Is RIS-Aided Massive MIMO Promising With ZF Detectors and Imperfect CSI?

Kangda Zhi, *Graduate Student Member, IEEE*, Cunhua Pan, *Member, IEEE*, Gui Zhou, *Graduate Student Member, IEEE*, Hong Ren, *Member, IEEE*, Maged Elkashlan, *Senior Member, IEEE*, and Robert Schober, *Fellow, IEEE*

***Abstract*— This paper provides a theoretical framework for understanding the performance of reconfigurable intelligent surface (RIS)-aided massive multiple-input multiple-output (MIMO) with zero-forcing (ZF) detectors under imperfect channel state information (CSI). We first introduce a low-overhead minimum mean square error (MMSE) channel estimator, and then derive and analyze closed-form expressions for the uplink achievable rate. Our analytical results demonstrate that: 1) regardless of the RIS phase shift design, the rate of all users scales at least on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, where M and N are the numbers of antennas and reflecting elements, respectively; 2) by aligning the RIS phase shifts to one user, the rate of this user can at most scale on the order of $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$; 3) either M or the transmit power can be reduced inversely proportional to N, while maintaining a given rate. Furthermore, we propose two low-complexity majorization-minimization (MM)-based algorithms to optimize the sum user rate and the minimum user rate, respectively, where closed-form solutions are obtained in each iteration. Finally, simulation results validate the accuracy of all derived analytical results. Our simulation results also show that the maximum sum rate can be closely approached by simply aligning the RIS phase shifts to an arbitrary user.**

***Index Terms*— Reconfigurable intelligent surface (RIS), intelligent reflecting surface (IRS), massive MIMO, majorization-minimization (MM), ZF, imperfect CSI.**

Manuscript received 30 September 2021; revised 22 January 2022; accepted 25 February 2022. Date of publication 3 August 2022; date of current version 16 September 2022. This work was supported in part by the National Key Research and Development Project (2019YFE0123600), in part by the National Natural Science Foundation of China under Grant 62101128, in part by the Basic Research Project of Jiangsu Provincial Department of Science and Technology (BK20210205), and in part by the High Level Personal Project of Jiangsu Province (JSSCBS20210105). The work of Kangda Zhi was supported by the China Scholarship Council. The work of Robert Schober was supported in part by the German Science Foundation (DFG) under Grant SCHO 831/15-1. An earlier version of this paper was presented in part at the IEEE International Conference on Communications (ICC) from May 16 to May 20, 2022 [1]. *(Corresponding author: Cunhua Pan.)*

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

Cunhua Pan and Hong Ren are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: cpan@seu.edu.cn; hren@seu.edu.cn).

Robert Schober is with the Institute for Digital Communications, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91054 Erlangen, Germany (e-mail: robert.schober@fau.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3196097.

Digital Object Identifier 10.1109/JSAC.2022.3196097

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) has been widely recognized as a cornerstone technology for the fifth-generation (5G) and beyond wireless communications [2]–[8]. Thanks to its spatial multiplexing gains, massive MIMO system can simultaneously provide high quality of service for multiple users on the same time-frequency resource. Massive MIMO also has some other appealing properties, e.g., the transmit power can be reduced inversely proportional to the number of antennas without sacrificing the achievable rate.

However, conventional massive MIMO still has some drawbacks. The first one is the blockage problem. Due to the complex environment and user mobility, communication links may be blocked, in which case the channel strength could be severely degraded. Another problem is the high cost and energy consumption of the active radio-frequency (RF) chains. Massive MIMO commonly employs hundreds of antennas, each of which will be connected to a RF chain. Hence, this system incurs high hardware cost and energy consumption.

The recently developed technology of reconfigurable intelligent surfaces (RISs) [9]–[14], also referred to as intelligent reflecting surfaces (IRSs), is a promising solution for tackling the above two issues in massive MIMO systems. On the one hand, since the RIS is a small, thin and light surface, it can be flexibly deployed at a carefully selected location with a favorable propagation environment. Therefore, RISs enable additional high-quality communication paths to overcome the blockage problem. On the other hand, RISs are comprised of low-cost passive reflecting elements, which are much cheaper than active RF chains. Therefore, it is envisioned that RISs are beneficial for improving the energy efficiency of conventional massive MIMO systems.

Due to these appealing features, RIS-aided massive MIMO has gained growing research interests with many activities, focusing on various applications and different perspectives, such as channel estimation [15], dual-polarized transmission [16], millimeter wave (mmWave) communications [17], hardware impairments [18], multi-RISs co-design [19], cell-free systems [20], antenna selection [21], and power scaling law analysis [22]–[24].

To fully understand the potential of RISs, it is essential to draw theoretical insights from information-theoretical expressions, which rigorously demonstrate the

impact of the various system parameters. Fundamental information-theoretical expressions for conventional massive MIMO systems have been provided in, e.g., [3]–[5]. It was shown that the achievable rate of conventional massive MIMO systems with M antennas scales on the order of $\mathcal{O}\left(\log_2\left(M\right)\right)$. This naturally raises the question what is the corresponding scaling law for massive MIMO systems with the integration of RISs. To answer this question, explicitly analytical rate expressions are required. It has already been shown that in RIS-aided single-user systems with N reflecting elements, the achievable rate could scale as $\mathcal{O}\left(\log_2\left(N^2\right)\right)$ [11], [25], or even $\mathcal{O}\left(\log_2\left(N^4\right)\right)$ [26] if two RISs cooperate. Similar scaling orders were also reported for some other RIS-aided communication scenarios, such as the RIS-aided relay [27], RIS with scattering parameter analysis [28], and RISs with hardware impairments [29], [30]. However, these works focused on the simple single-user case, and cannot be easily generalized to multi-user systems.

In fact, it is challenging to provide an insightful analysis for the rate scaling order of RIS-aided multi-user systems. This is because the resulting signal-to-interference-plus-noise ratio (SINR) expressions are more complicated and more involved than the interference-free signal-to-noise ratio (SNR) expressions for single-user systems, and also because the optimal RISs passive beamforming vectors cannot be given in closed form in the case of multiple users. Some initial results were provided in [23] and [24] by considering RIS-aided massive MIMO with simple maximal ratio combining (MRC). For uncorrelated Rayleigh fading channels, it was proved that the achievable rate scales only as $\mathcal{O}\left(\log_2\left(1\right)\right)$ with respect to N. This is due to the severe multi-user interference, since the common RIS-base station (BS) channel is used by all users. To tackle this issue, most recently, the authors in [31] firstly revealed that a rate scaling order $\mathcal{O}\left(\log_2\left(MN\right)\right)$ can be achieved with zero-forcing (ZF), which demonstrates the huge potential of ZF detectors in RIS-aided massive MIMO systems.

However, there are two main limitations in [31]. Firstly, ideal channel state information (CSI) of the aggregated channel including the superimposition of the direct channel and the reflected channel, was assumed. The impact of pilot overhead and channel estimation errors on the system performance are still unknown. Whether the rate scaling order obtained based on perfect CSI still holds in the presence of imperfect CSI deserves further study. Secondly, the authors in [31] only considered some initial performance analysis and RIS phase shift optimization, which lacks further insightful analysis. Therefore, this work aims to provide an analytical framework to gain an in-depth analysis for the performance of RIS-aided massive MIMO systems with ZF detectors under the realistic assumption of imperfect CSI.

Specifically, in this work, we first propose a low-overhead channel estimation scheme, in which the required pilot length is independent of N. We next perform a comprehensive theoretical analysis to reveal the explicit rate scaling order and answer the fundamental question whether the RIS-aided massive MIMO with ZF detectors is promising or not. Finally, based on majorization-minimization (MM) algorithms, we respectively optimize the RIS phase shifts to maximize the sum user rate and the minimum user rate. The detailed contributions are summarized as follows.

1) Low-overhead channel estimation: We propose a minimum mean square error (MMSE)-based method to estimate the aggregated channel in the systems, which is a superimposition of cascaded RIS channels and the direct channels. The length of pilots only needs to be no smaller than the number of users. We also analyze the impacts of various system parameters on the mean square error (MSE).

2) Reveal rate scaling orders: We derive the closed-form ergodic rate expression and its insightful lower and upper bounds. The lower bound shows that the data rates of all users are guaranteed to be on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, regardless of the RIS phase shift design. The upper bound shows that the data rate of a specific user can be on the order of $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$, if the RIS phase shift is designed to align its beamforming to that user. We also demonstrate that these two analytical results are robust to RIS phase shift quantization errors.

3) Answer the question whether the considered system is promising or not: Based on the analytical results, we prove that RIS-aided massive MIMO systems with ZF detector are promising for three applications. It can provide ultra-high network throughput according to the high data rate scaling order for all users; it can help reduce M inversely proportional to N without sacrificing the data rate, which helps avoid the power hungry RF chains and is promising for green communications; it can help all users reduce their transmit power inversely proportional to N while maintaining high data rates, which is promising for IoT applications.

4) Low-complexity RIS optimization: We design the RIS phase shifts to maximize the sum user rate and minimum user rate, relying on the MM algorithm with closed-form solution in each iteration. We also show that aligning RIS phase shifts to an arbitrary user is an effective heuristic approach for maximizing the sum user rate. In addition, we demonstrate that maximizing the sum user rate can also ensure a high minimum user rate.

The rest of this paper is organized as follows. Section II describes the system and channel model. Section III proposes the MMSE channel estimation scheme. Section IV theoretically proves that RIS-aided massive MIMO is promising with ZF detectors. Section V proposes the MM algorithm for solving the sum user rate and minimum user rate maximization problems. Section VI provides extensive simulations to verify the correctness of analytical results and the effectiveness of proposed optimization algorithms. Finally, Section VII concludes this work.

Notations: Boldface lower case and upper case letters denote the vectors and matrices, respectively. The inverse, conjugate transpose, conjugate and transpose of matrix $\mathbf{X}$ are denoted by $\mathbf{X}^{-1}$, $\mathbf{X}^H$, $\mathbf{X}^*$, $\mathbf{X}^T$, respectively. The (m,n)-th and (m,m)-th elements of the matrix are represented by $[\mathbf{X}]_{(m,n)}$ and $[\mathbf{X}]_{mm}$. $\mathbf{X} \succ \mathbf{0}$ and $\mathbf{X} \succeq \mathbf{0}$ respectively denote that $\mathbf{X}$ is definite positive and semi-positive. $\mathcal{O}$ denotes the standard

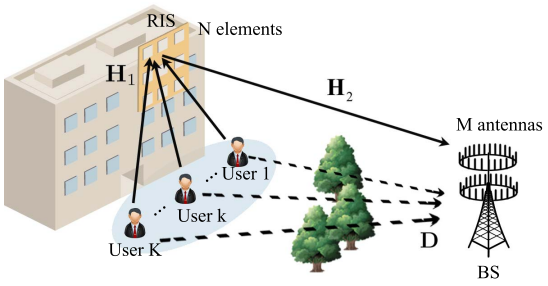


Fig. 1. Massive MIMO systems assisted by an RIS.

big-O notation. $\lambda_{\max}(\mathbf{X})$ and $\angle\mathbf{X}$ denote the maximal eigenvalue and the phase of matrix $\mathbf{X}$. $\mathbb{E}\{\cdot\}$ and $\mathrm{Cov}\{\cdot\}$ denote the mean and covariance operators.

II. System and Channel Model

As shown in Fig. 1, the uplink transmission of an RIS-assisted massive MIMO system is considered. The considered system consists of K users with a single antenna, a BS with $M > K$ antennas,[1] and an RIS with N reflecting elements. Besides, we assume a quasi-static channel model with each channel coherence interval (CCI) spanning τ_c time slots. In each CCI, the instantaneous channel between the users and the RIS, and that between the RIS and the BS are denoted by $\mathbf{H}_1 \in \mathbb{C}^{N\times K}$ and $\mathbf{H}_2 \in \mathbb{C}^{M\times N}$, respectively. Then, the cascaded user-RIS-BS channel is $\mathbf{G} = \mathbf{H}_2\mathbf{\Phi}\mathbf{H}_1$, where $\mathbf{\Phi} = \mathrm{diag}\left\{e^{j\theta_1},\ldots,e^{j\theta_N}\right\}$ is the RIS phase shift matrix.[2] Meanwhile, the direct channels between the users and the BS are denoted by $\mathbf{D} \in \mathbb{C}^{M\times K}$. Finally, in each CCI, the instantaneous *aggregated* channels from the users to the BS are given by $\mathbf{Q} = \mathbf{G} + \mathbf{D} \in \mathbb{C}^{M\times K}$.

It has been shown in [24] that it is better to place an RIS close to the users rather than close to the BS in the massive MIMO systems. Therefore, in this paper, we assume that the RIS is deployed on the facade of a tall building in the proximity of the users, as illustrated in Fig. 1. Since the RIS has a certain height and the distance between each user and the RIS is short, the strength of the non-line-of-sight (NLoS) components is much weaker than that of the line-of-sight (LoS) components for the user-RIS channels [34]. Accordingly, user-RIS channels are expected to be LoS-dominant. For analytical tractability, we assume that the user-RIS channels are purely LoS as follows[3]

$$\mathbf{H}_1 = \left[\sqrt{\alpha_1}\,\overline{\mathbf{h}}_1,\ldots,\sqrt{\alpha_K}\,\overline{\mathbf{h}}_K\right], \tag{1}$$

where $\alpha_k, \forall k$ is the large-scale path loss factor for user k, and $\overline{\mathbf{h}}_k \in \mathbb{C}^{N\times 1}$ is the deterministic LoS channel between user k and the RIS.

[1]If $M \le K$, user scheduling methods can be applied. In each time slot, we can select $\tilde{K} = M - 1$ users to be served. As a result, the condition $M > \tilde{K}$ holds, and all conclusions obtained in the rest of this paper still hold by substituting K with $\tilde{K}$.

[2]For analytical tractability, unit independent amplitudes are assumed in this work. The incorporation of more practical amplitude models, as used in [32], [33], will be left for our future work.

[3]The gap between the data rate achieved by a purely LoS channel and that achieved by a LoS-dominant channel (i.e., a Rician fading channel with a relatively large Rician factor) is negligible, as shown in [24], [25], [35].

Since the RIS is installed close to the users, it may be located far away from the BS. Therefore, both LoS and NLoS transmission paths would exist in $\mathbf{H}_2$. As a result, we characterize the RIS-BS channel by Rician fading, which is expressed as

$$\mathbf{H}_2 = \sqrt{\beta/(\delta+1)}(\sqrt{\delta}\,\overline{\mathbf{H}}_2 + \widetilde{\mathbf{H}}_2), \tag{2}$$

where β is the path loss factor, and δ is the Rician factor which represents the ratio between the power of LoS component $\overline{\mathbf{H}}_2$ and the power of NLoS component $\widetilde{\mathbf{H}}_2$. The elements of $\widetilde{\mathbf{H}}_2$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance. For a rich-scattering environment, there is $\delta \to 0$ and then the RIS-BS channel reduces to a Rayleigh fading channel containing only NLoS paths. For a scattering-free environment, there is $\delta \to \infty$ and then the RIS-BS channel is purely LoS.

Finally, since the users might be located far away from the BS, and rich scatterers (trees, cars, buildings and so on) are distributed on the ground, the channels between the users and the BS are assumed to be Rayleigh fading as in [25]. Thus, we have

$$\mathbf{D} \triangleq [\mathbf{d}_1,\ldots,\mathbf{d}_K] = \widetilde{\mathbf{D}}\mathbf{\Omega}^{1/2}, \tag{3}$$

where $\mathbf{d}_k = \sqrt{\gamma_k}\widetilde{\mathbf{d}}_k$ is the channel between user k and the BS with large-scale fading coefficient γ_k and small-scale fading vector $\widetilde{\mathbf{d}}_k$ comprised of i.i.d. complex Gaussian random variables with zero mean and unit variance. Here, $\mathbf{\Omega} = \mathrm{diag}\{\gamma_1,\ldots,\gamma_K\}$ and $\widetilde{\mathbf{D}} = [\widetilde{\mathbf{d}}_1,\ldots,\widetilde{\mathbf{d}}_K]$.

The two-dimensional uniform rectangular array (URA) model is considered in this paper to characterize the LoS channels [11]. For an $L\times 1$ array response vector $\mathbf{a}_L$, we have

$$\mathbf{a}_L(\vartheta^a,\vartheta^e) = \tilde{\mathbf{a}}_{L_x}(\sin\vartheta^e\sin\vartheta^a) \otimes \tilde{\mathbf{a}}_{L_y}(\cos\vartheta^e), \tag{4}$$

where ϑ^a and ϑ^e denote the azimuth and elevation angles of arrival (AoA) or the corresponding angles of departure (AoD), $L = L_x \times L_y$, and $\tilde{\mathbf{a}}_b(c) \triangleq [1,\ldots,e^{j2\pi\frac{d}{\lambda}(b-1)c}]^T$, $b \in \{L_x, L_y\}$. Besides, it is worth noting that $\mathbf{a}_L^H(\vartheta^a,\vartheta^e)\,\mathbf{a}_L(\vartheta^a,\vartheta^e) = L$.

Now, based on the definition of the array response vector in (4), we can respectively express the LoS channels $\overline{\mathbf{h}}_k$ and $\overline{\mathbf{H}}_2$ as follows

$$\overline{\mathbf{h}}_k = \mathbf{a}_N\left(\varphi_{k,r}^a,\varphi_{k,r}^e\right), 1 \le k \le K, \tag{5}$$

$$\overline{\mathbf{H}}_2 \triangleq \mathbf{a}_M\mathbf{a}_N^H = \mathbf{a}_M(\phi_r^a,\phi_r^e)\,\mathbf{a}_N^H(\varphi_t^a,\varphi_t^e), \tag{6}$$

where $\varphi_{k,r}^a, \varphi_{k,r}^e$ denote the AoA from user k to the RIS and ϕ_r^a, ϕ_r^e $(\varphi_t^a, \varphi_t^e)$ denote the AoA (AoD) from the RIS to the BS.[4]

III. Channel Estimation

In the considered system, there are two kinds of channel parameters, namely the fast varying small-scale/instantaneous parameters ($\widetilde{\mathbf{H}}_2$ and $\widetilde{\mathbf{D}}$) and the slowly varying

[4]Based on the adopted system model, this work serves as a useful preliminary tool to theoretically understand the properties and potentials of RIS-aided massive MIMO systems based on ZF detectors. For the extension to more realistic and complex models, machine learning techniques could be applied, which will be left for our future work.

large-scale/statistical parameters (AoA, AoD, path-loss factors, and Rician factors). Since the large-scale parameters vary much slower than the small-scale parameters, they are easier to measure. Specifically, the AoA and AoD can be calculated based on location information, and the path-loss and Rician factors can be measured by environmental sensors. Besides, since the required frequency of estimation of large-scale parameters is much lower than that of small-scale parameters, the corresponding overhead can be ignored. Therefore, only the overhead and estimation of the small-scale channel parameters are studied in this paper.

The small-scale channel parameters are estimated by the BS using a pilot-based method. For conventional massive MIMO systems, only the $M \times K$ user-BS direct channel $\mathbf{D}$ needs to be estimated, and the minimum pilot sequence length is $\tau = K$. In RIS-aided massive MIMO systems, the required pilot overhead could be prohibitive due to the extremely large channel dimension of $M \times N$ in the RIS-BS link. Since both M and N could be large in the considered system, a low-overhead channel estimation scheme is highly desirable. To reduce the pilot overhead, we extend a classic channel estimation method from conventional massive MIMO systems [4] to the considered system, by estimating only the instantaneous aggregated user-BS channel $\mathbf{Q} \in \mathbb{C}^{M\times K}$. As a result, the minimum pilot sequence length is still $\tau = K$, independent of M and N. Meanwhile, the estimated channel is sufficient for the design of ZF detectors at the BS.[5]

Specifically, in each CCI, the K users are assigned mutually orthogonal pilot sequences with length $\tau \geq K$. The pilot sequence of user k is denoted by $\mathbf{s}_k \in \mathbb{C}^{\tau\times 1}$. Let $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_K]$, where $\mathbf{S}^H\mathbf{S} = \mathbf{I}_K$ due to the orthogonality. Then, at the beginning of each CCI, τ time slots are used for the K users to transmit the pilot signal $\mathbf{S}$ to the BS. The received $M \times \tau$ pilot signal at the BS can be given by $\mathbf{Y}_p = \sqrt{\tau p}\mathbf{Q}\mathbf{S}^H + \mathbf{N}$, where p is the common average transmission power of each user during the channel estimation stage, and $\mathbf{N}$ is the noise matrix whose elements are i.i.d. Gaussian variables following $\mathcal{CN}(0, \sigma^2)$. Then, we can obtain the observation vector for the channel of user k by multiplying the term $\frac{1}{\sqrt{\tau p}}\mathbf{s}_k$ to $\mathbf{Y}_p$, as follows

$$\mathbf{y}_p^k = \tfrac{1}{\sqrt{\tau p}}\mathbf{Y}_p\mathbf{s}_k = \mathbf{q}_k + \tfrac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k, \tag{7}$$

where $\mathbf{q}_k$, the k-th column of $\mathbf{Q}$, denotes the aggregated channel of user k.

Lemma 1: Channel $\mathbf{q}_k$ and noise $\frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k$ in (7) are complex Gaussian distributed, where $\mathbf{q}_k \sim \mathcal{CN}(\sqrt{\frac{\alpha_k\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k, \left(N\frac{\alpha_k\beta}{\delta+1} + \gamma_k\right)\mathbf{I}_M)$, and $\frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma^2}{\tau p}\mathbf{I}_M)$.

Proof: Please refer to Appendix A. ∎

From Lemma 1, it is seen that the considered channel is still Gaussian distributed as conventional massive MIMO systems [7, Eq. (1)], but with the different mean and variance. Therefore, we can still apply the well-known MMSE estimator to obtain the channel estimate of $\mathbf{q}_k$.

[5]We note that the proposed channel estimation method is not sufficient for instantaneous CSI-based RIS design since the small-scale parameters inside $\mathbf{Q}$ are not estimated. However, this method is sufficient for statistical CSI-based RIS design which only requires knowledge of the large-scale parameters.

Theorem 1: Based on the observation vector, the MMSE estimate of channel $\mathbf{q}_k$ is given by

$$\hat{\mathbf{q}}_k = \sqrt{\frac{\alpha_k\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k + \kappa_k\left(\sqrt{\frac{\alpha_k\beta}{\delta+1}}\widetilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k + \mathbf{d}_k + \frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k\right), \tag{8}$$

where $\kappa_k = \frac{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}} \in (0,1)$. Denote the estimation error as $\mathbf{e}_k = \mathbf{q}_k - \hat{\mathbf{q}}_k$, where the error $\mathbf{e}_k$ is independent of the estimate $\hat{\mathbf{q}}_k$. Then, the MSE matrix for the channel estimation is

$$\mathbf{MSE}_k = \mathbb{E}\left\{\mathbf{e}_k\mathbf{e}_k^H\right\} = \frac{1}{\frac{1}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k}+\frac{\tau p}{\sigma^2}}\mathbf{I}_M \triangleq \epsilon_k\mathbf{I}_M. \tag{9}$$

Proof: Please refer to Appendix B. ∎

Based on (9), the MSE can be calculated as $\mathrm{MSE}_k = \mathrm{Tr}\left\{\mathbf{MSE}_k\right\} = \frac{M}{\frac{1}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k}+\frac{\tau p}{\sigma^2}}$. Clearly, the MSE is a decreasing function of τ, p, and δ, but an increasing function of M, N, α_k, β, γ_k, and σ^2. This is because $\frac{\tau p}{\sigma^2}$ represents the pilot SNR, and increasing its value improves the estimation quality. δ is the Rician factor, and increasing its value makes the RIS-aided channels more deterministic and therefore decreases the estimation error. Also, the increase of N introduces more communication paths between the users and the BS, which also increases the estimation error.

Note that in the absence of the RIS (i.e., $\alpha_k = \beta = 0, \forall k$) or for a purely LoS RIS-BS channel ($\delta \to \infty$), the MSE matrix in (9) reduces to $\mathbf{MSE}_k = \frac{\gamma_k}{1+\frac{\tau p}{\sigma^2}\gamma_k}\mathbf{I}_M$, which is the same as for conventional massive MIMO systems [4]. Let $\hat{\mathbf{Q}} = [\hat{\mathbf{q}}_1, \ldots, \hat{\mathbf{q}}_K]$ denote the estimated aggregated channel of the K users. Then, based on (8), we have

$$\hat{\mathbf{Q}} = \sqrt{\frac{\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1 + \sqrt{\frac{\beta}{\delta+1}}\widetilde{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1\mathbf{\Upsilon} + \widetilde{\mathbf{D}}\mathbf{\Omega}^{1/2}\mathbf{\Upsilon} + \frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{S}\mathbf{\Upsilon}, \tag{10}$$

where $\mathbf{\Upsilon} = \mathrm{diag}\left\{\kappa_1, \ldots, \kappa_K\right\}$.

IV. Ergodic Rate Analysis

In the transmission phase, the K users transmit symbols $\mathbf{x} = [x_1, \ldots, x_K]^T$ where $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, and the received signal at the BS can be expressed as

$$\mathbf{y} = \sqrt{p}\mathbf{Q}\mathbf{x} + \mathbf{n} = \sqrt{p}\hat{\mathbf{Q}}\mathbf{x} + \sqrt{p}\mathcal{E}\mathbf{x} + \mathbf{n}, \tag{11}$$

where $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_M\right)$ and $\mathcal{E} \triangleq [\mathbf{e}_1, \ldots, \mathbf{e}_K] = \mathbf{Q} - \hat{\mathbf{Q}}$. As in [4], [5], we assume that each user has a common average data transmission power p for simplicity, and this power is the same as that used in the channel estimation stage. To eliminate the multi-user interference, the BS adopts the linear ZF detectors $\mathbf{A} = \hat{\mathbf{Q}}(\hat{\mathbf{Q}}^H\hat{\mathbf{Q}})^{-1} = [\mathbf{a}_1, \ldots, \mathbf{a}_K]$, which leads to $\mathbf{A}^H\hat{\mathbf{Q}} = \mathbf{I}_K$.[6] Then, in each CCI, the BS detects the

[6]Even though ZF is a sub-optimal solution, it has low implementation complexity with respect to the number of antennas. Therefore, it is suitable for the considered massive MIMO system.

received signal as follows

$$\mathbf{r} = \mathbf{A}^H\mathbf{y} = \sqrt{p}\mathbf{x} + \sqrt{p}\mathbf{A}^H\mathcal{E}\mathbf{x} + \mathbf{A}^H\mathbf{n}, \tag{12}$$

whose k-th entry can be further expressed as

$$r_k = \sqrt{p}x_k + \sqrt{p}\sum\nolimits_{i=1}^{K}\mathbf{a}_k^H\mathbf{e}_i x_i + \mathbf{a}_k^H\mathbf{n}. \tag{13}$$

A. Achievable Data Rate

Based on (13), the accurate ergodic rate of user k can now be given by

$$R_k = \tau^o\,\mathbb{E}\left\{\log_2\left(1+\frac{p}{p\sum_{i=1}^{K}\left|\mathbf{a}_k^H\mathbf{e}_i\right|^2+\sigma^2\left\|\mathbf{a}_k^H\right\|^2}\right)\right\}, \tag{14}$$

where a factor $\tau^o \triangleq \frac{\tau_c-\tau}{\tau_c}$ captures the rate loss caused by pilot overhead, and the expectation is taken over random channel components in $\hat{\mathbf{Q}}$. It is difficult to derive an exact expression of (14) due to the expectation operator before the logarithm symbol. Since the function $f(x) = \log_2(1+1/x)$ is convex of x, we utilize the Jensen's inequality to obtain the following lower bound

$$\begin{aligned} R_k &\geq \underline{R_k}(\mathbf{\Phi}) \\ &\overset{(a)}{=} \tau^o\log_2\left(1+\frac{p}{p\sum_{i=1}^{K}\mathbb{E}\left\{\mathbf{a}_k^H\mathbb{E}\left\{\mathbf{e}_i\mathbf{e}_i^H\right\}\mathbf{a}_k\right\}+\sigma^2\mathbb{E}\{\left\|\mathbf{a}_k^H\right\|^2\}}\right) \qquad (15)\\ &\overset{(b)}{=} \tau^o\log_2\left(1+\frac{p}{(p\sum_{i=1}^{K}\epsilon_i+\sigma^2)\,\mathbb{E}\{[(\hat{\mathbf{Q}}^H\hat{\mathbf{Q}})^{-1}]_{kk}\}}\right), \qquad (16)\end{aligned}$$

where ϵ_i is defined in (9), (a) utilizes the independence between the channel estimate and the estimation errors, and (b) is due to the result in (9) and $\left\|\mathbf{a}_k^H\right\|^2 = \left[\mathbf{A}^H\mathbf{A}\right]_{kk} = [(\hat{\mathbf{Q}}^H\hat{\mathbf{Q}})^{-1}]_{kk}$.

Theorem 2: The achievable rate of user k is lower bounded by (17), as shown at the bottom of the next page, where $\mathbf{\Lambda} = \frac{\beta}{\delta+1}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon} + \mathbf{\Omega}\mathbf{\Upsilon}^2 + \frac{\sigma^2}{\tau p}\mathbf{\Upsilon}^2$.

Proof: Please refer to Appendix C. ■

The rate expression in Theorem 2 depends only on the slowly varying statistical CSI. Therefore, (17) enables us to design the phase shifts of the RIS only relying on the long-term CSI. Accordingly, we only need to update the RIS's phase shifts over a large time scale, which could effectively reduce overhead and computational complexity. Before the design of the phase shifts, we first analyze (17) to shed some light on the benefits of the RIS, and to answer the question whether RIS-aided massive MIMO is promising or not.

B. Conventional Systems Without RIS

Corollary 1: When the RIS is switched off (i.e., $\alpha_k = \beta = 0, \forall k$), the data rate (17) reduces to

$$\underline{R_k}^{w/o} = \tau^o\log_2\left(1+\frac{p(M-K)}{p\sum_{i=1}^{K}\frac{1}{\frac{\tau p}{\sigma^2}+\frac{1}{\gamma_i}}+\sigma^2}\times\frac{\gamma_k^2}{\gamma_k+\frac{\sigma^2}{\tau p}}\right). \tag{18}$$

When the RIS is switched off, the RIS-aided massive MIMO systems degrade to the conventional massive MIMO systems with Rayleigh fading channels ($\mathbf{Q} \to \mathbf{D}$), which has been studied in [4]. As expected, the obtained rate (18) is the same as [4, Eq. (42)]. Based on (18), it can be seen that the rate is on the order of $\mathcal{O}(\log_2(M))$, and the rate can maintain a non-zero value when the power is scaled down proportionally to $p = E_u/\sqrt{M}$, as the number of antennas $M \to \infty$, where E_u is a constant. Specifically, we have

$$\lim\nolimits_{p=\frac{E_u}{\sqrt{M}},M\to\infty}\ \underline{R_k}^{w/o} \to \tau^o\log_2\left(1+\tau E_u^2\gamma_k^2\sigma^{-4}\right). \tag{19}$$

Note that the achievable rate in (18) and power scaling law in (19) will serve as baselines and help us identify the benefits brought by introducing an RIS.

C. What's New After Integrating an RIS?

The order of magnitude of $\underline{R_k}(\mathbf{\Phi})$ in (17) with respect to M is $\mathcal{O}(\log_2(M))$, since ϵ_k and $\mathbf{\Lambda}$ are independent of M. However, it is challenging to determine how $\underline{R_k}(\mathbf{\Phi})$ scales with N, due to the unknown value of $\mathbf{\Phi}$ and the inverse operator. For tractability, we propose an insightful lower bound $\underline{\underline{R_k}}$ for $\underline{R_k}(\mathbf{\Phi})$ in the following.

Corollary 2: A $\mathbf{\Phi}$-independent lower bound $\underline{\underline{R_k}}$ is given by

$$\underline{R_k}(\mathbf{\Phi}) \geq \underline{\underline{R_k}} = \tau^o\log_2\left(1+\frac{p(M-K)}{(p\sum_{i=1}^{K}\epsilon_i+\sigma^2)[\mathbf{\Lambda}^{-1}]_{kk}}\right), \tag{20}$$

where equality holds when $\delta = 0$, and the gap $\underline{R_k}(\mathbf{\Phi}) - \underline{\underline{R_k}}$ enlarges after optimizing $\mathbf{\Phi}$. Besides, (20) can be approximated as

$$\underline{\underline{R_k}} \approx \tau^o\log_2\left(1+\frac{p(M-K)}{p\sum_{i=1}^{K}\epsilon_i+\sigma^2}\times\frac{\left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k\right)^2}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}}\right), \tag{21}$$

which scales on the order of $\mathcal{O}(\log_2(MN))$.

Proof: Please refer to Appendix D. ■

Interestingly, if we treat $N\frac{\alpha_k\beta}{\delta+1}+\gamma_k$ as a new path-loss factor, (21) possesses the same form as (18). This reveals two fundamental impacts of the RIS: i) Positive effect: RIS enhances the channel strength by a factor $N\frac{\alpha_k\beta}{\delta+1}$; ii) Negative effect: RIS results in larger channel estimation errors ϵ_k. However, the channel strength always increases with N since $\frac{(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k)^2}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}}$ grows without bound as $N \to \infty$, but the estimation error saturates to $\epsilon_k \to \frac{\sigma^2}{\tau p}$ as $N \to \infty$. Therefore, for large N, the benefits of the RIS outweigh its drawbacks in massive MIMO systems.

Corollary 2 proves that even with imperfect CSI, RIS-aided massive MIMO systems can achieve an ergodic rate at least on the order of $\mathcal{O}(\log_2(MN))$. This promising gain comes from the additional N paths contributed by the RIS for each user, such that more signals can be collected by the BS. Compared with the order $\mathcal{O}(\log_2(M))$ in conventional systems, Corollary 2 proves that much higher capacity can be achieved after integrating an RIS. More importantly, the scaling law $\mathcal{O}(\log_2(MN))$ indicates that if we want to maintain a fixed rate, the number of antennas can be reduced inversely proportional to the number of RIS elements. For

better understanding, we provide a quantitative relationship for a special case.

Corollary 3: When $\delta = 0$ and for large N, to achieve $\mathrm{SNR}_k = \mathrm{C}_0$ for a given N, the required number of antennas M is approximately given by

$$M \approx \frac{\mathrm{C}_0(K+\tau)\sigma^2}{\tau p\left(N\alpha_k\beta+\gamma_k\right)} + K = 2\mathrm{C}_0\frac{\sigma^2}{p} \times \frac{1}{N\alpha_k\beta+\gamma_k} + K, \text{ if } \tau = K. \quad (22)$$

Proof: When $\delta = 0$, we have $\underline{R_k}\left(\mathbf{\Phi}\right) = \underline{\underline{R_k}}$. Then, using (21), for large N, we have $\epsilon_k \approx \frac{\sigma^2}{\tau p}$, and $\mathrm{SNR}_k \approx \frac{p(M-K)}{K\frac{\sigma^2}{\tau}+\sigma^2}\left(N\alpha_k\beta+\gamma_k\right)$. Solving the equation $\mathrm{SNR}_k = \mathrm{C}_0$ completes the proof. ■

Corollary 3 corresponds to the scenarios with rich scattering. Eq. (22) clearly exhibits the inverse proportional relationship between M and N. Meanwhile, intuitively, M increases with C_0, K, and $\frac{\sigma^2}{p}$, but decreases with the link strengths $\alpha_k\beta$ and γ_k. Since the RIS's reflecting elements consume much less energy than RF chains, Corollary 3 states that the energy efficiency can be remarkably improved by integrating an RIS.

Corollary 4: If the RIS-BS channel is purely LoS ($\delta \to \infty$), RIS-aided massive MIMO systems perform no worse than conventional massive MIMO systems, i.e., $\underline{R_k}\left(\mathbf{\Phi}\right) \geq \underline{R_k}^{w/o}$.

Proof: Substituting $\delta \to \infty$ into (20), ϵ_k, and κ_k, it can be shown that $\underline{\underline{R_k}} = \underline{R_k}^{w/o}$. Then, we have $\underline{R_k}\left(\mathbf{\Phi}\right) \geq \underline{\underline{R_k}} = \underline{R_k}^{w/o}$. ■

Corollary 4 corresponds to the scenario where the RIS is carefully deployed to reduce the scatters and obstacles between the BS and the RIS. In this case, the additional channel estimation error in $\epsilon_k, \forall k$, caused by the RIS, vanishes. Therefore, the RIS only has the positive effect of enhancing the channel strength, which improves the achievable rate. We emphasize that even though we can only prove that RIS-aided systems are no worse than conventional systems when $\delta \to \infty$, in general, it could perform much better because the second lower bound $\underline{\underline{R_k}}$ is not as tight as the first lower bound $\underline{R_k}\left(\mathbf{\Phi}\right)$ if $\mathbf{\Phi}$ is carefully designed.

D. Power Scaling Law

In conventional massive MIMO systems, an attractive feature is that the transmit power can be scaled down proportionally by increasing M [3]–[5]. After introducing an RIS, we reveal a new power scaling law with respect to N, and compare it to (19).

Corollary 5: As $N \to \infty$, when the power is scaled proportionally to $p = E_u/N$, the achievable rate in (17) can maintain a non-zero value $\underline{\vec{R}_k}\left(\mathbf{\Phi}\right) \to \tau^o \log_2\left(1+\overrightarrow{\mathrm{SNR}_k}\right)$, where

$$\overrightarrow{\mathrm{SNR}_k} = \frac{E_u(M-K)}{\sum_{i=1}^{K}\frac{E_u}{\frac{\tau E_u}{\sigma^2}+\frac{\delta+1}{\alpha_i\beta}}+\sigma^2} \times \frac{1}{\left[\mathbf{\Xi}^{-1}\right]_{kk}} \geq \frac{E_u(M-K)}{\sum_{i=1}^{K}\frac{E_u}{\frac{\tau E_u}{\sigma^2}+\frac{\delta+1}{\alpha_i\beta}}+\sigma^2} \times \frac{\left(\frac{\alpha_k\beta}{\delta+1}\right)^2}{\frac{\alpha_k\beta}{\delta+1}+\frac{\sigma^2}{\tau E_u}}. \quad (23)$$

with $\mathbf{\Xi} = \mathrm{diag}\left\{\frac{(\frac{\alpha_1\beta}{\delta+1})^2}{\frac{\alpha_1\beta}{\delta+1}+\frac{\sigma^2}{\tau E_u}}, \ldots, \frac{(\frac{\alpha_K\beta}{\delta+1})^2}{\frac{\alpha_K\beta}{\delta+1}+\frac{\sigma^2}{\tau E_u}}\right\} + \frac{\beta\delta}{\delta+1}\frac{\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1}{N}$.

Proof: Substitute $p = \frac{E_u}{N}$ into (17). As $N \to \infty$, we have $\kappa_k \to \frac{\frac{\alpha_k\beta}{\delta+1}}{\frac{\alpha_k\beta}{\delta+1}+\frac{\sigma^2}{\tau E_u}}$, $\frac{E_u}{N}\epsilon_i \to \frac{E_u}{\frac{\delta+1}{\alpha_i\beta}+\frac{\tau E_u}{\sigma^2}}$, $\frac{\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon}}{N} \to \mathrm{diag}\left\{\kappa_1^2\alpha_1, \ldots, \kappa_K^2\alpha_K\right\}$, $\frac{\mathbf{\Omega}\mathbf{\Upsilon}^2}{N} \to \mathbf{0}$, and $\frac{\sigma^2}{\tau p}\frac{\mathbf{\Upsilon}^2}{N} \to \frac{\sigma^2}{\tau E_u}\mathbf{\Upsilon}^2$, which help us arrive at the first equation in (23). Then, the lower bound can be obtained by using the inequality in (51). ■

Comparing (23) with (19), it can be seen that this new scaling law has a high order of magnitude with respect to M. Besides, by comparing (23) with (18), it is interesting to find that (23) can be interpreted as the SNR achieved by a conventional massive MIMO system with transmit power E_u and path-loss $\frac{\alpha_k\beta}{\delta+1}$. To sum up, for large M and N, transmit power can be significantly reduced while achieving high data rates.

E. Comparison With MRC-Based Systems

Corollary 6: When p or M or N is large, ZF-based RIS-aided massive MIMO outperforms its MRC-based counterpart. Besides, the severe fairness problem in MRC-based RIS-aided massive MIMO system [24, Remark 2] does not exist in the considered ZF-based systems.

Proof: According to Corollary 2, when p or M grows without bound, it is found that $R_k \geq \underline{\underline{R_k}} \to \infty, \forall k$. Thus, all users can have infinite data rates. However, as proved in [24, Remark 2], when using MRC detectors, due to the mutual interference, the rate is still bounded when p or M is large. Meanwhile, the rates of all users in the considered system are at least on the order of $\mathcal{O}\left(\log_2\left(N\right)\right)$. However, when using MRC, the rate of only one user can be on the order of $\mathcal{O}\left(\log_2\left(N\right)\right)$, while the rates of all other users degrade to zero when N is large, which results in a serious fairness problem. ■

ZF-based RIS systems perform better since RIS-aided systems suffer from severe multi-user interference. This is because multiple users share the common RIS-BS channel, and thus the K users' channels are highly correlated. The highly correlated channels result in severe interference and low data rate. However, by using ZF, the severe multi-user interference

$$\underline{R_k}\left(\mathbf{\Phi}\right) = \tau^o \log_2\left(1 + \frac{p(M-K)}{\left(p\sum_{i=1}^{K}\epsilon_i+\sigma^2\right)\left[\left(\mathbf{\Lambda}+\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1\right)^{-1}\right]_{kk}}\right), \quad (17)$$

issue can be addressed, which leads to promising performance for various aspects.

F. The Upper Bound

The analysis based on the lower bound $\underline{\underline{R_k}}$ is rigorous but conservative, since it ignores the performance gain achieved by optimizing $\mathbf{\Phi}$. We next provide an upper bound to unveil the maximum gain achieved by optimizing $\mathbf{\Phi}$.

Corollary 7: The rate is upper bounded by $\underline{R_k}(\mathbf{\Phi}) \le \overline{R_k} = \tau^o \log_2\left(1+\overline{\mathrm{SNR}}_k\right)$*, where*

$$\overline{\mathrm{SNR}}_k = \frac{p(M-K)}{p\sum_{i=1}^K \epsilon_i + \sigma^2}\left\{\frac{(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k)^2}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}}+\left|\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k\right|^2\frac{\alpha_k\beta\delta}{\delta+1}\right\} \tag{24}$$

$$\le \frac{p(M-K)}{p\sum_{i=1}^K \epsilon_i + \sigma^2}\left\{\frac{(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k)^2}{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}}+N^2\frac{\alpha_k\beta\delta}{\delta+1}\right\}. \tag{25}$$

Based on (24), $\overline{R_k}$ *is at least on the order of* $\mathcal{O}\left(\log_2\left(MN\right)\right)$*. Based on (25),* $\overline{R_k}$ *is on the order of* $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$.

Proof: Please refer to Appendix E. ■

We emphasize that (24) holds for all K users but (25) does not. This is because (25) is achieved by aligning the RIS phase shifts to a specific user k so that $\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k = N$. However, when $\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k = N$, it is known that $\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_i$, $\forall i \ne k$, is bounded even for $N \to \infty$ [24]. Thus, the additional N-fold gain in (25) comes from the concentration of passive beamforming on user k. Combining the lower bound in Corollary 2 and this upper bound, we highlight the following conclusion:

Remark 1: If the RIS phase shifts are aligned to one specific user, the rate of this user will scale ***at most*** *on the order of* $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$*, while the rates of the other users scale* ***at least*** *on the order of* $\mathcal{O}\left(\log_2\left(MN\right)\right)$*, which is high as well.*

Based on these two achievable rate scaling laws, the sum user rate will be high for large M and N, if we simply align the RIS phase shifts for an arbitrary user, which constitutes a low-complexity heuristic approach for the sum-rate maximization problem.

Corollary 8: The quantization error caused by RIS discrete phase shifts does not impact the derived achievable rate scaling orders.

Proof: First, the lower bound $\underline{\underline{R_k}}$ does not depend on $\mathbf{\Phi}$, and hence, is not affected by quantization errors. Secondly, $\left|\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k\right|^2 \ge N^2\cos^2\left(\frac{\pi}{2^b}\right)$ holds for an RIS with b-bit quantization [25]. Therefore, scaling order $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$ still holds for $\overline{R_k}$. ■

G. Summary

We summarize that RIS-aided massive MIMO with ZF detectors is promising for

- **Green communications (Corollary 3)** : The number of BS antennas can be reduced inversely proportional to the number of RIS elements, while maintaining a constant rate.
- **Enhanced mobile broadband (Corollary 2, 7, 8, Remark 1)** : According to the rate scaling orders, ultra-high throughput requirement can be achieved for large M and N.
- **Internet of things (Corollary 5)** : For large M and N, all users can significantly reduce their transmit powers while maintaining high data rates.

V. RIS Phase Shift Design

In this section, based on the derived rate expression in (17) and the low-complexity MM technique [36], we aim to solve the sum user rate maximization (Max-Sum) and the minimum user rate maximization (Max-Min) problems, respectively. The Max-Sum problem maximizes the utility but may sacrifice fairness. On the contrary, the Max-Min problem guarantees fairness but may sacrifice utility. Thus, simultaneously investigating both problems can help us understand which optimization criterion is more suitable for the considered systems. For tractability, variable $\mathbf{\Phi}$ is rewritten as $\mathbf{\Phi} = \mathrm{diag}\left\{\mathbf{v}^H\right\}$, where $\mathbf{v} = \left[e^{j\theta_1},\ldots,e^{j\theta_N}\right]^H$. Then, we can transform the design of $\mathbf{\Phi}$ to the design of vector $\mathbf{v}$.

Lemma 2: The rate in (17) can be rewritten as $\underline{R_k}(\mathbf{v}) = \frac{\tau^o}{\ln(2)}\ln\left(1+\frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{C}_k\mathbf{v}}\right)$*, where*

$$\mathbf{B} = \frac{1}{N}\mathbf{I}_N + \frac{\beta\delta}{\delta+1}\mathrm{diag}\left\{\mathbf{a}_N^H\right\}\mathbf{H}_1\mathbf{\Lambda}^{-1}\mathbf{H}_1^H\,\mathrm{diag}\left\{\mathbf{a}_N\right\},$$

$$\mathbf{C}_k = \frac{p\sum_{i=1}^K\epsilon_i+\sigma^2}{p(M-K)}\left(\left[\mathbf{\Lambda}^{-1}\right]_{kk}\mathbf{B} - \frac{\beta\delta}{\delta+1}\mathbf{z}_k\mathbf{z}_k^H\right), \tag{26}$$

and $\mathbf{z}_k^H = \left[\mathbf{\Lambda}^{-1}\mathbf{H}_1^H\,\mathrm{diag}\left\{\mathbf{a}_N\right\}\right]_{(k,:)}$*. Besides, we have* $\mathbf{B} \succ \mathbf{0}$ *and* $\mathbf{C}_k \succeq \mathbf{0}$.

Proof: We can complete the proof by substituting the last equality in (51) into (17), and using $\mathbf{\Phi}^H\mathbf{a}_N = \mathrm{diag}\left\{\mathbf{a}_N\right\}\mathbf{v}$ and $1 = \frac{1}{N}\mathbf{v}^H\mathbf{I}_N\mathbf{v}$. Besides, we have $\mathbf{B} \succ \mathbf{0}$ due to $\mathbf{\Lambda}^{-1} \succ \mathbf{0}$, which results in $\mathbf{v}^H\mathbf{B}\mathbf{v} > 0$. Since the rate $\underline{R_k}(\mathbf{v})$ must be non-negative due to its definition in (15), we obtain $\mathbf{v}^H\mathbf{C}_k\mathbf{v} \ge 0$, which means that $\mathbf{C}_k \succeq \mathbf{0}$. ■

Define $f_k(\mathbf{v}) \triangleq \ln\left(1+\frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{C}_k\mathbf{v}}\right)$ for brevity. Since the same factor $\frac{\tau^o}{\ln(2)}$ is included in $\underline{R_k}(\mathbf{v}), \forall k$, we can ignore it and formulate the following two optimization problems

$$\textbf{Max-Sum}: \max_{\mathbf{v}} \sum\nolimits_{k=1}^K f_k(\mathbf{v}), \qquad s.t.\ \left|[\mathbf{v}]_{(n)}\right| = 1,\ \forall n. \tag{27}$$

$$\textbf{Max-Min}: \max_{\mathbf{v}} \min_{k} f_k(\mathbf{v}), \qquad s.t.\ \left|[\mathbf{v}]_{(n)}\right| = 1,\ \forall n. \tag{28}$$

To successfully solve the above two problems under the MM algorithm framework, tractable lower-bound surrogate functions need to be constructed for objective functions in (27) and (28), and then closed-form optimal solutions are expected to be derived via the surrogate functions.

A. Max-Sum Problem

Lemma 3: For a fixed point $\mathbf{v}_n$*, a lower bound of* $f_k(\mathbf{v})$ *is given by*

$$f_k(\mathbf{v}) \ge \underline{f_k}(\mathbf{v} \mid \mathbf{v}_n) = \mathrm{const}_k + 2\,\mathrm{Re}\left\{\left(\mathbf{f}_k^n\right)^H\mathbf{v}\right\}, \tag{29}$$

where

$$\begin{aligned}\text{const}_k &= f_k(\mathbf{v}_n) - \frac{\mathbf{v}_n^H \mathbf{B} \mathbf{v}_n}{\mathbf{v}_n^H \mathbf{C}_k \mathbf{v}_n} \\ &\quad - \psi_k \mathbf{v}_n^H \left(\lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right) \mathbf{I}_N - \left(\mathbf{C}_k + \mathbf{B}\right)\right) \mathbf{v}_n \\ &\quad - N \psi_k \lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right), \\ \left(\mathbf{f}_k^n\right)^H &= \omega_k \mathbf{v}_n^H \mathbf{B} - \psi_k \mathbf{v}_n^H \left(\left(\mathbf{C}_k + \mathbf{B}\right) - \lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right) \mathbf{I}_N\right), \\ \omega_k &= \frac{1}{\mathbf{v}_n^H \mathbf{C}_k \mathbf{v}_n}, \\ \psi_k &= \frac{\mathbf{v}_n^H \mathbf{B} \mathbf{v}_n}{\left(\mathbf{v}_n^H \mathbf{C}_k \mathbf{v}_n\right)\left(\mathbf{v}_n^H \mathbf{C}_k \mathbf{v}_n + \mathbf{v}_n^H \mathbf{B} \mathbf{v}_n\right)}. \end{aligned} \tag{30}$$

Proof: Please refer to Appendix F. ■

Then, the Max-Sum problem (27) can be directly solved based on the proposed surrogate function $\underline{f_k}(\mathbf{v} \mid \mathbf{v}_n)$ in Lemma 3. Denoted by $\mathbf{v}_n$ the solution in the n-th iteration, the closed-form optimal solution in the $(n+1)$-th iteration is given by

$$\begin{aligned}\mathbf{v}_{n+1} &= \arg\max_{\mathbf{v}} \sum\nolimits_{k=1}^{K} \underline{f_k}\left(\mathbf{v} \mid \mathbf{v}_n\right) \\ &= \exp\left\{ j \angle \left(\sum\nolimits_{k=1}^{K} \mathbf{f}_k^n \right) \right\}. \end{aligned} \tag{31}$$

B. Max-Min Problem

Next, we focus on the Max-Min problem (28), which is more challenging since the objective function $\min_k f_k(\mathbf{v})$ is non-differentiable. Therefore, we first adopt the log-sum-exp approximation in [37] to obtain a lower-bounded smooth objective function, as follows

$$\begin{aligned}\min_k f_k(\mathbf{v}) &\geq \min_k \underline{f_k}(\mathbf{v} \mid \mathbf{v}_n) \geq \widetilde{f}\left(\mathbf{v}\right) \\ &\triangleq -\frac{1}{\mu} \ln\left(\sum\nolimits_{k=1}^{K} \exp\left\{ -\mu \underline{f_k}(\mathbf{v} \mid \mathbf{v}_n) \right\} \right), \end{aligned} \tag{32}$$

where $\mu > 0$ is a constant for controlling the approximation accuracy, and the last inequality can be proved similar to [37, (15)].

Lemma 4: For a fixed point $\mathbf{v}_n$, $\widetilde{f}\left(\mathbf{v}\right)$ in (32) is lower bounded by

$$\begin{aligned}\widetilde{f}\left(\mathbf{v}\right) \geq \underline{\widetilde{f}}\left(\mathbf{v} \mid \mathbf{v}_n\right) = \widetilde{\text{const}} + 2\,\text{Re}\left\{ \left[\left(\sum\nolimits_{k=1}^{K} l_k^n \left(\mathbf{f}_k^n\right)^H \right) \right.\right. \\ \left.\left. + \left(2\mu \max_k \left\| \mathbf{f}_k^n \right\|^2 \right) \mathbf{v}_n^H \right] \mathbf{v} \right\}, \end{aligned} \tag{33}$$

where

$$\begin{aligned}\widetilde{\text{const}} &= \widetilde{f}\left(\mathbf{v}_n\right) - 2\,\text{Re}\left\{ \sum\nolimits_{k=1}^{K} l_k^n \left(\mathbf{f}_k^n\right)^H \mathbf{v}_n \right\} \\ &\quad + 2N \left(-2\mu \max_k \left\| \mathbf{f}_k^n \right\|^2 \right), \end{aligned} \tag{34}$$

$$l_k^n = \frac{\exp\left\{ -\mu \underline{f_k}(\mathbf{v}_n \mid \mathbf{v}_n) \right\}}{\sum_{k=1}^{K} \exp\left\{ -\mu \underline{f_k}(\mathbf{v}_n \mid \mathbf{v}_n) \right\}}. \tag{35}$$

Proof: Please refer to Appendix G. ■

Based on the MM algorithm, the Max-Min problem (28) can be solved by maximizing the lower bound $\underline{\widetilde{f}}\left(\mathbf{v} \mid \mathbf{v}_n\right)$ in each iteration. Given the solution $\mathbf{v}_n$ in the n-th iteration, the closed-form optimal solution in the $(n+1)$-th iteration is

$$\begin{aligned}\mathbf{v}_{n+1} &= \arg\max_{\mathbf{v}} \underline{\widetilde{f}}\left(\mathbf{v} \mid \mathbf{v}_n\right) \\ &= \exp\left\{ j \angle \left\{ \left(\sum\nolimits_{k=1}^{K} l_k^n \mathbf{f}_k^n \right) \right.\right. \\ &\quad \left.\left. + \left(2\mu \max_k \left\| \mathbf{f}_k^n \right\|^2 \right) \mathbf{v}_n \right\} \right\}. \end{aligned} \tag{36}$$

Finally, the framework for solving Max-Sum problem (27) and Max-Min problem (28) are summarized in Algorithm 1, where steps $4-9$ are used to accelerate the convergence speed of the MM technique [38].

Algorithm 1 MM Algorithm

1: Initialize $\mathbf{v}_0$, $n = 0$;
2: **repeat**
3: Given $\mathbf{v}_n$, obtain solution $\mathbf{v}_{n+1}^{(1)}$ from (31) or (36);
4: Given $\mathbf{v}_{n+1}^{(1)}$, obtain solution $\mathbf{v}_{n+1}^{(2)}$ from (31) or (36);
5: $\triangle_{\mathbf{v}1} = \mathbf{v}_{n+1}^{(1)} - \mathbf{v}_n$, and $\triangle_{\mathbf{v}2} = \mathbf{v}_{n+1}^{(2)} - \mathbf{v}_{n+1}^{(1)} - \triangle_{\mathbf{v}1}$;
6: $\rho = -\frac{\|\triangle_{\mathbf{v}1}\|}{\|\triangle_{\mathbf{v}2}\|}$, and $\mathbf{v}_{n+1} = -\exp\left\{ j \angle \left(\mathbf{v}_n - 2\rho \triangle_{\mathbf{v}1} + \rho^2 \triangle_{\mathbf{v}2} \right) \right\}$;
7: **while** $\mathbf{v}_{n+1}$ does not lead to an increasing objective value in (27) or (28) **do**
8: $\rho = (\rho - 1)/2$, and $\mathbf{v}_{n+1} = -\exp\left\{ j \angle \left(\mathbf{v}_n - 2\rho \triangle_{\mathbf{v}1} + \rho^2 \triangle_{\mathbf{v}2} \right) \right\}$;
9: **end while**
10: $n \leftarrow n + 1$;
11: **until** The objective value in (27) or (28) converges.

C. Convergence and Complexity Analysis

For both the Max-Sum and Max-Min problems, MM algorithms are utilized to optimize variable $\mathbf{v}$ in each iteration. First, the monotonicity of the MM algorithm has been proved in [39]. Second, in Corollary 7, we have proved that there is an upper bound for the objective function of the optimization problem. Therefore, the convergence of the proposed algorithm is guaranteed.

The computational complexity of solving the Max-Sum problem based on (31) is mainly caused by $\mathbf{f}_k^n, \forall k$, in (30). By neglecting the lower-order terms, the approximate computational complexity of $\lambda_{\max}\left(\mathbf{B} + \mathbf{C}_k\right), \forall k$, is $\mathcal{O}\left(KN^3 + K^3 + NK^2\right)$ [40, C.1]. Note that parameters $\mathbf{B}$ and $\mathbf{C}_k$ do not need to be updated in each iteration. Denote the number of iterations in Algorithm 1 as T. Then, given $\mathbf{B}$ and $\mathbf{C}_k$, the approximate computational complexity of the remaining terms in the expression for $\mathbf{f}_k^n, \forall k$, in (30) is $\mathcal{O}\left(KN^2\right)$ which needs to be computed T times. Therefore, the overall approximate complexity of solving the Max-Sum problem with Algorithm 1 is $\mathcal{O}\left(TKN^2 + KN^3 + K^3 + NK^2\right)$.

The computational complexity of solving the Max-Min problem based on (36) mainly comes from parameters $\mathbf{f}_k^n, \forall k$, in (30) and $l_k^n, \forall k$, in (35). The approximate computational complexity of $\mathbf{f}_k^n, \forall k$, is the same as that in solving the Max-Sum problem, i.e., $\mathcal{O}\left(TKN^2 + KN^3 + K^3 + NK^2\right)$.

Besides, the parameters $l_k^n, \forall k$, in (35) has the approximate computational complexity of $\mathcal{O}\left(TKN^2\right)$. Therefore, the overall approximate complexity of solving the Max-Min problem with Algorithm 1 remains on the order of $\mathcal{O}\left(TKN^2 + KN^3 + K^3 + NK^2\right)$.

VI. Numerical Results

In this section, we verify the correctness of our derived results and give insights. Unless otherwise stated, as in [24], we set $M = N = 64$, $\delta = 1$, $\tau_c = 196$, $\tau = K$, and $\mu = 10$. The transmit power of the pilot and data signals for each user is $p = 30$ dBm, and the noise power is $\sigma^2 = -104$ dBm. The BS and the RIS are located at $(0, 0)$ and $(0, 700\text{ m})$, respectively. The number of users is $K = 8$, and the users are randomly located in a circle centred at $(10\text{ m}, 700\text{ m})$ of radius 10 m. Without loss of generality, we denote the users nearest to and furthest from the RIS as users 1 and 8, respectively. The path-loss factors are calculated as $\alpha_k = 10^{-3} d_{\mathrm{uR},k}^{-2}$, $\beta = 10^{-3} d_{\mathrm{RB}}^{-2.5}$, and $\gamma_k = 10^{-3} d_{\mathrm{uB},k}^{-4}$, where $d_{\mathrm{uR},k}$, d_{RB}, and $d_{\mathrm{uB},k}$ denote the distances between user k and the RIS, the RIS and the BS, and user k and the BS, respectively. The AoA and AoD parameters used for the LoS channels are generated randomly from $[0, 2\pi)$. The convergence accuracy for Algorithm 1 is set to 10^{-6}. The theoretical result in (17) is verified via Monte-Carlo simulations based on (14), and the corresponding results are referred to as "Simulation" in the legends of the following figures. The MRC-based system for perfect and imperfect CSI are evaluated based on [41] and [24], respectively.

To begin with, we want to verify the accuracy of the derived rate scaling order $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$ (based on the upper bound) and $\mathcal{O}\left(\log_2\left(MN\right)\right)$ (based on the lower bound) in Remark 1. To this end, some simple RIS design schemes are needed. Recall that the order $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$ obtained in Corollary 7 requires that the phase shifts of the RIS are aligned to only one user. Therefore, the following two types of RIS designs are considered.

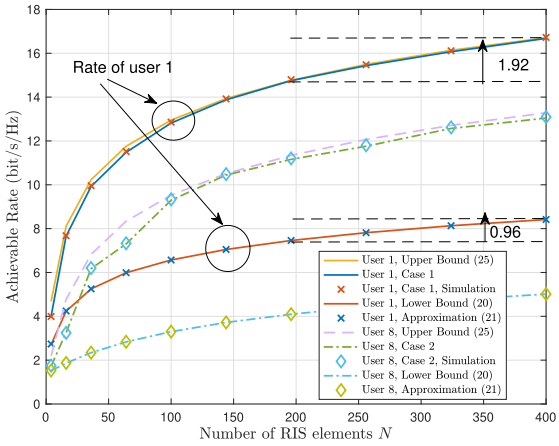
- **Case 1:** The phase shifts are aligned to the nearest user (user 1) so that $\mathbf{a}_N^H \boldsymbol{\Phi} \overline{\mathbf{h}}_1 = N$.
- **Case 2:** The phase shifts are aligned to the furthest user (user 8) so that $\mathbf{a}_N^H \boldsymbol{\Phi} \overline{\mathbf{h}}_8 = N$.

Besides, recall that the order $\mathcal{O}\left(\log_2\left(MN\right)\right)$ drawn in Corollary 2 corresponds to a $\boldsymbol{\Phi}$-independent lower bound, and this bound could be tighter when the RIS is not effectively optimized. Therefore, we also consider the following two RIS design schemes in which $\boldsymbol{\Phi}$ is not optimized.

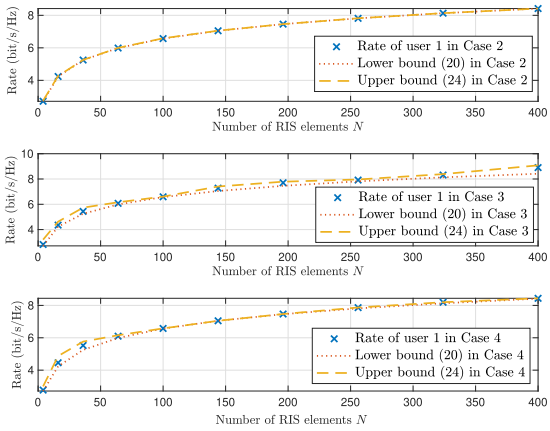
- **Case 3:** The phase shift of each RIS element is set randomly in $[0, 2\pi)$.
- **Case 4:** The phase shift matrix is simply set as an identity matrix, i.e., $\boldsymbol{\Phi} = \mathbf{I}_N$.

Furthermore, these four cases will serve as baseline schemes for the proposed Algorithm 1.

Fig. 2(a) illustrates the derived upper and lower bounds, and shows the rate of one user when the RIS phase shifts are aligned to it. To be specific, we respectively plot the rate of user 1 in Case 1, and the rate of user 8 in Case 2. Firstly, it can be observed that when the RIS phase shifts are aligned to user 1 or user 8, their rates tightly approach the upper bound in (25), which validates that the derived scaling order $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$ in (25) is achievable. Secondly, the theoretical results match well with the Monte-Carlo simulation results, which verifies the correctness of our derivatives. Besides, compared with the rate of user 8 achieved in Case 2, aligning the phase shifts to user 1 in Case 1 yields a higher achievable rate. This is because user 1 is located closer to the RIS and thus has a smaller path-loss. Thirdly, it can be seen that the approximate lower bound (21) perfectly matches with the accurate lower bound (20) for all considered values of N, which verifies the reliability of our previous analysis based on (21). Finally, when N is doubled from $N = 200$ to $N = 400$, the increment of the rate in lower bound and that in Case 1 are almost $\tau^o \log_2(2) = 0.96$ and $\tau^o \log_2\left(2^2\right) = 1.92$, respectively, which confirms the derived theoretical scaling orders of $\mathcal{O}\left(\log_2\left(MN\right)\right)$ and $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$.



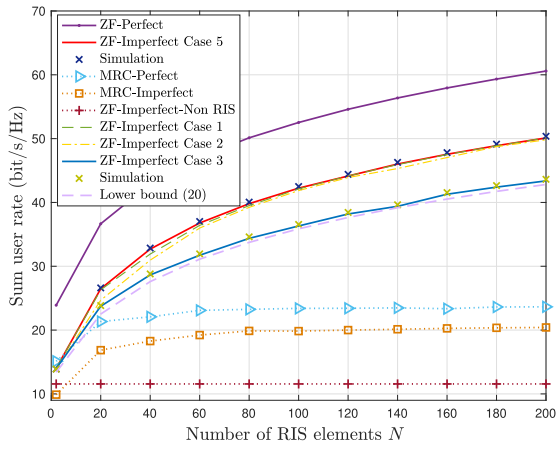
(a) Achievable rate of user 1 or user 8



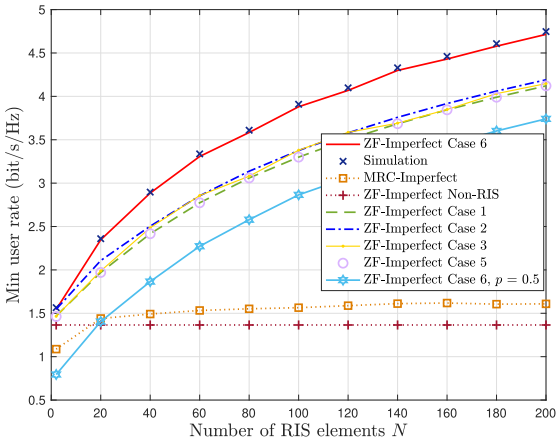
(b) Achievable rate of user 1 in Case 2 - 4

Fig. 2. Achievable rate of a user under different RIS designs.

Fig. 2(b) shows the achievable rate of user 1 when the RIS phase shifts *are not* aligned to it, i.e., in Case 2 - 4. It can be observed that in these three cases, the upper bound (24) and lower bound (20) are tight, which means that the rate scales accurately on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$. This is because the RIS phase shifts cannot be aligned simultaneously to many users. Then, only one user's rate can scale as $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$

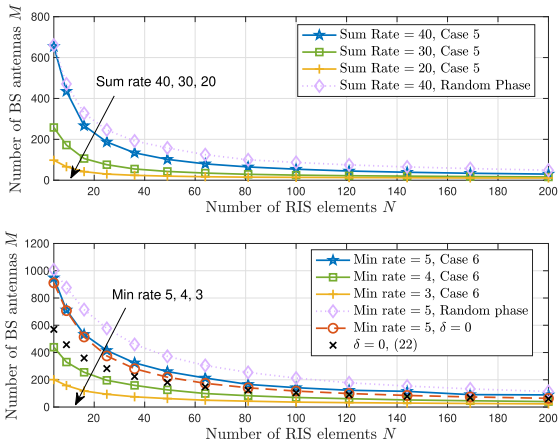


(a) Sum user rate

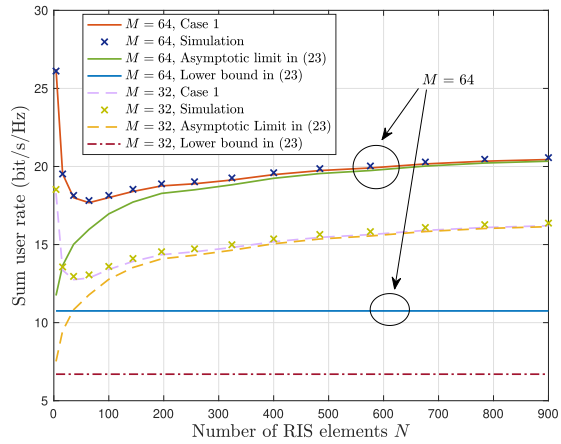


(b) Minimum user rate

Fig. 3. Sum user rate and minimum user rate.



(a) Trade-off between M and N



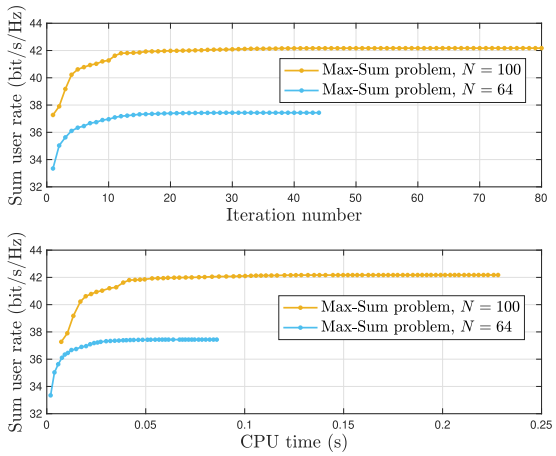
(b) The power scaling law, $p = 10/N$

Fig. 4. Trade-off between M and N and the power scaling law.

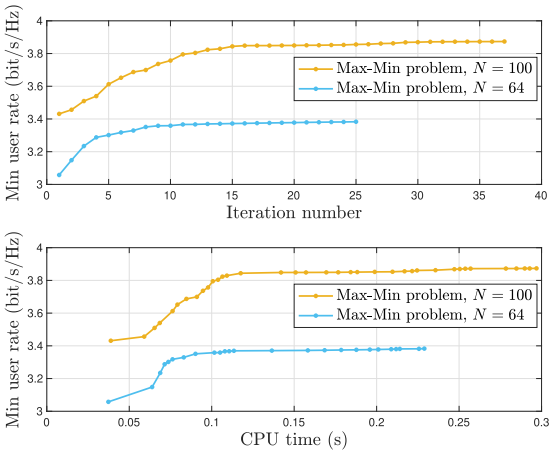
while the rates of all other users scale only as $\mathcal{O}\left(\log_2\left(MN\right)\right)$. Therefore, the scaling order $\mathcal{O}\left(\log_2\left(MN\right)\right)$ obtained based on the lower bound is appropriate for understanding the system capacity since it corresponds to the rate of most of the users.

Next, we respectively examine the sum user rate and the minimum user rate when using the proposed RIS design in Algorithm 1. Fig. 3(a) illustrates the sum user rate. The RIS's phase shifts are designed by solving the Max-Sum problem (27), denoted as **Case 5**. For comparison, the RIS's phase shifts designs based on Case 1 (aligned to user 1), Case 2 (aligned to user 8) and Case 3 (set randomly) are considered as well. Firstly, it can be observed that there exists some performance loss caused by channel estimation errors. This is because the length of the pilots is $\tau = K = 8$, which is very small compared to the large M and N. However, the ZF-based perfect and imperfect CSI cases have a similar growth rate (i.e., a nearly constant gap). This is because the channel estimation error ϵ_k saturates for large N and then does not degrade the scaling order. Secondly, it is seen that ZF-based systems perform much better than MRC-based and RIS-free systems, especially when N is large. This is consistent with our analytical results. Thirdly, the rate in Case 5 is much higher than that in Case 3. However, a near-optimal performance is achieved by Case 1 and Case 2. Especially, in Case 1 where the RIS phase shifts are aligned to the nearest user, the rate is almost the same as the optimal result. This is because by aligning the RIS's phase shifts to a user, the rate of this user scales on the order of $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$, while the rates of all other users scale still on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, which corresponds to a large sum user rate when both M and N are large. Since directly setting $\mathbf{a}_N^H\boldsymbol{\Phi}\overline{\mathbf{h}}_k = N$ is a very simple and low-complexity approach, aligning the RIS's phase shifts to an arbitrary user is a high-quality sub-optimal solution for practical systems. Finally, it can be again observed that the lower bound (20) is tight to the achievable rate realized by a randomly chosen $\boldsymbol{\Phi}$.

Fig. 3(b) evaluates the minimum user rate. We design the RIS phase shifts by solving the Max-Min problem (28), denoted as **Case 6**. The RIS designs based on Case 1 (aligned to user 1), Case 2 (aligned to user 8), Case 3 (set randomly), and Case 5 (Max-Sum) are also considered for comparison. It is seen that our optimal design in Case 6 yields better minimum user rates compared with other cases. However, despite some performance loss, Cases 1, 2, 3, and 5 also achieve relatively high minimum user rates. This is because the dominant limitation, namely the multi-user interference, is eliminated. Thus, even the lowest rate grows still on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, which is guaranteed to be high with large M and N. Meanwhile, it can be seen that the



(a) Max-Sum algorithm



(b) Max-Min algorithm

Fig. 5. The convergence behavior of the proposed MM algorithms, where $M = 64$.

minimum rates in Case 2 are better than that in Case 1, 3, and 5. This is because in Case 2, the RIS's phase shifts are aligned to the furthest user who has the lowest path-loss factor. Intuitively, compared with Case 1 which aligns the RIS phase shifts to the nearest user, Case 2 is more fair and then achieves a better minimum user rate. Besides, it can be observed that the achievable rate is degraded when the transmit power p is reduced from 30 dBm (1 W) to 27 dBm (0.5 W). This is because a lower SNR not only decreases the power of the desired signal received at the BS but also results in larger channel estimation errors which further decrease the received SINR.

Fig. 4(a) evaluates the trade-off between M and N with respect to the sum user rate and the minimum user rate, respectively. As expected, in both cases, M can be reduced inversely proportional to the increase of N while maintaining a constant rate. Meanwhile, after the optimization of $\mathbf{\Phi}$, M can be further decreased compared to the case with random phase shift design. Besides, it can be seen that the reduction of M is more obvious when the rate target is more stringent. This comes from the decreasing slope of the logarithm function. Without the RIS, the rate is on the order of $\mathcal{O}\left(\log_2\left(M\right)\right)$, and very large M is needed to achieve a high rate target. However, if the rate is on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, the high data rate target can be met with moderate M but large N, since the product MN is very large. Besides, when $\delta = 0$ (in Corollary 3), Fig. 4(a) verifies the derived theoretical relationship in (22) by using the path-loss of user 8. As can be observed, the derived results are accurate when $N > 40$.

Fig. 4(b) validates the derived power scaling law in (23), where the power is scaled proportionally to $p = 10/N$. As $N \to \infty$, it is verified that the rate tends to the derived asymptotic limit, and it is larger than the lower bound. Also, it can be observed that the asymptotic limit is improved significantly when M is doubled from 32 to 64. This is because the asymptotic limit in (23) is on the order of $\mathcal{O}\left(\log_2\left(M\right)\right)$.

Finally, Fig. 5 illustrates the convergence behavior of the proposed MM algorithms for solving Max-Sum problem (27) and Max-Min problem (28), respectively. Both the iteration number and the needed CPU time are shown for different numbers of RIS reflecting elements. As can be seen, our algorithms have a fast speed of convergence (within 0.3 s) even though the values of M and N are large, and only a few iterations are sufficient to achieve a large portion of the achievable rate after full convergence. This is because closed-form solutions (31) and (36) are exploited in each iteration of the proposed algorithms. Therefore, the computation has low complexity, and the convergence time is very short. Besides, it can be observed that the iteration number and CPU time increase with N due to the increased number of optimized variables.

VII. Conclusion

This work demonstrates that RIS-aided MIMO with ZF detectors is a promising system architecture for many applications. We derive theoretical expressions for the ergodic rate, based on which two low-complexity MM algorithms are proposed to respectively optimize the sum user rate and the minimum user rate. We demonstrate that by aligning the RIS phase shifts to a user, the rate scaling order of that user can approach $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$, while the rate scaling order of the other users is guaranteed to be $\mathcal{O}\left(\log_2\left(MN\right)\right)$. Therefore, high system capacity can be realized with a low-complexity RIS design. We also prove that by increasing N, the M required to maintain a constant achievable rate can be reduced inverse proportionally. Besides, we prove that as $N \to \infty$, the transmit power of all users can be scaled proportionally to $p = 1/N$ while maintaining high rates.

Appendix A

Based on the definitions of $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{D}$, the channel of user k can be expanded as

$$\mathbf{q}_k = \sqrt{\frac{\alpha_k \beta \delta}{\delta+1}}\overline{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k + \sqrt{\frac{\alpha_k \beta}{\delta+1}}\widetilde{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k + \sqrt{\gamma_k}\tilde{\mathbf{d}}_k. \quad (37)$$

Firstly, since $\widetilde{\mathbf{H}}_2$ consists of i.i.d. $\mathcal{CN}\left(0,1\right)$ elements, vector $\sqrt{\frac{\alpha_k \beta}{\delta+1}}\widetilde{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k$ is comprised of mutually independent elements. Secondly, the elements of vector $\sqrt{\frac{\alpha_k \beta}{\delta+1}}\widetilde{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k$ are linear combinations of independent Gaussian random variables. Therefore, vector $\sqrt{\frac{\alpha_k \beta}{\delta+1}}\widetilde{\mathbf{H}}_2 \mathbf{\Phi} \overline{\mathbf{h}}_k$ consists of i.i.d. Gaussian

variables, following $\sqrt{\frac{\alpha_k\beta}{\delta+1}}\widetilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k \sim \mathcal{CN}\left(\mathbf{0}, N\frac{\alpha_k\beta}{\delta+1}\mathbf{I}_M\right)$. Meanwhile, we have $\sqrt{\gamma_k}\tilde{\mathbf{d}}_k \sim \mathcal{CN}(\mathbf{0}, \gamma_k\mathbf{I}_M)$. Since the sum of independent Gaussian vectors is still a Gaussian vector [42], we have

$$\sqrt{\frac{\alpha_k\beta}{\delta+1}}\widetilde{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k+\sqrt{\gamma_k}\tilde{\mathbf{d}}_k \sim \mathcal{CN}\left(\mathbf{0}, \left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k\right)\mathbf{I}_M\right). \tag{38}$$

Combining (37) with (38), it is proved that $\mathbf{q}_k$ is a Gaussian distributed vector, where $\mathbb{E}\{\mathbf{q}_k\} = \sqrt{\frac{\alpha_k\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\overline{\mathbf{h}}_k$, and $\mathrm{Cov}\{\mathbf{q}_k\} = \mathbb{E}\{(\mathbf{q}_k - \mathbb{E}\{\mathbf{q}_k\})(\mathbf{q}_k - \mathbb{E}\{\mathbf{q}_k\})^H\} = \left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k\right)\mathbf{I}_M$.

Following a similar procedure, the distribution of the noise matrix $\frac{1}{\sqrt{\tau p}}\mathbf{N}\mathbf{s}_k$ can be derived, which is omitted here for brevity.

APPENDIX B

Since the channel $\mathbf{q}_k$ and the noise $\mathbf{N}\mathbf{s}_k$ are Gaussian distributed random variables, the considered observation vector $\mathbf{y}_p^k$ in (7) is consistent with the complex Bayesian linear model [43, Eq. (15.63)] [3, Lemma B.17]. Therefore, we can directly apply the results in [43] to obtain the MMSE channel estimate of $\mathbf{q}_k$ and the MSE matrix. In particular, applying [43, Eq. (15.64)] and using the distribution in Lemma 1, we can obtain (39), as shown at the bottom of the next page.

By substituting the observation vector $\mathbf{y}_p^k$ in (39) with its expression in (7), we complete the calculation of the MMSE channel estimate in (8). Next, applying [43, Eq. (15.67)], the MSE matrix is obtained as

$$\mathbf{MSE}_k = \left[\left(\left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k\right)\mathbf{I}_M\right)^{-1} + \frac{\tau p}{\sigma^2}\mathbf{I}_M\right]^{-1}. \tag{40}$$

After some straightforward simplifications, we can arrive at (9). Besides, since the channel $\mathbf{q}_k$ is a Gaussian vector, the channel estimate $\hat{\mathbf{q}}_k$ and the estimation error $\mathbf{e}_k$ are independent of each other, due to the orthogonality principle of the MMSE estimator [43].

APPENDIX C

To derive the lower bound in (16), term $\mathbb{E}\{[(\hat{\mathbf{Q}}^H\hat{\mathbf{Q}})^{-1}]_{kk}\}$ needs to be tackled, where $\hat{\mathbf{Q}}$ is given in (10). We begin by proving that channel $\hat{\mathbf{Q}}$ is Gaussian distributed.

Lemma 5: [44] A random matrix $\mathbf{X}$ is complex Gaussian distributed as $\mathbf{X} \sim \mathcal{CN}(\mathbf{E}, \mathbf{\Sigma}\otimes\mathbf{\Psi})$, if $\mathrm{vec}(\mathbf{X}^H) \sim \mathcal{CN}(\mathrm{vec}(\mathbf{E}^H), \mathbf{\Sigma}\otimes\mathbf{\Psi})$. If $\mathbf{X}_1 \sim \mathcal{CN}(\mathbf{E}_1, \mathbf{\Sigma}_1\otimes\mathbf{\Psi}_1)$ and $\mathbf{X}_2 \sim \mathcal{CN}(\mathbf{E}_2, \mathbf{\Sigma}_2\otimes\mathbf{\Psi}_2)$ are independent distributed, then $\mathbf{X}_1+\mathbf{X}_2 \sim \mathcal{CN}(\mathbf{E}_1+\mathbf{E}_2, \mathbf{\Sigma}_1\otimes\mathbf{\Psi}_1+\mathbf{\Sigma}_2\otimes\mathbf{\Psi}_2)$.

For notional brevity, we divide the estimated channel $\hat{\mathbf{Q}}$ into three independent parts $\hat{\mathbf{Q}} = \hat{\mathbf{Q}}_{RIS} + \hat{\mathbf{Q}}_{BS} + \hat{\mathbf{Q}}_{noise}$, where

$$\begin{aligned}
\hat{\mathbf{Q}}_{RIS}^H &= \sqrt{\frac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H + \sqrt{\frac{\beta}{\delta+1}}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{\Phi}^H\widetilde{\mathbf{H}}_2^H,\\
\hat{\mathbf{Q}}_{BS}^H &= \mathbf{\Upsilon}\mathbf{\Omega}^{1/2}\widetilde{\mathbf{D}}^H,\\
\hat{\mathbf{Q}}_{noise}^H &= \frac{1}{\sqrt{\tau p}}\mathbf{\Upsilon}\mathbf{S}^H\mathbf{N}^H.
\end{aligned} \tag{41}$$

Recall that $\widetilde{\mathbf{H}}_2$, $\widetilde{\mathbf{D}}$, and $\mathbf{N}$ are composed of i.i.d. Gaussian random variables. By observing (41), we can find that each column of matrices $\hat{\mathbf{Q}}_{RIS}^H$, $\hat{\mathbf{Q}}_{BS}^H$ and $\hat{\mathbf{Q}}_{noise}^H$ can be written as a linear transformation of mutually independent standard Gaussian random vectors. Therefore, the columns of $\hat{\mathbf{Q}}_{RIS}^H$, $\hat{\mathbf{Q}}_{BS}^H$, and $\hat{\mathbf{Q}}_{noise}^H$ are independent Gaussian vectors. As a result, after vectorization, the vectors $\mathrm{vec}(\hat{\mathbf{Q}}_{RIS}^H)$, $\mathrm{vec}(\hat{\mathbf{Q}}_{BS}^H)$, and $\mathrm{vec}(\hat{\mathbf{Q}}_{noise}^H)$ are still Gaussian distributed.

Next, we derive their mean vector and covariance matrices. First, consider the term $\mathrm{vec}(\hat{\mathbf{Q}}_{RIS}^H)$. Obviously, the mean vector is $\mathbb{E}\{\mathrm{vec}(\hat{\mathbf{Q}}_{RIS}^H)\} = \mathrm{vec}(\sqrt{\frac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H)$, and the covariance matrix is given by (42), as shown at the bottom of the next page, where (c) utilizes $\mathrm{vec}(\boldsymbol{ABC}) = \left(\boldsymbol{C}^T\otimes\boldsymbol{A}\right)\mathrm{vec}(\boldsymbol{B})$ and $(\boldsymbol{A}\otimes\boldsymbol{B})^H = \boldsymbol{A}^H\otimes\boldsymbol{B}^H$. (d) exploits $(\boldsymbol{A}\otimes\boldsymbol{C})(\boldsymbol{B}\otimes\boldsymbol{D}) = (\boldsymbol{AB})\otimes(\boldsymbol{CD})$ and $\mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}_N$. According to (42) and Lemma 5, the distribution of $\hat{\mathbf{Q}}_{RIS}$ is given by

$$\hat{\mathbf{Q}}_{RIS} \sim \mathcal{CN}\left(\sqrt{\frac{\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1, \mathbf{I}_M\otimes\frac{\beta}{\delta+1}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon}\right). \tag{43}$$

Similarly, the distribution of $\hat{\mathbf{Q}}_{BS}$ and $\hat{\mathbf{Q}}_{noise}$ can be calculated as follows

$$\hat{\mathbf{Q}}_{BS} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_M\otimes\mathbf{\Omega}\mathbf{\Upsilon}^2\right), \tag{44}$$

$$\hat{\mathbf{Q}}_{noise} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_M\otimes\frac{\sigma^2}{\tau p}\mathbf{\Upsilon}^2\right). \tag{45}$$

Then, using Lemma 5 and the property that $\boldsymbol{A}\otimes\boldsymbol{B}+\boldsymbol{A}\otimes\boldsymbol{C} = \boldsymbol{A}\otimes(\boldsymbol{B}+\boldsymbol{C})$, the estimated channel $\hat{\mathbf{Q}}$ is Gaussian distributed as follows

$$\begin{aligned}
\hat{\mathbf{Q}} \sim \mathcal{CN}\Bigg(&\sqrt{\frac{\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1,\\
&\mathbf{I}_M\otimes\left(\frac{\beta}{\delta+1}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon}+\mathbf{\Omega}\mathbf{\Upsilon}^2+\frac{\sigma^2}{\tau p}\mathbf{\Upsilon}^2\right)\Bigg).
\end{aligned} \tag{46}$$

Lemma 6: [45, Definition 5.1] Let $\mathbf{W} = \mathbf{X}^H\mathbf{X}$, with $n\times m$ matrix $\mathbf{X} \sim \mathcal{CN}(\mathbf{E}, \mathbf{I}_n\otimes\mathbf{\Psi})$. Then, $\mathbf{W}$ follows a complex non-central Wishart distribution with n degrees of freedom, covariance matrix $\mathbf{\Psi}$, and non-centrality parameter $\mathbf{\Sigma} = \mathbf{\Psi}^{-1}\mathbf{E}^H\mathbf{E}$, denoted by $\mathbf{W} \sim \mathcal{CW}_m(n, \mathbf{\Psi}, \mathbf{\Sigma})$. Besides, its mean is $\mathbb{E}(\mathbf{W}) = n\mathbf{\Psi}+\mathbf{\Psi}\mathbf{\Sigma}$ [42, 10.3]. In particular, if $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_n\otimes\mathbf{\Psi})$ has zero mean, $\mathbf{W}$ is complex central Wishart distributed, denoted by $\mathbf{W} \sim \mathcal{CW}_m(n, \mathbf{\Psi})$, where $\mathbb{E}(\mathbf{W}) = n\mathbf{\Psi}$ and $\mathbb{E}(\mathbf{W}^{-1}) = \frac{1}{n-m}\mathbf{\Psi}^{-1}$, $n > m$ [46].

Since $\hat{\mathbf{Q}}$ is Gaussian distributed, from Lemma 6, the product $\hat{\mathbf{Q}}^H\hat{\mathbf{Q}}$ follows a complex non-central Wishart distribution denoted by

$$\hat{\mathbf{Q}}^H\hat{\mathbf{Q}} \sim \mathcal{CW}_K(M, \mathbf{\Psi}_{RIS}, \mathbf{\Sigma}_{RIS}), \tag{47}$$

where $\mathbf{\Psi}_{RIS} = \frac{\beta}{\delta+1}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon}+\mathbf{\Omega}\mathbf{\Upsilon}^2+\frac{\sigma^2}{\tau p}\mathbf{\Upsilon}^2$ and $\mathbf{\Sigma}_{RIS} = (\mathbf{\Psi}_{RIS})^{-1}\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1$. The statistical property of (47) is very complex which encumbers further analysis and optimization. Fortunately, it has been proved that the non-central Wishart distribution can be closely approximated

by a central Wishart distribution [47]. Therefore, as in [5], [48], [49], we approximate the non-central Wishart distribution (47) by a central one with the same first order moment. With Lemma 6 and (6), the mean of (47) is given by

$$\begin{aligned}\mathbb{E}\left\{\hat{\mathbf{Q}}^H\hat{\mathbf{Q}}\right\} &= M\left(\frac{\beta}{\delta+1}\boldsymbol{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\boldsymbol{\Upsilon}+\boldsymbol{\Omega}\boldsymbol{\Upsilon}^2+\frac{\sigma^2}{\tau p}\boldsymbol{\Upsilon}^2\right)\\ &\quad+\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\boldsymbol{\Phi}\mathbf{H}_1\\ &= M\left(\frac{\beta}{\delta+1}\boldsymbol{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\boldsymbol{\Upsilon}+\boldsymbol{\Omega}\boldsymbol{\Upsilon}^2+\frac{\sigma^2}{\tau p}\boldsymbol{\Upsilon}^2\right)\\ &\quad+M\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1. \end{aligned}\tag{48}$$

Then, the central Wishart distribution with the same mean is given by (49), as shown at the bottom of the next page.

By using the property of complex central Wishart distribution in Lemma 6, we obtain (50), as shown at the bottom of the next page. The proof is completed by substituting (50) into (16).

APPENDIX D

Recall that $\boldsymbol{\Lambda}=\frac{\beta}{\delta+1}\boldsymbol{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\boldsymbol{\Upsilon}+\boldsymbol{\Omega}\boldsymbol{\Upsilon}^2+\frac{\sigma^2}{\tau p}\boldsymbol{\Upsilon}^2$. It is readily found that $\boldsymbol{\Lambda}=\boldsymbol{\Lambda}^H$. Note that we assume the existence of direct links, therefore we have $\boldsymbol{\Omega}\succ\mathbf{0}$. Meanwhile, we have $\frac{\beta}{\delta+1}\boldsymbol{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\boldsymbol{\Upsilon}\succeq\mathbf{0}$, $\boldsymbol{\Omega}\boldsymbol{\Upsilon}^2\succ\mathbf{0}$, and $\frac{\sigma^2}{\tau p}\boldsymbol{\Upsilon}^2\succ\mathbf{0}$. Therefore, we obtain that $\boldsymbol{\Lambda}\succ\mathbf{0}$, $\boldsymbol{\Lambda}^{-1}\succ\mathbf{0}$, and $\left(\boldsymbol{\Lambda}^{-1}\right)^H=\boldsymbol{\Lambda}^{-1}$. Then, applying the Woodbury's identity and using the fact that $\boldsymbol{\Lambda}^{-1}$ is positive definite and Hermitian, we have

$$\begin{aligned}&\left[\left(\boldsymbol{\Lambda}+\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\right)^{-1}\right]_{kk}\\ &=\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}-\frac{\frac{\beta\delta}{\delta+1}\left[\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\right]_{kk}}{1+\frac{\beta\delta}{\delta+1}\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N}\\ &=\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}-\frac{\frac{\beta\delta}{\delta+1}\left|\left[\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\right]_{(k,1)}\right|^2}{1+\frac{\beta\delta}{\delta+1}\left(\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\right)\boldsymbol{\Lambda}^{-1}\left(\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\right)^H}\\ &\le\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}. \end{aligned}\tag{51}$$

Substituting (51) into (17), the lower bound in (20) can be obtained.

Lemma 7: [5], [7], [24] When $N\to\infty$, the product of the LoS components $\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_i$ is still bounded, unless user i has the same AoA as user k.

We can respectively calculate the diagonal and non-diagonal elements of $\boldsymbol{\Lambda}$ as follows

$$[\boldsymbol{\Lambda}]_{(k,k)}=\left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}\right)\kappa_k^2,\tag{52}$$

$$[\boldsymbol{\Lambda}]_{(k,i)}=\frac{\beta}{\delta+1}\sqrt{\alpha_k\alpha_i}\kappa_k\kappa_i\overline{\mathbf{h}}_k^H\overline{\mathbf{h}}_i,\quad\forall i\neq k.\tag{53}$$

When N is small, due to the small product-distance path loss $\alpha_k\beta$ and $\sqrt{\alpha_i\alpha_k}\beta$ compared with γ_k, (53) is much smaller compared with (52). Therefore, $\boldsymbol{\Lambda}$ can be approximated as a diagonal matrix for small N. When N increases, based on Lemma 7, (52) grows much faster than (53). Thus, (52) is still much larger than (53) and we can approximate that $\boldsymbol{\Lambda}$ is dominated by diagonal elements. Finally, when $N\to\infty$, (52) tends to infinity but (53) does not. Therefore, $\boldsymbol{\Lambda}$ tends to a diagonal matrix for large N. Accordingly, for any N, $\boldsymbol{\Lambda}$ can be approximated as a diagonal matrix $\mathrm{diag}\{[\boldsymbol{\Lambda}]_{(1,1)},\ldots,[\boldsymbol{\Lambda}]_{(K,K)}\}$ and then the approximate lower bound in (21) can be obtained by using $\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}\approx\left([\boldsymbol{\Lambda}]_{kk}\right)^{-1}=\frac{N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}}{\left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k\right)^2}$. Finally, by observing the order of magnitude of the numerator and denominator of the SNR in (21), we can find that the numerator is on the order of $\mathcal{O}\left(MN^2\right)$, but the denominator is only on the order of $\mathcal{O}\left(N\right)$. Therefore, the rate is on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$. Besides, it can be readily found that $\underline{R_k}\left(\boldsymbol{\Phi}\right)=\underline{\underline{R_k}}$ when $\delta=0$. Meanwhile, for an optimal solution $\boldsymbol{\Phi}^{**}$ and a sub-optimal solution $\boldsymbol{\Phi}^{*}$, we have $\underline{R_k}\left(\boldsymbol{\Phi}^{**}\right)>\underline{R_k}\left(\boldsymbol{\Phi}^{*}\right)$. Since $\underline{\underline{R_k}}$ is independent of $\boldsymbol{\Phi}$, we have $\underline{R_k}\left(\boldsymbol{\Phi}^{**}\right)-\underline{\underline{R_k}}>\underline{R_k}\left(\boldsymbol{\Phi}^{*}\right)-\underline{\underline{R_k}}$, which indicates that the gap between $\underline{R_k}\left(\boldsymbol{\Phi}\right)$ and $\underline{\underline{R_k}}$ will be enlarged if $\boldsymbol{\Phi}$ is optimized. In other words, the proposed bound $\underline{\underline{R_k}}$ will be tight when we use unoptimized phase shifts.

APPENDIX E

Lemma 8: If $\mathbf{X}\succ\mathbf{0}$, $\left[\mathbf{X}^{-1}\right]_{kk}\ge\frac{1}{[\mathbf{X}]_{kk}}$. The equality holds only if $\mathbf{X}$ is diagonal [50].

Recall that we have $\boldsymbol{\Lambda}\succ\mathbf{0}$ and $\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\succeq\mathbf{0}$. Using Lemma 8 and (1), we have

$$\begin{aligned}&\left[\left(\boldsymbol{\Lambda}+\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\right)^{-1}\right]_{kk}\\ &\ge\frac{1}{\left[\boldsymbol{\Lambda}+\frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\right]_{kk}}\end{aligned}$$

$$\hat{\mathbf{q}}_k=\sqrt{\frac{\alpha_k\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_k+\left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k\right)\mathbf{I}_M\left(\left(N\frac{\alpha_k\beta}{\delta+1}+\gamma_k+\frac{\sigma^2}{\tau p}\right)\mathbf{I}_M\right)^{-1}\left(\mathbf{y}_p^k-\sqrt{\frac{\alpha_k\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\boldsymbol{\Phi}\overline{\mathbf{h}}_k\right).\tag{39}$$

$$\begin{aligned}&\mathrm{Cov}\left\{\mathrm{vec}\left(\hat{\mathbf{Q}}_{RIS}^H\right)\right\}=\mathbb{E}\left\{\mathrm{vec}\left(\sqrt{\tfrac{\beta}{\delta+1}}\boldsymbol{\Upsilon}\mathbf{H}_1^H\boldsymbol{\Phi}^H\widetilde{\mathbf{H}}_2^H\mathbf{I}_M\right)\mathrm{vec}\left(\sqrt{\tfrac{\beta}{\delta+1}}\boldsymbol{\Upsilon}\mathbf{H}_1^H\boldsymbol{\Phi}^H\widetilde{\mathbf{H}}_2^H\mathbf{I}_M\right)^H\right\}\\ &\overset{(c)}{=}\left(\mathbf{I}_M\otimes\sqrt{\tfrac{\beta}{\delta+1}}\boldsymbol{\Upsilon}\mathbf{H}_1^H\boldsymbol{\Phi}^H\right)\mathbb{E}\left\{\mathrm{vec}\left(\widetilde{\mathbf{H}}_2^H\right)\mathrm{vec}\left(\widetilde{\mathbf{H}}_2^H\right)^H\right\}\left(\mathbf{I}_M\otimes\sqrt{\tfrac{\beta}{\delta+1}}\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Upsilon}\right)\\ &=\left(\mathbf{I}_M\otimes\sqrt{\tfrac{\beta}{\delta+1}}\boldsymbol{\Upsilon}\mathbf{H}_1^H\boldsymbol{\Phi}^H\right)\left(\mathbf{I}_M\otimes\sqrt{\tfrac{\beta}{\delta+1}}\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Upsilon}\right)\overset{(d)}{=}\mathbf{I}_M\otimes\tfrac{\beta}{\delta+1}\boldsymbol{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\boldsymbol{\Upsilon},\end{aligned}\tag{42}$$

$$= \frac{1}{[\mathbf{\Lambda}]_{kk} + \frac{\alpha_k\beta\delta}{\delta+1}\left|\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k\right|^2} \geq \frac{1}{[\mathbf{\Lambda}]_{kk} + \frac{\alpha_k\beta\delta}{\delta+1}N^2} = \frac{1}{\left(N\frac{\alpha_k\beta}{\delta+1} + \gamma_k + \frac{\sigma^2}{\tau p}\right)\kappa_k^2 + \frac{\alpha_k\beta\delta}{\delta+1}N^2}, \tag{54}$$

where the last inequality holds by using the property that $\left|\mathbf{a}_N^H\mathbf{\Phi}\overline{\mathbf{h}}_k\right| \leq N$ from triangle inequality [24, (189)], and the equality holds when $\theta_n = -\angle\left\{\left[\mathbf{a}_N^H\right]_n\left[\overline{\mathbf{h}}_k\right]_n\right\}, \forall n$.

The proof is completed by substituting (54) into (17) with a few additional simplifications.

Appendix F

To begin with, we give a brief introduction to the optimization under the MM framework [12], [36]. To maximize a function $g(\mathbf{v})$ based on the MM algorithm, at a point $\mathbf{v}_n$, we need to construct a lower bound $\underline{g}(\mathbf{v}|\mathbf{v}_n)$ satisfying

$$g(\mathbf{v}_n) = \underline{g}\left(\mathbf{v}_n \mid \mathbf{v}_n\right), \tag{55}$$

$$g\left(\mathbf{v}\right) \geq \underline{g}\left(\mathbf{v} \mid \mathbf{v}_n\right), \tag{56}$$

$$\nabla_{\mathbf{v}} g\left(\mathbf{v}\right)\big|_{\mathbf{v}=\mathbf{v}_n} = \nabla_{\mathbf{v}}\underline{g}\left(\mathbf{v} \mid \mathbf{v}_n\right)\big|_{\mathbf{v}=\mathbf{v}_n}. \tag{57}$$

Then, we are able to increase the value of the original function from $g(\mathbf{v}_n)$ to $g(\mathbf{v}_{n+1})$ by finding the point $\mathbf{v}_{n+1}$ which maximizes the lower bound $\underline{g}(\mathbf{v}|\mathbf{v}_n)$. Therefore, the success of using the MM algorithm highly relies on the property of the constructed lower bound.

In the following, we derive a tractable lower bound for $f_k(\mathbf{v})$ which satisfies the above three conditions and can successfully produce a closed-form solution. We first rewrite $f_k(\mathbf{v})$ as

$$\begin{aligned} f_k(\mathbf{v}) &= \ln\left(1 + \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{C}_k\mathbf{v}}\right) = -\ln\left(\frac{\mathbf{v}^H\mathbf{C}_k\mathbf{v}}{\mathbf{v}^H\mathbf{C}_k\mathbf{v} + \mathbf{v}^H\mathbf{B}\mathbf{v}}\right) \\ &= -\ln\left(1 - \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{C}_k\mathbf{v} + \mathbf{v}^H\mathbf{B}\mathbf{v}}\right) \\ &= -\ln\left(1 - \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{t_k}\right) \triangleq f_k(\mathbf{v}, t_k), \end{aligned} \tag{58}$$

where $t_k = \mathbf{v}^H(\mathbf{C}_k + \mathbf{B})\mathbf{v} > 0$. Then, according to [36, (14)] and the composition rule [40, (3.10)], $f_k(\mathbf{v}, t_k)$ is jointly convex in $\mathbf{v}$ and t_k. Therefore, given a point $(\mathbf{v}_n, t_k^n)$, we can obtain a lower bound of $f_k(\mathbf{v}, t_k)$ by using its first-order Taylor expansion, which automatically meets the three conditions needed for MM algorithms. Specifically, we have

$$\begin{aligned} &f_k(\mathbf{v}, t_k) \\ &\geq f_k\left(\mathbf{v}_n, t_k^n\right) + \left.\frac{\partial f_k(\mathbf{v})}{\partial \mathbf{v}^T}\right|_{\mathbf{v}=\mathbf{v}_n}(\mathbf{v} - \mathbf{v}_n) \\ &\quad + \left.\frac{\partial f_k(\mathbf{v})}{\partial \mathbf{v}^H}\right|_{\mathbf{v}^*=\mathbf{v}_n^*}(\mathbf{v}^* - \mathbf{v}_n^*) + \left.\frac{\partial f_k(\mathbf{v})}{\partial t_k}\right|_{t_k=t_k^n}(t_k - t_k^n), \end{aligned} \tag{59}$$

where $\frac{\partial f_k(\mathbf{v},t_k)}{\partial \mathbf{v}^T} = \frac{\mathbf{v}^H\mathbf{B}}{t_k - \mathbf{v}^H\mathbf{B}\mathbf{v}}$, $\frac{\partial f_k(\mathbf{v},t_k)}{\partial \mathbf{v}^H} = \frac{\mathbf{v}^T\mathbf{B}^T}{t_k - \mathbf{v}^H\mathbf{B}\mathbf{v}}$, and $\frac{\partial f_k(\mathbf{v},t_k)}{\partial t_k} = -\frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{(t_k - \mathbf{v}^H\mathbf{B}\mathbf{v})}\frac{1}{t_k}$.

Substituting these three partial derivatives into (59) and using $t_k = \mathbf{v}^H(\mathbf{C}_k + \mathbf{B})\mathbf{v}$ and $t_k^n = \mathbf{v}_n^H(\mathbf{C}_k + \mathbf{B})\mathbf{v}_n$, after some simplifications, we can obtain

$$f_k(\mathbf{v}) \geq \mathrm{const1}_k + 2\,\mathrm{Re}\left\{\omega_k\mathbf{v}_n^H\mathbf{B}\mathbf{v}\right\} - \psi_k\mathbf{v}^H\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{v}, \tag{60}$$

where $\mathrm{const1}_k = f_k\left(\mathbf{v}_n\right) - \frac{\mathbf{v}_n^H\mathbf{B}\mathbf{v}_n}{\mathbf{v}_n^H\mathbf{C}_k\mathbf{v}_n}$, and ω_k and ψ_k are defined in (30). Next, according to the inequality in [36, (26)] and the property that $\mathbf{C}_k + \mathbf{B} \preceq \lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{I}_N$, we have

$$\begin{aligned} &\mathbf{v}^H\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{v} \\ &\quad\leq \mathbf{v}^H\lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{I}_N\mathbf{v} \\ &\qquad + 2\,\mathrm{Re}\left\{\mathbf{v}^H\left(\left(\mathbf{C}_k + \mathbf{B}\right) - \lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{I}_N\right)\mathbf{v}_n\right\} \\ &\qquad + \mathbf{v}_n^H\left(\lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{I}_N - \left(\mathbf{C}_k + \mathbf{B}\right)\right)\mathbf{v}_n. \end{aligned} \tag{61}$$

Substituting (61) into (60) and using the fact that $\mathbf{v}^H\lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right)\mathbf{I}_N\mathbf{v} = N\lambda_{\max}\left(\mathbf{C}_k + \mathbf{B}\right)$, we can arrive at (29).

Appendix G

Under the MM algorithm framework, given a point $\mathbf{v}_n$, we want to construct a quadratic form lower bound $\underline{\widetilde{f}}\left(\mathbf{v} \mid \mathbf{v}_n\right)$ of $\widetilde{f}\left(\mathbf{v}\right)$ as follows

$$\begin{aligned} \widetilde{f}(\mathbf{v}) \geq \underline{\widetilde{f}}\left(\mathbf{v} \mid \mathbf{v}_n\right) &= \widetilde{f}\left(\mathbf{v}_n\right) + 2\,\mathrm{Re}\left\{\mathbf{u}^H\left(\mathbf{v} - \mathbf{v}_n\right)\right\} \\ &\quad + \left(\mathbf{v} - \mathbf{v}_n\right)^H\mathbf{M}\left(\mathbf{v} - \mathbf{v}_n\right), \end{aligned} \tag{62}$$

where $\mathbf{u}$ and $\mathbf{M}$ are two parameters to be decided.

Since condition $\widetilde{f}(\mathbf{v}_n) = \underline{\widetilde{f}}\left(\mathbf{v}_n \mid \mathbf{v}_n\right)$ is already satisfied, we next construct parameters $\mathbf{u}$ and $\mathbf{M}$ satisfying conditions (56) and (57). We first use condition (57) to design $\mathbf{u}$. The differential of the left hand side of (62) at point $\mathbf{v}_n$ with arbitrary increment $\mathrm{d}\mathbf{v} = \mathbf{v} - \mathbf{v}_n$ is

$$\begin{aligned} &\mathrm{d}\widetilde{f}(\mathbf{v})\Big|_{\mathbf{v}=\mathbf{v}_n} \\ &= -\frac{1}{\mu}\frac{\sum\limits_k \mathrm{d}\left\{\exp\left\{-\mu\Big(\mathrm{const}_k + 2\,\mathrm{Re}\left\{\left(\mathbf{f}_k^n\right)^H\mathbf{v}\right\}\Big)\right\}\right\}\Big|_{\mathbf{v}=\mathbf{v}_n}}{\sum\limits_k \exp\left\{-\mu\Big(\mathrm{const}_k + 2\,\mathrm{Re}\left\{\left(\mathbf{f}_k^n\right)^H\mathbf{v}_n\right\}\Big)\right\}} \\ &= \sum_k 2\,\mathrm{Re}\left\{l_k^n\left(\mathbf{f}_k^n\right)^H\mathrm{d}\mathbf{v}\right\}, \end{aligned} \tag{63}$$

$$\hat{\mathbf{Q}}^H\hat{\mathbf{Q}} \sim \mathcal{CW}_K\left(M, \frac{\beta}{\delta+1}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon} + \mathbf{\Omega}\mathbf{\Upsilon}^2 + \frac{\sigma^2}{\tau p}\mathbf{\Upsilon}^2 + \frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1\right). \tag{49}$$

$$\mathbb{E}\left\{\left(\hat{\mathbf{Q}}^H\hat{\mathbf{Q}}\right)^{-1}\right\} = \frac{\left(\frac{\beta}{\delta+1}\mathbf{\Upsilon}\mathbf{H}_1^H\mathbf{H}_1\mathbf{\Upsilon} + \mathbf{\Omega}\mathbf{\Upsilon}^2 + \frac{\sigma^2}{\tau p}\mathbf{\Upsilon}^2 + \frac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1\right)^{-1}}{M - K}. \tag{50}$$

where l_k^n is defined in (35). Next, the differential of the right hand side of (62) at point $\mathbf{v}_n$ is

$$\mathrm{d}\underline{\widetilde{f}}(\mathbf{v} \mid \mathbf{v}_n)\Big|_{\mathbf{v}=\mathbf{v}_n} = 2\,\mathrm{Re}\left\{\mathbf{u}^H \mathrm{d}\mathbf{v}\right\}. \tag{64}$$

To satisfy condition (57), we need $\sum_k 2\,\mathrm{Re}\left\{l_k^n \left(\mathbf{f}_k^n\right)^H \mathrm{d}\mathbf{v}\right\} = 2\,\mathrm{Re}\left\{\mathbf{u}^H \mathrm{d}\mathbf{v}\right\}$, resulting in

$$\mathbf{u} = \sum_k l_k^n \mathbf{f}_k^n. \tag{65}$$

Next, we aim to construct $\mathbf{M}$ using condition (56). Letting $\mathbf{v} = \mathbf{v}_n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)$, $\varrho \in [0,1]$, and substituting it into (56), we need

$$\begin{aligned}\widetilde{f}\left(\mathbf{v}_n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right) &\geq \widetilde{f}\left(\mathbf{v}_n\right) + 2\varrho\,\mathrm{Re}\left\{\mathbf{u}^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\} \\ &\quad + \varrho^2\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)^H \mathbf{M}\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\end{aligned} \tag{66}$$

to be satisfied for any ϱ and any $\tilde{\mathbf{v}}$. Since we know that $\widetilde{f}(\mathbf{v})$ and $\underline{\widetilde{f}}(\mathbf{v} \mid \mathbf{v}_n)$ have the same value and differential at point $\mathbf{v}_n$, (56) can now be transformed to the condition that the second-order derivative of the left hand side of (66) is no smaller than that of the right hand side of (66) for any $\varrho \in [0,1]$ and any $\tilde{\mathbf{v}}$ [38].

Specifically, the second-order derivative of the right hand side of (66) is given by (67), as shown at the bottom of the page.

Then, we focus on the left hand side of (66). Its first-order derivative is

$$\begin{aligned}&\frac{\partial}{\partial\varrho}\widetilde{f}\left(\mathbf{v}_n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right) \\ &\quad = \sum_k 2\,\mathrm{Re}\left\{u_k^n(\varrho)\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\},\end{aligned} \tag{68}$$

where $u_k^n(\varrho) = \frac{\exp\{-\mu\tilde{l}_k(\varrho)\}}{\sum_k \exp\{-\mu\tilde{l}_k(\varrho)\}}$, $\tilde{l}_k(\varrho) = \mathrm{const}_k + 2\,\mathrm{Re}\left\{\left(\mathbf{f}_k^n\right)^H\left(\mathbf{v}_n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right)\right\}$, and $\frac{\partial\tilde{l}_k(\varrho)}{\partial\varrho} = 2\,\mathrm{Re}\left\{\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\}$. Then, the second-order derivative can be calculated as follows

$$\begin{aligned}&\frac{\partial}{\partial\varrho^2}\widetilde{f}\left(\mathbf{v}^n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right) \\ &\quad = \sum_k 2\,\mathrm{Re}\left\{\frac{\partial}{\partial\varrho}\left\{u_k^n(\varrho)\right\}\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\},\end{aligned} \tag{69}$$

where

$$\begin{aligned}\frac{\partial u_k^n(\varrho)}{\partial\varrho} &= -2\mu\,\mathrm{Re}\left\{u_k^n(\varrho)\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\} \\ &\quad + \mu u_k^n(\varrho)\left(\sum_k 2\,\mathrm{Re}\left\{u_k^n(\varrho)\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\}\right).\end{aligned} \tag{70}$$

Substituting (70) into (69), we obtain the second-order derivative as follows

$$\begin{aligned}&\frac{\partial}{\partial\varrho^2}\widetilde{f}\left(\mathbf{v}_n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right) \\ &\quad = -\mu\sum_k u_k^n(\varrho)\left(2\,\mathrm{Re}\left\{\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\}\right)^2 \\ &\qquad + \mu\left(\sum_k 2\,\mathrm{Re}\left\{u_k^n(\varrho)\left(\mathbf{f}_k^n\right)^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\}\right)^2.\end{aligned} \tag{71}$$

Define $\mathbf{t} = \tilde{\mathbf{v}} - \mathbf{v}_n$. (71) can be rewritten as a quadratic form of $\mathbf{t}$, as follows

$$\tfrac{\partial}{\partial\varrho^2}\widetilde{f}\left(\mathbf{v}_n + \varrho\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right) = \begin{bmatrix}\mathbf{t}^H & \mathbf{t}^T\end{bmatrix}\mathbf{W}\begin{bmatrix}\mathbf{t} \\ \mathbf{t}^*\end{bmatrix}, \tag{72}$$

where

$$\begin{aligned}\mathbf{W} &= -\mu\sum_k u_k^n(\varrho)\begin{bmatrix}\mathbf{f}_k^n \\ \left(\mathbf{f}_k^n\right)^*\end{bmatrix}\begin{bmatrix}\mathbf{f}_k^n \\ \left(\mathbf{f}_k^n\right)^*\end{bmatrix}^H \\ &\quad + \mu\begin{bmatrix}\sum_k u_k^n(\varrho)\mathbf{f}_k^n \\ \sum_k u_k^n(\varrho)\left(\mathbf{f}_k^n\right)^*\end{bmatrix}\begin{bmatrix}\sum_k u_k^n(\varrho)\mathbf{f}_k^n \\ \sum_k u_k^n(\varrho)\left(\mathbf{f}_k^n\right)^*\end{bmatrix}^H.\end{aligned} \tag{73}$$

Besides, we rewrite the second-order derivative in (67) as

$$\begin{aligned}&2\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)^H\mathbf{M}\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right) \\ &\quad = \begin{bmatrix}\mathbf{t}^H & \mathbf{t}^T\end{bmatrix}\begin{bmatrix}\mathbf{M} & 0 \\ 0 & \mathbf{M}^T\end{bmatrix}\begin{bmatrix}\mathbf{t} \\ \mathbf{t}^*\end{bmatrix}.\end{aligned} \tag{74}$$

To satisfy condition (56), according to (73), we can choose that $\mathbf{M} \preceq \lambda_{\min}(\mathbf{W})\mathbf{I}_N$, where

$$\begin{aligned}\lambda_{\min}(\mathbf{W}) &\overset{(e)}{\geq} -\mu\sum_k u_k^n(\varrho)\lambda_{\max}\left(\begin{bmatrix}\mathbf{f}_k^n \\ \left(\mathbf{f}_k^n\right)^*\end{bmatrix}\begin{bmatrix}\mathbf{f}_k^n \\ \left(\mathbf{f}_k^n\right)^*\end{bmatrix}^H\right) \\ &\overset{(f)}{=} -\mu\sum_k u_k^n(\varrho)\left(\left(\mathbf{f}_k^n\right)^H\mathbf{f}_k^n + \left(\mathbf{f}_k^n\right)^T\left(\mathbf{f}_k^n\right)^*\right) \\ &= -2\mu\sum_k u_k^n(\varrho)\left\|\mathbf{f}_k^n\right\|^2 \overset{(g)}{\geq} -2\mu\max_k\left\|\mathbf{f}_k^n\right\|^2,\end{aligned} \tag{75}$$

according to the following properties: (e) [51] : For Hermitian matrix $\mathbf{X}$ and rank one Hermitian matrix $\mathbf{T}$, we have $\lambda_{\min}(\mathbf{X}+\mathbf{T}) \geq \lambda_{\min}(\mathbf{X}) + \lambda_{\min}(\mathbf{T}) = \lambda_{\min}(\mathbf{X})$. (f) : If $\mathbf{X}$ is rank one, $\lambda_{\max}(\mathbf{X}) = \mathrm{Tr}\{\mathbf{X}\}$. (g) : For non-negative vector $[b_1, b_2, \ldots, b_n]$ and $[c_1, c_2, \ldots, c_n]$, if $c_i \in (0,1)$ and $\sum_{i=1}^n c_i = 1$, then $\sum_{i=1}^n c_i b_i \leq \sum_{i=1}^n c_i \max_{1\leq i\leq n} b_i = \max_{1\leq i\leq n} b_i$.

Based on (75), we can now construct $\mathbf{M} = \left(-2\mu\max_k\left\|\mathbf{f}_k^n\right\|^2\right)\mathbf{I}_N$. Substituting this $\mathbf{M}$ and $\mathbf{u}$ in (65) into (62) completes the proof.

$$\frac{\partial}{\partial\varrho^2}\left\{\widetilde{f}\left(\mathbf{v}_n\right) + 2\varrho\,\mathrm{Re}\left\{\mathbf{u}^H\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\} + \varrho^2\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)^H\mathbf{M}\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)\right\} = 2\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right)^H\mathbf{M}\left(\tilde{\mathbf{v}} - \mathbf{v}_n\right). \tag{67}$$

REFERENCES

[1] K. Zhi, C. Pan, G. Zhou, H. Ren, and K. Wang, "Analysis and optimization of RIS-aided massive MIMO systems with statistical CSI," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, Jul. 2021, pp. 153–158.

[2] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[3] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signals Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.

[4] 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.

[5] Q. Zhang, S. Jin, K.-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.

[6] L. Zhu, J. Zhang, Z. Xiao, and R. Schober, "Optimization of multi-UAV-BS aided millimeter-wave massive MIMO networks," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2020, pp. 1–6.

[7] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Massive MIMO with spatially correlated Rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, May 2019.

[8] P. Liu, K. Luo, D. Chen, and T. Jiang, "Spectral efficiency analysis of cell-free massive MIMO systems with zero-forcing detector," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 795–807, Oct. 2020.

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

[10] C. Pan *et al.*, "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.

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

[13] 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.

[14] 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.

[15] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[16] A. S. de Sena *et al.*, "IRS-assisted massive MIMO-NOMA networks: Exploiting wave polarization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7166–7183, Nov. 2021.

[17] 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.

[18] A. Papazafeiropoulos, C. Pan, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation and beamforming design," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2077–2092, Mar. 2022.

[19] 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.

[20] 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.

[21] J. He, K. Yu, Y. Shi, Y. Zhou, W. Chen, and K. B. Letaief, "Reconfigurable intelligent surface assisted massive MIMO with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4769–4783, Jul. 2022.

[22] E. Bjornson 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, 2020.

[23] 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," 2020, *arXiv:2010.13525*.

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

[25] Y. Han, W. Tang, S. Jin, C. 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.

[26] Y. Han, S. Zhang, L. Duan, and R. Zhang, "Cooperative double-IRS aided communication: Beamforming design and power scaling," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1206–1210, Aug. 2020.

[27] Z. Kang, C. You, and R. Zhang, "IRS-aided wireless relaying: Deployment strategy and capacity scaling," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 215–219, Feb. 2021.

[28] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1229–1243, Feb. 2022.

[29] X. Qian *et al.*, "Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 77–81, Sep. 2021.

[30] Z. Xing, R. Wang, J. Wu, and E. Liu, "Achievable rate analysis and phase shift optimization on intelligent reflecting surface with hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5514–5530, Sep. 2021.

[31] K. Zhi, C. Pan, H. Ren, and K. Wang, "Ergodic rate analysis of reconfigurable intelligent surface-aided massive MIMO systems with ZF detectors," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 264–268, Feb. 2022.

[32] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.

[33] A. Papazafeiropoulos, "Ergodic capacity of IRS-assisted MIMO systems with correlation and practical phase-shift modeling," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 421–425, Feb. 2022.

[34] B. Zheng *et al.*, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec. 2020.

[35] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 2–17, Jan. 2021.

[36] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

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

[38] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast miso communication systems," *IEEE Trans. Signal Process.*, vol. 68, pp. 3236–3251, 2020.

[39] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2411–2422, Oct. 2007.

[40] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[41] K. Zhi, C. Pan, H. Ren, and K. Wang, "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.

[42] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, vol. 197. Hoboken, NJ, USA: Wiley, 2009.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[44] S. Jin, X. Gao, and X. You, "On the ergodic capacity of rank-1 Ricean-fading MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 502–517, Feb. 2007.

[45] T. Ratnarajah, "Topics in complex random matrices and information theory," Ph.D. dissertation, Dept. Math. Statist., Univ. Ottawa, Ottawa, ON, Canada, 2003.

[46] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimensional Syst. Signal Process.*, vol. 5, no. 3, pp. 263–279, Jul. 1994.

[47] J. Steyn and H. S. Roux, "Approximations for the non-central Wishart distribution," *South Afr. Stat. J.*, vol. 6, no. 2, pp. 165–173, 1972.

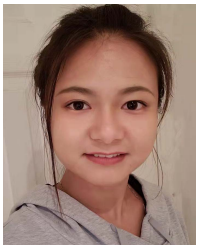
[48] C. Siriteanu, Y. Miyanaga, S. D. Blostein, S. Kuriki, and X. Shi, "MIMO zero-forcing detection analysis for correlated and estimated Rician fading," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3087–3099, Sep. 2012.

[49] C. Siriteanu, A. Takemura, S. Kuriki, D. S. P. Richards, and H. Shin, "Schur complement based analysis of MIMO zero-forcing for Rician fading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1757–1771, Apr. 2015.
[50] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 5000–5004.
[51] H. Lütkepohl, "Handbook of matrices," *Comput. Statist. Data Anal.*, vol. 2, no. 25, p. 243, 1997.

Kangda Zhi (Graduate Student Member, IEEE) 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 the 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 (Member, IEEE) 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 with the University of Kent, U.K. He held a post-doctoral position with the Queen Mary University of London, U.K., from 2016 and 2019, where he was a Lecturer from 2019 to 2021. Since 2021, he has been a Full Professor with Southeast University. He has published over 120 IEEE journal articles. His research interests mainly include reconfigurable intelligent surfaces (RIS), intelligent reflection surface (IRS), ultra-reliable low latency communication (URLLC), machine learning, UAV, the Internet of Things, and mobile edge computing. He is a Workshop Organizer in IEEE ICCC 2021 on the topic of Reconfigurable Intelligent Surfaces for Next Generation Wireless Communications (RIS for 6G Networks) and the 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 the Reconfigurable Intelligent Surfaces Emerging Technology Initiative. He is the Workshop Chair for IEEE WCNC 2024 and the 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 is currently an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE COMMUNICATIONS LETTERS, and IEEE ACCESS. He serves as the Guest Editor for the Special Issue on xURLLC in 6G: Next Generation Ultra-Reliable and Low-Latency Communications of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He also serves as a Leading Guest Editor for the Special Issue on Advanced Signal Processing for Reconfigurable Intelligent Surface-Aided 6G Networks of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (JSTSP), the Leading Guest Editor for the Special Issue on Backscatter and Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G of *IEEE Vehicular Technology Magazine*, the Leading Guest Editor for the Special Issue on Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G and Beyond of IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY, and the Leading Guest Editor for the Special Issue on Reconfigurable Intelligent Surface Aided Communications for 6G and Beyond of IEEE ACCESS.



Gui Zhou (Graduate Student Member, IEEE) received the B.S. and M.E. degrees from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2015 and 2019, respectively. She is currently pursuing the Ph.D. degree with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. Her major research interests include reconfigurable intelligent surface (RIS) and signal processing.

Hong Ren (Member, IEEE) 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 the 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.

Maged Elkashlan (Senior Member, IEEE) received the Ph.D. 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 with the University of New South Wales, the University of Sydney, and the University of Technology Sydney. In 2011, he joined the School of Electronic Engineering and Computer Science, Queen Mary University of London. He also holds a Visiting Faculty appointment with the Beijing University of Posts and Telecommunications. His research interests fall into the broad areas of communication theory and statistical signal processing. He 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 for 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 for 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*.

Robert Schober (Fellow, IEEE) received the Diploma and Ph.D. degrees in electrical engineering from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany, in 1997 and 2000, respectively.

From 2002 to 2011, he was a Professor and the Canada Research Chair with The University of British Columbia (UBC), Vancouver, Canada. Since January 2012, he has been 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.

Dr. Schober is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada and a member of the German National Academy of Science and Engineering. Currently, he serves as a member for the Editorial Board of the PROCEEDINGS OF THE IEEE, a Member at Large for the ComSoc Board of Governors, and a ComSoc Treasurer. He 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, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, the 2011 Alexander von Humboldt Professorship, the 2012 NSERC E. W. R. Stacie Fellowship, the 2017 Wireless Communications Recognition Award by the IEEE Wireless Communications Technical Committee, and the 2021 ACM NanoCom Milestone Award for "Fundamental contributions to the modeling, design, and analysis of molecular communication systems." Since 2017, he has been listed as a highly cited researcher by the Web of Science. He served as the Editor-in-Chief for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2012 to 2015 and as the VP Publications for the IEEE Communication Society (ComSoc) in 2020 and 2021.