

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/380326097>

Exploiting Multi-Layer Refracting RIS-Assisted Receiver for HAP-SWIPT Networks

Article in IEEE Transactions on Wireless Communications · May 2024

DOI: 10.1109/TWC.2024.3394214

CITATIONS

12

READS

106

8 authors, including:



Kang An

National University of Defense Technology

190 PUBLICATIONS 5,057 CITATIONS

[SEE PROFILE](#)



Yifu Sun

National University of Defense Technology

25 PUBLICATIONS 309 CITATIONS

[SEE PROFILE](#)



Zhi Lin

National University of Defense Technology

73 PUBLICATIONS 2,696 CITATIONS

[SEE PROFILE](#)



Wanli Ni

Tsinghua University

146 PUBLICATIONS 990 CITATIONS

[SEE PROFILE](#)

Exploiting Multi-Layer Refracting RIS-Assisted Receiver for HAP-SWIPT Networks

Kang An, Yifu Sun, Zhi Lin, Yonggang Zhu, Wanli Ni, *Member, IEEE*,
Naofal Al-Dahir, *Fellow, IEEE*, Kai-Kit Wong, *Fellow, IEEE*, and Dusit Niyato, *Fellow, IEEE*

Abstract—Aiming to circumvent the severe large-scale fading and the energy scarcity dilemma in high-altitude platform (HAP) networks, this paper investigates the benefits of the reconfigurable intelligent surface (RIS) and simultaneous wireless information and power transfer (SWIPT) on HAP communications. Specifically, we propose a concept of multi-layer refracting RIS-assisted receiver to achieve concurrent transmission of the information and energy, which is conducive to overcoming the severe fading effect induced by extreme long-distance HAP links and fully exploits RIS’s degrees-of-freedom (DoFs) for the SWIPT design. Based on the RIS-enhanced receiver, we then formulate a worst-case sum-rate maximization problem by considering the channel state information (CSI) error, the information rate requirements, and the energy harvesting constraint. To handle the intractable non-convex problem, a scalable robust optimization framework is proposed to obtain semi-closed-form solutions. Specifically, a discretization method is adopted to convert the imperfect CSI into a robust one. Then, by utilizing the LogSumExp inequality to smooth the objective and constraints, we develop a dual method to obtain the optimal solution for the HAP transmit precoder. In addition, a modified cyclic coordinate descent (M-CCD) is adopted to update the block-wise RIS coefficients. Moreover, closed-form solutions for power splitting (PS) ratios and the receive decoder are derived. Finally, the asymptotic performance of our proposed RIS-enhanced receiver is provided to reveal the substantial capacity gain for HAP communications. Numerical simulations demonstrate that the proposed architecture and optimization framework are capable of achieving superior performance with low complexity compared to state-of-the-art schemes in HAP networks.

Index Terms—HAP, RIS, SWIPT, imperfect channel state information, robust optimization framework.

This work is supported by the National NSFC under Grant 61901502, 62201592, 62131005; in part by the Research Plan Project of NUDT under Grant ZK21-33; in part by the Young Elite Scientist Sponsorship Program of China Association for Science and Technology (CAST) under Grant 2021-JCJQ-QT-048; in part by the Macau Young Scholars Program under Grant AM2022011; in part by the National Postdoctoral Program for Innovative Talents under Grant BX20200101; and in part by the Graduate Innovation Program of Hunan Province under Grant CX20220008. This paper was presented in part at IEEE Global Communications Conference (GLOBECOM), Kuala Lumpur, Malaysia, in 2023, doi: 10.1109/GLOBECOM54140.2023.10437169 [1]. (*Corresponding authors:* Yifu Sun; Yonggang Zhu.)

K. An, Y. Sun, and Y. Zhu are with Sixty-third Research Institute, National University of Defense Technology, Nanjing 210007, China (Email: ankang89@nudt.edu.cn; sunyifu_nudt@163.com; zhumaka1982@163.com).

Z. Lin is with College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China, and also with the School of Computer Science and Engineering, Macau University of Science and Technology, Macau 999078, China (Email: linzhi945@163.com).

W. Ni is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with the Beijing National Research Center for Information Science and Technology, Beijing 100084, China (e-mail: niwanli@tsinghua.edu.cn).

N. Al-Dahir is with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: aldhahir@utdallas.edu).

K.-K. Wong is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 6BT, U.K. (e-mail: kai-kit.wong@ucl.ac.uk).

D. Niyato is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: dniyato@ntu.edu.sg).

I. INTRODUCTION

HIGH-altitude platform (HAP) has emerged as an aerial base station to provide a sustainable services and broad coverage for remote areas or disaster recovery [2]–[4]. Specifically, HAP is a communication platform at the altitude of 20 to 50 km, where large-scale antenna arrays and wide solar panels are deployed to enable multiple-input multiple-output (MIMO) and sustain its services with the required energy [2]. As compared to the terrestrial networks, HAP communications enjoy lower scattering and wider coverage, and thus more economic and reliable transmission service can be provided for users in the rural areas than a multi-cell infrastructure [4]. Compared to the low-altitude platform (LAP) (e.g., unmanned aerial vehicle (UAV)), the quasi-stationary HAP avoids the introduction of a significant Doppler shift and the need for tracking devices. Moreover, HAP’s high altitude provides a large footprint and a wide wireless coverage [3]. Compared to satellite communications, HAP is subject to low channel attenuation and reduced transmission delay, such that offering a superior low-latency and high-throughput service for earth users [2]. In addition, HAP also enjoys high elevation angles, deployment flexibility, and hybrid connectivity with multi-band radio frequency. In this context, the utilization of HAP is an excellent design option for next-generation wireless communication and beyond.

However, since the battery-powered wireless devices in HAP networks are often deployed in rural, remote, and unpopulated areas (e.g., mobile terminals (MTs), medical implants, and sensors), it is inconvenient and infeasible to recharge or frequently replace the batteries, which constitutes an intractable bottleneck for HAP communications. Fortunately, energy harvesting technology can be adopted to alleviate the energy scarcity and prolong the lifetime of energy-constrained devices [5]. Simultaneous wireless information and power transfer (SWIPT) is a typical technique for wireless energy harvesting, where the ambient radio signals transmit information and convey energy concurrently [6]. As one of the key SWIPT techniques, the power splitting (PS) technique divides the received signals into two power streams, i.e., the information stream for signal decoding and the power one for energy harvesting, which brings great convenience to the deployment of energy-limited devices [5]. Although SWIPT has been widely adopted to provide perpetual energy supplies to terrestrial communications [6]–[8] and LAP communications [9]–[11], its energy transmission efficiency would significantly decrease with the increasing distance due to the severe large-scale fading effects, especially in the long-distance communication system. Massive MIMO with significant gain of element pattern has been utilized to overcome the severe fading effects in low earth orbit (LEO) satellites, such that SWIPT can be enabled in long-distance communications [12]. However, due to the prohibitive cost and size constraints, it may be impractical to employ a large-scale array at both the

aerial platform and receiver. Thus, investigating HAP-SWIPT networks is an open and challenging research direction.

Based on the above background, techniques which can boost the communication link quality and efficiently transmit energy from HAP to wireless devices are urgently needed. Recently, reconfigurable intelligent surface (RIS), as a revolutionary technique, can manipulate the wireless channel [13]. **The current state-of-the-art of RIS-aided wireless communications can be divided into two fold.** The **first one** is to use RIS as a **relay-like reflector** for reconfiguring the radio propagation environment, which has been widely adopted to enhance the service coverage [13]–[16], maximize the achievable rate [17]–[19], and improve the SWIPT performance of both the terrestrial and LAP communications [20]–[25]. The authors in [20] considered a passive RIS-aided SWIPT system, where the weighted sum rate was maximized by jointly optimizing the transmit precoder and passive RIS coefficients, while guaranteeing the linear energy harvesting requirement. Considering the severe “double fading” effect and practical nonlinear EH circuit, the authors of [21] proposed an active RIS-assisted PS-SWIPT systems, where the total transmit power was minimized by optimizing transmit beamformers, PS ratios, and RIS phase shifts/amplification factors. In addition, [22] further studied RIS-aided PS-SWIPT system under the imperfect CSI, and a new nonlinear EH model based Gaussian error function was introduced. Furthermore, the authors in [23] and [24] investigated RIS-aided UAV-SWIPT networks, where the user scheduling, UAV trajectory/velocity, and IRS phase shifts/reflection coefficients are jointly designed to improve the SWIPT performance. However, the above RIS-aided reflectors are not directly applicable to HAP-SWIPT networks. Specifically, HAP-SWIPT networks work at a height of several dozen kilometers, such that the HAP-SWIPT using RIS-reflector may suffer from extremely severe “double fading” effect. Although the active RIS can be utilized to overcome the above “double fading” effect in the terrestrial and LAP networks [19], long-distance HAP link’s large-scale fading must be compensated by a large amount of power, which is impractical in the energy-constrained HAP-SWIPT networks.

The **second one** is to use RIS as a **low-cost transmitter** for manipulating the electromagnetic waves, including a single-layer RIS-assisted transmitter [26]–[29] and a multi-layer one [30]–[32]. Recently, several multilayer RIS structures have been used in some research communities. Specifically, the authors in [30] proposed a multi-layer RIS transmitter to replace the a large-scale antenna array at the user side, and thus improving the **degrees of freedom** (DoFs) for beamformer design. The authors in [31] have further exploited multi-layer RIS to act as HAP’s transmitter for simultaneously defending against the jamming and eavesdropping attacks. Besides, multi-layer RIS have been adopted to execute various complex signal processing tasks, such as image classification and programmable AI machine [32]. Nevertheless, the above multi-layer RIS-related works focused on the transmitter design, which may not be tailored for HAP-SWIPT. To elaborate, the HAP-SWIPT networks with an RIS-transmitter should consider multi-user beamforming coordination at RIS, which limits the RIS’s potential in HAP-SWIPT networks. Thus, a new architecture should be proposed to fully unleash the potential of HAP-SWIPT networks.

Based on the above background, we further summarize our motivation from the following three perspectives:

► **Why does HAP networks need both SWIPT and RIS?** In principle, SWIPT can simultaneously transmit information and alleviate the energy scarcity of battery-powered users in

HAP networks [6]. However, the long-distance HAP links significantly degrade the energy transmission efficiency. Although employing large-scale array at transceiver can enable long-distance HAP-SWIPT, it leads to prohibitive hardware costs [12]. Thus, the low-cost and energy-efficient RIS can be adopted to overcome the abovementioned drawbacks, since it can boost the communication link quality by manipulating the electromagnetic wave [13].

► **Why does HAP-SWIPT networks need a new RIS architecture?** Although RIS has been adopted to improve the SWIPT performance of LAP networks, the existing RIS-reflector [13] and single-layer RIS [27] are not directly applicable to HAP-SWIPT networks, due to their “double fading” effect in long-distance links and limited aperture gain. Hence, a new RIS architecture should be harnessed to enable HAP-SWIPT.

► **How to efficiently implement resource management in RIS-assisted HAP-SWIPT networks?** The novel network designed needs efficient resource management. In the existing RIS-related works [33], the non-convex optimization problems with multiple constraints are always relaxed into convex ones, such that these convexified subproblems can be solved by the off-the-shelf optimization toolbox. However, the utilization of the optimization toolbox in the conventional methods leads to the extremely high computational complexity, which scales exponentially with the number of constraints and antennas. Besides, it is challenging to perform the untransparent computation process of optimization toolbox in FPGA hardware, especially in the energy-constrained HAP network whose computational capability is quite limited. Thus, these call for the design of scalable optimization framework.

Motivated by the aforementioned observations, we propose a novel multi-layer refracting RIS-assisted receiver to facilitate HAP-SWIPT networks, and propose a low-complexity scalable beamforming scheme to obtain semi-closed-form solutions. Specifically, the main contributions are summarized as follows:

- **Novel Multi-Layer Refracting RIS-Assisted Receiver for HAP-SWIPT networks:** We propose a multi-layer refracting RIS-assisted receiver architecture to overcome the practical energy-constrained challenges in HAP-SWIPT networks, which can also well circumvent the severe fading effect induced by the extreme long-distance HAP links. In addition, RIS’s DoFs can be fully unleashed for the beamforming design. In addition, we derive a rigorous performance analysis for the proposed multi-layer RIS-receiver architecture, and further compare it to those of passive/active RIS-reflectors and the single-layer RIS-receivers. The result reveals the capacity gain enabled by the proposed RIS-receiver for HAP-SWIPT networks.
- **Intractable Problem Formulation:** Considering the angular CSI imperfection, we formulate a worst-case optimization problem to maximize the sum achievable rate by jointly designing the transmit precoders, the PS ratios, the multi-layer refracting RIS’s coefficient matrices, and the received digital decoder. To the best of our knowledge, this is the first treatise considering worst-case sum rate maximization problem of multi-layer RIS-assisted HAP-SWIPT networks with non-linear energy-harvesting model, which is much more challenging than the total power minimization problem of single-layer RIS-aided SWIPT under bounded CSI imperfection [22], and the weighted sum rate maximization problem of single-layer RIS-aided SWIPT under linear EH model [20]. Thus, the new problem formulation must result in new algorithm.

- **Low-Complexity Scalable Beamforming Algorithm:** Since the formulated problem is non-convex, a scalable robust optimization algorithm based on the discretization method, the LogSumExp-dual scheme, and modified cyclic coordinate descent (M-CCD) is proposed to obtain closed-form solutions for the optimization variables. Different from semidefinite relaxation (SDR) (e.g., [34]), successive convex approximation (SCA) (e.g., [25]), and majorization-minimization (MM) (e.g., [20]) use off-the-shelf optimization toolbox to solve the intractable problem, the proposed algorithm does not rely on toolbox and can be easily implemented in practical massive large-scale array systems.
- **Novel LogSumExp-Dual Algorithm for Transmit Precoder:** For the optimization of the transmit precoder, we first transform the imperfect angular CSI into a robust one by using the discretization method. Then, we adopt the LogSumExp inequality to offer a tight approximation for the non-smooth objective and EH constraints. Thus, by using dual variables, we can add the EH constraints into the objective to formulate the subproblem as a general Rayleigh-quotient form. Next, by utilizing the multi-dimensional bisection method to search for the optimal dual variables, we establish the first-order optimality condition for the subproblem such that the semi-closed-form solution for the precoder can be obtained by the dual method. Different from the widely-adopted SCA in [25] introducing massive slack variable to relax the objective and constraints and the conventional generalized power iteration (GPI) in [35] addressing the optimization of transmit precoder without any QoS constraints, our proposed algorithm utilizes LogSumExp inequality and dual method to handle the objective and multiple constraints, such that obtain the optimal solution with low complexity, which is regarded as a more general and efficient GPI.
- **Improved M-CCD Algorithm for Multi-layer RIS Coefficients:** For the optimization of RIS coefficients, we first convert the original subproblem into a quadratic form by introducing Lagrange and Dinkelbach parameters. Then, we propose an iterative framework based on M-CCD to update the KKT-optimal closed-form solution for each variable. Different from the majorization-minimization (MM) framework (e.g., [20]) which utilizes the cost function to majorize all the variables and then adopts gradient projection (GP) method to update the step size parameter, the proposed M-CCD updates the block-wise variable without majorization, and thus achieving a satisfactory tradeoff between the SWIPT performance and computational complexity. This leads to effective and efficient MM implementations.
- **Optimality, Convergence, Complexity, and Simulations:** The optimality, convergence and complexity of the proposed framework are established by rigorous theoretical analysis. Besides, we compare its complexity with that of the existing RIS-related schemes, which shows that the scalability of our proposed framework is significantly better than that of existing schemes. Numerical results demonstrate that with the help of the multi-layer RIS-assisted receiver, our proposed framework outperforms the existing benchmarks.

Notation: \mathbf{X}^H , \mathbf{X}^T , \mathbf{X}^* , and $\|\mathbf{X}\|_F$ denote conjugate transpose, transpose, conjugate, and Frobenius norm of a matrix \mathbf{X} . The notations $E\{\cdot\}$, $\text{Tr}\{\cdot\}$, $\text{Re}\{\cdot\}$, and $\lambda\{\cdot\}$ denote the expectation, trace, real part, and eigenvalue of a complex

number or matrix, respectively. $\mathbb{C}^{m \times n}$ represents the complex space of $m \times n$ dimensions. The symbol $\mathbb{H}^{n \times n}$ is the Hermitian matrix of $n \times n$ dimensions. $[\cdot]_{n,n}$ represents the n th diagonal element of a matrix. $\mathbf{X} \succeq 0$ means that the matrix \mathbf{X} is positive semi-definite. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean vector x and covariance matrix Σ is denoted by $\mathcal{CN}(x, \Sigma)$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1 (a), we consider a RIS-assisted HAP-SWIPT network, where an HAP equipped with $N_H = N_{H1} \times N_{H2}$ uniform planar array (UPA) antennas transmits the desired signals to K MTs with the aid of the multi-layer refracting RIS-assisted receiver. At the MTs, the received signals are split for information decoding and energy harvesting, i.e., SWIPT. Here, the gateway acts as a control center to collect various kinds of information and implement resource allocation in the whole network [4], so that HAP can cooperate with multi-layer refracting RIS receivers for synchronizing the RIS's coefficients and transmit beamformer. It is assumed that the k -th MT's multi-layer refracting RIS is composed of A layers having $N_{E,ka} = N_{E,ka1} \times N_{E,ka2}$ units on the a -th layer, and the k -th MT is equipped with $N_{D,k}$ Rx antennas for digital decoding. Besides, the channels from HAP to the first RIS layer of k -th MT, from the a -th layer to the $(a+1)$ -th layer, and from the A -th layer to the k -th MT's Rx antennas are denoted as $\mathbf{G}_k \in \mathbb{C}^{N_{E,k1} \times N_H}$, $\mathbf{B}_{ka} \in \mathbb{C}^{N_{E,k(a+1)} \times N_{E,ka}}$, and $\mathbf{B}_{kA} \in \mathbb{C}^{N_{D,k} \times N_{E,kA}}$, respectively.

A. Multi-Layer Refracting RIS-Assisted Receiver

As shown in Fig. 1 (b), multiple refracting RISs are vertically stacked in front of the MT's Rx feeds to form a multi-layer receiver architecture, where the refracting RISs and the Rx antennas perform the analog and the digital decoding on the HAP's transmitted signals, respectively. Specifically, each RIS receives a superposition of the signals over the air, forwards them to the RIS-aided phase shifters, and transmits the phase-delayed signals to the transmit units via the microstrip [31]. Then, the feeds connected to the dedicated radio frequency (RF) chains and power amplifiers (PA) impose the digital decoding on the receive signals. Note that there exists an enclosure containing the entire multi-layer RIS receiver structure, which can use the surrounded absorbing materials to protect the internal channel from external interference and reduce the potential energy loss. Naturally, the gap between adjacent vertical RIS layers is flexible, but it is generally compact for implementation.

The a -th layer's refracting coefficient matrix¹ of the MT k is given by $\Xi_{ka} = \text{diag}\left(e^{j\theta_{ka,1}}, \dots, e^{j\theta_{ka,N_{E,ka}}}\right) \in \mathbb{C}^{N_{E,ka} \times N_{E,ka}}$, and the k -th MT's digital processing vector

¹The performance degradation introduced by the discrete phase shifts and its corresponding optimization methods have been thoroughly investigated in the existing works (e.g., [13]). Furthermore, the full characterization of this discrete phase shifts scenario requires a separate full investigation. Thus, due to the space limits and extensive related works, we leave investigation of discrete phase shifts for the future.

²Due to the fact that RIS elements have a nearly frequency-flat response [36] and half-wavelength-sized spacing between adjacent RIS units [13], this paper captures the widely-adopted dependence that the phase can be tuned in the sense of having a constant amplitude and no mutual coupling. However, the simplified RIS model might not satisfy any of these ideal conditions, such that more practical phase-dependent amplitude variation in [37] and mutual coupling in [38] should be considered, which is still quite a challenging task and worth in-depth investigation in our future works.

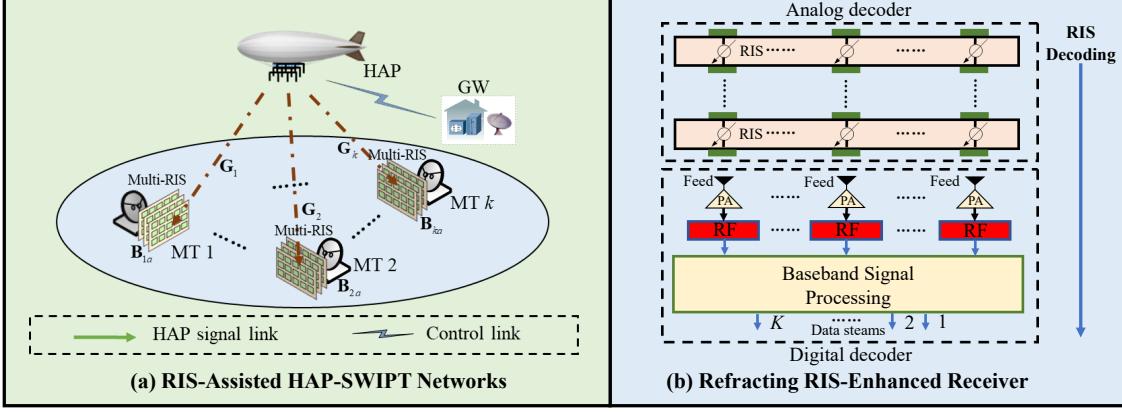


Fig. 1: The illustration of RIS-assisted HAP-SWIPT networks and RIS-enhanced receiver.

is expressed as $\mathbf{v}_k \in \mathbb{C}^{N_{D,k} \times 1}$. Here, $\theta_{ka,n} \in [0, 2\pi)$ denotes the phase shift of the n -th RIS unit. As such, the equivalent analog decoder generated by the k -th MT's RIS-aided receiver is given by $\Omega_{k,(1,A)}$, i.e.,

$$\Omega_{k,(u,v)} = \begin{cases} \prod_{a=1}^v \mathbf{B}_{ka} \mathbf{\Xi}_{ka}, u, v \in [A], \\ \mathbf{I}_{N_{E,k1}}, u = 1, v = 0, \\ \mathbf{I}_{N_{D,k}}, u = A + 1, v = A. \end{cases} \quad (1)$$

To clarify the characteristics of the multi-layer refracting RIS receiver, its key advantages are summarized as follows.

- **Simplified Hardware Architecture:** Different from the typical MIMO receiver requiring massive RF chains to achieve spatial gains for enhancing the energy transmission efficiency, the multi-layer refracting RIS receiver only requires low-cost tunnel diodes, whose power and hardware consumption are extremely low [39]. Thus, the hardware architecture of multi-layer refracting RIS receiver is simplified in comparison with massive MIMO receiver.

- **Miniaturization and High Scalability:** Due to the small size of RIS unit and the compact cascaded structure, the multi-layer refracting RIS receiver has much smaller physical size than those of the typical MIMO and single-layer RIS-aided receiver. Specifically, the single-layer RIS-aided and typical MIMO receiver can only deploy all $N_{\text{Tot}} = \sum_{a=1}^A N_{E,a}$ RIS units/antennas at the transversal side, the multi-layer RIS receiver can deploy A layers of $N_{E,ka} = N_{E,k1} \times N_{E,k2}$ RIS units at longitudinal side. Combined with the fact that the small gap between adjacent layers [31], the scalability and miniaturization of the multi-layer refracting RIS receiver are highly desirable.

- **Enhanced Aperture Gain for Long-Distance SWIPT:** In the MIMO receiver, the aperture gain asymptotically increases in the order of $N_{\text{Tot}} = \sum_{a=1}^A N_{E,a}$ [13]. Besides, as shown in **Theorem 2**, the aperture gains of single-layer passive RIS receiver and single-layer active RIS receiver are approximately increase in the order of $(\sum_{a=1}^A N_{E,a})^2$ and $\sum_{a=1}^A N_{E,a}$, respectively. In contrast, due to the multi-layer structure, the aperture gain of multi-layer refracting RIS receiver asymptotically increases in the order of $(\prod_{a=1}^A N_{E,a})^2$, thereby compensating for the severe path loss induced by long-distance path loss. Therefore, the multi-layer refracting RIS receiver can achieve efficient HAP-SWIPT.

- **Increased DoFs for Beamforming Design:** Different from typical MIMO receiver having N_{Tot} DoFs and

single-layer RIS receiver having $\min \left\{ \frac{N_{\text{Tot}}}{2} - \frac{1}{2}, N_D \right\}$ DoFs, the spatial DoFs of multi-layer refracting RIS receiver is $\min \left\{ \frac{\prod_{a=1}^A N_{E,a}}{2} - \frac{1}{2}, N_D \right\}$ [40]. Besides, as shown in **Theorem 3**, the multi-layer RIS can regulate the the amplitude of incident signal within a larger range than the existing receivers. The above findings suggest that our proposed receiver can generate increased DoFs for beamforming design.

B. Signal Transmission and Energy Harvesting Model

Denote s_k as the desired symbol transmitted by the HAP to the k -th MT, which satisfies $\mathbb{E}\{|s_k|^2\} = 1$. Prior to transmission, s_k is processed by the digital precoders $\mathbf{w}_k \in \mathbb{C}^{N_H \times 1}$, such that the signal transmitted by the HAP is $\sum_{k=1}^K \mathbf{w}_k s_k$. Then, after decoding the transmitted signals by the analog and the digital decoder $\Omega_{k,(1,A)}$ and \mathbf{v}_k , the received signal at the k -th MT can be expressed as

$$y_k = \mathbf{v}_k^H \left(\Omega_{k,(1,A)} \mathbf{G}_k \sum_{k=1}^K \mathbf{w}_k s_k + \mathbf{n}_k \right), \quad (2)$$

where $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma_k^2 \mathbf{I}_{N_{D,k}})$ is the thermal noise at the k -th MT's Rx antennas. Subsequently, the k -th MT divides the received signals into two portions, where γ_k portion is utilized for information decoding and $1 - \gamma_k$ one is used for energy harvesting. Hence, the k -th MT's received signals for information decoding and energy harvesting are, respectively, given by $y_{ID,k} = \sqrt{\gamma_k} y_k + z_k$ and $y_{EH,k} = \sqrt{1 - \gamma_k} y_k$, where $\gamma_k \in (0, 1]$ is the PS ratio and $z_k \sim \mathcal{CN}(0, \delta_k^2)$ is the additional noise induced by the k -th MT's information decoding circuit [25]. Thus, the achievable rate for information decoding at the k -th MT can be modeled as

$$R_{ID,k} = \log_2 \left(1 + \frac{|\mathbf{g}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k}^K |\mathbf{g}_k^H \mathbf{w}_i|^2 + \sigma_k^2 + \delta_k^2 / \gamma_k} \right), \quad (3)$$

where $\mathbf{g}_k = \mathbf{G}_k^H \Omega_{k,(1,A)}^H \mathbf{v}_k$.

Note that there are two major drawbacks in the practical energy-harvesting circuits. First, the energy harvesting cannot be activated when the input power is less than the sensitivity power. Second, the harvested energy increases with the input power non-linearly [6]. To account for these effects precisely, this paper adopts the nonlinear energy-harvesting model in [6]. As such, the harvested energy at k -th MT is

$$\zeta_{EH,k} = (\alpha p_k + \beta) / (p_k + \varepsilon) - \beta / \varepsilon, \quad (4)$$

where $p_k = \eta_k (1 - \gamma_k) \sum_{i=1}^K |\mathbf{g}_k^H \mathbf{w}_i|^2$ denotes the input power for energy harvesting, $\{\alpha, \beta, \varepsilon\} > 0$ captures the

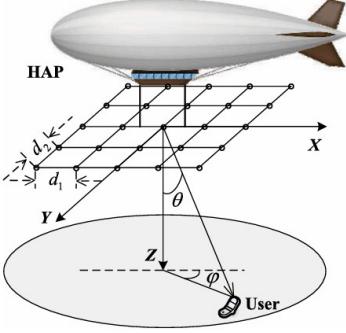


Fig. 2: Geometrical relation between HAP and any user.

sensitivity and saturation thresholds of the energy-harvesting circuits, and $\eta_k \in (0, 1]$ is the energy conversion efficiency.

C. Near/Far-Field Channel Model

Because of the extremely short distance between the adjacent vertical RIS layers, i.e., $2\lambda\sqrt{N_{E,a}}/\sqrt{\pi}$, the channels $\{\mathbf{B}_{ka}\}$ in multi-layer RIS-receiver are near-field channels. In contrast, due to the fact that HAP networks work at a height of several dozen kilometers, the channels $\{\mathbf{G}_k\}$ from HAP to the first RIS layer of k -th MT are far-field channels.

As for the near-field channels, the radar illumination model in [39] is adopted to characterize $\{\mathbf{B}_{ka}\}$, which is given by

$$\mathbf{B}_{ka} = \left[\frac{\lambda_f \sqrt{\hat{\rho}} G_{n,v}^D(\theta^R, \varphi^R) G_{n,v}^R(\theta^D, \varphi^D)}{4\pi d_{n,v}} e^{-j\frac{2\pi d_{n,v}}{\lambda}} \right]_{n,v}, \quad (5)$$

where $\hat{\rho}$ is the RIS's efficiency, λ_f is the carrier wavelength, $d_{n,v}$ is the distance between the n -th RIS unit on the a -th layer and the v -th unit on the $(a+1)$ -th layer, and $G_{n,v}^D(\theta^R, \varphi^R)$, $G_{n,v}^R(\theta^D, \varphi^D)$ are the active and passive gains, respectively.

As illustrated in Fig. 2, we assume that HAP employs UPA of dimension $N_H = N_{H1} \times N_{H2}$. Due to the quasi-optical and highly directional nature of the radio wave propagation at high frequency band, the far-field HAP channels $\{\mathbf{G}_k\}$ can be modeled as the superposition of a predominant line-of-sight component and a sparse set of single-bounce non-LoS ones [4], namely,

$$\mathbf{G}_k = \sqrt{G(\theta_0, \varphi_0)} g_{k,0} \mathbf{a}_P(\theta_{k,0}^{Rx}, \varphi_{k,0}^{Rx}) \mathbf{a}_P^H(\theta_{k,0}^{Tx}, \varphi_{k,0}^{Tx}) \quad (6)$$

$$+ \sqrt{\frac{1}{MP} \sum_{d=1}^{MP} \sqrt{G(\theta_d, \varphi_d)} g_{k,d} \mathbf{a}_P(\theta_{k,d}^{Rx}, \varphi_{k,d}^{Rx}) \mathbf{a}_P^H(\theta_{k,d}^{Tx}, \varphi_{k,d}^{Tx})},$$

where MP is the total number of paths, θ^{Tx} (θ^{Rx}) is the vertical AoD (AoA), and φ^{Tx} (φ^{Rx}) denotes the horizontal AoD (AoA). g represents the large-scale fading coefficients, and $g \sim \mathcal{CN}(0, \sigma_{PL}^2)$, where $\sigma_{PL}^2 = 10^{PL/10}$, $PL = -30.18 - 26 \log_{10}(d_s)$ [dB] and d_s is the link distance in meters. $\mathbf{a}(\theta, \varphi)$ is the steering vector of UPA in [31]. Besides, according to the model introduced by ITU, the element pattern $G(\theta, \varphi)$ in dB, i.e., $\widehat{G}(\theta, \varphi) = 10 \log_{10}(G(\theta, \varphi))$ can be expressed as [4]

$$\widehat{G}(\theta, \varphi) = E_{max} - \min\{G_x(\theta, \varphi) + G_y(\theta, \varphi), S_m\}, \quad (7)$$

where E_{max} is the maximum antenna gain, $G_x(\theta, \varphi)$ and $G_y(\theta, \varphi)$ are the relative patterns in x and y planes, respectively, which are given by [4]

$$G_i(\theta, \varphi) = \min \left\{ 12 \left(\frac{(\arctan(\cot \theta / \cos \varphi))}{\varphi_i^{3dB}} \right)^2, S_m \right\},$$

where φ_i^{3dB} denotes the 3 dB beamwidth, and S_m represents the side-lobe level of the antenna pattern.

Note that $\{\mathbf{B}_{ka}\}$ is a deterministic matrix due to the short distance between the adjacent vertical layers, such that it can be precisely measured in advance [27], [31], [39]. However, owing to the channel estimation error and feedback delay, we assume that only the angular information based imperfect CSI $\{\mathbf{G}_k\}$ can be obtained at the gateway³ [4]. To elaborate, the precise positions of MTs are unknown but the uncertainty region of the angular information can be obtained. Thus, $\{\mathbf{G}_k\}$ belongs to a given uncertainty set, i.e.,

$$\Delta = \{ \mathbf{G}_k | \theta_k \in [\theta_{k,L}, \theta_{k,U}], \varphi_k \in [\varphi_{k,L}, \varphi_{k,U}] \}, \quad (8)$$

where φ_U and φ_L are the upper and lower bounds of elevation angle, θ_U and θ_L denote the upper and lower bounds of azimuth angle, respectively.

Remark 1 (SWIPT Feasibility in RIS-Receiver Aided HAP Networks) : Due to the utilization of antenna array on HAP, the received power increases with the order of N_H [14]. Besides, according to the model introduced by 3GPP [45], the gain of element pattern $\widehat{G}(\theta, \varphi)$ in the HAP's transmitter is larger than 32 dB (i.e., $E_{max} = 52$ dB and $S_m = 20$ dB). Thus, the severe large-scale fading effects induced by the long-distance communications can be partially compensated. Moreover, as shown in **Theorem 2**, the received power can be further significantly enhanced, which is proportional to $(\prod_{a=1}^A N_{E,a})^2$. Therefore, the feasibility of HAP-SWIPT networks can be boosted by using the RIS-enhanced receiver.

D. Problem Formulation

In this paper, under the angular uncertainty Δ , we aim to jointly optimize the transmit precoders \mathbf{w}_k , PS ratios γ_k , multi-layer refracting RIS's coefficient matrices Ξ_{ka} , and the received digital decoder \mathbf{v}_k to maximize the worst-case sum achievable rate, while meeting the information rate requirements of the MTs, the harvested energy requirements, the transmit power constraints, and the RIS unit-modula constraints⁴. Thus, the corresponding optimization problem⁵ is

$$\max_{\mathbf{w}_k, \Xi_{ka}, \mathbf{v}_k, \gamma_k} \min_{\Delta} \sum_{k=1}^K R_{ID,k} \quad (9)$$

³Since the wireless channels undergo slow fading and due to the existence of RF chains at multi-layer refracting RIS receivers [22], the CSI is available at the gateway through feedback/training sent from the RIS-receivers via backhaul channel [41], [42]. To elaborate, the time division duplex (TDD) mode allows for channel estimation by exploiting channel reciprocity [22], where each coherence time block is divided into the channel estimation stage and data transmission stage. In the channel estimation stage, there are many efficient methods in the literature, such as the three-phase pilot-based channel estimation algorithm [43] and alternating least squares (ALS) [44].

⁴SWIPT aims to use the ambient radio signals to transmit information and convey energy concurrently [6]. Thus, there are three main research streams in the current state-of-the-art of SWIPT, i.e., maximizing the sum rate while meeting the energy harvesting constraints (e.g., [20]), maximizing harvested power while satisfying the data rate requirements (e.g., [46]), and minimizing total power while meeting the data rate requirements (e.g., [23]). Since HAP aims to provide ultra-high rate services and broad coverage for battery-powered users in sixth generation (6G) wireless networks [2], this paper considers a worst-case sum achievable rate maximization problem, while meeting the harvested energy requirements, which is consistent with [20]. The other two research streams will be left for our future works.

⁵Due to the nature of a quasi-stationary position in the stratosphere, the optimization only considered the HAP's transmit beamforming instead of the trajectory of HAP, which is consistent with the existing works, e.g., [4]. However, the optimization is still related to HAP-related parameters. Specifically, the height of HAP affects the propagation characteristics of the A2G communication link since the LoS condition and the environment between HAP and MTs alter as the height varies [47], which will indirectly affect the optimization of HAP's beamforming.

$$\begin{aligned}
& \text{s.t. C1 : } \min_{\Delta} R_{ID,k} \geq \Gamma_k, \forall k, \text{ C2 : } \min_{\Delta} \zeta_{EH,k} \geq \varsigma_{\max}, \forall k, \\
& \text{C3 : } \sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_{\max}, \text{ C5 : } \|\mathbf{v}_k\| = 1, \forall k, \\
& \text{C4 : } |\mathbf{\Xi}_{ka}]_{n,n}| = 1, \forall k, a, n,
\end{aligned}$$

where Γ_k is the minimum rate threshold, ς_{\max} denotes the harvested power requirement, and P_{\max} is the power limit at the HAP. Constraint C1 ensures that the achievable rate of each MT is above the quality-of-service (QoS) requirement Γ_k . Constraint C2 guarantees that the harvested power for self-sustainability is higher than the harvested power requirement ς_{\max} . Constraint C3 indicates that the total transmit power must be lower than the power budget P_{\max} . Constraints C4 and C5 specify trivial but important RIS unit-module and digital decoder's normalization requirement, respectively.

Problem (9) is very challenging for several reasons. First, problem (9) is complicated due to jointly coupled variables \mathbf{w}_k , γ_k , and $\mathbf{\Xi}_{ka}$. Second, the angular uncertainty Δ and the conflicting constraints C1-C4 make (9) non-convex. Third, the multi-layer RIS structure makes the optimization problem much more complex than the former works, e.g., [25], which means that the existing schemes are not directly applicable to (9). Most importantly, although the conventional optimization schemes such as SDR in [34], SCA in [25], and MM in [20], can be adopted to solve problem (9) by introducing multiple constraints into the optimization problem, their computational complexity scales exponentially with the number of constraints, especially for the case of large N_H and $N_{E,ka}$. Besides, an optimization toolbox is needed in the above-mentioned schemes, which is generally hard to implement in practical hardware due to its extremely high complexity. To avoid huge computational complexity, in the sequel, we propose a low-complexity beamforming scheme to obtain the semi-closed-form solutions.

III. SCALABLE ROBUST BEAMFORMING OPTIMIZATION FRAMEWORK FOR PROBLEM (9)

In this section, based on the alternating optimization framework, a scalable robust beamforming optimization scheme for solving problem (9) is proposed. To elaborate, we decouple problem (9) into four subproblems and then derive the semi-closed-form solution to each subproblem, thereby significantly improving the scalability of our proposed scheme.

A. LogSumExp-Dual Scheme for \mathbf{w}_k

Firstly, we focus on optimizing the transmit precoder \mathbf{w}_k under the angular uncertainty Δ . By defining $\bar{\varsigma}_{\max} = \varepsilon \varsigma_{\max} / (\alpha - \varsigma_{\max} - \beta/\varepsilon)$ and $\bar{\gamma}_k = \log_2(\bar{\varsigma}_{\max} / (1 - \gamma_k))$, the subproblem for \mathbf{w}_k can be expressed as

$$\begin{aligned}
& \max_{\mathbf{w}_k} \min_{\Delta} \sum_{k=1}^K R_{ID,k} \\
& \text{s.t. C1, } \bar{\text{C3}} : \sum_{k=1}^K \|\mathbf{w}_k\|^2 = P_{\max}, \\
& \bar{\text{C2}} : \min_{\Delta} R_{EH,k} = \log_2 \left(\sum_{i=1}^K |\mathbf{g}_k^H \mathbf{w}_i|^2 \right) \geq \bar{\gamma}_k, \forall k.
\end{aligned} \tag{10}$$

Note that the power constraint C3 must hold with equality when the objective function of (10) achieves the maximum, and thus we transform C3 into $\bar{\text{C3}}$ in this subproblem. Clearly, there still exists three major obstacles which prevent us from obtaining the semi-closed-form solution \mathbf{w}_k to (10), i.e., the continuous angular uncertainty Δ , the non-convexity of the objective function, and the non-smoothness of C1 and $\bar{\text{C2}}$. In what follows, we address the foregoing obstacles one-by-one.

At first, we turn to the continuous angular uncertainty Δ with infinite possibilities. A discretization method is exploited to convert the imperfect CSI with Δ into a tractable form. To elaborate, since the available CSI belongs to a given continuous angular uncertainty set, we select uniformly spaced angles as

$$\begin{aligned}
\theta^{(p)} &= \theta_L + (i-1)\Delta\theta, p = 1, \dots, Q_1, \\
\varphi^{(q)} &= \varphi_L + (j-1)\Delta\varphi, q = 1, \dots, Q_2,
\end{aligned} \tag{11}$$

where Q_1 and Q_2 are the numbers of samples for θ and φ , respectively, $\Delta\theta = (\theta_U - \theta_L)/(Q_1 - 1)$, and $\Delta\varphi = (\varphi_U - \varphi_L)/(Q_2 - 1)$. Hence, the equivalent CSI \mathbf{g}_k is converted into a robust one, i.e.,

$$\widehat{\mathbf{g}}_k \widehat{\mathbf{g}}_k^H = \sum_{p_1=1}^{N_{H1} N_{E1,k1}} \sum_{q_1=1}^{N_{H2} N_{E2,k1}} \sum_{p_2=1}^{N_{H1} N_{E1,k1}} \sum_{q_2=1}^{N_{H2} N_{E2,k1}} \frac{1}{(N_H N_{E,k1})^2} \widehat{\mathbf{g}}_k^{(p_1, q_1, p_2, q_2)} \widehat{\mathbf{g}}_k^{(p_1, q_1, p_2, q_2), H}, \tag{12}$$

where $\widehat{\mathbf{g}}_k^{(p_1, q_1, p_2, q_2)} = \mathbf{G}_k^{(p_1, q_1, p_2, q_2), H} \boldsymbol{\Omega}_{k,(1,A)}^H \mathbf{v}_k$, p_1, q_1 and p_2, q_2 are used to represent the phase for the discrete two sets $(\theta_{TX}^{\text{TX}}, \varphi_{TX}^{\text{TX}})$ and $(\theta_{RX}^{\text{RX}}, \varphi_{RX}^{\text{RX}})$, respectively. The interested readers can refer to our previous works [4], [17], [26], [27], [31] for more details, which are omitted here for brevity. With the robust CSI, the \min_{Δ} operation in both the objective function and the constraints can be removed. Next, we handle the non-smoothness of constraints and the non-convex objective. The constraints C1 and $\bar{\text{C2}}$ can be equivalently recast as

$$\widetilde{\text{C1}} : \min_{\forall k \in [K]} R_{ID,k} \geq \Gamma_k, \widetilde{\text{C2}} : \min_{\forall k \in [K]} R_{EH,k} \geq \bar{\gamma}_k. \tag{13}$$

Thus, the LogSumExp inequality is adopted to approximate the non-smooth minimum functions $\widetilde{\text{C1}}$ and $\widetilde{\text{C2}}$ as the smooth ones, which is given by [35]

$$\min_{\forall k \in [K]} [x_k] \approx -\rho \ln \left(\sum_{k=1}^K \exp \left(\log_2 \left(\frac{x_k}{-\rho} \right) \right) \right), \tag{14}$$

where $\rho > 0$ denotes the smoothing parameter. Substituting (14) in (13), we can reformulate $\widetilde{\text{C1}}$ and $\widetilde{\text{C2}}$ as

$$\begin{aligned}
\min_{\forall k \in [K]} R_{ID,k} &= \min_{\forall k \in [K]} \log_2 \left(\frac{\overline{\mathbf{w}}^H \widehat{\mathbf{A}}_k \overline{\mathbf{w}}}{\overline{\mathbf{w}}^H \widehat{\mathbf{B}}_k \overline{\mathbf{w}}} \right) \\
&\approx -\rho \ln \left\{ \sum_{k=1}^K \exp \left[\log_2 \left(\frac{\overline{\mathbf{w}}^H \widehat{\mathbf{A}}_k \overline{\mathbf{w}}}{\overline{\mathbf{w}}^H \widehat{\mathbf{B}}_k \overline{\mathbf{w}}} \right)^{-\frac{1}{\rho}} \right] \right\}, \tag{15}
\end{aligned}$$

$$\begin{aligned}
\min_{\forall k \in [K]} R_{EH,k} &= \min_{\forall k \in [K]} \log_2 \left(\frac{\overline{\mathbf{w}}^H \widehat{\mathbf{A}}_k \overline{\mathbf{w}}}{\overline{\mathbf{w}}^H \mathbf{C}_k \overline{\mathbf{w}}} \right) \\
&\approx -\rho \ln \left\{ \sum_{k=1}^K \exp \left[\log_2 \left(\frac{\overline{\mathbf{w}}^H \widehat{\mathbf{A}}_k \overline{\mathbf{w}}}{\overline{\mathbf{w}}^H \mathbf{C}_k \overline{\mathbf{w}}} \right)^{-\frac{1}{\rho}} \right] \right\}, \tag{16}
\end{aligned}$$

where $\overline{\mathbf{w}} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_K^T]^T \in \mathbb{C}^{KN_H \times 1}$, $\bar{\sigma}_k^2 = \sigma_k^2 + \delta_k^2 / \gamma_k$, $\mathbf{C}_k = (1/P_{\max}) \mathbf{I}_{KN_H}$,

$$\widehat{\mathbf{A}}_k = \text{Blkdiag} \left\{ \widehat{\mathbf{g}}_k \widehat{\mathbf{g}}_k^H, \dots, \widehat{\mathbf{g}}_k \widehat{\mathbf{g}}_k^H, \dots, \mathbf{0}_{N_H \times N_H} \right\}_{k\text{-th block}}.$$

We observe that (15) and (16) are Rayleigh quotients since both of the numerator and denominator can be normalized by $\|\overline{\mathbf{w}}\| = \sqrt{P_{\max}}$. For this reason, the power constraint $\bar{\text{C3}}$ can be ignored. Besides, $\bar{\text{C3}}$ can also be guaranteed by the derived semi-closed-form solution later. Hence, problem (10) can be reformulated as

$$\max_{\overline{\mathbf{w}}} \sum_{k=1}^K R_{ID,k} \text{ s.t. } \widetilde{\text{C1}}, \widetilde{\text{C2}}. \tag{17}$$

Similar to [20], the optimal solution to (17) can be obtained

$$\begin{aligned}\mathbf{N}(\bar{\mathbf{w}}) &= \mathcal{L}_{\text{num}}(\bar{\mathbf{w}}) \times \sum_{k=1}^K \left\{ 1 + \frac{\lambda \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_k \bar{\mathbf{w}}} \right)\right)}{\sum_{m=1}^K \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_m \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_m \bar{\mathbf{w}}} \right)\right)} + \frac{\mu \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_m \bar{\mathbf{w}}} \right)\right)}{\sum_{m=1}^K \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_m \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_k \bar{\mathbf{w}}} \right)\right)} \right\} \frac{\hat{\mathbf{A}}_k}{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_k \bar{\mathbf{w}}}, \quad (21) \\ \mathbf{M}(\bar{\mathbf{w}}) &= \mathcal{L}_{\text{den}}(\bar{\mathbf{w}}) \times \sum_{k=1}^K \left\{ \left(1 + \frac{\lambda \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_k \bar{\mathbf{w}}} \right)\right)}{\sum_{m=1}^K \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_m \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_m \bar{\mathbf{w}}} \right)\right)} \right) \frac{\hat{\mathbf{B}}_k}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_k \bar{\mathbf{w}}} + \frac{\mu \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_m \bar{\mathbf{w}}} \right)\right)}{\sum_{m=1}^K \exp\left(-\frac{1}{\rho} \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_m \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_k \bar{\mathbf{w}}} \right)\right)} \frac{\hat{\mathbf{C}}_k}{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_k \bar{\mathbf{w}}} \right\}.\end{aligned}$$

by solving its dual problem instead of its original one. As such, by adding $\tilde{C}1$ and $\tilde{C}2$ into the objective function in (10) with the non-negative Lagrange multipliers λ and μ , we have the partial Lagrangian function of problem (10), i.e.,

$$\begin{aligned}\mathcal{L}(\bar{\mathbf{w}}, \lambda, \mu) &= \sum_{k=1}^K \log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_k \bar{\mathbf{w}}} \right) \\ &\quad - \lambda \left(\Gamma_k + \rho \ln \left\{ \sum_{k=1}^K \exp \left[\log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{B}}_k \bar{\mathbf{w}}} \right)^{-\frac{1}{\rho}} \right] \right\} \right) \\ &\quad - \mu \left(\bar{\gamma}_k + \rho \ln \left\{ \sum_{k=1}^K \exp \left[\log_2 \left(\frac{\bar{\mathbf{w}}^H \hat{\mathbf{A}}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \hat{\mathbf{C}}_k \bar{\mathbf{w}}} \right)^{-\frac{1}{\rho}} \right] \right\} \right).\end{aligned}\quad (18)$$

The dual problem can be expressed as

$$\min_{\lambda, \mu} \max_{\bar{\mathbf{w}}} \mathcal{L}(\bar{\mathbf{w}}, \lambda, \mu) \quad s.t. \quad \lambda, \mu \geq 0. \quad (19)$$

Before handling the dual problem (19), we should derive the closed-form solution for $\mathcal{L}(\bar{\mathbf{w}})$ with given λ and μ . By calculating the gradient of the Lagrangian function (18) w.r.t. $\bar{\mathbf{w}}$, we can obtain the first-order optimality condition of (17), which is shown in the following lemma.

Lemma 1: The first-order optimality condition of (17) is satisfied when the following holds:

$$\mathbf{M}^\dagger(\bar{\mathbf{w}}) \mathbf{N}(\bar{\mathbf{w}}) \bar{\mathbf{w}} = \mathcal{L}(\bar{\mathbf{w}}) \bar{\mathbf{w}}, \quad (20)$$

where $\mathbf{M}(\bar{\mathbf{w}})$ and $\mathbf{N}(\bar{\mathbf{w}})$ are shown in (21) at the top of next page, and $\mathcal{L}_{\text{num}}(\bar{\mathbf{w}})$, $\mathcal{L}_{\text{den}}(\bar{\mathbf{w}})$ are the numerator and the denominator of $\mathcal{L}(\bar{\mathbf{w}})$.

Proof: Please refer to Appendix A. \blacksquare

However, it is still challenging to obtain the closed-form solution to $\bar{\mathbf{w}}$. Thus, we have the following proposition.

Proposition 1: Denoting the optimal point maximizing the objective function of (17) as $\bar{\mathbf{w}}^*$, then $\bar{\mathbf{w}}^*$ must be the leading eigenvector of $\mathbf{M}^\dagger(\bar{\mathbf{w}}^*) \mathbf{N}(\bar{\mathbf{w}}^*)$ satisfying

$$\mathbf{M}^\dagger(\bar{\mathbf{w}}^*) \mathbf{N}(\bar{\mathbf{w}}^*) \bar{\mathbf{w}}^* = \lambda_{\max} \bar{\mathbf{w}}^*, \quad (22)$$

where λ_{\max} is the maximum eigenvector. Besides, $\bar{\mathbf{w}}^*$ can be obtained by calculating the following closed-form solution in an iterative manner, which is given by

$$\bar{\mathbf{w}}^{(i_d+1)} = \sqrt{P_{\max}} \frac{\mathbf{M}^{-1}(\bar{\mathbf{w}}^{(i_d)}) \mathbf{N}(\bar{\mathbf{w}}^{(i_d)}) \bar{\mathbf{w}}^{(i_d)}}{\|\mathbf{M}^{-1}(\bar{\mathbf{w}}^{(i_d)}) \mathbf{N}(\bar{\mathbf{w}}^{(i_d)}) \bar{\mathbf{w}}^{(i_d)}\|}, \quad (23)$$

where $\bar{\mathbf{w}}^{(i_d+1)}$ denotes the solution obtained at the i_d -th iteration. For the initial point $\bar{\mathbf{w}}^{(0)}$, we can adopt the maximum ratio transmission (MRT). Obviously, the power constraint $\tilde{C}3$ is automatically guaranteed by (23). Finally, $\bar{\mathbf{w}}^{(i_d)}$ converges to the optimal solution.

Proof: Please refer to Appendix B. \blacksquare

However, **Proposition 1** is executed with fixed λ and μ . Hence, we optimize the optimal λ and μ of the dual problem (19) in the following. Defining the corresponding $\bar{\mathbf{w}}$ with given λ and μ as $\bar{\mathbf{w}}(\lambda, \mu)$, the optimal λ and μ can be found by the following complementary slackness condition, namely,

$$\lambda \mathcal{G}(\lambda) = \lambda \sum_{k=1}^K \left\{ \log_2 \left(\frac{\bar{\mathbf{w}}^H(\lambda) \hat{\mathbf{A}}_k \bar{\mathbf{w}}(\lambda)}{\bar{\mathbf{w}}^H(\lambda) \hat{\mathbf{B}}_k \bar{\mathbf{w}}(\lambda)} \right) - \Gamma_k \right\} = 0, \quad (24)$$

$$\mu \mathcal{P}(\mu) = \mu \sum_{k=1}^K \left\{ \log_2 \left(\frac{\bar{\mathbf{w}}^H(\mu) \hat{\mathbf{A}}_k \bar{\mathbf{w}}(\mu)}{\bar{\mathbf{w}}^H(\mu) \hat{\mathbf{C}}_k \bar{\mathbf{w}}(\mu)} \right) - \bar{\gamma}_k \right\} = 0. \quad (25)$$

Then, we have the following cases:

- 1) $\lambda = 0$ and $\mu = 0$, if $\mathcal{G}(0) \geq 0$ and $\mathcal{P}(0) \geq 0$;
- 2) Otherwise, (24) and (25) hold if and only if $\mathcal{G}(\lambda) = 0$ and $\mathcal{P}(\mu) = 0$.

As shown in [20, Lemma 1], $\mathcal{G}(\lambda)$ and $\mathcal{P}(\mu)$ are monotonically decreasing w.r.t. λ and μ . Thus, we utilize the multi-dimensional bisection search method to search for the optimal λ and μ along the line, i.e.,

$$\mathbf{p}_{\bar{\mathbf{w}}} = \mathbf{p}_{\bar{\mathbf{w}}}^L + \frac{\mathbf{p}_{\bar{\mathbf{w}}}^U - \mathbf{p}_{\bar{\mathbf{w}}}^L}{\mathcal{Q}(\mathbf{p}_{\bar{\mathbf{w}}}^L) - \mathcal{Q}(\mathbf{p}_{\bar{\mathbf{w}}}^U)} (\mathcal{Q}(\mathbf{p}_{\bar{\mathbf{w}}}^L) - \mathcal{Q}(\mathbf{p}_{\bar{\mathbf{w}}}^U)), \quad (26)$$

where $\mathbf{p}_{\bar{\mathbf{w}}} = [\lambda, \mu]$, $\mathcal{Q}(\mathbf{p}_{\bar{\mathbf{w}}}) = \mathcal{G}(\lambda) + \mathcal{P}(\mu)$, and $\mathbf{p}_{\bar{\mathbf{w}}}^L$, $\mathbf{p}_{\bar{\mathbf{w}}}^U$ denote the lower and upper bounds of $\mathbf{p}_{\bar{\mathbf{w}}}$, respectively. More details for the multi-dimensional bisection search method can be found in [27], which are omitted here for brevity.

Proposition 2 (Convergence and Optimality of Proposed LogSumExp-Dual Scheme): Under the alternative optimization framework, $\bar{\mathbf{w}}$ and $\mathbf{p}_{\bar{\mathbf{w}}}$ converge to the optimal solutions.

Proof: Based on the proof of **Proposition 1**, the proof of **Theorem 1** in [20] and the tight approximation of LogSumExp [35], **Proposition 2** can be easily proven and hence it is omitted for simplicity. \blacksquare

B. Optimal Closed-Form Solution for γ_k

After optimizing $\bar{\mathbf{w}}$, we turn to the design of PS ratio γ_k . Note that γ_k is only dependent on the k -th MT's individual utility $R_{ID,k}$, and $R_{ID,k}$ increases as γ_k decreases. As such, the constraint $\tilde{C}3$ must hold with equality for all MTs at the optimal solution, which is given by

$$\sum_{i=1}^K |\hat{\mathbf{g}}_k^H \mathbf{w}_i|^2 = \frac{\bar{\gamma}_{\max}}{(1 - \gamma_k)}. \quad (27)$$

Thus, the optimal closed-form solution for γ_k is obtained as

$$\gamma_k = 1 - \frac{\bar{\gamma}_{\max}}{\sum_{i=1}^K |\hat{\mathbf{g}}_k^H \mathbf{w}_i|^2}. \quad (28)$$

C. Modified Cyclic Coordinate Descent Algorithm for Ξ_{ka}

In this section, we focus on optimizing the multi-layer RIS's coefficients Ξ_{ka} . Note that Ξ_{ka} only determines the k -th MT's individual achievable rate $R_{ID,k}$, thus the designs of Ξ_{ka} at each MT are done simultaneously, and formulated as

$$\max_{\Xi_{ka}} \min_{\Delta} \frac{|\mathbf{g}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k}^K |\mathbf{g}_k^H \mathbf{w}_i|^2 + \bar{\sigma}_k^2} \quad (29)$$

$$s.t. \quad \text{C2 : } \min_{\Delta} \sum_{i=1}^K |\mathbf{g}_k^H \mathbf{w}_i|^2 \geq \frac{\bar{\gamma}_{\max}}{(1 - \gamma_k)}, \quad \text{C4.}$$

Clearly, the above problem is challenging to be solved due to Δ and the non-convexity of both the objective function and constraints. Similar to Section III-A, the discretization method

is adopted. By selecting uniformly spaced angles as in (11), we can obtain the robust CSI, which is given by

$$\begin{aligned} \widehat{\mathbf{H}}_{(k,i),a} &= \sum_{p_1=1}^{N_{\text{H}1}N_{\text{E}1,k1}} \sum_{q_1=1}^{N_{\text{H}2}N_{\text{E}2,k1}} \sum_{p_2=1}^{N_{\text{H}1}N_{\text{E}1,k1}} \\ &\quad \sum_{q_2=1}^{N_{\text{H}2}N_{\text{E}2,k1}} \frac{1}{N_{\text{H}}N_{\text{E},k1}} \widehat{\mathbf{h}}_{(k,i),a}^{(p_1,q_1,p_2,q_2)} \widehat{\mathbf{h}}_{(k,i),a}^{(p_1,q_1,p_2,q_2),H}, \\ \mathbf{h}_{(k,i),a}^{(p_1,q_1,p_2,q_2),H} &= \\ \mathbf{v}_k^H \boldsymbol{\Omega}_{k,(a+1,A)} \mathbf{B}_{ka} \text{diag} \left\{ \boldsymbol{\Omega}_{k,(1,a-1)} \mathbf{G}_k^{(p_1,q_1,p_2,q_2)} \mathbf{w}_i \right\}. \end{aligned} \quad (30)$$

Next, after some mathematical manipulations, we can reformulate the subproblem w.r.t $\boldsymbol{\Xi}_{ka} = \text{diag}(\boldsymbol{\xi}_{ka})$ as

$$\max_{\boldsymbol{\xi}_{ka}} \frac{\boldsymbol{\xi}_{ka}^H \widehat{\mathbf{H}}_{(k,k),a} \boldsymbol{\xi}_{ka}}{\boldsymbol{\xi}_{ka}^H \mathbf{R}_{ka} \boldsymbol{\xi}_{ka} + \bar{\sigma}_k^2} \quad (31)$$

$$\text{s.t. } \overline{\text{C}2} : \boldsymbol{\xi}_{ka}^H \mathbf{Q}_{ka} \boldsymbol{\xi}_{ka} \geq \tilde{\gamma}_k, \overline{\text{C}4} : |\boldsymbol{\xi}_{ka}|_n = 1, \forall a, n,$$

where $\tilde{\gamma}_k = \bar{\gamma}_{\max}/(1 - \gamma_k)$, $\mathbf{R}_{ka} = \sum_{i \neq k} \widehat{\mathbf{H}}_{(k,i),a}$, and $\mathbf{Q}_{ka} = \sum_{i=1}^K \widehat{\mathbf{H}}_{(k,i),a}$. Then, to facilitate the derivation of the closed-form solution to (31), we add $\overline{\text{C}2}$ into the objective function of (31) with the non-negative Lagrange multiplier τ , i.e.,

$$\max_{\boldsymbol{\xi}_{ka}} \frac{\boldsymbol{\xi}_{ka}^H \widehat{\mathbf{H}}_{(k,k),a} \boldsymbol{\xi}_{ka}}{\boldsymbol{\xi}_{ka}^H \mathbf{R}_{ka} \boldsymbol{\xi}_{ka} + \bar{\sigma}_k^2} + \tau (\boldsymbol{\xi}_{ka}^H \mathbf{Q}_{ka} \boldsymbol{\xi}_{ka} - \tilde{\gamma}_k) \text{ s.t. } \overline{\text{C}4}. \quad (32)$$

However, problem (32) is still challenging to be handled due to the fractional form of the objective function. Hence, we utilize the Dinkelbach's method to transform (31) to an equivalent form [48], which is expressed as

$$\max_{\boldsymbol{\xi}_{ka}} f(\boldsymbol{\xi}_{ka}, \tau, \vartheta) = \boldsymbol{\xi}_{ka}^H \mathbf{T}_{ka} \boldsymbol{\xi}_{ka} \text{ s.t. } \overline{\text{C}4}, \quad (33)$$

where $\mathbf{T}_{ka} = \widehat{\mathbf{H}}_{(k,k),a} - \vartheta \mathbf{R}_{ka} + \tau \mathbf{Q}_{ka}$, and ϑ is the non-negative Dinkelbach's parameter to be optimized. We observe that problem (33) is NP-hard owing to the multiplicative optimization variables $\boldsymbol{\xi}_{ka}$, τ , and ϑ . Thus, we first propose an efficient algorithm specializing on the M-CCD optimization framework [49] to obtain the closed-form solutions of $\boldsymbol{\xi}_{ka}$ and ϑ , and then use the bisection method to search for τ .

As for optimizing $\boldsymbol{\xi}_{ka}$ and ϑ with given τ , we do not optimize overall \mathbf{p}_{Ξ} like the existing SDR and SCA methods, while we concatenate all the variables into one vector $[\vartheta; \boldsymbol{\xi}_{ka}]$ and then addresses $N_{\text{E},ka}+1$ scalar subproblems for the block size chosen. Here, we choose the block size as 1. Note that we only solve the subproblem w.r.t. $\boldsymbol{\xi}_{ka,i}$ at one iteration, while the remaining variables are updated subsequently such that the closed-form solutions can be obtained, which again results in the improvement of CCD implementation over the state of art.

To elaborate, we can first expand the objective function $f(\boldsymbol{\xi}_{ka}, \vartheta)$ as

$$\begin{aligned} \boldsymbol{\xi}_{ka}^H \mathbf{T}_{ka} \boldsymbol{\xi}_{ka} &= \sum_{j=1}^{N_{\text{E},ka}} \sum_{i=1}^{N_{\text{E},ka}} \boldsymbol{\xi}_{ka,i}^* T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j} \\ &= \sum_{i=1}^{N_{\text{E},ka}} \boldsymbol{\xi}_{ka,i}^* T_{ka,(i,i)} \boldsymbol{\xi}_{ka,i} + \sum_{j \neq i}^{N_{\text{E},ka}} \sum_{i=1}^{N_{\text{E},ka}} \boldsymbol{\xi}_{ka,i}^* T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j} \\ &= \sum_{i=1}^{N_{\text{E},ka}} T_{ka,(i,i)} + \Re \left(\sum_{i=1}^{N_{\text{E},ka}} \boldsymbol{\xi}_{ka,i}^* T_{ka,(i,i)} \right), \end{aligned} \quad (34)$$

where $T_{ka,(i)} = \sum_{j=1}^{j < i} T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j} + \sum_{j > i}^{N_{\text{E},ka}} T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j}$. Note that the third equality above holds due to the the unit-modulus property and the fact that \mathbf{T}_{ka} is a Hermitian matrix. Hence, we can obtain $N_{\text{E},ka}+1$ scalar subproblems and update

them by the following M-CCD algorithm.

$$\begin{cases} \vartheta^{(i_d+1)} = o \left(\vartheta, \boldsymbol{\xi}_{ka,1}^{(i_d)}, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d)} \right), \\ \boldsymbol{\xi}_{ka,1}^{(i_d+1)} = \arg \max_{\boldsymbol{\xi}_{ka,1} \in \mathcal{B}} f \left(\vartheta^{(i_d+1)}, \boldsymbol{\xi}_{ka,1}, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d)} \right), \\ \vdots \\ \vartheta_i^{(i_d+1)} = o \left(\vartheta, \dots, \boldsymbol{\xi}_{ka,i}^{(i_d)}, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d)} \right), \\ \boldsymbol{\xi}_{ka,i}^{(i_d+1)} = \arg \max_{\boldsymbol{\xi}_{ka,i} \in \mathcal{B}} f \left(\vartheta^{(i_d+1)}, \dots, \boldsymbol{\xi}_{ka,i}, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d)} \right), \\ \vdots \\ \vartheta_{N_{\text{E},ka}}^{(i_d+1)} = o \left(\vartheta, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d+1)}, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d)} \right), \\ \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d+1)} = \arg \max_{\boldsymbol{\xi}_{ka,N_{\text{E},ka}} \in \mathcal{B}} f \left(\vartheta^{(i_d+1)}, \boldsymbol{\xi}_{ka,1}^{(i_d+1)}, \dots, \boldsymbol{\xi}_{ka,N_{\text{E},ka}}^{(i_d)} \right). \end{cases} \quad (35)$$

Given the value of $\boldsymbol{\xi}_{ka}$, ϑ can be solved by the optimal closed-form solution [48]

$$\vartheta^{(i_d+1)} = o \left(\vartheta, \boldsymbol{\xi}_{ka}^{(i_d)} \right) = \frac{\boldsymbol{\xi}_{ka}^{(i_d),H} \widehat{\mathbf{H}}_{(k,k),a} \boldsymbol{\xi}_{ka}^{(i_d)}}{\boldsymbol{\xi}_{ka}^{(i_d),H} \mathbf{R}_{ka} \boldsymbol{\xi}_{ka}^{(i_d)} + \bar{\sigma}_k^2}. \quad (36)$$

Next, we handle the optimization problem w.r.t. the each component $\boldsymbol{\xi}_{ka,i}$ of $\boldsymbol{\xi}_{ka}$. By ignoring the constant terms in (34), the subproblem w.r.t. $\boldsymbol{\xi}_{ka,i}$ is given by

$$\begin{aligned} \max_{\boldsymbol{\xi}_{ka,i}} & \Re \left(\boldsymbol{\xi}_{ka,i}^* \left(\sum_{j=1}^{j < i} T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j}^{(i_d+1)} + \sum_{j>i}^{N_{\text{E},ka}} T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j}^{(i_d)} \right) \right) \\ & \text{s.t. } \overline{\text{C}4} : \boldsymbol{\xi}_{ka,i} \in \mathcal{B}, \forall i. \end{aligned} \quad (37)$$

It is evident that problem (37) admits the closed-form solution of (33), which is given by

$$\begin{aligned} \boldsymbol{\xi}_{ka,i} &= \exp \left\{ j \arg \left(\sum_{j=1}^{j < i} T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j}^{(i_d+1)} \right. \right. \\ &\quad \left. \left. + \sum_{j>i}^{N_{\text{E},ka}} T_{ka,(i,j)} \boldsymbol{\xi}_{ka,j}^{(i_d)} \right) \right\}, \forall i. \end{aligned} \quad (38)$$

After solving for $\boldsymbol{\xi}_{ka}$ and ϑ , we turn to optimizing the Lagrange multiplier τ . Here, we can find the optimal τ by using the following complementary slackness condition, i.e.,

$$\tau \mathcal{M}(\tau) = \tau (\boldsymbol{\xi}_{ka}^H(\tau) \mathbf{Q}_{ka} \boldsymbol{\xi}_{ka}(\tau) - \tilde{\gamma}_k) = 0. \quad (39)$$

If the constraint $\overline{\text{C}2}$ holds, we can calculate the optimal $\tau = 0$. Otherwise, we have to find τ by using the following equation

$$\boldsymbol{\xi}_{ka}^H(\tau) \mathbf{Q}_{ka} \boldsymbol{\xi}_{ka}(\tau) - \tilde{\gamma}_k = 0. \quad (40)$$

Since (40) is monotonically decreasing w.r.t. τ [20], the bisection search method can be adopted to find τ , which is omitted here for brevity. Finally, by using the proposed M-CCD algorithm for all the layers' coefficients at each MT, semi-closed-form solutions for $\boldsymbol{\Xi}_{ka}$ can be obtained.

Proposition 3 (Convergence and Optimality of M-CCD):

The sequence $\{\boldsymbol{\xi}_{ka}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d)}\}$ generated by M-CCD algorithm can converge to KKT-optimal solution of problem (29), which is the coordinatewise optimal solutions [50].

Proof: Please refer to Appendix C. ■

D. MMSE Decoder for \mathbf{v}_k

In this subsection, we investigate the optimization of the digital decoder \mathbf{v}_k at the k -th MT. According to [51], the linear minimum-mean-square-error (MMSE) detector can be adopted for \mathbf{v}_k as the optimal digital decoder, which can balance the interference and noise at the receiver, i.e.,

$$\mathbf{v}_k = \frac{\left(\sum_{i \neq k}^K \widehat{\mathbf{w}}_{(k,i)} \widehat{\mathbf{w}}_{(k,i)}^H + \bar{\sigma}_k^2 \mathbf{I}_{N_{\text{D},k}} \right)^{-1} \widehat{\mathbf{w}}_{(k,k)}}{\left\| \left(\sum_{i \neq k}^K \widehat{\mathbf{w}}_{(k,i)} \widehat{\mathbf{w}}_{(k,i)}^H + \bar{\sigma}_k^2 \mathbf{I}_{N_{\text{D},k}} \right)^{-1} \widehat{\mathbf{w}}_{(k,k)} \right\|^2}, \quad (41)$$

Algorithm 1: Robust Optimization for solving (9)

```

1 Initialize  $(\bar{\mathbf{w}}^{(0)}, \{\gamma_k^{(0)}\}, \{\Xi_{ka}^{(0)}\}, \{\mathbf{v}_k^{(0)}\})$  and
    $i_d = 1$ ;
2 repeat
3   Compute  $\hat{\mathbf{A}}_k^{(i_d-1)}$  and  $\hat{\mathbf{B}}_k^{(i_d-1)}$  by (15) and (16);
4   Determine  $\mathcal{G}^{(i_d)}(\lambda)$  and  $\mathcal{P}^{(i_d)}(\mu)$  by (24) and (25);
5   if  $\mathcal{G}^{(i_d)}(\lambda) \geq 0$  and  $\mathcal{P}^{(i_d)}(\mu) \geq 0$  then
6      $\lambda^{(i_d)} = \mu^{(i_d)} = 0$ ;
7     Set  $i_w = 1$  and empty all the  $r^{(i_w)}$ ;
8     while  $r^{(i_w+1)} - r^{(i_w)} \geq \epsilon_w$  do
9       Compute  $r^{(i_w)} = \sum_{k=1}^K R_{ID,k}(\bar{\mathbf{w}}^{(i_w)})$ ;
10      Update  $\mathbf{N}(\bar{\mathbf{w}}^{(i_w)})$  and  $\mathbf{M}(\bar{\mathbf{w}}^{(i_w)})$  by
           (21);
11      Update  $\bar{\mathbf{w}}^{(i_w)}$  by (23) and  $\bar{\mathbf{w}}^{(i_d)} = \bar{\mathbf{w}}^{(i_w)}$ ;
12    end
13  else
14    Multi-dimensional bisection search method
        is adopted to search for  $\lambda^{(i_d)}$  and  $\mu^{(i_d)}$ ;
15    Update  $\bar{\mathbf{w}}^{(i_d)}$  with procedure of steps 8-12;
16  end
17 end
18 Determine  $\gamma_k^{(i_d)}$  and  $\mathcal{M}^{(i_d)}(\tau)$  by (28) and (39);
19 for  $k = 1$  to  $K$  do
20   if  $\mathcal{M}^{(i_d)}(\tau) \geq 0$  then
21      $\tau^{(i_d)} = 0$ ;
22     for  $a = 1$  to  $A$  do
23       for  $i = 1$  to  $N_{E,ka}$  do
24         Update  $\vartheta^{(i_d)}, \xi_{ka,i}^{(i_d)}$  by (36) and (38);
25       end
26     end
27   else
28     Bisection method is adopted for  $\tau^{(i_d)}$ ;
29     Update  $\xi_{ka,i}^{(i_d)}$  as steps 22-26;
30   end
31   Update  $\mathbf{v}_k^{(i_d)}$  by using (41).
32 end
33 Set  $i_d = i_d + 1$ ;
34 until some stopping criterion is satisfied;
Output:  $(\mathbf{w}_k^*, \gamma_k^*, \Xi_{ka}^*, \mathbf{v}_k^*)$ .

```

where $\hat{\mathbf{w}}_{(k,i)} = \Omega_{k,(1,A)} \hat{\mathbf{G}}_k \mathbf{w}_i$.

E. Optimality, Convergence and Complexity Analysis

To better illustrate the overall algorithm, we summarize the scalable robust optimization framework in Algorithm 1. In this subsection, the convergence of the proposed scalable robust optimization framework is first presented. As shown in **Proposition 2** and **Proposition 3**, the proposed LogSumExp-dual scheme and M-CCD algorithm converge to the KKT-optimal solution which maximizes the objective value of problem (9) at each iteration. Hence, we have

$$\begin{aligned} & r\left(\mathbf{w}_k^{(i_d)}, \Xi_{ka}^{(i_d)}, \mathbf{v}_k^{(i_d)}, \gamma_k^{(i_d)}\right) \\ & \leq r\left(\mathbf{w}_k^{(i_d+1)}, \Xi_{ka}^{(i_d+1)}, \mathbf{v}_k^{(i_d+1)}, \gamma_k^{(i_d+1)}\right), \end{aligned} \quad (42)$$

where r denotes the objective value of problem (9). Note that the inequalities hold from the optimization of various variables. Besides, $\mathbf{w}_k, \Xi_{ka}, \mathbf{v}_k, \gamma_k$ are bounded by the conflicting constraints C1-C4, $r\left(\mathbf{w}_k^{(i_d)}, \Xi_{ka}^{(i_d)}, \mathbf{v}_k^{(i_d)}, \gamma_k^{(i_d)}\right)$ is guaranteed to converge to the optimal solution.

Next, we present the computational complexity and scalability of the proposed framework. For the complexity of the LogSumExp-dual scheme for \mathbf{w}_k , its computational complexity is dominated by the calculation of $\mathbf{M}^\dagger(\bar{\mathbf{w}})$. Since $\mathbf{M}^\dagger(\bar{\mathbf{w}})$ is the sum of $K N_H \times N_H$ submatrices, the inverse matrix $\mathbf{M}^\dagger(\bar{\mathbf{w}})$ can be computed by using the inverse of each submatrix. Besides, the complexity for obtaining the optimal λ and μ is $\log_2(\Delta p_1 \Delta p_2 / \varepsilon_p)$, where Δp_i is the gap between the lower and upper bounds of the bisection method and ε_p denotes the accuracy. As such, the complexity of the LogSumExp-dual scheme is given by $\mathcal{O}(1/3 \log_2(\Delta p_1 \Delta p_2 / \varepsilon_p) K N_H^3)$. For the complexity of M-CCD, each $\xi_{ka,i}$ is computed in $\mathcal{O}(N_{E,ka})$ operations, thus $\mathcal{O}(N_{E,ka}^2)$ operations are required to compute ξ_{ka} . Considering the complexity of the bisection method, A RIS layers, and K ESSs, the total complexity of M-CCD is expressed as $\mathcal{O}\left(\sum_{k=1}^K \sum_{a=1}^A \log_2(\Delta p_3 / \varepsilon_p) N_{E,ka} (N_{E,ka} + 1)\right)$. The optimization complexities of \mathbf{v}_k and γ_k are $\mathcal{O}\left(\sum_{k=1}^K 1/3 N_{D,k}^3\right)$ and $\mathcal{O}(K)$, respectively. Accordingly, the total complexity of the proposed algorithm is

$$\begin{aligned} \mathcal{O}_{\text{Prop.}} = & \mathcal{O}\left\{\max\left(1/3 \log_2(\Delta p_1 \Delta p_2 / \varepsilon_p) K N_H^3,\right.\right. \\ & \left.\sum_{k=1}^K \sum_{a=1}^A \log_2(\Delta p_3 / \varepsilon_p) N_{E,ka} (N_{E,ka} + 1),\right. \\ & \left.\left.K, \sum_{k=1}^K 1/3 N_{D,k}^3\right)\right\}. \end{aligned} \quad (43)$$

The widely-adopted schemes with satisfactory performance for optimizing \mathbf{w}_k and Ξ_{ka} are SCA in [52] and MM in [20]. Thus, next we compare the complexity of the proposed algorithm with that of the SCA-MM scheme. As for SCA, the complexity for solving \mathbf{w}_k is $\mathcal{O}(K N_H C_1^2 C_2)$, where $C_1 = K(N_H + L + 8)$ is the total number of variables, $C_2 = K(L + 10)$ is total number of constraints, and L is the accuracy as in [52]. Besides, the complexity for obtaining Ξ_{ka} is $\mathcal{O}\left(\sum_{k=1}^K \sum_{a=1}^A \log_2(\Delta p_3 / \varepsilon_p) N_{E,ka}^2\right)$. Note that the complexities for \mathbf{v}_k and γ_k are the same as that of the proposed algorithm. Thus, the total complexity of SCA-MM scheme is given by

$$\begin{aligned} \mathcal{O}_{\text{SCA-MM}} = & \mathcal{O}\left\{\max\left(K N_H C_1^2 C_2, \sum_{k=1}^K 1/3 N_{D,k}^3,\right.\right. \\ & \left.\left.K, \sum_{k=1}^K \sum_{a=1}^A \log_2(\Delta p_3 / \varepsilon_p) N_{E,ka}^2\right)\right\}. \end{aligned} \quad (44)$$

Compared to (43), the complexity of our proposed algorithm is strictly smaller than the SCA-MM scheme due to usage of SCA. Moreover, the computation process of SCA-MM is not transparent since it heavily relies on an off-the-shelf optimization toolbox, which makes it hard to implement the algorithm on the HAP. Thus, in addition to its lower complexity, the scalability of proposed algorithm is better than that of SCA-MM scheme.

IV. PERFORMANCE ANALYSIS

To further investigate the notable potential gain enabled by the proposed multi-layer RIS-receiver, this section provides its asymptotic performance analysis. To make the problem analytically tractable, we consider a simple SISO scenario as in [14], where a single-antenna HAP communicates with a single-antenna MT with the assistance of the proposed multi-layer RIS-receiver. Here, for a fair comparison with the asymptotic performance of the passive and active RIS-reflectors, as in [14], we assume Rayleigh-fading channels.

A. Asymptotic SNR for Proposed and Existing Architectures

For a SISO system with the RIS-aided receiver, we redefine the HAP-MT channel matrix and the RIS-Rx antennas' inter-layer channel matrix as $\mathbf{G}_k : \mathbf{f}_1$ and $\mathbf{B}_{kA} : \mathbf{h}_1$, respectively. In addition, for a SISO system with the RIS-reflector, we define the HAP-RIS channel matrix and the RIS-MT channel vector as \mathbf{f}_2 and \mathbf{h}_2 , respectively. Next, we present the following theorems for the asymptotic SNR obtained by the RIS-receiver and the RIS-reflector.

Theorem 1 (Asymptotic SNR for RIS-reflectors): Denoting $N_{\text{Tot}} = \sum_{a=1}^A N_{E,a}$, and assuming $\mathbf{f}_2 \sim \mathcal{CN}(\mathbf{0}_{N_{\text{Tot}}}, \zeta_{f_2}^2 \mathbf{I}_{N_{\text{Tot}}})$, $\mathbf{h}_2 \sim \mathcal{CN}(\mathbf{0}_{N_{\text{Tot}}}, \zeta_{h_2}^2 \mathbf{I}_{N_{\text{Tot}}})$, the asymptotic SNR ψ_{passive} with a passive RIS-reflector and ψ_{active} with a active RIS-reflector are given, respectively, by

$$\psi_{\text{passive}} \rightarrow \left(\hat{\rho} \sum_{a=1}^A N_{E,a} \right)^2 \frac{P_{\max} \pi^2 \zeta_{h_2}^2 \zeta_{f_2}^2}{16\sigma^2} \quad (45)$$

$$\psi_{\text{active}} \rightarrow \quad (46)$$

$$\left(\hat{\rho} \sum_{a=1}^A N_{E,a} \right) \frac{P_{1,\max} P_{2,\max} \pi^2 \zeta_{h_2}^2 \zeta_{f_2}^2}{16 \left(P_{2,\max} \sigma_A^2 \zeta_{h_2}^2 + \sigma^2 P_{1,\max} \zeta_{f_2}^2 + \sigma_A^2 \sigma^2 \right)},$$

where σ_A^2 is the dynamic noise power induced by the active RIS, $P_{\max} = P_{1,\max} + P_{2,\max}$, $P_{1,\max}$ is the maximum HAP transmit power for the system with an active RIS-reflector, and $P_{2,\max}$ denotes the maximum power at the active RIS-reflector.

Proof: Please refer to [14] and [19]. ■

Theorem 2 (Asymptotic SNR for RIS-receivers and fully-digital MIMO receiver): Assuming $\mathbf{f}_1 \sim \mathcal{CN}(\mathbf{0}, \zeta_{f_1}^2 \mathbf{I})$, the asymptotic SNR $\chi_{\text{Prop.}}$ with the proposed RIS-receiver, $\chi_{\text{sing.-pass.}}$ with a single-layer passive RIS-receiver, $\chi_{\text{sing.-act.}}$ with a single-layer active RIS-receiver, and $\chi_{\text{digital-act.}}$ with fully-digital MIMO receiver are given, respectively, by

$$\chi_{\text{Prop.}} \rightarrow$$

$$\left(\prod_{a=1}^A \sum_{i_a=1}^{N_{E,a}} |b_{a,(i_{a+1}, i_a)}| \right)^2 \frac{P_{\max} (4 - \pi)^2 \zeta_{f_1}^2}{4\sigma^2}, \quad (47)$$

$$\chi_{\text{sing.-pass.}} \rightarrow$$

$$\left(\sum_{a=1}^A \sum_{i_a=1}^{N_{E,a}} |b_{a,(i_{a+1}, i_a)}| \right)^2 \frac{P_{\max} (4 - \pi)^2 \zeta_{f_1}^2}{4\sigma^2}, \quad (48)$$

$$\begin{aligned} \chi_{\text{sing.-act.}} \rightarrow & \left(\sum_{a=1}^A \sum_{i_a=1}^{N_{E,a}} |b_{a,(i_{a+1}, i_a)}| \right)^2 \\ & \times \frac{P_{1,\max} P_{2,\max} (4 - \pi)^2 \zeta_{f_1}^2}{4 \left(P_{2,\max} \sigma_A^2 + \sigma^2 P_{1,\max} \zeta_{f_1}^2 + \sigma_A^2 \sigma^2 \right)}, \end{aligned} \quad (49)$$

$$\chi_{\text{digital-act.}} \rightarrow \left(\sum_{a=1}^A N_{E,a} \right) \frac{P_{\max} \zeta_{f_1}^2}{16\sigma^2}, \quad (50)$$

where $b_{a,(i_{a+1}, i_a)}$ is the near-field channel element. Since the extremely short distance between the adjacent vertical RIS layers, i.e., $2\lambda\sqrt{N_{E,a}}/\sqrt{\pi}$, the near-field channels model (5) is almost the same as $\mathbf{B}_a = \hat{\rho}\mathbf{1}$ [53]. As such, $\chi_{\text{Prop.}}$, $\chi_{\text{sing.-pass.}}$, $\chi_{\text{sing.-act.}}$, and $\chi_{\text{digital-act.}}$ approximately increase in the order of $\left(\prod_{a=1}^A \hat{\rho} N_{E,a} \right)^2$, $\left(\hat{\rho} \sum_{a=1}^A N_{E,a} \right)^2$, $\left(\hat{\rho} \sum_{a=1}^A N_{E,a} \right)$, and $\left(\sum_{a=1}^A N_{E,a} \right)$, respectively.

Proof: Please refer to Appendix D. ■

Theorem 3: Assuming $A = 2$ and defining the radiated signal of the n -th RIS unit as y_n , y_n 's phase shifts can be flexibly adjusted with $[-\pi, \pi]$, and its amplitude can be

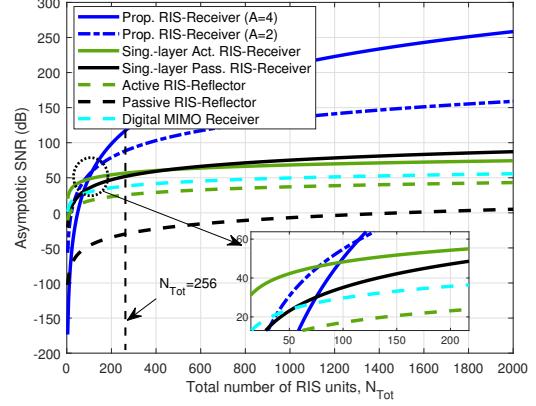


Fig. 3: Asymptotic SNR versus $N_{\text{Tot}} = \sum_{a=1}^A N_{E,a}$.

adjusted in the range $[0, y_{n,\max}]$, i.e., [54]

$$\begin{aligned} y_{n,\max}^2 = & \frac{1}{16\pi^2} \int_{-\frac{l_a\sqrt{N_E}}{2}}^{\frac{l_a\sqrt{N_E}}{2}} dp_x \int_{-\frac{l_a\sqrt{N_E}}{2}}^{\frac{l_a\sqrt{N_E}}{2}} dp_z \\ & l_1 l_2 (p_x^2 + l_1^2) [(p_x - x_n)^2 + l_2^2] \\ & \left\{ (p_x^2 + p_z^2 + l_1^2) [(p_x - x_n)^2 + (p_z - z_n)^2 + l_2^2] \right\}^{2.5}, \end{aligned} \quad (51)$$

where l_a is each RIS element's side-length, $(x_n, l_1 + l_2, z_n)$ is the the position of the n -th RIS unit on the second RIS layer, l_1 and l_2 are the distances from the feeds to the first RIS layer, and from the first layer to second layer, respectively.

Proof: See Appendix A in [30]. ■

Clearly, as compared to the single-layer RIS-receiver and typical MIMO receiver, the multi-layer RIS-receiver can adjust $y_{n,\max}$ within a longer range, which suggests that the RIS-aided receiver can generate more DoFs for enhancing SWIPT performance and facilitating beamforming design.

Next, we compare these asymptotic SNRs shown in **Theorem 1** and **Theorem 2** to reveal the superiority of our proposed multi-layer RIS-receiver in wireless communications.

B. Comparison Between Proposed and Existing Architectures

We can observe from **Theorem 1** and **Theorem 2** that, due to much shorter distance between RIS and Rx antennas, the RIS-receivers in **Theorem 2** would not experience the large fading of RIS-Rx link as compared with the RIS-reflectors in **Theorem 1**, i.e., $\zeta_{h_2}^2$, and thus the power attenuation is much smaller and the “double fading” effect can be well overcome by the RIS-receivers. Furthermore, compared to the asymptotic SNRs of the existing single-layer RIS-receiver and fully-digital MIMO receiver which are proportional to the function $\left(\hat{\rho} \sum_{a=1}^A N_{E,a} \right)^2$, $\left(\hat{\rho} \sum_{a=1}^A N_{E,a} \right)$, and $\left(\sum_{a=1}^A N_{E,a} \right)$, the asymptotic SNR of the proposed multi-layer RIS-receiver is proportional to $\left(\prod_{a=1}^A \hat{\rho} N_{E,a} \right)^2$, which suggests a notable enhanced aperture gain achieved by our proposed architectures for long-distance SWIPT. Thus, when $N_{E,a}$ and A are larger than a small value, the proposed multi-layer RIS-receiver can outperform the existing architectures. To illustrate this claim, we consider a SISO system with the proposed multi-layer RIS-receiver and the single-layer RIS-receiver.

Next, a specific setup for HAP communications has been adopted to compare the MT's SNRs in the above five systems, which is shown in Fig. 3. Here, we set the total power

TABLE I: Main Simulation Parameters

| Parameter | Values |
|--|--|
| Number of HAP's transmit antennas, N_H | $N_H = 8 \times 8$ |
| Number of MTs, K | $K = 3$ |
| Number of RIS layers, A | $A = 2$ |
| Number of MT's Rx antennas, $N_{D,k}$ | $N_{D,k} = 2 \times 2$ |
| Noise power at MTs, σ_k^2 | $\sigma_k^2 = -80 \text{ dBm}$ |
| Noise power at EH circuits, δ_k^2 | $\delta_k^2 = -80 \text{ dBm}$ |
| Carrier frequency, δ_k^2 | 18 GHz |
| Maximum power at HAP, P_{\max} | $P_{\max} = 40 \text{ dBm}$ |
| Minimum rate threshold, Γ_k | $\Gamma_k = 1 \text{ bps/Hz}$ |
| Harvested energy requirement, ς_{\max} | $\varsigma_{\max} = -20 \text{ dBm}$ |
| Nonlinear EH model parameters | $\alpha = 2.463, \beta = 1.635, \varepsilon = 0.826$ |
| Smoothing parameter, ρ | $\rho = 0.1$ |
| Bandwidth, W | $W = 1 \text{ MHz}$ |

consumption $P_{\max} = 2 \text{ W}$, $P_{1,\max} = P_{2,\max} = 1 \text{ W}$, and $\sigma^2 = \sigma_A^2 = -80 \text{ dBm}$ for a fair comparison. In addition, due to the long distance between HAP and RIS, we set $\varsigma_{f_1}^2 = \varsigma_{f_2}^2 = -90 \text{ dB}$, while $\varsigma_{h_2}^2 = -40 \text{ dB}$ for the short-distance RIS-MT link. Clearly, for a more practical condition $N_{\text{Tot}} = \sum_{a=1}^A N_{E,a} = 256$, the SNR achieved by the RIS-receivers is at least 26.33 dB higher than that achieved by the RIS-reflectors, which confirms the claim that the “double fading” effect can be well overcome by the RIS-receivers. Besides, the required number of RIS units at each layer N_E for the proposed multi-layer RIS-receiver to outperform the single-layer one is less than 36, which also decreases with A .

Thus, due to the notable potential gain $\left(\prod_{a=1}^A \hat{\rho} N_{E,a}\right)^2$, the proposed multi-layer RIS-receiver can achieve better performance than the single-layer one even when N_E and A are not large, which is consistent with **Theorem 2**.

Furthermore, as shown in Fig. 3, since the single-layer active RIS-receiver introduces an additional dynamic noise power σ_A^2 at the denominator of (48) and it would only experience the attenuation effects induced by $\hat{\rho}$ instead of the large fading of the RIS-Rx link, the single-layer active RIS-receiver outperforms the single-layer passive one only when $N_{\text{Tot}} = \sum_{a=1}^A N_{E,a} < 468$. When $N_{\text{Tot}} > 468$, the SNRs of both single-layer RIS-receiver architectures (passive and active) are approximately the same, which means that the amplitude gain achieved by the active RIS is nullified by the introduced dynamic noise. However, considering the fact that the amplitude of the signal penetrating multi-layer architecture can be also partially controlled in [31], our proposed multi-layer RIS-receiver not only can significantly amplify the desired signal power, but also does not introduce additional dynamic noise, which further confirms the superiority of our proposed multi-layer RIS-receiver architecture.

V. SIMULATION RESULTS

In this section, we provide numerical results to demonstrate the superiority of our proposed architecture and algorithms. The main parameters are listed in Table I⁶. The HAP channel

⁶The minimum value for harvesting energy guarantees the self-sustainability for MTs. According to [20], [21], [23], [55], the minimum EH thresholds in both the terrestrial and non-terrestrial networks range from -15 dBm to -25 dBm. Thus, this paper sets the minimum value for harvesting energy as -20 dBm as [21] and [55].

model parameters in (6) are the same as those in [4]. Moreover, the CSI uncertainty bound is defined as $\Delta = \theta_U - \theta_L$, and the HAP's altitude is D (meters). HAP is deployed at $[0, 0, D]$ m and three MTs are distributed at the directions $\{(\theta, \varphi) | (30^\circ, 30^\circ), (60^\circ, 150^\circ), (70^\circ, 75^\circ)\}$ of HAP with a radius of D m. Here, we compare the following architectures and algorithms:

- **Proposed architecture:** the proposed multi-layer refracting RIS-assisted receiver architecture with A layers having $N_{E,ka} = N_E$ on each layer is adopted for SWIPT, and the proposed scalable robust beamforming optimization framework is utilized to solve problem (9).
- **Digital architecture:** the fully-digital receiver architecture in [27] with $N = AN_E$ Rx antennas is adopted, and the MMSE decoder [51] is used to design the decoder.
- **Single-layer active architecture:** the single-layer active RIS-aided receiver architecture $N = AN_E$ RIS units is adopted, and the proposed algorithm is applied to obtain Ξ_k . To further highlight the advantages of our proposed architecture, we consider a favorable setting for the single-layer active one, where the maximum amplification factor is set as a constant, while ignoring active RIS's power constraints.
- **Single-layer passive architecture:** the single-layer passive RIS-aided receiver architecture having $N = AN_E$ RIS units is adopted, and the proposed algorithm is applied to obtain Ξ_k .
- **SCA-MM scheme:** under the proposed architecture, SCA in [52] and MM in [20] are utilized to handle \mathbf{w}_k and Ξ_{ka} in problem (9), respectively.
- **ZF-MCCD scheme:** under the proposed architecture, zero-forcing (ZF) in [4] and M-CCD in Section III-C are utilized to handle \mathbf{w}_k and Ξ_{ka} in problem (9).

Fig. 4 presents the transmit (Tx) and receive (Rx) beampattern with different architectures, where $N_E = 6 \times 6$, $\Delta = 2^\circ$, and $D = 20$ km. Here, we focus on investigating the Tx and Rx beampatterns of MT 1. As expected, even under the angular uncertainty Δ , the HAP's transmitter can still accurately generate the mainlobe towards the desired MT with SNR of 0 dB, and simultaneously align the nulls with at least -50 dB depth towards undesired targets, indicating the viability of the transmit precoder generated by our proposed LogSumExp-dual scheme. Besides, since the total RIS units are evenly divided at A layers, the lobe resolution of the proposed architecture is lower than those of both the single-layer active and the

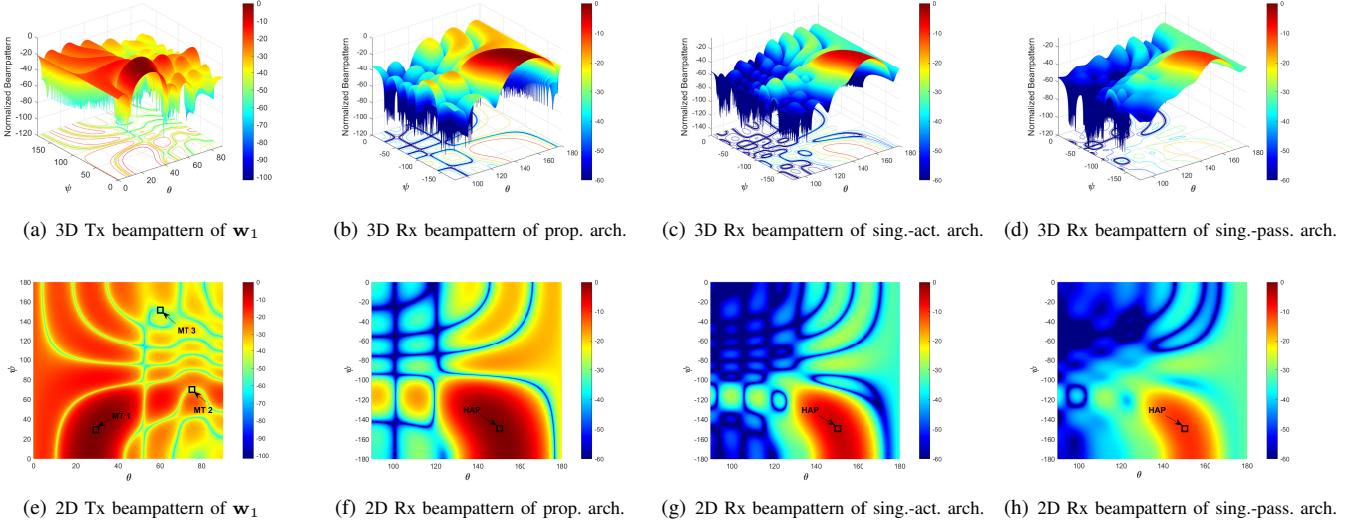


Fig. 4: Transmit and receive beampattern with different architectures (colorbar is unified for the last 3 subfigures, unit: dB).

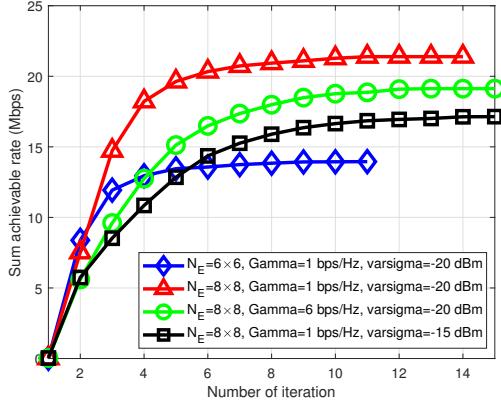


Fig. 5: Convergence performance.

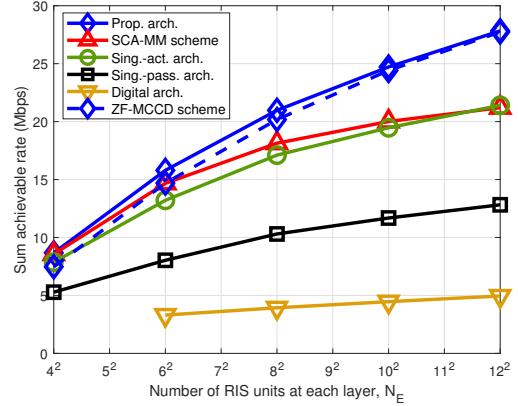


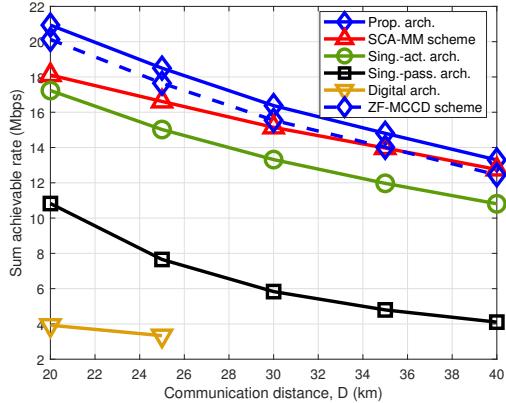
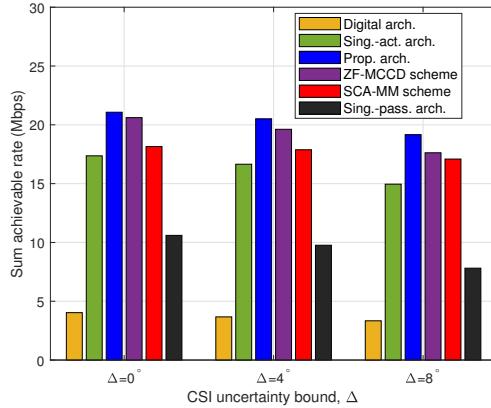
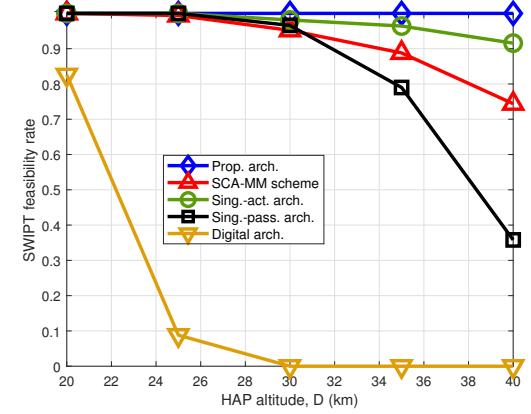
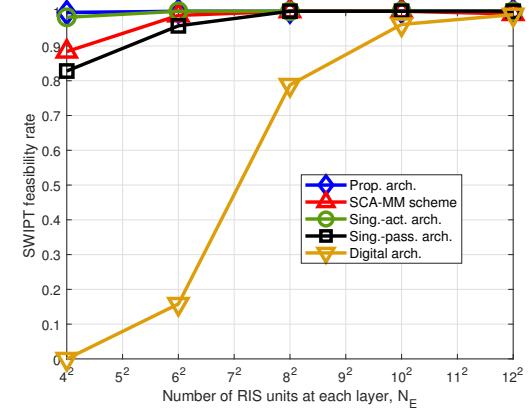
Fig. 6: Sum achievable rate versus N_E .

passive architectures with the same number of RIS units. However, the mainlobes of our proposed architecture can be still guaranteed to point to the desired MT with Δ . In addition, it is worth noting that the received MT's SNR achieved by the proposed architecture is about 0 dB, while those of the single-layer active and the passive architectures are about -4.2 dB and -10 dB, respectively. This phenomenon suggests that the proposed multi-layer RIS-assisted architecture can amplify the amplitude of the desired signals much more than both the single-layer active and passive ones, which is consistent with **Theorem 2**. Furthermore, due to the fact that only the phase can be adapted in the single-layer passive RIS-receiver such that the DoF for beamforming design is limited, its corresponding beampattern is not smooth. Conversely, the beampatterns of the proposed and the single-layer active architectures are both smooth, which suggests that the amplitude of the signal penetrating multi-layer architecture can be also partially controlled as in [31]. Thus, an additional DoF is introduced by proposed architecture for beamforming design.

Fig. 5 illustrates the convergence of our proposed optimization algorithm. It is seen that all algorithms converge rapidly within 15 iterations, and larger N_E or lower Γ_k, ζ_{\max} achieve

a higher sum rate at the cost of slower convergence speed.

Fig. 6 illustrates the sum achievable rate versus the number of RIS units at each layer $\sqrt{N_E}$. As we can see, the sum achievable rate increases with increasing N_E , especially when the number of RIS units at each layer increases from 4×4 to 8×8 . Furthermore, it can also be observed that, as N_E grows, the sum achievable rate of all the benchmark architectures and schemes would converge gradually, and the difference in sum achievable rate between the proposed architecture and the benchmarks becomes larger. This behavior can be explained by the following two reasons. First, as shown in **Theorem 2**, although the desired signals' power of the benchmark architectures are proportional to the function of $(\hat{\rho} \sum_{a=1}^A N_{E,a})$, employing more RIS units can also simultaneously boost the inter-user interference strength, thus their sum achievable rate would converge gradually. However, benefiting from the notable potential gain achieved by the proposed architecture, i.e., $(\prod_{a=1}^A \hat{\rho} N_{E,a})^2$, the desired signal is significantly enhanced as N_E increases, which is far greater than the inter-user interference. Thus, the gap in sum achievable rate between the architectures increases with N_E . Second, in the SCA-MM

Fig. 7: Sum achievable rate versus D .Fig. 8: Sum achievable rate versus Δ .Fig. 9: SWIPT feasibility versus D .Fig. 10: SWIPT feasibility versus N_E .

scheme, the tightness between the original problem and the solved ones loosens as N_E grows, such that the difference between SCA-MM and proposed algorithm becomes larger. In contrast, although the ZF precoder leads to higher inter-user interference strength than the proposed algorithm, the interference can be further mitigated as N_E grows, thus eventually decreasing the performance gap between ZF-MCCD and our proposed algorithm. These results confirm both the superiority and the scalability of the proposed architecture and optimization framework as the number of RIS units increases. On the other hand, due to the severe large-scale fading in the HAP-MT link, the digital receiver architecture cannot realize SWIPT for a small N_E , namely, $N_E < 6 \times 6$, while we see the converse for RIS-aided HAP communications. In the practice, it is challenging to deploy $N_{\text{Tot}} > 36$ digital antennas at the MT side, whereas RIS can facilitate the employment of large-scale arrays, which indicates that HAP-SWIPT can be realized with the assistance of RIS-receivers.

To further reveal the benefits of our proposed multi-layer RIS-aided receiver in the HAP communications, the sum achievable rate versus the HAP's altitude D in Fig. 7. Here, the other settings are the same as in Figs. 4 and 6. Note that the higher HAP's altitude implies broader coverage and higher maneuverability, which is an essential factor for HAP communications [3]. We can find that the sum achievable rate decreases with D due to the higher large-scale fading, and the proposed RIS-receiver can still achieve superior performance over the other receivers and SCA-MM/ZF-MCCD as D grows. Besides, the conventional digital receiver cannot communicate

with HAP when $D > 25$ km, while the RIS-aided receiver can still enable SWIPT even when $D > 40$ km, suggesting that the RIS-receiver can significantly enhance the coverage and maneuverability of HAP such that the MT's quality of service can be improved, which is consistent with **Theorem 2**. It can be also observed that the single-layer passive RIS-receiver is more sensitive to the HAP's altitude D than both the proposed multi-layer RIS-receiver and the single-layer active one. This is because the amplitudes of proposed multi-layer RIS-receiver and the single-layer active one can be also partially or fully adapted, and thus more DoFs are generated to reduce the negative impact of D , which is consistent with **Theorem 3**.

Fig. 8 shows the sum achievable rate versus the HAP's altitude D in Fig. 7. Here, the other settings are the same as in Figs. 4 and 6. Note that the higher HAP's altitude implies broader coverage and higher maneuverability, which is an essential factor for HAP communications [3]. We can find that the sum achievable rate decreases with D due to the higher large-scale fading, and the proposed RIS-receiver can still achieve superior performance over the other receivers and SCA-MM/ZF-MCCD as D grows. Besides, the conventional digital receiver cannot communicate

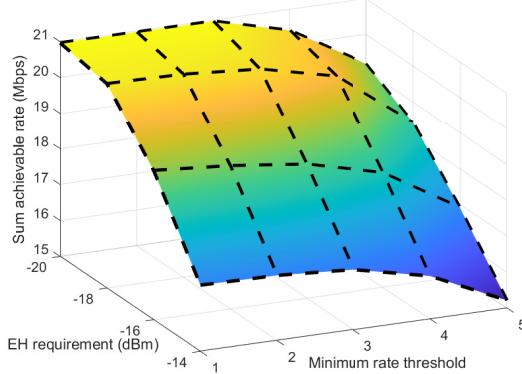


Fig. 11: Sum achievable rate versus Γ_k and ς_{\max} .

SWIPT in SCA-MM, such that only high sum achievable rate is counted, thereby leading to more stable performance. In addition, since the inter-user interference strength of the ZF precoder increases as the uncertainty region Δ , the ZF-MCCD scheme is more sensitive to Δ than both the SCA-MM scheme and our proposed algorithm. In summary, compared to the benchmarks, both our proposed multi-layer RIS-receiver and robust optimization framework can achieve the superior and stable performance under the CSI uncertainty.

Figs. 9 and 10 depict the SWIPT feasibility rate F versus D and N_E , when $N_E = 6 \times 6$ and $D = 20$ km, respectively. We observe that the SWIPT feasibility rate decreases with D , and increases with N_E . Similar to the findings in Fig. 4 and 6, the digital receiver achieves the worst performance, which further confirms the hardship of the conventional HAP-SWIPT and the positive impacts of RIS-receivers on HAP-SWIPT. Besides, due to the capability of amplitude regulation, the SWIPT feasibility rates of both the proposed and the single-layer active RIS are not sensitive to D and N_E , unlike that of single-layer passive one, especially with D . Furthermore, we find that the gap in the SWIPT feasibility rate between SCA-MM and our proposed optimization framework increases with increasing D and decreasing N_E . This is because of the loose approximations used in SCA-MM. These results further confirm the superiority and validity of our proposed optimization framework.

Fig. 11 shows the sum achievable rate versus the minimum threshold Γ_k and harvested energy requirement ς_{\max} . It can be observed that as ς_{\max} increases, the sum achievable rate significantly decreases. This is because the stringent EH requirement forces the optimization algorithm to allocate more resources for energy harvesting. On the other hand, the sum achievable rate remains stable when $\Gamma_k \leq 3$ bps/Hz, while it shows significant variation when $\Gamma_k \geq 4$ bps/Hz. This is due to the fact that as Γ_k increases, more resources should be allocated to ensure that each MT's rate is higher than Γ_k , which inevitably leads to a lower sum-rate.

VI. CONCLUSIONS

This paper investigated HAP-SWIPT networks with our proposed multi-layer refracting RIS-assisted receiver to overcome the severe large-scale fading and the energy scarcity dilemma. Utilizing the proposed architecture and taking the angular CSI imperfection into account, a worst-case sum rate maximization problem was formulated to achieve highly-efficient simultaneous transmission of information and energy.

To handle the intractable non-convex problem, a scalable robust optimization framework was developed by leveraging the discretization method, LogSumExp-dual scheme, and the M-CCD, which admits the semi-closed-form solutions of all optimization variables. Moreover, the asymptotic performance of our proposed RIS-receiver was derived to show the notable gain they achieve in HAP communications. The theoretical and simulation results showed that the multi-layer RIS-receiver can well overcome the severe “double fading” effect induced by the extreme long-distance HAP links and fully exploit RIS’s degrees-of-freedom (DoFs) for the SWIPT design. To elaborate, both the phase and amplitude of the signal processed by the multi-layer RIS-receiver can be adapted in a wide range, such that HAP-SWIPT networks assisted by the multi-layer RIS-receiver can achieve much better performance than those with the single-layer active/pассив RIS-receiver and the conventional digital receiver. Furthermore, the proposed algorithm achieves excellent performance with significantly lower complexity compared to the existing ones.

APPENDIX A PROOF OF LEMMA 1

Recalling $\mathcal{L}(\bar{\mathbf{w}}) = \mathcal{L}_1(\bar{\mathbf{w}}) + \mathcal{L}_2(\bar{\mathbf{w}}) + \mathcal{L}_3(\bar{\mathbf{w}})$ in (18), we first take the derivative of (18) w.r.t. $\bar{\mathbf{w}}$. As such, the first-order KKT condition holds if

$$\begin{aligned} \partial_{\bar{\mathbf{w}}} \mathcal{L}_1(\bar{\mathbf{w}}) + \partial_{\bar{\mathbf{w}}} \mathcal{L}_2(\bar{\mathbf{w}}) + \partial_{\bar{\mathbf{w}}} \mathcal{L}_3(\bar{\mathbf{w}}) &= 0 \\ \Rightarrow \mathbf{M}^\dagger(\bar{\mathbf{w}}) \mathbf{N}(\bar{\mathbf{w}}) \bar{\mathbf{w}} &= \mathcal{L}(\bar{\mathbf{w}}) \bar{\mathbf{w}}. \end{aligned} \quad (\text{A.1})$$

This completes the proof. ■

APPENDIX B PROOF OF PROPOSITION 1

Note that $\bar{\mathbf{w}}$ satisfying (20) is a stationary solution of (17) whose gradient is zero since (20) is the first-order optimality condition. On the other hand, we observe that the condition (20) is a non-linear eigenvalue problem, i.e., eigenvector-dependent non-linear eigenvalue problem (NEPv) [56], where $\bar{\mathbf{w}}$ is an eigenvector of $\mathbf{M}^\dagger(\bar{\mathbf{w}}) \mathbf{N}(\bar{\mathbf{w}})$ corresponding to the eigenvalue $\mathcal{L}(\bar{\mathbf{w}})$. Furthermore, λ_{\max} is equivalent to $\mathcal{L}(\bar{\mathbf{w}})$ in (18). As such, the leading eigenvector of $\mathbf{M}^\dagger(\bar{\mathbf{w}}) \mathbf{N}(\bar{\mathbf{w}})$ can be obtained as the optimal solution of (17), which maximizes the objective function among multiple eigenvectors. Hence, the proof to (22) is completed.

Next, we prove (23). Due to the basis's properties, we can define $\bar{\mathbf{w}}^{(0)} = \sum_{i=1}^N \kappa_i \mathbf{q}_i$, where \mathbf{q}_i is the i -th eigenvector and κ_i is the weight factor. Denoting $\bar{\mathbf{M}} = \mathbf{M}^\dagger(\bar{\mathbf{w}}) \mathbf{N}(\bar{\mathbf{w}})$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$, we have

$$\begin{aligned} \mathcal{D}(\mathbf{q}) &= \bar{\mathbf{M}} \bar{\mathbf{w}}^{(i_d)} = \sum_{i=1}^N \kappa_i \lambda_i^{(i_d)} \mathbf{q}_i \\ &= \kappa_1 \lambda_1^{(i_d)} \left(\mathbf{q}_1 + \underbrace{\sum_{i=2}^N \frac{\kappa_i}{\kappa_1} \left(\frac{\lambda_i}{\lambda_1} \right)^{(i_d)} \mathbf{q}_i}_{(a)} \right), \end{aligned} \quad (\text{B.1})$$

due to the fact that $\bar{\mathbf{M}} \mathbf{q}_i = \lambda_i \mathbf{q}_i$. Then, we prove that (a) will vanish as $i_d \rightarrow \infty$. For an arbitrary vector \mathbf{q} , the Taylor expansion of $\mathcal{D}(\mathbf{q})$ at \mathbf{q}_1 leads to

$$\begin{aligned} \mathcal{D}^H(\mathbf{q}) \mathbf{q}_i &= \mathcal{D}^H(\mathbf{q}_1) \mathbf{q}_i \\ &\quad + (\mathbf{q} - \mathbf{q}_1)^H \partial_{\mathbf{q}_1} \mathcal{D}(\mathbf{q}_1) \mathbf{q}_i + o(\|\mathbf{q} - \mathbf{q}_1\|). \end{aligned} \quad (\text{B.2})$$

As such, we have

$$(\mathcal{D}^H(\mathbf{q})\mathbf{q}_1)^2 = (\lambda_1 + o(\|\mathbf{q} - \mathbf{q}_1\|))^2, \quad (\text{B.3})$$

$$\sum_{i=2}^N (\mathcal{D}^H(\mathbf{q})\mathbf{q}_i)^2 \quad (\text{B.4})$$

$$\leq \sum_{i=2}^N \left(\lambda_i (\mathbf{q}^H \mathbf{q}_i)^2 + 2\lambda_i (\mathbf{q}^H \mathbf{q}_i) o(\|\mathbf{q} - \mathbf{q}_1\|) + o^2(\|\mathbf{q} - \mathbf{q}_1\|) \right) \\ \leq (\lambda_2 \|\mathbf{q} - \mathbf{q}_1\| + o(\|\mathbf{q} - \mathbf{q}_1\|))^2.$$

Since the premise $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$, by iteratively projecting \mathbf{q} onto $\mathcal{D}(\mathbf{q})$ with (23), (a) vanishes. Hence, (23) converges to the optimal leading eigenvector $\bar{\mathbf{w}}^*$. ■

APPENDIX C PROOF OF PROPOSITION 3

First, we prove the convergence of the proposed M-CCD algorithm. The ascent of the objective function (33) can be expressed as

$$\begin{aligned} & f\left(\xi_{ka,1}^{(i_d)}, \xi_{ka,2}^{(i_d)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d)}\right) \quad (\text{C.1}) \\ & \stackrel{a}{\leq} f\left(\xi_{ka,1}^{(i_d)}, \xi_{ka,2}^{(i_d)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d+1)}\right) \\ & \stackrel{b}{\leq} f\left(\xi_{ka,1}^{(i_d+1)}, \xi_{ka,2}^{(i_d)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d+1)}\right) \\ & \stackrel{c}{\leq} f\left(\xi_{ka,1}^{(i_d+1)}, \xi_{ka,2}^{(i_d)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d)}, \vartheta^{(i_d+1)}, \tau^{(i_d+1)}\right) \\ & \stackrel{b}{\leq} f\left(\xi_{ka,1}^{(i_d+1)}, \xi_{ka,2}^{(i_d+1)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d)}, \vartheta^{(i_d+1)}, \tau^{(i_d+1)}\right) \\ & \stackrel{c}{\leq} f\left(\xi_{ka,1}^{(i_d+1)}, \xi_{ka,2}^{(i_d+1)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d)}, \vartheta^{(i_d+2)}, \tau^{(i_d+1)}\right) \\ & \quad \vdots \\ & \stackrel{b}{\leq} f\left(\xi_{ka,1}^{(i_d+1)}, \xi_{ka,2}^{(i_d+1)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d+1)}, \vartheta^{(i_d+N_{E,ka}-1)}, \tau^{(i_d+1)}\right) \\ & \stackrel{c}{\leq} f\left(\xi_{ka,1}^{(i_d+1)}, \xi_{ka,2}^{(i_d+1)}, \dots, \xi_{ka,N_{E,ka}}^{(i_d+1)}, \vartheta^{(i_d+N_{E,ka})}, \tau^{(i_d+1)}\right). \end{aligned}$$

The inequality a holds since the duality gap approaches zero, i.e., the optimality of $\min_{\tau>0} \max_{\xi_{ka}} f(\xi_{ka}, \vartheta, \tau)$ is achieved by searching the optimal τ via the bisection search method, inequality c holds from the update of τ [48], and inequality b holds from the fact that $\xi_{ka,i}^{(i_d)}$ is a unique element-wise maximizer of $f(\xi_{ka}, \vartheta, \tau)$, which will be given next.

To further prove the convergence of the M-CCD algorithm, we discuss the uniqueness of the element-wise maximizer of subproblem w.r.t. $\xi_{ka,i}$. Recalling the solution (38) of problem (37), namely, $\xi_{ka,i} = \exp\left\{j \arg\left(\tilde{T}_{ka,(i)}\right)\right\}$, where $\tilde{T}_{ka,(i)} = \sum_{j=1}^{j< i} T_{ka,(i,j)} \xi_{ka,j}^{(i_d+1)} + \sum_{j>i}^{N_{E,ka}} T_{ka,(i,j)} \xi_{ka,j}^{(i_d)}$, we can show that (38) is the unique element-wise maximizer by contradiction. Denoting $\bar{\xi}_{ka,i} = \exp\left\{j \arg\left(\tilde{T}_{ka,(i)} + \theta_i\right)\right\}$ as another element-wise maximizer, where θ_i is not an integer multiple of 2π , we have

$$f(\xi_{ka}, \vartheta, \tau) = \sum_{i=1}^{N_{E,ka}} T_{ka,(i,i)} + \sum_{i=1}^{N_{E,ka}} |T_{ka,(i)}| \quad (\text{C.2})$$

$$f(\bar{\xi}_{ka}, \vartheta, \tau) = \sum_{i=1}^{N_{E,ka}} T_{ka,(i,i)} + \sum_{i=1}^{N_{E,ka}} |T_{ka,(i)}| \cos(\theta_i).$$

Since both ξ_{ka} and $\bar{\xi}_{ka}$ maximize $f(\xi_{ka}, \vartheta, \tau)$, we obtain $f(\xi_{ka}, \vartheta, \tau) = f(\bar{\xi}_{ka}, \vartheta, \tau)$ so that θ_i must be an integer multiple of 2π , which contradicts our assumption. Thus, (38) is the unique element-wise maximizer of $f(\xi_{ka}, \vartheta, \tau)$, such that inequality b holds, which implies that the sequence $\{\xi_{ka}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d)}\}$ is monotonically increasing. Furthermore, the objective function $\{\xi_{ka}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d)}\}$ is upper-bounded

by the constraint $\bar{C}4$. Hence, $\{\xi_{ka}^{(i_d)}, \vartheta^{(i_d)}, \tau^{(i_d)}\}$ increases to the limited point $\{\xi_{ka}^{(\infty)}, \vartheta^{(\infty)}, \tau^{(\infty)}\}$.

Next, we show that the limited point is KKT-optimal point of problem (29). Since the objective function of problem (33) is strictly convex, the bisection method can find the unique solution satisfying the complementary slackness condition (39). Then, we consider two cases: 1) $\tau^* = 0$; 2) $\tau^* > 0$.

In the first case, $\bar{C}4$ is not tight in the optimum. Thus, the optimal solution is equal to that with $\tau = 0$. Based on the fact that $\xi_{ka}^{(\infty)}$ is a block-wise maximizer of f w.r.t. block ξ_{ka} , the solution $\xi_{ka}^{(\infty)}(0)$ satisfies the partial KKT conditions of problem (33) w.r.t. $\xi_{ka,i}^{(\infty)}$. For the second case, the equality (40) should hold. Then, we use the method of contradiction to prove the KKT optimum of the second case. First, denote $\xi_{ka}^{(\infty)}(\tau^*)$ as the KKT solution corresponding to τ^* , and thus we have

$$\xi_{ka}^{(\infty),H}(\tau^*) \mathbf{Q}_{ka} \xi_{ka}^{(\infty)}(\tau^*) - \tilde{\gamma}_k = 0, \quad (\text{C.3})$$

Assume $\xi_{ka}^{(\infty)}(\tau^*)$ is not the globally optimal solution of problem (33), then we obtain

$$f\left(\xi_{ka}^{(\infty)}(\tau^*)\right) \leq f\left(\xi_{ka}^{*,(\infty)}\right). \quad (\text{C.4})$$

Since τ^* is the optimal Lagrange multiplier, $\xi_{ka}^{(\infty)}(\tau^*)$ is the KKT solution of (33) when $\tau = \tau^*$. Then, we have

$$\begin{aligned} & f\left(\xi_{ka}^{(\infty)}(\tau^*)\right) + \tau^* \xi_{ka}^{(\infty),H}(\tau^*) \mathbf{Q}_{ka} \xi_{ka}^{(\infty)}(\tau^*) \\ & \geq f\left(\xi_{ka}^{*,(\infty)}\right) + \tau^* \xi_{ka}^{*,(\infty),H} \mathbf{Q}_{ka} \xi_{ka}^{*,(\infty)}. \end{aligned} \quad (\text{C.5})$$

By substituting (40) and (C.3) into (C.5), we have $f\left(\xi_{ka}^{(\infty)}(\tau^*)\right) \geq f\left(\xi_{ka}^{*,(\infty)}\right)$, which contradicts (C.4). Thus, the optimum of the second case is the KKT solution. Finally, due to the fact that the duality gap between (33) and (29) approaches zero, problem (37) is equivalent to problem (29). Thus, the KKT solution $\{\xi_{ka}^{(\infty)}, \vartheta^{(\infty)}, \tau^{(\infty)}\}$ is also the KKT-optimal point of problem (29). Besides, according to Theorem 1 in [50], the KKT solutions of M-CCD are equivalent to the coordinatewise optimal solutions. Thus, the proposed M-CCD converges to the coordinatewise optimal solutions. ■

APPENDIX D PROOF OF THEOREM 2

Denoting $\mathbf{w}_k = \omega$, the SNR $\chi_{\text{Prop.}}$ maximization problem can be formulated as

$$\max_{\omega, \Xi_a} \chi_{\text{Prop.}} = \frac{|\mathbf{h}_1^H \Xi_A \Omega_{(1,A-1)} \mathbf{f}_1 \omega|^2}{\sigma^2} \quad (\text{D.1})$$

$$\text{s.t. } \text{C3: } |\omega|^2 \leq P_{\max}, \text{C4: } |\Xi_a|_{n,n} = 1, \forall a, n.$$

Then, by expanding the numerator of $\chi_{\text{Prop.}}$, we have

$$\mathbf{h}_1^H \Xi_A \Omega_{(1,A-1)} \mathbf{f}_1 \quad (\text{D.2})$$

$$= \sum_{i_A=1}^{N_{E,A}} h_{1,i_A} e^{j\theta_{A,i_A}} \prod_{a=1}^{A-1} \left(\sum_{i_a=1}^{N_{E,a}} b_{a,(i_{a+1},i_a)} e^{j\theta_{a,i_a}} \right) f_{1,i_1}$$

The optimal solution of problem (D.1) is given by

$$\omega^* = \sqrt{P_{\max}}, \theta_{1,i_1}^* = -\angle b_{1,(i_2,i_1)} - \angle f_{1,i_1} \quad (\text{D.3})$$

$$\theta_{a,i_a}^* = -\angle b_{a,(i_{a+1},i_a)}, \forall a = 2, \dots, A.$$

By substituting (D.3) into $\chi_{\text{Prop.}}$, the maximum SNR $\chi_{\text{Prop.}}$ can be expressed as

$$\chi_{\text{Prop.}} = P_{\max} / \sigma^2 \quad (\text{D.43})$$

$$\times \left| \prod_{a=2}^A \left(\sum_{i_a=1}^{N_{E,a}} |b_{a,(i_{a+1},i_a)}| \right) \sum_{i_1=1}^{N_{E,1}} |b_{1,(i_2,i_1)}| |f_{1,i_1}| \right|^2.$$

Since $\mathbf{f}_1 \sim \mathcal{CN}(\mathbf{0}, \zeta_{f_1}^2 \mathbf{I})$, $|f_{1,i_1}|$ follows the Rayleigh distribution with mean $\sqrt{\pi}\zeta_{f_1}/2$ and variance $(4 - \pi)\zeta_{f_1}^2/2$. By using the central limit theorem, we obtain $\sum_{i_1=1}^{N_{E,1}} |f_{1,i_1}| \sim \mathcal{CN}(N_{E,1}\sqrt{\pi}\zeta_{f_1}/2, N_{E,1}(4 - \pi)\zeta_{f_1}^2/2)$. Hence, we can transform (D.4) into (47). Equations (48) - (50) can be proved in similar way to proof above, and thus we omit the details here for brevity. The proof is complete. ■

REFERENCES

- [1] Y. Sun, K. An, Z. Lin, Y. Zhu, N. Al-Dhahir, and K.-K. Wong, "Scalable robust beamforming for multi-layer refracting RIS-assisted HAP-SWIPT networks," in *IEEE Global Communications Conference*, 2023, pp. 1656–1662.
- [2] M. S. Alam, G. K. Kurt, H. Yanikomeroglu, P. Zhu, and N. D. Dao, "High altitude platform station based super macro base station constellations," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 103–109, 2021.
- [3] X. Cao, P. Yang, M. Alzenad, X. Xi, D. Wu, and H. Yanikomeroglu, "Airborne communication networks: A survey," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 1907–1926, 2018.
- [4] Z. Lin, M. Lin, Y. Huang, T. d. Cola, and W.-P. Zhu, "Robust multi-objective beamforming for integrated satellite and high altitude platform network with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6384–6396, 2019.
- [5] Y. Alsaba, S. K. A. Rahim, and C. Y. Leow, "Beamforming in wireless energy harvesting communications systems: A survey," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 2, pp. 1329–1360, 2018.
- [6] Y. Chen, N. Zhao, and M.-S. Alouini, "Wireless energy harvesting using signals from multiple fading channels," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 5027–5039, 2017.
- [7] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, 2013.
- [8] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, "Artificial noise aided secure cognitive beamforming for cooperative MISO-NOMA using SWIPT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 918–931, 2018.
- [9] W. Feng, J. Tang, Y. Yu, J. Song, N. Zhao, G. Chen, K.-K. Wong, and J. Chambers, "UAV-enabled SWIPT in IoT networks for emergency communications," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 140–147, 2020.
- [10] X. Sun, W. Yang, Y. Cai, Z. Xiang, and X. Tang, "Secure transmissions in millimeter wave SWIPT UAV-based relay networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 785–788, 2019.
- [11] W. Wang, J. Tang, N. Zhao, X. Liu, X. Y. Zhang, Y. Chen, and Y. Qian, "Joint precoding optimization for secure SWIPT in UAV-aided NOMA networks," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5028–5040, 2020.
- [12] C. Wang and N. Liu, "Optimization and analysis of simultaneous wireless information and power transmission for LEO satellite," in *2018 IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, 2018, pp. 375–379.
- [13] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, 2020.
- [14] ——, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.
- [15] Z. Lin, H. Niu, K. An, and et al, "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 3717–3724, 2022.
- [16] A. Khalili, E. M. Monfared, S. Zargari, M. R. Javan, N. M. Yamchi, and E. A. Jorswieck, "Resource management for transmit power minimization in UAV-assisted RIS hetnets supported by dual connectivity," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1806–1822, 2022.
- [17] Y. Sun, K. An, J. Luo, Y. Zhu, G. Zheng, and S. Chatzinotas, "Intelligent reflecting surface enhanced secure transmission against both jamming and eavesdropping attacks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11 017–11 022, 2021.
- [18] L. You, J. Xiong, D. W. K. Ng, and et al, "Energy efficiency and spectral efficiency tradeoff in RIS-aided multiuser MIMO uplink transmission," *IEEE Trans. Signal Process.*, vol. 69, pp. 1407–1421, 2021.
- [19] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. Vincent Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Trans. Commun., Early Access*, 2022.
- [20] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, 2020.
- [21] S. Zargari, A. Hakimi, C. Tellambura, and S. Herath, "Multiuser MISO PS-SWIPT systems: Active or passive RIS?" *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1920–1924, 2022.
- [22] S. Zargari, S. Farahmand, B. Abolhassani, and C. Tellambura, "Robust active and passive beamformer design for IRS-aided downlink MISO PS-SWIPT with a nonlinear energy harvesting model," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 4, pp. 2027–2041, 2021.
- [23] S. Zargari, A. Hakimi, C. Tellambura, and S. Herath, "User scheduling and trajectory optimization for energy-efficient IRS-UAV networks with SWIPT," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 1815–1830, 2023.
- [24] H. Ren, Z. Zhang, Z. Peng, L. Li, and C. Pan, "Energy minimization in RIS-assisted UAV-enabled wireless power transfer systems," *IEEE Internet of Things Journal*, pp. 1–1, 2022.
- [25] Z. Chu, P. Xiao, D. Mi, W. Hao, Y. Xiao, and L.-L. Yang, "Multi-RIS assisted multi-cluster wireless powered IoT networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4712–4728, 2023.
- [26] Y. Sun, Y. Zhu, K. An, G. Zheng, S. Chatzinotas, K.-K. Wong, and P. Liu, "Robust design for RIS-assisted anti-jamming communications with imperfect angular information: A game-theoretic perspective," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7967–7972, 2022.
- [27] Y. Sun, K. An, Y. Zhu, G. Zheng, K.-K. Wong, S. Chatzinotas, H. Yin, and P. Liu, "RIS-assisted robust hybrid beamforming against simultaneous jamming and eavesdropping attacks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9212–9231, 2022.
- [28] Z. Yang, W. Xu, C. Huang, J. Shi, and M. Shikh-Bahaei, "Beamforming design for multiuser transmission through reconfigurable intelligent surface," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 589–601, 2021.
- [29] S. Alfattani, A. Yadav, H. Yanikomeroglu, and A. Yongacoglu, "Resource-efficient HAPS-RIS enabled beyond-cell communications," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2023.
- [30] K. Liu, Z. Zhang, L. Dai, and L. Hanzo, "Compact user-specific reconfigurable intelligent surfaces for uplink transmission," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 680–692, 2022.
- [31] Y. Sun, K. An, Y. Zhu, G. Zheng, K.-K. Wong, S. Chatzinotas, D. W. K. Ng, and D. Guan, "Energy-efficient hybrid beamforming for multilayer RIS-assisted secure integrated terrestrial-aerial networks," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4189–4210, 2022.
- [32] C. Liu and et al, "A programmable diffractive deep neural network based on a digital-coding metasurface array," *Nature Electro.*, vol. 5, no. 2, pp. 113–122, 2022.
- [33] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IIRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, 2022.
- [34] Y. Sun, K. An, J. Luo, and et al, "Outage constrained robust beamforming optimization for multiuser IRS-assisted anti-jamming communications with incomplete information," *IEEE Inter. Things J.*, vol. 9, no. 15, pp. 13 298–13 314, 2022.
- [35] E. Choi, M. Oh, J. Choi, J. Park, N. Lee, and N. Al-Dhahir, "Joint precoding and artificial noise design for MU-MIMO wiretap channels," *IEEE Trans. Commun.*, pp. 1–1, 2022.
- [36] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8627–8640, 2021.
- [37] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, 2020.
- [38] E. Bjoernson, "Optimizing a binary intelligent reflecting surface for OFDM communications under mutual coupling," in *International ITG Workshop on Smart Antennas*, 2021, pp. 1–6.
- [39] V. Jamali, A. M. Tulino, and et al, "Intelligent surface-aided transmitter architectures for millimeter-wave ultra massive MIMO systems," *IEEE Open J. Commun. Society*, vol. 2, pp. 144–167, 2021.
- [40] H. V. Cheng and W. Yu, "Degree-of-freedom of modulating information in the phases of reconfigurable intelligent surface," *IEEE Trans. Inf. Theory*, vol. 70, no. 1, pp. 170–188, 2024.
- [41] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, 2015.
- [42] R. Ma, W. Yang, X. Guan, X. Lu, Y. Song, and D. Chen, "Covert mmwave communications with finite blocklength against spatially random wardens," *IEEE Int. Things J.*, vol. 11, no. 2, pp. 3402–3416, 2024.
- [43] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, 2021.
- [44] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, 2020.
- [45] 3rd Generation Partnership Project, "Study on nr to support non-terrestrial networks," in *3rd Generation Partnership Project, Tech. Rep. 38.811*, 2017.
- [46] Q. Wu and R. Zhang, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 586–590, 2020.

- [47] M. M. Azari, F. Rosas, K.-C. Chen, and S. Pollin, "Ultra reliable UAV communication using altitude and cooperation diversity," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 330–344, 2018.
- [48] 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, 2018.
- [49] A. Arora, C. G. Tsinos, M. R. B. Shankar, S. Chatzinotas, and B. Ottersten, "Efficient algorithms for constant-modulus analog beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 756–771, 2022.
- [50] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, 2016.
- [51] V. Wong, R. Schober, D. W. K. Ng, and L. C. Wang, "Key technologies for 5G wireless systems," *Cambridge University Press*, 2017.
- [52] Z. Lin, M. Lin, J. Ouyang, W.-P. Zhu, A. D. Panagopoulos, and M.-S. Alouini, "Robust secure beamforming for multibeam satellite communication systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6202–6206, 2019.
- [53] W. Tang, J. Y. Dai, M. Z. Chen, K.-K. Wong, X. Li, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, "MIMO transmission through reconfigurable intelligent surface: System design, analysis, and implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2683–2699, 2020.
- [54] Y. Sun, Y. Zhu, K. An, Z. Lin, C. Li, D. W. K. Ng, and J. Wang, "Active-passive cascaded ris-aided receiver design for jamming nulling and signal enhancing," *IEEE Trans. Wireless Commun., Early Access*, 2023.
- [55] Z. Lin, M. Lin, J. Ouyang, W.-P. Zhu, and S. Chatzinotas, "Beamforming for secure wireless information and power transfer in terrestrial networks coexisting with satellite networks," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1166–1170, 2018.
- [56] Y. Cai, L.-H. Zhang, Z. Bai, and R.-C. Li, "On an eigenvector-dependent nonlinear eigenvalue problem," *SIAM J. Matrix Anal. Appl.*, vol. 39, no. 3, pp. 1360–1382, 2018.



Kang An received the B.E. degree in electronic engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2011, and PhD degree in communication engineering from Army Engineering University, Nanjing, China, in 2017. He is currently an associate professor with the Sixty-third Research Institute, National University of Defense Technology, Nanjing, China. His current research interests include reconfigurable intelligent surface, anti-jamming communications, satellite/aerial communications, physical-layer security, signal processing and machine learning for wireless communications.

He has published more than 100 peer-reviewed research papers in leading journals and flagship conferences and many of them are ESI highly cited papers. He was listed in the World's Top 2% Scientists identified by Stanford University in 2022 and 2023. He was a recipient of exemplary Reviewer for IEEE Transactions on Communications and IEEE Communications Letters in 2022. He was the recipient of the Outstanding Ph.D. Thesis Award of Chinese Institute of Command and Control in 2019. He is also serving as an Editor for Frontiers in Communications and Networks and Frontiers in Space Technologies. He was the corecipient of the Best Paper Awards at the IEEE IWCNC 2023 and the IEEE ICCT 2023.



Yifu Sun received the B.Eng. in Communications Engineering from National University of Defense Technology (NUDT), Changsha, China, in 2019, where he is currently pursuing the Ph.D. degree in Information and Communications Engineering with College of Electronic Science and Technology. His current research interests are in anti-jamming communications, reconfigurable intelligent surface, physical layer security, cooperative and cognitive communications, massive MIMO systems, and signal processing for wireless communications.



Zhi Lin received the B.E. and M.E. degrees in information and communication engineering from the PLA University of Science and Technology and the Ph.D. degree in electronic science and technology from the Army Engineering University of PLA, Nanjing, China, in 2013, 2016, and 2020, respectively. From March 2019 to June 2020, he was a visiting Ph.D. student with the Department of Electrical and Computer Engineering, McGill University, Montréal, Canada. Since February 2023, he is a Postdoctoral Fellow with the School of Computer Science and Engineering, Macau University of Science and Technology, Macau, China. Since January 2021, he has been with the College of Electronic Engineering, National University of Defense Technology, Hefei, China, where he is currently an Associate Professor.

Dr. Lin's research interests include array signal processing, physical layer security, reconfigurable intelligent surface and satellite-aerial-terrestrial integrated networks. He was the Symposium Co-Chair of IEEE WCSP'22 and TPC members of IEEE flagship conferences, including IEEE ICC, Globecom, Infocom, VTC, and so on. He was listed in the World's Top 2% Scientists identified by Stanford University in 2023. He was the recipient of the Outstanding Ph.D. Thesis Award of Chinese Institute of Electronics in 2022, the Macao Young Scholars Fellowship in 2022 and the Best Paper Award from IEEE IWCNC 2023 conference and the IEEE ICCT 2023. He has been serving as an Academic Editor for the *Wireless Communications and Mobile Computing* since 2023. He was also a Lead Guest Editor of the *IET Communications* Special Issues on Reconfigurable Intelligent Surfaces Aided Physical Layer Security in 6G Wireless Networks.



Yonggang Zhu received the B.S. degree in electronical engineering and Ph.D. degree in science of military equipment from PLA University of Science and Technology, Nanjing, China, in 2004, and 2009, respectively. Currently, he is a research associate professor with the Sixty-third Research Institute, National University of Defense Technology, Nanjing, China. His research interests include compressive sensing, statistical signal processing, reconfigurable intelligent surface, and anti-jamming communication.



Wanli Ni (Member, IEEE) received the B.Eng. and Ph.D. degrees in the School of Information and Communication Engineering from the Beijing University of Posts and Telecommunications (BUPT), China, in 2018 and 2023, respectively. Currently, he is a Postdoctoral Research Fellow in the Department of Electronic Engineering at Tsinghua University, China. His research interests include federated learning (FL), reconfigurable intelligent surface (RIS), over-the-air computation, semantic communication, performance analysis and optimization of machine learning in wireless networks.

He was a recipient of the Outstanding Doctoral Dissertation Awards from both China Education Society of Electronics (CESE) and BUPT in 2023, the Excellent Graduate of Beijing in 2023, the BUPT Top-10 Campus Pioneers on Academic in 2023, the National Scholarship in 2022 and 2021, and the Samsung Scholarship in 2019. He was a recipient of the Best Paper Award from the IEEE SAGC Conference in 2020. He was a recipient of the IEEE ComSoc Student Travel Grant from multiple international conferences including IEEE GLOBECOM, INFOCOM, and ICC. He was selected as an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2022, the IEEE COMMUNICATIONS LETTERS in 2022, and the IEEE WIRELESS COMMUNICATIONS LETTERS in 2021. He has served as a Guest Editor of the Electronic journal special issue "Edge Learning and Big AI Models in Wireless Communications and Networks".



Naofal Al-Dhahir (Fellow, IEEE) is Erik Jonsson Distinguished Professor & ECE Associate Head at UT-Dallas. He earned his PhD degree from Stanford University and was a principal member of technical staff at GE Research Center and AT&T Shannon Laboratory from 1994 to 2003. He is co-inventor of 43 issued patents, co-author of about 520 papers and co-recipient of 5 IEEE best paper awards. He is an IEEE Fellow, AAIA Fellow, received 2019 IEEE COMSOC SPCC technical recognition award, 2021 Qualcomm faculty award, and 2022 IEEE COMSOC RCC technical recognition award. He served as Editor-in-Chief of IEEE Transactions on Communications from Jan. 2016 to Dec. 2019. He is a Fellow of the US National Academy of Inventors and a Member of the European Academy of Sciences and Arts.

current research centers around 6G and beyond mobile communications. He is Fellow of IEEE and IET. He served as the Editor-in-Chief for IEEE Wireless Communications Letters between 2020 and 2023.



Kai-Kit Wong (Fellow, IEEE) received the BEng, the MPhil, and the PhD degrees, all in Electrical and Electronic Engineering, from the Hong Kong University of Science and Technology, Hong Kong, in 1996, 1998, and 2001, respectively. After graduation, he took up academic and research positions at the University of Hong Kong, Lucent Technologies, Bell-Labs, Holmdel, the Smart Antennas Research Group of Stanford University, and the University of Hull, UK. He is Chair in Wireless Communications at the Department of Electronic and Electrical Engineering, University College London, UK. His

current research centers around 6G and beyond mobile communications. He is Fellow of IEEE and IET. He served as the Editor-in-Chief for IEEE Wireless Communications Letters between 2020 and 2023.

Dusit Niyato (Fellow, IEEE) received the B.E. degree from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering and, by courtesy, the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. He has published more than 380 technical papers in the area of wireless and mobile networking. He is an inventor of four U.S. and German patents. He has authored four books including Game Theory in Wireless and Communication Networks: Theory, Models, and Applications (Cambridge University Press).

He won the Best Young Researcher Award of IEEE Communications Society (ComSoc) Asia Pacific (AP) and the 2011 IEEE Communications Society Fred W. Ellersick Prize Paper Award. He is serving as a Senior Editor for IEEE Wireless Communications Letters, an Area Editor for IEEE Transactions on Wireless Communications (Radio Management and Multiple Access) and IEEE Communications Surveys and Tutorials (Network and Service Management and Green Communication), an Editor for IEEE Transactions on Communications, and an Associate Editor for IEEE Transactions on Mobile Computing, IEEE Transactions on Vehicular Technology, and IEEE Transactions on Cognitive Communications and Networking. He was a Guest Editor of IEEE Journal on Selected Areas on Communications. He was a Distinguished Lecturer of the IEEE Communications Society from 2016 to 2017. He was named the 2017, 2018, and 2019 highly cited researcher in computer science.