Hybrid NOMA for STAR-RIS Enhanced Communication

Jiayi Lei, Tiankui Zhang, *Senior Member, IEEE,* Yuanwei Liu, *Senior Member, IEEE*

Abstract—In this paper, a hybrid non-orthogonal multiple access (NOMA) framework for the simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) enhanced cell-edge communication is investigated. Specifically, one transmitted user and one reflected user are paired as one NOMA-pair, while multiple NOMA-pairs are served via time division multiple access (TDMA). The objective is to maximize the minimum downlink rate by jointly optimizing the user pairing, decoding order, passive beamforming, power and time allocation. A novel two-layer iterative algorithm is proposed to solve the highly coupled problem. Simulation results show that: 1) the proposed framework outperforms the conventional reflecting-only-RIS-based and the OMA-based frameworks; 2) the beamforming design and power allocation dominate the achieved performance; 3) increasing the number of passive elements and shortening the distance between BS and STAR-RIS are two effective ways to further improve the performance.

Index Terms—Hybrid NOMA, resource allocation, STAR-RIS.

I. Introduction

In traditional wireless communication networks, cell-edge users always suffer from poor quality-of-service (QoS) due to severe channel fading and inter-cell interference. For such practical scenarios, the reconfigurable intelligent surface (RIS) is a promising technology to enhance communications for cell-edge users. With the ability to smartly reconfigure the wireless propagation environment, the RIS is able to provide additional and high-quality transmission links [1]. In addition, thanks to its small size, light weight, and high extensibility, the RIS is easy to be deployed in existing networks. However, the conventional RIS is designed to only reflect the incident signals, making the source and the destination have to be on the same side of the RIS. To address this issue, a novel simultaneous transmitting and reflecting RIS (STAR-RIS) is proposed in [2]. The incident signal can be transmitted and reflected by the STAR-RIS simultaneously, and thus a full-space smart radio environment is created. With apparent advantages, the STAR-RIS aided networks have attracted widespread attention [3, 4]. In [3], the coverage probability of a STAR-RIS assisted massive multiple-input multiple-output (MIMO) system was analyzed with Rayleigh fading and phase-shift errors. In [4], the transmit power for a STAR-RIS aided multiple-input single-output system was minimized by joint active and passive beamforming design.

Jiayi Lei and Tiankui Zhang are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: {leijiayi,zhangtiankui }@bupt.edu.cn).

Yuanwei Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk).

Non-orthogonal multiple access (NOMA) is another promising technology for future wireless communications. It has been proved that NOMA yields a significant gain over conventional orthogonal multiple access techniques in terms of spectral efficiency and user fairness [5]. Significantly, combining NOMA and STAR-RIS is a meaningful research topic. On the one hand, the STAR-RIS is able to reconfigure channels smartly and introduce additional degrees-of-freedom (DoFs) for system design, thus enhancing the NOMA gain. On the other hand, due to the simultaneous transmission and reflection characteristics of STAR-RIS, users are naturally divided into transmission users and reflection users, which provides a basis for user pairing in NOMA. Inspired by this, there emerge some works which combine the two techniques [6–8]. In [6], the effective capacity of a STAR-RIS assisted NOMA network with two users on different sides of the STAR-RIS was studied. In [7], the authors investigated the secrecy performance of the STAR-RIS assisted NOMA networks. A sum rate maximization problem for STAR-RIS-NOMA systems was investigated in [8], where the decoding order, power allocation, active beamforming, and passive beamforming were jointly optimized.

Although some excellent works on STAR-RIS assisted NOMA networks have been conducted, most of them considered pure NOMA, which will face the following bottleneck. With a large number of users, it is necessary to perform user clustering to ensure the effectiveness of NOMA. In this case, users served by pure NOMA suffer from not only intra-cluster interference but also inter-cluster interference, which makes a severely negative effect on QoS and increases the complexity of network design. Hybrid NOMA is one of the effective solutions to tackle this issue, but it has not yet received widespread attention [9–11]. In [9], the authors studied hybrid NOMA and time division multiple access (TDMA) for the uplink transmission in reflecting-only RIS assisted wireless powered communication networks. Similarly, a hybrid TDMA-NOMA scheme was developed to balance the performance and signalling overhead for RIS aided Internet-of-Things (IoT) system in [10]. However, as the functional limitation of reflecting-only RIS, only users located on one side of the RIS were considered in [9, 10]. Furthermore, few related works investigate hybrid NOMA for STAR-RIS aided networks [11]. A frequency division multiple access (FDMA)-NOMA mixed framework was applied in [11] to eliminate inter-cluster interference. However, it is worth noting that the passive beamforming at STAR-RIS is time-selective, but not frequency-selective. As a result, the advantages of the STAR-RIS can not be fully utilized.

Motivated by the above observation, a novel hybrid NOMA

framework for STAR-RIS assisted cell-edge networks is proposed in this paper, which reduces the complexity of network design while fully leveraging the advantages of STAR-RISs. The main contributions are summarized as follows: 1) We propose a novel hybrid NOMA-TDMA framework for STAR-RIS assisted cell-edge networks, where one transmitted user and one reflected user are paired and served via NOMA, while multiple NOMA-pairs are served by TDMA; 2) We formulate a max-min rate problem and solve it by a novel two-layer iterative algorithm, where user pairing, passive beamforming, power and time allocation are jointly optimized; 3) Numerical results show the superiority of the proposed network framework and the two-layer algorithm. Moreover, increasing the number of STAR-RIS elements and shortening the distance from STAR-RIS to BS are confirmed to be two effective ways to improve network performance.

Notation: Scalars, vectors and matrices are denoted by Italic letters, bold-face lower-case, and bold-face upper-case, respectively. For a complex-valued vector $\mathbf{a}$, $\mathbf{a}^{\mathrm{T}}$ means its transpose, $\mathbf{a}^{\mathrm{H}}$ means its conjugate transpose, and $diag(\mathbf{a})$ denotes a diagonal matrix with the elements of vector $\mathbf{a}$ on the main diagonal. Besides, $\|\cdot\|$ denotes a vector's Euclidean norm, and $arg(\cdot)$ denotes a complex number's argument.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model
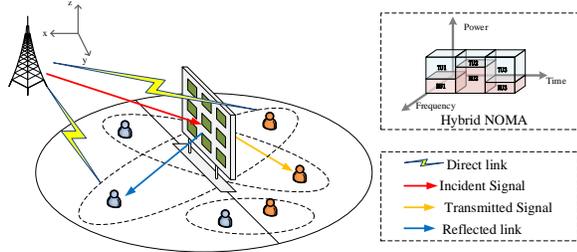
Fig. 1: STAR-RIS assisted communication networks.

As shown in Fig. 1, we consider a cell-edge area, where multiple users are randomly distributed, and a STAR-RIS is deployed in the center to enhance the communications. For the convenience of explanation, a 3D Cartesian coordinate system is established. Let $\mathbf{q}_S = [x_S, y_S, z_S]$, $\mathbf{q}_B = [x_B, y_B, z_B]$ and $\mathbf{q}_k = [x_k, y_k, 0]$ denote the 3D positions of the STAR-RIS first element, the base station (BS) and cell-edge users respectively. The BS and users are all equipped with a single antenna. The STAR-RIS consists of a uniform planar array (UPA) with $M = M_y M_z$ passive transmitting and reflecting elements, where M_y and M_z denote the number of elements along the $y-$ and $z-$axis, respectively. The STAR-RIS adopts the energy-splitting protocol, i.e., each element of the STAR-RIS splits the incident signal into transmitted and reflected signals to serve users at both sides of the surface simultaneously [1].

[1]There are three practical protocols for operating STAR-RISs: namely energy splitting (ES), mode switching (MS), and time switching (TS). Based on the insights obtained in [12], the ES protocol is the best option among the three protocols for commutations with high quality-of-service (QoS) requirements. Therefore, we adopt the ES protocol in this paper. Numerical simulation and comparison of the three protocols in STAR-RIS assisted wireless communications will be considered in our future study.

As the whole region is divided into the right-half sub-region and the left-half sub-region by the STAR-RIS, users can be divided into transmitted users (TU) and reflected users (RU), whose sets are denoted as $\mathcal{K}_T$ and $\mathcal{K}_R$ respectively. We assume that users are equally located in the two sub-regions, that is, $|\mathcal{K}_T| = |\mathcal{K}_R| = K$ [2].

We consider that one TU is paired with one RU, and then all users are grouped into K user-pairs. Let binary variable $c_{k_T, k_R}(c_{k_R, k_T}) \in \{0, 1\}, k_T \in \mathcal{K}_T, k_R \in \mathcal{K}_R$ denote whether the TU k_T and the RU k_R are paired up, which satisfies $\sum_{k_R=1}^{K} c_{k_T, k_R} = 1$. The two paired users receive data from the BS via non-orthogonal multiple access (NOMA). K user-pairs are served sequentially via TDMA. This multiple access scheme is referred to as hybrid NOMA in this paper. Given $k \in \mathcal{K}_T \cup \mathcal{K}_R$, let $\bar{k}$ denote the index of the user which is paired with k, i.e., $c_{k, \bar{k}} = 1$. Then, the time allocation coefficient of the user-pair $\{k, \bar{k}\}$ is denoted by $\tau_{\{k, \bar{k}\}} \in [0, 1]$, which is supposed to satisfy $\sum_{k \in \mathcal{K}_T(k \in \mathcal{K}_R)} \tau_{\{k, \bar{k}\}} = 1$.

As user-pairs occupy orthogonal time resource in the hybrid NOMA framework, the passive beamforming vectors of the STAR-RIS can be designed specifically and independently for each user-pair, so as to reconstruct the channel between the BS and the currently served user-pair more purposefully. Let $\mathbf{V}_k = \sqrt{\beta_k} diag\left(e^{j\theta_k^1}, e^{j\theta_k^2}, ..., e^{j\theta_k^M}\right)$ denote the transmission and reflection coefficient matrix of STAR-RIS for the user k, where $\sqrt{\beta_k} \in [0, 1]$ denotes the amplitude and θ_k^m denotes the phase shifts of m-th STAR-RIS element. We assume that there is no energy dissipated by STAR-RIS, that is, $\beta_k + \beta_{\bar{k}} = 1$.

B. Channel Model

There are two kinds of downlinks for each user, direct link and STAR-RIS assisted link. Let $h_{B,k} \in \mathbb{C}^{1 \times 1}$, $\mathbf{h}_{B,S} \in \mathbb{C}^{M \times 1}$ and $\mathbf{h}_{S,k} \in \mathbb{C}^{M \times 1}$ denote the channel vectors from the BS to user, from the BS to the STAR-RIS, and from the STAR-RIS to user respectively. Considering the blocked line-of-sight (LoS) link and potential extensive scattering from the BS to users, the propagation $h_{B,k} \in \mathbb{C}^{1 \times 1}$ is modeled as Rayleigh fading, while the transmission links from the BS to the STAR-RIS and from the STAR-RIS to users are modeled as LoS-dominated channels. Specifically, the channel coefficient of the direct link is expressed as

$$h_{B,k} = \sqrt{\frac{\delta_0}{d_{B,k}{}^{\alpha_1}}} \widetilde{h}_{B,k}. \tag{1}$$

where δ_0 is the channel power at the reference distance of 1m, α_1 is the path loss exponent, $d_{B,k} = \|\mathbf{q}_B - \mathbf{q}_k\|$ is the distance between the BS and the user k, and $\widetilde{h}_{B,k}$ is complex Gaussian random variable with zero mean and unit variance, i.e., $\widetilde{h}_{B,k} \sim \mathcal{CN}(0, 1)$.

The channel vectors from the BS to the STAR-RIS and from the STAR-RIS to users are given by

$$\mathbf{h}_{B,S} = \sqrt{\frac{\delta_0}{d_{B,S}{}^{\alpha_2}}} \mathbf{a}_{AoA}, \tag{2}$$

$$\mathbf{h}_{S,k} = \sqrt{\frac{\delta_0}{d_{S,k}{}^{\alpha_3}}} \mathbf{a}_{AoD}, \tag{3}$$

[2]The proposed network framework and algorithm in this paper can be extended to cases where the number of TUs is not equal to that of RUs, with small modifications on the design of user pairing.

where $d_{B,S} = \|\mathbf{q}_B - \mathbf{q}_S\|$ and $d_{S,k} = \|\mathbf{q}_S - \mathbf{q}_k\|$. $\mathbf{a}_{AoA}$ is the receive array response vector and given as $\mathbf{a}_{AoA} = e^{-j\frac{2\pi d_{U,S}}{\lambda}} \left[1, e^{-j\frac{2\pi d_y}{\lambda}\sin\varphi_r\cos\eta_r}..., e^{-j\frac{2(M_y-1)\pi d_y}{\lambda}\sin\varphi_r\cos\eta_r}\right]^{\mathrm{T}} \otimes$ $\left[1, e^{-j\frac{2\pi d_z}{\lambda}\sin\varphi_r\sin\eta_r}..., e^{-j\frac{2(M_z-1)\pi d_z}{\lambda}\sin\varphi_r\sin\eta_r}\right]^{\mathrm{T}}$, where φ_r and η_r are the zenith angle of arrival (AoA) and the azimuth AoA of the signal from the BS to the STAR-RIS respectively. $\mathbf{a}_{AoD}$ is the transmit array response vector and given as $\mathbf{a}_{AoD} = e^{-j\frac{2\pi d_{S,k}}{\lambda}} \left[1, e^{-j\frac{2\pi d_y}{\lambda}\sin\varphi_t\cos\eta_t}..., e^{-j\frac{2(M_y-1)\pi d_y}{\lambda}\sin\varphi_t\cos\eta_t}\right]^{\mathrm{T}} \otimes$ $\left[1, e^{-j\frac{2\pi d_z}{\lambda}\sin\varphi_t\sin\eta_t}..., e^{-j\frac{2(M_z-1)\pi d_z}{\lambda}\sin\varphi_t\sin\eta_t}\right]^{\mathrm{T}}$, where φ_t and η_t are the zenith angle of departure (AoD) and the azimuth AoD of the signal from the STAR-RIS to the user k respectively. d_y and d_z are the STAR-RIS element spacings along the y- and z-axis respectively, λ denotes the signal wavelength, and $\otimes$ denotes the Kronecker product.

For any user k, the downlink equivalent-combined channel gain is given as

$$h_k = h_{B,k} + \mathbf{h}_{S,k}{}^{\mathrm{H}}\mathbf{V}_k\mathbf{h}_{B,S}, \tag{4}$$

C. Hybrid NOMA signal model

As mentioned earlier, the two paired users are served simultaneously via NOMA, while K user-pairs are served sequentially via TDMA. Let $\pi(k) \in \{0,1\}$ denote the decoding order of user k. In downlink NOMA, to guarantee that successive interference cancelation (SIC) performs successfully, an optimal decoding order is to decode in the order of increasing equivalent-combined channel gain, regardless of the user power allocation, that is,

$$\pi(k) = \begin{cases} 0, if\ |h_k| \le |h_{\bar{k}}|, \\ 1, if\ |h_k| > |h_{\bar{k}}|. \end{cases} \tag{5}$$

Under the proposed hybrid NOMA framework, the signal-to-interference-plus-noise ratio (SINR) at user k is expressed as

$$\Gamma_k = \frac{|h_k|^2\rho_k P}{\pi(\bar{k})|h_k|^2\rho_{\bar{k}}P + n}, \tag{6}$$

where P is the total power of the BS. $\rho_k, \rho_{\bar{k}} \in [0,1]$ are the power allocation coefficients, which satisfy $\rho_k + \rho_{\bar{k}} = 1$. n is the complex circular i.i.d. additive Gaussian noise with $n \sim \mathcal{CN}(0,\sigma^2)$, where σ^2 is the noise power. Then, the downlink communication rate of the user is given by

$$R_k = \tau_{\{k,\bar{k}\}}\log_2(1+\Gamma_k) \tag{7}$$

D. Problem Formulation

Considering the communication fairness in cell-edge areas, the minimum transmission rate is taken as the performance metric of the proposed STAR-RIS assisted networks. The object is to maximize the minimum transmission rate among multiple users, by jointly optimizing the user pairing, the decoding order, the passive beamforming design, the user-pair time allocation and the user power allocation. The problem is formulated as follows,

$$(\text{P1}): \underset{\mathbf{V}_k,c,\pi_k,\tau_{\{k,\bar{k}\}},\rho_k}{\text{Max}} \underset{k\in\mathcal{K}_L\cup\mathcal{K}_R}{\text{Min}} R_k \tag{8a}$$

$$\text{s.t.} \quad \sqrt{\beta_k} \in [0,1], \beta_k + \beta_{\bar{k}} = 1, \tag{8b}$$

$$\theta_k^m \in [0,2\pi], \tag{8c}$$

$$c_{k_L,k_R} \in \{0,1\}, \sum_{k_R=1}^{K} c_{k_L,k_R} = 1, \tag{8d}$$

$$\pi(k) = \begin{cases} 0, if\ |h_k| \le |h_{\bar{k}}|, \\ 1, if\ |h_k| > |h_{\bar{k}}|, \end{cases} \tag{8e}$$

$$\tau_{\{k,\bar{k}\}} \in [0,1], \sum_{k\in\mathcal{K}_L(k\in\mathcal{K}_R)} \tau_{\{k,\bar{k}\}} = 1, \tag{8f}$$

$$\rho_k \in [0,1], \rho_k + \rho_{\bar{k}} = 1. \tag{8g}$$

(8b) and (8c) define the feasible ranges of STAR-RIS amplitude coefficients and phase shift respectively. Constraint (8d) illustrates the one-to-one mapping relationship between TUs and RUs. Constraint (8e) guarantees the SIC performs successfully. Constraint (8f) means the sum of the time allocation coefficients for user-pairs is 1 and constraint (8g) represents that the sum of the power allocation coefficients for two paired users should also be 1.

III. PROBLEM SOLUTION AND PROPOSED ALGORITHM

Evidently, the proposed optimization problem is nonconvex and highly coupled. Moreover, it is worth noting that the decoding order, beamforming vectors, power and time allocation can only be solved when the user pairing is given, due to the constraints (8b), (8e), (8f) and (8g). However, when the other variables are given, it is difficult to further optimize the user pairing through mathematical derivation. To tackle this issue, a novel iterative algorithm with a two-layer loop-nesting structure is proposed. Specifically, the outer layer is designed to determine the user pairing by a one-to-one swapping matching based approach. With given user pairing, the decoding order, passive beamforming, user-pair time allocation and user power allocation are solved by an alternating optimization (AO) based algorithm in the inner layer.

A. Outer layer: one-to-one matching based user-pairing

In this subsection, we describe the user pairing as a matching game between TUs and RUs, which is solved by a matching-theory based algorithm. First of all, some basic concepts are introduced as follows.

Definition 1. *(One-to-one Two-sides Matching):* A one-to-one matching Φ is defined as a function from $\mathcal{K}_T \cup \mathcal{K}_R$ to $\mathcal{K}_T \cup \mathcal{K}_R$ such that
1) $\Phi(k_T) \in \mathcal{K}_R, \Phi(k_R) \in \mathcal{K}_T$,
2) $|\Phi(k_T)| = |\Phi(k_R)| = 1$
3) $k_T = \Phi(k_R) \Leftrightarrow k_R = \Phi(k_T)$

Definition 2. *(Utility Function):* Given a matching state Φ, the utility function $W(\Phi)$ is defined as the possible maximum value of the minimum user rate, which is obtained in the subsequent inner layer. $W(\Phi) = \underset{\mathbf{V}_k,\pi_k,\tau_{\{k,\bar{k}\}},\rho_k}{\text{Max}} \underset{k\in\mathcal{K}_L\cup\mathcal{K}_R}{\text{Min}} R_k$.

Definition 3. *(Swapping Matching):* For matching state $\Phi(k_L) = k_R$ and $\Phi(\tilde{k}_L) = \tilde{k}_R$, a swap matching is $\Phi_{k_L}^{\tilde{k}_L} = \left\{\Phi\backslash\left\{(k_L,k_R),\left(\tilde{k}_L,\tilde{k}_R\right)\right\} \cup \left\{\left(k_L,\tilde{k}_R\right),\left(\tilde{k}_L,k_R\right)\right\}\right\}$.

Definition 4. *(Swap-blocking pair):* Given a matching state Φ with $\Phi(k_T) = k_R$ and $\Phi\left(\tilde{k}_T\right) = \tilde{k}_R$, $\left(k_T, \tilde{k}_T\right)$ is a swapping-blocking pair if and only if $W\left(\Phi_{k_T}^{\tilde{k}_T}\right) > W(\Phi)$.

Definition 1 indicates the one-to-one mapping relationship between TUs and RUs. Note that different from general matching algorithms, the utility function in Definition 2 is a global function, which makes the obtained matching state closer to the optimal solution, rather than just a stable matching. Definition 3 enables two TUs to exchange their paired RUs. Definition 4 implies that the utility function will increase with a swap-blocking pair. Based on the above definitions, the one-to-one swapping matching based user pairing in the outer layer is briefly described as: with a given initial matching state, the algorithm keeps searching for one TU and one RU to form a swapping-blocking pair and executes the swapping matching until there is no swapping-blocking pair.

B. Inner layer: AO based decoding order, passive beamforming, power and time allocation

In this subsection, we aim to obtain the utility function defined in the outer layer. With given user pairing, the problem for jointly optimizing the decoding order, the beamforming, the power and time allocation is expressed as

$$\text{(P2):} \quad \max_{S, \mathbf{V}_k, \pi_k, \tau_{\{k,\bar{k}\}}, \rho_k} S \tag{9a}$$

$$\text{s.t.} \quad S \leq R_k, k \in \mathcal{K}_T \cup \mathcal{K}_R, (8b), (8c), (8e) \sim (8g), \tag{9b}$$

where S is an introduced auxiliary variable which represents the minimum rate among users. To seek the solution of the still highly coupled nonconvex problem (P1), an AO based iteration algorithm is proposed, where one variable is solved with giving the other ones [3].

First of all, as considered in previous works [8, 13], the decoding order can be determined by the equivalent-combined channel gain when the other variables are given. Furthermore, the SIC constraint can be removed under the optimal decoding order $\pi^*(k) = \begin{cases} 0, if\ |h_k| \leq |h_{\bar{k}}|, \\ 1, if\ |h_k| > |h_{\bar{k}}|. \end{cases}$ This operation will not affect the solution of (P2) by iteratively updating the optimal decoding order and solving the problem (P2) without (8e).

Then, we focus on the passive beamforming design at the STAR-RIS. It is known that the beamforming is expected to maximize the equivalent-combined channel gain. Based on the triangle inequality, there exists $|h_k| = \left|h_{B,k} + \mathbf{h}_{S,k}{}^H \mathbf{V}_k \mathbf{h}_{B,S}\right| \overset{(a)}{\leq} |h_{B,k}| + \left|\mathbf{h}_{S,k}{}^H \mathbf{V}_k \mathbf{h}_{B,S}\right|$, where (a) holds if and only if $\arg(h_{B,k}) = \arg(\mathbf{h}_{S,k}{}^H \mathbf{V}_k \mathbf{h}_{B,S}) = \varphi_k^0$. Actually, this indicates that the optimal beamforming is supposed to align the signals from the direct link and that from the STAR-RIS assisted link. Let $\mathbf{\Theta}_k = \left[e^{j\theta_k^1}, e^{j\theta_k^2}, ..., e^{j\theta_k^M}\right]^H$ and $\mathbf{e}_k = diag(\mathbf{h}_{S,k}{}^H)\mathbf{h}_{B,S}$, it is easy to know $\mathbf{h}_{S,k}{}^H \mathbf{V}_k \mathbf{h}_{B,S} = \sqrt{\beta_k} \mathbf{\Theta}_k{}^H \mathbf{e}_k$. As a result, the optimal phase shift at the STAR-RIS is given as

$$\theta_k^m = \text{mod}\left[\arg(h_{B,k}) - \arg(\mathbf{h}_{S,k}{}^H[m]) - \arg(\mathbf{h}_{B,S}[m]), 2\pi\right]. \tag{10}$$

---

[3] With sufficient iterations, the order in which multiple variables are solved has little effect on the final results.

As for the beamforming amplitude β_k, to tackle the non-convex constraint $S \leq R_k$, a successive convex approximation (SCA) based approach is applied. Since the signals from two links are aligned, the channel gain can be rewritten as

$$|h_k|^2 = A_k + B_k\sqrt{\beta_k} + C_k\beta_k, \tag{11}$$

where $A_k = |h_{B,k}|^2$, $B_k = 2|h_{B,k}|\left|\mathbf{h}_{S,k}{}^H diag(\mathbf{\Theta}_k)\mathbf{h}_{B,S}\right|$ and $C_k = \left|\mathbf{h}_{S,k}{}^H diag(\mathbf{\Theta}_k)\mathbf{h}_{B,S}\right|^2$. With $L_{k,1} = P\rho_k$ and $L_{k,2} = \pi(\bar{k})P\rho_{\bar{k}}$, the constraint $S \leq R_k$ is transformed as

$$S + \tau_{\{k,\bar{k}\}}\log_2\left(L_{k,2}\left(A_k + B_k\sqrt{\beta_k} + C_k\beta_k\right) + n\right) \leq$$
$$\tau_{\{k,\bar{k}\}}\log_2\left((L_{k,1} + L_{k,2})\left(A_k + B_k\sqrt{\beta_k} + C_k\beta_k\right) + n\right). \tag{12}$$

As $\log_2(\sqrt{x} + x)$ is a concave function with regard to x, we consider taking the first-order Taylor expansion at the given feasible points β_k^ε to obtain the upper bound of $\log_2\left(L_{k,2}\left(A_k + B_k\sqrt{\beta_k} + C_k\beta_k\right) + n\right)$:

$$\Xi_{k,1}^{UB} = \log_2\left(L_{k,2}\left(A_k + B_k\sqrt{\beta_k^\varepsilon} + C_k\beta_k^\varepsilon\right) + n\right) +$$
$$\frac{L_{k,2}\left(B_k/2\sqrt{\beta_k^\varepsilon} + C_k\right)}{\log(2)\left(n + L_{k,2}\left(A_k + B_k\sqrt{\beta_k^\varepsilon} + C_k\beta_k^\varepsilon\right)\right)}(\beta_k - \beta_k^\varepsilon). \tag{13}$$

To make (12) more tractable, auxiliary variables $\xi_k \leq A + B\sqrt{\beta_k} + C\beta_k$ are introduced and the subproblem to optimize the beamforming amplitude is converted into

$$\text{(P2-1):} \quad \max_{\beta_k, \xi_k} S, \tag{14a}$$

$$\text{s.t.} \quad S + \tau_{\{k,\bar{k}\}}\Xi_{k,1}^{UB} \leq \tau_{\{k,\bar{k}\}}\log_2\left((L_{k,1} + L_{k,2})\xi_k + n\right), \tag{14b}$$

$$\xi_k \leq A_k + B_k\sqrt{\beta_k} + C_k\beta_k, \tag{14c}$$

$$(8b), \tag{14d}$$

which can be solved by CVX.

Next, given $\left\{\pi_k, \mathbf{V}_k, \tau_{\{k,\bar{k}\}}\right\}$, $S \leq R_k$ with regard to the user power allocation can be rewritten as

$$S + \tau_{\{k,\bar{k}\}}\log_2\left(L_{k,3}\rho_{\bar{k}} + n\right) \leq \tau_{\{k,\bar{k}\}}\log_2\left(L_{k,4}\rho_k + L_{k,3}\rho_{\bar{k}} + n\right), \tag{15}$$

where $L_{k,3} = \pi(\bar{k})P|h_k|^2$ and $L_{k,4} = P|h_k|^2$. Similar to (12), the upper bound of $\log_2(L_{k,3}\rho_{\bar{k}} + n)$ can be derived by the first-order Taylor expansion at the given feasible points ρ_k^ε and $\rho_{\bar{k}}^\varepsilon$:

$$\Xi_{k,2}^{UB} = \log_2\left(L_{k,3}\rho_{\bar{k}}^\varepsilon + n\right) + \frac{L_{k,3}}{\log 2\left(L_{k,3}\rho_k^\varepsilon + n\right)}\left(\rho_{\bar{k}} - \rho_{\bar{k}}^\varepsilon\right). \tag{16}$$

As a result, the subproblem to solve the user power allocation is expressed as

$$\text{(P2-2):} \quad \max_{\rho_k} S, \tag{17a}$$

$$\text{s.t.} \quad S + \Xi_{k,2}^{UB} \leq \tau_{\{k,\bar{k}\}}\log_2\left(L_{k,4}\rho_k + L_{k,3}\rho_{\bar{k}} + n\right), \tag{17b}$$

$$(8g). \tag{17c}$$

Finally, the subproblem to solve the user-pair time allocation can be expressed as a linear programming problem:

$$\text{(P2-3):} \quad \max_{\tau_{\{k,\bar{k}\}}} S, \tag{18a}$$

$$\text{s.t.} \quad S \leq \tau_{\{k,\bar{k}\}}\log_2\left(1 + \frac{|h_k|^2\rho_k P}{\pi(\bar{k})|h_k|^2\rho_{\bar{k}}P + n}\right), \tag{18b}$$

$$(8f). \tag{18c}$$

C. Discussion on the Proposed Algorithm

In summary, the proposed iterative algorithm has a two-layer loop-nesting structure. The details are shown in Algorithm 1.

Algorithm 1 The proposed two-layer iterative algorithm

1: **Initialization:** user pairing c.
2: **Repeat**
3:   **Initialization:** V_k^0, ρ_k^0 and $\tau_{\{k,\bar{k}\}}^0$,
4:   **Repeat**
5:     $\varepsilon = 1$,
6:     Calculate the equivalent-combined channel gain and determine the optimal decoding order π_k^* according to (5),
7:       Obtain the optimal phase shift θ_k^* according to (10),
8:       Obtain the optimal amplitude β_k^* by solving (P2-1),
9:       Obtain the optimal power allocation ρ_k^* by solving (P2-2),
10:      Obtain the optimal time allocation $\tau_{\{k,\bar{k}\}}^*$ by solving (P2-3),
11:     Update $\left\{V_k^\varepsilon, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^\varepsilon\right\} \leftarrow \left\{V_k^*, \rho_k^*, \tau_{\{k,\bar{k}\}}^*\right\}$,
12:     Update $\varepsilon \leftarrow \varepsilon + 1$,
13:   **Until** the inner layer algorithm reaches the accuracy or inner iterates beyond the permitted maximum times.
14:   Obtain the utility function by the inner layer algorithm.
15:   For any TU k_T, it search for all other TUs $\tilde{k}_T$
16:   **if** $\left(k_T, \tilde{k}_T\right)$ is a swap-blocking pair
17:     Update $\Phi = \Phi_{k_T}^{\tilde{k}_T}$,
18:   **end if**
19: **Until** the outer layer algorithm reaches the accuracy or outer iterates beyond the permitted maximum times.

---

1) *Convergence:* For the outer layer, the objective value is non-decreasing after each iteration due to the definition of the swap-blocking pair. For the alternating iteration in the inner layer, the objective value is also non-decreasing, which is proven as follows. Firstly, in steps 6, 7 and 10 of Algorithm 1, since the optimal solution can be obtained by (5), (10) and solving (P2-3), we have $S\left(\pi_k^\varepsilon, \theta_k^\varepsilon, \beta_k^\varepsilon, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^\varepsilon\right) \leq S\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^\varepsilon, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right)$. Secondly, define $S_{amp}^{lb,\varepsilon}$ as the objective value obtained by solving (P2-1) at the ε-th iteration. For step 8, it follows

$$S\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^\varepsilon, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right) \overset{(a)}{=} S_{amp}^{lb,\varepsilon}\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^\varepsilon, \xi_k^\varepsilon, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right)$$

$$\overset{(b)}{\leq} S_{amp}^{lb,\varepsilon}\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^{\varepsilon+1}, \xi_k^{\varepsilon+1}, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right)$$

$$\overset{(c)}{\leq} S\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^{\varepsilon+1}, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right),$$

where (a) holds since the first-order Taylor expansion in (13) is tight at the given local points; (b) holds because of the optimized $\beta_k^{\varepsilon+1}$ and $\xi_k^{\varepsilon+1}$; (c) holds since the objective value of (P2-1) is the lower bound of that of the original problem (P2) at $\beta_k^{\varepsilon+1}$. Similarly, for step 9, there is $S\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^{\varepsilon+1}, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right) \leq S\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^{\varepsilon+1}, \rho_k^{\varepsilon+1}, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right)$. Based on the above analysis, it can be driven that $S\left(\pi_k^\varepsilon, \theta_k^\varepsilon, \beta_k^\varepsilon, \rho_k^\varepsilon, \tau_{\{k,\bar{k}\}}^\varepsilon\right) \leq S\left(\pi_k^{\varepsilon+1}, \theta_k^{\varepsilon+1}, \beta_k^{\varepsilon+1}, \rho_k^{\varepsilon+1}, \tau_{\{k,\bar{k}\}}^{\varepsilon+1}\right)$, and the proof is completed. Since the achievable max-min rate is upper bounded by a finite value, the proposed algorithm is guaranteed to converge.

2) *Complexity:* The computational complexity of Algorithm 1 mainly depends on the number of users. The maximum number of swap operations in the outer layer is $K(K-1)/2$. The complexity for solving (P2) in the inner layer is $\mathcal{O}\left(I_{\text{ite}}\left(2(K)^{4.5} + K^2(K+1)\right)\right)$ if the interior point method is employed, where I_{ite} denotes the iterations taken to converge. Then the overall computational complexity of Algorithm 1 is $O\left(K(K-1)I_{\text{ite}}\left(2(K)^{4.5} + K^2(K+1)\right)/2\right)$, i.e.,$O\left(I_{\text{ite}}K^{6.5}\right)$.

IV. NUMERICAL RESULTS

In this section, numerical results are provided to demonstrate the effectiveness of the proposed framework and algorithm. We consider an area with the size of $1000\text{m} \times 1000\text{m}$, where the STAR-RIS is deployed at the center. The main simulation setups are shown in Table I.

TABLE I: Simulation Parameters

Parameter	Value
Communication bandwidth	$BW = 1$ MHz
BS transmission power	$P_{\max} = 30$ dBm
UAV fight height	$H_U = 200$ m
Carrier frequency	$f_c = 750$ MHz
Wavelength	$\lambda = 3 \times 10^8/f_c$
STAR-RIS element spacings	$d_y = d_z = \lambda/10$
Noise power dense	$\sigma^2 = -150$ dBm/Hz
Path loss exponents	$\alpha_1 = 3,\ \alpha_2 = 2,\ \alpha_3 = 2.3$
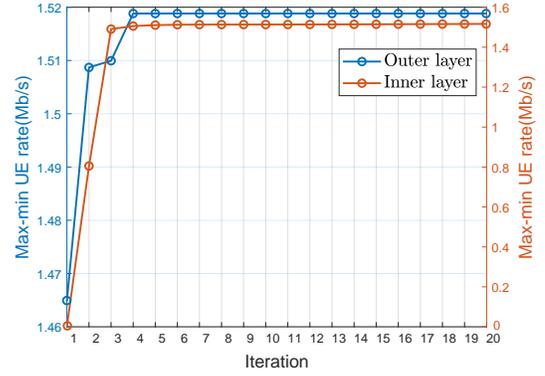
Fig. 2: Convergence of the proposed algorithm.

Fig. 2 illustrates the convergence process of the proposed two-layer iterative algorithm. We can see that both inner-layer iteration and outer-layer iteration show a growing trend and achieve convergence within 10 iterations, which verifies the feasibility of the proposed algorithm. In addition, it is shown that the performance gain obtained by the outer-layer iteration is significantly smaller than that obtained from the inner-layer iteration, which reveals that the user pairing optimization in the outer layer has a modest impact on network performance. This issue will be further demonstrated in Fig. 4.

In Fig. 3, we evaluate the performance of the proposed hybrid NOMA based STAR-RIS assisted network. There are three benchmark frameworks including TDMA based STAR-RIS assisted network, hybrid NOMA based reflecting-only RIS assisted network, and hybrid NOMA based network without RIS. It can be observed that regardless of the frameworks, the max-min user rate decreases as the number of users increases. This is caused by the wireless resource competition among multiple users. Compared with the three benchmarks, we can find that the combination of STAR-RIS and hybrid NOMA has the best performance. Specifically, the gain from the hybrid NOMA is greater than that from the STAR-RIS, with 6.08 times and 2.96 times respectively when $K = 8$. The max-min user rate of the proposed framework is also significantly
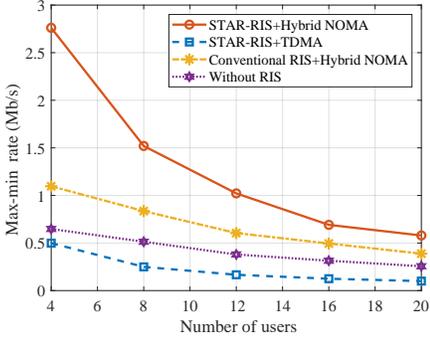
Fig. 3: Max-min rate with different frame-works vs. the number of users
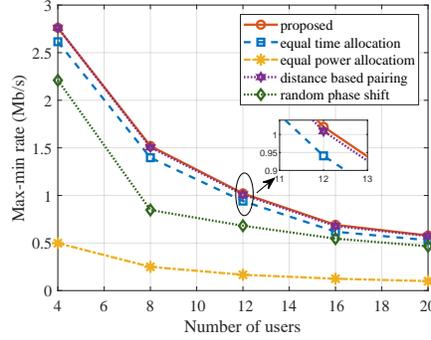
Fig. 4: Max-min rate with different algo-rithms vs. the number of users
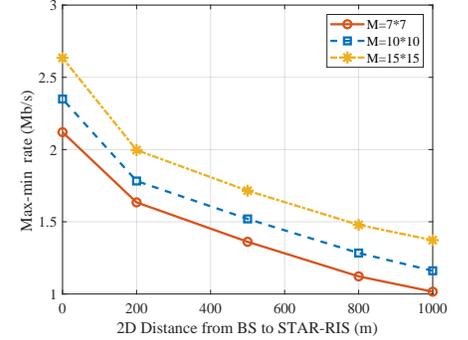
Fig. 5: Max-min rate with different M vs. the 2D distance from BS to STAR-RIS

improved compared with that of conventional reflecting-only RIS assisted network. This improvement exactly comes from the fact that the STAR-RIS can transmit and reflect signals simultaneously, allowing it to serve all users, regardless of its location.

In Fig. 4, the performance of the proposed two-layer iterative algorithm is compared with following benchmark algorithms: 1) equal user-pair time allocation, $\tau_{\{k,\bar{k}\}} = 1/K$; 2) equal user power allocation, $\rho_k = \rho_{\bar{k}} = 0.5$; 3) distance based user pairing, which means the matching state is determined by the distance from users to the STAR-RIS (the far TU is paired with the near RU); and 4) random phase shift at the STAR-RIS. As we can see, the max-min user rate of the proposed algorithm is the largest among all five algorithms, which verifies the superiority of the proposed algorithm. Furthermore, the gain from power optimization is the largest, followed by that from beamforming design, while the gains from other variables are relatively small. This result shows the critical importance of power allocation for NOMA and beamforming design for STAR-RIS.

Both Fig. 3 and Fig. 4 are obtained with $M_y = M_z = 10$ and $\mathbf{q}_S = [0, 0, 20]$ m, while Fig. 5 further investigates the impact of the number of STAR-RIS elements and the location of the STAR-RIS on the network performance [4]. Based on the results shown in Fig. 5, two major conclusions can be drawn. On the one hand, increasing the number of elements can improve the enhancement effect of the STAR-RIS. On the other hand, as the max-min rate increases with the shortening of the distance from BS to STAR-RIS, the deployment of STAR-RIS is of great importance.

V. CONCLUSION

In this paper, we investigated a STAR-RIS enhanced cell-edge network, where a novel hybrid NOMA framework was proposed to take full advantages of the STAR-RIS. A max-min user rate problem was solved by jointly optimizing the user pairing, decoding order, passive beamforming, power and time allocation. Numerical results verified the significant

superiority of the proposed framework in improving com-munication fairness. The importance of power allocation and passive beamforming design was also proved. Moreover, it was confirmed that increasing the STAR-RIS elements and shortening the distance of BS and STAR-RIS contributed to the improvement of network performance.

REFERENCES

[1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[2] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wirel. Commun.*, vol. 28, no. 6, pp. 102–109, Dec. 2021.

[3] A. Papazafeiropoulos, Z. Abdullah, P. Kourtessis, S. Kisseleff, and I. Krikidis, "Coverage probability of STAR-RIS-Assisted massive MIMO systems with correlation and phase errors," *IEEE Wirel. Commun. Lett.*, vol. 11, no. 8, pp. 1738–1742, Nov. 2022.

[4] Y. Lin, Y. Shen, and A. Li, "Simultaneous transmission and reflection beamforming design for RIS-Aided MU-MISO," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 4040–4045, Mar. 2023.

[5] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of Non-Orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[6] H. Liu, G. Li, X. Li, Y. Liu, G. Huang, and Z. Ding, "Effective capacity analysis of STAR-RIS-Assisted NOMA networks," *IEEE Wirel. Commun. Lett.*, vol. 11, no. 9, pp. 1930–1934, Sept. 2022.

[7] X. Li, Y. Zheng, M. Zeng, Y. Liu, and O. A. Dobre, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2684–2688, Feb. 2023.

[8] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint design for simultaneously transmitting and reflecting (STAR) ris assisted NOMA systems," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 1, pp. 611–626, Jan. 2023.

[9] D. Zhang, Q. Wu, M. Cui, G. Zhang, and D. Niyato, "Throughput maximization for IRS-assisted wireless powered hybrid NOMA and TDMA," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 9, pp. 1944–1948, Sept. 2021.

[10] 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. Wirel. Commun.*, vol. 22, no. 3, pp. 1677–1694, Mar. 2023.

[11] C. Wu, X. Mu, Y. Liu, X. Gu, and X. Wang, "Resource allocation in STAR-RIS-Aided networks: OMA and NOMA," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 7653–7667, Sept. 2022.

[12] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 5, pp. 3083–3098, May 2022.

[13] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, "Unsupervised machine learning-based user clustering in Millimeter-Wave-NOMA systems," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 11, pp. 7425–7440, Nov. 2018.

---

[4] The location of the STAR-RIS is considered in a simple way, that is, the distance to the BS. More accurate STAR-RIS deployment will be studied in our future works.