Two-Timescale Design for Reconfigurable Intelligent Surface-Aided URLLC

Qihao Peng, Hong Ren, Cunhua Pan, *Senior Member, IEEE*,

Maged Elkashlan, *Senior Member, IEEE*, Ana García Armada, *Fellow, IEEE*,

and Petar Popovski, *Fellow, IEEE*.

Abstract

In this paper, to tackle the blockage issue in massive multiple-input-multiple-output (mMIMO) systems, a reconfigurable intelligent surface (RIS) is seamlessly deployed to support devices with ultra-reliable and low-latency communications (URLLC). The transmission power of the base station and the phase shifts of the RIS are jointly devised to maximize the weighted sum rate while considering the spatially correlation and channel estimation errors. Firstly, the relationship between the channel estimation error and spatially correlated RIS's elements is revealed by using the linear minimum mean square error. Secondly, based on the maximum-ratio transmission precoding, a tight lower bound of the rate under short packet transmission is derived. Finally, the NP-hard problem is decomposed into two optimization problems, where the transmission power is obtained by geometric programming and phase shifts are designed by using gradient ascent method. Besides, we have rigorously proved that the proposed algorithm can rapidly converge to a sub-optimal solution with low complexity. Simulation results confirm the tightness between the analytic results and Monte Carlo simulations. Furthermore, the two-timescale scheme provides a practical solution for the short packet transmission.

Index Terms

Reconfigurable intelligent surface (RIS), massive MIMO, two-timescale design, spatial correlation, URLLC.

Q. Peng and M. Elkashlan are with the School of Electronic Engineering and Computer Science at Queen Mary University of London, U.K. (e-mail: {q.peng, maged.elkashlan}@qmul.ac.uk). H. Ren and C. Pan are with National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. (e-mail:{hren,cpan}@seu.edu.cn). Ana García Armada is with Department of Signal Theory and Communications, University Carlos III of Madrid, Avda. de la Universidad 30 (28911 Leganés) Spain (email: agarcia@tsc.uc3m.es). Petar Popovski is with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: petarp@es.aau.dk). (Corresponding author: Hong Ren)

I. INTRODUCTION

Industry 4.0 advocates a paradigm of smart factories where interconnected industrial Internet of Things (IIoT) enable the realization of product manufacturing and various tasks [1]. Through industrial control networks (ICN), various components, such as sensors and robots, can collaborate efficiently with the aid of real-time and precise commands. Conventionally, the wired network infrastructure (i.e., copper or fiber) is adopted to enable real-time and precise control, which results in high expenditure on installation and maintenance. In contrast to wired network infrastructure, wireless networks are not only less costly but also more flexible for mobile IIoTs. However, a small amount of data commands need to be exchanged with decoding error probability (DEP) less than 10^{-6} and latency below 1 ms [2], leading to challenges for wireless links to enable guaranteed connectivity performance.

Recently, the academic and industry communities have made significant contributions towards fulfilling the demanding requirements of ultra-reliable and low-latency communications (URLLC). The achievable data rate under point-to-point communications was derived and the relationship between achievable data rate and blocklength was revealed in [3]. Then, based on the approximated expression under finite blocklength, the feasibility of short packet transmission has been analyzed in several communication systems, such as orthogonal frequency-division multiplexing (OFDM) [4], massive multiple-input-multiple-output (mMIMO) [5], [6], and cell-free mMIMO systems [7]–[10]. Despite the significant potential of current systems, such as mMIMO and OFDM, in supporting low-latency and ultra-reliable services for multiple devices simultaneously, ensuring device's quality of service remains a challenge because of the detrimental effects of severe blockages induced by tall buildings and mobile objects.

To tackle the above-mentioned hurdle, a potential solution is to integrate reconfigurable intelligent surfaces (RISs) with communication systems. This is due to the fact that RIS can tune a large number of passive scattering elements to enhance the useful signals or mitigate interfering signals at low cost [11]–[15]. The benefits of integrating RISs into communication systems have been investigated, such as unmanned aerial vehicle communications [16], [17], and cell-free systems [18], [19]. However, the aforementioned contributions were primarily based on the Shannon Capacity, where the packet is assumed to be infinite.

Owing to the enhanced path gain with low energy consumption, the RIS-aided short packet transmission have attracted extensive attention [20]–[24]. Specifically, by designing the beam-

forming and optimizing blocklength, the authors of [20] proposed a Nelder–Mead simplex to minimize the DEP by leveraging the benefits of RIS with UAV. By using the related Gamma distribution and approximation, the average rate under short packet transmission was analyzed in [21]. Then, the authors of [22] further investigated the average rate under the correlated channel and electromagnetic interference. Ren *et al.* of [23] performed a comprehensive analysis of the single user's average DEP and achievable data rate under various scenarios, such as the Rayleigh fading and the Rician fading channels. To satisfy multiple users' diverse requirements, the total blocklength was minimized in [24].

Although previous studies have extensively investigated the advantages of RIS-enabled finite blocklength transmission, most of them focused on the instantaneous channel state information (CSI)-based beamforming. This scheme, however, is impractical under the finite blocklength regime. Firstly, based on the results given in [25], the blocklength used for channel estimation typically increases with the RIS's reflecting elements, which leads to prohibitively high pilot overhead. When the transmission blocklength is limited, the high pilot overhead means there are few blocklength left for data transmission, leading to severe data rate loss. As a result, it is more challenging to devise the optimal transmission schemes to strike a trade-off between channel estimation and data transmission. Secondly, the instantaneous CSI-based scheme demands that the base station (BS) determine the optimal resource allocation based on the instantaneous CSI. Therefore, information feedback and power control need to be executed frequently before each data transmission, which leads to high signal processing latency and makes it challenging to satisfy the low-latency requirements.

To address these practical challenges, a promising solution named the two-timescale scheme was proposed in [26], and was further analyzed in [27]–[29]. Rather than relying on individual channels, the two-timescale scheme devises phase shifts using the statistical CSI of the BS-RIS-user channel. Such an approach facilitates channel estimation akin to the RIS-free systems, thereby considerably reducing the pilot overhead. Moreover, the phase design of the RIS can be determined relying on the large-scale fading factors, which eliminates the requirement for information feedback and resource allocation operations until the long-term CSI has changed. This approach significantly reduces information exchange and processing latency. However, the above-mentioned contributions of [26]–[29] focused on the infinite blocklength, and the short packet transmission scheme under the two-timescale scheme remains an open problem.

Against the above background, we integrate the RIS with an mMIMO system and devise the

downlink two-timescale scheme while accounting for both channel estimation errors and the spatial correlation among the reflecting elements. First, we reveal the relationship between the normalized mean square error (NMSE) and spatially correlated RIS's elements by the linear minimum mean square error (LMMSE). Then, a tight lower bound for the ergodic data rate is derived in closed form by relying on statistical CSI, which is more applicable for mMIMO systems with large-scale antennas. Finally, an effective algorithm is proposed to maximize the system performance under the short packet transmission. The specific contributions are given as follows:

1) In view of the channel estimation errors and spatially correlated reflective elements, the two-timescale scheme is tailored for the RIS-enabled finite blocklength transmission under severe blockage scenarios, including indoor and urban communications. The double Gaussian distributed channel is estimated, and the expression of NMSE reveals that, in comparison to the independent elements, the spatial correlation of RIS reflective elements contributes to improving the accuracy of estimated channels. Furthermore, a tight lower bound based on the maximum ratio transmission (MRT) precoding is derived, which offers an explicit closed-form expression for the power allocation and phase shifts design.

2) Based on the statistical CSI, weighted sum rate under short packet transmission is maximized while satisfying per-user requirements of data rate and DEP. To address this issue, an effective alternating optimization is proposed, where the phase shifts and transmission power are optimized relying on the accelerated gradient ascent method and geometric programming (GP) [5], [30], respectively. Furthermore, the convergence of our algorithm is rigorously proved, which is verified in the simulation results.

3) Extensive numerical results demonstrate the tightness between the analytic expression and ergodic rate obtained from Monte-Carlo simulations, and insightful analysis is drawn. The obtained results provide evidence that our proposed method achieves performance close to the ideal with remarkably low complexity. Furthermore, the two-timescale scheme demonstrated its superiority to that relying on the instantaneous CSI, which provides a practical solution for short packet transmission.

The remaining sections are organized as follows. The system model is provided and a lower bound based on large-scale fading factors is presented in Section II. Then, the weighted sum rate under short packet transmission is maximized by devising the two-timescale scheme in Section
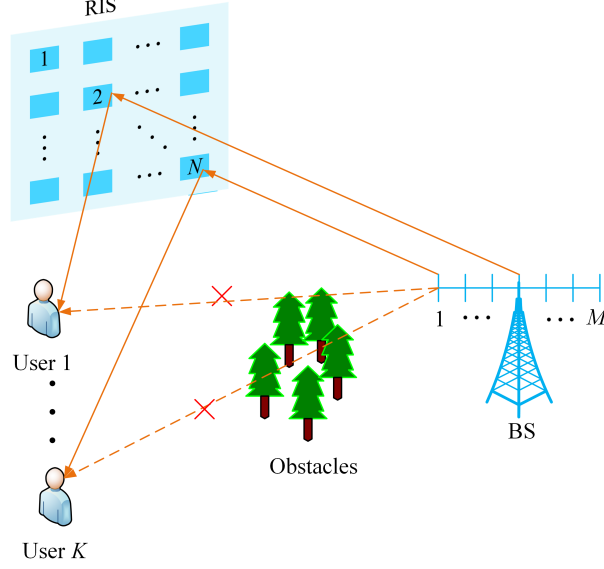
Fig. 1: System model for a RIS-enabled mMIMO system.

III. In Section IV, extensive numerical results are presented. Finally, our conclusions are drawn in Section V.

II. SYSTEM MODEL

A. Channel model

In the system depicted in Fig. 1, the M-antenna BS first estimates the uplink channel, and then performs beamforming techniques based on channel reciprocity to transmit signals to K single-antenna devices. However, the direct links are susceptible to obstruction owing to numerous blocking obstacles (i.e., robots, trees, and vehicles). To address this blockage issue, an strategically deployed RIS is utilized to reflect the effective signal by tuning the N passive and low-cost reflecting elements, thereby satisfying the diverse requirements of different devices.

The channel between the RIS and the kth device is denoted as $\mathbf{v}_k \in \mathbb{C}^{N \times 1}$, which is written as

$$\mathbf{v}_k = \sqrt{\beta_k^{\mathrm{RU}}} \big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}} \tilde{\mathbf{v}}_k, \tag{1}$$

where β_k^{RU} is the distance-based path loss, $\mathbf{C}_k^{\mathrm{RU}} \in \mathbb{C}^{N \times N}$ ($\mathbf{C}_k^{\mathrm{RU}} \succ \mathbf{0}$) represents the RIS transmit correlation matrix, and $\tilde{\mathbf{v}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is the small-scale fading factors. Furthermore, the

channel from the BS to the RIS $\mathbf{G} \in \mathbb{C}^{M \times N}$ is

$$\mathbf{G} = \sqrt{\beta^{\mathrm{BR}}} \big(\mathbf{C}^{\mathrm{B}}\big)^{\frac{1}{2}} \tilde{\mathbf{G}} \big(\mathbf{C}^{\mathrm{R}}\big)^{\frac{1}{2}}, \tag{2}$$

where β^{BR} is the large-scale fading factor, $\mathbf{C}^{\mathrm{B}} \in \mathbb{C}^{M \times M}$ ($\mathbf{C}^{\mathrm{B}} \succ \mathbf{0}$) and $\mathbf{C}^{\mathrm{R}} \in \mathbb{C}^{N \times N}$ ($\mathbf{C}_k^{\mathrm{R}} \succ \mathbf{0}$) are the BS transmit correlation matrix and RIS receive correlation matrix for $\mathbf{G}$, respectively. As described in previous works [28], [29], [31], the channel $\tilde{\mathbf{G}} \in \mathbb{C}^{M \times N}$ follows the Rayleigh fading, and each element of $\tilde{\mathbf{G}}$ follows a complex Gaussian distribution $\mathcal{CN}(0,1)$. The cascaded channel from the BS to the kth device is expressed as $\mathbf{h}_k = \mathbf{G}\mathbf{\Phi}\mathbf{v}_k \in \mathbb{C}^{M \times 1}$, where $\mathbf{\Phi} = \mathrm{diag}\big\{e^{j\theta_1}, e^{j\theta_2}, ..., e^{j\theta_N}\big\}$ is the phase shift matrix and $\theta_n \in [0, 2\pi)$ is the phase shift of the nth reflecting element. Then, the channel matrix is $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_K] \in \mathbb{C}^{M \times K}$.

B. Two-timescale and Channel Statistics

As can be seen from Fig. 2, there are two transmission schemes. The first approach necessitates the separate estimation of BS-RIS and RIS-user channels, resulting in a notable pilot overhead for channel estimation. Besides, under short packet transmission, the channel blocklength L is the product of bandwidth B and transmission latency T, i.e., $L = B \times T$. Once the bandwidth is determined, the channel blocklength is a linear function concerning the stringent transmission latency, leading to a finite blocklength. Therefore, it is challenging to devise the optimal transmission scheme based on the instantaneous CSI under short packet transmission.

In contrast, the two-timescale scheme can be implemented based on each user's effective channel in each coherence block, and thus the pilot overhead is equal to the number of users, which is the same as for mMIMO systems without RISs. Therefore, to satisfy the stringent latency, we investigate the two-timescale transmission scheme based on the channel statistics for the short packet transmission.

To facilitate the transmission design, we need to obtain the channel statistics of the BS and the RIS. According to previous studies given in [32], [33], the spatially correlation that is known as these channel statistics depends on the known correlation coefficient [32] and the distance between the RIS's reflecting elements [33]. Specifically, the BS transmit correlation matrix $\mathbf{C}^{\mathrm{B}}$ is based on the exponential correlation matrix [32]. The i-th row and j-th column element of
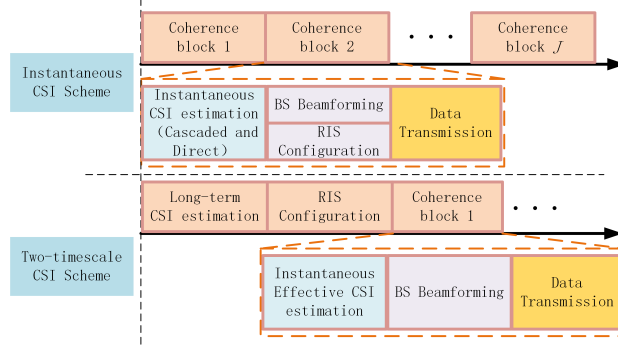
Fig. 2: Illustration of transmission designs based on Instantaneous CSI and long-term CSI (J channel coherence blocks).

the BS's correlation matrix $\mathbf{C}^{\mathrm{B}}$ is given by:

$$
c_{i,j}^{B} = \left\{ \begin{array}{ll} r^{j-i}, & i \leq j \\ \left(c_{i,j}^{B}\right)^{*}, & i > j \end{array} \right. , |r| \leq 1 \tag{3}
$$

where r is the (complex) correlation coefficient of neighboring receive branches. Furthermore, as stated in [33], the correlation $[\mathbf{C}^{\mathrm{RU}}]_{(i,j)}$ and $[\mathbf{C}^{\mathrm{R}}]_{(i,j)}$ between the ith element and jth element of the RIS is associated with the distance and wavelength λ, which is given by

$$
c_{i,j} = \mathrm{sinc}\left(\frac{2||\mathbf{u}_i - \mathbf{u}_j||}{\lambda})\right), 1 \leq i, j \leq N, \tag{4}
$$

where $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$, $\mathbf{u}_i = [0, \mathrm{mod}(i-1, \sqrt{N})d, \frac{\lfloor i-1 \rfloor}{\sqrt{N}}d]^T$, d is the width of each element. $||\mathbf{u}_i - \mathbf{u}_j||$ denotes the distance from the ith to jth elements of the RIS. Therefore, these channel statistics can be determined by its manufacturing process and can be treated as known prior information.

C. Channel Estimation

The uplink and downlink channels in time-division duplex (TDD) systems exhibit similar fading characteristics, known as channel reciprocity, and therefore the BS can utilize the results of uplink channel estimation to perform downlink precoding. For each transmission, the blocklength L is limited with the product of bandwidth B and transmission duration T. A segment of blocklength τ is used for estimating the uplink channel and the remaining portion for transmitting data. Furthermore, to prevent the K devices from pilot contamination, it is essential to ensure orthogonality between the pilot sequences of each device. Consequently, the pilot length τ should

be equal to or greater than K. Let us define $\mathbf{q}_k \in \mathbb{C}^{\tau \times 1}$ as the pilot sequence allocated for the kth device. Then, the expression of the received pilot signals $\mathbf{Y}_p \in M \times \tau$ at the BS is given by

$$\mathbf{Y}_p = \sqrt{\tau P_p} \sum_{k=1}^{K} \mathbf{h}_k \mathbf{q}_k^H + \mathbf{N}^p, \tag{5}$$

where P_p is the device's power for transmitting pilot. $\mathbf{N}^p$ is the additive noise, and each entry of $\mathbf{N}^p$ follows a complex normal distribution with zero mean and variance σ^2. The BS can obtain the cascaded channel to the kth device by using orthogonality of pilot sequence, which is given by

$$\mathbf{y}_k^p = \mathbf{h}_k + \frac{1}{\sqrt{\tau P_p}} \mathbf{N}^p \mathbf{q}_k. \tag{6}$$

Given the receiving signal $\mathbf{y}_k^p$, the widely utilized estimation technique in RIS-free mMIMO systems is the minimum mean square error (MMSE) criteria. However, this method may be no longer optimal for the RIS-enabled MIMIO systems, since the cascaded channel $\mathbf{H}$ is double Gaussian distributed, rather than Gaussian distributed. As a result, in the following theorem, we estimate the aggregated channel by adopting the LMMSE method. This is due to the fact that the LMMSE method can obtain the sub-optimal estimation of receiving signal only relying on the statistical information, without knowledge of the exact channel distribution.

Theorem 1: The estimated channel $\hat{\mathbf{h}}_k$ relying on LMMSE is given by

$$\begin{aligned}
\hat{\mathbf{h}}_k &= \beta_k^{\text{RU}} \beta^{\text{BR}} \text{Tr}\left\{\mathbf{Z}_k\right\} \mathbf{C}^{\text{B}} \\
&\quad \times \left(\beta_k^{\text{RU}} \beta^{\text{BR}} \text{Tr}\left\{\mathbf{Z}_k\right\} \mathbf{C}^{\text{B}} + \frac{\sigma^2}{\tau P_p} \mathbf{I}_M\right)^{-1} \mathbf{y}_k^p \\
&\triangleq \mathbf{R}_k \mathbf{y}_k^p,
\end{aligned} \tag{7}$$

where $\mathbf{Z}_k = \mathbf{\Phi} \mathbf{C}_k^{\text{RU}} \mathbf{\Phi}^H \mathbf{C}^{\text{R}}$. Then, the kth cascaded channel estimation error is

$$\text{NMSE}_k = \frac{\text{Tr}\left\{(\mathbf{I}_M - \mathbf{R}_k)\mathbf{C}^{\text{B}}\right\}}{\text{Tr}\left\{\mathbf{C}^{\text{B}}\right\}}. \tag{8}$$

Proof: Please refer to Appendix B. ∎

With the given result in (7), the estimated vector $\hat{\mathbf{h}}_k$ has the same dimension as the direct links in the RIS-free mMIMO systems. Different from [34], we estimate the effective cascaded channel rather than individual channels, and thus the pilot overhead for estimating the aggregated channel does not increase with the number of the reflecting elements N, but instead the same as the

number of devices K, which provides a more applicable solution for short packet transmission.

D. Achievable data rate under short packet transmission

Shannon's coding theorem states that when the channel blocklength is infinite, the DEP has any impact on the rate. Conversely, if the blocklength is finite, then it would cause a considerable performance loss on the rate. Specifically, by treating the interference as a part of noise [1], the data rate for small packet transmission is a function of blocklength, DEP, and SINR, which can be written as [3]

$$R_k \approx (1 - \eta) \log_2 (1 + \gamma_k) - \frac{Q^{-1}(\varepsilon_k)}{\ln 2} \sqrt{\frac{(1 - \eta) V_k (\gamma_k)}{L}}, \tag{9}$$

where $\eta = \tau/L$ means the ratio of pilot overhead, ε_k denotes decoding error rate, γ_k is the kth device's SINR, $V_k = 1 - (1 + \gamma_k)^{-2}$ represents the channel dispersion in terms of the kth device's SINR γ_k. $Q(\varepsilon_k)$ is $\frac{1}{\sqrt{2\pi}} \int_{\varepsilon_k}^{\infty} e^{-t^2/2} dt$, and $Q^{-1}(\varepsilon_k)$ is the inverse function of $Q(\varepsilon_k)$.

Then, we aim to calculate the average data rate for short packet transmission, which is given by

$$\bar{R}_k \approx \mathbb{E} \left\{ \frac{1 - \eta}{\ln 2} \left[\ln (1 + \gamma_k) - \alpha_k \sqrt{\frac{\frac{2}{\gamma_k} + 1}{\left(\frac{1}{\gamma_k} + 1\right)^2}} \right] \right\}$$

$$\triangleq \frac{1 - \eta}{\ln 2} \mathbb{E} \left\{ f_k \left(\frac{1}{\gamma_k} \right) \right\}, \tag{10}$$

where $\alpha_k = \frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}}$, $f_k(x) = \ln(1 + \frac{1}{x}) - \frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}} \sqrt{\frac{2x+1}{(1+x)^2}}$ is a function associated with the SINR and DEP of the kth device. Since the packet is delivered with extremely low latency, the distance-based path loss factors (i.e., large scale fading factors) remain constant with respect to the small-scale fading ones. Consequently, the expectation in (10) is adopted to obtain the closed-form expression only relying on the large-scale fading factors. Such work, however, is a challenge, owing to the non-convex and non-concave expression under finite blocklength. To tackle this issue, previous studies mainly focused on approximating the rate through concave function approximation when dealing with short packet transmission [35]–[37]. Similarly, we present a brief derivation for the lower bound of data rate.

---

[1]The channel hardening features of mMMIO result in SINR being heavily influenced by large-scale fading factors rather than small-scale ones. As a result, the instantaneous SINR can be considered equivalent to the expectation of SINR, such that we can approximate the equivalent channel as a Gaussian one, with SNR known at the transmitter, allowing the use of the finite blocklength formula for the Gaussian channels.

Considering that the kth device's data rate always satisfies $R_k \geq 0$, we can equivalently obtain the following inequality, which is given by

$$\alpha_k \leq \frac{\left(\frac{1}{\gamma_k}+1\right)\ln\left(1+\gamma_k\right)}{\sqrt{\frac{2}{\gamma_k}+1}} \triangleq g\left(\frac{1}{\gamma_k}\right), \tag{11}$$

where $g(x) = \frac{(1+x)\ln(1+\frac{1}{x})}{\sqrt{2x+1}}$. By checking the first-order derivative of $g(x)$, we can readily find that $g'(x) = \frac{x\ln(1+\frac{1}{x})-\frac{1}{x}-2}{(2x+1)^{\frac{3}{2}}}$ always satisfies $g'(x) < 0$, and thus $g(x)$ always decreases as x increases. In addition, by combining (11) with $f_k(x)$, we can readily obtain the feasible range of $f_k(x)$, represented as $0 \leq x \leq g^{-1}(\alpha_k)$. Based on these discussions, we prove that $f_k(x)$ is a decreasing and convex function when x satisfies $x \in (0, g^{-1}(\alpha_k)]$. Then, by using Jensen's inequality, we can formulate the expression for the lower bound $\hat{R}_k$ of the kth device, which is given by

$$\bar{R}_k \geq \hat{R}_k \triangleq \frac{1-\eta}{\ln 2} f_k\left(1/\hat{\gamma}_k\right), \tag{12}$$

where $\hat{\gamma}_k$ is $\hat{\gamma}_k = \frac{1}{\mathbb{E}(1/\gamma_k)}$.

In the following, a brief derivation for the lower bound using MRT precoding is presented. To eliminate the impact of small-scale fading parameters on the rate, we assume that the BS is equipped with large-scale antennas to achieve channel hardening and distance-based path loss is known in advance [29]. Based on the MRT precoding, the kth device's received signal is

$$r_k = \underbrace{\mathbb{E}\left\{\mathbf{h}_k^H\hat{\mathbf{h}}_k\sqrt{p_k^d}\right\}x_k}_{\text{Desired signal}}$$
$$-\underbrace{\left\{\mathbf{h}_k^H\hat{\mathbf{h}}_k\sqrt{p_k^d} - \mathbb{E}\left\{\mathbf{h}_k^H\hat{\mathbf{h}}_k\sqrt{p_k^d}\right\}\right\}x_k}_{\text{Signal leakage}} \tag{13}$$
$$+\sum_{k'\neq k}^{K}\underbrace{\mathbf{h}_k^H\hat{\mathbf{h}}_{k'}\sqrt{p_{k'}^d}}_{\text{Multi-user interference}}x_{k'} + n_k,$$

where p_k^d is the transmission power, and n_k is the additive noise with zero mean and variance σ^2. Then, the expression of the kth device's SINR is given by

$$\gamma_k = \frac{p_k^d\left|\mathbb{E}\left\{\mathbf{h}_k^H\hat{\mathbf{h}}_k\right\}\right|^2}{p_k^d\left|\mathbf{h}_k^H\hat{\mathbf{h}}_k - \mathbb{E}\left\{\mathbf{h}_k^H\hat{\mathbf{h}}_k\right\}\right|^2 + \sum_{k'\neq k}^{K}p_{k'}^d\left|\mathbf{h}_k^H\hat{\mathbf{h}}_{k'}\right|^2 + |n_k|^2}. \tag{14}$$

Since the mMIMO system has lots of transmission antennas, the law of large numbers is adopted to derive the lower bound of achievable data rate under short packet transmission.

Theorem 2: The lower bound for the kth device's averaged rate is

$$\hat{R}_k \triangleq \frac{1-\eta}{\ln 2} f_k \left(\frac{1}{\hat{\gamma}_k}\right),$$

$$\hat{\gamma}_k = \frac{p_k^d \left|\text{Tr}\left\{\mathbf{A}_k\right\}\right|^2}{\text{E}_{l,k} + \sum_{k'=1}^K p_{k'}^d \text{UI}_{k,k'} + \sigma^2},$$

(15)

where $\mathbf{A}_k$ is $\beta_k^{\text{RU}}\beta^{\text{BR}}\text{Tr}\left\{\mathbf{Z}_k\right\}\mathbf{R}_k\mathbf{C}^{\text{B}}$, $\text{E}_{l,k}$ is $p_k^d\left(\beta^{\text{BR}}\beta_k^{\text{RU}}\right)^2\text{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_k\right\}\text{Tr}\left\{\mathbf{C}^{\text{B}}\mathbf{R}_k^H\mathbf{C}^{\text{B}}\mathbf{R}_k\right\}$, and $\text{UI}_{k,k'}$ is given by

$$\text{UI}_{k,k'} = \beta_{k'}^{\text{RU}}\beta_k^{\text{RU}}\left(\beta^{\text{BR}}\right)^2 \Big(\left|\text{Tr}\left\{\mathbf{C}^{\text{B}}\mathbf{R}_{k'}\right\}\right|^2 \text{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_{k'}\right\}$$

$$+ \text{Tr}\left\{\mathbf{Z}_{k'}\right\}\text{Tr}\left\{\mathbf{C}^{\text{B}}\mathbf{R}_{k'}\mathbf{C}^{\text{B}}\mathbf{R}_{k'}^H\right\}\text{Tr}\left\{\mathbf{Z}_k\right\}\Big)$$

$$+ \frac{\sigma^2}{\tau P_p}\beta^{\text{BR}}\beta_k^{\text{RU}}\text{Tr}\left\{\mathbf{C}^{\text{B}}\mathbf{R}_{k'}\mathbf{R}_{k'}^H\right\}\text{Tr}\left\{\mathbf{Z}_k\right\}.$$

(16)

Proof: Please refer to Appendix C. ∎

The expression in (15) confirms that the SINR depends on the already known large-scale fading factors, in contrast to the uncertain ergodic rate through Monte Carlo simulations. Furthermore, owing to large-scale antenna transmission arrays in mMIMO systems, the derived lower bound is closer to the ergodic rate. As a result, the BS can allocate the transmission power and devise phase shifts according to the large-scale fading factors, which significantly reduce the information feedback and processing latency.

III. PHASE SHIFT DESIGN AND POWER ALLOCATION

According to the analytical result given in (15), we aim to design a scheme of power allocation and phase shifts to maximize the system performance by using the long-term statistical CSI.

A. Problem formulation

Since the weighted sum rate is one of the key performance indicators of wireless communication systems, our objective is to maximize the weighted sum rate of the RIS-aided mMIMO system while satisfying short packet transmission requirements of decoding error probability and latency.

First, the weight w_k is assigned to the kth device to provide priority for the specified devices. Then, by designing the RIS's phase shifts and optimizing the power consumption of BS, we aim

to maximize the weighted sum rate while satisfying the minimal per-user rate requirement. The problem is formulated as

$$\max_{\{\theta_n\},\{p_k^d\}} \sum_{k=1}^{K} w_k \hat{R}_k \tag{17a}$$

$$\text{s.t.} \quad \hat{R}_k \geq R_k^{\text{req}}, \forall k, \tag{17b}$$

$$\theta_n \in [0, 2\pi), \forall n, \tag{17c}$$

$$\sum_{K=1}^{K} p_k^d \leq P^d, \tag{17d}$$

where R_k^{req} is the kth device's minimal requirement of rate and DEP, and P^d denotes the total transmitting power. Specifically, the constraint (17b) implies that the kth device needs to satisfy two conditions simultaneously: it should receive a signal with the data rate not lower than R_k^{req}, and its decoding error rate should not be larger than ε_k.

Since the weighted sum rate maximization is an NP-hard problem, finding the optimal solution requires an exhaustive search, which is computationally demanding and impractical for communication systems. Moreover, the presence of a non-convex expression of data rate brings further complexity to solve Problem (17). As a result, maximizing the weighted sum rate requires efficient algorithms, which, even though sub-optimal, perform well in practice. To address this issue, we present an alternating optimization approach that provides a sub-optimal solution to the problem with low complexity.

B. Power Optimization

Owing to the non-convex expression in (9), finding an exact solution to transmission power is computationally challenging even with the fixed RIS's phase shifts matrix $\mathbf{\Phi}$. To solve this problem, we apply an equivalent transformation to convert it into a more tractable form. Firstly, using the lower bound given in (12), we can readily transform the requirements on data rate and DEP into the required SINR, which is given by

$$\hat{\gamma}_k \geq \frac{1}{f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{1-\eta}\right)}. \tag{18}$$

Then, we can equivalently reformulate Problem (17) by introducing the slack variables χ_k and we have

$$\max_{\{p_k^d\},\{\chi_k\}} \sum_{k=1}^{K} \bar{w}_k \left[\ln\left(1 + \chi_k\right) - \alpha_k \sqrt{1 - (1+\chi_k)^{-2}} \right] \tag{19a}$$

$$\text{s.t.} \quad \hat{\gamma}_k \geq \chi_k, \forall k, \tag{19b}$$

$$\chi_k \geq \frac{1}{f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{(1-\eta)}\right)}, \forall k, \tag{19c}$$

$$(17\text{d}), \tag{19d}$$

where $\bar{w}_k$ is $w_k \frac{\left(1-\frac{\tau}{L}\right)}{\ln 2}$. Finally, since each device's SINR can exceed $\frac{\sqrt{17}-3}{4}$ to meet the stringent requirements, we utilize Lemma 2 and Lemma 3 given in [38] to approximate the weighted lower bound of the data rate by an iterative approach, which is given by

$$\bar{w}_k \left[\ln(1 + \chi_k) - \alpha_k \sqrt{1 - (1+\chi_k)^{-2}} \right]$$
$$\geq \bar{w}_k \left[\rho_k^{(i)} \ln\left(\chi_k\right) + \delta_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \ln\left(\chi_k\right) - \alpha_k \hat{\delta}_k^{(i)} \right] \tag{20}$$
$$\geq \bar{w}_k \left[\ln\left(\chi_k\right)^{\left(\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)}\right)} + \left(\delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)}\right) \right],$$

where $\rho_k^{(i)}$ is $\frac{\chi_k^{(i)}}{1+\chi_k^{(i)}}$, $\delta_k^{(i)}$ is $\ln\left(1 + \chi_k^{(i)}\right) - \rho_k^{(i)} \ln\left(\chi_k^{(i)}\right)$, $\hat{\rho}_k^{(i)}$ is $\frac{\chi_k^{(i)}}{\sqrt{\chi_k^{(i)2}+2\chi_k^{(i)}}} - \frac{\chi_k^{(i)}\sqrt{\chi_k^{(i)2}+2\chi_k^{(i)}}}{\left(1+\chi_k^{(i)}\right)^2}$, and $\hat{\delta}_k^{(i)}$ is $\sqrt{1 - \frac{1}{\left(1+\chi_k^{(i)}\right)^2}} - \hat{\rho}_k^{(i)} \ln\left(\chi_k^{(i)}\right)$ in the ith iteration. Besides, the equality holds only when $\chi_k = \chi_k^{(i)}$. Substituting the lower bound of objective function into Problem (19), the original problem is rewritten as

$$\max_{\{p_k^d\},\{\chi_k\}} \sum_{k=1}^{K} \bar{w}_k \left[\ln\left(\chi_k\right)^{\left(\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)}\right)} + \left(\delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)}\right) \right] \tag{21a}$$

$$\text{s.t.} \quad (19\text{b}), (19\text{c}), (17\text{d}). \tag{21b}$$

By ignoring the constant terms $(\delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)})$ in (21a), the maximum exponential summation can be equivalently transformed into a product maximization problem, denoted as $\prod_{k=1}^{K} \chi_k^{\hat{w}_k^{(i)}}$, where $\hat{w}_k^{(i)}$ is $\bar{w}_k \left(\rho^{(i)} - \alpha_k \hat{\rho}^{(i)}\right)$. Based on these discussions, the sub-optimal problem in the ith

iteration can be expressed as

$$\max_{\{p_k^d\},\{\chi_k\}} \prod_{k=1}^{K} \chi_k^{\hat{w}_k^{(i)}} \tag{22a}$$

$$\text{s.t.} \quad (19b), (19c), (17d). \tag{22b}$$

Based on these discussions, it is feasible to efficiently solve Problem (22) using existing optimization tools like CVX.

C. Phase Shift Optimization

With fixed transmission power, the RIS can tune the phase shift to maximize the weighted sum rate, which can be written as

$$\max_{\{\theta_n\}} \sum_{k=1}^{K} w_k \hat{R}_k \tag{23a}$$

$$\text{s.t.} \quad (17b), (17c). \tag{23b}$$

To solve Problem (23), we employ the gradient ascent method to find the solution to the RIS's phase shifts. For ease of expression, we define $\boldsymbol{\Phi}$ as $\text{diag}(\mathbf{b})$, where $\mathbf{b} = e^{j\boldsymbol{\theta}}$ and $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_N]^T$ is a real vector of phase shift. It is noteworthy to mention that our method is based on the real phase shifts $\boldsymbol{\theta}$ rather than the complex variable $\mathbf{b}$, which is different from the existing work [27]. Specifically, the existing methods based on complex gradients may not satisfy the constraint (17c) and only provide the approximated solutions, which causes additional errors and performance loss. By contrast, our method directly obtains the real gradient of phase shifts without approximations, and thus there is no additional performance loss.

Then, the gradient of weighted sum rate in terms of $\boldsymbol{\theta}$ is derived by applying the chain rule, which is detailed as

$$\frac{\partial \sum_{k=1}^{K} w_k \hat{R}_k}{\partial \boldsymbol{\theta}} = \sum_{k=1}^{K} \left(\bar{w}_k \frac{1}{(1+\hat{\gamma}_k)} \frac{\partial \hat{\gamma}_k}{\partial \boldsymbol{\theta}} \right. $$
$$\left. - \tilde{w}_k \frac{1}{(1+\hat{\gamma}_k)^3 \sqrt{1 - \frac{1}{(1+\hat{\gamma}_k)^2}}} \frac{\partial \hat{\gamma}_k}{\partial \boldsymbol{\theta}} \right), \tag{24}$$

where $\tilde{w}_k$ is $\bar{w}_k \alpha_k$. The $\frac{\partial \hat{\gamma}_k}{\partial \boldsymbol{\theta}}$ is given by

$$
\frac{\partial \hat{\gamma}_k}{\partial \boldsymbol{\theta}} = \frac{\frac{\partial p_k^d |\text{Tr}\{\mathbf{A}_k\}|^2}{\partial \boldsymbol{\theta}}}{\mathrm{E}_{l,k} + \sum_{k'=1}^{K} p_{k'}^d \mathrm{UI}_{k,k'} + \sigma^2}
$$
$$
- \frac{p_k^d |\text{Tr}\{\mathbf{A}_k\}|^2 \left(\frac{\partial \mathrm{E}_{l,k}}{\partial \boldsymbol{\theta}} + \sum_{k'=1}^{K} \frac{\partial p_{k'}^d \mathrm{UI}_{k,k'}}{\partial \boldsymbol{\theta}} \right)}{\left(\mathrm{E}_{l,k} + \sum_{k'=1}^{K} p_{k'}^d \mathrm{UI}_{k,k'} + \sigma^2 \right)^2},
$$

(25)

Then, two useful lemmas are presented to calculate the gradient of terms.

Lemma 1: For the given matrices $\mathbf{A}$ and $\mathbf{B}$, the gradient of $\text{Tr}(\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H)$ relying on phase shifts can be expressed as

$$
\frac{\partial \text{Tr}\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\}}{\partial \boldsymbol{\theta}} = -j\boldsymbol{\Phi}^H \left(\mathbf{A} \odot \mathbf{B}^T \right) \mathbf{b}
$$
$$
+ j\boldsymbol{\Phi}^T \left(\mathbf{A}^T \odot \mathbf{B} \right) \mathbf{b}^* = \mathbf{u}_g(\mathbf{A}, \mathbf{B}).
$$

(26)

Proof: Please refer to Appendix D. ∎

Lemma 2: For a given matrix $\mathbf{X}$, $\frac{\text{Tr}\{\mathbf{X}\mathbf{R}_k\}}{\partial \boldsymbol{\theta}}$ is expressed as

$$
\frac{\text{Tr}\{\mathbf{X}\mathbf{R}_k\}}{\partial \boldsymbol{\theta}} = \beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \Big(\text{Tr}\{\mathbf{X}\mathbf{W}_k\}
$$
$$
- \beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \text{Tr}\{\mathbf{Z}_k\} \text{Tr}\{\mathbf{X}\mathbf{W}_k\mathbf{W}_k\} \Big)
$$
$$
\times \mathbf{u}_g(\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}) = \mathbf{z}_g(\mathbf{X}, \mathbf{W}_k),
$$

(27)

where $\mathbf{W}_k$ is equal to $\mathbf{C}^{\mathrm{B}} \left(\beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \text{Tr}\{\mathbf{Z}_k\} \mathbf{C}^{\mathrm{B}} + \frac{\sigma^2}{\tau P_p} \mathbf{I}_M \right)^{-1}$.

Proof: Please refer to Appendix E. ∎

In the following, we derive the gradients of other terms by using Lemma 1 and Lemma 2.

Lemma 3: Gradients of terms $|\text{Tr}\{\mathbf{A}_k\}|^2$, $\text{Tr}\{\mathbf{Z}_k\mathbf{Z}_k\}$, $\text{Tr}\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k^H\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\}$, and $\text{Tr}\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{R}_{k'}^H\}$ with respect to $\boldsymbol{\theta}$ are given by

$$
\frac{\partial |\text{Tr}\{\mathbf{A}_k\}|^2}{\partial \boldsymbol{\theta}} = 2\beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \text{Tr}\{\mathbf{A}_k\} \Big(\text{Tr}\{\mathbf{Z}_k\} \mathbf{z}_g(\mathbf{C}^{\mathrm{B}}, \mathbf{W}_k)
$$
$$
+ \text{Tr}\{\mathbf{R}_k\mathbf{C}^{\mathrm{B}}\} \mathbf{u}_g(\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}) \Big),
$$

(28)

$$
\frac{\partial \text{Tr}\{\mathbf{Z}_k\mathbf{Z}_k\}}{\partial \boldsymbol{\theta}} = \mathbf{u}_g \left(\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}} \boldsymbol{\Phi}^H \mathbf{C}^{\mathrm{R}} \boldsymbol{\Phi} \mathbf{C}_k^{\mathrm{RU}} \right)
$$
$$
+ \mathbf{u}_g \left(\mathbf{C}^{\mathrm{R}} \boldsymbol{\Phi} \mathbf{C}_k^{\mathrm{RU}} \boldsymbol{\Phi}^H \mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}} \right),
$$

(29)

$$\frac{\partial \mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k^H\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\right\}}{\partial\boldsymbol{\theta}} = 2\mathbf{z}_g\left(\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\mathbf{C}^{\mathrm{B}}, \mathbf{W}_k\right), \tag{30}$$

and

$$\frac{\partial \mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{R}_{k'}^H\right\}}{\partial\boldsymbol{\theta}} = 2\mathbf{z}_g\left(\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}, \mathbf{W}_{k'}\right). \tag{31}$$

Proof: The proof can be established by applying the chain rules and leveraging Lemma 2 and Lemma 3. Let us take $\mathrm{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_k\right\}$ for an example:

$$
\begin{aligned}
&\frac{\partial \mathrm{Tr}\left\{\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\right\}}{\partial\boldsymbol{\theta}}\\
&= \frac{\partial \mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial\boldsymbol{\theta}}\bigg|_{\mathbf{A}=\mathbf{C}^{\mathrm{R}},\mathbf{B}=\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}}\\
&\quad + \frac{\partial \mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial\boldsymbol{\theta}}\bigg|_{\mathbf{A}=\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}},\mathbf{B}=\mathbf{C}_k^{\mathrm{RU}}}\\
&= \mathbf{u}_g\left(\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\right)\\
&\quad + \mathbf{u}_g\left(\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}\right).
\end{aligned}
\tag{32}
$$

The other terms can be obtained by using a similar but cumbersome process, and therefore are omitted here for brevity. ∎

With the given results, we can readily obtain the gradient vectors $\frac{\partial \mathrm{E}_{l,k}}{\partial\boldsymbol{\theta}}$ and $\frac{\partial \mathrm{UI}_{k,k'}}{\partial\boldsymbol{\theta}}$ by using Lemmas 1, 2, 3 and chain rule. In the following, we calculate the gradient vector $\frac{\partial \mathrm{E}_{l,k}}{\partial\boldsymbol{\theta}}$, which is given by

$$
\begin{aligned}
\frac{\partial \mathrm{E}_{l,k}}{\partial\boldsymbol{\theta}} &= p_k^d\left(\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\right)^2\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k^H\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\right\}\\
&\quad \times \bigg(\mathbf{u}_g\left(\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\right)\\
&\quad\quad + \mathbf{u}_g\left(\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}\right)\bigg)\\
&\quad + 2p_k^d\left(\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\right)^2\mathrm{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_k\right\}\mathbf{z}_g\left(\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\mathbf{C}^{\mathrm{B}}, \mathbf{W}_k\right).
\end{aligned}
\tag{33}
$$

Then, following a similar process as (33) and after some algebraic manipulations, the gradient vector of $\frac{\partial \mathrm{UI}_{k,k'}}{\partial\boldsymbol{\theta}}$ can be obtained, which is given by

$$
\begin{aligned}
\frac{\partial \mathrm{UI}_{k,k'}}{\partial \boldsymbol{\theta}} = {} & 2\beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\big(\beta^{\mathrm{BR}}\big)^2 \mathrm{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_{k'}\right\} \mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\right\}\mathbf{z}_g(\mathbf{C}^{\mathrm{B}},\mathbf{W}_{k'}) \\
& + \beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\big(\beta^{\mathrm{BR}}\big)^2\big|\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\right\}\big|^2\mathbf{u}_g\left(\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}},\mathbf{C}_{k'}^{\mathrm{RU}}\right) \\
& + \beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\big(\beta^{\mathrm{BR}}\big)^2\big|\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\right\}\big|^2\mathbf{u}_g\left(\mathbf{C}^{\mathrm{R}},\mathbf{C}_k^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_{k'}^{\mathrm{RU}}\right) \\
& + \beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\big(\beta^{\mathrm{BR}}\big)^2\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}^H\right\}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathbf{u}_g(\mathbf{C}^{\mathrm{R}},\mathbf{C}_{k'}^{\mathrm{RU}}) \\
& + \beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\big(\beta^{\mathrm{BR}}\big)^2\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}^H\right\}\mathrm{Tr}\left\{\mathbf{Z}_{k'}\right\}\mathbf{u}_g(\mathbf{C}^{\mathrm{R}},\mathbf{C}_k^{\mathrm{RU}}) \\
& + 2\beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\big(\beta^{\mathrm{BR}}\big)^2\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathrm{Tr}\left\{\mathbf{Z}_{k'}\right\}\mathbf{z}_g\left(\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{C}^{\mathrm{B}},\mathbf{W}_{k'}\right) \\
& + \frac{\sigma^2}{\tau P_p}\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{R}_{k'}^H\right\}\mathbf{u}_g(\mathbf{C}^{\mathrm{R}},\mathbf{C}_k^{\mathrm{RU}}) \\
& + 2\frac{\sigma^2}{\tau P_p}\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathbf{z}_g\left(\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'},\mathbf{W}_{k'}\right).
\end{aligned}
\tag{34}
$$

Finally, by substituting the results of (28), (33), and $\frac{\partial \mathrm{UI}_{k,k'}}{\partial \boldsymbol{\theta}}$ into (24), we obtain the weighted sum rate's gradient with respect to $\boldsymbol{\theta}$. We can continuously tune the phase shift to obtain the maximum system performance as the functions always increase along with the gradient's direction. Furthermore, to reduce the number of iterations of the proposed algorithm, we adopt the accelerated gradient ascent algorithm, rather than the conventional gradient ascent one, and the detailed algorithm is presented in Algorithm 1.

Furthermore, to identify a feasible region for the NP-hard problem, the algorithm for initializing the phase shift design and transmission scheme is given by

$$
\max_{\{\theta_n\},\left\{p_k^d\right\},\{\chi_k\},\Gamma} \Gamma \tag{35a}
$$

$$
\text{s.t.} \quad \chi_k \geq \frac{\Gamma}{f_k^{-1}\left(\frac{R_k^{\mathrm{req}}\ln 2}{(1-\eta)}\right)}, \forall k, \tag{35b}
$$

$$
\big|e^{j\theta_n}\big| = 1, \forall n, \tag{35c}
$$

$$
\sum_{K=1}^{K} p_k^d \leq P^d. \tag{35d}
$$

Obviously, the feasible region can be determined only when $\Gamma \geq 1$. The process of finding the feasible region follows the same alternating optimization algorithm described in Algorithm 1, and therefore it is omitted for brevity.

D. Algorithm Analysis

1) Convergence: To ensure that the algorithm converges to a locally sub-optimal solution, it is necessary to demonstrate that the objective function consistently increases with the number of iterations, denoted as $\text{Obj}^{(i+1)} \geq \text{Obj}^{(i)}$. $\text{Obj}^{(i)}$ represents the objective function in the ith iteration. Given the phase shifts $\boldsymbol{\theta}^{(i)}$ in the ith iteration, the power allocation scheme is proved to approach the sub-optimal solution asymptotically. Defining the power vector $\mathbf{p}^{d,(i+1)}$ as the $(i+1)$th solution and the corresponding vector of SINRs as $\boldsymbol{\chi}_k^{(i+1)}(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)})$, we have

$$
\begin{aligned}
\sum_{k=1}^{K} & \bar{w}_k \bigg[\ln \left(\chi_k^{(i+1)}(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}) \right)^{(\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)})} \\
& + (\delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)}) \bigg] \\
\geq \sum_{k=1}^{K} & \bar{w}_k \bigg[\ln \left(\chi_k^{(i)}(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}) \right)^{(\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)})} \\
& + (\delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)}) \bigg] \\
= \sum_{k=1}^{K} & w_k \hat{R}_k(\mathbf{p}^{d,(i)}, \boldsymbol{\theta}^{(i)}).
\end{aligned}
\tag{36}
$$

Then, we combine (20) with (36) and prove that the objective function based on the $(i+1)$th power allocation is larger than that relying on the ith power allocation, which can be written as

$$
\begin{aligned}
\sum_{k=1}^{K} & w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}) \\
= \sum_{k=1}^{K} & \bar{w}_k \bigg[\ln \left(\chi_k^{(i+1)}(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}) \right)^{(\rho_k^{(i+1)} - \alpha_k \hat{\rho}_k^{(i+1)})} \\
& + (\delta_k^{(i+1)} - \alpha_k \hat{\delta}_k^{(i+1)}) \bigg] \\
\geq \sum_{k=1}^{K} & \bar{w}_k \bigg[\ln \left(\chi_k^{(i+1)}(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}) \right)^{(\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)})} \\
& + (\delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)}) \bigg] \\
= \sum_{k=1}^{K} & w_k \hat{R}_k(\mathbf{p}^{d,(i)}, \boldsymbol{\theta}^{(i)}).
\end{aligned}
\tag{37}
$$

With the given power allocation $\mathbf{p}^{d,(i+1)}$, we know that the objective function increases along

the direction of the gradient, and we have

$$\sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i+1)}) \geq \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}). \tag{38}$$

Finally, combining the results (38) with (37), the proof of the convergence of the proposed algorithm is completed, which is detailed as

$$\begin{aligned}
\mathrm{Obj}^{(i+1)} &= \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i+1)}) \\
&\geq \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}^{(i)}) \\
&\geq \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i)}, \boldsymbol{\theta}^{(i)}) = \mathrm{Obj}^{(i)}.
\end{aligned} \tag{39}$$

2) Algorithm Complexity: The complexity of our proposed algorithm primarily relies on two factors: calculating the gradient and solving the geometric programming (GP) problem in each iteration. Specifically, the complexity of the gradient ascent is to calculate the first-order derivative of the objective function, denoted as N_θ, while that of the GP problem depends on solving Problem (22) during each iteration [39]. Since GP problem involves $2K$ variables and $(2K + 1)$ constraints, the computational complexity can be approximated as $\mathcal{O}(N_a \times \max(2K + 1)(2K)^2, N_d)$, where N_a represents the number of alternating optimizations, and N_d pertains to the complexity for obtaining the first-order and second-order derivatives in Problem (22) [30]. Therefore, the complexity of our alternating optimization algorithm is $\mathcal{O}(N_a \times [\max\{(2K + 1)(2K)^2, N_d\} + N_\theta])$.

IV. Simulation Results

This section presents numerical results that validate the accuracy of our rigorous derivations compared to Monte Carlo simulations and demonstrate the performance improvement achieved by our proposed two-timescale scheme.

A. Simulation Setup

To tackle the blockage issue of cell-edge devices, a typical application scenario of the RIS-enabled mMIMO system is investigated. Specifically, an M-antenna BS is in central coordinates, and K devices randomly distributed on a semicircle of radius $r_d = 5$ m, with center at 10 meters

---

Algorithm 1 Alternating Iterative Optimization Algorithm

1: The iteration number i and error tolerance ζ is initialized as 1 and 0.001, respectively;

2: Solving Problem (35) to initialize the transmission power $\mathbf{p}^{d,(i)} = [p_1^{d,(i)}, p_2^{d,(i)}, ..., p_K^{d,(i)}]^T$ and phase shifts $\boldsymbol{\theta}^{(i)}$, calculate the weighted sum rate $\text{WSR}^{(i)} = \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i)}, \boldsymbol{\theta}^{(i)})$. Set $\text{WSR}^{(0)} = 0$;

3: **while** $\left(\text{WSR}^{(i)} - \text{WSR}^{(i-1)}\right)/\text{WSR}^{(i-1)} \geq \zeta$ **do**

4:     Update $\left\{\hat{w}_k^{(i)}, \forall k\right\}$;

5:     With given phase shifts $\boldsymbol{\theta}^{(i)}$, using the CVX package to solve Problem (22) and obtain transmission power $\mathbf{p}^{d,(i+1)}$;

6:     Set $n = 0$ and $a_0 = 0$. Initialize the iteration number $\boldsymbol{\theta}_n^{(i)}$ randomly and $\mathbf{t}_{-1} = \boldsymbol{\theta}_n^{(i)}$.

7:     **while** 1 **do**

8:         Calculate the gradient vector of $\boldsymbol{\theta}_n^{(i)}$, denoted as $\mathbf{f}'(\boldsymbol{\theta}_n^{(i)}) = \frac{\partial \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_n^{(i)}}$;

9:         Using the backtracking line search obtain the step size k_n;

10:        $\mathbf{t}_n = \boldsymbol{\theta}_n^{(i)} + k_n \mathbf{f}'(\boldsymbol{\theta}_n^{(i)})$, $a_{n+1} = \frac{1+\sqrt{4a_n^2+1}}{2}$;

11:        $\boldsymbol{\theta}_{n+1}^{(i)} = \mathbf{t}_n + \frac{(a_n-1)(\mathbf{t}_n - \mathbf{t}_{n-1})}{a_{i+1}}$;

12:        **if** $\sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}_{n+1}^{(i)}) - \sum_{k=1}^{K} w_k \hat{R}_k(\mathbf{p}^{d,(i+1)}, \boldsymbol{\theta}_n^{(i)}) < 1e^{-5}$ **then**

13:           $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}_{n+1}^{(i)}$, break;

14:        **end if**

15:        Update $n = n + 1$;

16:     **end while**

17:     Update $i = i+1$, obtain the weighted sum rate $\text{WSR}^{(i)}$ based on given transmission power $\mathbf{p}^{d,(i)}$ and phase shifts $\boldsymbol{\theta}^{(i)}$;

18: **end while**

---

from the RIS located in the two-dimensional plane [0,50]. The distance-based path-loss factors β^{BR} and β_k^{RU} are modeled as $\beta^{\text{BR}} = \beta_0(d^{\text{BR}})^{-2.2}$, $\beta_k^{\text{RU}} = \beta_0(d_k^{\text{RU}})^{-2.1}$, respectively, where $\beta_0 = 10^{-2}$, d^{BR} (m) and d_k^{RU} (m) represent the distance from the BS to the RIS, and from the RIS to the kth device. These parameters are similar to those in [25]. The correlation matrix model in [33] is employed. Furthermore, we assume that the communication is over the bandwidth of $B = 2$ MHz and the packet should be delivered below latency $T = 0.1$ ms, and therefore the blocklength is $L = B \times T = 200$. The transmission power and pilot power are $P^d = 1$ W and $p^p = 0.1$ W, and the power of noise is $\sigma^2 = B \times 1.381 \times 10^{-23} \times 290 \times 10^{\frac{9}{10}}$ W. The weight assigned to the kth device w_k is randomly generated over the interval (0, 1). The required data rate and decoding error rate are $R_k^{\text{req}} = 0.2$ bit/s/Hz and $\varepsilon_k = 10^{-7}, \forall k$.
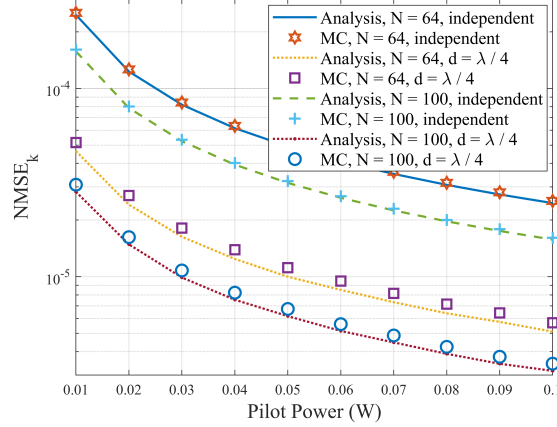
Fig. 3: NMSE of device versus the pilot power with $M = 100$.

B. Channel Estimation Error

In Fig. 3, the NMSE based on LMMSE is evaluated in the cases of independent (i.e., $\mathbf{C}^{\mathrm{B}} = \mathbf{I}_M$ and $\mathbf{C}^{\mathrm{R}} = \mathbf{C}_k^{\mathrm{RU}} = \mathbf{I}_N$) and correlated channels, respectively. Apparently, increasing the pilot power can significantly improve the estimation performance. Besides, we observe that increasing the number of reflective elements of the RIS can effectively enhance the accuracy of the estimated channel. This phenomenon can be attributed to the fact that the increasing reflective elements N can exploit the channel gain, which results in smaller channel estimation errors. More importantly, based on the results given in (6) and (7), both $\mathbf{R}_k$ and $\mathbf{C}^{\mathrm{B}}$ are non-diagonal matrices, and thus $\mathrm{Tr}\{\mathbf{R}_k\mathbf{C}^{\mathrm{B}}\}$ with correlated channel is larger than that of independent channel, which results in the improved estimation performance.

C. Lower Bound and Ergodic Rate

By averaging 10^4 Monte Carlo simulation results, we investigate the tightness between our derived lower bound and achievable ergodic rate under the conditions of spatial distance $d = \frac{\lambda}{4}$ and transmission power $p_k^d = 0.2$ W, $\forall k$. As illustrated in Fig. 4, increasing the number of the antennas of BS and the number of reflecting elements of RIS can enhance the system performance. More importantly, our derived lower bound can approach the achievable ergodic rate in any case. As a result, compared to phase shift design and power optimization relying on the intricate formulation of the ergodic rate, the resource allocation relying on the derived closed-form expression is more practical and applicable.
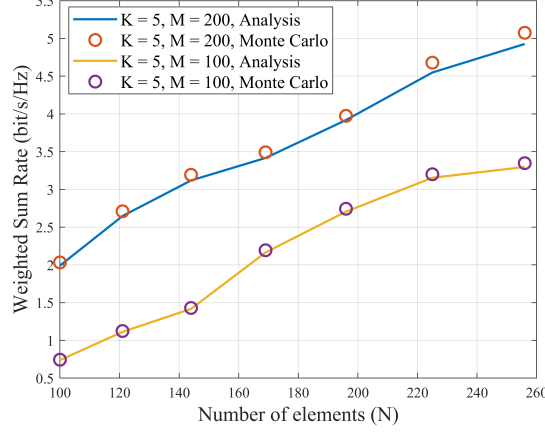
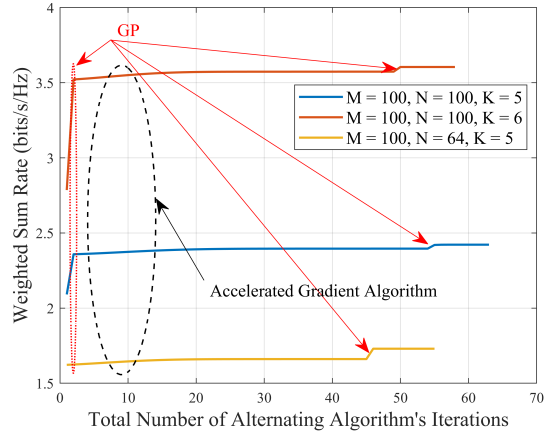Fig. 4: Weighted sum rate under versus number of reflective elements of RIS.

Fig. 5: Total Number of Alternating Algorithm's Iteration.

D. Convergence of the Alternating Optimization Algorithm

We depict the weighted sum rate of each iteration in Fig. 5 to validate our rigorous convergence proof with spatial distance of $d = \frac{\lambda}{4}$, transmission power of $P^d = 1$ W, and pilot power of $p_k^p = 0.1$ W, $\forall k$. It is observed from Fig. 5 that the proposed algorithm converges within about 55 iterations (i.e., about 2 or 3 alternating optimizations), which confirms the accuracy of our rigorous derivations and demonstrates the effectiveness of our proposed algorithm.

E. Effect of Phase Shifts

We evaluate the weighted sum rate's cumulative distribution function (CDF) to assess the system performance enhancement of our alternating optimization algorithm, as depicted in Fig.
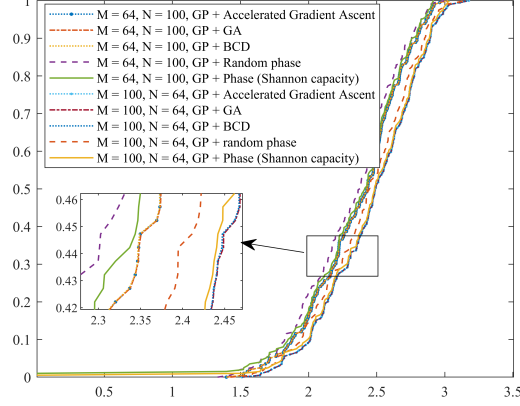
Fig. 6: CDF of weighted sum rate with $d = \frac{\lambda}{4}$ and $K = 5$.

6. In the scenario where no feasible region exists for our proposed algorithm, meaning that not all devices can fulfill the requirements of both the rate and DEP requirements, the corresponding weighted sum rates will be set to 0. Besides, we compare the proposed algorithm with the following algorithms:

- GP + genetic algorithm (GA) [40]: GP is adopt to optimize the transmission power and GA is used for tuning phase shifts of the RIS.

- GP + block coordinate descent (BCD) [41]: GP and BCD are employed to devise the transmission power and phase shifts, respectively.

- GP + phase (Shannon capacity): The transmission power based on short packet transmission is obtained by using GP and the phase shifts are tuned relying on Shannon capacity.

- GP + random phase: We devise the transmission power and phase shifts by using GP and random phase, respectively.

Obviously, our proposed algorithm achieves superior performance over the scheme that only relies on power allocation and approaches the complex benchmark algorithms (i.e., combining GP with genetic GA and combining GP with BCD). Furthermore, The baseline based on Shannon capacity cannot guarantee all devises' requirement, as the phase shifts are obtained by Shannon capacity, ignoring the impact of decoding error probability. More importantly, it is worth noting that, for the blocked devices, equipping more active antennas at the BS is more effective than equipping more passive reflective elements at the RIS.
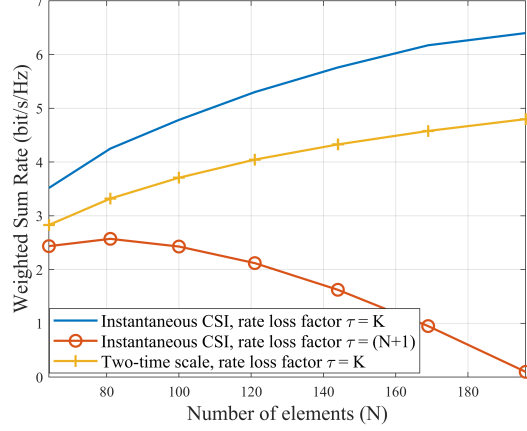
Fig. 7: System performance versus various numbers of reflective elements with $d = \frac{\lambda}{4}$ and $K = 5$.

F. Effect of Number of Elements

In Fig. 7, to better demonstrate the superiority of the two-time scale scheme, we compare its performance with that relying on instantaneous CSI scheme [29], where the phase shift design can be obtained by GA. Furthermore, we also depict the ideal but unrealistic result where instantaneous CSI can be obtained by using K blocklength and $N+1$ blocklength, respectively. For the ideal instantaneous scheme and the proposed two-timescale scheme (i.e., pilot overhead $\tau = K$), the weighted sum rates are both increasing with the increasing reflecting elements N, while the system performance relying on practical instantaneous CSI (i.e., $\tau = N + 1$) improves slightly as the RIS's reflecting elements increase from 64 to 81 but degrades beyond 81. Apparently, the ideal scheme based on instantaneous CSI exhibits superior performance, particularly for scenarios with large reflective elements N. This discrepancy arises because the instantaneous CSI-based scheme exploits optimal transmission design using both large-scale and small-scale fading factors in the $(L - K)$ blocklength. Furthermore, compared to the practical scheme upon instantaneous CSI (i.e., pilot overhead $\tau = N + 1$), the two-timescale scheme exhibits superior performance. This is because that when the number of reflective elements increases, the blocklength for channel estimation in the instantaneous CSI-based method grows as well. As a result, with a large of passive reflective elements N, a significant portion of the blocklength is allocated for channel estimation, leaving only a limited number of symbols for data transmission. This results in a significant rate loss in the regime of large N. Therefore, Fig. 7 demonstrates that the two-timescale scheme is a practical solution in the short packet

transmission.

V. CONCLUSION

The two-timescale scheme for the RIS-enabled finite blocklength transmission was investigated while considering the channel estimation error and spatial correlation of reflective elements. Firstly, the channel was estimated relying on the LMMSE method, and the performance in terms of NMSE was analyzed. Then, a tight lower bound of data rate was derived for facilitating resource allocation. To tackle the intrinsically weighted sum rate maximization problem, GP and gradient descent methods were combined to devise the transmission scheme and phase shifts design. Simulation results validated that the BS can perform resource allocation relying on our derived results, which significantly reduces the information exchange and processing latency. Furthermore, compared to the practical case upon instantaneous CSI, the two-timescale scheme exhibits its superiority in terms of performance improvement and effectiveness.

Since the decoding error probability and latency are also the key performance indicators of wireless communication, it is valuable to investigate the transmission scheme and phase shifts to minimize the decoding error probability while satisfying the latency and data rate requirements, which would be left for our future work.

APPENDIX A
USEFUL LEMMAS

Lemma 4: Let $\mathbf{V} \in \mathbb{C}^{M \times N}$ be a complex matrix, where the ith row and jth column entry $[\mathbf{V}]_{(i,j)}$ follows a complex Gaussian distribution $\mathcal{CN}(0,1)$. For a given matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$, the expectation of $\mathbf{V}\mathbf{X}\mathbf{V}^H$ is expressed as

$$\mathbb{E}\left\{\mathbf{V}\mathbf{X}\mathbf{V}^H\right\} = \mathrm{Tr}\left\{\mathbf{X}\right\}\mathbf{I}_M. \tag{40}$$

Proof: The expectation of $\left[\mathbf{V}\mathbf{X}\mathbf{V}^H\right]_{i,j}$, $(i \neq j)$, can be given by

$$
\begin{aligned}
&\mathbb{E}\left\{\left[\mathbf{V}\mathbf{X}\mathbf{V}^H\right]_{i,j}\right\} \\
&= \mathbb{E}\left\{\sum_{l=1}^{N}\sum_{k=1}^{N}[\mathbf{V}]_{i,k}[\mathbf{X}]_{k,l}\left[\mathbf{V}^H\right]_{l,i}\right\} \\
&= \sum_{l=1}^{N}\sum_{k=1}^{N}\mathbb{E}\left\{[\mathbf{V}]_{i,k}\left[\mathbf{V}^H\right]_{l,j}\right\}[\mathbf{X}]_{k,l} = 0.
\end{aligned}
\tag{41}
$$

For $i = j$, we have

$$
\mathbb{E}\left\{\left[\mathbf{V}\mathbf{X}\mathbf{V}^H\right]_{i,i}\right\}
$$
$$
=\mathbb{E}\left\{\sum_{l=1}^{N}\sum_{k=1}^{N}[\mathbf{V}]_{i,k}[\mathbf{X}]_{k,l}\left[\mathbf{V}^H\right]_{l,i}\right\} \tag{42}
$$
$$
=\sum_{k=1}^{N}\mathbb{E}\left\{[\mathbf{V}]_{i,k}\left[\mathbf{V}^H\right]_{i,k}\right\}[\mathbf{X}]_{k,k} = \mathrm{Tr}\left\{\mathbf{X}\right\}.
$$

By combining the results of (41) and (42) into a matrix, we complete this proof. ∎

Lemma 5: Consider a matrix $\mathbf{V} \in \mathbb{C}^{M \times N}$, and each element of $[\mathbf{V}]$ is a complex Gaussian distribution $\mathcal{CN}(0, 1)$. Given a matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$ and a hermitian matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$, we have

$$
\mathbb{E}\left\{\mathbf{V}^H\mathbf{C}\mathbf{V}\mathbf{X}\mathbf{V}^H\mathbf{C}\mathbf{V}\right\} = \mathrm{Tr}\left\{\mathbf{X}\right\}\mathrm{Tr}\left\{\mathbf{C}^2\right\}\mathbf{I}_N + |\mathrm{Tr}\left\{\mathbf{C}\right\}|^2\mathbf{X}. \tag{43}
$$

Proof: Define $\mathbf{v}_i \in \mathbb{C}^{M \times 1}$ as the ith column of matrix $\mathbf{V}$. The expectation of $[\mathbf{V}^H\mathbf{C}\mathbf{V}\mathbf{X}\mathbf{V}^H\mathbf{C}\mathbf{V}]_{i,j}$, $(i \neq j)$, is

$$
\mathbb{E}\left\{\left[\mathbf{V}^H\mathbf{C}\mathbf{V}\mathbf{X}\mathbf{V}^H\mathbf{C}\mathbf{V}\right]_{i,j}\right\}
$$
$$
=\mathbb{E}\left\{\sum_{l=1}^{N}\sum_{k=1}^{N}\mathbf{v}_i^H\mathbf{C}\mathbf{v}_k[\mathbf{X}]_{k,l}\mathbf{v}_l^H\mathbf{C}\mathbf{v}_j\right\} \tag{44}
$$
$$
=[\mathbf{X}]_{i,j}\mathbb{E}\left\{\mathbf{v}_i^H\mathbf{C}\mathbf{v}_i\right\}\mathbb{E}\left\{\mathbf{v}_j^H\mathbf{C}\mathbf{v}_j\right\}
$$
$$
=[\mathbf{X}]_{i,j}|\mathrm{Tr}\left\{\mathbf{C}\right\}|^2.
$$

For $i = j$, we have

$$
\mathbb{E}\left\{\left[\mathbf{V}^H\mathbf{C}\mathbf{V}\mathbf{X}\mathbf{V}^H\mathbf{C}\mathbf{V}\right]_{i,i}\right\}
$$
$$
= \mathbb{E}\left\{\mathbf{v}_i^H\mathbf{C}\mathbf{v}_i[\mathbf{X}]_{i,i}\mathbf{v}_i^H\mathbf{C}\mathbf{v}_i\right\}
$$
$$
+ \mathbb{E}\left\{\sum_{k \neq i}^{N}\mathbf{v}_i^H\mathbf{C}\mathbf{v}_k[\mathbf{X}]_{k,k}\mathbf{v}_k^H\mathbf{C}\mathbf{v}_i\right\} \tag{45}
$$
$$
\overset{(a)}{=} [\mathbf{X}]_{i,i}|\mathrm{Tr}\left\{\mathbf{C}\right\}|^2 + \mathrm{Tr}\left\{\mathbf{X}\right\}\mathrm{Tr}\left\{\mathbf{C}^2\right\},
$$

where (a) exploits $\mathbb{E}\left\{\left|\mathbf{v}_i^H\mathbf{C}\mathbf{v}_i\right|^2\right\} = |\mathrm{Tr}\left\{\mathbf{C}\right\}|^2 + \mathrm{Tr}\left\{\mathbf{C}^2\right\}$, the proof of which is detailed in Lemma 2 in [42].

By combining the results of (44) and (45) into a matrix, the result in (43) is obtained. ∎

APPENDIX B

PROOF OF THEOREM 1

With the given receiving signal $\mathbf{y}_k^p$, the estimated channel $\mathbf{h}_k$ based on the LMMSE method is given by

$$
\begin{aligned}
\hat{\mathbf{h}}_k = \mathbb{E}\left\{\mathbf{h}_k\right\} + \mathrm{Cov}\left\{\mathbf{h}_k, \mathbf{y}_k^p\right\} \mathrm{Cov}^{-1}\left\{\mathbf{y}_k^p, \mathbf{y}_k^p\right\} \\
\times\left(\mathbf{y}_k^p - \mathbb{E}\left\{\mathbf{y}_k^p\right\}\right),
\end{aligned}
\tag{46}
$$

where $\mathbb{E}\left\{\mathbf{h}_k\right\}$ and $\mathbb{E}\left\{\mathbf{y}_k^p\right\}$ are equal to 0. By using Lemma 5, we obtain the expressions of $\mathrm{Cov}\left\{\mathbf{h}_k, \mathbf{y}_k^p\right\}$ and $\mathrm{Cov}\left\{\mathbf{y}_k^p, \mathbf{y}_k^p\right\}$, which is given by

$$
\begin{aligned}
\mathrm{Cov}\left\{\mathbf{h}_k, \mathbf{y}_k^p\right\} &= \mathbb{E}\left\{\left(\mathbf{h}_k - \mathbb{E}\left\{\mathbf{h}_k\right\}\right)\left(\mathbf{y}_k^p - \mathbb{E}\left\{\mathbf{y}_k^p\right\}\right)^H\right\} \\
&= \beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \mathrm{Tr}\left\{\boldsymbol{\Phi} \mathbf{C}_k^{\mathrm{RU}} \boldsymbol{\Phi}^H \mathbf{C}^{\mathrm{R}}\right\} \mathbf{C}^{\mathrm{B}} \\
&\triangleq \beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \mathrm{Tr}\left\{\mathbf{Z}_k\right\} \mathbf{C}^{\mathrm{B}},
\end{aligned}
\tag{47}
$$

and

$$
\begin{aligned}
& \mathrm{Cov}\left\{\mathbf{y}_k, \mathbf{y}_k^p\right\} \\
=&\mathbb{E}\left\{\mathbf{h}_k \mathbf{h}_k^H + \frac{1}{\tau p^p}\mathbf{N}^p \mathbf{q}_k \mathbf{q}_k^H (\mathbf{N}^p)^H\right\} \\
=&\beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \mathrm{Tr}\left\{\mathbf{Z}_k\right\} \mathbf{C}^{\mathrm{B}} + \frac{\sigma^2}{\tau P_p}\mathbf{I}_M.
\end{aligned}
\tag{48}
$$

By substituting the results of (47) and (48) into (46), we obtain the result in (7).

APPENDIX C

PROOF OF THEOREM 2

The power of desired signal, leakage signal, and the interference are derived, respectively. The first term $\left|\mathbb{E}\left\{\mathbf{h}_k^H \hat{\mathbf{h}}_k\right\}\right|^2$ can be given by

$$
\begin{aligned}
& \left|\mathbb{E}\left\{\mathbf{h}_k^H \hat{\mathbf{h}}_k\right\}\right|^2 \\
=& \left|\mathbb{E}\left\{\mathbf{h}_k^H \mathbf{R}_k \mathbf{h}_k\right\}\right|^2 = \left|\mathrm{Tr}\left\{\mathbb{E}\left\{\mathbf{R}_k \mathbf{h}_k \mathbf{h}_k^H\right\}\right\}\right|^2 \\
=& \left|\mathrm{Tr}\left\{\mathbf{A}_k\right\}\right|^2,
\end{aligned}
\tag{49}
$$

where $\mathbf{A}_k$ is equal to $\beta_k^{\mathrm{RU}} \beta^{\mathrm{BR}} \mathrm{Tr}\left\{\mathbf{Z}_k\right\} \mathbf{R}_k \mathbf{C}^{\mathrm{B}}$.

The second term $\mathbb{E}\left\{\left|\mathbf{h}_k^H \hat{\mathbf{h}}_k\right|^2\right\}$ can be calculated as

$$
\mathbb{E}\left\{\left|\mathbf{h}_k^H \hat{\mathbf{h}}_k\right|^2\right\}
$$

$$
=\mathbb{E}\left\{\left|\mathbf{h}_k^H \mathbf{R}_k\left(\mathbf{h}_k + \frac{1}{\sqrt{\tau p^p}}\mathbf{N}^p \mathbf{q}_k\right)\right|^2\right\} \tag{50}
$$

$$
=\mathbb{E}\left\{\left|\mathbf{h}_k^H \mathbf{R}_k \mathbf{h}_k\right|^2 + \frac{1}{\tau P_p}\left|\mathbf{h}_k^H \mathbf{R}_k \mathbf{N}^p \mathbf{q}_k\right|^2\right\}.
$$

By using Lemma 2 in [42] and Lemma 5, the first term is given by

$$
\mathbb{E}\left\{\left|\mathbf{h}_k^H \mathbf{R}_k \mathbf{h}_k\right|^2\right\}
$$

$$
=\left(\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\right)^2\mathbb{E}\left\{\left|\tilde{\mathbf{v}}_k^H\left(\mathbf{C}_k^{\mathrm{RU}}\right)^{\frac{1}{2}}\boldsymbol{\Phi}^H\left(\mathbf{C}^{\mathrm{R}}\right)^{\frac{1}{2}}\tilde{\mathbf{G}}^H\left(\mathbf{C}^{\mathrm{B}}\right)^{\frac{1}{2}}\mathbf{R}_k\left(\mathbf{C}^{\mathrm{B}}\right)^{\frac{1}{2}}\tilde{\mathbf{G}}\left(\mathbf{C}^{\mathrm{R}}\right)^{\frac{1}{2}}\boldsymbol{\Phi}\left(\mathbf{C}_k^{\mathrm{RU}}\right)^{\frac{1}{2}}\tilde{\mathbf{v}}_k\right|^2\right\}
$$

$$
=\left(\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\right)^2\left(\left|\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\right\}\right|^2 + \mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k^H\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\right\}\right)\mathbb{E}\left\{\left|\tilde{\mathbf{v}}_k^H\left(\mathbf{C}_k^{\mathrm{RU}}\right)^{\frac{1}{2}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\left(\mathbf{C}_k^{\mathrm{RU}}\right)^{\frac{1}{2}}\tilde{\mathbf{v}}_k\right|^2\right\}
$$

$$
=\left(\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}\right)^2\left(\left|\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\right|^2 + \mathrm{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_k\right\}\right)\left(\left|\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k\right\}\right|^2 + \mathrm{Tr}\left\{\mathbf{Z}_k\right\}\right),
$$

$$
\tag{51}
$$

and the second term in (50) is

$$
\mathbb{E}\left\{\left|\mathbf{h}_k^H \mathbf{R}_k \mathbf{N}^p \mathbf{q}_k\right|^2\right\}
$$

$$
=\mathbb{E}\left\{\mathbf{h}_k^H \mathbf{R}_k \mathbf{N}^p \mathbf{q}_k \mathbf{q}_k^H (\mathbf{N}^p)^H \mathbf{R}_k \mathbf{h}_k\right\} \tag{52}
$$

$$
=\mathrm{Tr}\left\{\mathbb{E}\{\mathbf{h}_k \mathbf{h}_k^H\}\mathbf{R}_k \mathbf{R}_k^H\right\}
$$

$$
=\sigma^2 \beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_k \mathbf{R}_k^H\right\}.
$$

Then, the interfering term $\mathbb{E}\left\{\left|\mathbf{h}_k^H \hat{\mathbf{h}}_{k'}\right|^2\right\}$ is given by

$$
\mathbb{E}\left\{\left|\mathbf{h}_k^H \hat{\mathbf{h}}_{k'}\right|^2\right\} = \mathbb{E}\left\{\mathbf{h}_k^H\mathbb{E}\left\{\hat{\mathbf{h}}_{k'}\hat{\mathbf{h}}_{k'}^H\right\}\mathbf{h}_k\right\}
$$

$$
= \beta_{k'}^{\mathrm{RU}}\beta_k^{\mathrm{RU}}\left(\beta^{\mathrm{BR}}\right)^2\times
$$

$$
\mathbb{E}\left\{\left|\tilde{\mathbf{v}}_k^H\left(\mathbf{C}_k^{\mathrm{RU}}\right)^{\frac{1}{2}}\boldsymbol{\Phi}^H\left(\mathbf{C}^{\mathrm{R}}\right)^{\frac{1}{2}}\tilde{\mathbf{G}}^H\left(\mathbf{C}^{\mathrm{B}}\right)^{\frac{1}{2}}\mathbf{R}_{k'}\left(\mathbf{C}^{\mathrm{B}}\right)^{\frac{1}{2}}\tilde{\mathbf{G}}\left(\mathbf{C}^{\mathrm{R}}\right)^{\frac{1}{2}}\boldsymbol{\Phi}\left(\mathbf{C}_{k'}^{\mathrm{RU}}\right)^{\frac{1}{2}}\tilde{\mathbf{v}}_{k'}\right|^2\right\}
$$

$$
+ \frac{\beta^{\mathrm{BR}}\beta_k^{\mathrm{RU}}}{\tau P_p}\mathbb{E}\left\{\left|\tilde{\mathbf{v}}_k^H\left(\mathbf{C}_k^{\mathrm{RU}}\right)^{\frac{1}{2}}\boldsymbol{\Phi}^H\left(\mathbf{C}^{\mathrm{R}}\right)^{\frac{1}{2}}\tilde{\mathbf{G}}^H\left(\mathbf{C}^{\mathrm{B}}\right)^{\frac{1}{2}}\mathbf{R}_{k'}\mathbf{N}^p \mathbf{q}_{k'}\right|^2\right\}.
$$

$$
\tag{53}
$$

The first term in (53) is given by

$$
\mathbb{E}\left\{\left|\tilde{\mathbf{v}}_k^H\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\boldsymbol{\Phi}^H\big(\mathbf{C}^{\mathrm{R}}\big)^{\frac{1}{2}}\tilde{\mathbf{G}}^H\big(\mathbf{C}^{\mathrm{B}}\big)^{\frac{1}{2}}\mathbf{R}_{k'}\big(\mathbf{C}^{\mathrm{B}}\big)^{\frac{1}{2}}\tilde{\mathbf{G}}\big(\mathbf{C}^{\mathrm{R}}\big)^{\frac{1}{2}}\boldsymbol{\Phi}\big(\mathbf{C}_{k'}^{\mathrm{RU}}\big)^{\frac{1}{2}}\tilde{\mathbf{v}}_{k'}\right|^2\right\}
$$

$$
= \left|\mathrm{Tr}\left\{\mathbf{R}_{k'}\mathbf{C}^{\mathrm{B}}\right\}\right|^2 \mathbb{E}\left\{\tilde{\mathbf{v}}_k^H\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\mathbf{C}_{k'}^{\mathrm{RU}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\tilde{\mathbf{v}}_k\right\}
$$

$$
\quad + \mathrm{Tr}\left\{\mathbf{Z}_{k'}\right\}\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}^H\right\}\mathbb{E}\left\{\tilde{\mathbf{v}}_k^H\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\tilde{\mathbf{v}}_k\right\} \tag{54}
$$

$$
= \left|\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\right\}\right|^2\mathrm{Tr}\left\{\mathbf{Z}_k\mathbf{Z}_{k'}\right\} + \mathrm{Tr}\left\{\mathbf{Z}_{k'}\right\}\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}^H\right\}\mathrm{Tr}\left\{\mathbf{Z}_k\right\},
$$

and the second term in (53) is

$$
\mathbb{E}\left\{\left|\tilde{\mathbf{v}}_k^H\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\boldsymbol{\Phi}^H\big(\mathbf{C}^{\mathrm{R}}\big)^{\frac{1}{2}}\tilde{\mathbf{G}}^H\big(\mathbf{C}^{\mathrm{B}}\big)^{\frac{1}{2}}\mathbf{R}_{k'}\mathbf{N}^p\mathbf{q}_{k'}\right|^2\right\}
$$

$$
= \sigma^2\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{R}_{k'}^H\right\}\mathbb{E}\left\{\tilde{\mathbf{v}}_k^H\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\boldsymbol{\Phi}^H\mathbf{C}^{\mathrm{R}}\boldsymbol{\Phi}\big(\mathbf{C}_k^{\mathrm{RU}}\big)^{\frac{1}{2}}\tilde{\mathbf{v}}_k\right\} \tag{55}
$$

$$
= \sigma^2\mathrm{Tr}\left\{\mathbf{C}^{\mathrm{B}}\mathbf{R}_{k'}\mathbf{R}_{k'}^H\right\}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}.
$$

By substituting $\mathbb{E}\left\{|n_k|^2\right\} = \sigma^2$ and these expressions into (14), we obtain the expression of SINR.

APPENDIX D
PROOF OF LEMMA 1

With the given matrices $\mathbf{A}$ and $\mathbf{B}$, we substitute $\boldsymbol{\Phi} = \mathrm{diag}(\mathbf{b})$ into $\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}$ and have a more explicit form, which is given by

$$
\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\} = \sum_i\sum_j[\mathbf{A}]_{ij}[\boldsymbol{\Phi}]_{jj}[\mathbf{B}]_{ji}\big[\boldsymbol{\Phi}^H\big]_{ii}
$$

$$
= \mathbf{b}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\mathbf{b}. \tag{56}
$$

Then, based on the result in (56), it is convenient to derive the gradient in terms of the nth element θ_n by using the chain rule, which is given by

$$
\frac{\mathrm{Tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}^H\right\}}{\partial\theta_n} = -je^{-j\theta_n}\left[\left(\mathbf{A}\odot\mathbf{B}^T\right)\mathbf{b}\right]_n
$$

$$
\quad + j\left[\mathbf{b}^H\left(\mathbf{A}\odot\mathbf{B}^T\right)\right]_ne^{j\theta_n}. \tag{57}
$$

According to the above result of the nth element's gradient, the proof is completed by collecting (57) into a vector.

APPENDIX E

PROOF OF LEMMA 2

For ease of expression, we define $\mathbf{C}^{\mathrm{B}}\left(\beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathbf{C}^{\mathrm{B}} + \frac{\sigma^2}{\tau P_p}\mathbf{I}_M\right)^{-1}$ as $\mathbf{W}_k$. Then, we obtain the gradient of $\mathrm{Tr}\left\{\mathbf{X}\mathbf{R}_k\right\}$ with respect to the nth element θ_n by using the chain rule, which is detailed as

$$
\begin{aligned}
\frac{\mathrm{Tr}\left\{\mathbf{X}\mathbf{R}_k\right\}}{\partial\theta_n} &= \frac{\partial\left(\beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\right)\mathrm{Tr}\left\{\mathbf{X}\mathbf{W}_k\right\}}{\partial\theta_n} \\
&= \beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\frac{\partial\mathrm{Tr}\left\{\mathbf{Z}_k\right\}}{\partial\theta_n}\mathrm{Tr}\left\{\mathbf{X}\mathbf{W}_k\right\} \\
&\quad + \beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\frac{\partial\mathrm{Tr}\left\{\mathbf{X}\mathbf{W}_k\right\}}{\partial\theta_n}.
\end{aligned}
\tag{58}
$$

The first term in (58) can be readily obtained based on the result in (57). Then, the second term can be calculated by using $\partial(\mathbf{A})^{-1} = -(\mathbf{A})^{-1}\partial(\mathbf{A})(\mathbf{A})^{-1}$, which is given by

$$
\begin{aligned}
&\frac{\partial\mathrm{Tr}\left\{\mathbf{X}\mathbf{C}^{\mathrm{B}}\left(\beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathbf{C}^{\mathrm{B}} + \frac{\sigma^2}{\tau p^p}\mathbf{I}_M\right)^{-1}\right\}}{\partial\theta_n} \\
&= -\mathrm{Tr}\left\{\mathbf{X}\mathbf{W}_k\mathbf{W}_k\right\} \times \beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\frac{\partial\mathrm{Tr}\left\{\mathbf{Z}_k\right\}}{\partial\theta_n}.
\end{aligned}
\tag{59}
$$

Substituting the results of (59) into (58) and combining the nth element's gradient to a vector, we have

$$
\begin{aligned}
\frac{\mathrm{Tr}\left\{\mathbf{X}\mathbf{R}_k\right\}}{\partial\boldsymbol{\theta}} &= \beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}} \times \mathbf{u}_g(\mathbf{C}^{\mathrm{R}}, \mathbf{C}_k^{\mathrm{RU}}) \times \Big(\mathrm{Tr}\left\{\mathbf{X}\mathbf{W}_k\right\} \\
&\quad - \beta_k^{\mathrm{RU}}\beta^{\mathrm{BR}}\mathrm{Tr}\left\{\mathbf{Z}_k\right\}\mathrm{Tr}\left\{\mathbf{X}\mathbf{W}_k\mathbf{W}_k\right\}\Big).
\end{aligned}
\tag{60}
$$

The result in Lemma 2 is completed.

REFERENCES

[1] X. Li, D. Li, J. Wan, A. V. Vasilakos, C.-F. Lai, and S. Wang, "A review of industrial wireless networks in the context of industry 4.0," *Wireless Netw.*, vol. 23, no. 1, pp. 23–41, Jan. 2017.

[2] B. Chang, L. Zhang, L. Li, G. Zhao, and Z. Chen, "Optimizing resource allocation in URLLC for real-time wireless control systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8916–8927, Sep. 2019.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[4] J. M. Hamamreh, E. Basar, and H. Arslan, "OFDM-subcarrier index selection for enhancing security and reliability of 5G URLLC services," *IEEE Access*, vol. 5, pp. 25 863–25 875, Nov. 2017.

[5] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.

[6] J. Östman, A. Lancho, G. Durisi, and L. Sanguinetti, "URLLC with massive MIMO: Analysis and design at finite blocklength," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6387–6401, Oct. 2021.

[7] Q. Peng, H. Ren, C. Pan, N. Liu, and M. Elkashlan, "Resource allocation for uplink cell-free massive MIMO enabled URLLC in a smart factory," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 553–569, Jan. 2022.

[8] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5861–5871, Sep. 2021.

[9] E. Shi, J. Zhang, J. Zhang, D. W. K. Ng, and B. Ai, "Decentralized coordinated precoding design in cell-free massive MIMO systems for URLLC," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2638–2642, Feb. 2023.

[10] Q. Peng, H. Ren, C. Pan, N. Liu, and M. Elkashlan, "Resource allocation for cell-free massive MIMO enabled URLLC downlink systems," in *Proc. IEEE Int. Conf. Commun. China (ICCC)*, 2022, pp. 838–843.

[11] 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.

[12] 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.*, vol. 59, no. 6, pp. 14–20, Jun. 2021.

[13] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?" *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, May 2022.

[14] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. Di Renzo, A. L. Swindlehurst, R. Zhang *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[15] J. Wang, S. Gong, Q. Wu, and S. Ma, "RIS-aided MIMO systems with hardware impairments: Robust beamforming design and analysis," *IEEE Trans. Wireless Commun.*, 2023.

[16] 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.

[17] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in uav-ris wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, Jul. 2020.

[18] B. Al-Nahhas, M. Obeed, A. Chaaban, and M. J. Hossain, "RIS-aided cell-free massive MIMO: Performance analysis and competitiveness," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2021, pp. 1–6.

[19] P. Ni, M. Li, R. Liu, and Q. Liu, "Partially distributed beamforming design for RIS-aided cell-free networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13 377–13 381, Dec. 2022.

[20] A. Ranjha and G. Kaddoum, "URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and UAV positioning," *IEEE Internet Things Journal*, vol. 8, no. 6, pp. 4618–4627, Mar. 2021.

[21] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, "Average rate analysis of RIS-aided short packet communication in URLLC systems," in *IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2021, pp. 1–6.

[22] G. S. Chandra, R. K. Singh, S. Dhok, P. K. Sharma, and P. Kumar, "Downlink URLLC system over spatially correlated RIS channels and electromagnetic interference," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1950–1954, Sep. 2022.

[23] H. Ren, K. Wang, and C. Pan, "Intelligent reflecting surface-aided URLLC in a factory automation scenario," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 707–723, Jan. 2022.

[24] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, "Joint sum rate and blocklength optimization in RIS-aided short packet URLLC systems," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1838–1842, Aug. 2022.

[25] 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.

[26] 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.

[27] A. Kammoun, A. Chaaban, M. Debbah, M.-S. Alouini *et al.*, "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.

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

[29] K. Zhi, C. Pan, H. Ren, K. Wang, M. Elkashlan, M. Di Renzo, R. Schober, H. V. Poor, J. Wang, and L. Hanzo, "Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI," *IEEE Trans. Inf. Theory*, vol. 69, no. 5, pp. 3001–3033, May 2023.

[30] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.

[31] 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.

[32] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. lett.*, vol. 5, no. 9, pp. 369–371, 2001.

[33] 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. 2020.

[34] H. Guo and V. K. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser MISO systems," *IEEE Trans. Signal Processing*, vol. 70, pp. 3964–3977, 2022.

[35] H. Ren, C. Pan, K. Wang, Y. Deng, M. Elkashlan, and A. Nallanathan, "Achievable data rate for URLLC-enabled UAV systems with 3-D channel model," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1587–1590, Dec. 2019.

[36] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Maximizing the geometric mean of user-rates to improve rate-fairness: Proper vs. improper Gaussian signaling," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 295–309, Jan. 2022.

[37] M. Abughalwa, H. D. Tuan, D. N. Nguyen, H. V. Poor, and L. Hanzo, "Finite-blocklength RIS-aided transmit beamforming," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 12 374–12 379, Nov. 2022.

[38] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.

[39] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optim. Eng.*, vol. 8, no. 1, pp. 67–127, May 2007.

[40] 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.

[41] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, 2020.

[42] 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.