



Resource Allocation and MEAR Maximization for RIS-Aided eMBB/URLLC Traffic Multiplexing

Mengmeng Wang^{1(✉)}, Bei Liu^{2(✉)}, and Xin Su^{3(✉)}

¹ School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
wmmjust@163.com

² Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China
liubei@tsinghua.edu.cn

³ Department of Electronic Engineering, Tsinghua University, Beijing, China
suxin@tsinghua.edu.cn

Abstract. In future sixth generation (6G) cellular networks, service requirements for enhanced mobile broadband (eMBB) and ultra-reliable and low latency Communication (URLLC) will become more stringent, resulting in severe power consumption and spectrum scarcity. This paper studies the resource allocation problem in the case of Reconfigurable Intelligent Surface (RIS) assisted coexistence of eMBB and URLLC services, to maximize the eMBB user's minimum expected achieved rate (MEAR) while ensuring the quality of services (QoS) for URLLC users. We adopt the alternating direction multiplier method (ADMM) to optimize the RIS phase shift matrix and propose a heuristic algorithm for allocating eMBB resource blocks (RB). In addition, a proportional fairness (PF) algorithm is used to allocate URLLC services to ensure fairness among eMBB users. Compared to no RIS, simulation results show that the proposed scheme improves approximately 20% MEAR, requiring only 60 RIS elements, and it has significant advantages over other schemes.

Keywords: eMBB · MEAR · fairness · URLLC · RIS · RB allocation

1 Introduction

The services supported by the fifth-generation (5G) new radio are mainly divided into three categories, namely, enhancing mobile broadband (eMBB), ultra-reliable and low-latency communications (URLLC), and massive machine-type communications (mMTC). eMBB is a human-centered application scenario, which is characterized by ultra-high data transmission rate, wide coverage, and mobility guarantee. The URLLC not only requires a link delay of 1 ms but also supports high-reliability connection under high-speed mobile conditions. The

mMTC has a strong linkability, which can quickly promote the deep integration of various vertical industries (smart cities, intelligent furniture) [1]. These applications in future B5G and 6G networks are expected to support higher transmission rates, stricter latency, and reliability requirements [2].

Recently, More and more scholars begin to pay attention to how eMBB and URLLC allocate resources more effectively under the same physical resource block. The author in [3] proposed a coexistence mechanism of eMBB and URLLC to maximize the MEAR of eMBB users. The literature [4] adopted one-to-one matching superposition and puncturing technology to allocate URLLC resources, which greatly improves the MEAR of eMBB compared with single puncturing technology. [5]proposed a low-complexity algorithm for many-to-many pairing to increase the number of services for URLLC users and reduce the impact on eMBB rates. In [6], the authors proposed a modern risk-averse approach in portfolios that considers eMBB data rate variance to reduce the impact of immediately scheduled URLLC traffic on eMBB reliability. The authors of [7] introduced a risk-sensitive approach to protect low data rate eMBB users from puncturing. [8] introduced the same risk-sensitive scheme as literature [7] to study resource allocation under eMBB and URLLC coexistence under M/G/1 queuing model, which improved the fairness and throughput among eMBB users. [9] investigated the coexistence mechanism of eMBB and URLLC in MIMO-NOMA systems, introducing an eMBB user clustering mechanism to balance the throughput and computational complexity of the system. [10]considered the sub-carrier and power allocation issues in distributed MIMO system and improved the energy efficiency of eMBB and URLLC services by selecting orthogonal and non-orthogonal scheduling schemes. The Paper [11] proposed a hierarchical deep learning framework to address the dynamics of eMBB and URLLC services. The above literatures [3–11] considered the maximum eMBB and URLLC utility with limited resources at the base station (BS). However, as the quality of service for 6G applications increases, the applications will require more energy and spectrum resources. Thus, a new framework for RIS-assisted resource allocation in eMBB and URLLC coexistence scenarios was proposed in [12]. The eMBB sum rate can be increased to 70% gain. The RIS-assisted wireless communication networks have low hardware costs and energy consumption. The RIS has the advantages of small size and low cost, and can be flexibly deployed on the walls of buildings, light beads, etc [13].

As far as we know, most of the work in the above literatures focus on minimizing the impact on the eMBB users rate while satisfying the quality of service of URLLC, ignoring eMBB inter-user fairness and MEAR, and eMBB users in urban areas can suffer from poor direct path channel gain due to occlusions, which affects the user's service quality. However, the RIS improves the channel conditions of the wireless communication environment, so it is necessary for us to study the maximization of the MEAR and the guarantee of inter-user fairness in the case of RIS-assisted coexistence of eMBB and URLLC. Our main contributions are:

- 1) We introduce RIS into a new framework for resource allocation for eMBB and URLLC services with the optimization objective of maximizing eMBB users MEAR.
- 2) In order to solve the optimization problem efficiently, we divide it into two sub-problems to solve: the resource allocation phase for eMBB users and the resource allocation phase for URLLC users.
- 3) Firstly, for the eMBB resource scheduling phase, the ADMM algorithm is used to optimize the phase shift matrix of RIS to improve the channel gain for eMBB users. Secondly, we propose an efficient heuristic allocation algorithm to improve throughput and MEAR. In addition, to satisfy the URLLC users Qos and not to destroy the fairness of eMBB users, a proportional fairness (PF) algorithm is applied to allocate resources for URLLC users.
- 4) Finally, we simulate and analyse the proposed allocation scheme and compare it with the heuristic eMBB allocation scheme for maximizing MEAR [3] and the maximizing sum rate eMBB allocation scheme [12]. For a fair wireless communication environment, the comparison schemes also add RIS. the benefits of adding RIS and the advantages of the proposed scheme can be demonstrated.

The remaining sections are organized as follows. Section 2 introduces the system model and problem formulation. The solution to the above problem is presented in Sect. 3. Simulation results and discussion of the optimization problem are given in Sect. 4. Section 5 provides a summary of this paper.

2 System Model and Problem Formulation

2.1 System Model

As shown in Fig. 1, we consider the scenario of downlink radio access network transmission, where a BS serves a set of eMBB users and URLLC users randomly distributed in the coverage space, and the RIS supports only eMBB traffic. The set of eMBB users and the set of URLLC users can be denoted respectively as $E = \{1, 2, \dots, e\}$ and $U = \{1, 2, \dots, u\}$, The total bandwidth is split into $B = \{1, 2, \dots, b\}$ resource blocks (RB) with equal bandwidth f_b . A RB occupies 12 sub-carrier in frequency. The time domain is divided into time slots of the same time length Δ . The time slots is denoted as $T = \{1, 2, \dots, t\}$, To meet the latency requirements of URLLC traffic, Each time slot is divided into mini-slots denoted by $M = \{1, 2, \dots, m\}$. The URLLC traffic arrival each mini-slot follows Gaussian distribution, i.e., $U \sim \mathcal{N}(\lambda, \sigma^2)$, λ is the mean of U , σ^2 represents the variance of U

The RIS consists of N reflecting elements and a micro-controller to adjust the phase shift, and it dynamically adjusts the channel propagation environment between the BS and the eMBB users. The phase shift matrix of RIS is represented by $\Phi = \text{diag}[\phi_1, \phi_2, \dots, \phi_N] \in \mathbb{C}^{N \times N}$, where for all $n \in N$, $\phi_n = e^{j\theta_n}$, $\theta_n \in [0, 2\pi)$ denotes the phase shift of n th reflective element. We assume that the channel state information (CSI) between the BS to all users, the BS to the RIS,

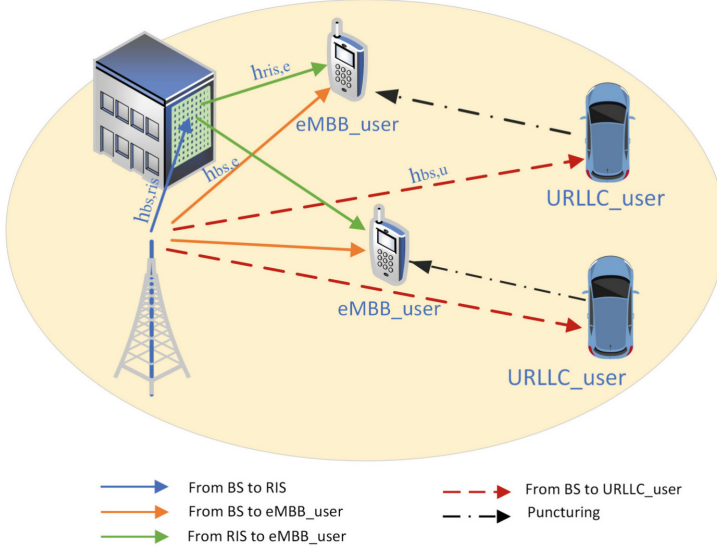


Fig. 1. System Model

and the RIS to the eMBB users is completely known. The channel coefficient between the BS and RIS is denoted by the channel vector $\mathbf{G} \in \mathbb{C}^{N \times 1}$, In addition, Let $\mathbf{h}_{ris,e} \in \mathbb{C}^{N \times 1}$ and $\mathbf{h}_{bs,e} \in \mathbb{C}^{1 \times 1}$ denote the channel vectors from the RIS to the e th eMBB user and from the BS to the e th eMBB user for all $e \in E$, respectively.

For all $e \in E$, The achieved data rate in [bps/Hz] of e th eMBB user at time slot t can be approximated as

$$R_e(t) = \sum_{b \in B} a_e^b(t) \log_2 \left(1 + \frac{P |h_{bs,e} + \mathbf{h}_{ris,e}^H \Phi \mathbf{G}|^2}{\mu^2} \right) \quad (1)$$

where P is transmission power from the BS to each RB, $\mu^2 = f_b \times N_0$ is channel noise power, N_0 is noise spectral density, $\alpha_e^b(t)$ is the eMBB user scheduling indicator at any time slot t as follows:

$$\alpha_e^b(t) = \begin{cases} 1, & \text{if RB } b \text{ is allocated for } e \in E \text{ at } t \in T \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Since URLLC packet is very small, the Shannon's capacity formulation can not be applied to calculate URLLC users data rate, The achievable data rate of a URLLC user over a RB b can be given by [17, 18] :

$$R_{u,b}^m(t) = \log_2 \left(1 + \frac{|h_u|^2 P}{N_0 f_b} \right) - \sqrt{\frac{V_u}{N_u}} Q^{-1}(\varepsilon) \quad (3)$$

where h_u is the channel gain of URLLC user u , $V_u = \left(1 - \left(1 + \frac{|h_u|^2 P}{N_0 f_b}\right)^{-2}\right)$ is the channel dispersion, $Q(\cdot)^{-1}$ is the inverse of $Q(x) = \int_x^\infty \frac{e^{-t^2}}{\sqrt{2\pi}} dt$, N_u is the finite blocklength of u . when a URLLC user requires multiple RBs, its achievable data rate looks like as follows:

$$R_u^m(t) = \sum_{b \in B} \gamma_{e,b}^m(t) R_{u,b}^m(t) \quad (4)$$

where $\gamma_{u,b}^m(t)$ is the URLLC user scheduling indicator at $m \in M$ of any time slot t as follows:

$$\gamma_{u,b}^m(t) = \begin{cases} 1, & \text{if RB } b \text{ is allocated for } u \in U \text{ at } m \in M \text{ of } t \in T \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The reliability constraint for the URLLC service is to ensure that its probability of disruption is less than or equal to a small positive number ϵ .

$$P \left[\sum_{u \in U} I_u^m(t) \leq U \right] \leq \epsilon \quad (6)$$

where $I_u^m(t)$ is the indicator for the serving URLLC users as follows:

$$I_u^m(t) = \begin{cases} 1, & \text{if } u \in U \text{ is served for at } m \in M \text{ of } t \in T \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The payload $\xi_u^m(t)$ of URLLC user $u \in U$ needs to be transferred within the stipulated period δ , it needs to satisfy the following:

$$I_u^m(t) \xi_u^m(t) \leq \delta R_u^m(t) \quad (8)$$

The eMBB rate loss is proportional to the number of URLLC punctures, so we adopt the linear rate loss model from the literature [14]. Therefore, the actual achievable rate of eMBB user $e \in E$ at time slot t is as follows.

$$R_e^{actual}(t) = \sum_{b \in B} \alpha_e^b(t) \left(1 - \sum_{m \in M} \sum_{u \in U} \mathbb{I}(\alpha_e^b(t) = \gamma_{u,b}^m(t))\right) \log_2 \left(1 + \frac{P|h_{bs,e} + \mathbf{h}_{ris,e}^H \Phi \mathbf{G}|^2}{\mu^2}\right) \quad (9)$$

The average actual achievable rate for the eMBB user e is expressed as:

$$R_e^{mactual}(t) = \frac{1}{|T|} \sum_{t=1}^T R_e^{actual}(t) \quad (10)$$

The MEAR among eMBB users is represented as follows:

$$R_e^{Mactual}(t) = \min_{e \in E} (R_e^{mactual}(t)) \quad (11)$$

Accordingly, the RIS-assisted eMBB/URLLC joint phase shift matrix and resource allocation problems are as follows:

$$\max_{\alpha, \gamma, \Phi} R_e^{Mactual}(t) \quad (12)$$

$$\text{s.t. } 0 \leq \theta_n < 2\pi, \forall n \in N, \quad (12a)$$

$$\sum_{e \in E} \alpha_e^b(t) \leq 1, \forall b \in B, \forall t \in T, \quad (12b)$$

$$\sum_{u \in U} \gamma_{u,b}^m(t) \leq 1, \forall b \in B, \forall m \in M, \quad (12c)$$

$$\sum_{b \in B} \alpha_e^b(t) \geq 1, \forall e \in E, \forall t \in T, \quad (12d)$$

$$\sum_{b \in B} I_u^m(t) \gamma_{u,b}^m(t) \geq 1, \forall u \in U, \forall m \in M, \forall t \in T, \quad (12e)$$

$$I_u^m(t) \xi_u^m(t) \leq \delta R_u^m(t), \forall u \in U, \forall m \in M, \forall t \in T, \quad (12f)$$

$$P \left[\sum_{b \in B} I_u^m(t) < U \right] \leq \epsilon, \forall m \in M, \forall t \in T, \quad (12g)$$

$$\sum_{e \in E} \sum_{b \in B} \alpha_e^b(t) + \sum_{u \in U} \sum_{b \in B} I_u^m(t) \gamma_{u,b}^m(t) \leq |B|, \forall m \in M, \forall t \in T, \quad (12h)$$

$$\alpha_e^b(t), \gamma_{e,b}^m(t), I_u^m(t) \in \{0, 1\}, \forall e \in E, \forall b \in B, \forall u \in U, \forall m \in M, \forall t \in T. \quad (12i)$$

In (12), the phase shift constrains of RIS is preserved by (12a). The orthogonality of RBs between eMBB and URLLC UEs is denoted by constraints (12b) and (12c), respectively. Constraints (12d) and (12e) indicate that eMBB and URLLC users require at least one RB. constraints (12f) and (12g) denote the latency and reliability limits of URLLC respectively. The total RBs limitation is presented by constrain (12h). Constrain (12i) denotes that per parameter of α , γ and I are binary. The above optimization problem is divided into two parts, eMBB resource allocation and URLLC resource scheduling, to be solved separately.

3 Solution Approach

3.1 eMBB Resource Allocation

We first consider the problem \mathcal{P}_1 of eMBB resource allocation. Due to \mathcal{P}_1 is a mixed-integer nonlinear programming (MINLP) problem, it can't be solved directly and need to converted to convex form. We divide the problem into the optimization of phase shift matrix $\mathcal{P}_{1.1}$ and the allocation of eMBB RBs \mathcal{P}_2 .

$$\mathcal{P}_1 \quad \max_{\alpha, \Phi} R_e^{Mactual}(t) \quad (13)$$

$$\text{s.t.} \quad \sum_{e \in E} \sum_{b \in B} \alpha_e^b(t) \leq |B|, \forall t \in T \quad (13a)$$

$$\sum_{e \in E} \alpha_e^b(t) \leq 1, \forall b \in B, \forall t \in T, \quad (13b)$$

$$\sum_{b \in B} \alpha_e^b(t) \geq 1, \forall e \in E, \forall t \in T, \quad (13c)$$

$$\alpha_e^b(t) \in \{0, 1\}, \forall e \in E, \forall b \in B, \forall t \in T, \quad (13d)$$

$$0 \leq \theta_n < 2\pi, \forall n \in N \quad (13e)$$

For a fixed RB, \mathcal{P}_1 reduces to a phase shift matrix optimization problem which can be denoted as:

$$\mathcal{P}_{1.1} \quad \max_{\Phi} R_e^{Mactual}(t) \quad (14)$$

$$\text{s.t.} \quad 0 \leq \theta_n < 2\pi, \forall n = 1, \dots, N \quad (14a)$$

Let us define a vector denoted by $\mathbf{x}^H = [x_1, x_2, \dots, x_N]^H$, where $x_n = e^{j\theta_n}$. We get $|h_{bs,e} + \mathbf{h}_{ris,e}^H \Phi \mathbf{G}|^2 = |h_{bs,e} + \mathbf{x}^H \Psi|^2$ by variable conversion, where $\Psi = \text{diag}(\mathbf{h}_{ris,e}^H) \mathbf{G}$, and $\Psi \in \mathbb{C}^{N \times 1}$. Thus the channel gain between the BS and eMBB user e can be rewritten as:

$$|h_{bs,e} + \mathbf{x}^H \Psi|^2 = \mathbf{x}^H \Psi \Psi^H \mathbf{x} + h_{bs,e} \Psi^H \mathbf{x} + \mathbf{x}^H \Psi h_{bs,e}^\dagger + |h_{bs,e}|^2 \quad (15)$$

where h^\dagger is the complex conjugate of h . through the introduction of the auxiliary variable ζ , $\mathcal{P}_{1.1}$ can be expressed as :

$$\mathcal{P}_{1.2} \quad \max_{\bar{\mathbf{x}}} R_e^{Mactual}(t) \quad (16)$$

$$\text{s.t.} \quad |\bar{\mathbf{x}}_n| = 1, \forall n = 1, \dots, N + 1 \quad (16a)$$

where channel gain of the objective function is $(\bar{\mathbf{x}}^H \Omega \bar{\mathbf{x}} + |h_{bs,e}|^2)$, and Ω can be written as follows:

$$\Omega = \begin{bmatrix} \Psi \Psi^H & \Psi h_{bs,e}^\dagger \\ h_{bs,e} \Psi^H & 0 \end{bmatrix}, \bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \zeta \end{bmatrix} \quad (17)$$

To maximize the objective function in problem $\mathcal{P}_{1.2}$, it is necessary to maximize $\bar{\mathbf{x}}$. Therefore, The optimization problem can be simplified as:

$$\mathcal{P}_{1.3} \quad \max_{\bar{\mathbf{x}}} \quad \bar{\mathbf{x}}^H \boldsymbol{\Omega} \bar{\mathbf{x}} \quad (18)$$

$$\text{s.t.} \quad |\bar{\mathbf{x}}_n| = 1, \forall n = 1, \dots, N + 1 \quad (18a)$$

In order to solve the constant modulus constraint problem, we adopt the ADMM [15] to obtain the optimal solution. The ADMM algorithm optimises the phase shift matrix as detailed in Algorithm 1. $\mathcal{P}_{1.3}$ need to organize into the ADMM optimization form as follows

$$\mathcal{P}_{1.4} \quad \max_{\bar{\mathbf{x}}} \quad \bar{\mathbf{x}}^H \boldsymbol{\Omega} \bar{\mathbf{x}} \quad (19)$$

$$\text{s.t.} \quad |\varphi_n| = 1, \forall n = 1, \dots, N + 1 \quad (19a)$$

$$\bar{\mathbf{x}} = \boldsymbol{\varphi} \quad (19b)$$

The scaled augmented Lagrangian function is written as:

$$L_p(\bar{\mathbf{x}}, \boldsymbol{\varphi}, \boldsymbol{\nu}) = \bar{\mathbf{x}}^H \boldsymbol{\Omega} \bar{\mathbf{x}} + \frac{\rho}{2} \left(\|\bar{\mathbf{x}} - \boldsymbol{\varphi} + \boldsymbol{\nu}\|^2 - \|\boldsymbol{\nu}\|^2 \right) \quad (20)$$

where $\rho > 0$ is a penalty factor. The update iteration form of ADMM is as follows:

$$\bar{\mathbf{x}}^{k+1} = \arg \min_{\bar{\mathbf{x}}} L_p(\bar{\mathbf{x}}, \boldsymbol{\varphi}^k, \boldsymbol{\nu}^k) \quad (21)$$

$$\boldsymbol{\varphi}^{k+1} = \arg \min_{\boldsymbol{\varphi}} L_p(\bar{\mathbf{x}}^{k+1}, \boldsymbol{\varphi}, \boldsymbol{\nu}^k) \quad (22)$$

$$\boldsymbol{\nu}^{k+1} = \bar{\mathbf{x}}^{k+1} - \boldsymbol{\varphi}^{k+1} + \boldsymbol{\nu}^k \quad (23)$$

Based on (20), subproblem (21) is equivalent to solve $\bar{\mathbf{x}}^{k+1} = \arg \min_{\bar{\mathbf{x}}} \bar{\mathbf{x}}^H \boldsymbol{\Omega} \bar{\mathbf{x}} + \frac{\rho}{2} \|\bar{\mathbf{x}} - \boldsymbol{\varphi}^k + \boldsymbol{\nu}^k\|^2$. We can achieve the solution as follows.

$$\bar{\mathbf{x}}^{k+1} = (2\boldsymbol{\Omega} + \rho\mathbf{I})^{-1} (\boldsymbol{\varphi}^k - \boldsymbol{\nu}^k) \quad (24)$$

Based on (20), subproblem (22) is equivalent to solve $\boldsymbol{\varphi}^{k+1} = \arg \min_{\boldsymbol{\varphi}} \frac{\rho}{2} \|\bar{\mathbf{x}}^{k+1} - \boldsymbol{\varphi} + \boldsymbol{\nu}^k\|^2$. We can achieve the solution as follows.

$$\boldsymbol{\varphi}^{k+1} = \boldsymbol{\nu}^k + \bar{\mathbf{x}}^{k+1} \quad (25)$$

Algorithm 1. ADMM Algorithm

- 1: **Input:** $\rho, \bar{\mathbf{x}}^0, \boldsymbol{\varphi}^0, \boldsymbol{\nu}^0, k = 0$
 - 2: While stopping condition is not satisfied
 - 3: Compute $\bar{\mathbf{x}}^{k+1}$ by (24)
 - 4: Compute $\boldsymbol{\varphi}^{k+1}$ by (25)
 - 5: Update $\boldsymbol{\nu}^{k+1}$ by (23)
 - 6: Update $k \leftarrow k + 1$
 - 7: **Output:** optimal $\bar{\mathbf{x}}$
-

For a feasible phase-shift solution, the RBs allocation problem is formulated as:

$$\mathcal{P}_2 \quad \max_{\alpha} \quad R_e^{Mactual}(t), \quad (26)$$

$$\text{s.t.} \quad \sum_{e \in E} \sum_{b \in B} \alpha_e^b(t) \leq |B|, \forall t \in T \quad (26a)$$

$$\sum_{e \in E} \alpha_e^b(t) \leq 1, \forall b \in B, \forall t \in T, \quad (26b)$$

$$\sum_{b \in B} \alpha_e^b(t) \geq 1, \forall e \in E, \forall t \in T, \quad (26c)$$

$$\alpha_e^b(t) \in \{0, 1\}, \forall e \in E, \forall b \in B, \forall t \in T, \quad (26d)$$

The optimization problem \mathcal{P}_2 is a non-convex problem due to the binary parameter α . In order to transfer more RBs in a limited time slot t , we propose a low-complexity heuristic algorithm for allocating RBs to eMBB users. In order to improve the MEAR of eMBB users, the eMBB user with poor channel condition needs more RBs. By optimizing the phase shift matrix of the worst eMBB user channel conditions, the number of RBs deprived from users with better channel conditions can be reduced, thus improving the transmission rate of the whole system. To make up for the rate loss suffered by eMBB users before, we consider both the average achievable rate of eMBB users in the past time slots and the channel conditions of the current time slot. The priority for RBs is given to eMBB users with poor channel conditions and low average achievable data rates. The specific details of eMBB RBs allocation are shown in Algorithm 2.

Algorithm 2. eMBB Resource Allocation

- 1: **Initialization:** $i, j, l, i = 1$
 - 2: Select eMBB users with the worst channel conditions to optimize the RIS phase shift matrix by Algorithm 1
 - 3: **while** $i \leq |T|$ **do**
 - 4: $E(e) \leftarrow RBs = 1$ and calculate $R_e^t, \forall e \in E$
 - 5: $rRBs = |B| - |E|$
 - 6: **for** $j = 1 : rRBs$ **do**
 - 7: **for** $l = 1 : |E|$ **do**
 - 8: $E(l) \leftarrow H = R_l^{mactual}(t-1) + R_l(t)$
 - 9: **end for**
 - 10: $id = \text{minArgc}(E(\cdot) \rightarrow H)$
 - 11: $E(id) \rightarrow RBs + 1$
 - 12: **end for**
 - 13: Update $i \leftarrow i + 1$
 - 14: **end while**
-

The interpretation of Algorithm 2 is as follows:

- (1) Initialization related parameters.
- (2) The eMBB user with the worst channel conditions is selected for optimization of the RIS phase shift matrix.
- (3) Allocate a RB to all eMBB users in order to calculate the achievable data rate for the current channel conditions
- (4) Calculate the sum of the actual average data rate for all past time slots of the eMBB user and the data rate achieved by the current channel conditions
- (5) Select the eMBB user corresponding to the minimum value according to the result obtained in step 10 and increase the number of RBs for this eMBB user by 1.
- (6) Repeat steps 4–7 until all RBs have been allocated.

3.2 URLLC Resource Allocation

The URLLC resource allocation problem can be expressed, for given RBs and Φ as follows

$$\mathcal{P}_3 \quad \max_{\gamma} \quad R_e^{Mactual}(t) \quad (27)$$

$$\text{s.t.} \quad \sum_{u \in U} \sum_{b \in B} I_u^m(t) \gamma_{u,b}^m(t) \leq |B|, \forall m \in M, \forall t \in T, \quad (27a)$$

$$\sum_{u \in U} \gamma_{u,b}^m(t) \leq 1, \forall b \in B, \forall m \in M, \forall t \in T, \quad (27b)$$

$$\sum_{b \in B} I_u^m(t) \gamma_{u,b}^m(t) \geq 1, \forall u \in U, \forall m \in M, \forall t \in T, \quad (27c)$$

$$I_u^m(t) \xi_u^m(t) \leq \delta R_u^m(t), \forall u \in U, \forall m \in M, \forall t \in T, \quad (27d)$$

$$P \left[\sum_{b \in B} I_u^m(t) < U \right] \leq \epsilon, \forall m \in M, \forall t \in T, \quad (27e)$$

$$\gamma_{u,b}^m(t), I_u^m(t) \in \{0, 1\}, \forall u \in U, \forall b \in B, \forall m \in M, \forall t \in T, \quad (27f)$$

\mathcal{P}_3 is still a non-convex problem because of its probability constraints and binary parameters. If \mathcal{P}_3 is treated in a relaxed way and then solved by convex optimization method, it will often violate the strict delay requirement of URLLC. Therefore, in order to meet the delay requirements and reliability constraints of URLLC, we adopt puncturing scheduling and PF algorithm to arrange the resource allocation problem of URLLC users. According to Algorithm 2, the eMBB users have completed resource allocation at the time slot boundary. All resource blocks are occupied by eMBB users, URLLC users choose the puncturing technique to seize spectrum resources from the eMBB users. The RBs

allocation of URLLC users is shown in Algorithm 3. The purpose of the PF scheduler is to maximize the data rates for eMBB users while ensuring fairness among eMBB users. $S_e^m(t)$ denotes the average data rate of the eMBB user e at mini-slot m of time slot t as follows:

$$S_e^m(t) = \left(1 - \frac{1}{m}\right) S_e^{m-1}(t) + \left(\frac{1}{m}\right) R_e^{m-1}(t) \quad (28)$$

where $R_e^{m-1}(t)$ represents the data rate of eMBB user e at mini-slot $(m-1)$ of time slot t . The objective function of the PF scheduler is expressed as:

$$\max \left(\frac{R_e^{m-1}(t)}{S_e^m(t)} \right) \quad (29)$$

Algorithm 3. URLLC Resource Allocation at m of t

- 1: **Initialization:** e, u, m
 - 2: **for** u in set U **do**
 - 3: **for** e in set E **do**
 - 4: Calculate $avgR_e^{m-1} = \frac{\sum_{m=1}^{m-1} R_e^{m-1}}{m-1}$ and the current mini-slot m R_e^m
 - 5: Calculate $Q(e) = \left(1 - \frac{1}{m}\right) avgR_e^{m-1} + \left(\frac{1}{m}\right) R_e^m$
 - 6: u selects $\max(Q(e))$ RBs for puncturing
 - 7: **end for**
 - 8: **end for**
-

The interpretation of Algorithm 3 is as follows:

1. Initialization related parameters.
2. Calculate the sum of the average data rate of eMBB users in all mini slots before m and current mini-slot m eMBB users data rate.
3. Select the eMBB user corresponding to the maximum value according to the result obtained in step 5 and URLLC users puncture this chosen eMBB user RBs
4. Repeat steps 2-3 until all URLLC users in m at t have been allocated.

4 Computational Complexity Analysis

This section presents the complexity of Algorithm 1, 2 and 3. The main computational complexity in Algorithm 1 focuses on matrix inversion in step 3, and its computational complexity is $O(|N+1|^3)$. Steps 4 and 5 are simple vector addition and subtraction which can be ignored. Therefore, the overall computational complexity of Algorithm 1 is $O(|N+1|^3)$. As for Algorithm 2 for per time slot, to allocate the remaining resource blocks to eMBB users is $(|B| - |E|)$. The complexity that per RB needs to be allocated is $O(|E|)$. The total computational complexity of Algorithm 2 in the whole time slot is $O(|E||B||T|)$. The overall computational complexity of Algorithm 3 in each mini-slot amounts to $O(|U||E|)$.

5 Simulation Results and Discussions

5.1 Simulation Setup

In this scenario, we consider a cellular radio network consisting of a BS and a RIS, and assume a coverage area of 200 m radius centered on the BS. The RIS is deployed at 50 m from the BS. The BS and each user (eMBB users and URLLC users) are respectively equipped with an antenna. Table 2 gives the important parameters for this simulation work. We assume that a physical resource block can satisfy the reliability service requirements of a URLLC user. It is assumed that the link between the base station and the users and the link between the RIS and the users are quasi-static flat fading Rayleigh channel. The large-scale fading model is expressed as $PL = \beta_0 (d_{bs,e/u})^{-\varrho_0}$, and $PL = \beta_1 (d_{bs,r_{is}})^{-\varrho_1} (d_{r_{is},e})^{-\varrho_2}$, where $d_{i,j}$ denotes the distance between i and j , β_0 and β_1 are path loss at the reference distance, $\varrho_0, \varrho_1, \varrho_2$ are the path loss exponents. We assume that BS-RIS channel follows Rician fading, The Rician fading channel can be denoted as

$$f = \left(\sqrt{\frac{\kappa}{1+\kappa}} f^{\text{LoS}} + \sqrt{\frac{1}{1+\kappa}} f^{\text{NLoS}} \right) \quad (30)$$

where κ denotes the Rician factor, f^{LoS} and f^{NLoS} are the line-of-sight components and the non-line-of-sight components, respectively. The channel f is a Rayleigh fading channel when $\kappa = 0$. Let $\kappa = 10$ [19], The channel f is a Rician fading channel (1).

Table 1. Summary of the simulation setup

Parameter	Value	Parameter	Value
P	30 dBm	N_0	-114 dBm
$ B $	50	M	8
$ E $	10	ϵ	0
λ	1, 2, ..., 9	δ	0.125 ms
f_b	180 kHz	T	1000
N	60	σ^2	1
β_0, β_1	-30 dBm, -40 dBm	$\varrho_0, \varrho_1, \varrho_2$	3.5, 2.2, 2.8

The simulation results are compared with the results of the following advanced allocation schemes. Baseline1: Resource allocation scheme for co-existence of eMBB and URLLC services in the literature [3]. Baseline2: Allocate resource blocks equally to each eMBB user at the beginning of each time slot [12]. The following are two performance evaluation parameters for eMBB users:

(1) The calculation of MEAR is as follows:

$$\text{MEAR} = \min \left(\frac{1}{T} \sum_{t=1}^{|T|} R_e^{\text{actual}}(t) \right), \forall e \in E, \quad (31)$$

(2) The calculation of Fairness is denoted by Jain's Fairness Index [16]:

$$\text{Fairness} = \frac{\left(\sum_{e \in E} \frac{1}{|T|} \sum_{t=1}^{|T|} R_e^{\text{actual}}(t) \right)^2}{|E| * \sum_{e \in E} \left(\frac{1}{|T|} \sum_{t=1}^{|T|} R_e^{\text{actual}}(t) \right)^2} \quad (32)$$

5.2 Simulation Results

We compare the MEAR of different methods with different λ , as shown in Fig. 2. The Proposed scheme produce a higher MEAR, compared to Baseline1 and Baseline2 schemes. This is because the Proposed scheme takes into account not only the loss of previous time slots for eMBB users, but also the channel conditions of the current slot. With the increase in λ , the proposed scheme and baseline1 decrease accordingly, as the number of resource blocks allocated to URLLC users increases. Baseline1 MEAR has not changed because the current number of URLLC users does not yet threaten the rate of the lowest user. In addition, when the RIS elements is 60, we observe that the Proposed_add_ris scheme improves MEAR the most, due to the fact that RIS improves the channel conditions for the lowest eMBB users, thus eliminating the need to deprive more resource blocks from users with better channel conditions to compensate for users with poor channel conditions.

Figure 3 shows the different λ comparing the fairness of the different schemes. The RIS elements is set as 60. Figure 3 reveals that the fairness score curve for the Proposed scheme is stable at 0.99. Compared to the Baseline1 and Baseline2 schemes, the Proposed scheme achieves significant fairness. Compared to Baseline1, Baseline1_add_ris scheme fairness score improves by 0.02 when $\lambda = 1$. Compared to Baseline2, Baseline2_add_ris scheme fairness score improves by 0.04 when $\lambda = 1$. Compared to Proposed, Proposed_add_ris scheme fairness score improves by 0.008 when $\lambda = 1$. This is because when RIS is introduced, the transmission data rate of eMBB users with worse channel conditions can be improved. thus reducing the data rate gap among eMBB users. The proposed scheme has the highest fairness, which is due to the consideration of the average rate of all the previous time slots and the current time slot channel, where users in the low average rate and poor channel gain are given priority in the allocation of resources.

The empirical cumulative distribution function (ECDF) and MEAR is shown in Fig. 4, We evaluated the results with $\lambda = 15$, When RIS is not

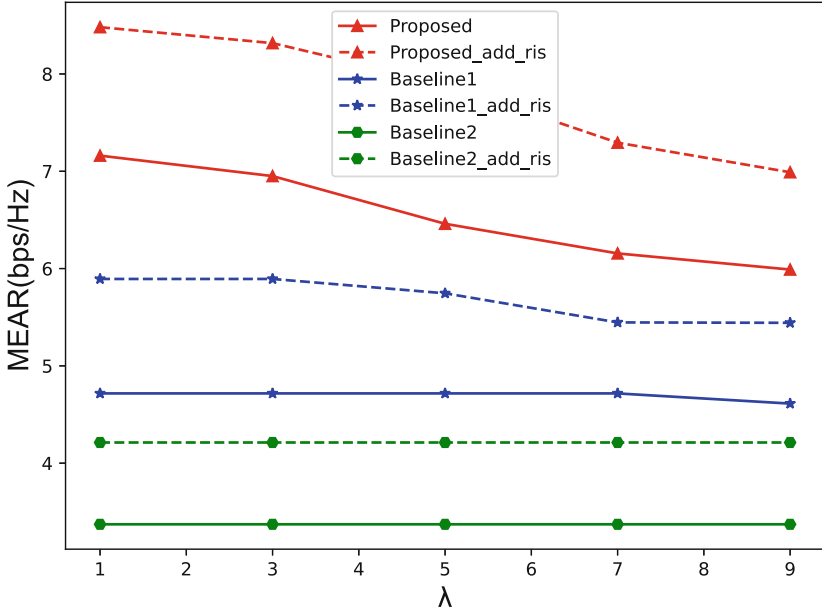


Fig. 2. λ versus MEAR

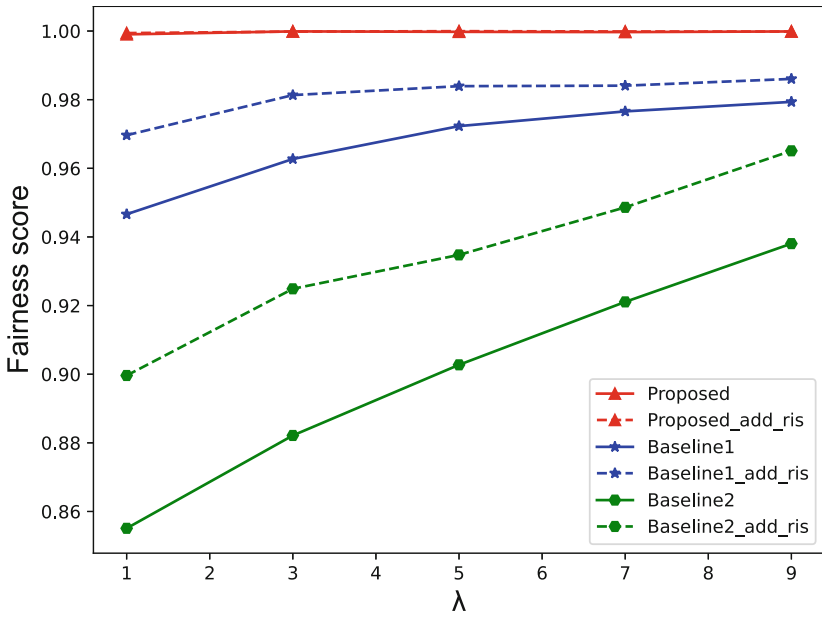


Fig. 3. λ versus Fairness score

added, the probability of the Proposed scheme below 5 bps/Hz is 0.08, while the probabilities for Baseline1 and Baseline2 are 0.18 and 0.23 respectively. The Proposed scheme ensures the goal of maximizing MEAR among eMBB users. When $N = 60$ RIS elements is added, the probability of the Proposed_add_ris scheme below 6 bps/Hz is 0.08, compared to 0.20 and 0.21 for Baseline1_add_ris and Baseline2_add_ris, respectively. It can be concluded that Proposed_add_ris scheme can achieve better MEAR performance.

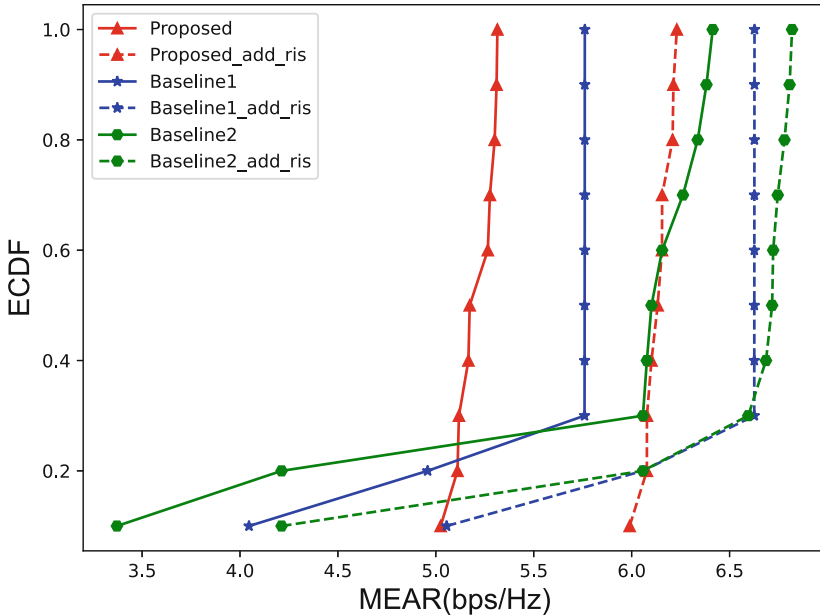


Fig. 4. MEAR versus ECDF

6 Conclusion

In this paper, we presented a new architecture for the coexistence of RIS-assisted eMBB and URLLC services. The RIS was deployed to improve the minimum eMBB channel conditions and thus increased its transmission rate, using a low-complexity heuristic algorithm to allocate RBs for eMBB users. The PF algorithm is used to allocate URLLC RBs by puncturing the eMBB users being transmitted under the condition that the URLLC QoS is satisfied. Simulation results verify that our proposed scheme has a better performance than other schemes. We will continue with the following research, because solving the phase shift matrix of RIS has a high computational complexity, it is not easy to apply in scenarios with high latency requirements. We plan to use deep reinforcement

learning algorithms to train the RIS phase shift matrix offline, making it possible to deploy RIS for URLLC services.

Acknowledgement. This work was supported by National Key R&D Program of China (No. 2020YFB1806702).

References

1. Popovski, P., Trillingsgaard, K.F., Simeone, O., et al.: 5G wireless network slicing for eMBB, URLLC, and mMTC: a communication-theoretic view. *IEEE Access* **6**, 55765–55779 (2018)
2. Saad, W., Bennis, M., Chen, M.: A vision of 6G wireless systems: applications, trends, technologies, and open research problems. *IEEE Netw.* **34**(3), 134–142 (2020)
3. Bairagi, A.K., et al.: Coexistence mechanism between eMBB and uRLLC in 5G wireless networks. *IEEE Trans. Commun.* **69**(3), 1736–1749 (2021)
4. Prathyusha, Y., Sheu, T.-L.: Coordinated resource allocations for eMBB and URLLC in 5G communication networks. *IEEE Trans. Veh. Technol.* **71**(8), 8717–8728 (2022)
5. Almekhlafi, M., Arfaoui, M.A., ASSi, C., et al.: Superposition-Based URLLC Traffic Scheduling in 5G and Beyond Wireless Networks. *IEEE trans. Commun.* **70**(9), 6295–6309(2022)
6. Alsenwi, M., Tran, N.H., Bennis, M., et al.: Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: a deep reinforcement learning based approach. *IEEE Trans. Wireless Commun.* **20**(7), 4585–4600 (2021)
7. Alsenwi, M., Tran, N.H., Bennis, M., et al.: eMBB-URLLC resource slicing: a risk-sensitive approach. *IEEE Commun. Lett.* **23**(4), 740–743 (2019)
8. SHi, B., Zheng, F. C., She, C.Y., et al.: Risk-resistant resource allocation for eMBB and URLLC coexistence under M/G/1 queueing model. *IEEE Trans. Veh. Technol.* **71**(6), 6279–6290(2022)
9. Chen, Q., Wu, J., Wang, J., et al.: Coexistence of URLLC and eMBB services in MIMO-NOMA systems. *IEEE Trans. Veh. Technol.* **72**(1), 839–851 (2023)
10. Liu, B., Zhu, P.C., Li, J.M., et al.: Energy-efficient optimization via joint power and subcarrier allocation for eMBB and URLLC services. *IEEE Wireless Commun. Lett.* **11**(11), 2340–2344 (2022)
11. Setayesh, M., Bahrami, S., Wong, V.W.S.: Resource slicing for eMBB and URLLC services in radio access network using hierarchical deep learning. *IEEE Trans. Wireless Commun.* **21**(11), 8950–8966 (2022)
12. Almekhlafi, M., Arfaoui, M.A., Elhattab, M., et al.: Joint resource allocation and phase shift optimization for ris-aided eMBB/URLLC traffic multiplexing. *IEEE Trans. Commun.* **70**(2), 1304–1319 (2022)
13. Wu, Q., Zhang, B., You, C., et al.: Intelligent reflecting surface aided wireless communications: a tutorial. *IEEE Trans. Commun.* **69**(5), 3313–3351 (2021)
14. Anand, A., Veciana, G., de Shakkottai, S.: Joint scheduling of URLLC and eMBB traffic in 5G wireless networks. *IEEE/ACM Trans. Netw.* **28**(2), 477–490 (2020)
15. Boyd, S., Parikh, N., Chu, E., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2010)

16. Jain, R., Chiu, D., Hawe, W.: A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. *Digit Equip Corp.* **21**, (1984)
17. Polyanskiy, Y., Poor, H. V., Verdú, S.: Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory.* **56**(5), 2307–2359 (2010)
18. Yang, W., Durisi, G., Korch, T., et al.: Quasi-static multiple-antenna fading channels at finite blocklength. *IEEE Trans. Inf. Theory* **60**(7), 4232–4265 (2014)
19. Zuo, J.K., Liu, Y. W., Al-Dhahir, N.: Reconfigurable intelligent surface assisted cooperative non-orthogonal multiple access systems. *IEEE Trans. Commun.* **69**(10), 6750–6764 (2021)