Resource Allocation for Two-Tier RIS-Assisted Heterogeneous NOMA Networks



XU Yongjun¹, YANG Zhaohui², HUANG Chongwen³, YUEN Chau⁴, GUI Guan⁵ (1. School of Communication and Information Engineering, Chongqing

 School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
 Department of Electronic and Electrical Engineering, University College London, London WC1E 6BT, UK;

3. College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China;

4. Engineering Product Development Pillar, Singapore University of Technology and Design, Singapore 487372, Singapore;

5. College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

DOI: 10.12142/ZTECOM.202201006

https://kns.cnki.net/kcms/detail/34.1294.TN.34.1294 TN.20220311.1430.002.html, published online March 14, 2022

Manuscript received: 2022-01-04

Abstract: Reconfigurable intelligent surface (RIS) as a promising technology has been proposed to change weak communication environments. However, most of the current resource allocation (RA) schemes have focused on RIS-assisted homogeneous networks, and there is still no open works about RA schemes of RIS-assisted heterogeneous networks (HetNets). In this paper, we design an RA scheme for a RISassisted HetNet with non-orthogonal multiple access to improve spectrum efficiency and transmission rates. In particular, we jointly optimize the transmit power of the small-cell base station and the phase-shift matrix of the RIS to maximize the sum rates of all small-cell users, subject to the unit modulus constraint, the minimum signal-to-interference-plus-noise ratio constraint, and the cross-tier interference constraint for protecting communication quality of microcell users. An efficient suboptimal RA scheme is proposed based on the alternating iteration approach, and successive convex approximation and logarithmic transformation approach. Simulation results verify the effectiveness of the proposed scheme in terms of data rates.

Keywords: heterogeneous networks; non-orthogonal multiple access; reconfigurable intelligent surface; resource allocation

Citation (IEEE Format): Y. J. Xu, Z. H. Yang, C. W. Huang, et al., "Resource allocation for two-tier RIS-assisted heterogeneous NOMA networks," *ZTE Communications*, vol. 20, no. 1, pp. 36 - 47, Mar. 2022. doi: 10.12142/ZTECOM.202201006.

1 Introduction

ecently, reconfigurable intelligent surface (RIS) as a promising technology has been proposed to improve energy efficiency (EE) and transmission quality by reconfiguring the wireless propagation environment^[1-4]. Particularly, a RIS has a large number of low-cost passive reflecting elements which can reflect the incident signals from the transmitters by changing the phase shift (PS) of the RIS in a passive way. Since the RIS is connected with a base station (BS) by a RIS controller, the reflected signal can be smartly configured to strengthen the desired signal and suppress the undesired signal for meeting the transmission requirements of communication systems.

1.1 Related Work

According to the above advantages of RIS, resource allocation (RA) problems for RIS-assisted networks have been concerned by many scholars. For example, the authors in Ref. [5] proposed a low-complexity iteration algorithm to maximize the spectrum efficiency (SE) by jointly optimizing the beamformer at the BS and the PS of the RIS in a point-to-point RIS-assisted multiple-input single-output (MISO) communication system. But only one user is considered. To overcome the effect of imperfect channel state information (CSI), a weighted sum-rate maximization problem for the multiuser scenario was investigated by jointly designing the active beamforming of the access point (AP) and the passive beamforming of the RIS under perfect/imperfect CSI in Ref. [6], where stochastic successive convex approximation (SCA) was used to obtain the solution. But the minimum signal-tointerference plus-noise ratio (SINR) constraint is not considered, and it fails to meet the required quality of service (QoS) of each user. Considering the same communication network as in Ref. [6], the transmit power and the PS of the RIS were jointly designed in Ref. [7] to maximize the total EE by using an alternat-

The work was partially supported by the China National Key R&D Program under Grant No. 2021YFA1000502, National Natural Science Foundation of China under Grant No. 62101492, Zhejiang Provincial Natural Science Foundation of China under Grant No. LR22F010002, Distinguished Young Scholars of the National Natural Science Foundation of China, Ng Teng Fong Charitable Foundation in the form of ZJU–SUTD IDEA Grant, Zhejiang University Education Foundation Qizhen Scholar Foundation, and Fundamental Research Funds for the Central Universities under Grant No. 2021FZZX001–21.

ing iteration approach and the sequential fractional programming. Moreover, in Ref. [8], the total transmit power of the AP was minimized by jointly optimizing the transmit beamforming of the AP and the reflect beamforming of the RIS, subject to the SINR of each user. In Ref. [9], a particle swarm optimization method was used to minimize the total transmit power. However, the above $\mathrm{works}^{[5\ -9]}$ consider the RIS-assisted MISO system. For a RIS-assisted multiple-input multiple-output (MIMO) network^[10], a weighted sum-rate maximization RA problem was studied by jointly optimizing the precoding matrices of the BS and the PS of the RIS, subject to the maximum power constraint of each BS and the unit modulus constraint of the RIS, where a RIS at the cellular edge was developed to assist the downlink transmission for cell-edge users. The classical block coordinate descent (BCD) algorithm, the Majorization-Minimization algorithm and the complex circle manifold method were used to deal with the non-convex problem. The authors in Ref. [11] proposed a local search method to maximize the sum rate in a RIS-assisted Terahertz communication system. For a RIS-assisted simultaneous wireless information and power transfer (SWIPT) system^[12], the authors studied the RA problem of the weighted sum power maximization of multiple energy-harvesting receivers. In Ref. [13], a theoretical outage performance comparison between nonorthogonal multiple access (NOMA) and orthogonal multiple access (OMA) in the RIS-assisted downlink communication was analyzed. Moreover, the power minimization problem and performance analysis for RIS-aided NOMA systems have been studied in Ref. [14] and Refs. [15 - 17], respectively. The performance comparisons between the new RIS system and the traditional decode-andforward relaying system were studied in terms of total transmit power minimization and the EE maximization^[18]. Moreover, other related works on the RIS-assisted scenario have been studied from space shift keying^[19], channel estimation^[20], etc.

1.2 Motivation and Contributions

However, the above works^[6-20] have not considered the RA problem in heterogeneous networks (HetNets). HetNet as an important technique in 6G communications^[21] can obviously improve communication coverage and network capacity by deploying small cells into the macrocell^[22-23]. However, the QoS of indoor users is heavily degraded by interior walls and strong fading channels. Thus, it is a natural idea to combine RIS and HetNet to exemplify received signals and expand communication coverage. Besides, the introduction of NOMA can improve SE and allow more accessing users. Different from the RA problem in traditional NOMA systems, our work needs to simultaneously optimize transmit power and phase shifts to further improve SE and signal received quality at the receivers.

In this paper, we study the RA problem for a downlink RISassisted heterogeneous NOMA network, where a small-cell BS (SBS) transmits wireless signals to the targeted users whose signals are enhanced with the help of the RIS. The RA problem for such a system is still in its infancy. The contributions of this paper are summarized as follows.

• A two-tier RIS-assisted heterogeneous NOMA network is formulated, where a single-antenna SBS transmits wireless signals to multiple single-antenna small-cell users (SCUs) with the assistance of the RIS. For the single-user scenario, the rate maximization RA problem is formulated by jointly optimizing the transmit power of the SBS and the PS of the RIS under the SINR constraint of the macro-cell user (MCU) and the maximum transmit power constraint. A Dinkelbach-based iteration scheme is proposed to globally obtain the optimal solutions.

• For the multiuser scenario, the total rate maximization problem is formulated under the cross-tier interference constraint, the maximum transmit power, and the unit modulus constraint of the RIS. To deal with the non-convex problem, we propose an efficient RA scheme based on a semi-definite programming (SDP) approach and SCA to obtain the suboptimal solution by an alternating iteration manner.

• Simulation results show that the proposed scheme has good convergence and transmission rates.

The rest of this paper is organized as follows. In Section 2, the system model and the single-user RA problem are introduced. Section 3 proposes a multi-user RA scheme for the multiuser scenario. Section 4 verifies the effectiveness of the proposed scheme. Section 5 concludes this paper.

2 System Model and Single-User RA Problem

2.1 System Model

A downlink two-tier RIS-assisted HetNet with NOMA is given in Fig. 1, where a RIS with M passive reflecting units is



▲ Figure 1. A downlink RIS-assisted heterogeneous non-orthogonal multiple access (NOMA) network

employed to enhance the transmission quality between the SBS and K NOMA SCUs. A macro base station (MBS) serves N MCUs. The SCUs can share the spectrum owned by MCUs via an underlying spectrum way. Since the RIS is developed in interior walls, where the reflected signals from the RIS caused by MCUs can be ignored. A separate wireless control link serves for information exchange between the RIS controller and the SBS, and other required information for implementing the transmit power^[10]. Motivated by the works^[5-7], perfect CSI and continuous phase-shift coefficients are assumed here. For the single-user scenario, defining $\boldsymbol{h} = [h_1, ..., h_M]^H \in \mathbb{C}^{M \times 1}$, $\boldsymbol{g} = [g_1, \dots, g_M]^H \in \mathbb{C}^{M \times 1}, \ \boldsymbol{f} = [f_1, \dots, f_M]^H \in \mathbb{C}^{M \times 1} \text{ and } \boldsymbol{\phi} =$ $[\phi_1,...,\phi_M]^H$ as the channel gain from the SBS to the RIS, the channel gain from the RIS to the SCU, the channel gain from the RIS to the MCU, and the reflecting coefficient (RC) of the RIS, where each $\phi_m = \beta_m e^{j\theta_m}$ comprises an amplitude coefficient $\beta_m \in [0,1]$. Similar to Refs. [5 – 9], a continuous PS coefficient is assumed, e.g., $\theta_m \in [0,2\pi]$ and $|\phi_m| = 1$. g^{SM} and $g^{\rm SS}$ are the channel gains from the SBS to the MCU and the SBS to the SCU. Define $\Phi = \text{diag}(\phi) \in \mathbb{C}^{M \times M}$ as the RC matrix of the RIS. Since we aim to obtain the maximum designed signal, the amplitude coefficient is set as $\beta_m = 1$ for simplicity^[24]. Ignoring signals reflected by the RIS for two and more times^[25], the received SINRs at the SCU and the MCU are

$$\gamma^{\text{SCU}} = \frac{p \left| \boldsymbol{g}^{H} \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{g}^{\text{SS}} \right|^{2}}{P h^{\text{MS}} + \sigma^{2}}, \qquad (1)$$

$$\gamma^{\text{MCU}} = \frac{p \left| \boldsymbol{g}^{H} \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{g}^{\text{SS}} \right|^{2}}{P h^{\text{MS}} + \sigma^{2}}, \qquad (2)$$

where σ^2 represents the noise power at the receiver.

2.2 Single-User RA Scheme

Motivated by the existing works in Refs. [5 - 10], the power allocation and PS ratio under the single-user scenario are jointly optimized by using an alternating iteration approach. However, the solution is just a suboptimal solution. In order to deal with this challenge, we try to get the optimal power allocation and PS solution under a single-user scenario.

Under this special case, we want to improve the transmission rate of the SCU by adjusting the transmit power and PS coefficients in a globally optimal solution way, while the QoS of the MCU is guaranteed. Therefore, the rate-maximization problem of the SCU is formulated as

$$\begin{aligned} \max_{p,\phi_m} \log_2 \left(1 + \gamma^{\text{SCU}} \right) \\ \text{s.t.} \ C_1: \gamma^{\text{MCU}} \ge \gamma^{\min}, \\ C_2: \left| \phi_m \right| = 1, \forall m, \\ C_3: 0 \le p \le p^{\max}, \end{aligned}$$
(3)

where C_1 denotes the QoS required by the MCU under the consideration of the interference power from the RIS and the SBS, γ^{\min} represents the SINR threshold of the MCU. C_2 accounts for the fact that each RIS reflecting element can only provide a phase shift, without amplifying the incoming signal. C_3 denotes the transmit power range of the SBS, and p^{\max} is the maximum transmit power at the SBS. Since we want to find the optimal p^* and ϕ^* , the maximum transmit power of the MBS is ignored.

Based on C_1 , the upper bound of p is

$$p \leq \frac{Ph^{MM} - \gamma^{\min}\sigma^{2}}{\gamma^{\min} |g^{SM} + f^{H}\boldsymbol{\Phi}\boldsymbol{h}|^{2}}.$$
(4)

Based on the monotonicity of p in the objective function in Problem (3), the optimal transmit power becomes

$$p^{*} = \min\left(p^{\max}, \frac{Ph^{MM} - \gamma^{\min}\sigma^{2}}{\gamma^{\min}|g^{SM} + f^{H}\boldsymbol{\Phi}\boldsymbol{h}|^{2}}\right)$$
(5)

To obtain the globally optimal solution, we substitute p^* into Problem (3), the PS coefficient optimization problem of Problem (3) can be reformulated as

$$\max_{\phi_m} \log_2 \left(1 + \frac{A \left| \boldsymbol{g}^H \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{g}^{\text{SS}} \right|^2}{\left| \boldsymbol{g}^{\text{SM}} + \boldsymbol{f}^H \boldsymbol{\Phi} \boldsymbol{h} \right|^2} \right)$$

s.t. $C_2: \left| \phi_m \right| = 1, \forall m,$ (6)

where $A = \frac{ph^{MM} - \gamma^{\min}\sigma^2}{\gamma^{\min}|f^H \boldsymbol{\Phi} \boldsymbol{h} + g^{SM}|}$ if $p^* < p^{\max}$, otherwise $A = p^{\max}$.

Since the data rate of Problem (6) is in proportion to its SINR, we have $\max_{\phi_m} \log_2 \left\{ 1 + \text{SINR}(\phi_m) \right\} \Leftrightarrow \max_{\phi_m} \text{SINR}(\phi_m)$. As a result, we can remove the logarithmic operation, Problem (6) becomes a nonlinear fractional programing problem^[26].

Based on Dinkelbach's method^[26], we have

$$\max_{\phi_m} A \left| g^H \boldsymbol{\Phi} \boldsymbol{h} + g^{SS} \right|^2 - \lambda \left| g^{SM} + \boldsymbol{f}^H \boldsymbol{\Phi} \boldsymbol{h} \right|^2$$

s.t. $C_2: \left| \phi_m \right| = 1, \forall m.$ (7)

where λ is an auxiliary variable. Obviously, the objective function of Problem (7) is a strictly continuous and decreasing function with λ .

Theorem 1: For any $\boldsymbol{\Phi}$ and λ , $Q(\lambda) = \max_{\boldsymbol{\Phi}_m} A \left| \boldsymbol{g}^H \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{g}^{\text{SS}} \right|^2 - \lambda \left| \boldsymbol{g}^{\text{SM}} + \boldsymbol{f}^H \boldsymbol{\Phi} \boldsymbol{h} \right|^2$ is a decreasing function with λ , and $Q(\lambda) \ge 0$.

Proof: See Appendix A.

Since $\boldsymbol{\phi} = [\phi_1, ..., \phi_M]^H$, $\bar{\boldsymbol{g}} = diag(\boldsymbol{g})\boldsymbol{h}$ and $\bar{\boldsymbol{f}} = diag(\boldsymbol{f})\boldsymbol{h}$, Problem (7) becomes

$$\begin{aligned} \max_{\boldsymbol{\phi}_{m}} \boldsymbol{\phi}^{H} (A \bar{\boldsymbol{g}} \bar{\boldsymbol{g}}^{H} - \lambda \bar{\boldsymbol{f}} \bar{\boldsymbol{f}}^{H}) \boldsymbol{\phi} + A \left| \boldsymbol{g}^{\text{SS}} \right|^{2} - \lambda \left| \boldsymbol{g}^{\text{SM}} \right|^{2} + \\ \operatorname{Re} \left\{ \left(A \boldsymbol{g}^{\text{SS}} \bar{\boldsymbol{g}}^{H} - \lambda \boldsymbol{g}^{\text{SM}} \bar{\boldsymbol{f}}^{H} \right) \boldsymbol{\phi} \right\} + \operatorname{Re} \left\{ \boldsymbol{\phi}^{H} \left(A \boldsymbol{g}^{\text{SS}} \bar{\boldsymbol{g}} - \lambda \boldsymbol{g}^{\text{SM}} \bar{\boldsymbol{f}} \right) \right\} \\ \text{s.t.} \ C_{2}: \left| \boldsymbol{\phi}_{m} \right| = 1, \forall m , \end{aligned}$$

$$(8)$$

where Re { \cdot } denotes the real part of a complex number.

Define $\tilde{\boldsymbol{g}} = A \bar{\boldsymbol{g}} \bar{\boldsymbol{g}}^H - \lambda \bar{f} \bar{f}^H$, $\bar{\boldsymbol{h}} = A g^{SS} \bar{\boldsymbol{g}} - \lambda g^{SM} \bar{f}$, and $\bar{A} = A |g^{SS}|^2 - \lambda |g^{SM}|^2$, Problem (8) becomes

$$\max_{\phi_m} \boldsymbol{\phi}^H \tilde{\boldsymbol{g}} \boldsymbol{\phi} + \operatorname{Re} \left\{ \boldsymbol{\phi}^H \bar{\boldsymbol{h}} \right\} + \operatorname{Re} \left\{ \bar{\boldsymbol{h}}^H \boldsymbol{\phi} \right\} + \bar{A}$$

s.t. $C_2: \left| \boldsymbol{\phi}_m \right| = 1, \forall m.$ (9)

Problem (9) is a non-convex quadratically constrained quadratic program (QCQP) problem^[27], which can be converted into a homogeneous QCQP problem. By introducing an auxiliary variable $t(t \ge 0)$, Problem (9) becomes

$$\max_{\phi_m} \bar{\phi}^H M \bar{\phi} + \bar{A}$$

s.t. $\tilde{C}_{2:} \left| \bar{\phi}_m \right| = 1, \forall m,$ (10)

where

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{\tilde{g}} & \boldsymbol{\bar{h}} \\ \boldsymbol{\bar{h}}^{H} & \boldsymbol{0} \end{bmatrix} \text{ and } \boldsymbol{\bar{\phi}} = \begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{t} \end{bmatrix},$$
(11)

and $\bar{\phi} = [\bar{\phi}_1, ..., \bar{\phi}_M] \in \mathbb{C}^{(M+1)\times 1}$. But problem (9) is still nonconvex. Note that $\bar{\phi}^H M \bar{\phi} = \text{Trace}(M \bar{\phi} \bar{\phi}^H)$. Define $X = \bar{\phi} \bar{\phi}^H$, which satisfies $X \ge 0$ and Rank(X) = 1. Thus, Problem (9) becomes

$$\max_{X} \operatorname{Trace}(\boldsymbol{MX}) + A$$

s.t. $\overline{C}_{3}: \boldsymbol{X}_{m} = 1, \forall m,$
 $C_{4}: \boldsymbol{X} \ge \boldsymbol{0}, \operatorname{Rank}(\boldsymbol{X}) = 1.$ (12)

By relaxing the rank-one constraint, Problem (12) is a convex SDP problem^[26], which can be efficiently solved by using the convex optimization tool, e.g., SeDuMi^[28]. Generally, the solution X of Problem (12) does not satisfy its rank constraint^[28], namely, Rank $(X) \neq 1$, while the objective function of Problem (12) only serves an upper bound of it. To obtain a rank-one solution, the Gaussian randomization scheme can be used^[29]. As a result, an iteration-based RA scheme is summarized in Algorithm 1. The convergence of Algorithm 1 is analyzed in Appendix B.

Moreover, the computational complexity of Algorithm 1 is analyzed as follows. The computational complexity is mainly decided by Dinkelbach's method. When the convergence precision ε and the maximum iteration number $L_{\rm max}$ are determined, the computational complexity of the Dinkelbach-based

iterative algorithm is $\mathcal{O}\left(\frac{\log (L_{\max})}{\varepsilon^2}\right)^{[30]}$. Besides, during each iteration, the complexity for obtaining ϕ_m by using SeDuMi tool is $\mathcal{O}\left(\sqrt{M+M^2} L \log \left(\frac{1}{l}\right)\right)$, where $L = M^5 + 3M^6 + M^7$. l denotes the complexity of SeDuMi tool.

notes the convergence precision of SeDuMi tool. As a result, the total complexity of Algorithm 1 is

$$\mathcal{O}\left(\frac{\sqrt{M+M^2}L}{\varepsilon^2}\log\left(\frac{1}{l}\right)\log\left(L_{\max}\right)\right)$$

Algorithm 1. A Dinkelbach-based RA Scheme

- Initialize the maximum number of iterations L_{max}, and the maximum tolerance ε;
- 2: Set the auxiliary variable $\lambda = 0$ and the iteration index l = 0;
- 3: repeat {Main loop}
- 4: Solve the inner loop problem in Eq. (10) for a given λ ;
- 5: Construct **X** by the Gaussian randomization scheme^[13] to satisfy the rank condition; Obtain ϕ_m by using SeDuMi tool^[41];

6: if
$$A \left| g^{H} \boldsymbol{\Phi} \boldsymbol{h} + g^{SS} \right|^{2} - \lambda \left| f^{H} \boldsymbol{\Phi} \boldsymbol{h} + g^{SM} \right|^{2} < \varepsilon$$
 then
7: Convergence = true

8: **return**
$$\phi_m^*, \forall m$$
, obtain p^* and $\lambda^* = \frac{A \left| g^H \Phi h + g^{SS} \right|^2}{\left| f^H \Phi h + g^{SM} \right|^2}$

9: else

10: Set
$$\lambda(l+1) = \frac{A \left| \boldsymbol{g}^{H} \boldsymbol{\Phi}(l) \boldsymbol{h} + \boldsymbol{g}^{SS} \right|^{2}}{\left| f^{H} \boldsymbol{\Phi}(l) \boldsymbol{h} + \boldsymbol{g}^{SM} \right|^{2}}$$
 and $l = l+1;$

12: end if

13: **until** Convergence = true or $l = L_{max}$

3 Multi-User RA Scheme

Since the single-user scenario is a specific case for practical systems, it is only helpful to analyze system performance, which is too ideal for practical two-tier RIS-aided communication systems. In order to improve system capacity and support massive connectivity, in this section, we extend the special case of Problem (3) to the multi-user NOMA scenario. Assuming that there are *K* NOMA SCUs and *N* MCUs, the sets are denoted as $\mathcal{K} = \{1,2,...,K\}, \forall k \in \mathcal{K} \text{ and } \mathcal{N} = \{1,2,...,N\}, \forall n \in \mathcal{N}.$ The signal of the SBS is $x = \sum_{k=1}^{\kappa} \sqrt{p_k} x_k$, where p_k and x_k denote the allocated power and the signal from the SBS to the *k*-th SCU, respectively. Therefore, the signal received at the *k*-th SCU is

$$y_k = g_k^{SS} x + g_k^H \boldsymbol{\Phi} \boldsymbol{h} x + h_k^{MS} \sqrt{P} s_n + n_k, \qquad (13)$$

where g_k^{SS} denotes the channel response from the SBS to SCU k, g_k denotes the channel vector from the RIS to SCU k; h_k^{MS} denotes the channel coefficient from the MBS to the k-th SCU; P and s_n are the transmit power and signal from the MBS to the

n-th MCU, respectively. $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the additive white Gaussian noise (AWGN) at the *k*-th SCU's receiver with zero mean and variance σ_k^2 . The transmission signals of the SBS and the MBS satisfy $E\left[\left|x_k\right|^2\right] = 1$ and $E\left[\left|s_n\right|^2\right] = 1$.

Assume that the channel coefficients are sorted in the descending order $|g_1^{SS} + g_1^H h|^2 \ge |g_2^{SS} + g_2^H h|^2 \ge ... \ge |g_k^{SS} + g_k^H h|^2$. For any *i*<*k*, the SINR of SCU *k* is

$$\gamma_{k}^{\text{SCU}} = \frac{p_{k} \left| g_{k}^{\text{SS}} + g_{k}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2}}{\left| g_{k}^{\text{SS}} + g_{k}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2} \sum_{i=1}^{k-1} p_{i} + \left| h_{k}^{\text{MS}} \right|^{2} P + \sigma_{k}^{2}}, \qquad (14)$$

where the first item in the denominator denotes the interference power from other SCUs, and the second item denotes the cross-tier interference from the MBS to the SCU k. Since the MBS is often far from SCUs, the interference power of the macrocell network can be omitted^[31].

Thus, the total sum rate for all SCUs is

$$R = \sum_{k=1}^{K} \log_2(1 + \gamma_k^{scu}).$$
 (15)

Since we only consider the transmit power optimization problem for SCUs due to the lower priority in spectrum usage, the joint optimization problem of the transmit power of SBS and the transmit power of MBS are beyond the scope of this paper. Thus, we assume that the MBS allocates the same transmit power P to each MCU. Thus, the received signal at the *n*th MCU can be formulated as

$$y_n = \sqrt{P} h_n^{\text{MM}} s_n + \sum_{k=1}^{K} \sqrt{p_k} \left(g_n^{\text{SM}} + f_n^H \boldsymbol{\Phi} \boldsymbol{h} \right) x_k + z_n , \quad (16)$$

where $z_n \sim CN(0,\sigma_n^2)$ denotes the received AWGN at the *n*-th MCU receiver with zero mean and variance σ_n^2 .

As a result, the SINR received at the *n*-th MCU is

$$\gamma_n^{\text{MCU}} = \frac{P \left| h_n^{\text{MM}} \right|^2}{\sum_{k=1}^{K} p_k \left| g_n^{\text{SM}} + f_n^H \boldsymbol{\Phi} \boldsymbol{h} \right|^2 + \sigma_n^2}, \qquad (17)$$

where h_n^{MM} denotes the channel response from the MBS to the *n*-th MCU, and g_n^{SM} is the channel response from the SBS to the *n*-th MCU. The first item in the denominator denotes the cross-tier interference from the SBS to the *n*-th MCU.

Thus, the sum-rate maximization RA problem of the considered RIS-assisted heterogeneous NOMA network becomes

$$\max_{p_k \phi_m} \sum_{k=1}^{\kappa} \log_2 \left(1 + \gamma_k^{\text{SCU}}\right)$$

s.t. $C_2: \left| \phi_m \right| = 1, \forall m,$
 $C_5: \sum_{k=1}^{\kappa} p_k \leq p^{\max},$
 $C_6: \gamma_n^{\text{MCU}} \geq \gamma_n^{\min},$ (18)

where γ_n^{\min} denotes the minimum SINR required by the *n*-th MCU. However, it is difficult to obtain the optimal solution to Problem (18), since the objective function is nonconcave with respect to either p_k or ϕ_m . Moreover, C_6 is nonconvex due to the coupled p_k and ϕ_m . To address this intractable problem, we propose a low-complexity and efficient scheme, which optimizes p_k and ϕ_m by leveraging an alternating iteration approach and the logarithmic transformation method, respectively. The details are given in the following.

3.1 Optimizing p_k with Given ϕ_m

Based on C_6 , we have

$$I_n^{th} = \frac{P \left| h_n^{MM} \right|^2}{\gamma_n^{\min}} - \sigma_n^2 \ge \sum_{k=1}^K p_k \left| g_n^{SM} + f_n^H \boldsymbol{\Phi} \boldsymbol{h} \right|^2, \tag{19}$$

where I_n^{th} denotes the interference power limit of MCU n.

As a result, the power allocation subproblem of Problem (18) becomes

$$\max_{p_{k}} \sum_{k=1}^{K} \log \left(1 + \gamma_{k}^{\text{SCU}}\right)$$

s.t. $C_{5}: \sum_{k=1}^{K} p_{k} \leq p^{\max},$
 $\overline{C}_{6}: \sum_{k=1}^{K} p_{k} |g_{n}^{\text{SM}} + f_{n}^{H} \boldsymbol{\Phi} \boldsymbol{h}|^{2} \leq I_{n}^{th}.$ (20)

It is noted that C_5 and \overline{C}_6 are linear constraints, but the objective function is nonconvex.

Based on the successive convex approximation for low complexity (SCALE) algorithm in Ref. [32], for any z > 0, we have the following bound:

$$\log_2(1+z) \ge \alpha \log_2 z + \beta , \qquad (21)$$

where $\alpha = \frac{z_0}{1 + z_0}$ and $\beta = \log_2(1 + z_0) - \alpha \log_2 z_0$ are the aux-

iliary variables. The relationship is tight at $z = z_0$, and α and β are the approximation constants.

As a result, we have the following sum rate

$$R = \sum_{k=1}^{K} \alpha_k \log_2 \gamma_k^{\text{SCU}} + \sum_{k=1}^{K} \beta_k , \qquad (22)$$

which is a lower bound of the sum rate achieved by SCUs. Note that the relaxation is still a non-convex form due to the difference of convex (d.c.) structure.

Defining $\bar{\sigma}_k = |h_k^{MS}|^2 P + \sigma_k^2$, with the logarithmic transformation $\bar{p}_k = \ln p_k, \forall k$, we have the following convex optimization problem

$$\max_{\bar{p}_{k}} \sum_{k=1}^{K} \alpha_{k} \log_{2}(|g_{k}^{SS} + g_{k}^{H}\boldsymbol{\Phi}\boldsymbol{h}|^{2}) - \sum_{k=1}^{K} \beta_{k} - \sum_{k=1}^{K} \alpha_{k} \log_{2}(\bar{\boldsymbol{\sigma}}_{k} + |g_{k}^{SS} + g_{k}^{H}\boldsymbol{\Phi}\boldsymbol{h}|^{2} \sum_{i=1}^{k-1} e^{\bar{p}_{i}})$$

s.t. $\bar{C}_{5}: \sum_{k=1}^{K} e^{\bar{p}_{k}} \leq p^{\max},$
 $\tilde{C}_{6}: \sum_{k=1}^{K} e^{\bar{p}_{k}} |g_{n}^{SM} + f_{n}^{H}\boldsymbol{\Phi}\boldsymbol{h}|^{2} \leq I_{n}^{\text{th}}, \forall n .$ (23)

Problem (23) is a standard concave maximization problem since the log-sum-exp is convex^[33]. Considering the convexity of Problem (23), the closed-form solution of power allocation can be obtained by using the Lagrange dual theory. Thus, the Lagrange function of Problem (23) can be written as

$$L\left(\bar{p}_{k}\boldsymbol{\mu},\boldsymbol{\vartheta}_{n}\right) = \sum_{k=1}^{K} \alpha_{k} \,\bar{p}_{k} + \sum_{k=1}^{K} \alpha_{k} \log_{2}\left(\left|\boldsymbol{g}_{k}^{\mathrm{SS}} + \boldsymbol{g}_{k}^{H}\boldsymbol{\Phi}\boldsymbol{h}\right|^{2}\right) - \sum_{k=1}^{K} \beta_{k} - \sum_{k=1}^{K} \alpha_{k} \log_{2}\left(\bar{\boldsymbol{\sigma}}_{k} + \left|\boldsymbol{g}_{k}^{\mathrm{SS}} + \boldsymbol{g}_{k}^{H}\boldsymbol{\Phi}\boldsymbol{h}\right|^{2} \sum_{i=1}^{k-1} e^{\bar{p}_{i}}\right) + \mu\left(p^{\max} - \sum_{k=1}^{K} e^{\bar{p}_{k}}\right) + \sum_{n=1}^{N} \vartheta_{n}\left(I_{n}^{\mathrm{th}} - \sum_{k=1}^{K} e^{\bar{p}_{k}}\left|\boldsymbol{g}_{n}^{\mathrm{SM}} + \boldsymbol{f}_{n}^{H}\boldsymbol{\Phi}\boldsymbol{h}\right|^{2}\right), (24)$$

where μ and ϑ_n are the non-negative Lagrange multipliers.

Based on Karush-Kuhn-Tucker (KKT) conditions, the transmit power is calculated by

$$p_{k}^{*} = \left[\frac{\alpha_{k}}{\mu + \sum_{n=1}^{N} \vartheta_{n} \left|g_{n}^{\text{SM}} + f_{n}^{H} \boldsymbol{\Phi} \boldsymbol{h}\right|^{2}}\right]^{*}, \qquad (25)$$

where $[x]^{+} = \max \{0, x\}.$

Based on the sub-gradient updating method, Lagrange multipliers can be calculated by

$$\mu(l+1) = \left[\mu(l) - d_1(l) \times \left(p^{\max} - \sum_{k=1}^{K} p_k(l)\right)\right]^+, \quad (26)$$

$$\boldsymbol{\vartheta}_{n}(l+1) = \left[\boldsymbol{\vartheta}_{n}(l) - \boldsymbol{d}_{2}(l) \times \left(\boldsymbol{I}_{n}^{\text{th}} - \sum_{k=1}^{K} \boldsymbol{p}_{k}(l) \left| \boldsymbol{g}_{n}^{\text{SM}} + \boldsymbol{f}_{n}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2}\right)\right]^{+},$$
(27)

where l is the iteration number. $d_1(l) \ge 0$ and $d_2(l) \ge 0$ denote the sufficiently small step sizes. When they satisfy the following relationship, we have

$$\sum_{l=1}^{\infty} d_i(l) = \infty, \lim_{l \to \infty} d_i(l) = 0, \forall i = \{1, 2\}.$$
(28)

The sub-gradient updating iteration algorithm can fast converge. The dual variables $\mu(l+1)$ and $\vartheta_n(l)$ will converge to the optimal values μ^* and $\vartheta_n^*, \forall n, \forall n \text{ as } l \to \infty$. Note that the duality gap for Problem (23) is zero and the solution is unique. As a result, the primal variable $p_k^*(\mu(l), \vartheta_n(l))$ can reach its targeted value $p_k^*, \forall k$.

3.2 Optimizing ϕ_m with Given p_k^*

Now, we perform the optimization over ϕ_m with the fixed p_k^* . Then, Problem (23) is equivalent to

$$\max_{\boldsymbol{\phi}_{m}} \sum_{k=1}^{K} \frac{\alpha_{k} p_{k} \left| \boldsymbol{g}_{k}^{\mathrm{SS}} + \boldsymbol{g}_{k}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2}}{\left| \boldsymbol{g}_{k}^{\mathrm{SS}} + \boldsymbol{g}_{k}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2} \sum_{i=1}^{k-1} p_{i} + \bar{\sigma}_{k}}$$

s.t. \bar{C}_{2} : $\left| \boldsymbol{\phi}_{m} \right| = 1, \forall m$
 \bar{C}_{6} : $\sum_{k=1}^{K} p_{k} \left| \boldsymbol{g}_{n}^{\mathrm{SM}} + \boldsymbol{f}_{n}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2} \leq I_{n}^{\mathrm{th}}$. (29)

Problem (29) is the sum of multiple-ratio fractional programming problems, and the non-convexity introduced by the fractional objective function and the non-convex unit-modulus constraint \bar{C}_2 . We adopt a relaxation method to solve Problem (29) efficiently.

Defining $\overline{F} = \overline{ff}^{H}$, based on $|x + y| \le |x| + |y|$ and $\overline{f} = \text{diag}(f_n)h, \overline{C}_6$ becomes

$$\left|g_{n}^{\mathrm{SM}}+f_{n}^{H}\boldsymbol{\Phi}\boldsymbol{h}\right|^{2} \leq \left|g_{n}^{\mathrm{SM}}\right|^{2}+\left|\boldsymbol{\phi}^{H}\bar{\boldsymbol{f}}\right|^{2}=\left|g_{n}^{\mathrm{SM}}\right|^{2}+\boldsymbol{\phi}^{H}\tilde{\boldsymbol{F}}\boldsymbol{\phi}\leq I_{n}^{h}/\sum_{k=1}^{K}p_{k}.$$
(30)

Thus, we have

$$\boldsymbol{\phi}^{H}\tilde{\boldsymbol{F}}\boldsymbol{\phi} \leqslant \bar{I}_{n}^{th}, \qquad (31)$$

where $\bar{I}_n^{th} = I_n^{th} / \sum_{k=1}^{K} p_k - |g_n^{\text{SM}}|^2$ is the equivalent interference level for the *n*-th MCU. Thus, Problem (29) becomes

$$\max_{\phi_{m}y_{k}} \sum_{k=1}^{k} y_{k}$$
s.t. \overline{C}_{2} : $\left| \phi_{m} \right| = 1, \forall m,$

$$C_{7}: \frac{\alpha_{k}p_{k} \left| g_{k}^{SS} + g_{k}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2}}{\left| g_{k}^{SS} + g_{k}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2}} \ge y_{k}, \forall k,$$

$$C_{8}: \phi^{H} \tilde{F} \phi \leq \overline{I}_{n}^{\text{th}}, \forall n,$$
(32)

where $y_k \ge 0$ is an auxiliary variable.

Defining $\bar{\boldsymbol{g}}_k = \operatorname{diag}(\boldsymbol{g}_k)\boldsymbol{h}, C_7$ can be rewritten as

$$\boldsymbol{\phi}^{H} \bar{\boldsymbol{g}}_{k} \bar{\boldsymbol{g}}_{k}^{H} \boldsymbol{\phi} + 2 \operatorname{Re} \left\{ \boldsymbol{g}_{k}^{\mathrm{SS},*} \bar{\boldsymbol{g}}_{k}^{H} \boldsymbol{\phi} \right\} + f_{k} \leq 0 , \qquad (33)$$

where $(\cdot)^*$ denotes the conjugate of vector and $f_k = \left|g_k^{\text{SS}}\right|^2 + \gamma_k \bar{\sigma}_k$

$$\frac{\overline{y_k - x_k}}{\overline{y_k} \sum_{i=1}^{k-1} p_i - \alpha_k p_k}.$$

In addition, C_2 can be rewritten as

$$\boldsymbol{\phi}^{H}\boldsymbol{U}\boldsymbol{\phi}=1, \qquad (34)$$

where $U \in M \times M$ is a symmetric matrix with elements of zeros, except for $u_{m,m} = 1$.

Thus, we have

$$\begin{split} \max_{\boldsymbol{\phi},\boldsymbol{y}_{k}} & \sum_{k=1}^{K} \boldsymbol{y}_{k} \\ \text{s.t. } & \bar{C}_{4} : \boldsymbol{\phi}^{H} \boldsymbol{U} \boldsymbol{\phi} = 1, \\ & \bar{C}_{7} : \bar{f}_{k} + f_{k} \leq 0, \\ & \bar{C}_{8} : \boldsymbol{\phi}^{H} \tilde{\boldsymbol{F}} \boldsymbol{\phi} \leq \bar{I}_{n}^{\text{th}}, \forall n, \end{split}$$
(35)

where $\bar{f}_k = \boldsymbol{\phi}^H \bar{\boldsymbol{g}}_k \bar{\boldsymbol{g}}_k^H \boldsymbol{\phi} + 2 \operatorname{Re} \{ \boldsymbol{g}_k^{ss,*} \bar{\boldsymbol{g}}_k^H \boldsymbol{\phi} \}$. Problem (35) is a convex optimization problem with the quadratic constraints.

As a result, the Lagrange function of Problem $\left(35\right)$ can be written as

$$L(\boldsymbol{\phi}, \boldsymbol{y}_{k}, \boldsymbol{\kappa}, \boldsymbol{\rho}_{n}, \boldsymbol{\varpi}_{k}) = -\sum_{k=1}^{K} \boldsymbol{y}_{k} + \boldsymbol{\kappa} \left(\boldsymbol{\phi}^{H} \boldsymbol{U} \boldsymbol{\phi} - 1\right) + \sum_{k=1}^{K} \boldsymbol{\varpi}_{k} \left(\bar{f}_{k} + f_{k}\right) + \sum_{n=1}^{N} \boldsymbol{\rho}_{n} \left(\boldsymbol{\phi}^{H} \tilde{\boldsymbol{F}} \boldsymbol{\phi} - \bar{I}_{n}^{\text{th}}\right),$$
(36)

where $\kappa, \overline{\omega}_k$, and ρ_n are the non-negative Lagrange multipliers.

Based on KKT conditions, the solution of y_k^* can be obtained as

$$y_k^* = \left[\frac{\alpha_k p_k + \sigma \sqrt{\varpi_k \alpha_k p_k}}{\sum_{i=1}^{k-1} p_i}\right]^+, \qquad (37)$$

where the Lagrange multipliers are updated by

$$\boldsymbol{\varpi}_{k}(l+1) = \left[\boldsymbol{\varpi}_{k}(l) + d_{3}(l) \times \left\{\bar{f}_{k} + f_{k}\right\}\right]^{+},$$
(38)

$$\kappa(l+1) = \left[\kappa(l) + d_4(l) \times \left(\boldsymbol{\phi}^H U \boldsymbol{\phi} - 1\right)\right]^*,\tag{39}$$

$$\boldsymbol{\rho}_{n}(l+1) = \left[\boldsymbol{\rho}_{n}(l) + d_{5}(l) \times (\boldsymbol{\phi}^{H}\boldsymbol{F}\boldsymbol{\phi} - \bar{\boldsymbol{I}}_{n}^{\text{th}})\right]^{+}, \tag{40}$$

where $d_3(l), d_4(l)$ and $d_5(l)$ are the step sizes. Based on the

same approach in Problem (28), the sub-gradient-based iterative method can guarantee the convergence of the proposed algorithm.

Algorithm 2. An Iterative RA Scheme

1: Input:

$$M,K,N,p^{\max},I_n^{\text{th}},\sigma_k^2,\boldsymbol{g}_k,\boldsymbol{h},\boldsymbol{f}_n,\boldsymbol{g}_n^{\text{SM}},\boldsymbol{g}_k^{\text{SS}},\boldsymbol{h}_n^{\text{MM}},\boldsymbol{\varepsilon},L_l,L_l, \text{ and } L_{\text{outer}}.$$

2: Initialization: $\alpha_k,\beta_k,l=0,p_k(0)=\frac{P^{\max}}{K},\mu(0)>0,\vartheta_n(0)>0,\kappa(0)>0,\rho_n(0)>0,\sigma_k(0)>0,d_3(0),d_4(0),d_5(0), \text{ and}$
 $\boldsymbol{\Phi}(0)=\frac{\pi}{2}\boldsymbol{I}_M.$
3: while $l < L_{\max}$ do
4: Given $\boldsymbol{\Phi}$ update $p_k,\forall k$:
5: for $i=1:1:L_l$ do
6: Update transmit power p_k by Problem (26);
7: Undate Lagrange multipliers μ and ϑ by Problems (27)

7: Update Lagrange multipliers μ and ϑ_n by Problems (27) and (28);

8: Update the auxiliary variable
$$\alpha_k(i+1) = \frac{\gamma_k^{SCU}(i)}{1+\gamma_k^{SCU}(i)}$$

and
$$\beta_k(i+1) = \log_2(1+\gamma_k^{\text{SCU}}(i)) - \alpha(i)\log_2(\gamma_k^{\text{SCU}}(i))$$

9: if
$$|| p(i+1) - p(i) ||^2 < \varepsilon$$
 then

10: Obtain the optimal transmit power
$$p^* = p(i + 1)$$
,
where $p(i + 1) = [p_1(i + 1),...,p_K(i + 1)]^T$, break

11: else

12: i = i + 1;

13: end if

15: Given p^* , $\forall k$ update $\boldsymbol{\Phi}$:

16: for
$$j = 1:1:L_J$$
 do

- 17: Update the auxiliary variable y_k by Problem (38);
- Update the Lagrange multipliers by Problems (39 41), respectively;

19: Until
$$\| \mathbf{y}(j+1) - \mathbf{y}(j) \|^2 < \varepsilon$$
, where $\mathbf{y}(j) = \begin{bmatrix} y_1(j), \dots, y_K(j) \end{bmatrix}^T$; Obtain $y_k, \kappa \boldsymbol{\rho}_n$, and $\boldsymbol{\varpi}_k$.

20: Obtain the RC matrix $\boldsymbol{\Phi}$ by solving the SDP problem in problem (42).

21: if
$$\| \boldsymbol{\Phi}(j+1) - \boldsymbol{\Phi}(j) \|^2 < \varepsilon$$
 then

22:
$$\boldsymbol{\Phi}^* = \boldsymbol{\Phi}(j+1)$$
, break;

23: else

24: j = j + 1;

25: end if

26: end for

27: **until** Convergence = **true** or $l = L_{outer}$.

28: end while 29: Output: p^* and Φ

In order to solve the PS coefficient, based on a Schur complement, the dual problem of Problem (35) becomes

4.1 Single-User Case

$$\begin{aligned} &\max \gamma \\ &\mathrm{s.t.}\, \gamma \geq 0, \\ &\tilde{M} \geq 0, \end{aligned} \tag{41}$$

where $\tilde{\boldsymbol{M}} = \begin{bmatrix} B_1 & B_2 \\ B_2^T & B_3 \end{bmatrix}, B_1 = \kappa \boldsymbol{U} + \sum_{k=1}^{\kappa} \boldsymbol{\varpi}_k \, \bar{\boldsymbol{g}} \bar{\boldsymbol{g}}^H + \sum_{n=1}^{N} \rho_n \tilde{\boldsymbol{F}}, B_2 = \sum_{k=1}^{\kappa} \boldsymbol{\varpi}_k \, \operatorname{Re} \left\{ g_k^{\mathrm{SS},*} \bar{\boldsymbol{g}}_k^H \right\}, \text{ and } B_3 = \sum_{k=1}^{\kappa} \boldsymbol{\varpi}_k \, f_k - \sum_{k=1}^{\kappa} \gamma_k - \kappa - \sum_{n=1}^{N} \rho_n \bar{I}_n^{\mathrm{th}} - \gamma,$ and γ is an auxiliary variable. Problem (41) is a convex SDP problem and the Slater's constraint qualification is satisfied^[28]. Particularly, an iterative-based RA scheme for the multiuser RIS-assisted heterogeneous NOMA network is summarized in Algorithm 2.

3.3 Computational Complexity

For the multi-user system, the complexity of the algorithm mainly depends on the number of users and the optimization approach. Regarding to Algorithm 2, it involves the inner and outer iterations. The inner layer is used to obtain variables pand $\boldsymbol{\Phi}$ respectively, while the outer layer is taken for alternating iterations. We denote the maximum alternating iteration number as L_{outer} , and the iteration numbers of obtaining p and $\boldsymbol{\Phi}$ as L_I and L_I respectively. There exists a polynomial-time complexity (K(N + 1)) to solve *p* by using sub-gradient updating methods. Therefore, its computational complexity is $(K(N + 1)L_i)$. Similarly, the computational complexity of the SDP problem to solve ϕ_m is $(\sqrt{M+2} L \log{(1/w)})^{[28]}$, where $L = M^3 + M^6 + M^3 (M + 1)^2 (M^3 + M + 1) + M^9$ and w is the convergence precision of the SDP problem. Based on the above discussion, the total complexity of Algorithm 2 is $((K(N + 1)L_{I}\sqrt{M + 2}L\log(1/w)L_{I})L_{outer}).$

4 Simulation Results

In this section, simulation results are given to demonstrate the effectiveness of the proposed RA scheme by comparing it with the rate-maximization based RA scheme without RIS^[33] which is defined as "the traditional RA scheme without RIS". A spectrum-sharing small cell is randomly distributed in the coverage area of the macrocell. SCUs are uniformly distributed in the coverage area of their associated SBS. The coverage radii of the macrocell and small cell are 500 m and 20 m, respectively. The distance-dependent path loss model for the largescale fading is given by $\frac{1}{(d_m r_k)^{\alpha}}$, where d_m is the distance be-

tween the SBS to the reflecting unit *m* of RIS, r_k is the distance between the RIS and the SCU *k*, and α denotes the path loss exponent. The distance between the SBS and the RIS is 10 m. The distance between RIS and the SCU is 5 m. The small-scale fading is considered as Rayleigh fading channel^[13]. The stopping criterion for convergence is $\varepsilon = 10^{-6}$. The noise power is $\sigma^2 = -100$ dBm, and $\alpha = 3$. The simulation setup is shown in Fig. 2. The detailed parameter settings are shown in Table 1.

Fig. 3 depicts the data rate of the SCU versus the transmit power at the MBS. The target SINR of MCU is $\gamma^{\min} = 2$ dB. From this figure, the data rate of the SCU improves quickly as the increasing transmit power *P*. The reason is that the large



Figure 2. Simulation setup

Table	1.	System	parameters
-------	----	--------	------------

Parameters	Values	Parameters	Values
М	4 ^[1]	K	3 ^[1]
N	2 ^[29]	Р	$20~\mathrm{dBm}^{[29]}$
I_n^{th}	10 ^{-3[21]}	p^{\max}	$20 \text{ mW}^{[21]}$
γ^{\min}	$2 \ dB^{[29]}$	r_k	5 m ^[6]
d_{m}	10 m ^[6]	ε	10 ^{-6[21]}
α	3 ^[6]	σ^2	$-100~\mathrm{dBm}^{[6]}$



Figure 3. Data rate of SCU versus transmit power of the MBS P

transmit power of the MBS increases the available transmit power from the SBS to the SCU according to the constraint C_1 . Besides, this figure also shows that the data rate under the proposed RA scheme is larger than that of the RA scheme without RIS, which indicates that the RIS can improve the rate performance in a passive reflecting way. Moreover, the data rate increases heavily with the increasing number of reflecting elements at the RIS, which indicates that a massive number of passive reflecting elements brings better performance improvement.

Fig. 4 shows the data rate of the SCU versus the minimum SINR requirement of the MCU (e.g., γ^{\min}). The transmit power of the MBS is assumed to be P = 20 dBm. With the increasing γ^{\min} , the data rate of the SCU decreases a lot. The reason is that, under the fixed transmit power at the MBS, the allowed transmit power from the SBS to the SCU becomes smaller to avoid less interference power to the MCU. It is noted that as for the achieved transmission performance, i.e., the achieved data rate, the proposed RA scheme is better than that of the RA scheme without RIS.

Fig. 5 gives the allocated power to the SCU versus the number of reflecting elements M under different transmit power of the MBS P. It is observed that the required transmit power of the proposed RA scheme is around 40% lower than that of the RA scheme without RIS under different transmit power P. This demonstrates that the energy consumption can be reduced a lot when the RIS is adopted. Moreover, the received transmit power at the SCU decreases with the increasing number of reflecting elements M. Because larger reflecting elements provide more multipath interference to the MCU. The available transmit power of the SBS decreases for guaranteeing the QoS of the MCU (namely, cross-tier interference constraint). When the transmit power at the MBS is increased, the available transmit power at the SBS is increased accordingly, since large transmit power at



Figure 4. Data rate of SCU versus SINR threshold of MCU γ^{min}

the MBS allows more transmit power to improve the communication quality of SCU via constraint (4).

4.2 Multi-User Case

The numbers of SCUs and MCUs are K=3 and N=2, respectively. The reflecting elements are M=4. Assume that each MCU receiver has the same interference power limit. The maximum transmit power of the SBS is $p^{\text{max}} = 20 \text{ mW}$, and the interference power limit is $I_{th}^n = 10^{-3} \text{ mW}^{[33]}$.

Fig. 6 shows the convergence performance of the proposed RA scheme. It is observed that the allocated power to each SCU can quickly reach the equilibrium points within eight iterations, which demonstrates that the proposed RA scheme has a good convergence.



▲ Figure 5. Allocated power from SBS to small-cell user (SCU) versus the number of reflecting elements M



▲ Figure 6. Convergence of proposed resource allocation (RA) scheme

Fig. 7 shows the sum rate of all SCUs versus the number of reflecting elements with different SCUs. It is observed that the total rate achieved by SCUs under the proposed RA scheme increases with the increasing number of reflecting elements. Since the RIS can provide more reflecting signals to strengthen the desired signals to SCUs. However, the total rate of SCUs under the traditional RA scheme always keeps stable with the increasing number of reflecting elements M since no RIS is deployed in this scheme. Besides, the total rate of SCUs with a large K is better than that with a small *K*. The gap of the total rate of SCUs under a small *m* is larger than that under a large M. Because the interference power among different SCUs is small due to less transmission path, the total rate is decided by the number of users. Thus, there is a big performance gap in the region with small reflecting elements. However, due to the effect of the cross-tier interference constraint, the overall performance of SCUs is limited. It is impossible to unrestrictedly improve the performance with the increasing number of reflecting elements M.

Fig. 8 shows the total rate of SCUs versus the different interference power limit of MCU I_n^{th} . The total rate of SCUs increases with the increasing I_n^{th} . The bigger I_n^{th} means that the MCU can tolerate more harmful interference power from SCUs and RIS. Correspondingly, the SBS allocates more transmit power to each SCU for pursuing a higher data rate. Moreover, the proposed RA scheme with a larger *M* achieves better performance than the other two cases. The reason is that the RIS can effectively enhance system performance by providing multiple flexible transmission paths.

5 Conclusions

In this paper, we have maximized the total rate of SCUs by jointly optimizing the transmit power at the SBS and PS coefficients at the RIS in RIS-assisted heterogeneous NOMA networks. Specifically, we firstly propose an optimal RA scheme to maximize the transmission rate of the SCU in a single-user scenario, subject to the SINR constraint of MCU, the maximum transmit power constraint of SBS, and the PS coefficient constraint. Then, we design a suboptimal RA scheme to maximize the total rate of SCUs for the multi-user scenario, subject to the sum transmit power constraint of the SBS, the cross-tier interference constraint of each MCU, and the unit modulus constraint of the RIS. The alternating iteration approach and the SCA approach are used to solve the originally non-convex optimization problem. Simulation results show that the proposed RA scheme has a good converge and transmission rate compared with the traditional RA scheme without RIS. In our future work, we will study the joint optimization of beamforming and PS coefficients under the discrete PS constraint and imperfect CSI in a multi-antenna heterogeneous NOMA network with RIS.



▲ Figure 7. Total rate of SCUs versus the number of reflecting elements M



A Figure 8. Total rate of SCUs versus the interference power limit of the MCU I_n^{th}

Appendix A: Proof of Theorem 1

Based on Problem (7), an equivalent objective function is given by

$$Q(\lambda) = \max_{\phi_m} A \left| \mathbf{g}^H \boldsymbol{\Phi} \mathbf{h} + \mathbf{g}^{SS} \right|^2 - \lambda \left| \mathbf{g}^{SM} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h} \right|^2.$$
(A.1)

Assume there are two optimal auxiliary variables λ_1 and $\lambda_2(\lambda_1 > \lambda_2)$ for the optimal phase shift policies $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, respectively, then

$$Q(\lambda_{2}) = A |g^{H} \boldsymbol{\Phi}_{2} \boldsymbol{h} + g^{SS}|^{2} - \lambda_{2} |g^{SM} + f^{H} \boldsymbol{\Phi}_{2} \boldsymbol{h}|^{2} >$$

$$A |g^{H} \boldsymbol{\Phi}_{1} \boldsymbol{h} + g^{SS}|^{2} - \lambda_{2} |g^{SM} + f^{H} \boldsymbol{\Phi}_{1} \boldsymbol{h}|^{2} >$$

$$A |g^{H} \boldsymbol{\Phi}_{1} \boldsymbol{h} + g^{SS}|^{2} - \lambda_{1} |g^{SM} + f^{H} \boldsymbol{\Phi}_{1} \boldsymbol{h}|^{2} =$$

$$Q(\lambda_{1}). \qquad (A.2)$$

Therefore, $Q(\lambda)$ is a strictly decreasing function with the variable λ . Define Φ' as any solution for Problem (7) and

$$\lambda' = \frac{A \left| \boldsymbol{g}^{H} \boldsymbol{\Phi}' \boldsymbol{h} + \boldsymbol{g}^{ss} \right|^{2}}{\left| \boldsymbol{g}^{sm} + \boldsymbol{f}^{H} \boldsymbol{\Phi}' \boldsymbol{h} \right|^{2}}, \text{ we have}$$

$$Q(\lambda') = \max_{\phi_{m}} A \left| \boldsymbol{g}^{H} \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{g}^{SS} \right|^{2} - \lambda' \left| \boldsymbol{g}^{SM} + \boldsymbol{f}^{H} \boldsymbol{\Phi} \boldsymbol{h} \right|^{2} \ge$$

$$A \left| \boldsymbol{g}^{H} \boldsymbol{\Phi}' \boldsymbol{h} + \boldsymbol{g}^{SS} \right|^{2} - \lambda' \left| \boldsymbol{g}^{SM} + \boldsymbol{f}^{H} \boldsymbol{\Phi}' \boldsymbol{h} \right|^{2} = 0. \quad (A.3)$$

As a result, we have $Q(\cdot) \ge 0$. The proof is complete.

Appendix B: Proof of Convergence

According to the algorithm procedure of Algorithm 1, the convergence is determined by the Dinkelbach-based approach. To show the convergence of Dinkelbach's method, we define $(\lambda(l), \phi_m(l))$ as the optimal solution in the *l*-th iteration, $\lambda(l) \neq \lambda^*$ and $\lambda(l+1) \neq \lambda^*$ as the EE at the *l*-th and (l + 1)-th iteration, respectively. We have the following proposition.

Proposition: If (λ', ϕ'_m) is an arbitrary feasible solution of Problem (7), we have

$$\lambda' = \frac{A \left| \boldsymbol{g}^{H} \boldsymbol{\Phi}' \boldsymbol{h} + \boldsymbol{g}^{SS} \right|^{2}}{\left| \boldsymbol{g}^{SM} + \boldsymbol{f}^{H} \boldsymbol{\Phi}' \boldsymbol{h} \right|^{2}}, \qquad (B.1)$$

then $F(\lambda', \Phi) \ge 0$. Proof: We first have

$$F(\lambda', \boldsymbol{\Phi}) = \max_{\phi_m} \left\{ A \left| \boldsymbol{g}^H \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{g}^{\text{SS}} \right|^2 - \lambda' \left| \boldsymbol{g}^{\text{SM}} + \boldsymbol{f}^H \boldsymbol{\Phi} \boldsymbol{h} \right|^2 \right\} \ge A \left| \boldsymbol{g}^H \boldsymbol{\Phi}' \boldsymbol{h} + \boldsymbol{g}^{\text{SS}} \right|^2 - \lambda' \left| \boldsymbol{g}^{\text{SM}} + \boldsymbol{f}^H \boldsymbol{\Phi}' \boldsymbol{h} \right|^2 = 0.$$
(B.2)

Based on the Proposition, we have $F(\lambda^{l}, \boldsymbol{\Phi}(l)) > 0$ and $F(\lambda^{l+1}, \boldsymbol{\Phi}(l+1)) > 0$ because of $\lambda(l) \neq \lambda^{*}$ and $\lambda(l+1) \neq \lambda^{*}$. As a result, $F(\lambda^{l}, \boldsymbol{\Phi}(l)) > 0$ can be rewritten as

$$F(\lambda(l), \boldsymbol{\Phi}(l)) = A \left| \boldsymbol{g}^{H} \boldsymbol{\Phi}(l) \boldsymbol{h} + \boldsymbol{g}^{SS} \right|^{2} - \lambda(l) \left| \boldsymbol{g}^{SM} + \boldsymbol{f}^{H} \boldsymbol{\Phi}(l) \boldsymbol{h} \right|^{2} = \left| \boldsymbol{g}^{SM} + \boldsymbol{f}^{H} \boldsymbol{\Phi}(l) \boldsymbol{h} \right|^{2} \{ \lambda(l+1) - \lambda(l) \} > 0 \Longrightarrow \lambda(l+1) > \lambda(l) .$$
(B.3)

Since $\lambda(l+1) > \lambda(l)$, while $F(\lambda, \Phi)$ is a strictly decreasing function in λ according to Theorem 1, we can show that with the increase of iterations, $F(\lambda(l), \Phi(l))$ can gradually approach to zero, namely $\lim_{l \to \infty} F(\lambda(l), \Phi(l)) = 0$. Accordingly, we have $\lambda(l+1) = \lambda(l), l \to \infty$. Thus, the Dinkelbach-based approach can guarantee the convergence of the algorithm. The proof is complete.

References

- XU Y, GAO Z, WANG Z, et al. RIS-enhanced WPCNs: joint radio resource allocation and passive beamforming optimization [J]. IEEE Transactions on vehicular technology, 2021, 70(8): 7980 - 7991
- [2] HUANG C W, HU S, ALEXANDROPOULOS G C, et al. Holographic MIMO surfaces for 6G wireless networks: opportunities, challenges, and trends [J]. IEEE wireless communications, 2020, 27(5): 118 - 125. DOI: 10.1109/ MWC.001.1900534transactions
- [3] HUANG C, MO R, YUEN C. Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning [J]. IEEE journal on selected areas in communications, 2020, 38(8):1839 – 1850. DOI: 10.1109/ JSAC.2020.3000835
- [4] GAO Z, XU Y, WANG Q, et al. Outage-constrained energy efficiency maximization for RIS-assisted WPCNs [J]. IEEE communications letters, 2021, 25(10): 3370 - 3374
- [5] YU X, XU D, SCHOBER R. Miso wireless communication systems via intelligent reflecting surfaces [C]//2019 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2019: 735 – 740
- [6] GUO H, LIANG Y C, CHEN J, et al. Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks [J]. IEEE transactions on wireless communications, 2020, 19(5):3064 – 3076
- [7] HUANG C, ZAPPONE A, ALEXANDROPOULOS G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication [J]. IEEE transactions on wireless communications, 2019, 18(8):4157 - 4170
- [8] JIANG W, ZHANG Y, WU J, et al. Intelligent reflecting surface assisted secure wireless communications with multiple-transmit and multiple-receive antennas
 [J]. IEEE access, 2020, 8: 86659 - 86673. DOI: 10.1109/ ACCESS.2020.2992613
- [9] SOUTO V D P, SOUZA R D, UCH^OOA-FILHO B F, et al. Beamforming optimization for intelligent reflecting surfaces without CSI [J]. IEEE Wireless Communications Letters, 2020, 9(9):1476 - 1480
- [10] PAN C H, REN H, WANG K Z, et al. Multicell MIMO communications relying on intelligent reflecting surfaces [J]. IEEE transactions on wireless communications, 2020, 19(8): 5218 – 5233. DOI: 10.1109/TWC.2020.2990766
- [11] CHEN W J, MA X Y, LI Z X, et al. Sum-rate maximization for intelligent reflecting surface based terahertz communication systems [C]//Proceedings of 2019 IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops). IEEE, 2019: 153 – 157. DOI: 10.1109/ICCChinaW.2019.8849960
- [12] WU Q Q, ZHANG R. Weighted sum power maximization for intelligent reflecting surface aided SWIPT [J]. IEEE wireless communications letters, 2020, 9 (5): 586 - 590. DOI: 10.1109/LWC.2019.2961656
- [13] DING Z G, SCHOBER R, POOR H V. On the impact of phase shifting designs on IRS-NOMA [J]. IEEE wireless communications letters, 2020, 9(10): 1596 – 1600. DOI: 10.1109/LWC.2020.2991116
- [14] WANG H, LIU C, SHI Z, et al. On power minimization for IRS-aided downlink NOMA systems [J] IEEE wireless communications letters, 2020, 9(11): 1808 – 1811. DOI: 10.1109/LWC.2020.2999097
- [15] FU M, ZHOU Y, SHI Y M, et al. Reconfigurable intelligent surface empowered downlink non-orthogonal multiple access [J]. IEEE transactions on communications, 2021, 69(6): 3802 - 3817. DOI: 10.1109/TCOMM.2021.3066587
- [16] WANG W, LIU Y W, LUO Z Q, et al. Toward cross-layer design for nonorthogonal multiple access: A quality-of-experience perspective [J]. IEEE wireless communications, 2018, 25(2): 118 - 124. DOI: 10.1109/ MWC.2018.1700081
- [17] HOU T W, LIU Y W, SONG Z Y, et al. Reconfigurable intelligent surface aided NOMA networks [J]. IEEE journal on selected areas in communications, 2020, 38(11): 2575 - 2588. DOI: 10.1109/JSAC.2020.3007039
- [18] BJÖRNSON E, ÖZDOGAN Ö, LARSSON E G. Intelligent reflecting surface versus decode-and-forward: how large surfaces are needed to beat relaying? [J]. IEEE wireless communications letters, 2020, 9(2): 244 – 248. DOI: 10.1109/LWC.2019.2950624
- [19] CANBILEN A E, BASAR E, IKKI S S. Reconfigurable intelligent surfaceassisted space shift keying [J]. IEEE wireless communications letters, 2020, 9 (9): 1495 - 1499. DOI: 10.1109/LWC.2020.2994930
- [20] WEI L, HUANG C W, ALEXANDROPOULOS G C, et al. Channel estimation

for RIS-empowered multi-user MISO wireless communications [J]. IEEE transactions on communications, 2021, 69(6): 4144 - 4157. DOI: 10.1109/ TCOMM.2021.3063236

- [21] XU Y J, LI G Q, YANG Y, et al. Robust resource allocation and power splitting in SWIPT enabled heterogeneous networks: A robust minimax approach [J]. IEEE Internet of Things journal, 2019, 6(6): 10799 - 10811. DOI: 10.1109/ JIOT.2019.2941897
- [22] DAMNJANOVIC A, MONTOJO J, WEI Y B, et al. A survey on 3GPP heterogeneous networks [J]. IEEE wireless communications, 2011, 18(3): 10 - 21. DOI: 10.1109/MWC.2011.5876496
- [23] XU Y J, GUI G, GACANIN H, et al. A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges [J]. IEEE communications surveys & tutorials, 2021, 23(2): 668 - 695. DOI: 10.1109/COMST.2021.3059896
- [24] DINKELBACH W. On nonlinear fractional programming [J]. Management science, 1967, 13(7): 492 – 498. DOI: 10.1287/mnsc.13.7.492
- [25] BOYD S, VANDENBERGHE L. Convex Optimization [M]. Cambridge: Cambridge University Press, 2004. DOI: 10.1017/cbo9780511804441
- [26] STURM J F. Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones [J]. Optimization methods and software, 1999, 11(1 4): 625 653. DOI: 10.1080/10556789908805766
- [27] CROUZEIX J P, FERLAND J A. Algorithms for generalized fractional programming [J]. Mathematical programming, 1991, 52(1 - 3): 191 - 207. DOI: 10.1007/BF01582887
- [28] ZHANG H J, HUANG S T, JIANG C X, et al. Energy efficient user association and power allocation in millimeter-wave-based ultra-dense networks with energy harvesting base stations [J]. IEEE journal on selected areas in communications, 2017, 35(9): 1936 - 1947. DOI: 10.1109/JSAC.2017.2720898
- [29] ZHANG H J, JIANG C X, BEAULIEU N C, et al. Resource allocation in spectrum-sharing OFDMA femtocells with heterogeneous services [J]. IEEE transactions on communications, 2014, 62(7): 2366 - 2377. DOI: 10.1109/ TCOMM.2014.2328574
- [30] PAPANDRIOPOULOS J, EVANS J S. SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks [J]. IEEE transactions on information theory, 2009, 55(8): 3711 - 3724. DOI: 10.1109/ TIT.2009.2023751
- [31] PALOMAR D P, CHIANG M. A tutorial on decomposition methods for network utility maximization [J]. IEEE journal on selected areas in communications, 2006, 24(8): 1439 - 1451. DOI: 10.1109/JSAC.2006.879350
- [32] ZHU Z Y, WANG N, HAO W M, et al. Robust beamforming designs in secure MIMO SWIPT IoT networks with a nonlinear channel model [J]. IEEE Internet of Things journal, 2021, 8(3): 1702 - 1715. DOI: 10.1109/JIOT.2020.3014933
- [33] OGUEJIOFOR O S, ZHANG L X, NAWAZ N. UE-centric clustering and resource allocation for practical two-tier heterogeneous cellular networks [J]. IET communications, 2018, 12(18): 2384 – 2392. DOI: 10.1049/iet-com.2018.5422

Biographies

XU Yongjun received his Ph.D. degree (Hons.) in communication and information system from Jilin University, China in 2015. He received the Outstanding Doctoral Thesis of Jilin Province in 2016. He is currently an associate professor with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, China. From 2018 to 2019, he was a visiting scholar with Utah State University, USA. He has authored or coauthored more than 100 papers for renowned journals, including *IEEE COMMUN SURVI TVT/TCOM/TII/TCCN/TGCN/CL/WCL*, etc. His recent interests include heterogeneous networks, resource allocation, intelligent reflecting surface, energy harvesting, backscatter communications, mobile edge computing, and UAV communications. He serves as an associate editor of EURASIP Journal on Wireless Communications and Networking, an editor of Physical Communications, and an academic editor of Digital Communications and Networks. He serves as a reviewer for IEEE TWC/TSP/TCOM/TVT/CL, etc. He is a senior member of IEEE.

YANG Zhaohui received his Ph.D. degree in communication and information system with the National Mobile Communications Research Laboratory, Southeast University, China in May 2018. From May 2018 to October 2020, he was a postdoctoral research associate with the Center for Telecommunications Research, Department of Informatics, King's College London, UK. He is currently a visiting associate professor with College of Information Science and Electronic Engineering Zhejiang Key Lab of Information Processing Communication and Networking, Zhejiang University, China and also a research fellow with the Department of Electronic and Electrical Engineering, University College London, UK. He is an associate editor for the *IEEE Communications Letters, IET Communications* and *EURASIP Journal on Wireless Communications and Networking*. He was an exemplary reviewer for *IEEE Transactions on Communications* in 2019 and 2020.

HUANG Chongwen (chongwenhuang@zju.edu.cn) obtained his B.Sc. degree in 2010 from Nankai University, China, and the M.Sc. degree from the University of Electronic Science and Technology of China in 2013, and Ph.D. degree from Singapore University of Technology and Design (SUTD). From Oct. 2019 to Sept. 2020, he is a Postdoc in his mother University SUTD. Since Sept. 2020, he joined Zhejiang University as a tenure-track young professor. Dr. HUANG is the recipient of IEEE Marconi Prize Paper Award in Wireless Communications, and IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2021. He has served as an editor of *IEEE Communications Letter, Elsevier Signal Processing, EURASIP Journal on Wireless Communications and Networking* and *Physical Communication* since 2021. His main research interests focus on holographic MIMO surface/reconfigurable intelligent surface, B5G/6G wireless communication, mmWave/THz communications, deep learning technologies for wireless communications, etc.

YUEN Chau received his B.Eng. and Ph.D. degrees from Nanyang Technological University (NTU), Singapore in 2000 and 2004, respectively. He was a Post-Doctoral Fellow with Lucent Technologies Bell Labs, USA in 2005, and a visiting assistant professor with The Hong Kong Polytechnic University, China in 2008. From 2006 to 2010, he was with the Institute for Infocomm Research (I2R), Singapore, where he was involved in an industrial project on developing an 802.11n Wireless LAN system and participated actively in 3Gpp LTE and LTE-Advanced (LTE-A) standardization. Since 2010, he has been with the Singapore University of Technology and Design. He was a recipient of the Lee Kuan Yew Gold Medal, the Institution of Electrical Engineers Book Prize, the Institute of Engineering of Singapore Gold Medal, the Merck Sharp and Dohme Gold Medal, and twice a recipient of the Hewlett Packard Prize. He is a fellow of IEEE.

GUI Guan received his doctoral degree in information and communication engineering from University of Electronic Science and Technology of China in 2012. Since 2015, he has been a professor with Nanjing University of Posts and Telecommunications (NJUPT), China. His recent research interests include artificial intelligence, deep learning, non-orthogonal multiple access, wireless power transfer, and physical layer security. He received the IEEE Communications Society Heinrich Hertz Award in 2021 and Highly Cited Chinese Researchers by Elsevier in 2020. In addition, he served as the IEEE VTS Ad Hoc Committee Member in AI Wireless, Executive Chair of VTC 2021-Fall, Vice Chair of WCNC 2021, TPC Chair of PHM 2021, Symposium Chair of WCSP 2021, and General Co-Chair of Mobimedia 2020.