Min-max Decoding Error Probability Optimization in RIS-Aided Hybrid TDMA-NOMA Networks

Tra Huong Thi Le, and Yan Kyaw Tun, *Member, IEEE.*

Abstract—One of the primary objectives for future wireless communication networks is to facilitate the provision of ultra-reliable and low-latency communication services while simultaneously ensuring the capability for vast connection. In order to achieve this objective, we examine a hybrid multi-access scheme inside the finite blocklength (FBL) regime. This system combines the benefits of non-orthogonal multiple access (NOMA) and time-division multiple access (TDMA) schemes with the aim of fulfilling the objectives of future wireless communication networks. In addition, a reconfigurable intelligent surface (RIS) is utilized to facilitate the establishment of the uplink transmission between the base station and mobile devices in situations when impediments impede their direct communication linkages. This paper aims to minimize the worst-case decoding-error probability for all mobile users by jointly optimizing power allocation, receiving beamforming, blocklength, RIS reflection, and user pairing. To deal with the coupled variables in the formulated mixed-integer non-convex optimization problem, we decompose it into three sub-problems, namely, 1) decoding order determination problem, 2) joint power allocation, receiving beamforming, RIS reflection, and blocklength optimization problem, and 3) optimal user pairing problem. Then, we provide the sequential convex approximation (SCA) and semidefinite relaxation (SDR)-based algorithms as potential solutions for iteratively addressing the deconstructed first two sub-problems at a fixed random user pairing. In addition, the Hungarian matching approach is employed to address the challenge of optimizing user pairing. In conclusion, we undertake a comprehensive simulation, which reveals the advantageous qualities of the proposed algorithm and its superior performance compared to existing benchmark methods.

Index Terms—Reconfigurable intelligent surface (RIS), non-orthogonal multiple access (NOMA), time-division multiple access (TDMA), finite blocklength (FBL), Ultra-reliable and low latency communication (URLLC), successive convex approximation (SCA), semidefinite relaxation (SDR), Hungarian matching.

I. INTRODUCTION

Ultra-reliable and low latency communication (URLLC) has recently garnered increased interest because of its potential to support a wide range of essential applications [1]. In the context of URLLC, the transmission of small packets, referred to as finite blocklengths (FBL), is commonly employed in order to adhere to stringent reliability and latency criteria. Within the context of FBL (Flexible Block Length) systems, there exists a trade-off between the benefits derived by coding and the associated latency. Toward this end, the Shannon capacity formula is not inappropriate for FBL. Polyanskiy et al. have derived an approximate equation for the achievable rate with the finite blocklength codes, which is a function of the signal-to-noise (SNR), the blocklength, and the decoding error probability [2]. The findings of the study indicate that the Shannon rate serves as the maximum attainable rate within the finite blocklength regime, taking into account both the chance of decoding errors and the transmission blocklength. Furthermore, (FBL) systems are susceptible to deterioration in transmission rates and inevitable mistakes in packet decoding, particularly in environments with severe propagation characteristics that result in poor channel conditions for users.

On the other hand, reconfigurable intelligent surface (RIS) technology holds promise for improving URLLC system efficiency by counteracting the negative effects of the wireless channel's random propagation nature [3], [4]. The wireless environment undergoes reconfiguration with the assistance of RIS, resulting in the transformation of random wireless channels into somewhat deterministic beamforming designs. In contrast to conventional relays, the RIS operates without any delay due to the absence of analog-to-digital conversion through radio frequency (RF) chains [5], [6]. Moreover, if the direct communication link is disrupted, RIS can create a communication link between the transceivers. As a result, in the event of a disruption in the direct communication link, the RIS has the capability to establish a communication link between the transceivers. Consequently, the integration of RIS into the system has the potential to enhance reliability, minimize delay, and support URLLC-based wireless systems. Due to its promising potential, several previous studies have been dedicated to examining the effects of RIS on URLLC. For instance, in [7], the optimization of user grouping, blocklength allocation, and reflective beamforming was performed in a combined manner. The objective of this optimization was to reduce the overall delay experienced by all users while ensuring that their particular dependability needs were met. The works in [8], [9] explained one alternative performance metric for the performance gain of the RIS-aided URLLC system. The study in [10] aimed to reduce the transmit power of base stations (BS) in a secure system that utilizes reconfigurable intelligent surfaces (RIS) to aid in multiple-input-single-output (MISO) URLLC. This was achieved by developing techniques for constructing artificial noise, BS beamformers, and RIS passive beamforming. The aforementioned studies have provided evidence that the optimization of intelligent reflecting surface (IRS) beamforming can enhance the channel quality of wireless communication networks.

Moreover, offering URLLC to large numbers of users with constrained resources is one of the most difficult issues in

Tra. H. T. Le is with the Department of Information Technology, FPT University, FPT Da Nang Technology Urban Area, Hoa Hai Ward, Ngu Hanh Son District, Da Nang City, Vietnam, e-mail: tralth4@fe.edu.vn.

Yan Kyaw Tun is with the Department of Electronic Systems, Aalborg University, 2450 København, Denmark, e-mail: ykt@es.aau.dk

6G compared to 5G. By sharing the bandwidth among users (and avoiding the requirement to split the blocklength in the process), NOMA may achieve great spectrum efficiency in comparison to regular OMA. This lowers communication latency for the users. Recent research has examined the effectiveness of NOMA in the URLLC system [7], [11]–[15]. The authors of [7] put up a very dependable resource allocation strategy to address the NOMA-related dependability issue for combined uplink and downlink transmission in URLLC. A static multiuser NOMA-URLLC framework based on hybrid automated repeat request re-transmissions was created by the authors in [11]. To illustrate the functionality of NOMA-URLLC, a range of network setups with persistent user connectivity were simulated inside this framework. Four distinct transmission strategies were examined in a NOMA-based URLLC scenario in [12]. While guaranteeing the error probability of the user under the ideal channel condition, the author optimized the user's error probability in the presence of the worst channel condition. NOMA-URLLC systems perform better than OMA-URLLC schemes under all circumstances, as shown by [11] and [12]. The study in [13] examined the ET maximization problem of a multicarrier NOMA system in NOMA-URLLC using a dynamic programming technique. Based on error probabilities and the maximum transmit power, the simulation results showed that NOMA-URLLC systems operate effectively under all given conditions. The fading of wireless channels, which might result in low channel quality, makes it impossible to constantly satisfy the rigorous criteria of dependability and latency in NOMA-URLLC networks. The reflecting amplitudes and phase shifts of an RIS's reflecting elements can be adjusted to fix this problem. Furthermore, RIS-reflecting beamforming optimization can offer extra degrees of freedom to improve the communication performance of URLLC and NOMA networks.

Recently, integrating NOMA and RIS for URLLC system has drawn a lot of interest. Effective beamforming in RIS-aided networks provides 90% greater rate performance than its orthogonal equivalent, according to the authors of [16]. Furthermore, using an RIS-aided network for the system under consideration improves rate performance by 40–50% and 56–80%, respectively, over using random phase-shifting at the RIS and using a system without an RIS. [17] shown that, compared to its orthogonal multiple access counterpart, the optimal phase-shift determination at the RIS in RIS-aided short-packet NOMA systems under perfect and imperfect consecutive interference cancellation may increase block error rate and throughput. Then, [18] revealed that the considered system achieves a higher achievable sum rate than RIS-based short-packet orthogonal multiple access ones. Furthermore, the work in [19] showed that RIS-reflecting beamforming can assist in increasing the significant sum throughput of all users in NOMA-URLLC. In addition, the RIS-NOMA-aided URLLC communication was also introduced into the different scenarios such as the presence of hardware impairments [20], mobile edge computing [21] and smart grid [22].

However, the NOMA scheme introduces interference among the users and depends on the execution of successive interference cancellation (SIC) [23]. This will significantly influence the QoS of the transmissions, e.g., the reliability and the delay due to retransmissions. Based on the findings above, combining the benefits of both NOMA and TDMA is critical to achieve a compromise between longer blocklength and low/no interference to improve massive access while offering URLLC services. Recently, researchers have started to study the integration of RIS and hybrid NOMA-TDMA [24]–[27] and the incorporation of hybrid NOMA-TDMA and URLLC [23], [28]. [24] showed that RIS assignment can reduce the power consumption in RIS-aided NOMA-TDMA. The work in [25] demonstrated that RIS beamforming optimizing could improve the minimum beampattern gain (MBBG) in the hybrid NOMA-TDMA-aided integrated sensing and communication (ISAC) system. In [26], the combination of STAR-RIS and hybrid NOMA-TDMA outperforms only RIS-based and the OMA-based frameworks in terms of maximal of minimum downlink rate. [27] showed the tradeoff between performance and cost can be more effectively balanced in the hybrid NOMA-TDMA with the assistance of active RIS. In [23], the benefits of NOMA and TDMA manifest the reduction of the total delay in the finite blocklength (FBL) regime. [28] showed that the timely throughput can be improved up to 46% by the hybrid NOMA compared to conventional OMA. However, to the best of our knowledge, the study of optimizing RIS reflecting beamforming in the hybrid NOMA-TDMA-aided URLLC networks is still missing in the literature.

Therefore, in this paper, we consider combining the uplink NOMA scheme within the user pairs and the TDMA scheme between the user pairs. In particular, how to address the trade-off mentioned above between performance and complexity by optimally allocating the transmit power, blocklength of users with limited radio resources, receiving beamforming, and passive beamforming is still an open problem. This issue is especially critical in multi-user scenarios, where the scalability of the allocation schemes should also be considered. Additionally, because URLLC is mostly used to send measurement data or control signals with small packet sizes, throughput optimization is less important. Therefore, our goal is to generalize the min-max decoding-error probability-based resource allocation, which simultaneously reflects the minimization of the decoding-error probability for all devices and the fairness among them—a factor that has not yet been taken into account in previous research on RIS-URLLC. Our main contributions are summarized as follows:

- In order to minimize the maximum decoding error probability for all users, we consider joint power allocation, transmission blocklength, receiving beamforming, RIS reflection, and user pairing optimization. This optimization is conducted while taking into account limits on transmit power and latency. The formulated problem above poses significant challenges because of the non-convex nature and the complexity associated with throughput expression in the small blocklength regime. In order to achieve this objective, the formulated problem is divided into three separate subproblems.
- The first subproblem is decoding order determinations. The SDR was devised in order to determine the decoding

orders by identifying the RIS reflection that maximizes the combined channel gain. Subsequently, utilizing the established decoding sequence, the whole user population is categorized into two distinct cohorts: proficient users and less proficient users. In the non-orthogonal multiple access (NOMA) system, data transmission to the base station (BS) occurs in each time slot through the random pairing of one strong and one weak user.

- The second subproblem involves the combined optimization of power allocation, receiving beamforming, RIS reflection, and transmission blocklength while maintaining fixed user pairing. In this study, we present a novel approach for addressing the subproblem in an efficient manner. Our proposed algorithm combines the methodologies of sequential convex approximation (SCA) and semidefinite relaxation (SDR) to achieve this goal.

- The third subproblem is determining user pairing for each time slot. In order to address the problem, we suggest employing a Hungarian matching-based iterative method.

- Finally, by the implementation of a comprehensive simulation, we are able to demonstrate the notable benefits of our suggested framework in comparison to alternative benchmark methods.

The rest of this paper is organized as follows. Our system model is presented in Section II. Problem formulation is provided in Section III, and Section V represents our proposed solution. The numerical results are presented in Section V. The paper is finally concluded in Section VI.

Notation: Boldface lower-case and upper-case characters stand for vectors and matrices, respectively. For real-valued elements, the set of $n \times m$ matrices is represented by $\mathbb{R}^{n \times m}$, and for complex-valued entries, by $\mathbb{C}^{n \times m}$. $\mathbb{S}_+^n$ is a representation of the set of $n \times n$ semidefinite positive matrices. The diagonal matrix with a diagonal vector $\boldsymbol{a}$ is represented by $\text{diag}(\boldsymbol{a})$, the (n, m)th entry of a matrix $\boldsymbol{A}$ is indicated by $[\boldsymbol{A}]_{n,m}$, and the nth entry of a vector $\boldsymbol{a}$ is shown by $[\boldsymbol{a}]_n$.

II. SYSTEM MODEL

In this paper, we consider a RIS-URLLC, as shown in Fig. 1, whereby several users operate covertly from the base station (BS) and employ short blocklengths for transmission, aided by a RIS. There exists a set $\mathcal{K}$ of K users, each equipped with a single antenna, engaging in transmission to a BS that possesses N_t antennas. All users transmit status update packets, respectively with D_k-bits ($k \in \mathcal{K} = \{1, \cdots, K\}$) information to the BS. The direct connections between users and BS are impeded due to several constraints. In order to improve the communication link between users and BS, RIS consisting of N components is employed. The K users are partitioned into $K/2$ groups, with each group consisting of two individuals. Simultaneous transmission occurs among users within the same group, whereas distinct groups employ a TDMA approach for transmission. In our paper, we make the assumption that the channel gain remains constant during the specified time length.

User k encodes the D_k information into a single packet (codeword) with a length of M^2 symbols. The channels
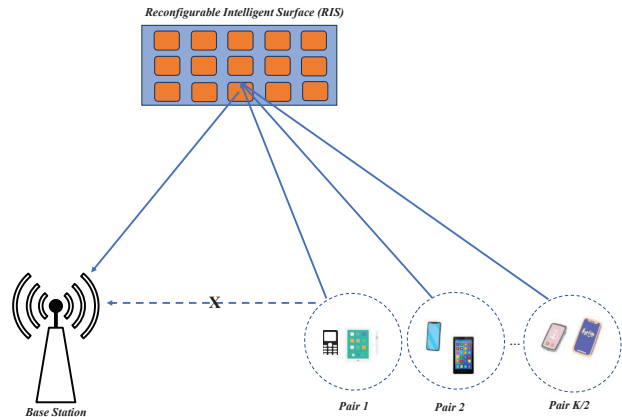

Fig. 1: System model.

between BS and RIS, and between RIS and user k are denoted by $\boldsymbol{H} \in \mathbb{C}^{N \times N_t}$, and $\boldsymbol{f}_k \in \mathbb{C}^{N \times 1}$, respectively. We can model $\boldsymbol{H}$ as the following Rician fading channel

$$\boldsymbol{H} = \sqrt{\frac{\rho_0}{d_0^{\alpha_1}}} \left(\sqrt{\frac{K_1}{1 + K_1}} \boldsymbol{H}^{LOS} + \sqrt{\frac{1}{1 + K_1}} \boldsymbol{H}^{NLOS} \right), \tag{1}$$

where the deterministic line-of-sight (LoS) component is denoted as $\boldsymbol{H}^{LOS}$, while the random non-line-of-sight (NLoS) component is represented as $\boldsymbol{H}^{NLOS}$. Both components are modeled using Rayleigh fading. d_0 represents the distance between BS and RIS. $\alpha_1 \geq 2$ is used to express the path loss exponent, whereas ρ_0 is used to describe the path loss at a reference distance of 1 m. K_1 represents the Rician factor. Similarly, the Rician fading channel from the RIS to user k is expressed as:

$$\boldsymbol{f}_k = \sqrt{\frac{\rho_0}{d_k^{\alpha_2}}} \left(\sqrt{\frac{K_2}{1 + K_2}} \boldsymbol{f}_k^{LOS} + \sqrt{\frac{1}{1 + K_2}} \boldsymbol{f}_k^{NLOS} \right), \tag{2}$$

where $\boldsymbol{f}_k^{LOS}$ and $\boldsymbol{f}_k^{NLOS}$ are the LoS and NLoS components, respectively, and d_k denotes the distance between the RIS and user k. α_2 and K_2 are path loss exponent and Rician factor, respectively. The reflection vector $\boldsymbol{\phi} \in \mathbb{C}^{N \times 1}$ of RIS is determined by the phase shift vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_N]^T \in \mathbb{R}^{N \times 1}$ as

$$\boldsymbol{\phi} = [\phi_1, \phi_2, ..., \phi_N]^T = [e^{j\theta_1}, e^{j\theta_2}, ..., e^{j\theta_N}]^T, \tag{3}$$

where $\theta_n \in [0, 2\pi)$ leading to the unit modulus reflection, or equivalently $|\phi_n| = 1$ and $n \in \mathcal{N} \triangleq \{1, 2, \cdots, N\}$. The received signal at BS can be written as

$$z_k = \sum_{k \in \mathcal{K}} \boldsymbol{w}_k^H \left(\left\{ \boldsymbol{H}^H \text{diag}(\boldsymbol{\phi}) \boldsymbol{f}_k \right\} s_k + \boldsymbol{n}_0 \right), \tag{4}$$

where s_k is the symbol delivering the model of user k, $\boldsymbol{n}_0 \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) with the noise power σ^2 at BS, $\boldsymbol{w}_k$ is receiving beamforming at BS of user k. We denote k, k' are the same group if $b_{k,k'} = 1$.

Because each group has at most 2 users, we have

$$\sum_{k \in \mathcal{K}|k'} b_{k',k} \leq 1, \ \forall k' \in \mathcal{K}. \quad (5)$$

The BS employs a technique known as successive interference cancellation (SIC) to accomplish symbol detection with (4). The SIC order is denoted as $\pi(k)$ for user $k \in \mathcal{K}$, where $\pi(k) = n$ represents the detection of the symbol of user k at the n-th order in the SIC process. Assuming that the decoding order π corresponds to the decreasing order of direct channel gain, it can be inferred that the user signal with the greatest channel gain within each group is initially decoded at the BS. With $\pi(k) \leq \pi(k')$ and k, k' are in same group, we have

$$\gamma_k = \frac{\left| \boldsymbol{w}_k^H \boldsymbol{H}^H \mathrm{diag}(\boldsymbol{\phi}) \boldsymbol{f}_k \right|^2 p_k}{\left| \boldsymbol{w}_{k'}^H \boldsymbol{H}^H \mathrm{diag}(\boldsymbol{\phi}) \boldsymbol{f}_{k'} \right|^2 p_{k'} + \sigma^2}, \quad (6)$$

and

$$\gamma_{k'} = \frac{\left| \boldsymbol{w}_{k'}^H \boldsymbol{H}^H \mathrm{diag}(\boldsymbol{\phi}) \boldsymbol{f}_{k'} \right|^2 p_{k'}}{\sigma^2}, \quad (7)$$

where p_k and $p_{k'}$ are the transmit power of user k and k', respectively. Therefore, generally, the SINR of user k is

$$\gamma_k = \frac{\left| \boldsymbol{w}_k^H \boldsymbol{H}^H \mathrm{diag}(\boldsymbol{\phi}) \boldsymbol{f}_k \right|^2 p_k}{\sum\limits_{k' \in \mathcal{K}, \pi(k') > \pi(k)} \left| \boldsymbol{w}_k^H \boldsymbol{H}^H \mathrm{diag}(\boldsymbol{\phi}) \boldsymbol{f}_{k'} \right|^2 p_{k'} b_{k,k'} + \sigma^2}, \quad (8)$$

which leads to coding rate (in bits/sec/Hz) by user k under short blocklength regime as follows [2]

$$R_k = \log_2(1 + \gamma_k) - \frac{V_k}{M} \frac{Q^{-1}(\epsilon_k)}{\ln 2}, \quad (9)$$

where $Q(\cdot)$ is the Q-function, $V_k = (1 - (1 + \gamma_k)^{-2})^{1/2} \sim 1$ [29]. If a fixed number of D_k bits are transmitted during the channel blocklength M^2 from user k, its coding rate becomes $R_k = \frac{D_k}{M^2}$ bits/sec/Hz. Then the block error rate probability ϵ_k for user k is given

$$\epsilon_k = Q(g_k), \quad (10)$$

where

$$g_k = \frac{M \ln 2}{V_k} \left(\log_2(1 + \gamma_k) - \frac{D_k}{M^2} \right), \quad (11)$$

The average energy of user k is

$$\bar{E}_k = T_{sym} M^2 p_k, \quad (12)$$

where T_{sym} is the time for transmitting one symbol.

III. PROBLEM FORMULATION

This paper aims to formulate joint power allocation, receiving beamforming, blocklength, RIS reflection, and user pairing problem, which minimizes the maximal error decoding probability subject to the energy and the RIS's practical reflection constraints. Thus, we write the formulated problem mathematically as follows:

$$\mathbf{P}: \min_{\boldsymbol{\phi}, p_k, \boldsymbol{w}_k, b_{k,k'}, M} \max_k \epsilon_k \quad (13)$$

$$\text{s.t.} \quad 0 < p_k \leq p_k^{\max}, \ \forall k \in \mathcal{K}, \quad (13a)$$

$$|\phi_n| = 1, \ \forall n \in \mathcal{N}, \quad (13b)$$

$$||\boldsymbol{w}_k||^2 = 1, \ \forall k \in \mathcal{K}, \quad (13c)$$

$$\sum_{k \in \mathcal{K}|k'} b_{k',k} \leq 1, \ \forall k, k' \in \mathcal{K}, \quad (13d)$$

$$\sum_{k \in \mathcal{K}} \bar{E}_k \leq E_0, \quad (13e)$$

$$M^2 \leq \frac{T}{T_{symbol}}, \quad (13f)$$

$$b_{k',k} \in \{0, \ 1\}, \ \forall k, k' \in \mathcal{K}, \quad (13g)$$

$$M^2 \in \mathbb{Z}^+, \quad (13h)$$

where T is the duration for one-time slot, $p_k^{\max}$ is the maximum transmit power of user k. Constraints (13a) and (13b) are the power constraint for each user and the RIS reflection constraints of the RIS's elements, respectively. Constraint (13c) presents the receiving beamforming constraint, and constraint (13d) ensures each user can be paired with at most one other user. Constraint (13e) expresses the total energy constraint for all users, and constraint (13f) is the latency constraint. Constraint (13g) ensures the user pairing constraint. Finally, constraint (13h) ensures that the blocklength allocated to each user is a positive integer. On the other hand, $\epsilon_k = Q(g_k)$ and Q-function is a monotonically decreasing function. Then, problem $\mathbf{P}$ can be replaced by the max-min function of g_k.

$$\mathbf{P1}: \max_{\boldsymbol{\phi}, p_k, \boldsymbol{w}_k, b_{k,k'}, M} \min_k g_k \quad (14)$$

s.t. (13a), (13b), (13c), (13d), (13g), (13e), (13f), (13h).

The formulated problem $\mathbf{P1}$ involves the continuous and integer variables. Furthermore, a strong coupling of the optimization variables and the intra-pair interference of a NOMA pair exists. As a consequence, $\mathbf{P1}$ is a non-convex MINLP problem that is challenging to solve. As a result, we develop an efficient solution to solve problem $\mathbf{P1}$.

IV. PROPOSED SOLUTION

In this regard, we first relax the mixed integer problem $\mathbf{P1}$ by removing the integer constraints (13h) on the variables M. To facilitate the solution design and to smooth the objective function of $\mathbf{P1}$, an auxiliary variable χ is introduced so that the equivalent form of $\mathbf{P1}$

$$\mathbf{P2}: \max_{\boldsymbol{\phi}, p_k, \boldsymbol{w}_k, b_{k,k'}, M, \chi} \chi \quad (15)$$

$$\text{s.t.} \quad g_k \geq \chi, \ \forall k \in \mathcal{K}, \quad (15a)$$

(13a), (13b), (13c), (13d), (13g), (13e), (13f).

In order to implement the NOMA communication system, it is important to arrange the users in accordance with their respective effective channel gains. However, In the case of RIS-assisted NOMA, it is important to note that the combined gain $\boldsymbol{H}^H \mathrm{diag}(\boldsymbol{\phi}) \boldsymbol{f}_k$ s influenced by the reflection coefficient

ϕ. previous studies have demonstrated that NOMA systems employing fixed power distribution techniques exhibit superior performance when pairing two users with differing channel conditions. Therefore, we propose an approach to determine the decoding orders by finding RIS reflection vectors that maximize the sum of combined channel gains. Based on the objective of accomplishing decoding orders, the entire user population is categorized into two distinct groups: strong users and weak users. Subsequently, each pair is formed by pairing one strong user with one weak user. Furthermore, due to the inherent challenges associated with handling both continuous and integer variables in **P2** is rather difficult to tackle, we employ the bi-level optimization framework. This allows us to decompose **P2** into two sub-problems: the inner problem focuses on optimizing power allocation, receiving beamforming, blocklength, and RIS reflection, while the outer problem deals with user pairing. The inner problem corresponds to a max-min g_k problem concerning the power allocation, receiving beamforming, blocklength, and RIS reflection variables given certain user pairing relationships. The outer problem corresponds to max-min g_k considering the user pairing variable. Consequently, the proposed algorithm has three steps, which are shown as follows.

- The first step is to determine the decoding order and then divide the total users into strong user subsets and weak user subsets.
- The second step is to optimize power allocation, receiving beamforming, RIS reflection, and blocklength under a fixed user pairing variable.
- The third step is to achieve the optimal user pairing relationships between a weak user and a strong user with the optimizing variables obtained by the second step.

A. Decoding Order Determination

This subsection presents a low-complexity algorithm to determine the decoding orders based on maximizing the combined channel gain. The formulated optimization problem is as follows:

$$\textbf{P3}: \quad \max_{\phi} |\boldsymbol{H}^H \text{diag}(\phi)\boldsymbol{f}_k|^2 \qquad (16)$$
$$\text{s.t.} \quad (13b).$$

In problem **P3**, the quadratic form $|\boldsymbol{H}^H\text{diag}(\phi)\boldsymbol{f}_k|^2$ can be rewritten as $\phi^H \boldsymbol{R}_k \phi$, where $\boldsymbol{R}_k = \text{diag}((f_k)^H)\boldsymbol{H}^H\boldsymbol{H}\text{diag}(f_k)$. By introducing slack matrices $\boldsymbol{S} = \phi\phi^H$, which are rank-one and positive semidefinite (PSD), we have $|\boldsymbol{H}^H\text{diag}(\phi)\boldsymbol{f}_k|^2 = \text{tr}(\boldsymbol{R}_k\boldsymbol{S})$. Thus, problem **P3** can be transformed to

$$\textbf{P3-1}: \quad \max_{\boldsymbol{S}} \text{tr}(\boldsymbol{R}_k\boldsymbol{S}) \qquad (17)$$
$$\text{s.t.} \quad \text{rank}(\boldsymbol{S}) = 1, \qquad (17a)$$
$$\boldsymbol{S} \succeq 0,. \qquad (17b)$$

As a standard SDP problem, problem **P3-1** can be solved by existing solvers in CVX [30]. However, to deal with the challenge of the rank-one constraint (17a), we can apply

the Gaussian randomization to obtain a suboptimal solution [31]. Based on the searched reflection matrix ϕ^*, if the combined channel gains experienced by any two users (k, k') associated with BS can be arranged as $|\boldsymbol{H}^H\text{diag}(\phi)\boldsymbol{f}_{k'}|^2 \leq |\boldsymbol{H}^H\text{diag}(\phi)\boldsymbol{f}_k|^2$. Then, we choose $K/2$ users with the most effective channel gain as the strong users and the remaining $K/2$ users as the weak users. In each time slot of TDMA, one strong user k and one weak user k' are paired and transmit data to BS in the NOMA scheme with the decoding order $\pi(k) \leq \pi(k')$.

B. Power Allocation, Receiving Beamforming, RIS Reflection, and Blocklength Optimization with Random User Paring

In this section, we investigate Alternating Optimization (AO) approach to handle problem **P1** efficiently by dividing it into four subproblems.

1) Power Allocation Optimization with Fixed Receiving Beamforming, RIS Reflection, and Blocklength:

$$\textbf{P4}: \quad \max_{\boldsymbol{p},\chi} \chi \qquad (18)$$
$$\text{s.t.} \quad (15a), \ (13a), \ (13e),$$

where $\boldsymbol{p} = \{p_k, \forall k \in \mathcal{K}\}$. The constraint (15a) is equivalent to

$$2^{\frac{\chi}{M\ln 2} + \frac{D_k}{M^2}} - 1 \leq \gamma_k. \qquad (19)$$

In addition, (8) is equivalent to

$$\sum_{k' \in \mathcal{K}, \pi(k') > \pi(k)} X_{k'}^k p_{k'}\gamma_k + \sigma^2 = X_k p_k, \qquad (20)$$

where $X_k = \left|\boldsymbol{w}_k^H \boldsymbol{H}^H \text{diag}(\phi)\boldsymbol{f}_k\right|^2$ and $X_{k'}^k = \left|\boldsymbol{w}_{k'}^H \boldsymbol{H}^H \text{diag}(\phi)\boldsymbol{f}_{k'}\right|^2$. We replace $p_{k'}\gamma_k$ with its upper bound defined as [32]

$$p_{k'}\gamma_k \leq \frac{\lambda_{k,k'}}{2}\gamma_k^2 + \frac{1}{2\lambda_{k,k'}}p_{k'}^2, \qquad (21)$$

where $\boldsymbol{\lambda} = \{\lambda_{k,k'}|k,k' \in \mathcal{K}\}$. To make problem (18) be convex, we replace (20) by

$$X_k p_k \geq \sum_{k' \in \mathcal{K}, \pi(k') > \pi(k)} X_{k'}^k \left(\frac{\lambda_{k,k'}}{2}\gamma_k^2 + \frac{1}{2\lambda_{k,k'}}p_{k'}^2\right) + \sigma^2. \qquad (22)$$

Then problem (18) is equivalent to

$$\textbf{P4-1}: \quad \max_{\boldsymbol{p},\chi,\boldsymbol{\gamma}} \chi \qquad (23)$$
$$\text{s.t.} \quad (19), (22), \ (13a), (13e).$$

where $\boldsymbol{\gamma} = \{\gamma_k, \forall k \in \mathcal{K}\}$.

It can be observed that the objective function and all constraints in problem **P4-1** exhibit convexity. As a result, the Karush-Kuhn-Tucker (KKT) solution of problem **P4-1** may be updated iteratively until convergence by optimally addressing its convex approximation problem using CVX [30]. The algorithm presented in Algorithm 1 summarizes the proposed power allocation algorithm, which incorporates an adjustable convergence accuracy denoted as ϵ.

Algorithm 1 SCA-Based Algorithm for Power Allocation

1: **Input:** Receiving beamforming matrix $\{w_k\}$, RIS reflection ϕ, and blocklength M
2: Set maximum number of iterations $I_{\max}^p$ and the maximum error tolerance ϵ^p
3: Initialize $p_s^{(0)}$ and $\chi_s^{(0)}$. Set $i = 0$.
4: **repeat**
5: $\quad i \leftarrow i + 1$.
6: $\quad$ Obtain $\chi_s^{(i)}$ and $p_s^{(i)}$ by solving problem **P4-1**.
7: **until** $(|\chi_s^{(i)} - \chi_s^{(i-1)}| <= \epsilon^p$) or $(i \geq I_{\max}^p)$
8: **Output:** $p^o = p_s^{(i)}$ and $\chi^o = \chi_s^{(i)}$

Algorithm 2 Bisection Search and SDR-Based Algorithm for Receiving Beamforming Matrix Optimization

---

1: **Input:** Transmission power p, RIS reflection ϕ, and blocklength M
2: Set maximum error tolerance ϵ^b
3: Initialize $\chi_{\min}$ and $\chi_{\max}$.
4: **while** $\chi_{\max} - \chi_{\min} \geq \epsilon^b$ **do**
5: $\quad$ Solve problem (25) with $\bar{\chi} = \frac{\chi_{\min} + \chi_{\max}}{2}$.
6: $\quad$ **if** problem (25) is feasible **then**
7: $\quad\quad W^\dagger \leftarrow W_{\bar{\chi}}$.
8: $\quad\quad \chi_{\min} \leftarrow \bar{\chi}$.
9: $\quad$ **else**
10: $\quad\quad \chi_{\max} \leftarrow \bar{\chi}$.
11: $\quad$ **end if**
12: **end while**
13: **for** $k = 1, 2, \cdots, K$ **do**
14: $\quad$ **for** $c = 1, 2, \cdots, C$ **do**
15: $\quad\quad$ Generate a Gaussian random vector $\tilde{w}_k^{(c)} \sim \mathcal{CN}(\mathbf{0}, W_k^\dagger)$.
16: $\quad\quad$ Obtain the objective value $\chi^{(c)}$ of (25) with $w_k = w_k^{(c)}$.
17: $\quad$ **end for**
18: $\quad$ Find the best candidate $c^\dagger = \arg\max_{1 \leq c \leq C} \chi^{(c)}$.
19: $\quad w_k^\dagger = w_k^{(c^\dagger)}$
20: **end for**
21: **Output:** $w^\dagger = \{w_k^{(\dagger)}, \forall k \in \mathcal{K}\}$

---

2) Receiving Beamforming Matrix Optimization: For given power allocation, blocklength, and RIS reflection, the beamforming vectors $w = \{w_k, \forall k \in \mathcal{K}\}$ can be optimized by solving

$$\textbf{P5}: \quad \max_{w, \chi} \chi \tag{24}$$
$$\text{s.t.} \quad (13c), (15a).$$

As (24) is non-convex due to the non-convex constraint (15a), we adopt the SDR technique to obtain an efficient approximate solution. Denote the combined channel $q_k = H^H \text{diag}(\phi) f_k$. By introducing $Q_k = q_k q_k^H$ and $W_k = w_k w_k^H$, the term $|w_k^H H^H \text{diag}(\phi) f_k|^2$ can be written as $\text{Tr}(W_k Q_k)$, where W_k needs to satisfy $W_k \succeq 0$ and $\text{rank}(W_k) = 1$. Since the constraint $\text{rank}(W_k) = 1$ is nonconvex, we relax this constraint and (24) can be transformed as the following problem

$$\textbf{P5-1}: \quad \max_{W, \chi} \chi \tag{25}$$
$$\text{s.t.} \quad \eta_k \left(\sum_{\pi(k') > \pi(k)} \text{Tr}(W_{k'} Q_{k'}) p_{k'} + \sigma^2 \right)$$
$$\leq \text{Tr}(W_k Q_k) p_k, \forall k \in \mathcal{K}, \tag{25a}$$
$$W_k \succeq 0, \forall k \in \mathcal{K}, \tag{25b}$$
$$\text{Tr}(W_k) = 1, \forall k \in \mathcal{K}, \tag{25c}$$

where $W = \{W_k, \forall k \in \mathcal{K}\}$. However, (25) is still non-convex due to the non-convex constraint (25a) and (25b). We first drop the rank-one constraint (25b). The bisection-search method is used to solve the coupling problem between χ and W, and then the SDR technique is used to obtain the optimal W. Specifically, with certain χ_{max} and χ_{min}, we replace χ in η_k in (25) by $\frac{\chi_{max} + \chi_{min}}{2}$, and solve the resulting feasibility problem reduced from (25) as

$$\textbf{P5-2}: \quad \text{find } W = \{W_k, \forall k \in \mathcal{K} | W_k \in \mathbb{S}_+^{N_t}\} \tag{26}$$
$$\text{s.t.} \quad (25a), (25b), (25c). \tag{26a}$$

The determination of the updated values for χ_{max} and χ_{min} is contingent upon the identification of a feasible solution W. It can be verified that when χ_{max} is sufficiently big and χ_{min} is sufficiently small, the bisection search described above for χ has the capability to produce a globally optimum matrix

W that is connected to beamforming in the i-th alternative iteration. Next, it is necessary to contemplate the process of transforming the optimal solution W^* into a viable solution $\{\hat{w}_k\}$ for the problem (26). The eigenvalue decomposition approach may be used to derive the optimum beamforming vectors $\{w_k^*\}$ when the optimal solutions W^* of problem (26) adhere to the rank-one constraint. The Gaussian randomization approach is employed to acquire the approximate solutions of equation (26). To begin, we build a random vector $\hat{w}_k$ that conforms to a complex normal distribution with mean 0 and variance W_k^*. Next, we utilize the operator $\hat{w}_k$ to address problem (26) and assess its viability. In this context, it is necessary to produce an ample quantity of feasible beamforming vectors $\hat{w}_k$, in an independent manner. Subsequently, the most optimal vector, denoted as $\hat{w}_k^*$, is selected from the pool of randomly generated vectors. Based on the above information, the final phase shift vector can be approximated as $w_k^* = \hat{w}_k^*$. The Algorithm 2 provides a step-by-step procedure for solving equation (24) using an iterative approach.

3) RIS Reflection Optimization: For given blocklength, receiving beamforming, and power allocation, the reflected vectors ϕ can be optimized by solving

$$\textbf{P6}: \quad \text{find } \phi \tag{27}$$
$$\text{s.t.} \quad (13b), (15a).$$

The SDR technique, which is comparable to the approach outlined in Section IV-B-2, can be used to solve problem (27). However, it takes a lot of time to solve the SDR problem, particularly when the matrix size is big. Therefore, we create

an effective method based on SCA as follows in order to lower the computational complexity and ensure performance. Denote $l_k = w_k^H H^H \text{diag}(f_k)$ and $L_k = l_k l_k^H$. By defining $S = \phi\phi^H \in \mathbb{S}_+^N$, With $\eta_k = 2^{\frac{\chi}{M \ln 2} + \frac{D_k}{M^2}} - 1$, Problem (P2) can be written as

$$\textbf{P6-1}: \quad \text{find } S \tag{28}$$

$$\text{s.t.} \quad \eta_k \left(\sum_{\pi(k') > \pi(k)} \text{tr}(L_{k'} S) p_{k'} + \sigma^2 \right) \leq \text{tr}(L_k S) p_k, \forall k \in \mathcal{K},$$
$$\tag{28a}$$

$$\text{rank}(S) = 1, \tag{28b}$$

$$[S]_{n,n} = 1, \ n = 1, 2, \cdots, N, \tag{28c}$$

$$S \succeq 0. \tag{28d}$$

The equivalent form of the rank-one constraint (28b) can be expressed as

$$||S||_* - ||S||_2 \leq 0, \tag{29}$$

where $||S||_* = \sum_i \sigma_i(S)$ and $||S||_2 = \sigma_1(S)$ denote the nuclear norm and spectral norm, respectively, and $\sigma_i(S)$ is the ith largest singular value of matrix S. For any $S \in \mathbb{S}_+^N$, we have $||S||_* - ||S||_2 \leq 0$ and the equality holds if and only if S is a rank-one matrix. However, (28) is still a nonconvex constraint. In order to take use of the penalty-based approach, we incorporate this restriction into the objective function of equation (28), which results in the optimization problem that follows:

$$\textbf{P6-2}: \quad \min_S \frac{1}{2\mu}(||S||_* - ||S||_2) \tag{30}$$

$$\text{s.t.} \quad (28a), (28c), (28d).$$

where μ is a penalty factor for (29). As the result, using a modest μ will yield the minimum value of (29). It is important to remember that even if the rank-one constraint is relaxed in (30), the answer obtained by solving (30) is an the best answer for $\mu \to 0$. Conversely, if μ is small enough, the solution to (30) is a rank-one solution. However, because of the objective function's DC form, (30) is not yet convex. We specify a lower bound of $||S||_2$, in order to address the DC form, and this may be done by

$$||S||_2 \geq \bar{S}^{(n)} = ||S^{(n)}||_2 \\ + \text{tr}[u_{max}(S^{(n)})(u_{max}(S^{(n)}))^H(S - S^{(n)})], \tag{31}$$

where $u_{max}(S^{(n)})$ denotes the eigenvector corresponding to the largest eigenvalue of $S^{(n)}$. Accordingly, the optimization problem can be written as follows:

$$\textbf{P6-3}: \quad \min_S \frac{1}{2\mu}(||S||_* - ||\bar{S}^{(n)}||) \tag{32}$$

$$\text{s.t.} \quad (28a), (28c), (28d).$$

It is noted that the problem in (32) is convex, meaning optimization solvers like CVX may be used to solve it well. Algorithm 3 provides a summary of the suggested penalty-based strategy for resolving the issue in (32).

Algorithm 3 SCA-Based Algorithm for RIS Reflection Optimization

1: **Input:** transmission power p, beamforming matrix $\{w_k\}$ and blocklength M
2: Set maximum number of iterations I_{max}^r and the maximum error tolerance ϵ^r
3: Initialize $S^{(0)}$. Set $i = 0$
4: **repeat**
5: $\quad i \leftarrow i + 1$.
6: $\quad$ Obtain $S^{(i)}$ by solving problem (32),
7: **until** convergence or $(i \geq I_{max}^r)$
8: Decompose $S^{(i)} = \phi^{(i)}(\phi^{(i)})^H$
9: **Output:** $\phi^o = \phi^{(i)}$.

---

4) Blocklength Optimization: For a given power allocation p, the beamforming vectors w and reflect vector ϕ, we have the problem:

$$\textbf{P7}: \quad \max_{\chi, M} \chi \tag{33}$$

$$\text{s.t.} \quad (13e), (13f), (15a).$$

It is observed that the problem in (33) is a convex optimization problem that the CVX optimization toolbox can efficiently solve. Since the integer constraint (13h) was relaxed in problem **P1**, we adopt a greedy search method to solve the integer conversion problem [29]. Firstly, we initialize the integer solution as a floor function of M^2. Next, we keep on allocating one blocklength to the user as long as the constraints (13e), (13f), (15a) are met.

C. User Pairing

Existing research proved that the NOMA systems with fixed power allocation schemes could perform better when pairing two users in distinct channel conditions. Then we consider the user pairing where one group has one strong user and one weak user. With the given power allocation, beamforming, phase shift, and blocklength optimization, problem **P1** can be transformed to

$$\textbf{P8}: \quad \max_b \chi \tag{34}$$

$$\text{s.t.} \quad (13d), (13g), (15a).$$

The transformed problem can be solved by the iterative bisection method to maximize the variable χ. In each iteration, i, with the given χ_i, the pairing scheme can be viewed as a matching problem in a bipartite graph $\Omega_i(U_s, U_w, E)$, where U_s, U_v are the strong and weak users, respectively and E is the edges with vertexes from U_s and U_v. The weight $e_{k,k'}$ of the edge between user $k \in U_s$ and user $k' \in U_w$ is based on the value of g_k and $g_{k'}$ if k and k' are in the same group. Specially, $e_{k,k'}$ is equal to $\min\{g_k, g_{k'}\}$ if $\min\{g_k, g_{k'}\} \geq \chi_i$ and 0, otherwise. Then, the maximum match of the bipartite graph $\Omega_i(U_s, U_w, E)$ with the given χ_i threshold is determined by employing the Hungarian algorithm [33]. The process is iterative until the difference between the upper bound χ_U and lower bound χ_L falls below the iteration accuracy ϵ^u or the maximum

---

Algorithm 4 Bisection and Hungarian Matching-Based User Pairing Algorithm

---

1: **Input:** transmission power $\boldsymbol{p}$, beamforming matrix $\{\boldsymbol{w}_k\}$, RIS reflection $\boldsymbol{\phi}$, blocklength M, and the corresponding χ^o.
2: Set maximum number of iterations $I_{\max}^u$ and the maximum error tolerance ϵ^u
3: Compute the weight matrix $\boldsymbol{\Omega}$ where each element is $e_{k,k'} = \min\{g_k, g_{k'}\}$ between the strong user k and weak user k'.
4: Initialize χ_H and χ_L.
5: Set $i = 1$
6: **while** $\chi_H - \chi_L \geq \epsilon^u$ or $i \leq I_{\max}^u$ **do**
7:     Set $\chi_i = \frac{\chi_H + \chi_L}{2}$
8:     Build the edge V between a strong user k and a weak user k' if $e_{k,k'} \geq \chi_i$, generate the bipartite graph Ω
9:     Obtain the maximum match Ω_{max}^i of $\Omega_i(U_s, U_w, E)$ by employing the Hungarian algorithm
10:     **if** Ω_{max}^i is not a perfect match **then**
11:         $\chi_H = \chi_i$.
12:     **else**
13:         $\chi_L = \chi_i$.
14:     **end if**
15:     $i \leftarrow i + 1$
16: **end while**
17: **Output:** $\Omega_{max}^o = \Omega_{max}^i$.

---

match turns into a perfect match, or the predefined maximum number of iterations is reached. Based on the matching when the iteration is converged, the NOMA groups are obtained, each of which is a pair of a strong user and a weak user. The Hungarian matching-based user pairing optimization is summarized in Algorithm 4.

D. Overall Algorithm, Convergence and Complexity Analysis

1) Overall Algorithm: The overall algorithm is summarized in Algorithm 5, which consists of three steps: design of decoding order and random grouping, resource allocation,

2) Convergence: Firstly, we prove the convergence of step 2. For given decoding order $\boldsymbol{\pi}$ and random grouping, we have

$$\chi(\boldsymbol{p}^{(i-1)}, \boldsymbol{w}^{(i-1)}, \boldsymbol{\phi}^{(i-1)}, M^{(i-1)})$$
$$\overset{(a)}{\leq} \chi(\boldsymbol{p}^{(i)}, \boldsymbol{w}^{(i-1)}, \boldsymbol{\phi}^{(i-1)}, M^{(i-1)})$$
$$\overset{(b)}{\leq} \chi(\boldsymbol{p}^{(i)}, \boldsymbol{w}^{(i)}, \boldsymbol{\phi}^{(i-1)}, M^{(i-1)}) \qquad (35)$$
$$\overset{(c)}{\leq} \chi(\boldsymbol{p}^{(i)}, \boldsymbol{w}^{(i)}, \boldsymbol{\phi}^{(i)}, M^{(i-1)})$$
$$\overset{(d)}{\leq} \chi(\boldsymbol{p}^{(i)}, \boldsymbol{w}^{(i)}, \boldsymbol{\phi}^{(i)}, M^{(i)}).$$

where $\chi(\boldsymbol{p}^{(i-1)}, \boldsymbol{w}^{(i-1)}, \boldsymbol{\phi}^{(i-1)}, M^{(i-1)})$ is the value of the objective function of problem (15) at iteration $(i-1)$ and (a), (b), (c), (d) results from Algorithm 1, 2, 3 and 4, respectively. As a result, the objective function is non-decreasing after each iteration. Moreover, since the objective is continuous over the

---

Algorithm 5 Proposed Three-Step Algorithm for Minimizing Maximal Decoding Error Probability

---

1: - **Step 1: Design of decoding order and random user pairing**
2: Solve problem (17) to obtain the combined channel gain of each user $|\boldsymbol{H}^H \text{diag}(\boldsymbol{\phi}) \boldsymbol{f}_k|^2$.
3: Obtain $\boldsymbol{\pi} = \{\pi(k), \forall k \in \mathcal{K}\}$ according to derived $|\boldsymbol{H}^H \text{diag}(\boldsymbol{\phi}) \boldsymbol{f}_k|^2$ and $K/2$ users with the most effective channel gain as the strong users and the remaining $K/2$ users as the weak users.
4: Randomly divide K users into $K/2$ groups, each has one strong user and one weak user.
5: - **Step 2: Joint power allocation, receiving beamforming, RIS reflection, and blocklength optimization**
6: Set maximum number of iterations I_{max}, and iteration index $i = 0$.
7: Initialize $\boldsymbol{p}^{(0)}$, $\boldsymbol{w}^{(0)}$, $\boldsymbol{\phi}^{(0)}$ and $M^{(0)}$.
8: **repeat**
9:     $i \leftarrow i + 1$
10:     Given $M^{(i-1)}$, $\boldsymbol{w}^{(i-1)}$ and $\boldsymbol{\phi}^{(i-1)}$, solve problem (23) by applying Algorithm 1, and obtain $\boldsymbol{p}^{(i)} = \boldsymbol{p}^o$ and χ^o.
11:     Given $\boldsymbol{p}^{(i)}$, $M^{(i-1)}$ and $\boldsymbol{\phi}^{(i-1)}$, solve problem (25) by applying Algorithm 2 with $\chi_{\min} = \chi^o$, and obtain $\boldsymbol{w}^{(i)}$.
12:     Given $\boldsymbol{p}^{(i)}$, $M^{(i-1)}$ and $\boldsymbol{w}^{(i)}$, solve problem (32) by applying Algorithm 3, and obtain $\boldsymbol{\phi}^{(i)}$.
13:     Given $\boldsymbol{p}^{(i)}$, $\boldsymbol{w}^{(i)}$ and $\boldsymbol{\phi}^{(i)}$, solve problem (33) by using CVX, and obtain $M^{(i)}$.
14: **until** convergence or ($i \geq I_{max}$)
15: - **Step 3: User pairing optimization**
16: Given $\boldsymbol{p}^{(0)}$, $\boldsymbol{w}^{(0)}$, $\boldsymbol{\phi}^{(0)}$ and $M^{(0)}$, solve problem (33) by Algorithm 4, and obtain $\boldsymbol{b}$.

---

compact feasible set of problem **P1**, it is upper-bounded by some finite positive number [34]. Hence, Step 2 is ensured to converge. Step 3 involves both bisection and Hungarian, and then it converges.

3) Complexity Analysis: The computational complexity of Algorithm 5 is analyzed as follows. Step 1 deals with solving the semidefinite program (SDP) problem, then, the complexity of Algorithm (15) is $\mathcal{O}(M^{4.5})$ [35]. In the step 2, the complexities of Algorithm 1, 2, 3 are $\mathcal{O}(I_{\max}^p (K + 1)^3)$, $\mathcal{O}((M + 1)^{4.5} \log(1/\epsilon^b))$, $\mathcal{O}(I_{max}^r (M + 1)^{4.5})$, respectively. The complexity of obtaining optimal blocklength is $\mathcal{O}(I_M)$, where I_M is the number of iterations needed for convergence. Thus, the complexity of step 2 is $\mathcal{O}(I_{\max}(I_{\max}^p (K + 1)^3 + (M + 1)^{4.5} \log(1/\epsilon^b) + \mathcal{O}(I_{max}^r (M + 1)^{4.5})))$. he computational complexity of the iterative bisection approach in step 3 is $\mathcal{O}(\log(1/\epsilon^u))$ while the Hungarian method's computational complexity in each iteration is $\mathcal{O}(\frac{K^3}{8})$ [36]. Thus, the computational complexity of Algorithm 4 in step 3 is $\mathcal{O}(\frac{K^3}{8} \log(1/\epsilon^u))$. Totally, the complexity of Algorithm 5 is $\mathcal{O}(M^{4.5} + I_{\max}(I_{\max}^p (K + 1)^3 + (M + 1)^{4.5} \log(1/\epsilon^b) + \mathcal{O}(I_{max}^r (M + 1)^{4.5})) + \frac{K^3}{8} \log(1/\epsilon^u))$.

V. SIMULATION RESULTS

This section presents numerical findings that illustrate how well the suggested NOMA-TDMA transmission system performs while working with RIS to reduce the maximum error coding error. We examine a coordinate system that is three-dimensional and has the axes (x, y, z). We assume that the RIS and BS have elevations of 10 m, and that their horizontal coordinates are $(0, 0)$ and $(5, 15)$. The users are uniformly and randomly dispersed across the rectangular region created by the lines $x = 90, x = 190, y = -10$, and $y = 10$. Their elevations are set to 0 m. In every simulation, user coordinates are first generated at random and then fixed. In addition, we assume $\sigma^2 = -110$ dBm for the noise power at the BS and $p_{max} = 0.3$ W for the maximum transmit power of the user. The channel models include path loss and Rayleigh fading. The path loss exponents for the BS-RIS link and RIS-user are set to $\alpha_1 = 2.2, \alpha_2 = 2.6$, and the corresponding Rician factors are $K_1 = K_2 = 5$ dB. The maximum symbols of each user in the one-time slot $\frac{T}{T_{symbol}}$ is 20.

In Fig. 2, a comparison is made between the worst-case error probability of the proposed scheme and many alternative baselines. These baselines include random phase, pure NOMA, NOMA without Hungarian pairing, location-based successive interference cancellation (SIC), and equal power. The comparison is conducted under the conditions where $K = 4, N_t = 3$, and $E_0 = 10$ W. In the context of pure NOMA, it is assumed that all four users are belonging to a single group, which simultaneously broadcasts their signals to the base station (BS). In the location-based SIC system, users are categorized into two categories according to the proximity of their locations to the base station (BS). Furthermore, power is distributed evenly in the equal power scheme. In contrast, in the Hungarian pairing scheme, power allocation, receiving beamforming, RIS reflection, and blocklength optimization are concurrently optimized, but users are paired randomly. The initial observation reveals that the worst-case decoding error in all schemes exhibits a decrease as the parameter N increases. This may be attributed to the fact that a bigger size of reconfigurable intelligent surface (RIS) can leverage the potential advantage provided by a higher reflecting array. Furthermore, the proposed scheme has the lowest minimum-maximum error when compared to the six designs that have been taken into consideration. This is due to its ability to enhance power optimization, receiving beamforming, RIS reflection, and user pairing. Furthermore, the proposed scheme demonstrates a reduction in the worst-case decoding error when compared to the pure NOMA scheme. This improvement may be attributed to the fact that the proposed method encounters a lower level of inter-group interference in comparison to the pure NOMA scheme. The results indicate that the NOMA-TDMA scheme has superior performance compared to pure NOMA in terms of minimizing the maximal decoding error.

Fig. 3 shows the worst-case error decoding error when the number of element N varies under different total energy constraints E_0. It is observed that the decoding error in the worst-case scenario lowers as the values of N and E_0 grow. The increase in power allocation to users as E0 grows leads
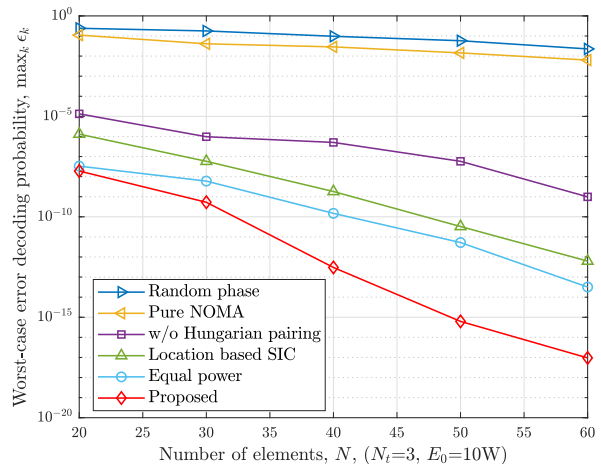

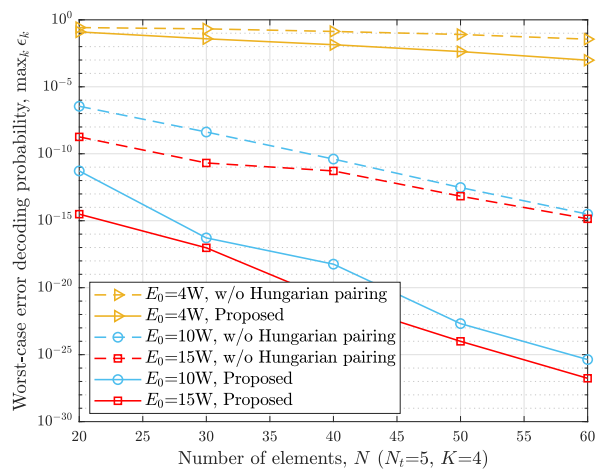Fig. 2: Optimized min-max decoding-error probability under different schemes.

Fig. 3: Optimized min-max decoding-error probability versus the number of elements of the RIS when E_0 varies.

to an increase in the signal-to-interference-plus-noise ratio (SINR) of users in the worst-case scenario. Consequently, the performance of the min-max decoding algorithm deteriorates. The efficiency of the suggested Hungarian matching-based user pairing is demonstrated in comparison to random pairing.

Fig. 4 shows the worst decoding error probability when the number of elements in the reconfigurable intelligent surface (RIS), denoted as N, is varied across different numbers of base station (BS) antennas, denoted as N_t. When N_t increases, there are higher receiving-beamforming gains. That is the reason why the worst-case decoding error goes down when N_t goes up. It has also been shown that increasing the number of elements N in the reconfigurable intelligent surface (RIS) has a positive impact on improving the signal-to-interference-plus-noise ratio (SINR) for the worst-case users. Increasing the value of N results in more reflection beamforming gains, hence reducing the min-max decoding error.

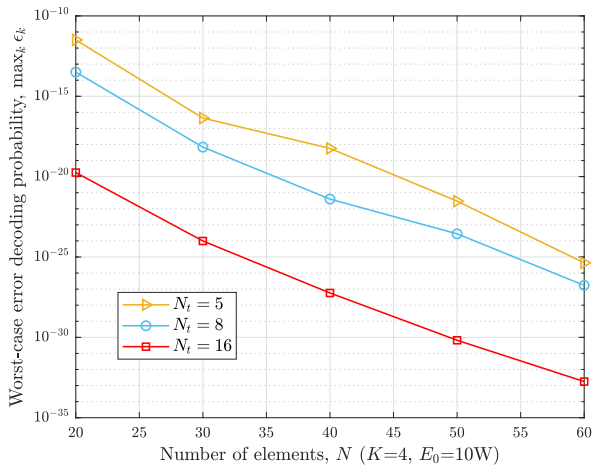In Fig. 5, we can witness the variation in the worst-

Fig. 4: Optimized min-max decoding-error probability versus the number of elements of the RIS when N_t varies.
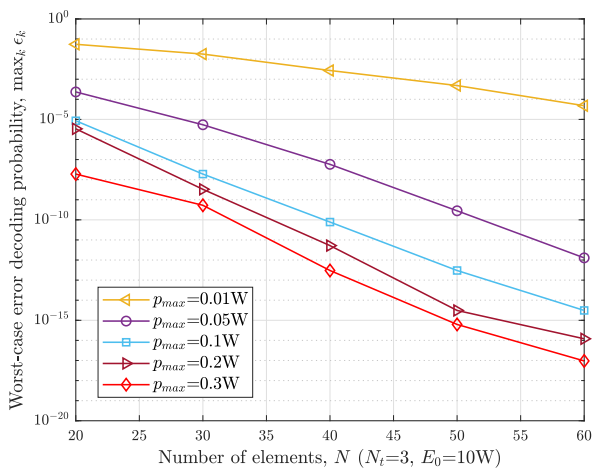

Fig. 6: Optimized min-max decoding-error probability versus a number of users when N_t varies.

Fig. 5: Optimized min-max decoding-error probability versus the number of elements of RIS when p_{max} varies.
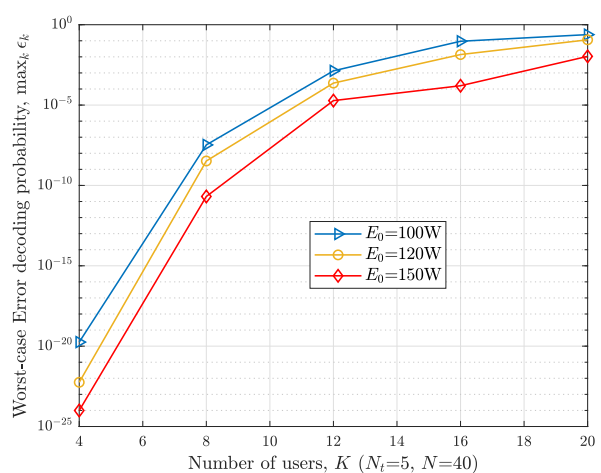
Fig. 7: Optimized min-max decoding-error probability versus a number of users when E_0 varies.

case error decoding error as the number of reconfigurable intelligent surface (RIS) elements changes while considering different maximal power levels of the user, denoted as p_{max}. It shows that when p_{max} increases, the worst-case decoding error decreases. That is because higher SINR can be obtained by users when p_{max} increases. Fig. 6 depicts the variation in worst-case error decoding error as the number of users changes for different numbers of antennas at the base station N_t. In addition, Fig. 7 shows the worst-case error decoding error when the number of users K varies under various total energy limitations E_0. Both Fig. 6 and Fig. 7 show that the worst-case error decoding error increases as the value of K grows.

VI. CONCLUSION

This paper investigated a RIS-empowered uplink hybrid NOMA-TDMA communication system under short packet transmission. We formulated an optimization problem to minimize t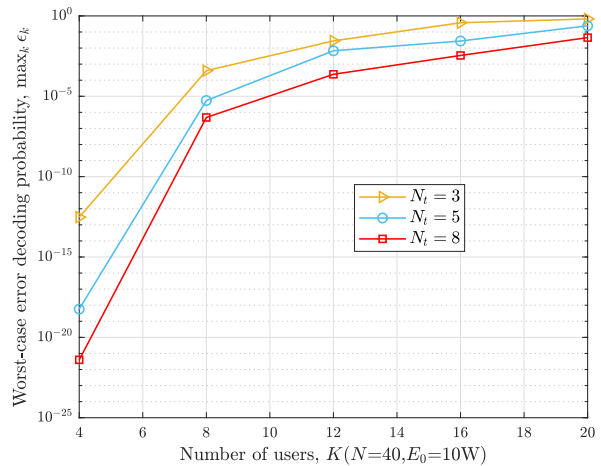he maximum decoding error probability. Then, we proposed an algorithm to jointly optimize the users' transmit power, receiving beamforming vector at the BS, RIS reflection, blocklength, and user pairing. Numerical results have shown that the proposed algorithm can perform better than benchmark schemes, and the proposed system outperforms the pure NOMA system. In future work, we may target other metrics, such as sum rate maximization and average sum of age of information (AoI) [37]. Since we assume that the channel gain is fixed during the total transmission time in this work, the TDMA time slot duration optimization can be added in future work.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, May 2019.

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

[3] Y. Liu, X. Mu, X. Liu, M. Di Renzo, Z. Ding, and R. Schober, "Reconfigurable intelligent surface-aided multi-user networks: Interplay between noma and ris," *IEEE Wireless Commun.*, vol. 29, no. 2, pp. 169–176, Apr. 2022.

[4] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[5] X. Li, Z. Xie, Z. Chu, V. G. Menon, S. Mumtaz, and J. Zhang, "Exploiting benefits of irs in wireless powered noma networks," *IEEE Trans. Green Commun. Networking*, vol. 6, no. 1, pp. 175–186, March 2022.

[6] J. Yaswanth, S. K. Singh, K. Singh, and M. F. Flanagan, "Energy-efficient beamforming design for ris-aided mimo downlink communication with swipt," *IEEE Trans. Green Commun. Networking*, May 2023.

[7] H. Xie, J. Xu, Y.-F. Liu, L. Liu, and D. W. K. Ng, "User grouping and reflective beamforming for irs-aided urllc," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2533–2537, 2021.

[8] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, "Deep reinforcement learning for practical phase shift optimization in ris-aided miso urllc systems," *IEEE Internet Things J.*, 2022.

[9] W. R. Ghanem, V. Jamali, and R. Schober, "Joint beamforming and phase shift optimization for multicell irs-aided ofdma-urllc systems," in *Proc. IEEE ICC*, Nanjing, China, 2021, pp. 1–7.

[10] M. N. Tehrani and S. Farahmand, "Resource allocation for irs-enabled secure multiuser multi-carrier downlink urllc systems," pp. 1–7, 2022.

[11] R. Kotaba, C. N. Manchon, N. M. K. Pratas, T. Balercia, and P. Popovski, "Improving spectral efficiency in urllc via noma-based retransmissions," in *Proc. IEEE ICC*, Shanghai, China, 2019, pp. 1–7.

[12] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint power and blocklength optimization for urllc in a factory automation scenario," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1786–1801, Mar. 2019.

[13] J. Chen, L. Zhang, Y.-C. Liang, and S. Ma, "Optimal resource allocation for multicarrier noma in short packet communications," *IEEE Trans. Veh. Tech.*, vol. 69, no. 2, pp. 2141–2156, Feb. 2019.

[14] S. Han, X. Xu, Z. Liu, P. Xiao, K. Moessner, X. Tao, and P. Zhang, "Energy-efficient short packet communications for uplink noma-based massive mtc networks," *IEEE Trans. Veh. Tech.*, vol. 68, no. 12, pp. 12 066–12 078, Dec. 2019.

[15] W. Ahsan, W. Yi, Y. Liu, and A. Nallanathan, "A reliable reinforcement learning for resource allocation in uplink noma-urllc networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 5989–6002, Aug. 2022.

[16] R. Deshpande, M. Katwe, K. Singh, and Z. Ding, "Resource allocation design for spectral-efficient urllc using ris-aided fd-noma system," *IEEE Wireless Commun. Lett.*, 2023.

[17] T.-H. Vu, T.-V. Nguyen, D. B. da Costa, and S. Kim, "Intelligent reflecting surface-aided short-packet non-orthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4500–4505, 2022.

[18] T.-H. Vu, T.-V. Nguyen, Q.-V. Pham, D. B. da Costa, and S. Kim, "Star-ris-enabled short-packet noma systems," *IEEE Trans. Veh. Technol.*, 2023.

[19] C. Zhang, M. Cui, and G. Zhang, "Throughput optimization for irs-assisted multi-user noma urllc systems," *Wireless Networks*, pp. 1–13, 2023.

[20] C. Xia, Z. Xiang, B. Su, H. Liu, J. Meng, and G. Pan, "Ris-noma-assisted short-packet communication with hardware impairments," *IEEE Internet Things J.*, 2023.

[21] Y. Yang and M. C. Gursoy, "Energy-efficient scheduling in ris-aided mec networks with noma and finite blocklength codes," in *Proc. ISWCS*. IEEE, 2022, pp. 1–6.

[22] J. Yang, G. Liu, J. Ren, Y. Liu, L. Yao, Y. Zhou, and J. Chen, "Resource allocation for intelligent reflecting surface-assisted cooperative noma-urllc networks in smart grid," in *Proc. IEEE SmartGridComm*. IEEE, 2022, pp. 83–89.

[23] Y. Zhu, X. Yuan, Y. Hu, T. Wang, M. C. Gursoy, and A. Schmeink, "Low-latency hybrid noma-tdma: Qos-driven design framework," *IEEE Trans. Wireless Commun.*, 2022.

[24] H. Al-Obiedollah, H. B. Salameh, K. Cumanan, Z. Ding, and O. A. Dobre, "Self-sustainable multi-irs-aided wireless powered hybrid tdma-noma system," *IEEE Access*, 2023.

[25] J. Lei, T. Zhang, and Y. Liu, "Hybrid noma for star-ris enhanced communication," *IEEE Trans. Veh. Technol.*, 2023.

[26] W. Lyu, Y. Xiu, X. Li, S. Yang, P. L. Yeoh, Y. Li, and Z. Zhang, "Hybrid noma assisted integrated sensing and communication via ris," *arXiv preprint arXiv:2309.06080*, 2023.

[27] G. Chen, Q. Wu, C. He, W. Chen, J. Tang, and S. Jin, "Active irs aided multiple access for energy-constrained iot systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1677–1694, 2022.

[28] O. B. Gamgam and E. Karasan, "Satisfying strict deadlines for cellular internet of things through hybrid multiple access," *Trans. Emerging Telecommun. Technol.*, vol. 34, no. 8, p. e4814, 2023.

[29] A. A. Nasir, "Min-max decoding-error probability-based resource allocation for a urllc system," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2864–2867, 2020.

[30] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[31] B. Yang, X. Cao, C. Huang, C. Yuen, L. Qian, and M. Di Renzo, "Intelligent spectrum learning for wireless networks with reconfigurable intelligent surfaces," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3920–3925, May 2021.

[32] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optimzation*, vol. 47, pp. 29–51, 2010.

[33] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics (NRL)*, vol. 52, no. 1, pp. 7–21, 2005.

[34] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[35] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[36] T. Kim and M. Dong, "An iterative hungarian method to joint relay selection and resource allocation for d2d communications," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 625–628, 2014.

[37] Z. Gao, A. Liu, C. Han, and X. Liang, "Non-orthogonal multiple access-based average age of information minimization in leo satellite-terrestrial integrated networks," *IEEE Trans. Green Commun. Networking*, vol. 6, no. 3, pp. 1793–1805, Sep. 2022.