Northumbria Research Link

Northumbria University NEWCASTLE

University Library

Ergodic Rate Analysis of Reconfigurable Intelligent Surface-Aided Massive MIMO Systems with ZF Detectors

Kangda Zhi, Cunhua Pan, Hong Ren and Kezhi Wang

Abstract—This letter investigates the reconfigurable intelligent surface (RIS)-aided massive multiple-input multiple-output (MIMO) systems with a two-timescale design. First, the zero-forcing (ZF) detector is applied at the base station (BS) based on instantaneous aggregated channel state information (CSI), which is the superposition of the direct channel and the cascaded user-RIS-BS channel. Then, by leveraging the channel statistical property, we derive the closed-form ergodic achievable rate expression. Using a gradient ascent method, we design the RIS passive beamforming relying only on the long-term statistical CSI. We prove that the ergodic rate scales on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, where M and N denote the number of BS antennas and RIS elements, respectively. We also prove the striking superiority of the considered RIS-aided system with ZF detectors over the RIS-free systems and RIS-aided systems with maximum-ratio combining (MRC).

Index Terms—Reconfigurable intelligent surface (RIS), intelligent reflecting surface (IRS), statistical CSI, massive MIMO, ZF.

I. INTRODUCTION

As an emerging technique, reconfigurable intelligent surface (RIS) has been widely investigated and recognized as a cost-effective complement for future systems [1], [2]. The RIS mainly consists of a large number of passive reflecting elements that have low hardware cost and energy consumption. Besides, RIS can help conventional systems overcome the blockage issue and assist the transmission by creating high-quality transmission paths.

A well-acknowledged challenge for the RIS is that it may introduce heavy channel estimation overhead. Fortunately, a novel and more practical countermeasure, named as two-timescale beamforming design, has been proposed and validated by some contributions [3]–[9]. On one hand, the two-timescale scheme aims at designing the passive RIS beamforming based on purely statistical channel state information (CSI), and this could effectively reduce the overhead and the energy consumption in the operation of the RIS [5]. On the other hand, the two-timescale scheme designs the BS beamforming based on the instantaneous aggregated channel,

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

H. Ren is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (hren@seu.edu.cn).

K. Wang is with Department of Computer and Information Sciences, Northumbria University, UK. (e-mail: kezhi.wang@northumbria.ac.uk).
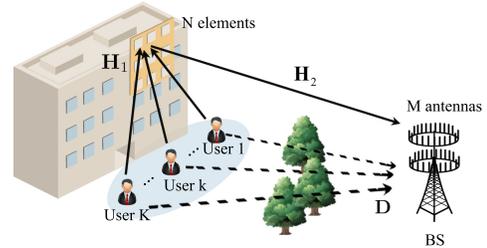
Fig. 1. Uplink transmission in RIS-aided massive MIMO systems.

which is the superposition of the direct channel and cascaded user-RIS-BS channel. As a result, the aggregated channel has the same dimension as that in conventional systems. Thus, two-timescale schemes possess the same channel estimation overhead as conventional systems.

Inspired by the above benefits, the two-timescale design has been recently exploited in RIS-aided massive MIMO systems [10]. It has demonstrated that by integrating an RIS into conventional massive MIMO systems, the rate performance can be significantly improved, especially when the original direct links are weak due to the blockage. However, only the simple maximum-ratio combining (MRC) detector was considered in [10], and it was revealed that the achieved gains are limited by the multi-user interference. Therefore, it is expected that the zero-forcing (ZF) detector, which can effectively mitigate this interference, is more suitable for RIS-aided massive MIMO systems. Different from [10], when using ZF detectors, matrix inversion operator introduces the additional technical challenges of deriving the ergodic capacity.

Against the above background, in this letter, we consider an RIS-aided massive MIMO system with ZF detectors. We derive the closed-form expression for the ergodic rate, which only relies on the long-term CSI. We then design the RIS based on a gradient ascent algorithm. By analyzing the rate expression, we find that it scales on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$, which indicates that the RIS-aided massive MIMO system with ZF detectors has the ability to achieve ultra-high system capacity.

II. SYSTEM MODEL

As illustrated in Fig. 1, the uplink transmission of a massive MIMO system is considered. Different from conventional systems, an RIS is introduced and equipped at the facade of a tall building close to K single-antenna users to improve their channel conditions. The considered model is especially suitable for the scenario where some cell-edge users suffer from service degradation. Denote the number of BS antennas

and RIS elements as M and N, respectively, where $M > K$. Then, we can define the channel between the users and the RIS, the channel between the RIS and the BS, the direct channel between the users and the BS as $\mathbf{H}_1 \in \mathbb{C}^{N \times K}$, $\mathbf{H}_2 \in \mathbb{C}^{M \times N}$ and $\mathbf{D} \in \mathbb{C}^{M \times K}$, respectively.

Define the phase shift matrix of the RIS as $\boldsymbol{\Phi} = \mathrm{diag}\left\{e^{j\theta_1}, \ldots, e^{j\theta_N}\right\}$ where θ_n is the phase shift of the n-th RIS element. Herein, we can express the cascaded user-RIS-BS channel as $\mathbf{G} = \mathbf{H}_2\boldsymbol{\Phi}\mathbf{H}_1 \in \mathbb{C}^{M \times K}$, and then express the aggregated channel from users to the BS as $\mathbf{Q} = \mathbf{G} + \mathbf{D} \in \mathbb{C}^{M \times K}$. It is worth noting that this aggregated channel $\mathbf{Q}$ possesses the same dimension as conventional massive MIMO systems.

Based on the above definitions, we next present the detailed channel model for $\mathbf{Q}$. Firstly, considering that the direct links may be easily blocked [3], we adopt the Rayleigh channel model for $\mathbf{D}$ as follows

$$\mathbf{D} = \tilde{\mathbf{D}}\boldsymbol{\Omega}_d^{1/2}, \tag{1}$$

where $\boldsymbol{\Omega}_d = \mathrm{diag}\left\{\gamma_1, \ldots, \gamma_K\right\}$, and γ_k denotes the distance-dependent path-loss factor. Each element of matrix $\tilde{\mathbf{D}} \in \mathbb{C}^{M \times K}$ is independent and identically distributed (i.i.d.) complex Guassian random variables, whose mean is zero and variance is unit.

Next, since we consider that the RIS is deployed close to the users, and according to the fact that the RIS is often installed above the ground, we assume that the user-RIS channels have purely line-of-sight (LoS) paths. Then, we denote

$$\mathbf{H}_1 = \left[\sqrt{\alpha_1}\,\overline{\mathbf{h}}_1, \ldots, \sqrt{\alpha_K}\,\overline{\mathbf{h}}_K\right], \tag{2}$$

where α_k denotes the path loss. To specify the LoS channel $\overline{\mathbf{h}}_k$, we utilize the two-dimensional uniform squared planar array (USPA) model [11]. Then, the array response vector for a $\sqrt{X} \times \sqrt{X}$ USPA can be expressed as follows

$$\mathbf{a}_X\left(\vartheta^a, \vartheta^e\right) = \left[1, \ldots, e^{j2\pi\frac{d}{\lambda}(x\sin\vartheta^e\sin\vartheta^a + y\cos\vartheta^e)},\right.$$
$$\left.\ldots, e^{j2\pi\frac{d}{\lambda}\left(\left(\sqrt{X}-1\right)\sin\vartheta^e\sin\vartheta^a + \left(\sqrt{X}-1\right)\cos\vartheta^e\right)}\right]^T, \tag{3}$$

where $0 \leq x, y \leq \sqrt{X} - 1$ are element indices in the two-dimensional planar array, d and λ denote the element spacing and wavelength, ϑ^a and ϑ^e are azimuth and elevation angles in the propagation path, respectively. Therefore, denoting by the azimuth and elevation AoA of user k as φ_{kr}^a and φ_{kr}^e, we can now express $\overline{\mathbf{h}}_k = \mathbf{a}_N\left(\varphi_{kr}^a, \varphi_{kr}^e\right)$.

Since the RIS is placed near the users, the distance between the RIS and the BS could be a bit large. Even though both the RIS and the BS have certain heights, it is still not guaranteed that the RIS-BS channel is purely LoS. As a result, the Rician model is suitable for the considered RIS-BS channel. Besides, by adjusting the value of Rician factors, we can study the impacts of scatterers in RIS-aided systems. This feature is important, since many works have proven that rich scattering environment is beneficial in conventional massive MIMO systems [12], while the corresponding impact in RIS-aided massive MIMO systems with ZF detector is still unknown. Thus, we define

$$\mathbf{H}_2 = \sqrt{\frac{\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2 + \sqrt{\frac{\beta}{\delta+1}}\tilde{\mathbf{H}}_2, \tag{4}$$

where β is the path loss, δ represents the Rician factor. Note that the Rician factor, varying from 0 to ∞, characterizes the strength ratio between LoS and non-LoS (NLoS) paths. The NLoS path $\tilde{\mathbf{H}}_2$ contains i.i.d. complex Gaussian random variables with zero mean and unit variance. Recalling USPA model (3), the LoS path, $\overline{\mathbf{H}}_2$, is written as

$$\overline{\mathbf{H}}_2 = \mathbf{a}_M\left(\phi_r^a, \phi_r^e\right)\mathbf{a}_N^H\left(\varphi_t^a, \varphi_t^e\right) \triangleq \mathbf{a}_M\mathbf{a}_N^H, \tag{5}$$

where a notational simplification $\overline{\mathbf{H}}_2 \triangleq \mathbf{a}_M\mathbf{a}_N^H$ is applied in the sequel of this paper. Note that the rank of matrix $\overline{\mathbf{H}}_2$ is one. This means that when $\delta \to \infty$, the cascaded channel $\mathbf{G}$ may become rank-deficient, and then the achievable spatial multiplexing gains may degrade.

We can now express the $M \times 1$ received signal vector at the BS as

$$\mathbf{y} = \sqrt{p}\mathbf{Q}\mathbf{x} + \mathbf{n} = \sqrt{p}\left(\mathbf{H}_2\boldsymbol{\Phi}\mathbf{H}_1 + \mathbf{D}\right)\mathbf{x} + \mathbf{n}, \tag{6}$$

where $\mathbf{x} = \left[x_1, \ldots, x_K\right]^T \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_K\right)$ includes the transmit symbols from K users, and $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_M\right)$ is the noise vector. For simplicity, we assume that all users transmit with the same power p.

To facilitate the analysis, we assume that in each channel coherence time, the instantaneous aggregated channel $\mathbf{Q}$ is perfectly known at the BS, which serves as an upper bound for practical systems. Based on the two-timescale design framework, we need to design the BS beamforming based on instantaneous aggregated CSI, i.e., $\mathbf{Q}$. Thus, the ZF detector at the BS is designed as $\mathbf{A} = \mathbf{Q}\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}$, which results in $\mathbf{A}^H\mathbf{Q} = \mathbf{I}_K$. Thus, the detected symbol vector is given by

$$\mathbf{r} = \mathbf{A}^H\mathbf{y} = \sqrt{p}\mathbf{x} + \left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\mathbf{Q}^H\mathbf{n}. \tag{7}$$

For ZF, there is no multi-user interference. As a result, the signal-to-interference-plus-noise ratio (SINR) reduces to the ratio of transmit power and noise. Based on (7), the SINR of user k is given by

$$\mathrm{SINR}_k = \frac{p\left[\mathbb{E}_\mathbf{x}\left\{\mathbf{x}\mathbf{x}^H\right\}\right]_{kk}}{\left[\mathbb{E}_\mathbf{n}\left\{(\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H\mathbf{n}\mathbf{n}^H\mathbf{Q}(\mathbf{Q}^H\mathbf{Q})^{-1}\right\}\right]_{kk}}$$
$$= \frac{p}{\sigma^2\left[(\mathbf{Q}^H\mathbf{Q})^{-1}\right]_{kk}}. \tag{8}$$

Then, the k-th user's ergodic rate is lower bounded by

$$R_k = \mathbb{E}\left\{\log_2\left(1 + \mathrm{SINR}_k\right)\right\} \tag{9}$$
$$\overset{(a)}{\geq} \log_2\left(1 + \frac{p}{\sigma^2\mathbb{E}\left\{[(\mathbf{Q}^H\mathbf{Q})^{-1}]_{kk}\right\}}\right), \tag{10}$$

where (a) utilizes the Jensen's inequality based on the fact that function $f(x) = \log_2\left(1 + \frac{1}{x}\right)$ is convex with respect to x.

III. RATE ANALYSIS AND RIS DESIGN

In this section, we first derive the closed-form expression for the rate R_k, and then use the derived expression to propose a statistical CSI-based RIS design.

To derive R_k, we need to compute $\mathbb{E}\left\{\left[(\mathbf{Q}^H\mathbf{Q})^{-1}\right]_{kk}\right\}$. To this end, we expand matrix $\mathbf{Q}^H$ as

$$\mathbf{Q}^H = \sqrt{\frac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\boldsymbol{\Phi}^H\overline{\mathbf{H}}_2^H + \sqrt{\frac{\beta}{\delta+1}}\mathbf{H}_1^H\boldsymbol{\Phi}^H\tilde{\mathbf{H}}_2^H + \boldsymbol{\Omega}_d^{1/2}\tilde{\mathbf{D}}^H. \tag{11}$$

For ease of exposition, we define $\mathbf{Q}^H \triangleq [\mathbf{q}_1, \ldots, \mathbf{q}_M]$, $\overline{\mathbf{H}}_2^H \triangleq [\overline{\mathbf{c}}_1, \ldots, \overline{\mathbf{c}}_M]$, $\hat{\mathbf{H}}_2^H \triangleq [\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_M]$, and $\tilde{\mathbf{D}}^H \triangleq \left[\tilde{\mathbf{d}}_1, \ldots, \tilde{\mathbf{d}}_M\right]$. Recalling that $\tilde{\mathbf{H}}_2$ and $\tilde{\mathbf{D}}$ are all comprised of i.i.d. complex Gaussian variables, and $\tilde{\mathbf{H}}_2$ and $\tilde{\mathbf{D}}$ are mutual independent, we therefore have

$$\tilde{\mathbf{c}}_m \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_N\right), 1 \leq m \leq M, \tag{12}$$

$$\tilde{\mathbf{d}}_m \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_K\right), 1 \leq m \leq M, \tag{13}$$

where $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{c}}_j$ are mutual independent for $i \neq j$; $\tilde{\mathbf{d}}_i$ and $\tilde{\mathbf{d}}_j$ are mutual independent, for $i \neq j$; $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{d}}_j$ are mutual independent for all i and j. Then, since the linear transformation for a standard Gaussian random vector is still a Gaussian random vector [13], we obtain

$$\begin{aligned} &\sqrt{\tfrac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{c}}_m + \sqrt{\tfrac{\beta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\tilde{\mathbf{c}}_m \\ &\sim \mathcal{CN}\left(\sqrt{\tfrac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{c}}_m, \tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1\right), \forall m \end{aligned} \tag{14}$$

and $\mathbf{\Omega}_d^{1/2}\tilde{\mathbf{d}}_m \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Omega}_d\right), \forall m$, where the facts $\mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}_N$ and $\mathbf{\Omega}_d = \mathbf{\Omega}_d^H$ were used.

Next, taking into account that the sum of independent Gaussian vectors is still Gaussian distributed [13, Theorem 1.2.14], we can obtain the statistics of the m-th column of aggregated channel $\mathbf{Q}^H$ as follows

$$\mathbf{q}_m \sim \mathcal{CN}\left(\sqrt{\tfrac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{c}}_m, \tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d\right), \tag{15}$$

where $\mathbf{q}_m$, $1 \leq m \leq M$, are mutual independent. Therefore, $\mathrm{vec}\left(\mathbf{Q}^H\right)$ is a complex Gaussian vector with the following mean and covariance matrices

$$\begin{aligned} \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{Q}^H\right)\right\} &= \mathrm{vec}\left(\sqrt{\tfrac{\beta\delta}{\delta+1}}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\right), \\ \mathrm{Cov}\left\{\mathrm{vec}\left(\mathbf{Q}^H\right)\right\} &= \mathbf{I}_M \otimes \left(\tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d\right), \end{aligned} \tag{16}$$

where vec and $\otimes$ denote the vectorization by column stacking and Kronecker product, respectively.

Then, using the distribution of $\mathrm{vec}\left(\mathbf{Q}^H\right)$ and following the notations in [12, Page 2], matrix $\mathbf{Q}$ is a complex Gaussian distributed matrix, written as

$$\mathbf{Q} \sim \mathcal{CN}\left(\sqrt{\tfrac{\beta\delta}{\delta+1}}\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1, \mathbf{I}_M \otimes \left(\tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d\right)\right). \tag{17}$$

Therefore, the product $\mathbf{Q}^H\mathbf{Q}$ has a complex non-central Wishart distribution [13, Definition 10.3.1], which can be repressed as

$$\begin{aligned} \mathbf{Q}^H\mathbf{Q} \sim \mathcal{W}_K\Big(&M, \tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d \\ &, (\tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d)^{-1}\tfrac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1\Big). \end{aligned} \tag{18}$$

Even though the non-central Wishart distribution (18) is accurate, its statistics are very complicated, and then we cannot obtain a tractable expression for insightful analysis. To facilitate the analysis, as in contributions [14]–[16], we next approximate the non-central Wishart distribution (18) as a central Wishart distribution with the same first-order moment.

To begin with, the first-order moment for the considered non-central Wishart distribution is [15, Eq. (45)]

$$\begin{aligned} &\mathbb{E}\left\{\mathbf{Q}^H\mathbf{Q}\right\} \\ &= M\left(\tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d\right) + \tfrac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\overline{\mathbf{H}}_2^H\overline{\mathbf{H}}_2\mathbf{\Phi}\mathbf{H}_1 \\ &= M\left(\tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d\right) + M\tfrac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1, \end{aligned} \tag{19}$$

where the last equality is obtained by using (5) and $\mathbf{a}_M^H\mathbf{a}_M = M$.

Therefore, a virtual central Wishart distribution with this moment is given by [15, Sec. V. A]

$$\mathbf{Q}^H\mathbf{Q} \sim \mathcal{W}_K\Big(M, \tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d + \tfrac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1\Big). \tag{20}$$

Based on the obtained complex central Wishart distribution (20), with the help of [17, Table I], we can obtain the expectation of the matrix inverse as follows

$$\mathbb{E}\left\{\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\right\} = \frac{\left(\tfrac{\beta}{\delta+1}\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{\Omega}_d + \tfrac{\beta\delta}{\delta+1}\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1\right)^{-1}}{M - K}. \tag{21}$$

Substituting (21) into (10), we obtain the lower bound of the ergodic rate of user k as follows

$$R_k \geq \log_2\left(1 + \frac{p(M-K)}{\sigma^2(\delta+1)\left[(\mathbf{\Lambda} + \beta\delta\mathbf{H}_1^H\mathbf{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\mathbf{\Phi}\mathbf{H}_1)^{-1}\right]_{kk}}\right), \tag{22}$$

where $\mathbf{\Lambda} = \beta\mathbf{H}_1^H\mathbf{H}_1 + (\delta+1)\mathbf{\Omega}_d$.

Note that the derived expression, (22), depends only on the statistical CSI, since the instantaneous CSI-related variables have been averaged out. Therefore, based on the two-timescale design framework, we can use (22) to design the phase shifts of the RIS only relying on statistical CSI. Since the statistical CSI-based phase shifts design only needs to be done on a large time-scale, the overhead can be effectively reduced. Besides, it is clear that (22) is an increasing function of p and the RIS-BS channel strength β, but it is a decreasing function of noise power σ^2.

Corollary 1 *As $M \to \infty$, the rate can maintain non-zero when the power is scaled down proportionally to $p = 1/M$.*

Proof: It can be proved by noticing that all the matrices in the denominator of (22) do not depend on M. ∎

Corollary 2 *When $M \to \infty$ or $p \to \infty$, RIS-aided massive MIMO systems with ZF detectors perform much better than that with MRC detectors.*

Proof: Based on (22), when $M \to \infty$ or $p \to \infty$, we have $R_k \to \infty$, while the rate in RIS-aided massive MIMO systems with MRC detectors is still bounded due to the multi-user interference, as proved in [10, Eq. (7)]. ∎

Corollary 3 *When $\beta = 0$, i.e., without the existence of the RIS, the rate of user k reduces to*

$$R_k \geq \log_2\left(1 + p(M-K)\gamma_k/\sigma^2\right), \tag{23}$$

which is the same rate as [18, Eq. (20)], and scales on the order of $\mathcal{O}\left(\log_2\left(M\right)\right)$.

Corollary 4 *The ergodic rate of user k in (22) is further lower bounded by*

$$R_k \geq \log_2\left(1 + \frac{p\left(M - K\right)}{\sigma^2(\delta + 1)\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}}\right) \tag{24}$$

$$\approx \log_2\left(1 + \frac{p\left(M - K\right)}{\sigma^2}\left(\frac{N\alpha_k\beta}{\delta + 1} + \gamma_k\right)\right), \; as \; N \to \infty. \tag{25}$$

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

Proof: Since we consider the existence of direct links, there is $\boldsymbol{\Omega}_d \succ \mathbf{0}$ and then $\boldsymbol{\Lambda} \succ \mathbf{0}$ and $\boldsymbol{\Lambda}^{-1} \succ \mathbf{0}$. Besides, we have $\boldsymbol{\Lambda}^H = \boldsymbol{\Lambda}$. Based on the Woodbury's identity, we have

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

$$= \left[\boldsymbol{\Lambda}^{-1}\right]_{kk} - \frac{\beta\delta\left[\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\right]_{kk}}{1 + \beta\delta\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N} \tag{26}$$

$$= \left[\boldsymbol{\Lambda}^{-1}\right]_{kk} - \frac{\beta\delta\left|\left[\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N\right]_k\right|^2}{1 + \beta\delta\mathbf{a}_N^H\boldsymbol{\Phi}\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\boldsymbol{\Phi}^H\mathbf{a}_N} \leq \left[\boldsymbol{\Lambda}^{-1}\right]_{kk}. \tag{27}$$

Substituting (27) into (22), we arrive at (24). Note that the k-th diagonal element of $\mathbf{H}_1^H\mathbf{H}_1$ equals $\alpha_k N$, while the non-diagonal elements are not proportional to N. When $N \to \infty$, we can approximate $\mathbf{H}_1^H\mathbf{H}_1$ as $N\mathrm{diag}\left\{\alpha_1, \ldots, \alpha_K\right\}$, which results in the approximation in (25). ∎

Corollary 4 reveals a very promising capacity gain. It is well-known that the ergodic rate of RIS-aided systems scales as $\mathcal{O}\left(\log_2\left(MN^2\right)\right)$ in the single-user scenario [3]. Here, we prove that by using ZF detectors in setup of multiple users, the rate of each user could still scale as $\mathcal{O}\left(\log_2\left(MN\right)\right)$, which demonstrates that the considered systems can achieve a promising sum user rate. Besides, comparing (25) with (23), it is shown that the RIS-aided massive MIMO systems with ZF detectors always outperform RIS-free massive MIMO systems. Meanwhile, it can be observed that the lower bound (25) tends to (22) when $\delta \to 0$ and tends to (23) when $\delta \to \infty$.

Next, we design the RIS phase shifts based on (22), which depends only on the statistical CSI. The sum-rate maximization problem can be formulated as follows

$$\max_{\boldsymbol{\Phi}} \quad R^s = \sum_{k=1}^{K} R_k, \tag{28a}$$

$$\text{s.t.} \quad \left|\left[\boldsymbol{\Phi}\right]_{nn}\right| = 1, \; 1 \leq n \leq N. \tag{28b}$$

Problem (28) is non-convex due to the non-convex unit modulus constraint. However, we can still obtain a sub-optimal solution based on the gradient ascent method. For tractability, we rewrite $\boldsymbol{\Phi} = \mathrm{diag}\left\{\mathbf{v}^H\right\}$, where $\mathbf{v} = \left[e^{j\theta_1}, \ldots, e^{j\theta_N}\right]^H$. Then, we provide the gradient vector with respect to $\mathbf{v}$ in the following lemma.

Lemma 1 *The gradient of the objective function in (28) is*

$$\frac{\partial R^s(\mathbf{v})}{\partial \mathbf{v}^*} = \sum_{k=1}^{K} \frac{\frac{\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}} - \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}\mathbf{A}_k\mathbf{v}}{\left(\mathbf{v}^H\mathbf{A}_k\mathbf{v}\right)^2}}{\ln(2)\left(1 + \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)}, \tag{29}$$

where

$$\mathbf{A}_k = \frac{\sigma^2(\delta + 1)}{p(M - K)}\left(\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}\mathbf{B} - \beta\delta\mathbf{s}_k\mathbf{s}_k^H\right), \tag{30}$$

$$\mathbf{B} = \frac{1}{N}\mathbf{I}_N + \beta\delta\,\mathrm{diag}\left(\mathbf{a}_N^H\right)\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\,\mathrm{diag}\left(\mathbf{a}_N\right), \tag{31}$$

with $\mathbf{s}_k^H \triangleq \left[\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\,\mathrm{diag}\left(\mathbf{a}_N\right)\right]_{(k,:)}$ corresponds to the k-th row vector.

Proof: Substituting $\boldsymbol{\Phi}^H\mathbf{a}_N = \mathrm{diag}\left(\mathbf{a}_N\right)\mathbf{v}$ into (26) and utilize (31), we have

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

$$= \left[\boldsymbol{\Lambda}^{-1}\right]_{kk} - \frac{\beta\delta\left[\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\,\mathrm{diag}(\mathbf{a}_N)\mathbf{v}\mathbf{v}^H\,\mathrm{diag}\left(\mathbf{a}_N^H\right)\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\right]_{kk}}{1 + \beta\delta\mathbf{v}^H\,\mathrm{diag}\left(\mathbf{a}_N^H\right)\mathbf{H}_1\boldsymbol{\Lambda}^{-1}\mathbf{H}_1^H\,\mathrm{diag}\left(\mathbf{a}_N\right)\mathbf{v}}$$

$$= \frac{\mathbf{v}^H\left\{\left[\boldsymbol{\Lambda}^{-1}\right]_{kk}\mathbf{B} - \beta\delta\mathbf{s}_k\mathbf{s}_k^H\right\}\mathbf{v}}{\mathbf{v}^H\mathbf{B}\mathbf{v}}. \tag{32}$$

Then, substituting (32) into (22), the sum rate can be rewritten as $R^s = \sum_{k=1}^{K}\log_2\left(1 + \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)$. Based on the chain rule, the gradient of a real function with respect to complex vector variable is given by [19]

$$\frac{\partial R^s(\mathbf{v})}{\partial \mathbf{v}^*} = \sum_{k=1}^{K} \frac{1}{\ln(2)\left(1 + \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)}\frac{\partial\left(\frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)}{\partial \mathbf{v}^*}. \tag{33}$$

Using $\frac{\partial\left\{\mathbf{v}^H\mathbf{B}\mathbf{v}\right\}}{\partial \mathbf{v}^*} = \mathbf{B}\mathbf{v}$, and $\frac{\partial\left\{\mathbf{v}^H\mathbf{A}_k\mathbf{v}\right\}}{\partial \mathbf{v}^*} = \mathbf{A}_k\mathbf{v}$, we have

$$\frac{\partial\left(\frac{\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)}{\partial \mathbf{v}^*} = \frac{\left\{\frac{\partial\left(\mathbf{v}^H\mathbf{B}\mathbf{v}\right)}{\partial \mathbf{v}^*}\right\}\mathbf{v}^H\mathbf{A}_k\mathbf{v} - \mathbf{v}^H\mathbf{B}\mathbf{v}\left\{\frac{\partial\left(\mathbf{v}^H\mathbf{A}_k\mathbf{v}\right)}{\partial \mathbf{v}^*}\right\}}{\left(\mathbf{v}^H\mathbf{A}_k\mathbf{v}\right)^2}$$

$$= \frac{\mathbf{B}\mathbf{v}\mathbf{v}^H\mathbf{A}_k\mathbf{v} - \mathbf{v}^H\mathbf{B}\mathbf{v}\mathbf{A}_k\mathbf{v}}{\left(\mathbf{v}^H\mathbf{A}_k\mathbf{v}\right)^2} = \frac{\mathbf{B}\mathbf{v}}{\mathbf{v}^H\mathbf{A}_k\mathbf{v}} - \frac{\mathbf{v}^H\mathbf{B}\mathbf{v}\mathbf{A}_k\mathbf{v}}{\left(\mathbf{v}^H\mathbf{A}_k\mathbf{v}\right)^2}. \tag{34}$$

Substituting (34) into (33) completes the proof. ∎

Assume the variable in the t-th iteration is $\mathbf{v}^t$. Then, the next variable $\mathbf{v}^{t+1}$ in the $(t + 1)$-th iteration is given by

$$\tilde{\mathbf{v}}^{t+1} = \mathbf{v}^t + \mu\left.\frac{\partial R^s(\mathbf{v})}{\partial \mathbf{v}^*}\right|_{\mathbf{v} = \mathbf{v}^t}, \tag{35}$$

$$\mathbf{v}^{t+1} = \exp\left(j\arg\left(\tilde{\mathbf{v}}^{t+1}\right)\right), \tag{36}$$

where μ is the step size which can be chosen by using backtracking line search [4], [9]. (36) is a projection operation for meeting the unit modulus constraint (28b).

IV. SIMULATION RESULTS

Unless otherwise stated, we consider $K = 4$ users evenly located on the half-circle centered of an RIS with a radius $d_{UI} = 20$ m. The distance between the RIS and the BS is $d_{IB} = 700$ m. Using d_{UI} and d_{IB}, the distance between the users and the BS can be calculated by their geometric relationship as [10]. Based on the distances, the path loss factors α_k, β and γ_k are calculated the same as [10]. Besides, we set $M = N = 64$, $p = 30$ dBm, $\delta = 1$ and $\sigma^2 = -104$ dBm. The angles in the LoS channels are generated randomly from $[0, 2\pi]$. The Monte Carlo simulations are obtained based on (9) with 10^4 times average.
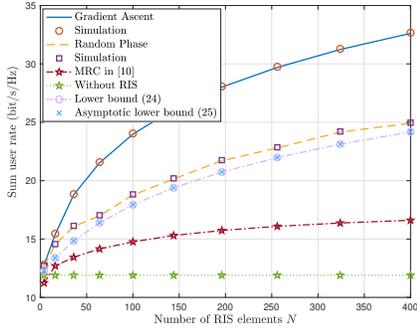
Fig. 2. Rate versus the number of RIS elements N.
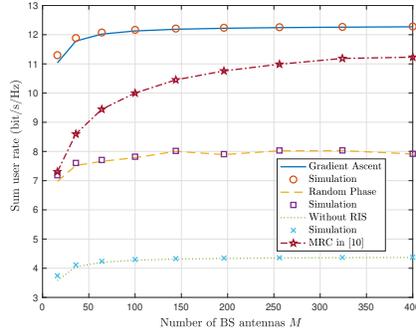
Fig. 3. Rate versus M, where power is scaled down as $p = 10/M$.
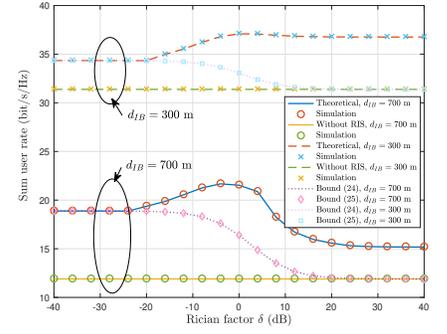
Fig. 4. Rate versus the Rician factor δ.

In Fig. 2, it can be observed that ZF-based RIS design outperforms the random phase shifts-based design, the MRC-based design, and the RIS-free systems. The superiority of ZF over MRC lies in the fact that RIS-aided systems suffer from severe multi-user interference [10]. This is because users share the same RIS-BS channel and then their cascaded channels are highly correlated. Therefore, by effectively eliminating the interference, ZF can achieve a higher ergodic rate than MRC. In addition, Fig. 2 validates the accuracy of the approximate in (25). In Fig. 3, we verify the power scaling law as expected in Corollary 1. It again emphases the advantages of ZF-based RIS systems. Besides, all numerical results show that the derived lower bound (22) is very tight with the Monte Carlo simulations.

Fig. 4 plots the rate versus Rician factor δ. It can be seen that when $d_{IB} = 700$ m, the rate with large δ performs worse, while a contrary result is observed when $d_{IB} = 300$ m. This is because when d_{IB} is large, the direct channel becomes weak, and then the cascaded channel $\mathbf{G}$ becomes a dominant factor. In this case, when δ is large, the channel $\mathbf{G}$ becomes rank-deficient, which degrades the rate. However, when d_{IB} is small, the direct links are strong. Since the direct links have full-rank, the aggregate channel could always have full-rank. As shown in (22), the variable $\mathbf{\Phi}$ can play more roles when δ is large, which results in large performance gains. Furthermore, Fig. 4 validates the tightness of (25) under all Rician factors.

V. Conclusion

An RIS-aided massive MIMO system with ZF detectors was considered in this paper. We first derived the closed-form ergodic rate expressions, whose lower bound demonstrates that the rate can scale on the order of $\mathcal{O}\left(\log_2\left(MN\right)\right)$. Then, using the gradient ascent algorithm, we optimized the phase shifts of the RIS based on statistical CSI. Finally, simulation results validated the correctness of our analytical results.

References

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

[2] 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.

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

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

[5] 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.

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

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

[8] 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.*, early access, 2021.

[9] A. Papazafeiropoulos *et al.*, "Asymptotic analysis of max-min weighted SINR for IRS-assisted MISO systems with hardware impairments," *IEEE Wireless Commun. Lett.*, early access, 2021.

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

[11] S. Zhou, W. Xu *et al.*, "Spectral and energy efficiency of IRS-assisted MISO communication with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1366–1369, Sep. 2020.

[12] 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.

[13] R. J. Muirhead, *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009, vol. 197.

[14] C. Siriteanu, Y. Miyanaga, S. D. Blostein *et al.*, "MIMO zero-forcing detection analysis for correlated and estimated rician fading," *IEEE Trans. Veh. Tech.*, vol. 61, no. 7, pp. 3087–3099, Sep. 2012.

[15] C. Siriteanu, A. Takemura, S. Kuriki *et al.*, "Schur complement based analysis of MIMO zero-forcing for rician fading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1757–1771, Apr. 2015.

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

[17] 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.

[18] 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.

[19] R. Hunger, "An introduction to complex differentials and complex differentiability," 2007.