\right|^2\right\},\end{aligned} \tag{163}$$

where

$$\begin{aligned}&\mathbb{E}\left\{\left|\left(\underline{\hat{\mathbf{q}}}_{c,k}^H + \sqrt{\gamma_k}\tilde{\mathbf{d}}_k^H\boldsymbol{\Upsilon}_k^H\right)\mathbf{q}_{c,k}\right|^2\right\}\\ &= \mathbb{E}\left\{\left|\underline{\hat{\mathbf{q}}}_{c,k}^H\mathbf{q}_{c,k}\right|^2\right\} + \mathbb{E}\left\{\left|\sqrt{\gamma_k}\tilde{\mathbf{d}}_k^H\boldsymbol{\Upsilon}_k^H\mathbf{q}_{c,k}\right|^2\right\}\\ &+ 2\,\mathrm{Re}\left\{\sqrt{\gamma_k}\mathbb{E}\left\{\underline{\hat{\mathbf{q}}}_{c,k}^H\mathbf{q}_{c,k}\mathbf{q}_{c,k}^H\boldsymbol{\Upsilon}_k\tilde{\mathbf{d}}_k\right\}\right\}.\end{aligned} \tag{164}$$

The calculation of the signal leakage follows from (163) and (164).

APPENDIX I

Recall that $\boldsymbol{\Phi} = \mathrm{diag}\{\boldsymbol{c}\}$ and $\boldsymbol{c} = e^{j\boldsymbol{\theta}}$. Then, we can express $\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}$ as

$$\begin{aligned}&\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\} = \sum_i\left[\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right]_{ii}\\ &= \sum_i\sum_a[\mathbf{A}]_{ia}[\boldsymbol{\Phi}]_{aa}[\mathbf{B}]_{ai}\left[\boldsymbol{\Phi}^H\right]_{ii} = \boldsymbol{c}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\boldsymbol{c}.\end{aligned} \tag{165}$$

Applying the chain rule, the gradient of $\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}$ with respect to the n-th element of $\boldsymbol{\theta}$, i.e., θ_n, is

$$\begin{aligned}&\frac{\partial\,\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial\theta_n}\\ &= \frac{\partial\boldsymbol{c}^H}{\partial\theta_n}\left(\mathbf{A}\odot\mathbf{B}^T\right)\boldsymbol{c} + \mathbf{c}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\frac{\partial\boldsymbol{c}}{\partial\theta_n}\\ &= -je^{-j\theta_n}\left[\left(\mathbf{A}\odot\mathbf{B}^T\right)\boldsymbol{c}\right]_n + j\left[\mathbf{c}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\right]_n e^{j\theta_n}.\end{aligned} \tag{166}$$

$$\begin{aligned}&\mathrm{Cov}\left\{\mathbf{y}_{c,p}^k, \mathbf{y}_{c,p}^k\right\} = \mathbb{E}\left\{\left(\mathbf{y}_{c,p}^k - \mathbb{E}\left\{\mathbf{y}_{c,p}^k\right\}\right)\left(\mathbf{y}_{c,p}^k - \mathbb{E}\left\{\mathbf{y}_{c,p}^k\right\}\right)^H\right\}\\ &= \mathbb{E}\left\{\widehat{c}_k\tilde{\mathbf{H}}_{c,2}\boldsymbol{\Phi}\overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H + \gamma_k\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k^H + \frac{\mathbf{H}_{c,2}\boldsymbol{\Phi}\mathbf{V}\mathbf{s}_k{\mathbf{s}_k}^H\mathbf{V}^H\boldsymbol{\Phi}^H\mathbf{H}_{c,2}^H}{\tau p} + \frac{\mathbf{N}\mathbf{s}_k{\mathbf{s}_k}^H\mathbf{N}^H}{\tau p}\right\}\\ &= \left(\widehat{c}_k\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\mathbf{R}_{ris}\boldsymbol{\Phi}\overline{\mathbf{h}}_k + \gamma_k + \frac{\sigma^2}{\tau p}\right)\mathbf{I}_M + \mathbb{E}\left\{\frac{\frac{\beta\delta}{\delta+1}\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{V}\mathbf{s}_k{\mathbf{s}_k}^H\mathbf{V}^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H}{\tau p}\right\} + \mathbb{E}\left\{\frac{\frac{\beta}{\delta+1}\tilde{\mathbf{H}}_{c,2}\boldsymbol{\Phi}\mathbf{V}\mathbf{s}_k{\mathbf{s}_k}^H\mathbf{V}^H\boldsymbol{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H}{\tau p}\right\}\\ &= \left(\widehat{c}_k\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\mathbf{R}_{ris}\boldsymbol{\Phi}\overline{\mathbf{h}}_k + \gamma_k + \frac{\sigma^2}{\tau p}\right)\mathbf{I}_M + \frac{\sigma_e^2\beta\delta\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H}{\tau p(\delta+1)} + \mathbb{E}\left\{\frac{\sigma_e^2\beta\tilde{\mathbf{H}}_{c,2}\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\tilde{\mathbf{H}}_{c,2}^H}{\tau p(\delta+1)}\right\}\\ &= \left(\widehat{c}_k\overline{\mathbf{h}}_k^H\boldsymbol{\Phi}^H\mathbf{R}_{ris}\boldsymbol{\Phi}\overline{\mathbf{h}}_k + \gamma_k + \frac{\sigma^2}{\tau p}\right)\mathbf{I}_M + \frac{\sigma_e^2\beta\delta\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{R}_{emi}\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H}{\tau p(\delta+1)} + \frac{\sigma_e^2\beta\,\mathrm{Tr}\left\{\mathbf{R}_{emi}\boldsymbol{\Phi}^H\mathbf{R}_{ris}\boldsymbol{\Phi}\right\}}{\tau p(\delta+1)}\mathbf{I}_M.\end{aligned} \tag{155}$$

Equation (166) is the n-th element of $\frac{\partial \operatorname{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial \boldsymbol{\theta}}$. Thus, the proof of (98) is completed by casting (166) for all values of n in a vector. The proof of (99) follows by noting that $\left\{\boldsymbol{\Phi}^T\left(\mathbf{A}^T \odot \mathbf{B}\right)\boldsymbol{c}^*\right\}^* = \boldsymbol{\Phi}^H\left(\mathbf{A}^H \odot \mathbf{B}^*\right)\boldsymbol{c} = \boldsymbol{\Phi}^H\left(\mathbf{A} \odot \mathbf{B}^T\right)\boldsymbol{c}$ if $\mathbf{A}$ and $\mathbf{B}$ are unitary matrices.

REFERENCES

[1] K. Zhi, C. Pan, H. Ren, K. Wang, and M. Elkashlan, "Reconfigurable intelligent surface-aided MISO systems with statistical CSI: Channel estimation, analysis and optimization," *Proc. IEEE Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 576–580, 2021.

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

[3] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst *et al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, 2021.

[4] A. Kammoun, L. Sanguinetti, M. Debbah, and M. Alouini, "Asymptotic analysis of RZF in large-scale MU-MIMO systems over rician channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7268–7286, Nov. 2019.

[5] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.

[6] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3880–3900, Jul. 2016.

[7] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[8] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[9] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[10] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

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

[12] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[13] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.

[14] W. Mei and R. Zhang, "Multi-beam multi-hop routing for intelligent reflecting surfaces aided massive MIMO," *IEEE Trans.Wireless Commun.*, vol. 21, no. 3, pp. 1897–1912, Mar. 2022.

[15] L. Dai, B. Wang, M. Wang, X. Yang, J. Tan, S. Bi, S. Xu, F. Yang, Z. Chen, M. Di Renzo, C. B. Chae, and L. Hanzo, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45 913–45 923, Mar. 2020.

[16] Q. Tao, J. Wang, and C. Zhong, "Performance analysis of intelligent reflecting surface aided communication systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2464–2468, Nov. 2020.

[17] A. Sirojuddin, D. D. Putra, and W.-J. Huang, "Low-complexity sum-capacity maximization for intelligent reflecting surface-aided MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1354–1358, Jul. 2022.

[18] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376–1393, Feb. 2021.

[19] P. Wang, J. Fang, L. Dai, and H. Li, "Joint transceiver and large intelligent surface design for massive MIMO mmwave systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052–1064, Feb. 2021.

[20] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[21] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3D beam flattening," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4128–4143, Jul. 2021.

[22] Y. Zhang, B. Di, H. Zhang, J. Lin, C. Xu, D. Zhang, Y. Li, and L. Song, "Beyond cell-free MIMO: Energy efficient reconfigurable intelligent surface aided cell-free MIMO communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 412–426, Jun. 2021.

[23] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-aided secure MIMO wireless communications via intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7851–7866, Dec. 2020.

[24] H. M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's CSI," *IEEE Signal Process. Lett.*, vol. 27, pp. 1300–1304, Aug. 2020.

[25] Z. Chu, W. Hao, P. Xiao, D. Mi, Z. Liu, M. Khalily, J. R. Kelly, and A. P. Feresidis, "Secrecy rate optimization for intelligent reflecting surface assisted MIMO system," *IEEE Trans. on Information Forensics and Security*, vol. 16, pp. 1655–1669, Dec. 2020.

[26] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020.

[27] G. Chen, Q. Wu, W. Chen, D. W. K. Ng, and L. Hanzo, "IRS-aided wireless powered MEC systems: TDMA or NOMA for computation offloading?" *IEEE Trans. Wireless Commun.*, early access, 2022.

[28] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for internet of things," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 619–623, March 2021.

[29] S. Jia, X. Yuan, and Y. C. Liang, "Reconfigurable intelligent surfaces for energy efficiency in D2D communication network," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 683–687, Mar. 2021.

[30] Y. Chen, B. Ai, H. Zhang, Y. Niu, L. Song, Z. Han, and H. Vincent Poor, "Reconfigurable intelligent surface assisted device-to-device communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 2792–2804, May 2021.

[31] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 5092–5106, Sep. 2020.

[32] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[33] H. Shen, W. Xu, S. Gong, C. Zhao, and D. W. K. Ng, "Beamforming optimization for IRS-aided communications with transceiver hardware impairments," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1214–1227, Feb. 2021.

[34] S. Zhou, W. Xu, K. Wang, M. Di Renzo, and M.-S. Alouini, "Spectral and energy efficiency of IRS-assisted MISO communication with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1366–1369, Sep. 2020.

[35] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.

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

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

[38] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

[39] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

[40] M. M. Zhao, Q. Wu, M. J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming opti-

mization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 2–17, Jan. 2021.

[41] M.-M. Zhao, A. Liu, Y. Wan, and R. Zhang, "Two-timescale beamforming optimization for intelligent reflecting surface aided multiuser communication with QoS constraints," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6179–6194, Sep. 2021.

[42] H. Guo, Y. C. Liang, and S. Xiao, "Intelligent reflecting surface configuration with historical channel observations," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1821–1824, Nov. 2020.

[43] A. Abrardo, D. Dardari, and M. Di Renzo, "Intelligent reflecting surfaces: Sum-rate optimization based on statistical position information," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7121–7136, Oct. 2021.

[44] Y. Gao, J. Xu, W. Xu, D. W. K. Ng, and M. S. Alouini, "Distributed IRS with statistical passive beamforming for MISO communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 221–225, Feb. 2021.

[45] Y. Chen, Y. Wang, J. Zhang, and M. Di Renzo, "QoS-driven spectrum sharing for reconfigurable intelligent surfaces (RISs) aided vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5969–5985, Sep. 2021.

[46] Q. U. A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, Dec. 2020.

[47] 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. Netw.*, vol. 7, no. 2, pp. 354–368, Jun. 2021.

[48] Y. Jia, C. Ye, and Y. Cui, "Analysis and optimization of an intelligent reflecting surface-assisted system with interference," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8068–8082, Dec. 2020.

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

[50] ——, "Statistical CSI-based design for reconfigurable intelligent surface-aided massive MIMO systems with direct links," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 1128–1132, May 2021.

[51] T. Van Chien, H. Q. Ngo, S. Chatzinotas, M. Di Renzo, and B. Ottersten, "Reconfigurable intelligent surface-assisted cell-free massive MIMO systems over spatially-correlated channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5106–5128, Jul. 2022.

[52] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next?: Five promising research directions for antenna arrays," *Digital Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.

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

[54] Ö. T. Demir and E. Björnson, "Is channel estimation necessary to select phase-shifts for RIS-assisted massive MIMO?" *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9537–9552, Nov. 2022.

[55] E. Björnson and L. Sanguinetti, "Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1306–1324, Sep. 2020.

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

[57] A. de Jesus Torres, L. Sanguinetti, and E. Björnson, "Electromagnetic interference in RIS-aided communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 668–672, Apr. 2022.

[58] ——, "Intelligent reconfigurable surfaces vs. decode-and-forward: What is the impact of electromagnetic interference?" *Proc. IEEE Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 1–5, 2022.

[59] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Massive MIMO with spatially correlated Rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, May 2019.

[60] Ö. Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive MIMO with rician fading and phase shifts," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5299–5315, Nov. 2019.

[61] Q. Zhang, S. Jin, K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[62] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

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

[64] N. O'Donoughue and J. M. F. Moura, "On the product of independent complex gaussians," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1050–1063, Mar. 2012.

[65] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[66] T. L. Marzetta and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[67] J. Nam, G. Caire, M. Debbah, and H. V. Poor, "Capacity scaling of massive MIMO in strong spatial correlation regimes," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 3040–3064, May 2020.

[68] X. Hu, C. Zhong, Y. Zhang, X. Chen, and Z. Zhang, "Location information aided multiple intelligent reflecting surface systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7948–7962, Dec. 2020.

[69] 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, May 2021.

[70] C. You, B. Zheng, W. Mei, and R. Zhang, "How to deploy intelligent reflecting surfaces in wireless network: BS-side, user-side, or both sides?" *J. Commun. Inf. Netw.*, vol. 7, no. 1, pp. 1–10, Mar. 2022.

[71] 3GPP, *TR 25.996, Spatial channel model for Multiple Input Multiple Output (MIMO) simulations*, Sep. 2012.

[72] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

[73] X. Qian, M. Di Renzo, V. Sciancalepore, and X. Costa-Pérez, "Joint optimization of reconfigurable intelligent surfaces and dynamic metasurface antennas for massive MIMO communications," in *IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2022, pp. 450–454.

[74] X. Li, "An entropy-based aggregate method for minimax optimization," *Engineering Optimization*, vol. 18, no. 4, pp. 277–285, Aug. 1992.

[75] K. Zhi *et al.*, "Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI," *arXiv preprint arXiv:2108.07622*, 2021.

[76] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," in *Doklady Akademii Nauk*, vol. 269, no. 3. Russian Academy of Sciences, 1983, pp. 543–547.

[77] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[78] A. Winkelbauer, "Moments and absolute moments of the normal distribution," 2012. [Online]. Available: https://arxiv.org/abs/1209.4340

Kangda Zhi received the B.Eng degree from the School of Communication and Information Engineering, Shanghai University (SHU), Shanghai, China, in 2017, and the M.Eng degree from School of Information Science and Technology, University of Science and Technology of China (USTC), Hefei, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests include Reconfigurable Intelligent Surface (RIS) and massive MIMO. He received the Exemplary Reviewer Certificate of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2021.

Cunhua Pan received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he was a Research Associate at the University of Kent, U.K. He held a post-doctoral position at Queen Mary University of London, U.K., from 2016 and 2019.From 2019 to 2021, he was a Lecturer in the same university. From 2021, he is a full professor in Southeast University.

His research interests mainly include reconfigurable intelligent surfaces (RIS), intelligent reflection surface (IRS), ultrareliable low latency communication (URLLC) , machine learning, UAV, Internet of Things, and mobile edge computing. He has published over 120 IEEE journal papers. He is currently an Editor of IEEE Transactions on Vehicular Technology, IEEE Wireless Communication Letters, IEEE Communications Letters and IEEE ACCESS. He serves as the guest editor for IEEE Journal on Selected Areas in Communications on the special issue on xURLLC in 6G: Next Generation Ultra-Reliable and Low-Latency Communications. He also serves as a leading guest editor of IEEE Journal of Selected Topics in Signal Processing (JSTSP) Special Issue on Advanced Signal Processing for Reconfigurable Intelligent Surface-aided 6G Networks, leading guest editor of IEEE Vehicular Technology Magazine on the special issue on Backscatter and Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G, leading guest editor of IEEE Open Journal of Vehicular Technology on the special issue of Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G and Beyond, and leading guest editor of IEEE ACCESS Special Issue on Reconfigurable Intelligent Surface Aided Communications for 6G and Beyond. He is Workshop organizer in IEEE ICCC 2021 on the topic of Reconfigurable Intelligent Surfaces for Next Generation Wireless Communications (RIS for 6G Networks), and workshop organizer in IEEE Globecom 2021 on the topic of Reconfigurable Intelligent Surfaces for future wireless communications. He is currently the Workshops and Symposia officer for Reconfigurable Intelligent Surfaces Emerging Technology Initiative. He is workshop chair for IEEE WCNC 2024, and TPC co-chair for IEEE ICCT 2022. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019. He received the IEEE ComSoc Leonard G. Abraham Prize in 2022, IEEE ComSoc AsiaPacific Outstanding Young Researcher Award, 2022.

Hong Ren received the B.S. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2014 and 2018, respectively. From 2016 to 2018, she was a Visiting Student with the School of Electronics and Computer Science, University of Southampton, U.K. From 2018 to 2020, she was a Post-Doctoral Scholar with Queen Mary University of London, U.K. She is currently an associate professor with Southeast University. Her research interests lie in the areas of communication and signal processing, including ultra-low latency and high reliable communications, Massive MIMO and machine learning.

Kezhi Wang received Ph.D. degree in Engineering from the University of Warwick, U.K. He was with University of Essex and Northumbria University, U.K. Currently he is a Senior Lecturer with Department of Computer Science, Brunel University London, U.K. His research interests include wireless communications, signal processing, mobile edge computing and machine learning. He is a co-recipient of the IEEE ComSoc Leonard G. Abraham Prize in 2022.

Maged Elkashlan received the PhD degree in Electrical Engineering from the University of British Columbia in 2006. From 2007 to 2011, he was a Scientist at the Commonwealth Scientific and Industrial Research Organization (CSIRO) Australia. During this time, he held visiting faculty appointments at University of New South Wales, University of Sydney, and University of Technology Sydney. In 2011, he joined the School of Electronic Engineering and Computer Science at Queen Mary University of London. He also holds a visiting faculty appointment at Beijing University of Posts and Telecommunications. His research interests fall into the broad areas of communication theory and statistical signal processing.

Dr. Elkashlan is an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS. He served as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2013 to 2018 and the IEEE COMMUNICATIONS LETTERS from 2012 to 2016. He also served as a Guest Editor of the special issue on "Location Awareness for Radios and Networks" of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, the special issue on "Energy Harvesting Communications" of IEEE COMMUNICATIONS MAGAZINE, the special issue on "Green Media: The Future of Wireless Multimedia Networks" of IEEE WIRELESS COMMUNICATIONS, and the special issue on "Millimeter Wave Communications for 5G" of IEEE COMMUNICATIONS MAGAZINE.

Marco Di Renzo (Fellow, IEEE) received the Laurea (cum laude) and Ph.D. degrees in electrical engineering from the University of L'Aquila, Italy, in 2003 and 2007, respectively, and the Habilitation à Diriger des Recherches (Doctor of Science) degree from University Paris-Sud (now Paris-Saclay University), France, in 2013. Currently, he is a CNRS Research Director (Professor) with the Laboratory of Signals and Systems (L2S) of Paris-Saclay University – CNRS and CentraleSupelec, Paris, France. He serves as the Coordinator of the Communications and Networks Research Area of the Laboratory of Excellence DigiCosme at Paris-Saclay University, as a Member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies at Paris-Saclay University, and as the Head of the Intelligent Physical Communications group with the Laboratory of Signals and Systems at CentraleSupelec. He serves as the Editor-in-Chief of IEEE Communications Letters. He is a founding member and a Vice Chair of the Industry Specification Group (ISG) on Reconfigurable Intelligent Surfaces (RIS) within the European Telecommunications Standards Institute (ETSI), where he serves as the Rapporteur for the work item on communication models, channel models, and evaluation methodologies. He is a Fellow of the IEEE, IET, AAIA, and Vebleo; an Ordinary Member of the European Academy of Sciences and Arts, and the Academia Europaea; and a Highly Cited Researcher. Also, he is a Fulbright Fellow, and was a Nokia Foundation Visiting Professor and a Royal Academy of Engineering Distinguished Visiting Fellow. His recent research awards include the 2021 EURASIP Best Paper Award, the 2022 IEEE COMSOC Outstanding Paper Award, and the 2022 Michel Monpetit Prize from the French Academy of Sciences.

Robert Schober (S'98, M'01, SM'08, F'10) received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from Friedrich-Alexander University of Erlangen-Nuremberg (FAU), Germany, in 1997 and 2000, respectively. From 2002 to 2011, he was a Professor and Canada Research Chair at the University of British Columbia (UBC), Vancouver, Canada. Since January 2012 he is an Alexander von Humboldt Professor and the Chair for Digital Communication at FAU. His research interests fall into the broad areas of Communication Theory, Wireless and Molecular Communications, and Statistical Signal Processing.

Robert received several awards for his work including the 2002 Heinz Maier Leibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, a 2006 UBC Killam Research Prize, a 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, a 2011 Alexander von Humboldt Professorship, a 2012 NSERC E.W.R. Stacie Fellowship, a 2017 Wireless Communications Recognition Award by the IEEE Wireless Communications Technical Committee, and the 2021 ACM NanoCom Milestone Award for "fundamental contributions to the modelling, design, and analysis of molecular communication systems". Since 2017, he has been listed as a Highly Cited Researcher by the Web of Science. Robert is a Fellow of the Canadian Academy of Engineering, a Fellow of the Engineering Institute of Canada, and a Member of the German National Academy of Science and Engineering.

He served as Editor-in-Chief of the IEEE Transactions on Communications from 2012 to 2015 and as VP Publications of the IEEE Communication Society (ComSoc) in 2020 and 2021. Currently, he serves as Member of the Editorial Board of the Proceedings of the IEEE , as Member at Large of the ComSoc Board of Governors, and as ComSoc Treasurer.

H. Vincent Poor (F'87) received the Ph.D. degree in EECS from Princeton University, Princeton, NJ, USA, in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 until 2016, he served as a Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, including most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the forthcoming book Advanced Data Analytics for Power Systems (Cambridge University Press, 2020). Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences, the Royal Society and other national and international academies. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, the 2019 ASEE Benjamin Garver Lamme Award, a D.Sc. honoris causa from Syracuse University, awarded in 2017, and a D.Eng. honoris causa from the University of Waterloo, awarded in 2019.

Jiangzhou Wang (Fellow, IEEE) is a Professor with the University of Kent, U.K. He has published more than 400 papers and four books. His research focuses on mobile communications. He was a recipient of the 2022 IEEE Communications Society Leonard G. Abraham Prize and IEEE Globecom2012 Best Paper Award. He was the Technical Program Chair of the 2019 IEEE International Conference on Communications (ICC2019), Shanghai, Executive Chair of the IEEE ICC2015, London, and Technical Program Chair of the IEEE WCNC2013. He is/was the editor of a number of international journals, including IEEE Transactions on Communications from 1998 to 2013. Professor Wang is a Fellow of the Royal Academy of Engineering, U.K., Fellow of the IEEE, and Fellow of the IET.

Lajos Hanzo (FIEEE'04) (http://www-mobile.ecs.soton.ac.uk, https://en.wikipedia.org/wiki/Lajos_Hanzo) received his Master degree and Doctorate in 1976 and 1983, respectively from the Technical University (TU) of Budapest. He was also awarded the Doctor of Sciences (DSc) degree by the University of Southampton (2004) and Honorary Doctorates by the TU of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served several terms as Governor of both IEEE ComSoc and of VTS. He is also a Fellow of the Royal Academy of Engineering (FREng), of the IET and of EURASIP. He is the recipient of the 2022 Eric Sumner Field Award. He has published 19 John Wiley research monographs and 2000+ contributions at IEEE Xplore.