# QoS-Aware Resource Allocation of RIS-Aided Multi-User MISO Wireless Communications

Ya Gao, *Member IEEE*, Chengzhuang Lu, Yuhang Lian, Xingwang Li, *Senior Member IEEE*,
Gaojie Chen, *Senior Member IEEE*, Daniel Benevides da Costa, *Senior Member IEEE*,
Arumugam Nallanathan, *Fellow IEEE*

***Abstract*—In this paper, it is investigated a multiuser multiple-input single-output wireless communication scenario (MISO) in which a base station (BS) serves multiple users with the assistance of a reconfigurable intelligent surface (RIS). Specifically, the delay sensitivity QoS aware total effective capacity (TEC) is maximized. By employing Zero-Forcing (ZF) beamforming at the BS, the beamforming optimization is converted the power allocation. Then, we alternately optimize QoS provisioning power allocation schemes and RIS phase shifts. In particular, we derive a closed-form expression for the optimal power allocation scheme based on the Lagrangian method and the Karoosh-Koon-Tucker (KKT) condition. Meanwhile, the RIS phase shift strategy is optimized by employing the gradient descent method. Finally, the performance of the proposed joint optimal schemes for RIS-assisted wireless communication is verified by simulations. The simulation results demonstrate that the optimization scheme effectively guarantees the user's delay QoS requirements and the TEC significantly outperforms the random reflecting phase shift scheme.**

***Index Terms*—Reconfigurable intelligent surface, statistical QoS provisioning, effective capacity, convex optimization, gradient descent.**

## I. Introduction

RECONFIGURABLE intelligent surface (RIS) has recently been considered as a promising paradigm to provide intelligent and reconfigurable wireless transmission environment for B5G/6G systems [1]. RIS consists of a large number of low-cost passive reflective elements, each of which is able to adjust the amplitude and/or phase of the incident signal independently. Massively deployed in wireless networks, RIS can assist to create a virtual line-of-sight (LoS) link and bypass obstacles between transmitters and receivers via smart reflection [2], thereby appearing as a new solution to efficiently tackle the wireless channel fading attenuation and interference effects, and potentially achieving a tremendous network capacity improvement and energy consumption reduction [3], [4]. On the other hand, the explosive growth of latency-sensitive multimedia services, i.e., automatic driving, medical observation and so on, has also promoted the requirements for delay Quality of Service (QoS) provisionings. However, due to the time-varying nature and stochastic characteristic of wireless channels, especially in the reflected surfaces assisted wireless communications, traditional deterministic delay QoS guarantees are unrealistic in practice. The statistical delay-bounded QoS provisionings, which is characterized by the queue length violation probabilities, have drawn many attentions [5], [6]. However, RIS was not considered in these previous works. In [7], the authors derived accurate, asymptotic, and approximate closed-form expressions for the effective rate using probability density functions. However, it is worth noting that both the source node and destination have a single antenna, and the number of RIS reflection elements is small. In [8], it was analysed the ergodic channel capacity performance of a RIS assisted Rician channel under simultaneous transmission from follower UAVs (FUAV) to a leader UAV (LUAV). However, the impact of different latency service quality requirements on the results still remains to be investigated in the literature.

Motivated by the aforementioned problems, we consider a heterogeneous statistical QoS guaranteed RIS-assisted MISO system. In addition, we construct the total effective capacity (TEC) maximization problem subject to the total transmit power constraint. To solve this problem, we employ the alternating optimization method. Specifically, we derive the optimal transmit power allocation schemes by employing Lagrangian method and KKT-conditions. Moreover, we iteratively obtain the RIS phase shift strategy using the gradient descent algorithm. The performance of the obtained joint optimal power allocation and RIS phase shift schemes are evaluated by simulations. Simulation results show that the proposed scheme can effectively guarantee the heterogeneous statistical QoS provisioning for RIS-assisted wireless communications.

The rest of this paper is organized as follows. Section II describes the signal model and QoS mechanism for RIS-assisted MISO heterogeneous networks. In Section III, we formulate the effective capacity maximization problem. Section IV derives the joint optimal power allocation strategy and RIS phase shift. Section V conducts some numerical simulations. The paper concludes in Section VI.

"*Notation*: $a$ is a variable, $\mathbf{a}$ is a vector, and $\mathbf{A}$ is a matrix. $\mathbb{C}^{p\times q}$ denotes a complex matrix of dimension $p\times q$. $\mathrm{diag}\left(\cdot\right)$

---

Ya Gao, Chengzhuang Lu and Yuhang Lian are with School of Physics and Electronic Information, Luoyang Normal University, Luoyang, 471934, China. Ya Gao is also with State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, 710071, China. Chengzhuang Lu is also with School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China. (e-mail: gaoya@lynu.edu.cn, 499441737@qq.com, 1319592700@qq.com).

Xingwang Li is with School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, 454150, China. (e-mail: lixingwangbupt@gmail.com).

Gaojie Chen is with the 5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford, GU2 7XH, United Kingdom (e-mail: gaojie.chen@surrey.ac.uk).

Daniel Benevides da Costa is with the Digital Science Research Center, Technology Innovation Institute, 9639 Masdar City, Abu Dhabi, United Arab Emirates. (e-mail: danielbcosta@ieee.org).

Arumugam Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK (e-mail: a.nallanathan@qmul.ac.uk).

Corresponding author: Xingwang Li.

is a diagonal matrix, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^{-1}$, and $(\cdot)^+$ stand for the conjugate-transpose, the transpose, the inverse and the pseudo-inverse of a matrix, respectively. Moreover, $\|\cdot\|$, $\|\cdot\|_{\mathrm{F}}$, $\mathrm{tr}\,(\cdot)$ and $\mathrm{vec}\,(\cdot)$ denote the Euclidian norm of a vector, the Frobenius norm, the trace and a vector stacking all the columns of a matrix, respectively. $\mathrm{Re}\,(\cdot)$, $|\cdot|$ and $(\bar{\cdot})$ denote the real part, the modulus and the conjugate of a complex number. Also, $\mathbf{A}\otimes\mathbf{B}$ and $\mathbf{A}\circ\mathbf{B}$ denote the Kronecker and Hadamard products of A and B. Finally, $\mathbb{E}\{\cdot\}$ represents the operations of expectation.

## II. SYSTEM MODEL

We consider an RIS-assisted multi-user multiple-input and single-output (MISO) wireless communication system, where a base station (BS) transmits wireless information to $K$ teminal users by utilizing reconfigurable intelligent surface (RIS). It is assumed that the BS is equipped with $M$ antennas, each user is a single-antenna device, and RIS has $N$ reflection elements. As illustrated in Fig. 1, configured at the BS, data packets are divided into frames at the data link layer and then decomposed into bit streams at the physical layer. It is considered that the channel state information (CSI) is perfectly known at the BS [9]. The BS jointly optimizes the power control and RIS phase shift based on the QoS index of the service request and the CSI fedback provided by the user. The optimal phase shift is then sent back to the RIS controller.

### A. Signal Model

Assume that the channel links experience a quasi-static flat fading. We denote $\mathbf{\Phi} \in \mathbb{C}^{N\times N}$ the diagonal reflection matrix of the RIS, with $\mathbf{\Phi} = \mathrm{diag}\,[\phi_1, \phi_2, ..., \phi_{\mathrm{n}}]$, where $\phi_n = e^{j\varphi_n}(n \in N)$. Let $\mathbf{x} \in \mathbb{C}^{M\times 1}$ be the signal emitted by the BS, which can be written as $\mathbf{x} = \sum_{i=1}^{K} \mathbf{w}_i s_i$. Here, $\mathbf{w}_i$ and $s_i$ represent the corresponding transmit beamforming vector and the flow of information received by the user $i$ , respectively. The signal received at user $i$ $(i \in K)$, denoted by $y_i$, is expressed as follows

$$y_i = (\mathbf{h}_i \mathbf{\Phi} \mathbf{G})\,\mathbf{x} + n_i, \tag{1}$$

where $\mathbf{h}_i \in \mathbb{C}^{1\times N}$, $\mathbf{G} \in \mathbb{C}^{N\times M}$, and $n_i \sim \mathcal{CN}(0, \sigma_i^2)$ denote the channel vector between RIS and user $i$, the channel matrix between BS and RIS, and the zero-mean complex white Gaussian noise with variance $\sigma_i^2$, respectively. Both links BS-to-RIS and RIS-to-users undergo Nakagami-$m$ fading channel model.

The signal-to-interference-plus-noise ratio (SINR) experienced at the $i$th mobile user, denoted by $\rho_i$, can be written as

$$\rho_i = \frac{\left|(\mathbf{h}_i \mathbf{\Phi} \mathbf{G})\,\mathbf{w}_i\right|^2}{\sum\limits_{k=1,k\neq i}^{K} \left|(\mathbf{h}_i \mathbf{\Phi} \mathbf{G})\,\mathbf{w}_k\right|^2 + \sigma_i^2}. \tag{2}$$

The transmit beamforming vector at the BS needs to satisfy

$$\mathrm{tr}\left(\mathbf{W}\mathbf{W}^H\right) \le P_{\mathrm{T}}, \tag{3}$$

where $\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K] \in \mathbb{C}^{M\times K}$ and $P_{\mathrm{T}}$ is the maximum transmit power at the BS.

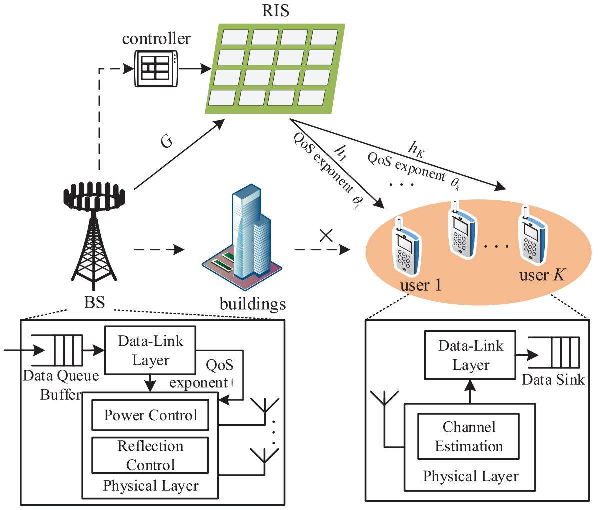

Fig. 1. System model of RIS-assisted QoS-aware downlink multi-user communication system.

### B. QoS Provisionin for RIS-Aided Wireless Communications

Due to the unstable characteristics of wireless channels and the various QoS demands of HetNets, deterministic QoS guarantees alone are unable to meet the heterogeneous QoS requirements. For latency-sensitive services, statistical QoS provisioning offers a solution by describing the probability of delay violation, which occurs when service latency exceeds a predetermined threshold. This approach provides a means to achieve heterogeneous QoS guarantees.

For a queuing system with stationary ergodic arrival properties and service processes, the queue length process $Q\,(t)\,(t > 0)$ is distributedly convergent to a finite random variable $Q\,(\infty)$ satisfying [10]

$$-\lim_{\alpha\to\infty} \frac{\log \Pr\left(Q\,(\infty) \ge \alpha\right)}{\alpha} = \theta, \tag{4}$$

where $\alpha$ is the queue length threshold and $\theta$ is QoS exponent indicating the exponential decay rate controlled by the queue length bound. The value of QoS exponent $\theta$ reflects the delay QoS requirements of the wireless traffics.

Let us denote the QoS exponent for HetNets by $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_K]$, where $\theta_i\,(1 \le i \le K)$ represents the QoS demands corresponding to the $i$th user. We construct the instantaneous data rate, denoted as $R\,[t]\,(t = 1, 2, ...)$, where $t$ is the time exponent. The instantaneous service rate can be written as $R_i\,[t] = B\log_2\,(1 + \rho_i\,[t])$. Considering a stationary and ergodic service process represented by $S_i\,[u] = \sum_{t=1}^{u} R_i\,[t]$, which is the partial sum of instantaneous data rates, the Gartner–Ellis $S\,[t]$ limit can be writen as $\Lambda_C\,(\theta_i) = \lim\limits_{u\to\infty} \frac{1}{u}\log\left(\mathbb{E}\left\{e^{-\theta_i S_i[u]}\right\}\right)$, which exists for all $\theta_i\,(1 \le i \le K)$ [10]. Then, the effective capacity (EC) of the $i$th downlink, represented by $E_C\,(\theta_i)$, can be derived as

$$E_C\,(\theta_i) = \frac{-\Lambda_C\,(\theta_i)}{\theta_i} = -\frac{1}{\theta_i}\log\left(\mathbb{E}\left\{e^{-\theta_i R_i[t]}\right\}\right), \tag{5}$$

we omit the index t below for simplifying calculations.

$$E_C\,(\theta_i) = -\frac{1}{\theta_i}\log\left(\mathbb{E}\left\{e^{-\theta_i B\log_2\,(1+\rho_i)}\right\}\right). \tag{6}$$

Assuming downlink transmission, the total effective capacity for RIS-aided communications can be written as

$$\widetilde{E}_C(\boldsymbol{\theta}) = \sum_{i=1}^{K} -\frac{1}{\theta_i} \log\left(\mathbb{E}\left\{\log_2(1+\rho_i)^{-\beta_i}\right\}\right), \tag{7}$$

where $\beta_i = \theta_i B/\ln 2$ is the normalized QoS index of the $i$th downlink. The total effective capacity is non-convex due to the heterogeneous characteristic of $\theta_i$. Our previous work [11] has proved the existence of unique real valued numbers $\theta_0 \in \{\theta_{\min}, \theta_{\max}\}$. Thus, the following equation holds

$$\widetilde{E}_C(\theta) = -\frac{1}{\theta_0} \sum_{i=1}^{K} \log\left(\mathbb{E}\left\{\log_2(1+\rho_i)^{-\beta_i}\right\}\right), \tag{8}$$

which is obviously a convex function.

## III. EFFECTIVE CAPACITY MAXIMIZATION FOR RIS-AIDED MISO WIRELESS COMMUNICATIONS

For the considered system setup, we aim to jointly optimize the BS beamforming vector and RIS phase shift to maximize the total effective capacity while taking into account the QoS requirements. The heterogeneous QoS aware effective capacity maximization problem is formulated as follows

$$\begin{aligned}
\boldsymbol{P1}: \quad & \arg\max_{(\mathbf{W},\boldsymbol{\Phi})} \left\{-\frac{1}{\theta_0}\sum_{i=1}^{K}\log\left(\mathbb{E}\left\{\log_2(1+\rho_i)^{-\beta_i}\right\}\right)\right\} \\
\text{s.t.} \quad & C1: \operatorname{tr}\left(\mathbf{W}\mathbf{W}^H\right) \le P_{\mathrm{T}}, \\
& C2: |\phi_n| = 1,\ \forall n = 1,2,...,N,
\end{aligned}$$

where $P_{\mathrm{T}}$ is the total transmit power and $C2$ is the unit-modulus constraint imposed on the phase shifters.

To optimize the problem, we employ Zero-Forcing (ZF) beamforming at the BS, which can reduce or even eliminate the interference between users [12], [13]. Let us denote $\mathbf{P} = \operatorname{diag}[P_1, P_2, ..., P_K] \in \mathbb{C}^{K\times K}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K]^T \in \mathbb{C}^{K\times M}$ the power allocation matrix and the equivalent channel matrix, respectively. Here, $P_i$ denotes the transmit power for the $i$th user and $\mathbf{v}_i = \mathbf{h}_i\boldsymbol{\Phi}\mathbf{G} \in \mathbb{C}^{1\times M}$. Then, beamforming vector can be expressed as follows

$$\mathbf{W} = \mathbf{V}^H\left(\mathbf{V}\mathbf{V}^H\right)^{-1}\mathbf{P}^{\frac{1}{2}}. \tag{9}$$

Plugging (9) into the constraint $C1$, we can obtain

$$\operatorname{tr}\left(\mathbf{P}^{\frac{1}{2}}\mathbf{A}\mathbf{P}^{\frac{1}{2}}\right) \le P_{\mathrm{T}}, \tag{10}$$

where $\mathbf{A} \triangleq \left(\mathbf{V}\mathbf{V}^H\right)^{-1} \in \mathbb{C}^{K\times K}$.

In the ZF beamforming [13], we have $|\mathbf{v}_i\mathbf{w}_i| = \sqrt{P_i}$, and $|\mathbf{v}_i\mathbf{w}_k| = 0$, $\forall k \ne i$. Then, $\rho_i$ can be reduced to $\rho_i = P_i/\sigma_i^2$. Let us define $\gamma_i = 1/\sigma_i^2$. Thus, the transmit rate for the $i$th user can be shown as

$$R_i = B\log_2(1+P_i\gamma_i). \tag{11}$$

In addition, since all downlinks are independent of each other, problem $\boldsymbol{P1}$ can be equivalently transformed as

$$\begin{aligned}
\boldsymbol{P1'}: \quad & \arg\min_{(\mathbf{P},\boldsymbol{\Phi})} \left\{\mathbb{E}_{\gamma_i}\left[\prod_{i=1}^{K}(1+P_i\gamma_i)^{-\beta_i}\right]\right\} \\
\text{s.t.} \quad & C1: \operatorname{tr}\left(\mathbf{P}^{\frac{1}{2}}\mathbf{A}\mathbf{P}^{\frac{1}{2}}\right) \le P_{\mathrm{T}}, \\
& C2: |\phi_n| = 1,\ \forall n = 1,2,...,N.
\end{aligned}$$

To maximize the total effective capacity, we jointly optimize the transmit power allocation schemes at the BS and the phase shift of the RIS.

## IV. HETEROGENEOUS QOS AWARE JOINT OPTIMAL POWER CONTROL AND RIS PHASE SHIFTS FOR RIS-AIDED MISO WIRELESS COMMUNICATIONS

### *A. Optimization of Transmit Power Allocation* $\mathbf{P}$

To solve the total effective capacity maximization problem, we employ an alternating optimization approach, where we iteratively optimize the transmit power allocation schemes $\mathbf{P}$ and RIS reflecting phase shift $\boldsymbol{\Phi}$. In this section, we first optimize transmit power allocation scheme with fixed RIS reflecting phase shift. Then, problem $\boldsymbol{P1'}$ can be converted to problem $\boldsymbol{P2}$, which is described as follows

$$\begin{aligned}
\boldsymbol{P2}: \quad & \arg\min_{(\mathbf{P})} \left\{\mathbb{E}_{\gamma_i}\left[\prod_{i=1}^{K}(1+P_i\gamma_i)^{-\beta_i}\right]\right\} \\
\text{s.t.} \quad & C1: \sum_{i=1}^{K}\mathbb{E}\{P_iA_{ii}\} \le P_{\mathrm{T}}.
\end{aligned}$$

where $A_{ii}$ is the $i$th column and $i$th row element of Matrix $\mathbf{A}$. To simplify the subsequent derivations, we transform the constraint $C1$ into the expectation of transmit power. Problem $\boldsymbol{P2}$ is convex and has an optimal solution. To solve prblem $\boldsymbol{P2}$, we formulate the Lagrangian function of problem $\boldsymbol{P2}$, denoted by $L$, shown as follows

$$L = \mathbb{E}_{\gamma_i}\left\{\prod_{i=1}^{K}(1+P_i\gamma_i)^{-\beta_i}\right\} + \sum_{i=1}^{K}\lambda_i\left(\mathbb{E}\{P_iA_{ii}\} - P_{\mathrm{T}}\right), \tag{12}$$

where $\lambda_i (i \in [1, K])$ are Lagrangian multipliers. The optimal transmission power of the BS is denoted by $P^*$, and it can be obtained by differentiating (12) with respect to $P_i$.

The KKT conditions for problem $P2$ are given by

$$\left\{\begin{aligned}
\frac{\partial L}{\partial P_i} =& \mathbb{E}_{\gamma_i}\left\{-\beta_i(1+P_i\gamma_i)^{-\beta_i-1}\cdot\gamma_i\cdot\prod_{j=1,j\ne i}^{K}(1+P_j\gamma_j)^{-\beta_j}\right\} \\
&+\lambda_i\cdot\mathbb{E}\{A_{ii}\} = 0, \\
\frac{\partial L}{\partial \lambda_i} =& \ \mathbb{E}\{P_iA_{ii}\} - P_{\mathrm{T}} = 0.
\end{aligned}\right. \tag{13}$$

Solving (13), one can obtain

$$(1+P_i\gamma_i)^{-1} = \frac{\lambda_iA_{ii}}{\beta_i\gamma_i\left[\prod_{j=1}^{K}(1+P_j\gamma_j)^{-\beta_j}\right]}, \forall i \in [1, K]. \tag{14}$$

Multiplying the $K$ derivatives of (14), we have

$$\prod_{i=1}^{K}(1+P_i\gamma_i)^{-1} = \frac{\prod_{i=1}^{K}\lambda_iA_{ii}}{\prod_{i=1}^{K}\beta_i\gamma_i\left[\prod_{j=1}^{K}(1+P_j\gamma_j)^{-\beta_j}\right]^K}, \tag{15}$$

in which, after simplications, it yields

$$\prod_{j=1}^{K}(1+P_j\gamma_j)^{-\beta_i} = \prod_{j=1}^{K}\left[\frac{\lambda_iA_{ii}}{\beta_i\gamma_i}\right]^{\frac{-\beta_i}{-1-\beta_iK}}. \tag{16}$$

Superseding (16) into (14), one can attain the optimal transmit power from BS to $i$-th user, denoted by $P_i^*$, which is shown as follows

$$P_i^* = \frac{\beta_i}{\lambda_i A_{ii} \left[ \prod_{i=1}^{K} \left( \frac{\beta_i \gamma_i}{\lambda_i A_{ii}} \right)^{\frac{\beta_i}{1+\beta_i K}} \right]} - \frac{1}{\gamma_i}, \tag{17}$$

where $\lambda_i$ can be derived by plugging (17) into $\sum_{i=1}^{K} \mathbb{E}\left\{P_i A_{ii}\right\} - P_{\mathrm{T}} = 0$. Then, we find that $P_i^*$ is a function of $\theta_i$, $\gamma_i$ and $\varphi_i$, which indicates that the optimal transmit power is corresponding to the QoS requirements, CSI and RIS phase shifts. To thoroughly analyze the insights of optimal power allocation scheme, we give remarks 1 and 2, corresponding to the optimal power allocation schemes under the very stringent and the very loose QoS requirements, respectively.

*Remark 1:* When the QoS requirement of the $i$th downlink is very stringent ($\theta_i \to 0$), the optimal power allocation is reduced to

$$\lim_{\theta_i \to 0} P_i^* = \frac{\frac{\beta_i}{\lambda_i A_{ii}}}{\prod_{j=1, j\neq i}^{K} \left( \frac{\beta_j \gamma_j}{\lambda_j A_{jj}} \right)^{\frac{\beta_j}{1+\beta_j K}}} - \frac{1}{\gamma_i}, \tag{18}$$

which is the water-filling scheme for the RIS-aided wireless communications.

*Remark 2:* When the QoS requirement of the $i$th downlink is very loose ($\theta_i \to \infty$), the optimal power allocation is simplified to

$$\lim_{\theta_i \to \infty} P_i^* = \frac{\sigma_0}{\gamma_i}, \tag{19}$$

where $\sigma_0 = \prod_{j=1, j\neq i}^{K} \left( \frac{\beta_j \gamma_j}{\lambda_j A_{jj}} \right)^{-\frac{\beta_j}{1+\beta_j K}} - 1$. This is the channel inversion scheme for the RIS-aided wireless communications.

### *B. Optimization of RIS Phase Shift* $\mathbf{\Phi}$

Plugging optimal power allocation scheme, specified by (17), into the total effective capacity maximization problem $\boldsymbol{P1'}$, we can obtain problem $\boldsymbol{P3}$ as

$$\begin{aligned} \boldsymbol{P3}: \quad & \arg\min_{(\mathbf{\Phi})} \left\{ \mathbb{E}_{\gamma_i} \left[ \prod_{i=1}^{K} (1 + P_i^* \gamma_i)^{-\beta_i} \right] \right\} \\ \text{s.t.} \quad & C1: \mathrm{tr}\left( \mathbf{P}^{\frac{1}{2}} \mathbf{A} \mathbf{P}^{\frac{1}{2}} \right) \leq P_{\mathrm{T}}, \\ & C2: |\phi_n| = 1, \ \forall n = 1, 2, ..., N. \end{aligned}$$

In this Subsection, the transmit power matrix $\mathbf{P}$ is fixed. Therefore, the objective function of problem $\boldsymbol{P3}$ can be treated as a constant value objective function subject to the constraints $C1$ and $C2$. Thus, the optimization problem $\boldsymbol{P3}$ can be further converted to the total trasmission power minimization problem, denoted by $\boldsymbol{P3'}$, shown as follows

$$\begin{aligned} \boldsymbol{P3'}: & \min_{(\mathbf{\Phi})} \ \mathrm{tr}\left( \mathbf{P}^{\frac{1}{2}} \mathbf{A} \mathbf{P}^{\frac{1}{2}} \right) \overset{(a)}{=} \mathrm{tr}\left( \mathbf{V}^{+} \mathbf{P} \mathbf{V}^{+H} \right) \\ & \text{s.t.} \quad |\phi_n| = 1, \ \forall n = 1, 2, ..., N. \end{aligned}$$

**Algorithm 1: TEC maximization algorithm based on gradient descent method :**

```
1)  **Input:** $K, \mathbf{H}_0, \mathbf{G}, \boldsymbol{\theta}, B, \gamma, P_{\mathrm{T}}, \mu$, and $\varepsilon > 0$.
2)  **Initialization:** Random $\Phi$, $\mathbf{V} = \mathbf{H_0 \Phi G}$, $\mathbf{A} = \left(\mathbf{V}\mathbf{V}^H\right)^{-1}$, $\mathbf{d}^{(0)} = -\nabla_{\mathbf{\Phi}} \left( \left(\mathbf{f}^{(0)}\right)^H \mathbf{Q}\mathbf{f}^{(0)} \right)$, $\mathbf{f}^{(0)} = \mathrm{vec}\left(\mathbf{\Phi}^{-1}\right)$.
3)  while $|\widetilde{E}_C^{(s+1)} - \widetilde{E}_C^{(s)}|^2 > \varepsilon$, do
4)   Update $\mathbf{P}$ when $\mathbf{\Phi}$ is fixed:
5)      for $i = 1, 2, ...$ do
6)          Calculate Lagrangian Multipliers $\lambda_i$ based on equation $\sum_{i=1}^{K} \mathbb{E}\left\{P_i A_{ii}\right\} - P_{\mathrm{T}} = 0$.
7)      end for
8)      for $i = 1, 2, ...$ do
9)          Calculate $P_i^*$ based on Eq. (13).
10)     end for
11)  Update $\mathbf{\Phi}$ when $\mathbf{P}$ is fixed:
12)     for $g = 1, 2, ...$ do
13)         $\mathbf{f}_1^{(g)} = \mathbf{f}^{(g)} \circ e^{j\mu \mathbf{d}^{(g)}}$.
14)         $S = \left(\mathbf{f}_1^{(g)}\right)^H \mathbf{Q}\mathbf{f}_1^{(g)} - \left(\mathbf{f}^{(g)}\right)^H \mathbf{Q}\mathbf{f}^{(g)}$.
15)         if $S < \varepsilon$
16)             obtain $\varphi_i$, break.
17)         end if
18)         $\mathbf{f}^{(g)} = \mathbf{f}_1^{(g)}$.
19)         Gain $\mathbf{\Phi} = \mathbf{\Phi}^{(g+1)}$.
20)      end for
21) end while
22) Output: $\mathbf{\Phi}^*$ and $\mathbf{P}^*$.
```

In problem $\boldsymbol{P3'}$, $(a)$ holds by substituting $\mathbf{A} = \left(\mathbf{V}\mathbf{V}^H\right)^{-1}$ into the objective function of $\boldsymbol{P3'}$, $\mathbf{V}^{+}$ denotes pseudo-inverse of $\mathbf{V}$. Let us denote $\mathbf{H}_0 \triangleq \left[\mathbf{h}_1^T, \mathbf{h}_2^T, ..., \mathbf{h}_K^T\right]^T \in \mathbb{C}^{K \times N}$. Assume that the equivalent channel matrix $\mathbf{V}{=}\mathbf{H_0 \Phi G}$ has a right inverse. Additionally, problem $\boldsymbol{P3'}$ can be referred as an unconstrained problem. Thus, problem $\boldsymbol{P3'}$ can be rewritten as follows

$$\boldsymbol{P3''}: \min_{(\mathbf{\Phi})} \ F(\mathbf{\Phi}) \triangleq \mathrm{tr}\left( (\mathbf{H_0 \Phi G})^{+} \mathbf{P} (\mathbf{H_0 \Phi G})^{+H} \right).$$

For simplicity, The expression for problem $\boldsymbol{P3''}$ can be reformulated as

$$\begin{aligned} F(\mathbf{\Phi}) & \triangleq \mathrm{tr}\left( (\mathbf{H}_0 \mathbf{\Phi} \mathbf{G})^{+} \mathbf{P} (\mathbf{H}_0 \mathbf{\Phi} \mathbf{G})^{+H} \right) \\ & = \mathrm{tr}\left( \left(\mathbf{P}^{-\frac{1}{2}} \mathbf{H}_0 \mathbf{\Phi} \mathbf{G}\right)^{+} \left(\mathbf{P}^{-\frac{1}{2}} \mathbf{H}_0 \mathbf{\Phi} \mathbf{G}\right)^{+H} \right). \end{aligned} \tag{20}$$

Let us denote $\hat{\mathbf{H}} \triangleq \mathbf{P}^{-\frac{1}{2}} \mathbf{H_0}$. (20) can be reformulated as

$$F(\mathbf{\Phi}){=}\mathrm{tr}\left( \left(\hat{\mathbf{H}} \mathbf{\Phi} \mathbf{G}\right)^{+} \left(\hat{\mathbf{H}} \mathbf{\Phi} \mathbf{G}\right)^{+H} \right) \overset{(b)}{=} ||\mathbf{G}^{+} \mathbf{\Phi}^{-1} \hat{\mathbf{H}}^{+}||_{\mathrm{F}}^2, \tag{21}$$

where the equality $(b)$ holds based on the properties of the Frobenius norm and the laws associated with the pseudo-inverse and inverse of matrix products.

Then, to separate variable $\mathbf{\Phi}$ from $F(\mathbf{\Phi})$, we vectorize (21) as follows

$$\begin{aligned} F(\mathbf{\Phi}) & = ||\mathrm{vec}\left( \mathbf{G}^{+} \mathbf{\Phi}^{-1} \hat{\mathbf{H}}^{+} \right)||^2 \\ & = || \left( \hat{\mathbf{H}}^{+} \otimes \mathbf{G}^{+} \right) \mathrm{vec}\left( \mathbf{\Phi}^{-1} \right) ||^2 \\ & = \mathrm{vec}(\mathbf{\Phi}^{-1})^H \left( \hat{\mathbf{H}}^{+H} \otimes \mathbf{G}^{+} \right)^H \left( \hat{\mathbf{H}}^{+H} \otimes \mathbf{G}^{+} \right) \mathrm{vec}\left(\mathbf{\Phi}^{-1}\right). \end{aligned} \tag{22}$$

Since problem $\boldsymbol{P3''}$ is an unconstraint problem, the gradient decent approach can be employed to monotonically decrease its objective function until it converges to a stationary point. To clarify the gradient decent approach, we make the following replacements

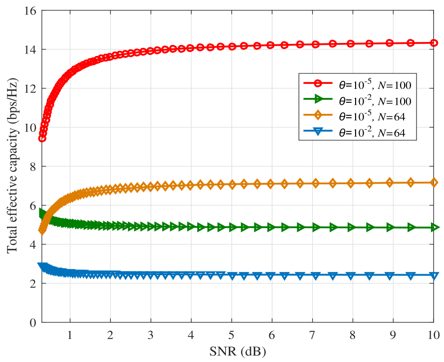

Fig. 2. TEC versus SNR.

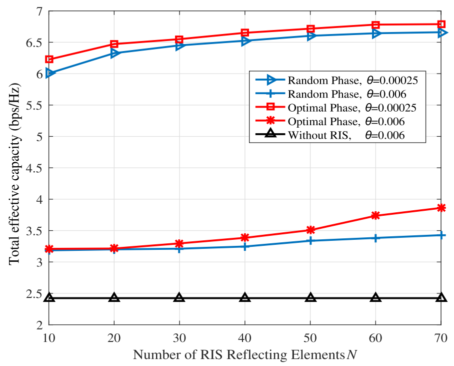

Fig. 3. TEC versus RIS element numbers.

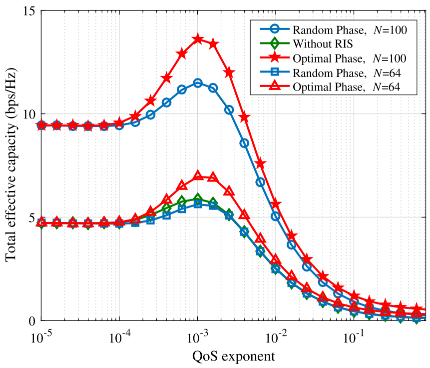

Fig. 4.TEC versus QoS exponent, $\gamma$=0.3 dB.

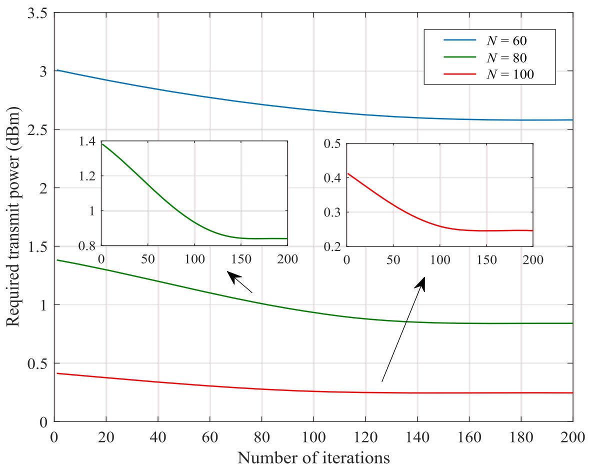

Fig. 5. Minimum transmission power of BS under increased number of iterations.

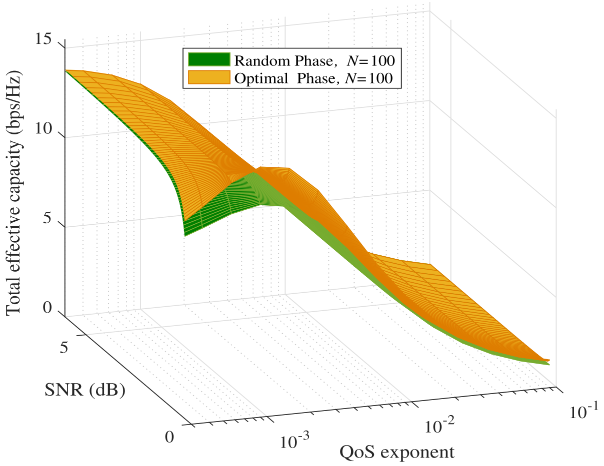

Fig. 6. Comparison with random phase scheme.

$$\begin{cases}\mathbf{Q} \triangleq \left(\hat{\mathbf{H}}^{+H} \otimes \mathbf{G}^{+}\right)^{H}\left(\hat{\mathbf{H}}^{+H} \otimes \mathbf{G}^{+}\right) \in \mathbb{C}^{N^{2}\times N^{2}},\\ \mathbf{f} \triangleq \mathrm{vec}\left(\mathbf{\Phi}^{-1}\right) \in \mathbb{C}^{N^{2}\times 1}.\end{cases} \tag{23}$$

Then, (20) can be rewritten as $\mathrm{F}\left(\mathbf{\Phi}\right) = \mathbf{f}^{H}\mathbf{Q}\mathbf{f}$. To calculate the gradient $\nabla_{\mathbf{\Phi}}\left(\mathbf{f}^{H}\mathbf{Q}\mathbf{f}\right)$ of $F\left(\mathbf{\Phi}\right)$ with regard to $\mathbf{\Phi}$, according to (23) we unfold $\mathbf{f}^{H}\mathbf{Q}\mathbf{f}$ as follows

$$\begin{aligned}\mathbf{f}^{H}\mathbf{Q}\mathbf{f} =& \sum_{a=1}^{N}\sum_{b=1}^{K}\sum_{c=1}^{M}\hat{H}_{b,a}\bar{\hat{H}}_{b,a}G_{c,a}\bar{G}_{c,a}\\ &+2\mathrm{Re}\left\{\sum_{a=1}^{N}\sum_{i>a}^{N}\sum_{b=1}^{K}\sum_{c=1}^{M}\hat{H}_{b,a}\bar{\hat{H}}_{b,i}G_{c,a}\bar{G}_{c,i}e^{j\left(\varphi_{a}-\varphi_{i}\right)}\right\}.\end{aligned} \tag{24}$$

Next, we take the partial derivatives of (24) pertaining to $\varphi_{i}\left(\forall i = 1, 2, ..., N\right)$ can be written as follows

$$\begin{aligned}\frac{\partial\left(\mathbf{f}^{H}\mathbf{Q}\mathbf{f}\right)}{\partial\varphi_{i}} =& 2\mathrm{Re}\left\{je^{j\varphi_{i}}\sum_{m>i}^{N}\sum_{b=1}^{K}\sum_{c=1}^{M}\bar{\hat{H}}_{b,i}H_{b,m}\bar{G}_{c,i}G_{c,m}e^{-j\varphi_{m}}\right.\\ &\left.-je^{-j\varphi_{i}}\sum_{a<i}^{N}\sum_{b=1}^{K}\sum_{c=1}^{M}\bar{\hat{H}}_{b,a}H_{b,i}\bar{G}_{c,a}G_{c,i}e^{j\varphi_{a}}\right\}.\end{aligned}$$

Then, each update can be performed according to the rules shown below

$$\begin{cases}\mathrm{vec}\left(\mathbf{\Psi}\right)^{(g+1)} = \mathrm{vec}\left(\mathbf{\Psi}\right)^{(g)} + \mu\mathbf{d}^{(g)},\\ \mathbf{f}^{(g+1)} = e^{j\mathrm{vec}\left(\mathbf{\Psi}\right)^{(g+1)}}\circ\mathrm{vec}\left(\mathbf{I}_{N}\right) = \mathbf{f}^{(g)}\circ e^{j\mu\mathbf{d}^{(g)}},\end{cases} \tag{25}$$

wherein $\mathrm{vec}\left(\mathbf{\Psi}\right)^{(g)}$ $\left(\mathbf{\Psi} = \mathrm{diag}\left[\varphi_{1}, \varphi_{2}, ..., \varphi_{N}\right]\right)$, $\mathbf{d}^{g}$ and $\mu$ represent the phase of $\mathbf{f}$ at repetition $g$, the descending direction and step size of the iteratively updated $g$, respectively.

The TEC maximization algorithm based on gradient descent method is given in Algorithm 1.

The complexity of step 9 is much less complex than multiplication. Its complexity is negligible. Thus, the computational complexity of the Algorithm hinges on iteration numbers, and the asymptotic sophistication of Algorithm 1 is $\mathcal{O}\left(I_{\mathrm{A}}\left(MK\left(N^{2}+N\right)\right)\right)$, where $I_{\mathrm{A}}$ is the number of iterations of the alternating.

## V. NUMERICAL RESULTS

To evaluate the performance of the proposed optimal schemes, we conduct numerical simulations in this section. We set the location of BS to be (0 m, 0 m) and the distance from BS to RIS to be 200 m. The BS is deployed with 4 antennas ($M = 4$). There are 4 single-antenna users ($K = 4$) distributing randomly in a 10-meter radius circle centering at (200 m, 30 m). Besides, we set the bandwidth $B$ = 10 MHz, the Nakagami-m parameter $m = 2$. According to the 3GPP propagation environment, the path-loss $\mathbf{h}_i$ and $\mathbf{G}$ are set to be $35.26 + 22.0 \lg d$ (dB).

Fig. 2 compares the perfomance of the proposed statistical QoS aware joint optimal schemes under the relatively stringent QoS demand (i.e., $\theta = 10^{-2}$) and the relatively loose QoS requirement (i.e., $\theta = 10^{-5}$). As illustrated in Fig. 2, it is found that under the relatively loose QoS requirement, the TEC increases with the increasing of SNR. On the contrary, under the relatively stringent QoS requirement, the TEC diminishes when SNR increases. The simulation results verify the theoretical analyses for our proposed optimal power allocation schemes. It can also be observed that an RIS with a large number of reflection elements can achieve better performance than one with a small number of reflecting elements.

Fig. 3 shows the relationship between the TEC and the number of RIS reflection elements. It is clear that the TEC increases as the number of RIS reflection elements increases. Under the same QoS requirement, optimal phase scheme has a larger TEC compared to the random phase scheme and the scheme without phase shifting. In Fig. 4, SNR is fixed to 0.3 dB. Obviously, it can be seen that the advanced joint optimal schemes can acquire a better property than the random phase shift schemes and the system without RIS. The scheme without RIS represents a direct link between the BS and the users. Due to the shorter distance between the BS and the users, the path loss is smaller compared to the indirect link involving the BS-RIS-users. Therefore, when the QoS index is moderate (i.e., $10^{-4} \sim 10^{-2}$), the performance of the system without RIS is higher than that of the random phase-shift scheme when RIS Number of reflectors $N = 64$. In addition, the TEC is first increasing and then decreasing as the QoS exponent increases. In [14], [15], the authors also observed a similar phenomenon in the simulation results. We conducted a preliminary analysis and identified that it could be attributed to the different values of the Nakagami-m parameter and the rate at which the transmit power of the BS increases. However, it is important to note that this aspect is not the focus of the current paper and does not affect the main conclusions. Further investigation will be conducted to explore the underlying reasons and draw corresponding conclusions.

In Fig. 5, we plot the relationship between the minimum transmitting power required by the BS and the number of iterations when the number of RIS reflection elements $N$ is 60, 80 and 100. It can be seen that as the number of iterations increases, the required transmist power of BS gradually decreases and the trend slows down, eventually reaching a convergence value. Additionally, it can be observed that the transmit power required by BS $N = 100$ is lower than that required when $N = 60, 80$. This further confirms the potential of RIS to improve the performance of wireless communication systems.

Fig. 6 depicts the TEC versus the statistical QoS exponent and SNR. In this simulation, the RIS element number is set to be 100. As depicted in Fig. 6, It can be seen that our developed schemes exhibit a higher TEC compared to those with random phase shifts. Furthermore, it can be observed that the difference between our proposed optimal schemes and the random phase shift is more significant when the QoS exponent falls within the moderately strict or moderately lenient region.

## VI. CONCLUSIONS

This paper considered the RIS aided wireless communications with herogeneous statistical QoS provisionings. We developed the joint optimal power allocation and phase shift scheme to maximize the total effective capacity. Firstly, we designed the RIS beamforming and heterogeneous statistical QoS mechanism separately. Then, we established the total effective capacity maximization problem subject to the total transmission power restraint. Solving this problem, we obtained the joint optimal schemes by employing KKT conditions and gradient iteration method. Simulation results validated the better performance of the proposed heterogeneous QoS aware joint optimal schemes for RIS aided wireless communciations.

## REFERENCES

[1] B. Rana, S.-S. Cho, and I.-P. Hong, “Review Paper on Hardware of Reconfigurable Intelligent Surfaces,” *IEEE Access*, vol. 11, pp. 29 614–29 634, Mar. 2023.

[2] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, “Reconfigurable Intelligent Surfaces: Principles and Opportunities,” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1546–1577, Aug. 2021.

[3] Q. Wu and R. Zhang, “Beamforming Optimization for Wireless Network Aided by Intelligent Reflecting Surface With Discrete Phase Shifts,” *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[4] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, “Wireless Communications Through Reconfigurable Intelligent Surfaces,” *IEEE Access*, vol. 7, pp. 116 753–116 773, Aug. 2019.

[5] D. Wu and R. Negi, “Effective Capacity: A Wireless Link Model for Support of Quality of Service,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, Jun. 2003.

[6] L. Musavian and Q. Ni, “Effective Capacity Maximization With Statistical Delay and Effective Energy Efficiency Requirements,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3824–3835, Jul. 2015.

[7] L. Kong, Y. Ai, S. Chatzinotas, and B. Ottersten, “Effective Rate Evaluation of RIS-Assisted Communications Using the Sums of Cascaded $\alpha - \mu$ Random Variates,” *IEEE Access*, vol. 9, pp. 5832–5844, Jan. 2021.

[8] B. E. Tegicho and T. E. Bogale, “RIS Assisted MIMO UAV Swarm Communication: Channel Capacity Margin Analysis,” in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, Dec. 2022, pp. 0347–0351.

[9] Q. Zhang, Y.-C. Liang, and H. V. Poor, “Reconfigurable Intelligent Surface Assisted MIMO Symbiotic Radio Networks,” *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4832–4846, Jul. 2021.

[10] C.-S. Chang, “Stability, Queue Length, and Delay of Deterministic and Stochastic Queueing Networks,” *IEEE Trans. Autom. Control.*, vol. 39, no. 5, pp. 913–931, May. 1994.

[11] W. Cheng, X. Zhang, and H. Zhang, “Heterogeneous Statistical QoS Provisioning for Downlink Transmissions over Mobile Wireless Cellular Networks,” in *2014 IEEE Global Communications Conference*, Austin, TX, USA, Dec. 2014, pp. 4622–4628.

[12] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[13] C. Peel, B. Hochwald, and A. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization,” *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[14] L. Musavian and Q. Ni, “Effective Capacity Maximization With Statistical Delay and Effective Energy Efficiency Requirements,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3824–3835, Jul. 2015.

[15] W. Cheng, X. Zhang, and H. Zhang, “Optimal Power Allocation With Statistical QoS Provisioning for D2D and Cellular Communications Over Underlaying Wireless Networks,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 151–162, Jan. 2016.