



**QUEEN'S
UNIVERSITY
BELFAST**

Active RIS-aided NOMA-enabled space-air-ground integrated networks with cognitive radio

Li, J., Yang, L., Wu, Q., Lei, X., Zhou, F., Shu, F., Mu, X., Liu, Y., & Fan, P. (2025). Active RIS-aided NOMA-enabled space-air-ground integrated networks with cognitive radio. *IEEE Journal on Selected Areas in Communications*, 43(1), 314-333. <https://doi.org/10.1109/JSAC.2024.3460067>

Published in:

IEEE Journal on Selected Areas in Communications

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2024 the authors.

This is an accepted manuscript distributed under a Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

Active RIS-Aided NOMA-Enabled Space-Air-Ground Integrated Networks with Cognitive Radio

Junjie Li, Liang Yang, Qingqing Wu, Xianfu Lei, Fuhui Zhou, Feng Shu, Xidong Mu, Yuanwei Liu, Pingzhi Fan

Abstract—In this work, we investigate an active reconfigurable intelligent surface (RIS)-aided non-orthogonal multiple access (NOMA)-enabled space-air-ground integrated network (SAGIN) with cognitive radio, leveraging the flexible deployment of an unmanned aerial vehicle (UAV) and the ubiquitous coverage of satellite networks. The UAV serves uplink and downlink users in the secondary network via NOMA and time division multiple access mechanisms, respectively, while satellites provide wireless backhaul for the UAV and primary users. We aim to maximize the weighted sum mean rate and energy efficiency for the secondary network by jointly the optimizing power allocation, the RIS reflection coefficients (RC), the user matching factors, and the UAV trajectory. We propose an alternating optimization framework based on the block coordinate ascent (BCA) technique, which decouples the problem into multiple variable blocks for alternating optimization until convergence. Moreover, we investigate the performance of energy-efficient active RIS with a sub-connected architecture, decoupling the RIS RC optimization into amplification factor and phase shift subproblems to be solved separately. Finally, simulation results validate the effectiveness of the proposed schemes, and demonstrate weakness of passive RIS and rationality and economics of sub-connected active RIS architecture.

Index Terms—Space-air-ground-integrated network (SAGIN), active reconfigurable intelligent surface (RIS), cognitive radio, weighted sum mean rate, weighted sum mean energy efficiency.

I. INTRODUCTION

Recently, emerging network technologies, including the internet of things (IoT), cloud computing, and artificial intelligence (AI), have significantly advanced technological development and enhanced daily convenience, but have imposed stricter requirements on network traffic, spectrum resources,

and coverage breadth [1]. With the development of fifth-generation technology in full swing, terrestrial communication systems have been fairly sophisticated. Nevertheless, competition for spectrum resources in low-frequency bands has intensified, restricting the expansion of network capacity [2]. Concurrently, IoT technology, which is expected to drive future demand, envisions a future where everything is connected, overcoming traditional limitations to achieve coverage in remote and harsh environments and high-altitude areas [3].

As a leading architecture for the development of sixth-generation (6G) technology, the space-air-ground integrated network (SAGIN) offers a promising solution [4], [5]. Specifically, unmanned aerial vehicles (UAVs) serve as a crucial link between satellite and terrestrial communications, typically equipped with network in a box (NIB) systems that function as high-speed mobile servers that dynamically facilitate communication, collaboration, and networking in an ad hoc manner [6]. They can route packets for clients and relay sensor data to centralized control centers, thereby improving operational efficiency. The SAGIN leverages the strengths of both UAVs and satellites, enabling seamless global coverage. Notably, the dynamic regulation of UAVs makes them well-suited for scenarios with limited energy supply, as they can significantly reduce the communication link distance with ground equipment, thereby improving energy efficiency (EE) [7]. Furthermore, smart devices can leverage cognitive radio (CR) technology to share the high-frequency bands of satellites, effectively overcoming the challenge of spectrum scarcity [8].

However, the signal stability of the primary user (PU) remains vulnerable to degradation resulting from unpredictable wireless electromagnetic wave transmission. Given the higher priority of spectrum utilization by the PU, the performance of the secondary user (SU) is critically limited for improved communication of the PU [9]. As an innovative technology for 6G, active reconfigurable intelligent surface (RIS) provide an effective solution for directional control of electromagnetic wave radiation. The internal design of active RIS comprises intricate phase-shifting circuitry for precise control over the radiation direction of incoming signals, coupled with a power amplifier to bolster these signals [10]. Consequently, it shines in CR networks by enhancing network capacity and suppressing interference from secondary networks to the primary network [11]. Crucially, active RIS effectively overcomes the multiplicative fading effect inherent to passive RIS and is suitable for serving UAVs and ground devices over long distances [12]. Motivated by the prospective superiority of active

J. Li and L. Yang are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: iejjli@hnu.edu.cn; liangy@hnu.edu.cn).

Q. Wu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, 200240, China (e-mail: qingqingwu@sjtu.edu.cn).

X. Lei and P. Fan are with the School of Information Science and Technology, Institute of Mobile Communications, Southwest Jiaotong University, Chengdu 610031, China (e-mail: xfleil@home.swjtu.edu.cn; p.fan@ieee.org).

F. Zhou is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: zhoufuhui@ieee.org).

F. Shu is with the School of Information and Communication Engineering and Collaborative Innovation Center of Information Technology, Hainan University, Haikou 570228, China, and also with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China (e-mail: shufeng0101@163.com).

X. Mu is with the Centre for Wireless Innovation (CWI), Queen's University Belfast, Belfast, BT3 9DT, U.K. (e-mail: x.mu@qub.ac.uk).

Y. Liu is with the Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong (e-mail: yuanwei@hku.hk).

RIS, this paper investigates the communication performance of SAGINs based on active RIS with CR.

A. State-of-the-Art

It has been demonstrated that SAGINs can enhance network throughput and reduce energy consumption via the dynamic deployment of UAVs and resource optimization [13]-[16]. For instance, by considering the non-orthogonal multiple access (NOMA) technique for UAVs, a block coordinate descent (BCD) method was proposed to maximize the EE of the SAGIN [13]. The authors in [14] developed a SAG-IoT framework with multiple UAVs as relays, and proposed an algorithm to jointly optimize device scheduling, power control, and trajectory optimization to maximize system capacity. To simultaneously satisfy the demand for high-rate and low-computing latency, a scheme was proposed where UAVs function as base stations (BSs) or mobile edge computing servers, while satellites serve as clouds for continuous computation [15]. In [16], a secure spectrum sharing strategy based on a lightweight UAV blockchain was proposed, utilizing the sequential Vickrey auction mechanism of blockchain. However, these solutions remain susceptible to performance degradation in uncontrolled radio environments.

Moreover, the emerging RIS technology enhances UAV-assisted communications by intelligently configuring the radiative phase of the meta-atoms [17]-[19]. For example, the authors in [17] designed an aerial RIS with flexible deployment capabilities, and proposed a scheme based on 3D beam spreading and flattening techniques. In [18], a UAV system equipped with RIS operating under the downlink NOMA mechanism was studied, and an algorithm based on double deep Q-Network was proposed for ensuring the endurance of the UAV and the capacity of the system. Additionally, the authors in [19] explored a hybrid full-duplex RIS-UAV communication that determines the operating states of the RIS and the UAV by selecting combinations. However, these approaches are unable to utilize high-frequency band resources from satellite communications and have limited coverage.

Recently, initial attempts on RIS-aided SAGINs investigations are underway [20], [21]. For instance, in [20], the authors investigated SAGINs utilizing a hybrid free-space optics/radio frequency (RF) high-altitude platform, inferred a mathematical model for the comprehensive effects of atmospheric turbulence and hover, and proposed a link-switching mechanism that adapts to varying weather turbulence. In view of the substantial Doppler associated with RIS, the authors in [21] converted the time-varying fading encountered in the time domain into the delay-Doppler domain, and proposed a robust 2D orthogonal time-frequency space modulation scheme. Unfortunately, due to the multiplicative fading factor of passive RIS, its efficacy in long-distance SAGINs is limited, particularly with fewer meta-atoms. In contrast, active RIS offers an effective solution.

B. Our Contributions

In view of the above analysis, recent research has primarily focused on the model design and signal processing of passive

RIS-aided SAGINs, while resource allocation remains under-explored. Notably, there is a significant gap in studying active RIS-aided NOMA-enabled SAGINs with CR, particularly in jointly optimizing the two key performance metrics of rate and EE. To fill these gaps, this paper investigates the problem of maximizing the weighted sum mean rate (WSMR) and weighted sum mean EE (WSMEE) for active RIS-aided NOMA-enabled SAGINs with CR. Our main contributions can be summarized as follows:

- We construct an architecture for an active RIS-aided NOMA-enabled SAGIN with CR, where PUs are served directly via satellite, while secondary uplink users (SUUs) and secondary downlink users (SDUs) access the UAV for relay to the satellite by NOMA and time-division multiple access (TDMA), respectively. We formulate the optimization problem aimed at maximizing WSMR and WSMEE by jointly optimizing power allocation, UAV flight trajectory, RIS reflection coefficients (RC), and matching factors under the constraints of kinematics, backhaul link capacity, maximum power, amplification factor, interference tolerance, and minimum quality of service (QoS) requirements. The optimization problem is challenging due to the intricate interdependencies among these variables.
- To this end, we propose a joint optimization framework based on the block coordinate ascent (BCA) technique to decouple the original problem. We employ the Dinkelbach technique to subtract the objective function and transform the blocks into tractable convex problems based on Lagrangian dual reformulation, fractional programming (FP) and successive convex approximation (SCA), and constrained concave-convex procedure (CCCP) methods. Additionally, we optimize matching factors with mixed-integer non-linear programming (MINLP) based on the linear conic relaxation (LCR) technique.
- Moreover, to enhance EE and reduce hardware costs, we implement a sub-connected active RIS architecture to reduce assembled amplifiers. We decouple the optimization of RIS RC into an amplification factor block and a phase shift block, and propose an energy-efficient optimization framework. Furthermore, we establish a no-fly zone (NFZ) within the PU distribution region.
- Finally, the numerical results unveil that: 1) the proposed schemes significantly enhance WSMR and WSMEE compared to orthogonal and fixed trajectory optimization schemes, while balancing performance and overhead with a reasonable number of loaded UAV users; 2) passive RIS exhibits limited effectiveness in the network, whereas active RIS yields substantial improvements; 3) the sub-connected architecture of the active RIS has significantly improved WSMEE compared to the fully-connected architecture, and the impact of the freedom sacrifice is very limited.

The rest of this paper is organized as follows. Section II proposes a system model of an active RIS-aided NOMA-enabled SAGIN with CR, and formulates the problem of maximizing WSMR and WSMEE. A BCA-based joint optimization

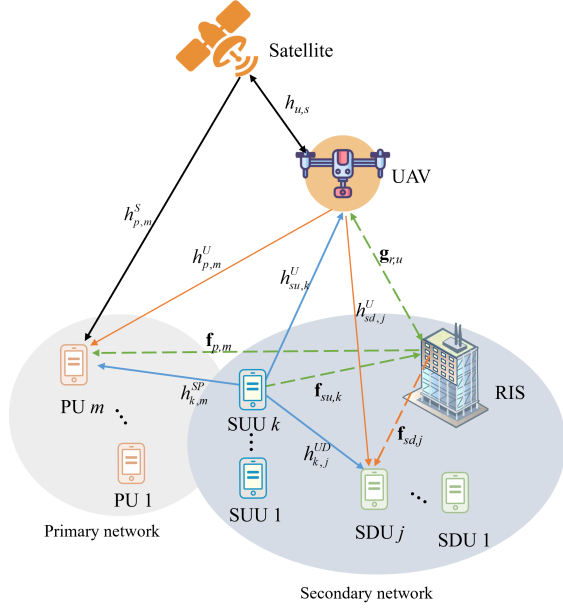


Fig. 1. Active RIS-aided NOMA-enabled SAGINs with CR.

framework is proposed in Section III. Section IV proposes an optimization framework for sub-connected active RIS. The frameworks are analyzed and supplemented in Section V. Simulations are performed in Section VI. Finally, conclusions are given in Section VII.

Notation: In this paper, boldface letters indicate matrices or vectors. $\mathbb{R}^{M \times N}$ and $\mathbb{C}^{M \times N}$ symbolize real and complex-valued matrix of size $M \times N$, respectively. $[\cdot]^T$, $[\cdot]^H$, $|\cdot|$, $\|\cdot\|$, and $\Re(x)$ symbolize transpose, conjugate transpose, modulus, Euclidean norm, and real part, respectively. $\angle$, $\otimes$, and $\odot$ symbolize the angle, Kronecker, and Hadamard product, respectively. For a vector $\mathbf{x}$, $[\mathbf{x}]_i$ represents its i -th element, and $\text{diag}(\mathbf{x})$ represents the diagonal operation. For a matrix $\mathbf{A}$, $\mathbf{A}_{i,j}$ is the element in its i -th row and j -th column, and $\mathbf{A} \succeq 0$ denotes the semi-definite matrix. $\mathbf{1}_N$ is a vector of length N with all elements 1, and ∇_a is the gradient of a .

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we introduce the RIS-aided NOMA-enabled SAGIN with CR and formulate its resource allocation optimization problem.

A. System Architecture

As illustrated in Fig. 1, an active RIS-aided SAGIN with NOMA is considered. This practically corresponds to the case that BS power systems are not serviceable in remote or post-disaster outbreak areas, where SAGINs are promising to ensure the communication quality of the devices in the area [22]–[24]. On the basis of the different needs of users, they can be divided into real-time users and non-real-time users, which are categorized as authorized primary networks and unauthorized secondary networks, respectively. Specifically, the primary network consists of a low-earth-orbit satellite (primary BS) directly serving M PUs, while the secondary

TABLE I
SUMMARY OF NOTATIONS

Notation	Description
$h_{su,k}^U, h_{sd,j}^U, h_{p,m}^U \in \mathbb{C}$	Channel between UAV and k -th SUU, j -th SDU, m -th PU
$h_{k,j}^{UD}, h_{k,m}^{SP} \in \mathbb{C}$	Channel between k -th SUU and m -th PU, j -th SDU
$\mathbf{g}_{r,u}, \mathbf{f}_{su,k}, \mathbf{f}_{sd,j}, \mathbf{f}_{p,m} \in \mathbb{C}^{N_t \times 1}$	Channel between RIS and UAV, k -th SUU, j -th SDU, m -th PU
$h_{p,m}^S, h_{u,s} \in \mathbb{C}$	Channel between satellite and m -th PU, UAV
$\boldsymbol{\theta}, \boldsymbol{\Theta}$	Vector and matrix of RIS RC
$\bar{\boldsymbol{\theta}}, \mathbf{b}$	Phase shift and amplification factor vectors for RIS
$s_{su,k}, s_{sd,j}, s_{p,m}$	Information symbols for k -th SUU, j -th SDU, and m -th PU
$p_{su,k}, p_{sd,j}, p_{p,m}$	Transmit power for k -th SUU, UAV to j -th SDU, and satellite to m -th PU
$\mathbf{n}_R, P_R$	Thermal noise and maximum transmit power of RIS
$\mathbf{q}[l], v^{xy}[l]$	2D horizontal coordinate and speed of the UAV at the l -th slot
$n_U, n_{sd,j}, n_{p,m}$	Complex additive white Gaussian noise (AWGN) for UAV, j -th SDU, and m -th PU
$\sigma_U^2, \sigma_{sd,j}^2, \sigma_{p,m}^2, \sigma_R^2$	Noise variance at UAV, j -th SDU, m -th PU, and RIS

network involves a UAV acting as a relay between the satellite and ground equipment, facilitating communication for K SUUs and J SDUs.¹ Let $\mathcal{M} = \{1, \dots, M\}$, $\mathcal{K} = \{1, \dots, K\}$ and $\mathcal{J} = \{1, \dots, J\}$ denote the index of PUs, SUUs and SDUs, respectively. The main parameters of the system are summarized in TABLE I. Assuming that the satellite operates within the same frequency band as the ground user and UAV communications,² and SUUs and SDUs are permitted to utilize the spectrum licensed to PUs subject to the interference tolerance of the PUs. Without loss of generality, PUs, SUUs and SDUs are randomly distributed within a circle area, where the 3D coordinates of the m -th PU, k -th SUU and j -th SDU can be denoted as $\mathbf{S}_{p,m} = (x_{p,m}, y_{p,m}, 0)$, $\mathbf{S}_{su,k} = (x_{su,k}, y_{su,k}, 0)$ and $\mathbf{S}_{sd,j} = (x_{sd,j}, y_{sd,j}, 0)$, respectively, and the coordinates of RIS are $\mathbf{S}_R = (x_R, y_R, z_R)$. The horizontal position and altitude of the UAV vary with time, and its coordinate at the t -th time are denoted as $\mathbf{S}_U(t) = (x_U(t), y_U(t), z_U(t))$, $0 \leq t \leq T$.

For convenience, we discretize T into L equal slots, each of duration Δt . Let $\mathcal{L} = \{1, \dots, L\}$ denote the index of L . When L is large enough, the UAV's position can be approximated as constant within each slot. Thus, the 3D and horizontal positions of the UAV in the l -th slot can be denoted as $\mathbf{S}_U[l] = (x_U[l], y_U[l], z_U[l])$ and $\mathbf{q}[l] = (x_U[l], y_U[l])^T$, respectively. Without loss of generality, assume that the UAV flies at a fixed altitude $z_U[l] = z'_U, \forall l \in \mathcal{L}$, with a maximum horizontal velocity of $v_{\max}^{xy}$. To accommodate hardware limi-

¹For PUs, we focus on the received signal quality, and thus uplink transmissions are not considered [9], [25]. SDUs and SUUs are designed for single receive/transmit capabilities. However, these devices can be extended to support simultaneous transmitting and receiving capabilities, where the interference from the SUU to the SDU would be replaced by corresponding self-interference.

²Standardization groups (e.g., ETSI) and researchers have proposed that cognitive radio satellite-terrestrial networks in the L/C band allow satellite networks to coexist with terrestrial networks within the same frequency band, thereby addressing the spectrum resource shortage [26].

tations and users' periodicity requirements, the UAV adheres to the following constraints:

$$C1: v^{xy}[l] = \frac{\|\mathbf{q}[l+1] - \mathbf{q}[l]\|}{\Delta t} \leq v_{\max}^{xy}, \forall l \in \mathcal{L}, \quad (1)$$

with $\mathbf{q}[0] = \mathbf{q}[L]$.

To suppress interference and enhance the quality of received signals, an active RIS technique is employed. Consider the ideal RIS case with RC matrix $\Theta = \text{diag}(\theta)$, where $\theta = [\theta_1, \dots, \theta_{N_t}]^T$, $\theta_n = b_n \bar{\theta}_n$, and $\bar{\theta}_n = e^{j\vartheta_n}$. Here, b_n and ϑ_n denote the amplitude and phase of the n -th meta-atom of the RIS, respectively, and N_t denotes the number of meta-atoms indexed by $\mathcal{N} = \{1, \dots, N_t\}$. Notably, the RC of the RIS can be adjusted to enhance the system performance of the secondary network while mitigating the interference to the primary network.

B. Channel Model

1) *Air-Ground Access Link*: All ground users are considered quasi-static within each slot, while the UAV flies at a constant speed. When communicating between the UAV and a ground user, the signal received is a combination of line-of-sight (LoS) and non-line-of-sight (NLoS) paths [27]. Since the LoS path predominates in air-ground communication, we approximate the LoS path probability as 1 for simplicity.³ Therefore, the channel model between the RIS and the UAV can be represented as

$$\mathbf{g}_{r,u}[l] = c/(4\pi d_R^U[l] f_c) \eta_R^U \mathbf{g}_{r,u}^*[l], \quad (2)$$

where f_c , c , and η_R^U symbolize the carrier frequency, the speed of light, and the attenuation weight, respectively, and $d_R^U[l]$ is the distance between the RIS and the UAV denoted by $d_R^U[l] = \sqrt{\|\mathbf{q}[l] - \mathbf{t}_R\|^2 + (z'_U - z_R)^2}$, wherein $\mathbf{t}_R$ symbolizes the horizontal coordinate of the RIS denoted by $\mathbf{t}_R = (x_R, y_R)^T$. Assuming that the RIS is a uniform planar array with $N_t = N_1 \times N_2$, $\mathbf{g}_{r,u}^*[l]$ can be defined as

$$\mathbf{g}_{r,u}^*[l] = \left[1, e^{-j\frac{2\pi}{\lambda_w} d_f n_R^U[l]}, \dots, e^{-j\frac{2\pi}{\lambda_w} (N_1-1) d_f n_R^U[l]} \right]^T \otimes \left[1, e^{-j\frac{2\pi}{\lambda_w} d_f m_R^U[l]}, \dots, e^{-j\frac{2\pi}{\lambda_w} (N_2-1) d_f m_R^U[l]} \right]^T, \quad (3)$$

where d_f and λ_w denote the element spacing and carrier wavelength, respectively. Additionally, $n_R^U = \sin(\theta_R^U) \cos(\phi_R^U)$ and $m_R^U = \sin(\phi_R^U)$, where θ_R^U and ϕ_R^U denote the azimuth and elevation angles, respectively.

2) *Ground-Ground Access Link*: Here, the links between the RIS and the ground users and between the ground users are considered. Assume that all channel state information is known and quasi-static flat fading over one cycle. We consider the channels as Rician fading, in which case $\mathbf{f}_{su,k}[l]$ can be expressed as

$$\mathbf{f}_{su,k}[l] = \sqrt{L_0} (d_{r,k})^{-\tau/2} \mathbf{f}_{su,k}^*[l], \quad (4)$$

³Based on [28] and [29], for suitable UAV altitude and carrier frequency, the LoS path probability approximation can be achieved as 1. Additionally, for less dense environments (e.g., remote areas), satisfactory accuracy is attained with the LoS path approximation [30]. The assumption of a LoS channel has been widely adopted in UAV trajectory optimisation work [14], [30].

where τ and L_0 denote the path loss exponent and path loss at reference 1 m, respectively, and $d_{r,k}$ denotes the distance. $\mathbf{f}_{su,k}^*[l]$ symbolises the small-scale fading denoted by $\mathbf{f}_{su,k}^*[l] = \sqrt{\frac{\alpha^*}{1+\alpha^*}} \mathbf{f}_{su,k}^{\text{LoS}}[l] + \sqrt{\frac{1}{1+\alpha^*}} \mathbf{f}_{su,k}^{\text{NLoS}}[l]$, where α^* is the Rician factor, $\mathbf{f}_{su,k}^{\text{LoS}}$ corresponds to (3), and $\mathbf{f}_{su,k}^{\text{NLoS}}[l]$ represents Rayleigh fading.

3) *Space-Air/Ground Backhaul Link*: Since the altitude and cycle time of the UAV is significantly smaller than that of the satellite, the effects of the satellite's speed and frequent hand-offs can be disregarded for one cycle, and the UAV is recognized as being approximately stationary relative to the satellite [15], [24]. Considering the effects of antenna gain, path loss, and rain attenuation [15], the channel between the UAV and the satellite can be expressed as

$$h_{u,s}[l] = \frac{c}{4\pi f_c \tilde{h}_S} \sqrt{G_u^t G_s^r \psi_{ra}}, \quad (5)$$

where G_u^t and G_s^r symbolize the transmitting antenna gain of the UAV and the receiving antenna gain of the satellite, respectively, and ψ_{ra} symbolizes the rain attenuation, $\tilde{h}_S$ symbolizes the altitude difference between the satellite and the UAV.

C. Transmission Schemes

To enhance the uplink access sum rate, the UAV serves the SUUs via the power domain NOMA protocol and operates in full duplex mode. Assuming the UAV is load limited, it can receive transmissions from up to K_s users per slot, where $K_s \leq K$. Let $a_k[l]$ denote the matching factor, such that $a_k[l] = 1$ if and only if the UAV serves the k -th SUU in the k -th slot, otherwise vice versa. Therefore, SUUs uplink access to the UAV has the following constraints:

$$C2: \sum_{k \in \mathcal{K}} a_k[l] \leq K_s, a_k[l] \in \{0, 1\}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \quad (6)$$

For convenience, we define

$$\begin{aligned} \bar{h}_{su,k}[l] &= h_{su,k}^U[l] + \mathbf{g}_{r,u}^H[l] \Theta^H[l] \mathbf{f}_{su,k}[l], \\ \bar{h}_{sd,j}[l] &= h_{sd,j}^U[l] + \mathbf{f}_{sd,j}^H[l] \Theta^H[l] \mathbf{g}_{r,u}[l], \\ \bar{h}_{k,j}^{UD}[l] &= h_{k,j}^{UD}[l] + \mathbf{f}_{sd,j}^H[l] \Theta^H[l] \mathbf{f}_{su,k}[l], \\ \bar{h}_{p,m}^U[l] &= h_{p,m}^U[l] + \mathbf{f}_{p,m}^H[l] \Theta^H[l] \mathbf{g}_{r,u}[l], \\ \bar{h}_{k,m}^{SP}[l] &= h_{k,m}^{SP}[l] + \mathbf{f}_{p,m}^H[l] \Theta^H[l] \mathbf{f}_{su,k}[l]. \end{aligned} \quad (7)$$

Neglecting self-interference [13], the signal received by the UAV in the secondary network can be expressed as

$$\begin{aligned} y_U[l] &= \underbrace{\sum_{k \in \mathcal{K}} a_k[l] \sqrt{p_{su,k}[l]} \bar{h}_{su,k}[l] s_{su,k}[l]}_{\text{Desired signal}} \\ &\quad + \underbrace{\mathbf{g}_{r,u}^H[l] \Theta^H[l] \mathbf{n}_R}_{\text{Interference from RIS}} + n_U, \end{aligned} \quad (8)$$

where $\mathbb{E}\{|s_{su,k}[l]|^2\} = 1$, $n_U \sim \mathcal{CN}(0, \sigma_U^2)$ means the AWGN at the UAV, and $\mathbf{n}_R \sim \mathcal{CN}(0, \sigma_R^2 \mathbf{I})$.

The information symbol $s_{sd,j}$ transmitted by the satellite to the j -th SDU is forwarded via the UAV relay operating decode and forward (DF) protocol [16]. Given the UAV's antenna limitation and the suppression of inter-user interference, and

assuming SDUs access the UAV under a TDMA protocol,⁴ the received signal at the j -th SDU is denoted as

$$\gamma_{sd,j}[l] = \underbrace{a_j[l] \sqrt{p_{sd,j}[l]} \bar{h}_{sd,j}[l] s_{sd,j}[l]}_{\text{Desired signal}} + \underbrace{\mathbf{f}_{sd,j}^H[l] \mathbf{\Theta}^H[l] \mathbf{n}_R}_{\text{Interference from RIS}} + \underbrace{\sum_{k \in \mathcal{K}} \sqrt{p_{su,k}[l]} \bar{h}_{k,j}^{UD}[l] s_{su,k}[l]}_{\text{Interference from SUUs}} + n_{sd,j}, \quad (9)$$

where $n_{sd,j} \sim \mathcal{CN}(0, \sigma_{sd,j}^2)$, and a_j means the matching factor between the UAV and the j -th SDU, satisfying

$$C3: \sum_{j \in \mathcal{J}} a_j[l] = 1, a_j[l] \in \{0, 1\}, \forall j \in \mathcal{J}, \forall l \in \mathcal{L}. \quad (10)$$

The signal received by the m -th PU can be expressed as

$$y_{p,m}[l] = \underbrace{h_{p,m}^S[l] \sqrt{p_{p,m}[l]} s_{p,m}[l]}_{\text{Signal from satellite}} + \underbrace{\mathbf{f}_{p,m}^H[l] \mathbf{\Theta}^H[l] \mathbf{n}_R}_{\text{Interference from RIS}} + \underbrace{\sum_{j \in \mathcal{J}} a_j[l] \sqrt{p_{sd,j}[l]} \bar{h}_{p,m}^U[l] s_{sd,j}[l]}_{\text{Interference from UAV}} + \underbrace{\sum_{k \in \mathcal{K}} \sqrt{p_{su,k}[l]} \bar{h}_{k,m}^{SP}[l] s_{su,k}[l]}_{\text{Interference from SUUs}} + n_{p,m}, \quad (11)$$

where $n_{p,m} \sim \mathcal{CN}(0, \sigma_{p,m}^2)$ means the AWGN at the receiver.

We employ successive interference cancellation (SIC) at the UAV receiver to eliminate co-channel interference [33], and the decoding order of the k -th SUU is defined by $\pi(k)$. Specifically, SUUs with higher channel gain power are prioritized for decoding, while the i -th SUU with $\pi(i) > \pi(k)$ is treated as the interference for the k -th SUU. The signal-to-interference-plus-noise ratio (SINR) by the k -th SUU accessing the UAV is expressed as

$$\gamma_{su,k}[l] = \frac{\bar{p}_{su,k}[l]}{\sum_{\pi(i) > \pi(k), i \in \mathcal{K}} \bar{p}_{su,i}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2}, \quad (12)$$

where $\bar{p}_{su,k}[l] = a_k[l] p_{su,k}[l] |\bar{h}_{su,k}[l]|^2$ and $\mathcal{I}_{R,U}^{\text{int}}[l] = \sigma_R^2 \|\mathbf{g}_{r,u}^H[l] \mathbf{\Theta}[l]\|^2$. The SINR for the j -th SDU is indicated as

$$\gamma_{sd,j}[l] = \frac{\bar{p}_{sd,j}[l]}{\mathcal{I}_{j,su}^{\text{int}}[l] + \mathcal{I}_{R,j}^{\text{int}}[l] + \sigma_{sd,j}^2}, \quad (13)$$

where $\bar{p}_{sd,j}[l] = a_j[l] p_{sd,j}[l] |\bar{h}_{sd,j}[l]|^2$, $\mathcal{I}_{R,j}^{\text{int}}[l] = \sigma_R^2 \|\mathbf{f}_{sd,j}^H[l] \mathbf{\Theta}[l]\|^2$, and $\mathcal{I}_{j,su}^{\text{int}}[l] = \sum_{k \in \mathcal{K}} p_{su,k}[l] |\bar{h}_{k,j}^{UD}[l]|^2$.

⁴For uplink transmission, employing the NOMA technique with a single-antenna receiver for multiple users is a common assumption, which improves spectral efficiency and capacity [31]. Conversely, a single-antenna transmitter with downlink NOMA technology poses challenges in terms of interference management, power control, and signal processing, which are exacerbated by the simultaneous implementation of NOMA in both uplink and downlink. To this end, we employ the TDMA protocol for UAV downlink transmission to reduce interference and decoding complexity, and improve spectral efficiency [32].

To ensure the QoS for PUs, the secondary network restricts the interference power transmitted to the m -th PU as

$$C4: \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] |\bar{h}_{p,m}^U[l]|^2 + \sigma_R^2 \|\mathbf{f}_{p,m}^H[l] \mathbf{\Theta}[l]\|^2 + \sum_{k \in \mathcal{K}} p_{su,k}[l] |\bar{h}_{k,m}^{SP}[l]|^2 \leq P_{tol,m}, \forall m \in \mathcal{M}, \forall l \in \mathcal{L}, \quad (14)$$

where $P_{tol,m}$ represents the maximum interference tolerance from the m -th PU to the secondary network.

The achievable throughput for SUUs and SDUs can be denoted as

$$B_{su} R_{su}[l] = B_{su} \sum_{k \in \mathcal{K}} \log(1 + \gamma_{su,k}[l]), \quad (15a)$$

$$B_{sd} R_{sd}[l] = B_{sd} \sum_{j \in \mathcal{J}} \log(1 + \gamma_{sd,j}[l]), \quad (15b)$$

where B_{su} and B_{sd} symbolize the transmit bandwidth of the SUUs and SDUs, respectively.

The achievable throughput for the UAV-to-satellite backhaul link can be expressed as

$$\hat{R}_S[l] = B_S \log(1 + \gamma_S[l]) = B_S R_S[l], \quad (16)$$

where B_S symbolizes the bandwidth of the backhaul link, $\gamma_S[l] = p_S[l] |h_{s,u}[l]|^2 / \sigma_S^2$ symbolizes the signal-to-noise ratio (SNR) at the satellite, and $p_S[l]$ symbolizes the transmit power of the UAV.

D. Energy Consumption Pattern

The energy consumption of UAV during flight is formulated as [30]

$$P_{\text{flight}}[l] = P_{\text{profile}}[l] + P_{\text{induced}}[l] + P_{\text{parasite}}[l], \quad (17)$$

where P_{profile} , P_{induced} and P_{parasite} symbolize the profile power, induced power and parasitic power, respectively. They are utilized to overcome resistance to the rotation of the propeller blades, to generate downward air thrust, and to counteract resistance to forward motion, respectively, as denoted by

$$P_{\text{profile}}[l] = c_0 \left(1 + \frac{3(v^{xy}[l])^2}{\Omega^2 R_a^2} \right), \quad (18a)$$

$$P_{\text{induced}}[l] = c_1 k_0 \left(\sqrt{k_0^2 + \frac{(v^{xy}[l])^4}{4v_0^4}} - \frac{(v^{xy}[l])^4}{2v_0^2} \right)^{\frac{1}{2}}, \quad (18b)$$

$$P_{\text{parasite}}[l] = c_2 (v^{xy}[l])^3, \quad (18c)$$

where c_0 , c_1 and v_0 symbolize the profile power parameter, induced power parameter and mean induced velocity in the hovering state, respectively, k_0 and c_2 denote the thrust-to-weight ratio and parasitic power parameter, respectively, and Ω and R_a symbolize the blade angular velocity and rotor radius, respectively.

The transmit power used for communication by SUUs, UAV and RIS in the secondary network is

$$P_{\text{comm}} = \sum_{k \in \mathcal{K}} p_{su,k}[l] + \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] + \sigma_R^2 \|\mathbf{\Theta}\|_F^2 + \sum_{k \in \mathcal{K}} p_{su,k}[l] \|\mathbf{\Theta}[l] \mathbf{f}_{su,k}[l]\|^2 + \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] \|\mathbf{\Theta}[l] \mathbf{g}_{r,u}[l]\|^2. \quad (19)$$

The hardware-dissipated power in the secondary network is

$$P_{\text{diss}} = KP_{su}^c + JP_{sd}^c + N_t P_{ps}^r + BP_{pa}^r, \quad (20)$$

where P_{su}^c , P_{sd}^c , P_{ps}^r and P_{pa}^r denote the dissipated power by each SUU, SDU, RIS phase shifter and RIS amplifier, respectively, and B denotes the number of amplifiers.

The WSMEE for one cycle in the secondary network is defined by

$$\eta_{\text{EE}} \triangleq \frac{\sum_{l \in \mathcal{L}} R[l]}{\sum_{l \in \mathcal{L}} P_{\text{sum}}[l]}, \quad (21)$$

where $P_{\text{sum}}[l] = \varsigma_U P_{\text{flight}}[l] + P_{\text{comm}}[l] + P_{\text{diss}}[l]$, ς_U symbolizes the weight of UAV flight energy consumption, $R[l] = \sum_k \eta_{su}[l] \log(1 + \gamma_{su,k}[l]) + \sum_j \eta_{sd}[l] \log(1 + \gamma_{sd,j}[l])$ symbolizes the weighted sum rate in the l -th slot, wherein $\eta_{su}[l]$ and $\eta_{sd}[l]$ are weights assigned to SUUs and SDUs for prioritization.

E. Problem Formulation

Based on the above analysis, we jointly optimize the transmit power $\mathbf{p}_{su}$ and $\mathbf{p}_{sd}$ of SUUs and UAV, the RC $\boldsymbol{\theta}$ of the active RIS, the matching factors $\mathbf{a}_{su}$ and $\mathbf{a}_{sd}$ for SUUs and SDUs access to the UAV, and the flight trajectory $\mathbf{q}$ of the UAV, and formulate the problem of maximizing the WSMR

in the secondary network as

$$\mathcal{P}_R^o : \max_{\mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{p}_{su}, \mathbf{p}_{sd}, \boldsymbol{\theta}, \mathbf{q}} \mathcal{T}(\mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{p}_{su}, \mathbf{p}_{sd}, \boldsymbol{\theta}, \mathbf{q}) \quad (22a)$$

$$\text{s.t. } C1, C2, C3, C4, \quad (22b)$$

$$C5 : \sum_{j \in \mathcal{J}} p_{sd,j}[l] \leq p_{sd,\max}, \forall l \in \mathcal{L}, \quad (22c)$$

$$C6 : 0 \leq p_{su,k}[l] \leq p_{su,\max}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (22d)$$

$$C7 : |\theta_n| \leq \omega_{n,\max}, \forall n \in \mathcal{N}, \quad (22e)$$

$$C8 : \sum_{k \in \mathcal{K}} p_{su,k}[l] \|\boldsymbol{\Theta}[l] \mathbf{f}_{su,k}[l]\|^2 + \sigma_R^2 \|\boldsymbol{\Theta}[l]\|_F^2 \quad (22f)$$

$$+ \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] \|\boldsymbol{\Theta}[l] \mathbf{g}_{r,u}[l]\|^2 \leq P_R, \forall l \in \mathcal{L},$$

$$C9 : B_{su} R_{su}[l] \leq B_S R_S[l], \forall l \in \mathcal{L}, \quad (22g)$$

$$C10 : \sum_{l \in \mathcal{L}} a_k[l] \geq L_{su}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (22h)$$

$$C11 : \sum_{l \in \mathcal{L}} a_j[l] \geq L_{sd}, \forall j \in \mathcal{J}, \forall l \in \mathcal{L}, \quad (22i)$$

where the objective function $\mathcal{T}(\mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{p}_{su}, \mathbf{p}_{sd}, \boldsymbol{\theta}, \mathbf{q}) \triangleq \frac{1}{L} \sum_{l \in \mathcal{L}} R[l]$. Here, $p_{sd,\max}$, $p_{su,\max}$, and P_R in $C5$, $C6$, and $C8$ symbolize the maximum transmit power of SUUs, UAV, and active RIS, respectively. $C7$ defines the upper bound $\omega_{n,\max}$ for the RIS amplifier. $C9$ ensures that the information rate of SUUs transmitting to the UAV is less than the maximum access rate of the UAV to the satellite backhaul link to prevent blocking. $C10$ and $C11$ guarantee that the UAV is accessed by each SUU (L_{su}) and SDU (L_{sd}) at least a minimum number of times within one flight cycle to ensure service fairness among users.

Additionally, the problem of maximizing the WSMEE of the secondary network can be formulated based on the above analysis as

$$\mathcal{P}_E^o : \max_{\mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{p}_{su}, \mathbf{p}_{sd}, \boldsymbol{\theta}, \mathbf{q}} \eta_{\text{EE}} \quad (23a)$$

$$\text{s.t. } C1 - C11, \quad (23b)$$

$$C12 : \sum_{k \in \mathcal{K}} \log(1 + \gamma_{su,k}[l]) \geq R_{su}^{\min}, \forall l \in \mathcal{L}, \quad (23c)$$

$$C13 : \sum_{j \in \mathcal{J}} \log(1 + \gamma_{sd,j}[l]) \geq R_{sd}^{\min}, \forall l \in \mathcal{L}, \quad (23d)$$

$$\begin{aligned} \hat{\mathcal{T}}(\boldsymbol{\Xi}, \boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd}) &= \sum_{l \in \mathcal{L}} (\eta_{su}[l] \log(1 + \rho_{su}[l]) - \eta_{su}[l] \rho_{su}[l] + \eta_{su}[l] (1 + \rho_{su}[l]) \Upsilon_{su}(\boldsymbol{\Xi})) \\ &\quad + \sum_{l \in \mathcal{L}} \sum_{j \in \mathcal{J}} (\eta_{sd}[l] \log(1 + \rho_{sd,j}[l]) - \eta_{sd}[l] \rho_{sd,j}[l] + \eta_{sd}[l] (1 + \rho_{sd,j}[l]) \Upsilon_{sd,j}(\boldsymbol{\Xi})). \end{aligned} \quad (29)$$

$$\begin{aligned} \tilde{\mathcal{T}}(\boldsymbol{\Xi}, \boldsymbol{\zeta}_{su}, \boldsymbol{\zeta}_{sd}) &= \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \underbrace{\left(2\Re \left\{ \sqrt{\eta_{su}[l] \bar{p}_{su,k}} \zeta_k^H[l] \right\} - |\zeta_k[l]|^2 \left(\sum_{k \in \mathcal{K}} \bar{p}_{su,k} + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2 \right) \right)}_{\tilde{R}_{su}[l]} \\ &\quad + \sum_{l \in \mathcal{L}} \sum_{j \in \mathcal{J}} \underbrace{\left(2\Re \left\{ \sqrt{\eta_{sd,j}[l] \bar{p}_{sd,j}} \zeta_j^H[l] \right\} - |\zeta_j[l]|^2 (\bar{p}_{sd,j} + \mathcal{I}_{j,su}^{\text{int}}[l] + \mathcal{I}_{R,j}^{\text{int}}[l] + \sigma_{sd,j}^2) \right)}_{\tilde{R}_{sd}[l]}. \end{aligned} \quad (33)$$

where $C12$ and $C13$ specify the minimum QoS requirements $R_{su}^{\min}$ and $R_{sd}^{\min}$ for SUUs and SDUs in each slot, respectively. Notably, when $P_{\text{sum}}[l] = 1$, and constraints $C12$ and $C13$ are disregarded, $\mathcal{P}_E^o$ is equivalent to $\mathcal{P}_R^o$. Without loss of generality, we focus on solving $\mathcal{P}_E^o$.

III. PROPOSED JOINT OPTIMIZATION SCHEME

Solving $\mathcal{P}_R^o$ and $\mathcal{P}_E^o$ is extremely challenging due to the complex non-convexity of the objective function, the severe coupling of $C4$, $C8$, $C9$, $C12$ and $C13$, the zero-one constraints of $C2$ and $C3$, and the decoding order of NOMA. Hence, in this section, we propose a joint optimization framework to solve $\mathcal{P}_E^o$. Specifically, we decompose $\mathcal{P}_E^o$ into several variable blocks to be solved by alternating optimization, as outlined in **Algorithm 1**.

For simplicity, we initially examine the sum rate of SUUs as

$$\begin{aligned} & \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \eta_{su}[l] \log \left(1 + \frac{\bar{p}_{su,k}[l]}{\sum_{\pi(i) > \pi(k)} \bar{p}_{su,i}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2} \right) \\ & \stackrel{(a)}{=} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \eta_{su}[l] \log \left(\frac{\sum_{\pi(i) \geq \pi(k)} \bar{p}_{su,i}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2}{\sum_{\pi(i) > \pi(k)} \bar{p}_{su,i}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2} \right) \\ & \stackrel{(b)}{=} \sum_{l \in \mathcal{L}} \eta_{su}[l] \log \left(1 + \frac{\sum_{k \in \mathcal{K}} a_k[l] p_{su,k}[l] |\bar{h}_{su,k}|^2}{\underbrace{\mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2}_{\gamma_{su}[l]}} \right). \end{aligned} \quad (24)$$

where (b) performs mutual elimination of K multiplicative fractions. One can notice that the sum rate of SUUs is independent of the decoding order of NOMA.

A. Transform Objective Function

Due to the fractional structure of the objective function, $\mathcal{P}_E^o$ poses exceptional challenges in solution. To this end, we employ the FP method to reformulate it into a subtractive form for solving, leading to the following proposition.

Proposition 1: By applying Dinkelbach's transform and introducing the auxiliary variable $\xi \in \mathbb{R}$ [35], the original problem $\mathcal{P}_E^o$ in (23) can be equivalently represented as

$$\tilde{\mathcal{P}}^o : \max_{\xi, \mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{p}_{su}, \mathbf{p}_{sd}, \boldsymbol{\theta}, \mathbf{q}} \sum_{l \in \mathcal{L}} R[l] - \xi \sum_{l \in \mathcal{L}} P_{\text{sum}}[l] \quad (25a)$$

$$\text{s.t. } C1 - C13, \quad (25b)$$

where ξ is updated to

$$\xi^* = \left(\sum_{l \in \mathcal{L}} R'[l] \right) / \left(\sum_{l \in \mathcal{L}} P'_{\text{sum}}[l] \right), \quad (26)$$

wherein $R'[l]$ and $P'_{\text{sum}}[l]$ symbolize the values of $R[l]$ and $P_{\text{sum}}[l]$ from the last iteration, respectively.

Based on the Lagrangian dual reformulation [34], one can obtain

$$\begin{aligned} \sum_i \eta_i \log \left(1 + \frac{\nu_{i,1}}{\nu_{i,2}} \right) & \geq \sum_i \eta_i \log(1 + \rho_i) \\ & - \sum_i \eta_i \rho_i + \sum_i \eta_i (\rho_i + 1) \frac{\nu_{i,1}}{\nu_{i,1} + \nu_{i,2}}, \end{aligned} \quad (27)$$

where $\eta_i, \nu_{i,1}, \rho_i \geq 0, \nu_{i,2} > 0$. By introducing auxiliary variables $\boldsymbol{\rho}_{su} \in \mathbb{R}^L$ and $\boldsymbol{\rho}_{sd} \in \mathbb{R}^{J \times L}$ with $\boldsymbol{\rho}_{sd}[l] = [\rho_{sd,1}[l], \dots, \rho_{sd,J}[l]]^T$, $\mathcal{P}_E^o$ can be reformulated with ξ fixed based on the above inequality as

$$\hat{\mathcal{P}}^o : \max_{\Xi, \boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd}} \hat{\mathcal{T}}(\Xi, \boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd}) - \xi \sum_{l \in \mathcal{L}} P_{\text{sum}}[l] \quad (28a)$$

$$\text{s.t. } C1 - C13, \quad (28b)$$

where $\Xi \triangleq \{\mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{p}_{su}, \mathbf{p}_{sd}, \boldsymbol{\theta}, \mathbf{q}\}$, and $\hat{\mathcal{T}}(\Xi, \boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd})$ is defined as (29), wherein functions $\Upsilon_{su}(\Xi)$ and $\Upsilon_{sd,j}(\Xi)$ are denoted by

$$\Upsilon_{su}(\Xi) = \frac{\sum_{k \in \mathcal{K}} \bar{p}_{su,k}[l]}{\sum_{k \in \mathcal{K}} \bar{p}_{su,k}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2}, \quad (30a)$$

$$\Upsilon_{sd,j}(\Xi) = \frac{\bar{p}_{sd,j}[l]}{\bar{p}_{sd,j}[l] + \mathcal{I}_{j,su}^{\text{int}}[l] + \mathcal{I}_{R,j}^{\text{int}}[l] + \sigma_{sd,j}^2}. \quad (30b)$$

To solve $\hat{\mathcal{P}}^o$, we utilize the BCA technique that alternately optimizes each variable block while keeping the other blocks invariant, as specified in the following subsections.

B. Optimizing $(\boldsymbol{\rho}_{su}^*, \boldsymbol{\rho}_{sd}^*)$

Fixing Ξ , the function $\hat{\mathcal{T}}(\Xi, \boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd})$ is a differentiable concave function w.r.t. $\boldsymbol{\rho} \in \{\boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd}\}$. By solving for the zero partial derivative, i.e., $\partial \hat{\mathcal{T}}(\Xi, \boldsymbol{\rho}) / \partial \boldsymbol{\rho} = 0$, the optimal solutions can be given as

$$\rho_{su}^*[l] = \gamma_{su}[l], \forall l \in \mathcal{L}. \quad (31a)$$

$$\rho_{sd,j}^*[l] = \gamma_{sd,j}[l], \forall j \in \mathcal{J}, \forall l \in \mathcal{L}. \quad (31b)$$

Bringing $\boldsymbol{\rho}_{su}^*$ and $\boldsymbol{\rho}_{sd}^*$ back to $\hat{\mathcal{T}}$ reveals that the value of $\hat{\mathcal{T}}$ is only pertinent to $\Upsilon_{su}(\Xi)$ and $\Upsilon_{sd,j}(\Xi)$. Unfortunately, the problem remains challenging to solve due to the non-convex nature of the sum of multiple fractions. To address this, we apply the multiple-ratio FP method to reformulate $\hat{\mathcal{P}}^o$ [35], leading to the following *Proposition 2*.

Proposition 2: By applying quadratic variation and introducing two auxiliary variables, $\boldsymbol{\zeta}_{su} \in \mathbb{R}^{K \times L}$ with $\boldsymbol{\zeta}_{su}[l] = [\zeta_1[l], \dots, \zeta_K[l]]^H$, and $\boldsymbol{\zeta}_{sd} \in \mathbb{R}^{J \times L}$ with $\boldsymbol{\zeta}_{sd}[l] = [\zeta_1[l], \dots, \zeta_J[l]]^H$, $\forall l \in \mathcal{L}$, $\hat{\mathcal{T}}(\Xi, \boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd})$ can be defined in the following form:

$$\tilde{\mathcal{P}}^o : \max_{\Xi, \boldsymbol{\zeta}_{su}, \boldsymbol{\zeta}_{sd}} \tilde{\mathcal{T}}(\Xi, \boldsymbol{\zeta}_{su}, \boldsymbol{\zeta}_{sd}) + h(\boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd}) \quad (32a)$$

$$\text{s.t. } C1 - C13, \quad (32b)$$

where $\tilde{\mathcal{T}}(\Xi, \boldsymbol{\zeta}_{su}, \boldsymbol{\zeta}_{sd})$ and $h(\boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd})$ are defined by (33) and (34), respectively, wherein $\tilde{\eta}_{su}[l] = \eta_{su}[l](1 + \rho_{su}[l])$, $\tilde{\eta}_{sd,j}[l] = \eta_{sd}[l](1 + \rho_{sd,j}[l])$.

$$\begin{aligned} h(\boldsymbol{\rho}_{su}, \boldsymbol{\rho}_{sd}) &= \sum_{l \in \mathcal{L}} \underbrace{(\eta_{su}[l](\log(1 + \rho_{su}[l]) - \rho_{su}[l]))}_{\tilde{h}(\boldsymbol{\rho}_{su}[l])} \\ &+ \sum_{l \in \mathcal{L}} \sum_{j \in \mathcal{J}} \underbrace{(\eta_{sd,j}[l](\log(1 + \rho_{sd,j}[l]) - \rho_{sd,j}[l]))}_{\tilde{h}(\boldsymbol{\rho}_{sd}[l])}. \end{aligned} \quad (34)$$

As a result, $\tilde{\mathcal{P}}^o$ is transformed into a tractable subtractive form, which can be solved by updating $(\boldsymbol{\zeta}_{su}, \boldsymbol{\zeta}_{sd})$ and Ξ .

C. Optimizing (ζ_{su}^* , ζ_{sd}^*)

Given Ξ , by setting $\partial \tilde{T}(\Xi, \zeta_{su}, \zeta_{sd}) / \partial \zeta_{su} = 0$ and $\partial \tilde{T}(\Xi, \zeta_{su}, \zeta_{sd}) / \partial \zeta_{sd} = 0$, the optimal $(\zeta_{su}^*, \zeta_{sd}^*)$ can be determined as

$$\zeta_k^*[l] = \frac{\sqrt{\tilde{\eta}_{su}[l] \tilde{p}_{su,k}[l]}}{\sum_{k \in \mathcal{K}} \tilde{p}_{su,k}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_U^2}, \quad (35a)$$

$$\zeta_j^*[l] = \frac{\sqrt{\tilde{\eta}_{sd,j}[l] \tilde{p}_{sd,j}[l]}}{\tilde{p}_{sd,j}[l] + \mathcal{I}_{j,su}^{\text{int}}[l] + \mathcal{I}_{R,j}^{\text{int}}[l] + \sigma_{sd,j}^2}. \quad (35b)$$

D. Power Allocation Optimization ($\mathbf{p}_{su}$, $\mathbf{p}_{sd}$)

Fixing the other variables to optimize $\mathbf{p}_{su}$ and $\mathbf{p}_{sd}$, (33) becomes $\tilde{T}(\mathbf{p}_{su}, \mathbf{p}_{sd})$. Removing the irrelevant terms, $\mathcal{P}_E^o$ can be reformulated as

$$\mathcal{P}_1 : \max_{\mathbf{p}_{su}, \mathbf{p}_{sd}} \tilde{T}(\mathbf{p}_{su}, \mathbf{p}_{sd}) - \xi^* \sum_{l \in \mathcal{L}} P_{\text{sum}}[l] \quad (36a)$$

$$\text{s.t. } C4, C5, C6, C8, \quad (36b)$$

$$C9 : \gamma_{su}[l] \leq 2^{B_{RS}[l]/B_{su}} - 1, \forall l \in \mathcal{L}, \quad (36c)$$

$$\hat{C}12 : \tilde{R}_{su}[l] + \tilde{h}(\rho_{su}[l]) \geq R_{su}^{\min}, \forall l \in \mathcal{L}, \quad (36d)$$

$$\hat{C}13 : \tilde{R}_{sd}[l] + \tilde{h}(\rho_{sd}[l]) \geq R_{sd}^{\min}, \forall l \in \mathcal{L}, \quad (36e)$$

where $\hat{C}12$ and $\hat{C}13$ are obtained by relaxation via (33) and (34). Consequently, $\mathcal{P}_1$ is a standard convex problem, and the optimal solution can be readily obtained by modern solvers, e.g., CVX.

E. RIS Reflection Coefficient Optimization θ

Fixing the other variables to optimize θ , (33) becomes $\tilde{T}(\theta)$. Then, $\mathcal{P}_E^o$ can be reformulated as

$$\mathcal{P}_2 : \max_{\theta} \tilde{T}(\theta) - \xi^* \sum_{l \in \mathcal{L}} P_{\text{sum}}[l] \quad (37a)$$

$$\text{s.t. } C4, C7, C8, C9, C12, C13. \quad (37b)$$

For clarity, by defining $\tilde{\mathbf{f}}_{su,k}[l] = \text{diag}(\mathbf{g}_{r,u}^H[l]) \mathbf{f}_{su,k}[l]$, $\tilde{\mathbf{f}}_{sd,j}[l] = \text{diag}(\mathbf{f}_{sd,j}^H[l]) \mathbf{g}_{r,u}[l]$, $\tilde{\mathbf{f}}_{k,j}^{UD}[l] = \text{diag}(\mathbf{f}_{sd,j}^H[l]) \mathbf{f}_{su,k}[l]$, $\tilde{\mathbf{f}}_{p,m}[l] = \text{diag}(\mathbf{f}_{p,m}^H[l]) \mathbf{g}_{r,u}[l]$, $\tilde{\mathbf{f}}_{k,m}^{SP}[l] = \text{diag}(\mathbf{f}_{p,m}^H[l]) \mathbf{f}_{su,k}[l]$, and

$$\begin{aligned} \Lambda[l] &= \Lambda_{su}[l] + \Lambda_{sd}[l] + \xi^* \Lambda_R[l], \\ \mathbf{v}[l] &= \mathbf{v}_{su}[l] + \mathbf{v}_{sd}[l], \quad o[l] = o_{su}[l] + o_{sd}[l], \end{aligned} \quad (38)$$

where the individual symbol meanings are defined as (39), wherein

$$\mathbf{w}_k^u[l] = \sqrt{a_k[l] p_{su,k}[l]} \tilde{\mathbf{f}}_{su,k}[l], \quad \mathbf{w}_j^d[l] = \sqrt{a_j[l] p_{sd,j}[l]} \tilde{\mathbf{f}}_{sd,j}[l],$$

$$c_k^u[l] = \sqrt{a_k[l] p_{su,k}[l] h_{su,i}^U[l]}, \quad c_j^d[l] = \sqrt{a_j[l] p_{sd,j}[l] h_{sd,j}^U[l]},$$

$$\mathbf{G}_{r,u}[l] = \text{diag}(\mathbf{g}_{r,u}^H[l]) \text{diag}(\mathbf{g}_{r,u}[l]),$$

$$\mathbf{F}_{sd,j}[l] = \text{diag}(\mathbf{f}_{sd,j}^H[l]) \text{diag}(\mathbf{f}_{sd,j}[l]),$$

$$\mathbf{F}_{su,k}[l] = \text{diag}(\mathbf{f}_{su,k}^H[l]) \text{diag}(\mathbf{f}_{su,k}[l]).$$

Then, substituting (38) for (37), $\mathcal{P}_2$ can be reformulated as

$$\min_{\theta} \sum_l \left(\theta^H[l] \Lambda[l] \theta[l] - 2\Re \left\{ \theta^H[l] \mathbf{v}[l] \right\} - o[l] \right) \quad (40a)$$

$$\text{s.t. } C4, C7, C9, \bar{C}8 : \theta^H[l] \Lambda_R[l] \theta[l] \leq P_R, \quad (40b)$$

$$\bar{C}12 : 2\Re \left\{ \mathbf{b}^H[l] \tilde{\mathbf{v}}_{su}[l] \right\} - \mathbf{b}^H[l] \tilde{\Lambda}_{su}[l] \mathbf{b}[l] \geq \tilde{o}_{su}[l], \quad (40c)$$

$$\bar{C}13 : 2\Re \left\{ \mathbf{b}^H[l] \tilde{\mathbf{v}}_{sd}[l] \right\} - \mathbf{b}^H[l] \tilde{\Lambda}_{sd}[l] \mathbf{b}[l] \geq \tilde{o}_{sd}[l], \quad (40d)$$

where $\tilde{o}_d[l] = R_d^{\min} - o_d[l] - \tilde{h}(\rho_d[l])$, $d \in \{su, sd\}$, $\forall l \in \mathcal{L}$.

Equivalently, $C4$ can be rewritten as

$$\bar{C}4 : \theta^H[l] \Lambda_m[l] \theta[l] + 2\Re \left\{ \theta^H[l] \mathbf{v}_m[l] \right\} + o_m[l] \leq P_{\text{tol},m}, \quad (41)$$

where

$$\begin{aligned} \Lambda_m[l] &= \sum_{k \in \mathcal{K}} p_{su,k}[l] \tilde{\mathbf{f}}_{k,m}^{SP}[l] \left(\tilde{\mathbf{f}}_{k,m}^{SP}[l] \right)^H + \sigma_R^2 \mathbf{F}_{p,m}[l] \\ &\quad + \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] \tilde{\mathbf{f}}_{p,m}[l] \tilde{\mathbf{f}}_{p,m}^H[l], \end{aligned}$$

$$\begin{aligned} \mathbf{v}_m[l] &= \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] h_{p,m}^U[l] \tilde{\mathbf{f}}_{p,m}[l] \\ &\quad + \sum_{k \in \mathcal{K}} p_{su,k}[l] h_{k,m}^{SP}[l] \tilde{\mathbf{f}}_{k,m}^{SP}[l], \end{aligned}$$

$$o_m[l] = \sum_{k \in \mathcal{K}} p_{su,k}[l] |h_{k,m}^{SP}[l]|^2 + \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] |h_{p,m}^U[l]|^2,$$

wherein $\mathbf{F}_{p,m}[l] = \text{diag}(\mathbf{f}_{p,m}^H[l]) \text{diag}(\mathbf{f}_{p,m}[l])$.

Similarly, $C9$ can be rewritten as

$$\begin{aligned} &\theta^H[l] \Lambda_U[l] \theta[l] + 2\Re \left\{ \theta^H[l] \mathbf{v}_U[l] \right\} + o_U[l] \\ &\leq \varepsilon_U[l] \left(\sigma_R^2 \theta^H[l] \mathbf{G}_{r,u}[l] \theta[l] + \sigma_U^2 \right), \quad \forall l \in \mathcal{L}, \end{aligned} \quad (42)$$

$$\Lambda_{su}[l] = \sum_{k \in \mathcal{K}} |\zeta_k[l]|^2 \left(\sum_{i \in \mathcal{K}} \mathbf{w}_i^u[l] (\mathbf{w}_i^u[l])^H + \sigma_R^2 \mathbf{G}_{r,u}[l] \right), \quad (39a)$$

$$\Lambda_{sd}[l] = \sum_{j \in \mathcal{J}} |\zeta_j[l]|^2 \left(\mathbf{w}_j^d[l] (\mathbf{w}_j^d[l])^H + \sigma_R^2 \mathbf{F}_{sd,j}[l] + \sum_{k \in \mathcal{K}} p_{su,k}[l] \tilde{\mathbf{f}}_{k,j}^{UD}[l] (\tilde{\mathbf{f}}_{k,j}^{UD}[l])^H \right), \quad (39b)$$

$$\Lambda_R[l] = \sum_{k \in \mathcal{K}} p_{su,k}[l] \mathbf{F}_{su,k}[l] + \sigma_R^2 \mathbf{I}_{N_t} + \sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] \mathbf{G}_{r,u}[l], \quad (39c)$$

$$\mathbf{v}_{su}[l] = \sum_{k \in \mathcal{K}} \left(\sqrt{\tilde{\eta}_{su}[l]} \mathbf{w}_k^u[l] \zeta_k^H[l] - |\zeta_k[l]|^2 \sum_{i \in \mathcal{K}} c_i^u[l] \mathbf{w}_i^u[l] \right), \quad (39d)$$

$$\mathbf{v}_{sd}[l] = \sum_{j \in \mathcal{J}} \left(\sqrt{\tilde{\eta}_{sd,j}[l]} \mathbf{w}_j^d[l] \zeta_j^H[l] - |\zeta_j[l]|^2 \left(c_j^d[l] \mathbf{w}_j^d[l] + \sum_{k \in \mathcal{K}} p_{su,k}[l] h_{k,j}^{UD}[l] \tilde{\mathbf{f}}_{k,j}^{UD}[l] \right) \right), \quad (39e)$$

$$o_{su}[l] = 2\Re \left\{ \sum_{k \in \mathcal{K}} \sqrt{\tilde{\eta}_{su}[l]} \zeta_k^u[l] \zeta_k^H[l] \right\} - \sum_{k \in \mathcal{K}} |\zeta_k[l]|^2 \sum_{i \in \mathcal{K}} \left(|c_i^u[l]|^2 + \sigma_U^2 \right), \quad (39f)$$

$$o_{sd}[l] = 2\Re \left\{ \sum_{j \in \mathcal{J}} \sqrt{\tilde{\eta}_{sd,j}[l]} c_j^u[l] \zeta_j^H[l] \right\} - \sum_{j \in \mathcal{J}} |\zeta_j[l]|^2 \left(|c_j^d[l]|^2 + \sum_{k \in \mathcal{K}} p_{su,k}[l] |h_{k,j}^{UD}[l]|^2 + \sigma_{sd,j}^2 \right). \quad (39g)$$

$$\begin{aligned}
R[l] = & \underbrace{\eta_{su}[l] \ln \left(\sum_{k \in \mathcal{K}} \left(\frac{\chi_{su,k}^D[l]}{(d_{su,k}^U[l])^2} + \frac{\chi_{su,k}^R[l]}{(d_R^U[l])^2} + \frac{2\Re\{D_{su,k}^R[l]\}}{d_{su,k}^U[l]d_R^U[l]} \right) + \frac{\chi_{r,u}^R[l]}{(d_R^U[l])^2} + \sigma_U^2 \right)}_{f_1(\mathbf{q}[l])} - \underbrace{\eta_{su}[l] \ln \left(\frac{\chi_{r,u}^R[l]}{(d_R^U[l])^2} + \sigma_U^2 \right)}_{f_2(\mathbf{q}[l])} \\
& + \underbrace{\eta_{sd}[l] \sum_{j \in \mathcal{J}} \ln \left(1 + \frac{\chi_{sd,j}^D[l]}{(d_{sd,j}^U[l])^2} + \frac{\chi_{sd,j}^R[l]}{(d_R^U[l])^2} + \frac{2\Re\{D_{sd,j}^R[l]\}}{d_{sd,j}^U[l]d_R^U[l]} \right)}_{f_3(\mathbf{q}[l])} = f_1(\mathbf{q}[l]) - f_2(\mathbf{q}[l]) + f_3(\mathbf{q}[l]), \forall l \in \mathcal{L}.
\end{aligned} \tag{51}$$

where

$$\begin{aligned}
\Lambda_U[l] &= \sum_{k \in \mathcal{K}} \mathbf{w}_k^u[l] (\mathbf{w}_k^u[l])^H, \\
\mathbf{v}_U[l] &= \sum_{k \in \mathcal{K}} c_k^u[l] \mathbf{w}_k^u[l], o_U[l] = \sum_{k \in \mathcal{K}} |c_k^u[l]|^2, \\
\varepsilon_U[l] &= \exp \left(\frac{B_S}{B_{su}} \log(1 + \gamma_S[l]) \right) - 1.
\end{aligned}$$

Since the right-hand side of inequality (42) is convex, we apply a first-order Taylor expansion to the transformation as

$$\boldsymbol{\theta}^H[l] \mathbf{G}_{r,u}[l] \boldsymbol{\theta}[l] \geq (\boldsymbol{\theta}^H[l])^{[t]} \mathbf{G}_{r,u}[l] (2\boldsymbol{\theta}[l] - \boldsymbol{\theta}^{[t]}[l]), \quad (43)$$

where $\boldsymbol{\theta}^{[t]}[l]$ represents the value from the previous iteration of $\boldsymbol{\theta}[l]$. By Substituting (43) into (42), we obtain

$$\begin{aligned}
\bar{C}9 : & \boldsymbol{\theta}^H[l] \Lambda_U[l] \boldsymbol{\theta}[l] + 2\Re\{\boldsymbol{\theta}^H[l] \mathbf{v}_U[l]\} + o_U[l] \leq \\
\varepsilon_U[l] & \left(\sigma_R^2 \boldsymbol{\theta}^H[l] \mathbf{G}_{r,u}[l] (2\boldsymbol{\theta}[l] - \boldsymbol{\theta}^{[t]}[l]) + \sigma_U^2 \right), \forall l \in \mathcal{L}.
\end{aligned} \tag{44}$$

Synchronizing the above, $\mathcal{P}_2$ can be reformulated as

$$\bar{\mathcal{P}}_2 : \min_{\boldsymbol{\theta}} \sum_{l \in \mathcal{L}} \left(\boldsymbol{\theta}^H[l] \Lambda[l] \boldsymbol{\theta}[l] - 2\Re\{\boldsymbol{\theta}^H[l] \mathbf{v}[l]\} \right) \tag{45a}$$

$$\text{s.t. } \bar{C}4, C7, \bar{C}8, \bar{C}9, \bar{C}12, \bar{C}13. \tag{45b}$$

Noticeably, $\bar{\mathcal{P}}_2$ constitutes a quadratically constrained quadratic programming (QCQP) problem, for which the optimal solution can be efficiently obtained by optimization tools such as CVX and primal-dual subgradient (PDS) [36].

F. UAV Trajectory Optimization $\mathbf{q}$

Here, with fixed other variables optimizing the UAV trajectory, $\mathcal{P}_E^o$ can be rewritten as

$$\mathcal{P}_3 : \max_{\mathbf{q}} \tilde{\mathcal{T}}(\mathbf{q}) - \xi^* \sum_{l \in \mathcal{L}} P_{\text{sum}}[l] \tag{46a}$$

$$\text{s.t. } C1, C4, C8, C9, C12, C13. \tag{46b}$$

Since the energy model of the UAV involves a complex function of its coordinates, it necessitates suitable deformation. Practically, the thrust-to-weight ratio k_0 is approximated to be 1 due to the vertical force equilibrium. Hence, it can be determined by

$$P_{\text{induced}}[l] = c_1 \left(\sqrt{1 + \frac{(v^{xy}[l])^4}{4v_0^4}} - \frac{(v^{xy}[l])^4}{2v_0^2} \right)^{\frac{1}{2}} = c_1 \varpi[l], \tag{47}$$

where $\varpi[l]$ is an element of the auxiliary variable $\boldsymbol{\varpi} \in \mathbb{R}^L$ with $\boldsymbol{\varpi} = [\varpi[1], \dots, \varpi[L]]^T$. By conversion, one obtains

$$\frac{1}{(\varpi[l])^2} = (\varpi[l])^2 + \frac{(v^{xy}[l])^2}{v_0^2}. \tag{48}$$

Let $f_{\text{in}}(\varpi[l]) = (\varpi[l])^2 - 1/(\varpi[l])^2 + (v^{xy}[l])^2/v_0^2$, which is a monotonically increasing function about $\varpi[l]$ for $\varpi[l] > 0$. For tractability, let

$$f_{\text{in}}(\varpi[l]) \geq 0. \tag{49}$$

(49) provides a suitable relaxation of (48) without affecting the optimal value of the objective function. The constraint in (48) achieves equality when the objective function attains its optimal value, and thus this is equivalent.

However, (49) remains jointly non-convex for $v^{xy}[l]$ and $\varpi[l]$. Therefore, we apply a first-order Taylor expansion to derive its global lower bound as

$$\begin{aligned}
& \frac{2(\mathbf{q}'[l] - \mathbf{q}'[l-1])^T (\mathbf{q}[l] - \mathbf{q}[l-1])}{v_0^2 \Delta t} - \frac{\|\mathbf{q}'[l] - \mathbf{q}'[l-1]\|}{v_0^2 \Delta t} \\
& + \varpi'[l] (2\varpi[l] - \varpi'[l]) - 1/(\varpi[l])^2 \geq 0, \forall l \in \mathcal{L},
\end{aligned} \tag{50}$$

where $\mathbf{q}'$ and ϖ' represent the values from the last iteration of $\mathbf{q}$ and ϖ , respectively.

Next, we formulate $R[l]$ equivalently as (51), where

$$\begin{aligned}
C_D &= \frac{c}{4\pi f_c}, \chi_{su,k}^D[l] = a_k[l] p_{su,k}[l] C_D^2, \\
\chi_{su,k}^R[l] &= a_k[l] p_{su,k}[l] |C_D \eta_R^U \mathbf{e}_{r,u}^H[l] \boldsymbol{\Theta}^H[l] \mathbf{f}_{su,k}[l]|^2, \\
\chi_{r,u}^R[l] &= \sigma_R^2 |C_D \eta_R^U \mathbf{e}_{r,u}^H[l] \boldsymbol{\Theta}[l]|^2, \\
\chi_{sd,j}^D[l] &= \frac{a_j[l] p_{sd,j}[l] C_D^2}{\mathcal{I}_{j,su}^{\text{int}}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_{sd,j}^2}, \\
\chi_{sd,j}^R[l] &= \frac{a_j[l] p_{sd,j}[l] |C_D \eta_R^U \mathbf{f}_{sd,j}^H[l] \boldsymbol{\Theta}^H[l] \mathbf{e}_{r,u}[l]|^2}{\mathcal{I}_{j,su}^{\text{int}}[l] + \mathcal{I}_{R,U}^{\text{int}}[l] + \sigma_{sd,j}^2}, \\
D_{su,k}^R[l] &= a_k[l] p_{su,k}[l] C_D^2 \eta_R^U \mathbf{e}_{r,u}^H[l] \boldsymbol{\Theta}^H[l] \mathbf{f}_{su,k}[l], \\
D_{sd,j}^R[l] &= a_j[l] p_{sd,j}[l] C_D^2 \eta_R^U \mathbf{f}_{sd,j}^H[l] \boldsymbol{\Theta}^H[l] \mathbf{e}_{r,u}[l],
\end{aligned} \tag{52}$$

wherein $\mathbf{e}_{r,u}[l]$ symbolizes the array response vector of $\mathbf{g}_{r,u}^H[l]$.

Remark 1: It is noticed that $\mathbf{e}_{r,u}$ and $\mathbf{q}$ are highly coupled, rendering $\mathcal{P}3$ exceptionally challenging to solve. Fortunately, phase alignment can be achieved in practice by optimal pre-coding design, resulting in a very limited effect of small-scale fading [17] [37]. To simplify, we approximate $\mathbf{e}_{r,u}$ using its

last iteration value $\mathbf{e}'_{r,u}$. Since $\mathbf{q}$ typically changes minimally per iteration, we approximate $\mathbf{e}_{r,u} \approx \mathbf{e}'_{r,u}$, which has a negligible impact on the optimization outcome. Subsequently, optimizing $\mathbf{q}$, we can evaluate the actual performance with the actual $\mathbf{e}_{r,u}$.

Since $1/(d_i^U[l])^2$ is convex w.r.t. $d_i^U[l]$, $i \in \{(su, k), (sd, j), (R)\}$, one can derive its lower bound as

$$\begin{aligned} 1/(d_i^U[l])^2 &\geq 1/(\hat{d}_i^U[l])^2 - (\hat{d}_i^U[l] - d_i^U[l]) / (\hat{d}_i^U[l])^3 \\ &\triangleq \Gamma_i(d_i^U[l]), i \in \{(su, k), (sd, j), (R)\}, \end{aligned} \quad (53)$$

where $\hat{d}_i^U[l]$ symbolizes the last iteration value of $d_i^U[l]$. Introducing auxiliary variables $\alpha_{su} \in \mathbb{R}^{K \times L}$, $\alpha_{sd} \in \mathbb{R}^{J \times L}$, and $\alpha_R \in \mathbb{R}^L$ with $\alpha_{su}[l] \in [\alpha_{su,1}[l], \dots, \alpha_{su,K}[l]]^T$ and $\alpha_{sd}[l] \in [\alpha_{sd,1}[l], \dots, \alpha_{sd,J}[l]]^T$ such that

$$\Gamma_i(d_i^U[l]) \geq \alpha_i[l], i \in \{(su, k), (sd, j), (R)\}. \quad (54)$$

Remark 2: For a RIS-aided single-input single-output (SISO) system, optimal performance occurs when the cascade phase of the reflecting link coincides with the phase of the direct link. However, for a complex system, $\Re\{D_{su,k}^R[l]\}$ or $\Re\{D_{sd,j}^R[l]\}$ may be negative in (51), which destroys the concavity of $f_1(\mathbf{q}[l]) + f_3(\mathbf{q}[l])$ and necessitates separate analysis.

Case I: $\Re\{D_i^R[l]\} \geq 0$, $i \in \{(su, k), (sd, j)\}$. In this case, $1/(d_{su,k}^U[l]d_R^U[l])$ and $1/(d_{sd,j}^U[l]d_R^U[l])$ in (51) can be approximated by introducing auxiliary variables $\alpha_{su}^{R1} \in \mathbb{R}^{K \times L}$ and $\alpha_{sd}^{R1} \in \mathbb{R}^{J \times L}$ with $\alpha_{su}^{R1}[l] \in [\alpha_{su,1}^{R1}[l], \dots, \alpha_{su,K}^{R1}[l]]^T$ and $\alpha_{sd}^{R1}[l] \in [\alpha_{sd,1}^{R1}[l], \dots, \alpha_{sd,J}^{R1}[l]]^T$, combining (54) yields

$$(\alpha_{su,k}^{R1}[l])^2 / \alpha_{su,k}[l] \leq \alpha_R[l], \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (55a)$$

$$(\alpha_{sd,j}^{R1}[l])^2 / \alpha_{sd,j}[l] \leq \alpha_R[l], \forall j \in \mathcal{J}, \forall l \in \mathcal{L}. \quad (55b)$$

Then, the lower bounds of $f_1(\mathbf{q}[l])$ and $f_3(\mathbf{q}[l])$ are expressed respectively as

$$\begin{aligned} f_1(\mathbf{q}[l]) &\geq \eta_{su}[l] \ln \left(\sum_k (\chi_{su,k}^D[l] \alpha_{su,k}[l] + \chi_{su,k}^R[l] \alpha_R[l] \right. \\ &\quad \left. + 2\Re\{D_{su,k}^R[l]\} \alpha_{su,k}^{R1}[l] + \chi_{r,u}^R[l] \alpha_R[l] + \sigma_U^2) \right) \\ &= \tilde{f}_1(\alpha_{su,k}[l], \alpha_R[l], \alpha_{su,k}^{R1}[l]), \end{aligned} \quad (56a)$$

$$\begin{aligned} f_3(\mathbf{q}[l]) &\geq \eta_{sd}[l] \sum_j \ln(1 + \chi_{sd,j}^D[l] \alpha_{sd,j}[l] \\ &\quad + \chi_{sd,j}^R[l] \alpha_R[l] + 2\Re\{D_{sd,j}^R[l]\} \alpha_{sd,j}^{R1}[l]) \\ &= \tilde{f}_3(\alpha_{sd,j}[l], \alpha_R[l], \alpha_{sd,j}^{R1}[l]). \end{aligned} \quad (56b)$$

Evidently, $\tilde{f}_1$ and $\tilde{f}_3$ are joint concave functions for their corresponding variables. Significantly, the constraints $\tilde{f}_1$ and $\tilde{f}_3$ are equivalent to $f_1(\mathbf{q}[l])$ and $f_3(\mathbf{q}[l])$, respectively, i.e., the inequalities in (53), (54) and (55) must be satisfied at the optimal solution for the equation to hold. Alternatively, it is always feasible to increase α_i , α_i^{R1} , α_R , $i \in \{su, sd\}$ to increase the value of $\tilde{f}_1 + \tilde{f}_3$ without violating the constraints of $f_1(\mathbf{q}[l]) + f_3(\mathbf{q}[l])$.

Case II: $\Re\{D_i^R[l]\} < 0$, $i \in \{(su, k), (sd, j)\}$. We aim to maximize $R[n]$, and $2\Re\{D_i^R[l]\}/(d_i^U[l]d_R^U[l])$, $i \in \{su, sd\}$ are negative overall need to find $1/(d_i^U[l]d_R^U[l])$ upper bound.

For this, the following lemma can be invoked to transform $f_1(\mathbf{q}[l])$ and $f_3(\mathbf{q}[l])$ into solvable functions [38].

Lemma 1: For any a and b satisfying $a > 0$ and $b > 0$, we have

$$ab \leq \frac{a^{[t]}}{2b^{[t]}}b^2 + \frac{b^{[t]}}{2a^{[t]}}a^2, \quad (57)$$

where $a^{[t]}$ and $b^{[t]}$ correspond to fixed points around a and b , respectively.

Applying Lemma 1, one can obtain

$$\begin{aligned} \frac{1}{d_i^U[l]d_R^U[l]} &\leq \frac{(d_R^U[l])^{[t]}}{2(d_i^U[l])^{[t]}(d_R^U[l])^2} + \frac{(d_i^U[l])^{[t]}}{2(d_R^U[l])^{[t]}(d_i^U[l])^2} \\ &= \tilde{\Gamma}(d_i^U[l], d_R^U[l]), i \in \{(su, k), (sd, j)\}. \end{aligned} \quad (58)$$

Proposition 3: By introducing two auxiliary variables $\alpha_{su}^{R2} \in \mathbb{R}^{K \times L}$ and $\alpha_{sd}^{R2} \in \mathbb{R}^{J \times L}$ with $\alpha_{su}^{R2}[l] \in [\alpha_{su,1}^{R2}[l], \dots, \alpha_{su,K}^{R2}[l]]^T$ and $\alpha_{sd}^{R2}[l] \in [\alpha_{sd,1}^{R2}[l], \dots, \alpha_{sd,J}^{R2}[l]]^T$, $f_1(\mathbf{q}[l])$ and $f_3(\mathbf{q}[l])$ can be reformulated as

$$\begin{aligned} \tilde{f}_1(\alpha_{su,k}[l], \alpha_R[l], \alpha_{su,k}^{R2}[l]) &= \eta_{su}[l] \ln \left(\sum_k (\chi_{su,k}^D[l] \alpha_{su,k}[l] \right. \\ &\quad \left. + \chi_{su,k}^R[l] \alpha_R[l] + 2\Re\{D_{su,k}^R[l]\} \alpha_{su,k}^{R2}[l] \right. \\ &\quad \left. + \chi_{r,u}^R[l] \alpha_R[l] + \sigma_U^2) \right), \end{aligned} \quad (59a)$$

$$\begin{aligned} \tilde{f}_3(\alpha_{sd,j}[l], \alpha_R[l], \alpha_{sd,j}^{R2}[l]) &= \eta_{sd}[l] \sum_j \ln(\chi_{sd,j}^D[l] \alpha_{sd,j}[l] \\ &\quad + \chi_{sd,j}^R[l] \alpha_R[l] + 2\Re\{D_{sd,j}^R[l]\} \alpha_{sd,j}^{R2}[l] + 1), \end{aligned} \quad (59b)$$

where

$$\tilde{\Gamma}(\beta_{su,k}[l], \beta_R[l]) \leq \alpha_{su,k}^{R2}[l], \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (60a)$$

$$\tilde{\Gamma}(\beta_{sd,j}[l], \beta_R[l]) \leq \alpha_{sd,j}^{R2}[l], \forall j \in \mathcal{J}, \forall l \in \mathcal{L}, \quad (60b)$$

and

$$\begin{aligned} (\beta_{su,k}[l])^2 &\leq 2(\mathbf{S}'_U[l] - \mathbf{S}_{su,k}[l])(\mathbf{S}_U[l] - \mathbf{S}'_U[l])^T \\ &\quad + (d_{su,k}^U[l])^2, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \end{aligned} \quad (61a)$$

$$\begin{aligned} (\beta_{sd,j}[l])^2 &\leq 2(\mathbf{S}'_U[l] - \mathbf{S}_{sd,j}[l])(\mathbf{S}_U[l] - \mathbf{S}'_U[l])^T \\ &\quad + (d_{sd,j}^U[l])^2, \forall j \in \mathcal{J}, \forall l \in \mathcal{L}, \end{aligned} \quad (61b)$$

$$\begin{aligned} (\beta_R[l])^2 &\leq 2(\mathbf{S}'_U[l] - \mathbf{S}_R[l])(\mathbf{S}_U[l] - \mathbf{S}'_U[l])^T \\ &\quad + (d_R^U[l])^2, \forall l \in \mathcal{L}, \end{aligned} \quad (61c)$$

wherein $\beta_{su} \in \mathbb{R}^{K \times L}$, $\beta_{sd} \in \mathbb{R}^{J \times L}$, and $\beta_R \in \mathbb{R}^L$ are the introduced auxiliary variables with $\beta_{su}[l] = [\beta_{su,1}[l], \dots, \beta_{su,K}[l]]^T$ and $\beta_{sd}[l] = [\beta_{sd,1}[l], \dots, \beta_{sd,J}[l]]^T$, and $\mathbf{S}'_U[l]$ symbolizes the value of the last iteration of $\mathbf{S}_U[l]$.

Here, the approximation of $f_1(\mathbf{q}[l])$ and $f_3(\mathbf{q}[l])$ is completed. While $f_2(\mathbf{q}[l])$ is convex for $d_R^U[l]$, certain tools (e.g., CVX) do not support logarithmic functions to be internally non-concave, and thus we leverage the concavity of the logarithmic function for relaxation.

Proposition 4: Applying the concavity of the logarithmic function, $f_2(\mathbf{q}[l])$ in (51) can be redefined as

$$\tilde{f}_2(\beta_R[l]) = f_2(\hat{\mathbf{q}}[l]) + \eta_{su}[l] \frac{1/(\beta_R[l])^2 - 1/(\hat{d}_R^U[l])^2}{1/(\hat{d}_R^U[l])^2 + \sigma_U^2/\chi_{r,u}^R[l]}, \quad (62)$$

where $\hat{\mathbf{q}}[l]$ and $\hat{d}_R^U[l]$ symbolize the values of the last iteration of $\mathbf{q}[l]$ and $d_R^U[l]$, respectively, and (62) is constrained to (61c).

The constraint C8 can be reformulated as

$$\sum_{j \in \mathcal{J}} a_j[l] p_{sd,j}[l] \|\Theta[l] C_D \eta_R^U \mathbf{e}_{u,r}[l]\|^2 / (\beta_R[l])^2 + \sum_{k \in \mathcal{K}} p_{su,k}[l] \|\Theta[l] \mathbf{f}_{su,k}[l]\|^2 + \sigma_R^2 \|\Theta[l]\|_F^2 \leq P_R, \forall l \in \mathcal{L}. \quad (63)$$

For C4, this can be reformulated as

$$f_{p,m}(\mathbf{q}) = \sum_{j \in \mathcal{J}} \left(\frac{\chi_{j,m}^D[l]}{(d_{p,m}^U[l])^2} + \frac{\chi_{j,m}^R[l]}{(d_R^U[l])^2} + \frac{2\Re\{D_{j,m}^R[l]\}}{d_{p,m}^U[l]d_R^U[l]} \right) + \mathcal{I}_{m,sr}^{\text{int}}[l] \leq P_{tol,m}, \forall m \in \mathcal{M}, \quad (64)$$

$$\text{where } \mathcal{I}_{m,sr}^{\text{int}}[l] = \sum_k p_{su,k}[l] |\bar{h}_{k,m}^{SP}[l]|^2 + \sigma_R^2 \|\mathbf{f}_{p,m}^H[l] \Theta[l]\|^2,$$

$$\begin{aligned} \chi_{j,m}^D[l] &= a_j[l] p_{sd,j}[l] C_D^2, \\ \chi_{j,m}^R[l] &= a_j[l] p_{sd,j}[l] |C_D \eta_R^U \mathbf{f}_{p,m}^H[l] \Theta^H[l] \mathbf{e}_{r,u}[l]|^2, \\ D_{j,m}^R[l] &= a_j[l] p_{sd,j}[l] C_D^2 \eta_R^U \mathbf{f}_{p,m}^H[l] \Theta^H[l] \mathbf{e}_{r,u}[l]. \end{aligned} \quad (65)$$

Similar to the above manipulations, by introducing auxiliary variables β_p , β_p^{R1} , and β_p^{R2} with $\beta_p[l] = [\beta_{p,1}[l], \dots, \beta_{p,M}[l]]^T$, $\beta_p^{R1}[l] = [\beta_{p,1}^{R1}[l], \dots, \beta_{p,M}^{R1}[l]]^T$ and $\beta_p^{R2}[l] = [\beta_{p,1}^{R2}[l], \dots, \beta_{p,M}^{R2}[l]]^T$, then one obtains

$$(\beta_{p,m}[l])^2 \leq 2(\mathbf{S}'_U[l] - \mathbf{S}_{p,m}[l])(\mathbf{S}_U[l] - \mathbf{S}'_U[l])^T + (d_{p,m}^U[l])^2, \forall m \in \mathcal{M}, \forall l \in \mathcal{L}, \quad (66)$$

where $\beta_{p,m}[l]$ is a lower bound for $d_{p,m}^U[l]$. Next, two possible solutions are performed.

Case 1: $\Re\{D_{p,m}^R[l]\} \geq 0$. The function $f_{p,m}(\mathbf{q})$ in (64) can be reformulated as

$$\begin{aligned} \tilde{f}_{p,m}(\beta_{p,m}[l], \beta_R[l], \beta_{p,m}^{R1}[l]) &= \sum_{j \in \mathcal{J}} \left(\frac{\chi_{j,m}^D[l]}{(\beta_{p,m}[l])^2} \right. \\ &\quad \left. + \frac{\chi_{j,m}^R[l]}{(\beta_R[l])^2} + \frac{2\Re\{D_{j,m}^R[l]\}}{(\beta_{p,m}^{R1}[l])^2} \right) + \mathcal{I}_{m,sr}^{\text{int}}[l], \end{aligned} \quad (67)$$

where

$$(\beta_{p,m}^{R1}[l])^2 / \beta_{p,m}[l] \leq \beta_R[l], \forall m \in \mathcal{M}, \forall l \in \mathcal{L}. \quad (68)$$

Case 2: $\Re\{D_{p,m}^R[l]\} < 0$. Applying the first-order Taylor expansion, $f_{p,m}(\mathbf{q})$ can be reformulated as

$$\begin{aligned} \tilde{f}_{p,m}(\beta_{p,m}[l], \beta_R[l], \beta_{p,m}^{R2}[l]) &= \sum_{j \in \mathcal{J}} \left(\frac{\chi_{j,m}^D[l]}{(\beta_{p,m}[l])^2} \right. \\ &\quad \left. + \frac{\chi_{j,m}^R[l]}{(\beta_R[l])^2} + 2\Re\{D_{j,m}^R[l]\} \beta_{p,m}^{R2}[l] \right) + \mathcal{I}_{m,sr}^{\text{int}}[l], \end{aligned} \quad (69)$$

where

$$\beta_{p,m}^{R2}[l] \leq \frac{1}{\hat{d}_{p,m}^U[l] \hat{d}_R^U[l]} - \frac{d_{p,m}^U[l] - \hat{d}_{p,m}^U[l]}{(\hat{d}_{p,m}^U[l])^2 \hat{d}_R^U[l]} - \frac{d_R^U[l] - \hat{d}_R^U[l]}{\hat{d}_{p,m}^U[l] (\hat{d}_R^U[l])^2}, \quad (70)$$

wherein $\forall m \in \mathcal{M}, \forall l \in \mathcal{L}$, and $\hat{d}_{p,m}^U[l]$ and $\hat{d}_R^U[l]$ are the last iteration values of $d_{p,m}^U[l]$ and $d_R^U[l]$, respectively.

Based on the auxiliary variables introduced above, C9 can be redefined as

$$\begin{aligned} \sum_{k \in \mathcal{K}} \left(\frac{\chi_{su,k}^D[l]}{(\beta_{su,k}[l])^2} + \frac{\chi_{su,k}^R[l]}{(\beta_R[l])^2} + 2\Re\{D_{su,k}^R[l]\} \bar{\alpha}[l] \right) \\ \leq \varepsilon_U[l] (\chi_{r,u}^R[l] \alpha_R[l] + \sigma_U^2), \text{ if } \Re\{D_{su,k}^R[l]\} \geq 0, \\ \text{then } \bar{\alpha} = \alpha_{su,k}^{R2}, \text{ other } \bar{\alpha} = \alpha_{su,k}^{R1}. \end{aligned} \quad (71)$$

In light of the above, $\mathcal{P}_3$ can be reformulated as

$$\bar{\mathcal{P}}_3 : \max_{\mathbf{q}, \Xi_U, \varpi} \sum_{l \in \mathcal{L}} \left(\tilde{f}_1[l] - \tilde{f}_2[l] + \tilde{f}_3[l] - \xi^* P_{\text{sum}}(\varpi)[l] \right) \quad (72a)$$

$$\text{s.t. } C1, (50), (53-55), (60-61), (63), (66), (68), (70-71), \quad (72b)$$

$$\dot{C}4 : \tilde{f}_{p,m}[l] \leq P_{tol,m}, \forall m \in \mathcal{M}, \forall l \in \mathcal{L}, \quad (72c)$$

$$\dot{C}12 : \tilde{f}_1[l] - \tilde{f}_2[l] \geq R_{su}^{\min}, \forall l \in \mathcal{L}, \quad (72d)$$

$$\dot{C}13 : \tilde{f}_3[l] \geq R_{sd}^{\min}, \forall l \in \mathcal{L}, \quad (72e)$$

where Ξ_U symbolizes $\Xi_U \in \{\alpha_{su}, \alpha_{sd}, \alpha_R, \alpha_{su}^{R1}, \alpha_{su}^{R2}, \alpha_{sd}^{R1}, \alpha_{sd}^{R2}, \beta_{su}, \beta_{sd}, \beta_R, \beta_p, \beta_p^{R1}, \beta_p^{R2}\}$, $\tilde{f}_1[l]$, $\tilde{f}_2[l]$, $\tilde{f}_3[l]$, and $\tilde{f}_{p,m}[l]$ correspond to (56a), (62), (56a), and (69), respectively. Evidently, $\bar{\mathcal{P}}_3$ is a solvable convex problem that can be solved by standard optimization tools.

$$\begin{aligned} \mathcal{F}(\mathbf{a}_{su}, \mathbf{a}_{sd}) &= \sum_l \left(\eta_{su}[l] \ln(1 + \mathbf{a}_{su}^T[l] \tilde{\gamma}_{su}[l]) + \eta_{sd}[l] \mathbf{a}_{sd}^T[l] \tilde{\mathbf{R}}_{sd}[l] \right) - \xi^* \sum_l \mathbf{a}_{sd}^H[l] \left(\tilde{\mathbf{P}}_{R,sd}[l] + \mathbf{p}_{sd}[l] \right), \\ \tilde{\gamma}_{su}[l] &= [\gamma_{su,1}[l], \dots, \gamma_{su,K}[l]]^T, \gamma_{su,k}[l] = p_{su,k}[l] |\bar{h}_{su,k}[l]|^2 / (I_{R,U}^{\text{int}}[l] + \sigma_U^2), \tilde{\mathbf{R}}_{sd}[l] = [\tilde{R}_{sd,1}[l], \dots, \tilde{R}_{sd,J}[l]]^T, \\ \tilde{R}_{sd,1}[l] &= \ln \left(\frac{p_{sd,j}[l] |\bar{h}_{sd,j}[l]|^2}{I_{j,su}^{\text{int}}[l] + I_{R,U}^{\text{int}}[l] + \sigma_{sd,j}^2} \right), \tilde{\mathbf{P}}_{R,sd}[l] = [\tilde{P}_{R,1}[l], \dots, \tilde{P}_{R,J}[l]]^T, \tilde{P}_{R,j}[l] = p_{sd,j}[l] \|\Theta[l] \mathbf{g}_{r,u}[l]\|^2, \\ P_{R,su} &= \sum_k p_{su,k}[l] \|\Theta[l] \mathbf{f}_{su,k}[l]\|^2 + \sigma_R^2 \|\Theta[l]\|_F^2, \mathbf{E}_{sd,m}[l] = [E_{1,m}[l], \dots, E_{J,m}[l]]^T, E_{j,m} = p_{sd,j}[l] |\bar{h}_{p,m}^U[l]|^2, \\ E_{su,m}^R &= \sigma_R^2 \|\mathbf{f}_{p,m}^H[l] \Theta[l]\|^2 + \sum_k p_{su,k}[l] |\bar{h}_{k,m}^{SP}[l]|^2, \mathbf{a}_{L,k} = [a_k[1], \dots, a_k[L]]^T, \mathbf{a}_{L,j} = [a_j[1], \dots, a_j[L]]^T. \end{aligned} \quad (74)$$

G. Matching Factor Optimization ($\mathbf{a}_{su}, \mathbf{a}_{sd}$)

Fixing the other variables and optimizing $(\mathbf{a}_{su}, \mathbf{a}_{sd})$, $\mathcal{P}^o$ can be rewritten as

$$\max_{\mathbf{a}_{su}, \mathbf{a}_{sd}} \mathcal{F}(\mathbf{a}_{su}, \mathbf{a}_{sd}) \quad (73a)$$

$$\text{s.t. } C2: \mathbf{1}_K^T \mathbf{a}_{su}[l] \leq K_s, [\mathbf{a}_{su}[l]]_k^2 = [\mathbf{a}_{su}[l]]_k, \forall k, \forall l, \quad (73b)$$

$$C3: \mathbf{1}_J^T \mathbf{a}_{sd}[l] \leq 1, [\mathbf{a}_{sd}[l]]_j^2 = [\mathbf{a}_{sd}[l]]_j, \forall j, \forall l, \quad (73c)$$

$$C4: \mathbf{a}_{sd}^H[l] \mathbf{E}_{sd,m}[l] + E_{su,m}^R \leq P_{tol,m}, \forall m, \forall l, \quad (73d)$$

$$C8: P_{R,su}[l] + \mathbf{a}_{sd}^H[l] \tilde{\mathbf{P}}_{R,sd}[l] \leq P_R, \forall l, \quad (73e)$$

$$C9: \mathbf{a}_{su}^H[l] \tilde{\gamma}_{su}[l] \leq \varepsilon_U[l], \forall l, \quad (73f)$$

$$C10: \mathbf{1}_L^T \mathbf{a}_{L,k} \geq L_{su}, \forall k, \quad (73g)$$

$$C11: \mathbf{1}_L^T \mathbf{a}_{L,j} \geq L_{sd}, \forall j, C12, C13, \quad (73h)$$

where the meaning of all symbols is explained by (74).

Indeed, (73) is a MINLP problem for which the solution is NP-hard [39]. Here, we employ the LCR technique to relax (73) and find a suboptimal solution instead [40].

Proposition 5: Applying the LCR technique and introducing the symmetric auxiliary matrices $\mathbf{A}_{su}[l] \in \mathbb{R}^{K \times K}$ and $\mathbf{A}_{sd}[l] \in \mathbb{R}^{J \times J}$, $\forall l \in \mathcal{L}$, (73) can be reformulated as

$$\bar{\mathcal{P}}_4: \max_{\mathbf{a}_{su}, \mathbf{a}_{sd}, \mathbf{A}_{su}, \mathbf{A}_{sd}} \mathcal{F}(\mathbf{a}_{su}, \mathbf{a}_{sd}) \quad (75a)$$

$$\text{s.t. } C4, C8 - C11, \quad (75b)$$

$$\bar{C}2: \mathbf{1}_K^T \mathbf{a}_{su}[l] \leq K_s, [\mathbf{a}_{su}[l]]_k = [\mathbf{A}_{su}[l]]_{k,k},$$

$$\begin{bmatrix} 1 & \mathbf{a}_{su}^T[l] \\ \mathbf{a}_{su}[l] & \mathbf{A}_{su}[l] \end{bmatrix} \succeq 0, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (75c)$$

$$\bar{C}3: \mathbf{1}_J^T \mathbf{a}_{sd}[l] \leq 1, [\mathbf{a}_{sd}[l]]_j = [\mathbf{A}_{sd}[l]]_{j,j},$$

$$\begin{bmatrix} 1 & \mathbf{a}_{sd}^T[l] \\ \mathbf{a}_{sd}[l] & \mathbf{A}_{sd}[l] \end{bmatrix} \succeq 0, \forall j \in \mathcal{J}, \forall l \in \mathcal{L}, \quad (75d)$$

$$\hat{C}12: \ln(1 + \mathbf{a}_{su}^T[l] \tilde{\gamma}_{su}[l]) \geq R_{su}^{\min}, \forall l \in \mathcal{L}, \quad (75e)$$

$$\hat{C}13: \mathbf{a}_{sd}^T[l] \tilde{\mathbf{R}}_{sd}[l] \geq R_{sd}^{\min}, \forall l \in \mathcal{L}. \quad (75f)$$

Notice that, (75) is a solvable semidefinite programming (SDP) problem, and the optimal solution is easily obtained by the SDP solver. To map $\mathbf{a}_{su}$ and $\mathbf{a}_{sd}$ from the continuous domain to the discrete domain, the first K largest terms of $\mathbf{a}_{su}[l]$ and the largest term of $\mathbf{a}_{sd}[l]$ can be set to 1, and the rest to 0, thereby completing the solution.

To summarize, suitable initial points are defined, $\mathbf{a}_{su}$, $\mathbf{a}_{sd}$, $\mathbf{p}_{su}$, $\mathbf{p}_{sd}$, $\boldsymbol{\theta}$, and $\mathbf{q}$ are optimized by alternating until η_{EE} converges, as outlined in **Algorithm 1**.

IV. PROPOSED ENERGY-EFFICIENT FRAMEWORK

In this section, we delve into energy-efficient RIS to align with the green and energy-efficient future network development theme.

A. Overview of Energy-Efficient RIS Framework

Typically, the phase shifters and amplifiers in an active RIS are paired, allowing independent tuning of phase shift and amplification for each meta-atom (fully-connected architecture), as shown in Fig. 2(a). Nonetheless, oriented to the

Algorithm 1: Proposed Joint Optimization Framework.

Initialize: Set initial points $\{\mathbf{a}_{su}^{[0]}, \mathbf{a}_{sd}^{[0]}, \mathbf{p}_{su}^{[0]}, \mathbf{p}_{sd}^{[0]}, \boldsymbol{\theta}^{[0]}, \mathbf{q}^{[0]}\}$ and iteration index $t = 0$.

1: **while** η_{EE} is not convergent **do**

2: $t = t + 1$;

3: Update $\xi^{[t]}$ by solving (26);

4: Update $\{\rho_{su}^{[t]}, \rho_{sd}^{[t]}\}$ by solving (31);

5: Update $\{\zeta_{su}^{[t]}, \zeta_{sd}^{[t]}\}$ by solving (35);

6: Update $\{\mathbf{p}_{su}^{[t]}, \mathbf{p}_{sd}^{[t]}\}$ by solving (36);

7: Update $\boldsymbol{\theta}^{[t]}$ by solving (45);

8: Update $\mathbf{q}^{[t]}$ by solving (72);

9: Update $\{\mathbf{a}_{su}^{[t]}, \mathbf{a}_{sd}^{[t]}\}$ by solving (75);

10: **end while**

Output: WSMR, WSMEE, and UAV trajectory $\mathbf{q}$.

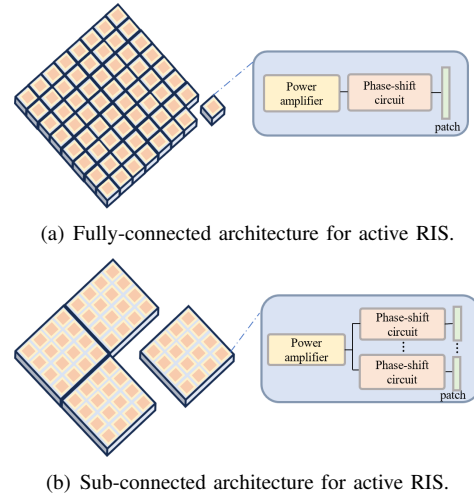


Fig. 2. Comparison of different architectures for active RIS.

trend of ultra-large-scale antenna arrays in 6G communications, thereby presenting a substantial challenge in terms of economic feasibility and power consumption.

To this end, we introduce an active RIS featuring a sub-connected architecture [41], as illustrated in Fig. 2(b). Notably, the sub-arrays within this architecture are modulated by a single amplifier, with multiple meta-atoms equipped with distinct phase shifters. This configuration enables independent phase shifts while sharing the amplitude, significantly reducing the number of amplifiers required per sub-array. This approach not only lowers power consumption but also facilitates the design and deployment of the hardware, making it a promising solution for future large-scale antenna arrays in 6G communication systems.

Denoting the number of amplifiers required by $B = N_t/N_p$, the RC vector of the sub-connected active RIS can be symbolized as $\boldsymbol{\theta} = \boldsymbol{\Theta}\boldsymbol{\Gamma}\mathbf{b} = \text{diag}(\boldsymbol{\Gamma}\mathbf{b})\boldsymbol{\theta}$, where $\boldsymbol{\Theta} \in \mathbb{C}^{N_t \times N_t}$ and $\boldsymbol{\theta} \in \mathbb{C}^{N_t}$ denote the phase shift matrix and phase shift vector of the phase shifter, respectively, $\boldsymbol{\Gamma} \in \mathbb{R}^{N_t \times B}$ is the matching matrix between the amplifier and the phase shifter denoted as $\boldsymbol{\Gamma} = \mathbf{I}_B \otimes \mathbf{1}_{N_p}$, and $\mathbf{b} \in \mathbb{R}^B$ with $\mathbf{b} = [b_1, \dots, b_B]^T$ represents the amplification factor vector. As a trade-off, the sub-connected architecture sacrifices some flexibility in beam

amplitude selection. However, as demonstrated in simulations, this impact is minimal.

Since sub-connected architectures are distinguished from fully connected architectures, the optimization of the RC is different. Consider a basic SISO system, where direct links and NLoS paths are ignored, the optimization results for the RC are summarized in *Lemma 2*.

Lemma 2: In the maximum SNR problem for a sub-connected RIS-aided SISO system, the optimal solution satisfies $b_1 = \dots = b_B$ when it holds.

Proof: Refer the Appendix A [42]. ■

Remark 3: The aforementioned system can directly optimize the RC θ before recovering $\bar{\theta}$ and $\mathbf{b}$. However, in our proposed complex EE optimization system, the amplitude of the RC cannot be guaranteed to be consistent in view of meeting the demands of multiple devices. Consequently, direct optimization of the RC cannot guarantee the same amplitude of the meta-atoms corresponding to the same amplifier. Accordingly, we will discuss the amplification factor and phase shift, respectively.

B. RIS Amplification Factor Optimization

Firstly, fixing the other variables to optimize $\mathbf{b}$, $\bar{\mathcal{P}}_2$ can be reformulated as

$$\mathcal{P}_5 : \min_{\mathbf{b}} \sum_l \left(\mathbf{b}^H[l] \tilde{\mathbf{A}}[l] \mathbf{b}[l] - 2\Re \{ \mathbf{b}^H[l] \tilde{\mathbf{v}}[l] \} \right) \quad (76a)$$

$$\text{s.t. } \bar{C}4\{\tilde{\mathbf{A}}_m, \tilde{\mathbf{v}}_m\}, \bar{C}8\{\tilde{\mathbf{A}}_R, \tilde{\mathbf{v}}_U, \tilde{\mathbf{G}}_{r,u}\}, \quad (76b)$$

$$\bar{C}12\{\tilde{\mathbf{A}}_{su}, \tilde{\mathbf{v}}_{su}\}, \bar{C}13\{\tilde{\mathbf{A}}_{sd}, \tilde{\mathbf{v}}_{sd}\}, \quad (76c)$$

$$\bar{C}7 : [\mathbf{b}[l]]_i \leq \bar{\omega}_{i,\max}, \forall i \in \mathcal{B}, \forall l \in \mathcal{L},$$

where

$$\begin{aligned} \tilde{\mathbf{A}}[l] &= \mathbf{\Gamma}^H \bar{\mathbf{\Theta}}^H[l] \mathbf{\Lambda}[l] \bar{\mathbf{\Theta}}[l] \mathbf{\Gamma}, \tilde{\mathbf{v}}[l] = \mathbf{\Gamma}^H \bar{\mathbf{\Theta}}^H[l] \mathbf{v}[l], \\ \tilde{\mathbf{G}}_{r,u}[l] &= \mathbf{\Gamma}^H \bar{\mathbf{\Theta}}^H[l] \mathbf{G}_{r,u}[l] \bar{\mathbf{\Theta}}[l] \mathbf{\Gamma}, \\ \tilde{\mathbf{A}}_i[l] &= \mathbf{\Gamma}^H \bar{\mathbf{\Theta}}^H[l] \mathbf{\Lambda}_i[l] \bar{\mathbf{\Theta}}[l] \mathbf{\Gamma}, i \in \{m, R, U, su, sd\}, \\ \tilde{\mathbf{v}}_i[l] &= \mathbf{\Gamma}^H \bar{\mathbf{\Theta}}^H[l] \mathbf{v}_i[l], i \in \{m, U, su, sd\}. \end{aligned} \quad (77)$$

Obviously, $\mathcal{P}_5$ is an intractable standard convex problem.

C. RIS Phase Shift Optimization

Due to $\|\mathbf{\Theta}\|_F^2 = \|\text{diag}(\mathbf{\Gamma}\mathbf{b}) \bar{\mathbf{\Theta}}\|_F^2 = \|\text{diag}(\mathbf{\Gamma}\mathbf{b})\|_F^2$ and $\|\mathbf{g}_{r,u}^H \text{diag}(\mathbf{\Gamma}\mathbf{b}) \bar{\mathbf{\Theta}}\|^2 = \|\mathbf{g}_{r,u}^H \text{diag}(\mathbf{\Gamma}\mathbf{b})\|^2$, it can be found that the power consumption of the RIS and the interference caused by RIS thermal noise in communication are independent of the RIS phase shift. Then, by fixing other variables to optimize the RIS phase shift $\bar{\theta}$, $\bar{\mathcal{P}}_2$ can be reformulated as

$$\mathcal{P}_6 : \min_{\bar{\theta}} \sum_l \left(\bar{\theta}^H[l] \tilde{\mathbf{A}}[l] \bar{\theta}[l] - 2\Re \{ \bar{\theta}^H[l] \tilde{\mathbf{v}}[l] \} \right) \quad (78a)$$

$$\text{s.t. } \bar{C}4\{\tilde{\mathbf{A}}_m, \tilde{\mathbf{v}}_m\}, \bar{C}9\{\tilde{\mathbf{A}}_U, \tilde{\mathbf{v}}_U, \tilde{\mathbf{G}}_{r,u}\}, \quad (78b)$$

$$\bar{C}12\{\tilde{\mathbf{A}}_{su}, \tilde{\mathbf{v}}_{su}\}, \bar{C}13\{\tilde{\mathbf{A}}_{sd}, \tilde{\mathbf{v}}_{sd}\}, \quad (78c)$$

$$\bar{C}7 : [\bar{\theta}[l]]_n \leq 1, \forall n \in \mathcal{N}, \forall l \in \mathcal{L},$$

Algorithm 2: Proposed Energy-Efficient Framework.

Initialize: Initial $\{\mathbf{a}_{su}^{[0]}, \mathbf{a}_{sd}^{[0]}, \mathbf{p}_{su}^{[0]}, \mathbf{p}_{sd}^{[0]}, \mathbf{b}^{[0]}, \bar{\mathbf{\theta}}^{[0]}, \mathbf{q}^{[0]}\}$ and iteration index $t = 0$.

1: **while** η_{EE} is not convergent **do**

2: Steps 2-6 and 8-9 of Algorithm 1;

3: Update $\mathbf{b}^{[t]}$ by solving (76);

4: Update $\bar{\mathbf{\theta}}^{[t]}$ by solving (78);

5: **end while**

Output: WSMR, WSMEE, and UAV trajectory $\mathbf{q}$.

where

$$\begin{aligned} \bar{\mathbf{A}}[l] &= \text{diag}(\mathbf{\Gamma}\mathbf{b}[l])^H \mathbf{\Lambda}[l] \text{diag}(\mathbf{\Gamma}\mathbf{b}[l]), \\ \bar{\mathbf{v}}[l] &= \text{diag}(\mathbf{\Gamma}\mathbf{b}[l])^H \mathbf{v}[l], \\ \bar{\mathbf{G}}_{r,u}[l] &= \text{diag}(\mathbf{\Gamma}\mathbf{b}[l])^H \mathbf{G}_{r,u}[l] \text{diag}(\mathbf{\Gamma}\mathbf{b}[l]), \\ \bar{\mathbf{A}}_i[l] &= \text{diag}(\mathbf{\Gamma}\mathbf{b}[l])^H \mathbf{\Lambda}_i[l] \text{diag}(\mathbf{\Gamma}\mathbf{b}[l]), \\ \bar{\mathbf{v}}_i[l] &= \text{diag}(\mathbf{\Gamma}\mathbf{b}[l])^H \mathbf{v}_i[l], i \in \{m, U, su, sd\}. \end{aligned} \quad (79)$$

Notably, we relax the unit-modulus constraint on the phase shift for convenience, denoted as $\bar{C}7$. Obviously, $\mathcal{P}_6$ is a solvable QCQP problem. Upon obtaining the optimal $\bar{\mathbf{\theta}}^{\text{opt}}$, the phase shift $\bar{\theta}^*$ with unit-modulus is obtained by normalization. In fact, the modulus of $\bar{\mathbf{\theta}}^{\text{opt}}$ can be complemented by optimizing the amplification factor to obtain the optimal amplitude of the RC. Hence, the relaxation of the unit-modulus constraint for the phase shift is justified here.

In summary, the optimization framework for the sub-connected RIS architecture is outlined in Algorithm 2.

V. FRAMEWORK SUPPLEMENTS

In this section, we provide further elaboration on the proposed joint optimization framework. Specifically, this includes the expansion of the UAV NFZ and the convergence and complexity analysis of the proposed algorithms.

A. Additional No-Fly Zone

For reasons such as security and interference suppression, we introduce a NFZ within the primary network coverage area. To facilitate this, we assume that the NFZ is a circle centered at (x_N, y_N) with radius r_N , i.e., $D_N = \{(x, y) | (x - x_N)^2 + (y - y_N)^2 < r_N^2\}$. Then, the UAV flight trajectory is constrained as

$$(x_U[l] - x_N)^2 + (y_U[l] - y_N)^2 \geq r_N^2, \forall l \in \mathcal{L}. \quad (80)$$

By Taylor expansion based on fixed points $x'_U[l]$ and $y'_U[l]$ of $x_U[l]$ and $y_U[l]$, (80) can be linearized as

$$\begin{aligned} (x'_U[l] - x_N)(2x_U[l] - x'_U[l] - x_N) + \\ (y'_U[l] - y_N)(2y_U[l] - y'_U[l] - y_N) \geq r_N^2, \forall l \in \mathcal{L}. \end{aligned} \quad (81)$$

Apparently, (81) is only related to $\mathbf{q}$, and thus we complete the restriction by simply bringing the constraint (81) to the subproblem of UAV trajectory optimization.

B. Algorithm Convergency

In Algorithm 1, the iterations (26), (31), (35), (36), (45) and (75) are easily proved to be monotone non-decreasing, and the approximation error of the array response $\mathbf{e}_{r,u}$ introduced into $\mathbf{q}$ optimization in (72) has a very limited effect on the result. Similarly, for Algorithm 2, the iterations (76) and (78) are monotonic and non-decreasing. Notably, since the flight energy consumption directly affects the EE, the change of $\mathbf{q}$ is more stringent, and the error introduced by the $\mathbf{e}_{r,u}$ approximation is more limited. Consequently, both algorithms converge to feasible solutions, as will be verified in the simulations in the next section.

C. Algorithmic Complexity

In this subsection, we provide the computational complexity analysis of the two proposed algorithms. Firstly, the computational complexity of closed-form ξ , $\{\rho_{su}, \rho_{sd}\}$, and $\{\zeta_{su}, \zeta_{sd}\}$ can be approximated as $\mathcal{O}(L)$, $\mathcal{O}(L(K+J))$, and $\mathcal{O}(L(K^2+J))$, respectively. The computational complexity of subproblems (36), (45), (72), and (75) can be approximated as $\mathcal{O}((LK)^{3.5} + (LJ)^{3.5})$, $\mathcal{O}((LN_t)^{3.5})$, $\mathcal{O}((L(4K+4J+3M+5))^{3.5})$, and $\mathcal{O}((LK)^2(LK+1)^{2.5} + (LJ)^2(LJ+1)^{2.5})$, respectively. Therefore, the overall computational complexity of **Algorithm 1** is $\mathcal{O}\{I_R\{L + (LK)^{3.5} + (LJ)^{3.5} + (LN_t)^{3.5} + (L(4K+4J+3M+5))^{3.5} + (LK)^2(LK+1)^{2.5} + (LJ)^2(LJ+1)^{2.5} + L(K+J) + L(K^2+J)\}\}$, where I_R symbolizes the number of iterations required for convergence. Similarly, the computational complexity of subproblems (77) and (79) are approximated as $\mathcal{O}((LB)^{3.5})$ and $\mathcal{O}((LN_t)^{3.5})$, respectively, the overall computational complexity of **Algorithm 2** can be approximated as $\mathcal{O}\{I_E\{L + L(K+J) + L(K^2+J) + (LK)^{3.5} + (LJ)^{3.5} + (LB)^{3.5} + (LN_t)^{3.5} + (LK)^2(LK+1)^{2.5} + (LJ)^2(LJ+1)^{2.5} + (L(4K+4J+3M+5))^{3.5}\}\}$, where I_E symbolizes the number of iterations required for convergence.

VI. SIMULATION RESULTS

In this section, we provide extensive simulation results to evaluate the performance of the proposed schemes.

A. Simulation Setup

Without loss of generality, we consider a 3D simulation scenario, where PUs are randomly distributed within a circle area centered at the origin with radius $r_N = 80$ m, SUs are randomly distributed between the circles with radii r_N and $r_O = 200$ m, and the active RIS is deployed at $(0, 200, 10)$ m.⁵ Assuming that the UAV charging control station is positioned at $(100, 0, 100)$ m, the UAV departs from this location and completes a cycle returning for maintenance and charging. The number of PUs, SUUs, and SDUs are set to $M = 2$, $K = 3$, and $J = 2$, respectively, and the maximum number of SUUs accessing the UAV is $K_s = 2$. We set the carrier frequency as

⁵Notably, the division between PUs and SUs relies on spectrum prioritization and transmission source characteristics, rather than geographical location. Actually, all users can be distributed arbitrarily.

TABLE II
KEY SYSTEM PARAMETERS

Parameters	Values
Path loss exponent	$\tau_{r,su} = \tau_{r,sd} = \tau_{r,p} = 2.3$, $\tau_{k,m}^{SP} = \tau_{k,j}^{UD} = 3.5$
Path loss at 1 meter	$L_0 = -44.48$ dB
Transmit power	$p_S = 0.5$ W, $p_{sd,\max} = 0.1$ W, $p_{su,\max} = p_R = 0.01$ W
Noise power	$\sigma_S^2 = -110$ dBm, $\sigma_{sd,j}^2 = -80$ dBm, $\sigma_R^2 = \sigma_U^2 = -90$ dBm,
Transmit bandwidth	$B_S = 30$ MHz, $B_{su} = B_{sd} = 10$ MHz
Dissipated power	$P_{su}^c = P_{sd}^c = P_{ps}^r = 0$ dBm, $P_{pa}^r = 3$ dBm
Minimum QoS	$R_{su}^{\min} = 3$ bps/Hz, $R_{sd}^{\min} = 2$ bps/Hz
Main parameters of the UAV [30]	$v_0 = 4.03$, $\Omega = 300$, $R_a = 0.4$, $c_0 = 79.8563$, $c_1 = 88.6279$, $c_2 = 0.0092$
Maximum horizontal speed of the UAV	$v_{\max}^{xy} = 25$ m/s
Antenna gain between UAV and satellite	$G_u^t G_s^r = 30$ dB
Rain attenuation	$\psi_{ra} = -2.3$ dB
Number and duration of time slots	$L = 40$, $\Delta t = 1$ s
Number of accesses to UAV during a cycle	$L_{su} = L_{sd} = L/3$
Maximum interference tolerance	$P_{tol,m} = -80$ dBW

$f_c = 4$ GHz, the number of meta-atoms and the amplification factor for the RIS as $N_t = 20$ and $\omega_{n,\max} = 20$ dB, the Rician factor $\alpha^* = 10$ dB, the inter-antenna separation distance $d_f = \lambda_w/2$, the number of amplifiers $B = N_t/10$, and the weights as $\eta_{su}[l] = \eta_{sd}[l] = 1$ and $\zeta_U = 1$.⁶ Other parameters are shown in TABLE II. The following parameters remain invariant unless otherwise noted.

For clarity, we initially define the following legend:

- **Proposed WSMR**: Maximizing WSMR in secondary networks is based on Algorithm 1.
- **Proposed WSMEE**: Maximizing WSMEE in secondary networks based on Algorithm 2.
- **Circular trajectory**: A fixed circular flight trajectory centered at the origin with a radius of 100 maximizes WSMEE based on Algorithm 2.
- **Fixed position**: The UAV hovers at the initial point with the aim of maximizing WSMEE.
- **Without RIS**: Maximizing WSMR without RIS.
- **Passive RIS**: Passive RIS-aided scheme aimed at maximizing WSMR.
- **OMA**: Only one SUU for each slot is allowed to access the UAV aiming to maximize the WSMR.
- **NFZ**: There is a NFZ constraint designed to maximize the WSMR scheme.

In Fig. 3, we show the WSMR versus the number of iterations for the proposed schemes and the baseline schemes. Apparently, all curves converge to stability rapidly, which proves the feasibility of the proposed model and schemes. Especially, the updating strategies of the ‘Proposed WSMR’ and ‘Proposed WSMEE’ schemes do not cause significant fluctuations, which verifies the effectiveness of the proposed

⁶Here, we have not chosen to optimize the weighted factors. This is because its easy to cause $\{n_{su}[l] = 1, n_{sd}[l] = 0\} / \{n_{su}[l] = 0, n_{sd}[l] = 1\}$, which compromises the fairness and universality of the service.

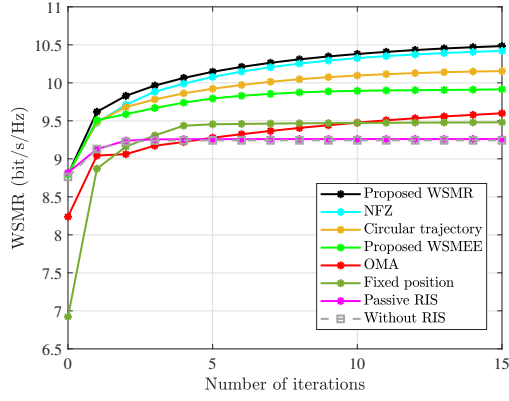


Fig. 3. WSMR versus number of iterations.

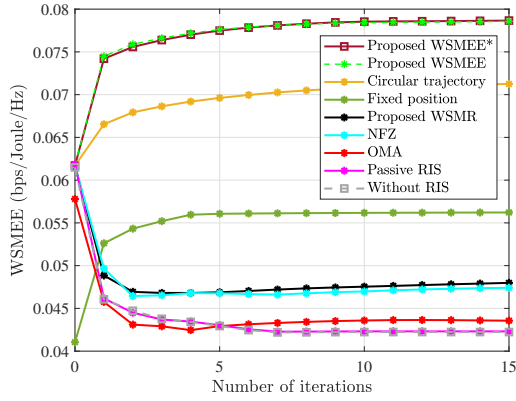
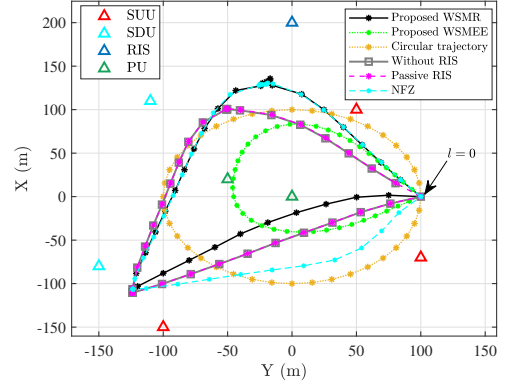


Fig. 4. WSMEE versus number of iterations.

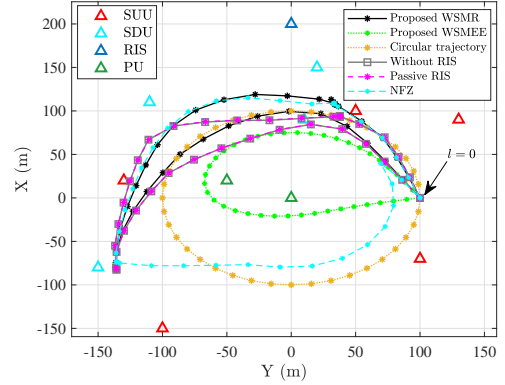
joint optimization design. Moreover, the performance of the ‘Without RIS’ scheme closely approaches that of the passive RIS scheme. This proves that the performance gain of passive RIS is weak by multiplicative fading in long-range scenarios, while active RIS can overcome this drawback. Besides, the scheme to avoid the NFZ has little impact on WSMR.

Fig. 4 illustrates the relationship between the number of iterations for WSMEE comparison under different schemes, where ‘Proposed WSMEE*’ refers to the fully-connected active RIS scheme operating at the same dissipated power. Intuitively, curves aiming to maximize WSMEE exhibit faster convergence, while other schemes display minor fluctuations, which is expected. Notably, the superior performance of the ‘Proposed WSMEE’ scheme compared to the ‘Circular trajectory’ and ‘Fixed position’ schemes, which confirms the effectiveness of the designed trajectory optimization. Schemes focusing on maximizing WSMR generally show lower WSMEE due to increased energy consumption. Furthermore, the sub-connected architecture achieves performance comparable to the fully-connected architecture, which verifies that sacrificing beam flexibility has minimal impact on performance.

Fig. 5 shows the horizontal flight trajectory of the UAV under different schemes with varying number of SUs. We observe in Fig. 5(a) that the UAV trajectory of the ‘proposed WSMR’ scheme generally favors the RIS side and



(a) $K = 3, J = 2$.

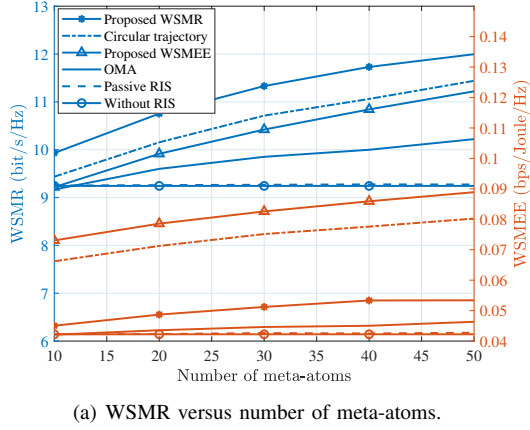


(b) $K = 5, J = 4$.

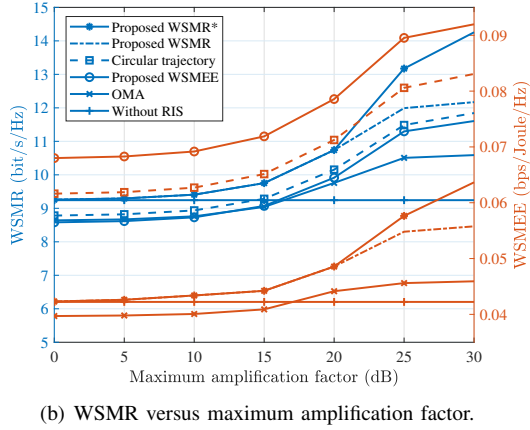
Fig. 5. Projections of optimized trajectories for the UAV.

lingers longer. This phenomenon becomes more pronounced with an increasing number of SUs in the upper half of the region, as shown in Fig. 5(b), which indicates that proximity to the RIS enhances performance. The ‘Proposed WSMEE’ scheme trajectory also moves closer to the RIS side, appearing smoother and closer to the initial point. This is because traversing longer distances at higher speeds consumes more energy, thereby reducing WSMEE. The curves for the ‘Without RIS’ and ‘Passive RIS’ schemes largely overlap, demonstrating that the effect of multiplicative fading causes passive RIS to have minimal impact on performance. Additionally, the ‘NFZ’ scheme intentionally avoids the primary network region, which proves the effectiveness of the design.

To illustrate the impact of active RIS parameters on system performance, Fig. 6(a) shows the WSMR and WSMEE versus the number of meta-atoms under different schemes. Note that the legend style is only shape-dependent and not color-dependent. Obviously, the WSMR and WSMEE of the schemes with RIS assistance increase with the number of meta-atoms, as expected. Moreover, the WSMR gap between the ‘Proposed WSMR’ and ‘OMA’ schemes widens as the number of meta-atoms increases, attributed to the improved suppression of inter-user interference with more meta-atoms, weakening the ‘OMA’ scheme’s advantage of free inter-user interference. Importantly, for WSMR and WSMEE metrics, the schemes’ performance ordering is different, which explains



(a) WSMR versus number of meta-atoms.



(b) WSMR versus maximum amplification factor.

Fig. 6. WSMR versus active RIS parameters.

that higher WSMR does not necessarily correspond to higher WSMEE.

Fig. 6(b) presents the WSMR and WSMEE versus the maximum amplification factor, with $B_S = 50$ MHz and ‘Proposed WSMR*’ featuring $p_R = 0.1$ W. As the maximum amplification factor increases, the beamforming degrees of freedom of schemes with RIS assistance improve, consequently boosting their WSMR and WSMEE. It is noteworthy that increasing the maximum transmit power of the RIS is not associated with the performance of the ‘Proposed WSMR’ scheme when the amplification factor is less than 20 dB. However, the performance improvement of the ‘Proposed WSMR’ scheme is limited when the amplification factor exceeds 20 dB, while the ‘Proposed WSMR*’ scheme exhibits significant enhancement. This can be attributed to the fact that the optimal RC is slack for the RIS maximum power at low amplification factors, while it becomes constrained by the active RIS’s maximum supply at amplification factors above a certain threshold. Besides, the ‘Proposed WSMEE’ scheme significantly outperforms others in terms of the WSMEE metric.

To illustrate the relationship between the maximum interference tolerance of PUs and WSMR, we plot Fig. 7. Intuitively, the WSMR of all schemes increases as the maximum interference tolerance, stabilizing around the threshold of -90 dBW. This phenomenon arises because higher interference tolerance enhances the flexibility in adjusting transmit power,

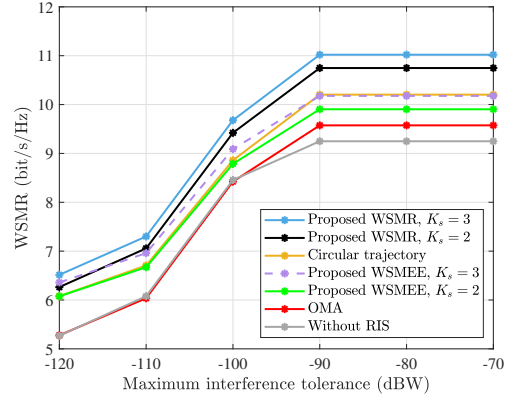
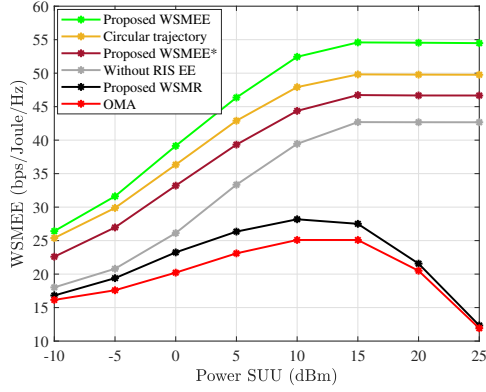


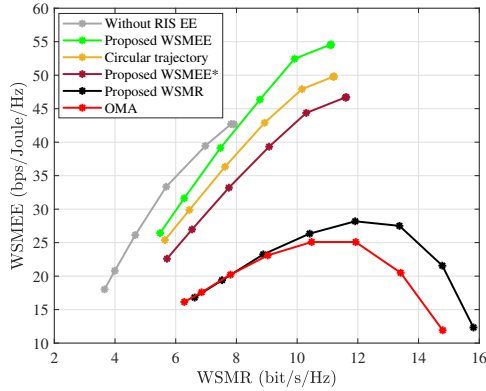
Fig. 7. WSMR versus PU maximum interference tolerance.

RC, and UAV position, thereby improving performance in the secondary network. Moreover, the difference between the proposed scheme and the ‘Without RIS’ scheme validates that RIS can effectively suppress the interference of secondary networks to PUs. Notably, the proposed schemes achieve similar performance at load finite $K_s = 2$ as at infinite $K_s = 3$. This indicates that some users with better QoS can be selected instead of the full matching scheme in practical applications, which reduces load, signaling consumption, and decoding complexity in SIC, while maintaining satisfactory performance.

In Fig. 8(a), we set $\varsigma_U = -30$ dB to scale down the UAV energy consumption weights and plot the WSMEE versus maximum transmit power of SUU with $\omega_{n,\max} = 25$ dB and $B_S = 50$ MHz, where ‘Without RIS EE’ denotes the without RIS scheme aiming to maximize WSMEE. Due to the challenges of meeting minimum rate requirements at lower power levels, it is ignored to emphasise the upper bound of WSMEE. Obviously, the WSMEE of the schemes that initially aim to maximize the WSMEE increases with the SUU’s maximum transmit power and then stabilizes. This trend arises because higher transmit power enhances spectral efficiency, thereby boosting WSMEE. However, beyond a critical threshold, further increases in transmit power lead to diminishing returns as additional power consumption outweighs the spectral efficiency gains. Moreover, the proposed scheme improves WSMEE despite the additional power dissipation by the RIS circuit compared to the ‘Without RIS EE’ scheme. Notably, the performance of ‘Proposed WSMEE’ significantly improves compared to the fully-connected ‘Proposed WSMEE*’. To further illustrate the relationship between WSMR and WSMEE with increasing SUU’s transmit power, Fig. 8(b) is presented. Noticeably, schemes optimizing for WSMEE demonstrate an initial increase in both WSMEE and WSMR with increasing SUU transmit power, followed by stabilization. In contrast, schemes aiming at WSMR initially increase WSMEE but subsequently observe a decline, while WSMR keeps increasing. This behavior highlights the trade-off where WSMR-centric schemes increase power consumption to boost WSMR, but at higher power levels, the incremental gain in WSMR diminishes relative to the increase in energy consumption (e.g.,



(a) WSMEE versus SUU maximum transmit power.



(b) WSMEE versus WSMR.

Fig. 8. System performance versus SUU maximum transmit power.

beyond 10 dBm). Thus, careful consideration is essential in designing the maximum transmit power, balancing between maximizing performance metrics and managing energy consumption effectively.

VII. CONCLUSION

In this work, an active RIS-aided NOMA-enabled SAGIN with CR was investigated. We proposed a joint optimization framework for power allocation, user matching, RC, and trajectory design to maximize WSMR and WSMEE. To further enhance EE, we introduced an optimization framework for active RIS with sub-connected architecture. Simulation results affirmed the effectiveness of the proposed schemes, while the optimization of UAV trajectories and the adoption of sub-connected active RIS achieve better performance compared to conventional schemes.

APPENDIX A PROOF OF LEMMA 2

For an active RIS-aided SISO system, assuming that the transmitter-to-RIS and RIS-to-receiver channels are $\mathbf{g} \in \mathbb{C}^{N_t}$

and $\mathbf{f} \in \mathbb{C}^{N_t}$, respectively, the SNR maximization problem is formulated as

$$\max_{p_{sd}, \bar{\theta}, \mathbf{b}} \eta_R = \frac{p_{sd} |\mathbf{f}^H \boldsymbol{\Theta}^H \mathbf{g}|^2}{\sigma_R^2 \|\mathbf{f}^H \boldsymbol{\Theta}^H\|^2 + \sigma_{sd}^2} \quad (\text{A.1a})$$

$$\text{s.t. } p_{sd} \leq p_{sd, \max}, \quad (\text{A.1b})$$

$$p_{sd} \|\boldsymbol{\Theta}^H \mathbf{g}\|^2 + \sigma_R^2 \|\boldsymbol{\Theta}^H\|^2 \leq P_R, \quad (\text{A.1c})$$

$$[\bar{\theta}]_n = 1, \forall n \in \mathcal{N}. \quad (\text{A.1d})$$

We assume that $\mathbf{g}$ and $\mathbf{f}$ are LoS links, and satisfy

$$\mathbf{g} = |g| \mathbf{1}_{N_t} \odot \exp(j\angle \mathbf{g}), \quad (\text{A.2a})$$

$$\mathbf{f} = |f| \mathbf{1}_{N_t} \odot \exp(j\angle \mathbf{f}). \quad (\text{A.2b})$$

Then, the optimal transmit power and RC can be obtained by applying the Lagrange multiplier method as

$$p_{sd}^{\text{opt}} = p_{sd, \max}, \|\boldsymbol{\Theta}^{\text{opt}}\| = \sqrt{\frac{P_R}{g^2 p_{sd, \max} + \sigma_{sd}^2}}. \quad (\text{A.3})$$

Next, applying the Cauchy-Schwartz inequality, one obtains

$$|\mathbf{f}^H \boldsymbol{\Theta}^H \mathbf{g}|^2 = \left| \sum_{n=1}^{N_t} f_n^* g_n \theta_n^* \right|^2 \leq \|\mathbf{f}^* \odot \mathbf{g}\|^2 \|\boldsymbol{\Theta}\|^2, \quad (\text{A.4})$$

where equality holds if and only if

$$\bar{\theta}_n = \angle(f_n - g_n), b_1 = \dots = b_B. \quad (\text{A.5})$$

REFERENCES

- [1] G. Wang, S. Zhou, and S. Zhang *et al.*, "SFC-based service provisioning for reconfigurable space-air-ground integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 7, pp. 1478-1489, Jul. 2020.
- [2] F. Tang, C. Wen, and L. Luo *et al.*, "Blockchain-based trusted traffic offloading in space-air-ground integrated networks (SAGIN): A federated reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 12, pp. 3501-3516, Dec. 2022.
- [3] H. Cui *et al.*, "Space-air-ground integrated network (SAGIN) for 6G: Requirements, architecture and challenges," *China Commun.*, vol. 19, no. 2, pp. 90-108, Feb. 2022.
- [4] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714-2741, May 2018.
- [5] N. Kato *et al.*, "Optimizing space-air-ground integrated networks by artificial intelligence," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 140-147, Aug. 2019.
- [6] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surv. Tut.*, vol. 18, no. 2, pp. 1123-1152, 2016.
- [7] N. Cheng *et al.*, "Air-ground integrated mobile edge networks: Architecture, challenges, and opportunities," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 26-32, Aug. 2018.
- [8] N. Yang, P. Li, and D. Guo *et al.*, "Detection of interference constraint violation based on heterogeneous data fusion in satellite-earth integrated networks," *IEEE Access*, vol. 8, pp. 37645-37656, 2020.
- [9] D. Xu, X. Yu, and Y. Sun *et al.*, "Resource allocation for IRS-assisted full-duplex cognitive radio systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7376-7394, Dec. 2020.
- [10] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4962-4975, Aug. 2021.
- [11] R. Allu, O. Taghizadeh, and S. K. Singh *et al.*, "Robust beamformer design in active RIS-assisted multiuser MIMO cognitive radio networks," *IEEE Trans. Cog. Commun. Netw.*, vol. 9, no. 2, pp. 398-413, Apr. 2023.
- [12] J. Sun, J. Li, and W. Hao *et al.*, "Secure transmission design with strong channel correlation for passive/active RIS communications," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1394-1398, Aug. 2023.
- [13] S. Wang *et al.*, "Energy-efficiency optimization for multiple access in NOMA-enabled space-air-ground networks," *IEEE Internet Things J.*, vol. 10, no. 17, pp. 15652-15665, Sep. 2023.

- [14] Y. Wang *et al.*, "Joint resource allocation and UAV trajectory optimization for space-air-ground Internet of remote things networks," *IEEE Syst. J.*, vol. 15, no. 4, pp. 4745-4755, Dec. 2021.
- [15] Z. Hu *et al.*, "Joint resources allocation and 3D trajectory optimization for UAV-enabled space-air-ground integrated networks," *IEEE T. Veh. Technol.*, vol. 72, no. 11, pp. 14214-14229, Nov. 2023.
- [16] N. Yang *et al.*, "Lightweight blockchain-based secure spectrum sharing in space-air-ground-integrated IoT network," *IEEE Internet Things J.*, vol. 10, no. 23, pp. 20511-20527, Dec. 2023.
- [17] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3D beam flattening," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4128-4143, Jul. 2021.
- [18] H. Zhang, M. Huang, H. Zhou, X. Wang, N. Wang, and K. Long, "Capacity maximization in RIS-UAV networks: A DDQN-based trajectory and phase shift optimization approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2583-2591, Apr. 2023.
- [19] S. K. Singh *et al.*, "NOMA enhanced hybrid RIS-UAV-assisted full-duplex communication system with imperfect SIC and CSI," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7609-7627, Nov. 2022.
- [20] T. V. Nguyen, H. D. Le, and A. T. Pham, "On the design of RIS-UAV relay-assisted hybrid FSO/RF satellite-aerial-ground integrated network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 757-771, Apr. 2023.
- [21] C. Xu *et al.*, "OTFS-aided RIS-assisted SAGIN systems outperform their OFDM counterparts in doubly selective high-doppler scenarios," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 682-703, Jan. 2023.
- [22] Q. Chen, W. Meng, T. Q. S. Quek, and S. Chen, "Multi-tier hybrid offloading for computation-aware IoT applications in civil aircraft-augmented SAGIN," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 399-417, Feb. 2023.
- [23] M. D. Nguyen, L. B. Le, and A. Girard, "Integrated computation offloading, UAV trajectory control, edge-cloud and radio resource allocation in SAGIN," *IEEE Trans. Cloud Comput.*, early access, Dec. 2023.
- [24] S. Mao, *et al.*, "Joint UAV position optimization and resource scheduling in space-air-ground integrated networks with mixed cloud-edge computing," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3992-4002, Sep. 2021.
- [25] K. An, M. Lin, J. Ouyang, and W.-P. Zhu, "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025-3037, Nov. 2016.
- [26] S. Kandeepan, L. De Nardis, M.-G. Di Benedetto, A. Guidotti, and G. E. Corazza, "Cognitive satellite terrestrial radios," in *Proc. 2010 IEEE Glob. commun. Conf.*, 2010, pp. 1-6.
- [27] F. Zeng *et al.*, "Resource allocation and trajectory optimization for QoE provisioning in energy-efficient UAV-enabled wireless networks," *IEEE T. Veh. Technol.*, vol. 69, no. 7, pp. 7634-7647, Jul. 2020.
- [28] A. Colpaert, E. Vinogradov, and S. Pollin, "Aerial coverage analysis of cellular systems at LTE and mmWave frequencies using 3D city models," *Sensors*, vol. 18, no. 12, pp. 4311-4327, Dec. 2018.
- [29] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569-572, Dec. 2014.
- [30] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329-2345, Apr. 2019.
- [31] X. Mu and Y. Liu, "Exploiting semantic communication for non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2563-2576, Aug. 2023.
- [32] N. Zhao *et al.*, "Joint trajectory and precoding optimization for UAV-assisted NOMA networks," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3723-3735, May 2019.
- [33] M. Zeng *et al.*, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2413-2424, Oct. 2017.
- [34] K. Shen and W. Yu, "Fractional programming for communication systems-part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631-2644, May. 2018.
- [35] K. Shen and W. Yu, "Fractional programming for communication systems-part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616-2630, May. 2018.
- [36] S. Boyd, "Subgradient methods," May 2014. [Online]. Available: https://web.stanford.edu/class/ee364b/lectures/subgrad_method_notes.pdf
- [37] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313-3351, May 2021.
- [38] P. Song, *et al.*, "D3M: Distributed multi-cell multigroup multicasting," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 3741-3745.
- [39] K. G. Murty and S. N. Kabadi, "Some NP-complete problems in quadratic and nonlinear programming," *Math. Program.*, vol. 39, pp. 117-129, Jun. 1987.
- [40] T. Nie, S.-C. Fang, Z. Deng, and J. E. Lavery, "On linear conic relaxation of discrete quadratic programs," *Optim. Methods Softw.*, vol. 31, pp. 737-754, Jan. 2016.
- [41] K. Liu, Z. Zhang, L. Dai, S. Xu, and F. Yang, "Active reconfigurable intelligent surface: Fully-connected or sub-connected?," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 167-171, Jan. 2022.
- [42] Z. Zhang *et al.*, "Active RIS vs. passive RIS: Which will prevail in 6G?," *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707-1725, Mar. 2023.